

Algoritmos adaptativos con estadística de alto
orden. Identificación ciega de canales

TESIS DOCTORAL
por
Enrique Alameda Hernández

Departamento de Física Aplicada. Universidad de Granada

-
Directores: María del Carmen Carrión Pérez
Diego Pablo Ruiz Padillo

-
Febrero 2003

Dña. María del Carmen Carrión Pérez, Doctora en Ciencias Físicas y Profesora Titular del Departamento de Física Aplicada de la Facultad de Ciencias de la Universidad de Granada, y

D. Diego Pablo Ruiz Padillo, Doctor en Ciencias Físicas y Profesor Titular del Departamento de Física Aplicada de la Facultad de Ciencias de la Universidad de Granada,

MANIFIESTAN:

Que la presente Memoria titulada ‘Algoritmos adaptativos con estadística de alto orden. Identificación ciega de canales’, presentada por Enrique Alameda Hernández para optar al Grado de Doctor en Ciencias Físicas, ha sido realizada bajo nuestra dirección en la Facultad de Ciencias de la Universidad de Granada.

Con esta fecha autorizamos la presentación de la misma.

Granada, 21 de noviembre de 2002

Fdo: María del Carmen Carrión Pérez

Fdo: Diego Pablo Ruiz Padillo

Memoria presentada por Enrique Alameda Hernández para optar al Grado de Doctor en Ciencias Físicas por la Universidad de Granada.

Fdo: Enrique Alameda Hernández

Agradecimientos

En primer lugar me gustaría destacar que la realización de esta Tesis no hubiera sido posible sin la ayuda y el apoyo de mis Directores M. Carmen y Diego Pablo; de la mano de los cuales entré a formar parte del Grupo de Investigación Sistemas, Señales y Ondas del Departamento de Física Aplicada de la Universidad de Granada, en el curso académico 97/98. Desde entonces, y hasta ahora, no han dudado en dirigir mi tarea investigadora, sobre todo en los momentos de mayor dificultad y en los pasos más delicados. Hacia ellos va mi más sincero agradecimiento y reconocimiento por lo que han hecho por mí.

Durante cuatro años he tenido la enorme suerte de compartir el despacho con David, que no sólo me aventaja en edad sino en carácter y en otras muchas cualidades. Ha sido un estímulo en mi labor investigadora, me ha dado ánimos y me ha alegrado los días (sobre todo los lunes por la mañana cuando me contaba el fin de semana) y me ha resuelto la infinidad de dudas que han aparecido. David se merece un lugar destacado en estos agradecimientos.

También el ambiente del Departamento de Física Aplicada ha influido positivamente en la consecución de este Trabajo. Las visitas de Juan Antonio Morente al laboratorio para contarnos ‘batallitas’ de la Universidad; la orientación de Fernando Gonzalez; los desayunos con Pablo Hurtado; los descansos con mi primo Julián; las charlas con el resto de becarios, tanto recientes como Susana e Isabel (y esperemos que Pepe lo sea pronto), como los de mi misma promoción como Antonio, Juan, Ana Belén, Marisa,... Va hacia todos ellos mi gratitud por el ambiente tan sano que han creado y su gran disposición a ayudarme que siempre han mostrado.

En el plano personal hay que destacar el apoyo incondicional que toda mi familia me ha prestado durante la realización de esta Tesis. Les estoy profundamente agradecidos por sus esfuerzos y por su cariño a mis padres Enrique y Ángela y a mis hermanos Ángela, Belén y Pedro. Ellos han sido los verdaderos motores ‘invisibles’ de este trabajo.

La alegría y ánimo necesarios para realizar esta Tesis Doctoral me los ha infundido Belén. Los momentos de descanso junto a ella suponían no sólo retomar fuerzas físicas sino también psicológicas. Gracias por tu comprensión durante todos estos duros meses finales y por tu ‘insistencia’ de vez en cuando para que dejara de trabajar un poco antes.

Parte del trabajo aquí presentado se debe a la buena acogida que tuve por parte de Des en la Escuela de Ingeniería Electrónica y Eléctrica de la Universidad de Leeds. Durante tres meses dirigió la investigación, compartió sus ideas y se volcó para que mi estancia allí fuera lo más agradable posible. Gracias a su ayuda pude realizar el bloque tercero de esta memoria.

Part of the research presented here is owed to the warm welcome that Des gave me at the School of Electronic and Electrical Engineering at the University of Leeds. During three months he supervised my research, shared his ideas and made all within his power to make my stay as comfortable as possible. The third part of this thesis came to life thanks to his help.

Special Acknowledgement

I would like to thank the School of Electronic and Electrical Engineering of the University of Leeds, and in particular the Institute of Integrated Information Systems, for having hosted my stay from 14th February 2002 to 14th May 2002. The research carried out there under the supervision of Senior Lecturer Desmond C. McLernon is contained in this Thesis.

Título de Doctor con Mención Europea

Con el fin de obtener la Mención Europea en el Título de Doctor (aprobada en Junta de Gobierno de la Universidad de Granada el 5 de febrero de 2001), se han cumplido, en lo que atañe a esta Tesis Doctoral y a su Defensa, los siguientes requisitos:

1. Se ha incluido un amplio resumen en inglés al final de esta Memoria.
2. Uno de los miembros del tribunal, el Doctor Desmond C. McLernon, proviene de una universidad europea no española, en particular de la University of Leeds.
3. Una parte de la Defensa se ha realizado en inglés.
4. Una parte de esta Tesis Doctoral se ha realizado en el Reino Unido, en la University of Leeds.

Financiación

Las principales fuentes de financiación que han permitido llevar a cabo esta Tesis Doctoral han sido las siguientes:

- ◇ Beca de Formación de Personal Docente e Investigador de la Consejería de Educación y Ciencia de la Junta de Andalucía, en su orden del 17 de junio de 1998, BOJA num. 78 de 14 de julio. El periodo de disfrute de la misma ha sido desde el 1 de mayo de 1999 al 30 de abril de 2003.
- ◇ Proyecto denominado *‘Desarrollo de algoritmos de tratamiento de señal en comunicaciones. Aplicación a la identificación ciega de canales y el análisis tiempo frecuencia de su respuesta.’*, del Ministerio de Ciencia y Tecnología con código TIC 2001-2902, con una duración de 2001-2003.
- ◇ Ayuda de la Consejería de Educación y Ciencia de la Junta de Andalucía, para llevar a cabo una estancia en la Universidad de Texas en Austin durante marzo y abril del 2001, en la resolución 1/2001 de la Convocatoria para la adjudicación de Ayudas a la Investigación (Orden de 19 de diciembre de 2000, BOJA nº8 del 20 de enero de 2001).

Índice General

Notación	xxi
Introducción	xxiii
1 Fundamentos	1
1.1 Sistemas lineales y señales discretas	2
1.1.1 Estabilidad y causalidad	4
1.2 La transformada Z	5
1.2.1 Propiedades más útiles de la transformada Z	8
1.2.2 La función del sistema	9
1.2.3 Condición de fase mínima	10
1.3 Momentos y cumulantes	12
1.3.1 Propiedades de momentos y cumulantes	14
1.3.2 Sistemas LTI de fase no mínima	17
1.4 Identificación de sistemas	19
1.4.1 Modelos paramétricos	19
1.4.2 Modelos de media móvil MA	21
1.4.3 Modelos autorregresivos AR	23
1.4.4 Métodos para modelos ARMA	25
1.5 Algoritmos adaptativos	26
1.5.1 Algoritmos del cuadrado de media mínima	27
1.5.2 Algoritmos recursivos de mínimos cuadrados	28
1.5.3 Algoritmos adaptativos y estadística de alto orden. Ge- neralizaciones	29
1.6 Aplicación: Canales de comunicaciones	30
1.6.1 Comunicación digital	30
1.6.2 Conversión Analógico-Digital	30
1.6.3 Componentes de un sistema de comunicación digital	31
1.6.4 Formulación del problema	32

I Algoritmos adaptativos de mínimos cuadrados con estadística de alto orden. Análisis y aplicación a identifi-

cación ciega	35
Introducción	37
2 El algoritmo recursivo de mínimos cuadrados RLS	41
2.1 Motivación y obtención del algoritmo	42
2.2 Convergencia en ambientes estacionarios	45
2.2.1 Convergencia en media	45
2.2.2 La desviación en media cuadrática	46
2.2.3 Curva de aprendizaje	47
2.3 Convergencia en ambientes no estacionarios	49
2.3.1 Capacidad de rastreo	50
3 El algoritmo recursivo de variable instrumental RIV	53
3.1 Obtención del algoritmo	54
3.2 Convergencia en ambientes estacionarios	57
3.2.1 Convergencia en media. Efecto de las condiciones iniciales	57
3.2.2 Convergencia en media cuadrática con ruido de estimación	60
3.2.3 Convergencia en media cuadrática con ruido de estimación II	62
3.3 Convergencia en ambientes no estacionarios	63
3.3.1 Capacidad de rastreo	63
3.3.2 Convergencia en media cuadrática	64
3.3.3 Convergencia óptima	67
4 El algoritmo sobredeterminado recursivo de variable instru-	
mental ORIV	69
4.1 Obtención del algoritmo	70
4.2 Análisis de la convergencia para sistemas estacionarios	73
4.2.1 Convergencia en media	74
4.2.2 Error cuadrático medio	77
4.2.3 Error cuadrático medio en presencia de ruido	79
4.2.4 MSD en ambientes ruidosos II	81
4.3 Análisis de la convergencia para sistemas no estacionarios	83
4.3.1 Convergencia en media	83
4.3.2 Desviación en media cuadrática en ambientes no estacio-	
narios	85
4.4 Condiciones óptimas para el rastreo	87
4.5 Equivalencia de las aproximaciones utilizadas	88
5 Identificación ciega de sistemas MA	91
5.1 Hipótesis de partida para la identificación ciega	92
5.2 Identificación ciega mediante cumulantes	93
5.2.1 Obtención de ecuaciones mediante la transformada Z	94
5.2.2 Obtención de ecuaciones en el dominio del tiempo	96
5.2.3 Características de las ecuaciones obtenidas	96
5.3 Resolución mediante el algoritmo ORIV	97

5.3.1	Ergodicidad	98
5.3.2	Aplicando ORIV a ecualización ciega	100
5.4	Introducción al análisis experimental realizado	104
5.4.1	Aspectos generales de los modelos a identificar	104
5.4.2	Algunos conceptos útiles	104
5.5	Los métodos ORIV-C2C3 y ORIV-C3C4 en ambientes estacionarios ruidosos	106
5.5.1	Resultados de las simulaciones	107
5.5.2	Análisis de las simulaciones	115
5.6	Los métodos ORIV-C2C3 y ORIV-C3C4 en ambientes no estacionarios	120
5.6.1	Resultados de las simulaciones con variación temporal lineal	120
5.6.2	Análisis de las simulaciones con evolución temporal lineal	128
5.6.3	Resultados de las simulaciones con variación temporal aleatoria	132
5.6.4	Análisis de las simulaciones con evolución temporal aleatoria	135
5.7	Los métodos ORIV-C2C3 y ORIV-C3C4 en ambientes no estacionarios ruidosos	136
5.7.1	Resultados de las simulaciones	136
5.7.2	Análisis de las simulaciones	141
5.7.3	Resultado de las simulaciones relacionadas con el seguimiento límite	144
5.7.4	Análisis de los resultados obtenidos relacionados con el seguimiento límite	147
6	Estudio comparativo de los algoritmos RIV y ORIV	149
6.1	Comparación teórica	150
6.1.1	Cuadro comparativo	150
6.1.2	Análisis teórico de cada faceta. Ambientes estacionarios .	152
6.1.3	Análisis teórico de cada faceta. Ambientes no estacionarios	157
6.1.4	Conclusiones del análisis teórico	160
6.2	Comparación experimental	161
6.2.1	Comparación en ambientes estacionarios	162
6.2.2	Comparación en ambientes no estacionarios	164
 II Algoritmos adaptativos de mínimo error cuadrático medio con estadística de alto orden. Análisis y aplicación a identificación ciega		169
Introducción		171
7	El algoritmo del cuadrado de media mínima LMS	175
7.1	Estimación lineal de variables aleatorias. Filtrado Wiener	175
7.2	El algoritmo LMS	178
7.3	Convergencia de la media	179

7.4	Convergencia en media cuadrática	181
7.4.1	Convergencia de $D(n)$	181
7.4.2	Convergencia de $J(n)$	183
8	El Principio de Ortogonalidad Sobredeterminado y Generalizado. El algoritmo LMS generalizado	187
8.1	El Principio de Ortogonalidad Sobredeterminado y Generalizado	188
8.2	Convergencia en media de OGLMS	190
8.3	Algunas variantes de OGLMS	191
8.4	El algoritmo GLMS	193
8.5	Convergencia de la media	193
8.6	Convergencia en media cuadrática a orden μ^2	195
8.6.1	Convergencia en media cuadrática. Promediado directo	196
8.6.2	Convergencia en media cuadrática. Hipótesis de gaussianidad	198
8.6.3	Convergencia en media cuadrática. Tercera aproximación	200
8.7	Convergencia en media cuadrática a orden μ	201
8.8	Unificación de resultados	202
8.9	Evolución temporal de la desviación del error al cuadrado	204
8.10	Aplicación de GLMS	205
8.11	Aplicación numérica. Ejemplo	207
8.11.1	Presentación de los resultados. Evolución de MSD con el tiempo	207
8.11.2	Presentación de los resultados. Tiempo de convergencia	210
8.11.3	Presentación de los resultados. MSD en convergencia . . .	211
8.11.4	Comentario y discusión	213
8.11.5	Tiempo de convergencia	218
8.11.6	Error cuadrático medio en convergencia	221
8.11.7	Comentario final	225
9	Análisis de los algoritmos sobredeterminados y generalizados tipo LMS, OGLMS	227
9.1	Análisis del algoritmo de gradiente estocástico basado en la 1ª opción, OGLMS1	228
9.1.1	Convergencia en media. Ambientes estacionarios	229
9.1.2	Convergencia en media cuadrática. Ambientes estacionarios	231
9.1.3	Convergencia en media. Ambientes no estacionarios . . .	234
9.1.4	Ecuación para controlar la evolución de μ	236
9.1.5	Convergencia en media. Ambientes no estacionarios II . .	241
9.1.6	Convergencia en media cuadrática en ambientes no estacionarios	243
9.2	Algoritmo de gradiente estocástico basado en la 3ª opción, OGLMS3	244
9.2.1	Convergencia en media	244
9.2.2	Convergencia en media cuadrática. Ambientes estacionarios	245

10 El algoritmo promediado y generalizado de mínima media cuadrática, AOGMLS. Obtención y análisis	249
10.1 El algoritmo AOGMLS	251
10.2 Convergencia en media. Aproximación extrema	254
10.3 Convergencia en media cuadrática. Promediado directo	256
10.4 Convergencia en media. Segunda aproximación	257
10.5 Convergencia en media cuadrática. Segunda aproximación	261
10.6 Fundamentos de la segunda aproximación	264
10.6.1 En el cálculo de $E[\Delta \mathbf{w}(n)]$:	265
10.6.2 En el cálculo de $E[\Delta \mathbf{w}(n) \Delta \mathbf{w}^t(n)]$	267
10.7 Convergencia en media utilizando la segunda aproximación	274
10.8 Simulaciones	276
10.8.1 Comportamiento de AOGMLS cuando la matriz asociada no es definida	277
10.8.2 Identificación ciega de modelos MA en ambientes ruidosos mediante AOGMLS	278
10.8.3 Aplicando el algoritmo AOGMLS a ecualización ciega	283
10.8.4 Conclusión	283
11 Algunas versiones directas del algoritmo GLMS	285
11.1 El algoritmo GLMS cuadrático, CGLMS	286
11.1.1 Convergencia en media	287
11.1.2 Convegencia en media cuadrática	288
11.2 El algoritmo A1GLMS	289
11.2.1 Convergencia en media	290
11.2.2 Análisis de la convergencia en media cuadrática	291
11.3 El algoritmo PNGLMS	291
11.3.1 Convergencia en media.	293
11.3.2 Convergencia en media cuadrática	297
11.4 Simulaciones	299
11.4.1 Algoritmo CGLMS	299
11.4.2 Algoritmo A1GLMS	300
11.4.3 Algoritmo PNGLMS	300
11.5 Diagrama resumen	300
 III Identificación ciega con estadística de segundo orden. Métodos de elección de la fase	 305
Introducción	307
12 Sistemas de primer orden	309
12.1 Exposición del problema	310
12.2 Relación entre estadística de segundo orden y coeficientes del sistema	310
12.2.1 Características de la solución	312

12.3	Sensibilidad de las soluciones a pequeñas variaciones en la secuencia de autocorrelación	312
12.3.1	Sensibilidad del coeficiente h_0	312
12.3.2	Sensibilidad del coeficiente h_1	315
12.4	Localización de los ceros para problemas mal condicionados	317
13	Sistemas de segundo orden	319
13.1	Relación entre estadística de segundo orden y coeficientes del sistema	320
13.1.1	Discusión de la solución	321
13.1.2	Características de la solución	322
13.2	Sistemas espectralmente equivalentes	324
13.3	Sensibilidad de las soluciones	327
13.3.1	Sensibilidad del coeficiente h_1	327
13.3.2	Sensibilidad del coeficiente h_0	329
13.3.3	Sensibilidad del coeficiente h_2	334
13.4	Interpretación de las situaciones de sensibilidad	336
13.4.1	Interpretación para \hat{h}_1	336
13.4.2	Interpretación para \hat{h}_0	337
13.4.3	Interpretación para \hat{h}_2	338
13.4.4	Relación entre sensibilidad y degeneración de las soluciones	338
13.5	Sensibilidad de la localización de los ceros a pequeñas variaciones de los coeficientes del filtro	339
14	El método de identificación ciega de sistemas usando estadística de segundo orden y mínima de alto	341
14.1	Ilustrando el método para sistemas de primer orden	342
14.2	Aplicación de ICSMA a modelos de segundo orden	344
14.3	Sistemas de tercer orden	347
14.3.1	Una solución iterativa basada en ICSMA	348
15	Análisis de los métodos de elección de la fase	351
15.1	Acerca de la estimación de la secuencia de autocorrelación	352
15.2	Una magnitud nueva: Mínima Diferencia Relativa	354
15.2.1	MRD para sistemas espectralmente equivalentes	354
15.3	Descripción de los métodos estudiados	356
15.3.1	Técnicas basadas en el equalizador	357
15.3.2	Métodos directos	358
15.4	El comportamiento esperado de los métodos	359
15.5	Descripción del entorno y de los modelos utilizados	361
15.5.1	Características de los modelos	362
15.6	Resultados de las simulaciones	363
15.6.1	Primer grupo. Métodos AC y ACS _m	363
15.6.2	Segundo grupo. Comparación general	371

IV Summary in English 395

Introduction 397

16 Analysis of the Overdetermined Recursive Instrumental Variable ORIV algorithm. Applications to system identification 401

16.1	The ORIV algorithm	402
16.2	Convergence analysis for stationary systems	403
16.2.1	Convergence in the mean	403
16.2.2	Mean Square Deviation	406
16.2.3	MSD in noisy environments	408
16.2.4	MSD in noisy environments II	409
16.3	Convergence analysis for non-stationary systems	411
16.3.1	Convergence in the mean	411
16.3.2	Mean Square Deviation	413
16.4	Optimum tracking condition	415
16.5	Applying ORIV to blind system identification	416
16.5.1	Relations between cumulants and system coefficients . . .	416
16.5.2	Solving the equations using ORIV	417
16.5.3	Simulation results	419

17 The Averaged and Overdetermined Generalised LMS algorithm AOGLMS. A new algorithm 425

17.1	The Overdetermined and Generalised Orthogonality Principle . .	426
17.2	Some versions of the OGLMS algorithm	427
17.3	The GLMS algorithm	429
17.3.1	Convergence in the mean	429
17.4	Analysis of the stochastic gradient algorithm based on the 1st option, OGLMS1	431
17.4.1	Convergence in the mean for stationary systems	432
17.5	The Averaged, Overdetermined and Generalised LMS algorithm. A new algorithm	434
17.5.1	Convergence in the mean	435
17.5.2	Convergence in the mean square	437
17.5.3	Analysis of the simulations	440

18 Blind identification of FIR filters using second-order statistics. Phase selection methods 445

18.1	Problem statement	446
18.2	First order model	446
18.2.1	Characteristics of the solutions	447
18.3	Sensitivity of the solutions to small variations in the autocorrelation sequence estimate	448
18.3.1	Variations for the h_0 coefficient	448
18.4	Second order systems	450
18.4.1	Discussion of the solution	451

18.4.2	Characteristics of the solution	452
18.5	Sensitivity of the solution to the estimation of the autocorrelation sequence	453
18.6	Meaning of the sensitivity conditions	454
18.6.1	Meaning for \hat{h}_1	455
18.6.2	Meaning for \hat{h}_0	455
18.6.3	Meaning for \hat{h}_2	456
18.7	A new method. Identifying a first order channel	456
18.8	A new method. Identifying a second order channel	458
18.9	Analysis of the phase selection methods	461
18.9.1	On the estimation of the autocorrelation sequence	461
18.9.2	The minimum relative difference. A new magnitude	463
18.9.3	Description of the studied methods	464
18.9.4	Description of the models	465
18.9.5	Simulation results	466
	Conclusions	471
	Conclusiones	475
	Apéndices	485
	A Lema de inversión matricial	485
	B Aproximaciones de la matriz $\Gamma(n)$	487
	C Las hipótesis de independencia	489
	D Resolución de ecuaciones en diferencias de primer orden	491
	D.1 Coeficientes constantes	491
	D.1.1 Tiempos de convergencia	492
	D.2 Coeficientes no constantes. Caso 1	493
	D.3 Coeficientes no constantes. Caso 2	494
	E Obtención del algoritmo RIV a partir del ORIV	497
	E.1 Análisis del producto $\Gamma^{-1}(n)\mathbf{X}(n)\Lambda^{-1}(n)$	497
	E.2 Análisis de $\boldsymbol{\alpha}(n)$	498
	E.3 Recuperación del algoritmo RIV	498
	Bibliografía	499

Notación utilizada

Todas las variables están en cursiva. Además:

- a) Vectores en minúscula y negrita. Eje: $\mathbf{w}(n)$, $\mathbf{v}(n)$, $\mathbf{h}(n)$, etc.
- b) Matrices en mayúscula y negrita. Eje: $\mathbf{\Phi}(n)$, $\mathbf{K}(n)$, etc.
- c) Las componentes de un vector se dan entre corchetes $[\]$.
- d) La transpuesta se nota por el superíndice t .
- e) La transpuesta conjugada se nota por el superíndice H .
- f) Una norma convenientemente definida se nota por $\| \cdot \|$.
- g) Todos los vectores son columna.
- h) Las referencias a ecuaciones de otros capítulos: (capítulo.nº ec.).
- i) La parte real de un número complejo se nota por $\Re\{\cdot\}$.
- j) La parte imaginaria de un número complejo se nota por $\Im\{\cdot\}$.

Introducción

El objetivo en un problema de tratamiento de señal es procesar un número finito de muestras de datos y extraer la información que pueda estar ‘escondida’ en la misma. A esa serie de muestras se le denomina señal y puede venir generada por un sistema físico o fenómeno que se desea estudiar. Por ejemplo, el hombre, desde siempre ha convivido con fenómenos periódicos que ha estudiado y así ha podido construir calendarios, predecir mareas, fases de la luna, etc.

Quizás la estimación lineal fue uno de los primeros problemas que se desarrollaron dentro del campo del procesado de señal. El impulso principal lo recibió de los estudios astronómicos del movimiento de planetas y cometas a partir de datos tomados con el telescopio. Los comienzos de la teoría de estimación se pueden atribuir a Galileo Galilei en 1632, el cual intentó minimizar varias funciones de error. Sin embargo, el origen de la teoría de la estimación lineal se puede atribuir a Gauss quien, a la edad de 18 años en 1795, inventó el método de mínimos cuadrados para estudiar el movimiento de cuerpos celestes [Gau09]. Sin embargo, a comienzos del siglo XIX, había considerable controversia acerca del verdadero inventor del método de mínimos cuadrados. La controversia se desató porque Gauss no publicó su descubrimiento en 1795. Más bien, fue Legendre, que inventó independientemente el método, quien lo publicó en 1810 [Leg10].

Sin embargo, el problema que auspició la invención de una de las herramientas más útiles del procesado de señal, la transformada de Fourier, se estudiaba desde hacía muchos siglos antes. Fue Pitágoras el que primero estableció empíricamente una relación entre

la periodicidad de una vibración sinusoidal pura de notas musicales producida por una tensión constante y un número representativo de la longitud de las cuerdas. Daniel Bernoulli (1738) resolvió la ecuación de onda correspondiente a una cuerda musical vibrante, encontrando un desarrollo en serie de funciones armónicas. Euler, en 1755, encontró las expresiones para los coeficientes del desarrollo, que se llamarán posteriormente series de Fourier. Más tarde, el mismo Fourier, en sus estudios sobre el calor, extendió el desarrollo anterior para cualquier función incluso con un número finito de discontinuidades. Determinar los coeficientes de Fourier para una función dada se conoció como análisis armónico.

El uso de los coeficientes de Fourier disfrutó de gran fama durante bastantes años debido a la gran abundancia de fenómenos naturales que presentan algún tipo de periodicidad y a que estos coeficientes eran capaces de determinar una función continua a partir de un conjunto finito de parámetros. Sin embargo su utilidad era limitada en el caso de señales contaminadas por algún tipo de ruido o que no presentasen una periodicidad manifiesta. Casi en los comienzos del siglo XX, Schuster introdujo el periodograma para analizar el espectro de potencia de una serie temporal [Sch98]. El periodograma se define como el módulo cuadrado de la transformada discreta de Fourier de la serie temporal, y fue la única herramienta numérica disponible para el análisis espectral hasta los trabajos de Yule en 1927, estos trabajos consiguieron eliminar en parte las limitaciones comentadas del análisis basado en los coeficientes de Fourier.

Yule propuso su nuevo método a raíz de su estudio de periodicidades en series temporales con especial orientación al problema del número de manchas solares de Wolfer [Yul27]. De hecho, Yule creó un modelo estocástico con realimentación en el cual la muestra presente de la serie se asumía que consistía en una combinación lineal de muestras pasadas más un término de error. Este modelo se conoce como autorregresivo y el método de análisis espectral basado en tal modelo se conoce como análisis espectral autorregresivo.

Los primeros estudios acerca de la estimación por mínima media cuadrática de procesos estocásticos fueron llevados a cabo por Kolmogorov, Krein y Wiener durante el final de la década de 1930 y toda la década de 1940 [Kol39, Kre45, Wie49]. En 1947 Levinson

formula el problema del filtrado Wiener para series discretas, que toma una forma matricial. Para entradas estacionarias la matriz (de correlación) tiene una forma especial conocida como Toeplitz, así llamada en honor al matemático O. Toeplitz. Explotando las propiedades de una matriz Toeplitz, Levinson derivó un procedimiento recursivo muy elegante para resolver el problema [Lev47]. En 1960 Durbin redescubrió el procedimiento recursivo de Levinson como una manera de ajustar modelos autorregresivos a series temporales [Dur60]. Aquí se pone de manifiesto que herramientas originariamente empleadas en un campo del procesado de señal se reutilicen en otro, dando lugar a un enriquecimiento mutuo. Este enriquecimiento será mayor cuanto mayor sea el número de áreas que exploten los algoritmos del procesado de señal. Con la llegada de los ordenadores y demás componentes electrónicos digitales las aplicaciones de los métodos propios del tratamiento de señales se han disparado.

Hasta aproximadamente 1970 todo el procesado de señal se llevaba a cabo mediante equipamiento analógico. Las únicas excepciones, ya evidentes en 1950, fueron en campos donde las técnicas necesarias eran de una gran complejidad. Éste era el caso, por ejemplo, del análisis de datos geofísicos que se grababan en cintas para un posterior procesado. Este tipo de problemas son uno de los primeros ejemplos en los cuales se emplean computadoras digitales para el tratamiento de señales. De todas maneras no se podía llevar a cabo en tiempo real, ya que requería de minutos o incluso horas el poder procesar unos pocos segundos de cinta.

También por esa época se fue popularizando la simulación de sistemas de procesado de señal en ordenadores antes de implementarlo en hardware analógico. Esto permitía estudiar con detalle el sistema y ajustar los parámetros antes de emplear tiempo y recursos en la construcción del filtro analógico.

De todas maneras, debido a la lentitud de los ordenadores de aquella época, ninguna tarea se podía llevar a cabo en tiempo real y los ordenadores se consideraban sólo útiles como simuladores o aproximadores de componentes analógicos. Pero debido a esto, al extenderse el uso de ordenadores, se fueron creando algoritmos pensados exclusivamente para sistemas digitales, naciendo así el procesado

digital de señales, aunque como se ha indicado sólo se consideraban ideas interesantes pero imprácticas. Aportaciones de este tipo fueron el cepstrum y el filtrado homomórfico que ya no tenían equivalente en sistemas continuos y que dotaban al procesado digital de identidad propia.

El gran empujón para la reducción de la carga computacional de los algoritmos fue la invención de la transformada rápida de Fourier en 1965, ya que muchas tareas de procesado de señal requerían la estimación del espectro. Esto permitió el desarrollo de algoritmos digitales con un tiempo de ejecución que admitía la interacción con el sistema y muchos de los algoritmos que se habían considerado como poco prácticos empezaron a utilizarse. Dada la superioridad en algunos aspectos del tratamiento digital de señales con respecto al empleo de filtros analógicos, se empezaron a construir los primeros sistemas completamente digitales.

Con el tremendo avance experimentado en la velocidad de cálculo de los procesadores, que permite llevar a cabo tareas de procesado de señal cada vez más sofisticadas a un coste cada vez menor, la cantidad de campos donde se aplican las técnicas del procesado digital de señales incluyen a los clásicos del procesado de señal y amplía a otros nuevos, como pueden ser las comunicaciones digitales.

Ya se ha comentado que una de las herramientas más utilizadas en el tratamiento de señales es el análisis espectral. Este análisis consiste en determinar la energía asociada a cada componente armónica de la que está compuesta la señal. Normalmente se lleva a cabo mediante transformadas discretas de Fourier para series periódicas o transformadas Z si la serie no es periódica. Sin embargo no se puede proceder de esta manera para un tipo de señales importantísimo, para las señales aleatorias, que debido a su naturaleza requieren el uso de algún tipo de promedio al llevar a cabo el análisis. Como se recordará Schuster ya propuso promediar el periodograma, aunque existen definiciones alternativas al espectro en función de la secuencia de autocorrelación, que al ser una función determinista permite un análisis tal y como el descrito en las líneas anteriores.

La secuencia de autocorrelación da cuenta de la dependencia existente entre dos instantes temporales de una señal y se calcula como

el valor esperado del producto de ambos: promediando el valor esperado del producto de ambos para distintas realizaciones. El espectro de potencia se define como la transformada de Fourier de la secuencia de autocorrelación. La utilización del espectro de potencia y la función de autocorrelación ha sido la base de lo que se conoce como análisis espectral clásico. Un desarrollo detallado de este análisis espectral clásico basado en estadística de segundo orden se puede encontrar en numerosos libros y artículos como los escritos por S. L. Marple, M. S. Kay y otros [Kay88], [Mar87], [Gar88], [Chi78], [Kes86] y [Hay83].

En la estimación del espectro de potencia, el proceso bajo consideración se trata como una superposición de componentes armónicos estadísticamente decorrelacionados. A continuación se estima la distribución de energía entre estas componentes de frecuencia. Como tal, las relaciones de fase entre componentes de frecuencia se suprimen [HKGD71]. La información contenida en el espectro de potencia está contenida esencialmente en la secuencia de autocorrelación; esto sería suficiente para una descripción estadística completa de un proceso gaussiano de media conocida. Sin embargo, hay situaciones de interés donde se debería mirar más allá del dominio de la autocorrelación para obtener información acerca de la desviación respecto a la gaussianidad y acerca de la presencia de no-linealidades. El espectro de orden superior definido en términos de la estadística de orden superior, consistente en tomar valor esperado del producto de más de dos variables aleatorias, contiene esa información.

En general, hay tres motivaciones detrás del uso de la estadística de orden superior [NR87]. Estos son: en primer lugar, suprimir ruido gaussiano de características espectrales desconocidas en problemas de detección, estimación de parámetros y clasificación; segundo, reconstruir la respuesta en fase y en frecuencia de señales y sistemas; y tercero, detectar y caracterizar no linealidades en series temporales.

La primera motivación se basa en la propiedad de que solamente para procesos gaussianos, todos los cumulantes de orden mayor que dos son idénticamente nulos. Si una señal no gaussiana se recibe junto con ruido aditivo gaussiano, una transformación al dominio de los cumulantes de alto orden eliminará (en teoría) el ruido. Por tanto, en estas situaciones de procesamiento de señal, existirán ciertas

ventajas para detectar y/o estimar parámetros de señales a partir de cumulantes de los datos observados.

La segunda motivación se basa en el hecho de que los cumulantes y momentos de alto orden preservan la verdadera fase de las señales. En el modelado de series temporales en problemas de procesado de señal, se utiliza casi exclusivamente la estadística de segundo orden porque es normalmente el resultado de criterios de optimización de mínimos cuadrados. Sin embargo, el dominio de la autocorrelación suprime información de fase. Una reconstrucción adecuada de la fase en el dominio de la autocorrelación sólo se puede conseguir si la señal es de fase mínima. Por otro lado, reconstrucción de señales de fase no mínima se puede conseguir en el dominio de la estadística de alto orden debido a su capacidad de preservar tanto información de la magnitud como de la fase, sea mínima o no.

En tercer lugar, la introducción de la estadística de alto orden es bastante natural cuando se intenta analizar la no linealidad de un sistema operando bajo una entrada aleatoria. Se han estudiado con bastante detalle relaciones generales para datos aleatorios, estacionarios y arbitrarios que alimenten sistemas lineales arbitrarios, durante muchos años. En principio, la mayoría de las relaciones se basan en un criterio de ajuste de la autocorrelación. Por otro lado, no se dispone de relaciones generales para datos aleatorios, estacionarios y arbitrarios que alimenten sistemas no lineales arbitrarios. En su lugar, se ha investigado como un caso especial cada tipo de no linealidad. La estadística de alto orden puede tener un papel fundamental en la detección y caracterización del tipo de no linealidad, presente en un sistema, a partir de su salida.

El procesado de señal, en general, tiene una historia muy rica, y su importancia es evidente en campos tan diversos como ingeniería biomédica [Bru01], [CL93], acústica, sonar, radar [BRAC01], [BRAC02a], [BRAC02b], [RBAC99b], [RBAC00], [RBC99], seismología, comunicaciones de voz [OM00], comunicaciones de datos, tecnología nuclear y muchos otros. En muchas aplicaciones, como, por ejemplo en sistemas para la transmisión y reconocimiento de voz, se desea extraer algunos parámetros característicos. Alternativamente, se puede desear eliminar de la señal interferencias tales como ruido, o modificar la señal para presentarla en una forma en la cual

sea más fácilmente interpretada por un experto. Considérese otro ejemplo, una señal transmitida por un canal de comunicaciones que sufre perturbaciones por infinidad de mecanismos, como distorsión debida al canal, la inclusión de ruido de fondo y pérdida de la señal debido a interferencias destructivas. Uno de los objetivos del receptor es reducir al máximo los efectos de estas alteraciones. En cada caso se requiere el procesado de la señal.

Los problemas del procesado de señal no están restringidos, por supuesto, a señales unidimensionales. Muchas aplicaciones del procesado de imágenes requieren el empleo de técnicas de procesado de señal bidimensionales. Este es el caso del análisis de fotos aéreas para la detección de incendios forestales o daños en los cultivos, el análisis de fotos meteorológicas vía satélite y la potenciación de las transmisiones de televisión emitidas por sondas lunares y del espacio exterior. También utilizan técnicas de procesado de señal multidimensional el análisis de datos sísmicos como se requiere para explotación petrolífera, la medida de terremotos y el seguimiento de test nucleares.

En particular, en el área del procesado de señales mediante estadística de alto orden, se han publicado en los últimos 40 años innumerables artículos con aplicaciones en oceanografía, geofísica, radioastronomía, procesado de imágenes, mecánica de fluidos, series económicas temporales, física de plasmas, datos de manchas solares, etc.

Para llevar a cabo tales aplicaciones se han tenido que desarrollar técnicas que resolvieran una gran cantidad de problemas. Métodos para la estimación del espectro, tanto convencionales como el periodograma de Schuster [Sch98] o más ‘modernos’ como el modelado paramétrico autorregresivo de Yule [Yul27], junto con infinidad de otras técnicas desarrolladas al efecto. Formalización de los filtros homomórficos [Opp69] y el cepstrum [BHT63] como caso particular, junto con un estudio de todas sus posibles aplicaciones a reconstrucción de señales de fase no mínima y deconvolución [TJ99], entre otras. Técnicas basadas en estadística de alto orden, no paramétricas, para la reconstrucción de señales, para recuperar su fase, su magnitud o ambas [Bri77], [MU84], [Pet90] y [PN91]. Modelado paramétrico de una señal dada, tanto mediante modelado autorre-

gresivo AR [Tug87], de media móvil MA [Gia87], [ZZ94] o autorregresivo de media móvil ARMA [SM92], [ZSL96]; y sus innumerables aplicaciones a estimación espectral, identificación de sistemas, predicción, reconstrucción de series, etc. Técnicas para la estimación del tiempo de retraso a partir de medidas de dos sensores [ZR91], determinación de la dirección en la que se encuentra una fuente, estimación de parámetros de exponenciales amortiguadas, detección de señales transitorias [Hin90], ... Detección y caracterización de no linealidades en series temporales, como ya se comentó previamente, mediante el uso de estadística de alto orden y modelado de Volterra [Sch80] y [KP88], detección de acoplos de fase [RBAC99a], entre otros.

Toda la teoría desarrollada para resolver los problemas antes mencionados constituye toda una formulación matemática. Esta formulación matemática requiere de una implementación práctica que luego se aplicará a señales reales. La implementación práctica es sumamente importante porque puede dotar al método resultante de características que de por sí la formulación matemática no tenía. Supóngase que la estadística del problema cambia con el tiempo, por ejemplo porque los coeficientes que definen al sistema lineal no sean estacionarios. La llegada de un nuevo dato lleva información sobre cómo va variando dicha estadística y debe ser aprovechado para actualizar el conocimiento que hasta ese momento se tenía del sistema. Se necesita por tanto un método que se vaya adaptando a los cambios que se van produciendo y que no requiera resolver de nuevo todo el problema a la luz del nuevo dato; esta tarea la llevan a cabo los algoritmos adaptativos. La ecuación alrededor de la que giran el resto de ecuaciones en un algoritmo adaptativo es la de actualización de los parámetros incógnita del problema, que tiene la forma genérica siguiente:

$$\text{parámetros nuevos} = \text{parámetros antiguos} + \text{actualización}$$

donde la ‘actualización’ es función de los nuevos datos recogidos.

Existen dos familias principales de algoritmos adaptativos: los algoritmos de mínima media cuadrática (Least Mean-Square, LMS), pertenecientes al tipo de algoritmos de gradiente estocástico y los al-

goritmos recursivos de mínimos cuadrados (Recursive-Least-Squares, RLS). Los primeros trabajos acerca de los algoritmos adaptativos se pueden encontrar en los últimos años de la década de 1950, durante ese tiempo gran cantidad de investigadores estaban trabajando independientemente en distintas aplicaciones de filtros adaptativos. Desde esta fecha, el algoritmo LMS surgió como un método simple y a la vez efectivo para el funcionamiento de filtros transversales adaptativos. El algoritmo LMS fue diseñado por Widrow y Hoff en 1959 en sus estudios de un esquema de reconocimiento de patrones, conocido como elemento lineal adaptativo, conocido comunmente en la literatura como Adaline (Adaptive Linear Element) [WH60], [Wid70]. El LMS es un algoritmo de gradiente estocástico porque itera cada uno de los coeficientes del filtro transversal, en la dirección del gradiente del módulo cuadrado de una señal de error con respecto a los coeficientes. Como tal, el algoritmo LMS está íntimamente relacionado con el concepto de aproximación estocástica desarrollado por Robbins y Monro (1951) en estadística para resolver ciertos problemas de estimación secuencial de parámetros. La principal diferencia entre ellos es que el algoritmo LMS usa un parámetro del tamaño del paso fijo para controlar la corrección (actualización) aplicada a cada coeficiente de una iteración a otra, mientras que en métodos de aproximación estocástica el parámetro del tamaño del paso se hace inversamente proporcional al tiempo n o a una potencia de n .

En 1981, Zames introdujo la norma H^∞ (o criterio minimax) como un índice robusto de comportamiento para resolver problemas en estimación y control, y con él el campo de control robusto tomó una nueva dirección de investigación. En este contexto, merece especialmente mencionar que Hassibi et al. [HSK96] han demostrado que el LMS es óptimo bajo el criterio H^∞ . Así, por primera vez, se han presentado pruebas teóricas del comportamiento robusto del algoritmo LMS.

Volviendo a la familia de algoritmos RLS de algoritmos de filtrado adaptativo, el artículo original sobre el algoritmo RLS estándar parece ser el de Plackett [Pla50], aunque se debe decir que muchos otros investigadores han derivado y rederivado el algoritmo RLS. En 1974, Godard usó la teoría de Kalman de filtros para derivar una variante del RLS, que se conoce a veces en la literatura co-

mo algoritmo Godard [God74]. Aunque antes de esa fecha, varios investigadores habían aplicado la teoría de Kalman de filtros para resolver el problema del filtrado adaptativo, la aproximación de Godard fue ampliamente aceptada como la aplicación de mayor éxito de la teoría de Kalman de filtros por un periodo de dos décadas. Después, Sayed y Kailath [SK94], véase también [Joh95], publicaron un artículo en el cual se dibujó por primera vez la relación exacta entre el algoritmo RLS y la teoría de Kalman de filtros, estableciendo por tanto las bases de cómo explotar la amplísima literatura sobre filtrado Kalman en la resolución de problemas de filtrado lineal adaptativo.

El filtrado adaptativo se ha empleado con éxito en una gran cantidad de aplicaciones, destacando entre ellas la identificación de sistemas, modelado de las capas de la tierra, deconvolución, ecualización, ecualización ciega, modulación, estimación espectral, detección de señales, supresión de ruido, cancelación de eco, etc. Véase [Hay96] y referencias allí citadas.

Recapitulando lo expuesto hasta ahora se puede afirmar que si a la creciente velocidad de los microprocesadores se le une la flexibilidad de los algoritmos adaptativos, el abanico de posibles aplicaciones del procesamiento de señal se abre hasta límites amplísimos y de gran importancia actual. Piénsese en la necesidad de ecualización que presentan las comunicaciones móviles en nuestros días. Por ello, comprender en profundidad cómo funcionan los algoritmos adaptativos de los que se dispone es de una prioridad absoluta, por lo cual buena parte de esta Memoria analizará algunos algoritmos adaptativos que carecían de análisis y completará el análisis parcial que existía de otros.

Esta Memoria comienza en el capítulo 1 repasando algunos conceptos y métodos necesarios para el posterior desarrollo de la misma. En muchos casos supone una explicitación cuantitativa de lo expuesto en esta introducción: sistemas lineales y señales discretas que serán analizados mediante la transformada Z , en particular se tratará sobre la importantísima condición de fase mínima, momentos y cumulantes de alto orden con sus propiedades más útiles, técnicas paramétricas de identificación de sistemas, bases de los algoritmos adaptativos LMS y RLS y finalizará con una pequeña aplicación al

campo de las comunicaciones.

En la parte I se aborda el estudio de la familia de algoritmos RLS. El algoritmo RLS estándar ya se ha estudiado ampliamente en la bibliografía, su análisis se presenta en el capítulo 2 para familiarizarse con él y con las técnicas más usuales empleadas en su estudio. La adaptación del algoritmo RLS al empleo de estadística de orden superior se denomina algoritmo recursivo de variable instrumental (Recursive Instrumental Variable, RIV). Las propiedades de convergencia de este algoritmo fueron estudiadas por Swami [Swa96] para situaciones estacionarias, como aportación de esta Memoria se incluye un análisis de RIV para situaciones no estacionarias y el efecto que produce el ruido de estimación. Tanto el análisis de Swami como la aportación de esta Memoria se detallan en el capítulo 3 junto con la obtención del algoritmo. Sin embargo hay situaciones en las que la formulación matemática proporciona sistemas con más ecuaciones que incógnitas, en las cuales ni el algoritmo RLS ni el RIV pueden operar. Para solventar este problema se derivó el algoritmo RIV sobredeterminado, ORIV (Overdetermined Recursive Instrumental Variable). Este algoritmo se presenta en el capítulo 4 junto con un completo análisis de su convergencia, siendo dicho análisis aportación propia de esta Memoria. Dos consecuencias principales se pueden obtener de él: las condiciones bajo las cuales se produce un seguimiento óptimo de la evolución de los parámetros (tracking capabilities) y la obtención de una condición necesaria para convergencia (condición de ortogonalidad) válida para problemas sobredeterminados y que incluyan estadística de alto orden.

Dado que el algoritmo ORIV es el más completo de los analizados hasta ahora, en el capítulo 5 se aplicará al problema de la identificación ciega de sistemas. La identificación ciega de sistemas, como parte integrante del procesado de señal, se beneficia de una gran cantidad de métodos existentes en esta área como puede ser la estimación paramétrica del espectro, dicha estimación paramétrica se puede adaptar sin problemas para resolver la identificación de sistemas. Pero también se puede beneficiar de otros campos como el de la teoría de control, a través del modelado Kalman, p.e. Además, está íntimamente relacionado con el interesantísimo problema de la separación ciega de fuentes, con el que comparte aspectos comunes [Hay00], [Cru99], [CC99]. En este capítulo la identificación de sis-

temas se realizará mediante un modelado paramétrico MA, para lo cual en primer lugar se rederivarán en el dominio del tiempo, por un lado la conocida ecuación de Giannakis-Mendel y por otro se propone una generalización de ella empleando exclusivamente estadística de orden superior [ARBC02b], [ARBC02a]. Los resultados obtenidos mediante el algoritmo ORIV se explican satisfactoriamente a la luz del análisis llevado a cabo en el capítulo anterior. Finalmente, para acabar esta primera parte, en el capítulo 6 se compara minuciosamente el algoritmo RIV con el ORIV, tanto teóricamente como mediante simulaciones.

Como se aprecia, la familia RLS cubre una gran cantidad de problemas, tanto sobredeterminados como con estadística de alto orden. No ocurre lo mismo para la familia del LMS, donde sólo existen algoritmos adaptados al empleo de estadística de alto orden, como el LMS generalizado o GLMS, pero no hay ninguno capaz de resolver sistemas sobredeterminados ya sea con estadística de segundo orden o de alto orden. En la parte II se cubren estas deficiencias, con nuevas contribuciones. En primer lugar, tal y como se procedió en la parte de la familia RLS, aquí también se presenta en el capítulo 7 el algoritmo LMS junto con un análisis de su convergencia, por completitud y por presentar las técnicas e hipótesis más empleadas en el estudio de este tipo de algoritmos. El hilo conductor del resto de esta parte es el Principio de Ortogonalidad Sobredeterminado y Generalizado deducido en la parte I. De él se obtienen tres algoritmos distintos propuestos en esta Memoria. El primero de ellos, el GLMS, se rederiva dentro de este nuevo formalismo en el capítulo 8 y se analiza mediante teoría y simulaciones en profundidad, análisis no realizado anteriormente. En el capítulo 9 se analizan los otros dos algoritmos, los algoritmos Sobredeterminados y Generalizados tipo LMS, OGLMS (Overdetermined and Generalized LMS) 1 y 3. Se demuestra que el algoritmo OGLMS1 presenta mejores propiedades en convergencia que OGLMS3 y por tanto se selecciona para reducir su carga computacional y así asemejarse al típico LMS, dando lugar a un nuevo algoritmo. Este nuevo algoritmo, el algoritmo promediado OGLMS, AOG LMS (Averaged OGLMS) se analiza en el capítulo 10 junto con unas breves simulaciones que confirmen su buen comportamiento. Para finalizar esta segunda parte en el capítulo 11 se presentan nuevas versiones del algoritmo GLMS, con un pequeño

análisis de su convergencia tanto teórico como experimental.

La última parte de esta Memoria se centra exclusivamente en el problema de la identificación ciega, dejando a un lado los algoritmos adaptativos. El principal objetivo es beneficiarse todo lo posible de los métodos basados en estadística de segundo orden, que proporcionan estimadores menos ruidosos con menor cantidad de datos que los estimadores obtenidos con estadística de alto orden. Así, en los capítulos 12 y 13 se obtienen las expresiones explícitas que relacionan los coeficientes de la respuesta impulso de un sistema lineal, de orden 1 y 2 respectivamente, con la autocorrelación de la salida del mismo. Dado que de esta manera sólo se pueden obtener los sistemas que compartan el espectro de potencia, ya que estas técnicas no llevan información sobre la fase, en el capítulo 14 se propone un nuevo método que modifica ligeramente las expresiones obtenidas en los capítulos 12 y 13 de forma que incluya la información justa sobre la fase para que la solución tenga además la fase correcta. Este método contrasta con los métodos tradicionales, en los cuales se llevaba a cabo un proceso de búsqueda del sistema correcto dentro del conjunto de sistemas espectralmente equivalentes. Esta búsqueda a veces se lleva a cabo mediante un ajuste de los cumulantes reales (estimados) con los cumulantes de los sistemas candidatos. En el capítulo 15 se compara mediante simulaciones el nuevo método con estos métodos basados en ajustes por cumulantes (cumulant matching).

Para finalizar se presentan las conclusiones más relevantes obtenidas en la Memoria.

Capítulo 1

Fundamentos

Introducción

Este primer capítulo está dedicado a presentar los conceptos y técnicas que se emplearán en el resto de la Memoria. Su principal objetivo es explicar los fundamentos de cada una de las partes que constituyen el problema a tratar: la identificación ciega y adaptativa de sistemas lineales mediante estadística de alto orden. Por este motivo es necesario, en primer lugar, decir qué se entiende por sistemas lineales y enumerar sus propiedades. La caracterización de sistemas lineales cuenta con la transformada Z como una herramienta muy poderosa, en particular permite clasificar de forma instantánea a los sistemas como de fase mínima, máxima y mixta, con todas las implicaciones que eso conlleva para poder definir sistemas estables y causales. Dado que los sistemas estarán alimentados por procesos aleatorios es imprescindible contar con herramientas que permitan identificar su estadística; en este capítulo se presentan tanto los clásicos momentos de la distribución, como los menos clásicos cumulantes, incluyendo sus propiedades y utilidad a la hora de trabajar con sistemas de fase no mínima.

Los modelos paramétricos permiten definir un sistema lineal concreto a partir de un conjunto finito de parámetros. Los modelos paramétricos más usuales son los de media móvil MA, los autorregresivos AR y los autorregresivos de media móvil ARMA. Para cada uno de ellos existen algoritmos capaces de estimar los parámetros que definen a los modelos paramétricos a partir de la estadística de la entrada y la salida. En el caso de que sólo se tenga acceso a la salida del sistema se debe llevar a cabo lo que se conoce como identificación ciega. Además, la forma de resolver un problema concreto se puede implementar de forma que con la llegada de nueva información, una nueva muestra de la salida

del sistema por ejemplo, se actualicen los estimadores de los parámetros; esta tarea la realizan los algoritmos adaptativos.

Un área en el que la adaptatividad del algoritmo y la necesidad de llevar a cabo una identificación ciega (para una ecualización posterior) son ineludibles es en comunicaciones. Además, como se verá, el empleo de cumulantes en esta tarea reporta innumerables beneficios.

Este capítulo está estructurado de la siguiente manera. En el apartado 1.1 se introducen los sistemas lineales y sus propiedades, los cuales se estudian en el apartado 1.2 mediante la transformada Z. El apartado 1.3 define los momentos y cumulantes de procesos aleatorios y deriva sus propiedades. En particular los cumulantes serán la base para los algoritmos de identificación de sistemas, que se presentan en el apartado 1.4, para modelos paramétricos. Los algoritmos adaptativos se discuten en el apartado 1.5 y como punto final en el apartado 1.6 se describe el problema de la ecualización de un canal de comunicaciones como ejemplo en el que concurren todos los aspectos descritos a lo largo del capítulo.

1.1 Sistemas lineales y señales discretas

En la teoría de los sistemas discreto-temporales se está interesado en el procesamiento de señales que vienen representadas por series. Una serie de números x , en el cual el número n -ésimo de la serie se denota por $x(n)$, se escribe formalmente como

$$x = \{x(n)\}, \quad -\infty < n < \infty$$

Normalmente a $x(n)$ se le conoce como muestra n -ésima y por simplificar la notación se suele hablar de la serie $x(n)$. Además hay que tener en cuenta que n es un número entero, para números no enteros la serie no está definida.

Un ejemplo muy útil de serie es el de serie impulso discreta $\delta(n)$ o simplemente impulso:

$$\delta(n) = \begin{cases} 0 & n \neq 0 \\ 1 & n = 0 \end{cases}$$

su relevancia se verá más adelante.

La suma y el producto de series se definen muestra a muestra, a saber

$$x \cdot y = \{x(n)y(n)\}$$

$$x + y = \{x(n) + y(n)\}$$

Una secuencia y se dice que es una versión retrasada o trasladada de la serie x si y toma los valores

$$y(n) = x(n - n_0)$$

donde n_0 es un entero. Según estas definiciones, una serie arbitraria se puede expresar como suma de series impulsos:

$$x(n) = \sum_{k=-\infty}^{\infty} x(k)\delta(n-k)$$

Matemáticamente se define un sistema como un operador que relaciona una serie de entrada $x(n)$ con una serie de salida $y(n)$. Esto se nota por

$$y(n) = T[x(n)]$$

y se suele representar como aparece en la figura 1.1.

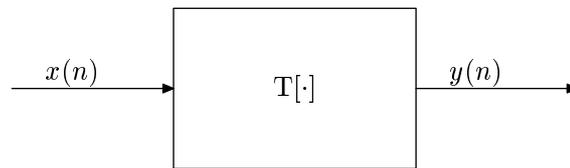


Figura 1.1 Representación gráfica usual de un sistema

Las clases de sistemas discretos se definen imponiendo restricciones a la transformación $T[\cdot]$. Dada su fácil caracterización matemática y porque se pueden diseñar de forma que lleven a cabo tareas de procesamiento útiles, la clase de sistemas lineales invariantes bajo traslaciones se va a estudiar en profundidad.

La clase de sistemas lineales se define a partir del principio de superposición. Si $y_1(n)$ e $y_2(n)$ son las respuestas cuando $x_1(n)$ y $x_2(n)$ son las entradas respectivas, el sistema es lineal si y sólo si

$$T[ax_1(n) + bx_2(n)] = aT[x_1(n)] + bT[x_2(n)] = ay_1(n) + by_2(n)$$

para cualesquiera escalares constantes arbitrarios a y b . Uniendo esta propiedad a la representación de una serie como combinación de impulsos retrasados se puede deducir que un sistema lineal se puede caracterizar completamente por su respuesta impulso. En particular, sea $h_k(n)$ la respuesta del sistema a $\delta(n-k)$, una muestra unitaria en el instante $n = k$. Por tanto,

$$y(n) = T \left[\sum_{k=-\infty}^{\infty} x(k)\delta(n-k) \right]$$

y por la linealidad de $T[\cdot]$

$$y(n) = \sum_{k=-\infty}^{\infty} x(k)T[\delta(n-k)] = \sum_{k=-\infty}^{\infty} x(k)h_k(n) \quad (1.1)$$

Por tanto, de acuerdo con (1.1) la respuesta del sistema se puede expresar en términos de la respuesta del sistema a $\delta(n - k)$. Si sólo se impone linealidad, $h_k(n)$ dependerá de n y k , en tal caso la utilidad computacional de la expresión (1.1) será limitada. Se obtiene una expresión mucho más útil si se impone la ligadura adicional de invarianza traslacional.

La clase de sistemas invariantes bajo traslaciones se caracteriza por la propiedad de que si $y(n)$ es la respuesta a $x(n)$, entonces $y(n - k)$ es la respuesta a $x(n - k)$, donde k es un entero negativo o positivo. Cuando el índice n está asociado al tiempo, invarianza traslacional se corresponde con invarianza temporal. La propiedad de invarianza traslacional implica que si $h(n)$ es la respuesta a $\delta(n)$, entonces la respuesta a $\delta(n - k)$ es simplemente $h(n - k)$. Por tanto, la expresión (1.1) se convierte en

$$y(n) = \sum_{k=-\infty}^{\infty} x(k)h(n - k) \quad (1.2)$$

En definitiva, cualquier sistema lineal invariante temporal está completamente caracterizado por su respuesta impulso $h(n)$.

A la ecuación (1.2) se le llama normalmente suma convolutiva. Si $y(n)$ es una serie cuyos valores están relacionados a los valores de dos secuencias $h(n)$ y $x(n)$ como indica la ecuación (1.2), se dice que $y(n)$ es la convolución de $x(n)$ con $h(n)$ y se nota por

$$y(n) = x(n) * h(n) \quad (1.3)$$

Esta nueva operación es conmutativa.

1.1.1 Estabilidad y causalidad

Se ha visto que las ligaduras de linealidad e invarianza temporal definen una clase de sistemas que vienen representados por la suma convolutiva. Las ligaduras adicionales de estabilidad y causalidad definen una clase de sistemas lineales invariantes temporales más restrictiva de importancia práctica.

Sistemas estables

Se define un sistema estable como uno para el cual toda entrada acotada produce una salida acotada. Los sistemas lineales invariante temporales (LTI) son estables si y sólo si:

$$S \equiv \sum_{k=-\infty}^{\infty} |h(k)| < \infty \quad (1.4)$$

Que se puede demostrar como sigue. Si la expresión (1.4) es cierta y x está acotada, es decir, $|x(n)| < M \forall n$, entonces a partir de la ecuación (1.2) se puede poner que

$$|y(n)| = \left| \sum_{k=-\infty}^{\infty} h(k)x(n-k) \right| \leq M \sum_{k=-\infty}^{\infty} |h(k)| < \infty$$

En consecuencia y está acotada. La implicación inversa se prueba mostrando que si $S = \infty$, entonces se puede encontrar una entrada acotada que generará una salida no acotada. Una serie de entrada tal tiene valores:

$$x(n) = \begin{cases} \frac{h^*(-n)}{|h(-n)|} & h(n) \neq 0 \\ 0 & h(n) = 0 \end{cases}$$

donde $h^*(n)$ es el complejo conjugado de $h(n)$. Claramente $x(n)$ está acotada. El valor de la salida para $n = 0$ es

$$y(0) = \sum_{k=-\infty}^{\infty} x(-k)h(k) = \sum_{k=-\infty}^{\infty} \frac{|h(k)|^2}{|h(k)|} = S$$

Y por tanto si $S = \infty$, la serie de salida no está acotada.

Sistemas causales

Un sistema causal es aquel para el cual la salida para cualquier $n = n_0$ depende sólo en la entrada para $n \leq n_0$. Un sistema LTI es causal si y sólo si la respuesta impulso es cero para $n < 0$.

1.2 La transformada Z

En la teoría de sistemas continuos en el tiempo la transformada de Laplace se puede considerar como una generalización de la transformada de Fourier. De manera similar es posible generalizar la transformada de Fourier de sistemas y señales discretas, resultando en lo que se conoce como transformada Z. La transformada Z juega un papel importante en el análisis y representación de sistemas discretos LTI.

La transformada Z $X(z)$ de una serie $x(n)$ se define como

$$X(z) = \sum_{n=-\infty}^{\infty} x(n)z^{-n} \quad (1.5)$$

donde z es una variable compleja. También es útil a veces notar la transformada Z de una serie $x(n)$ como $Z[x(n)]$.

Expresando la variable compleja z en forma polar como $z = re^{iw}$, la ecuación (1.5) se puede interpretar en términos de la transformada de Fourier. Con z expresado de esta manera (1.5) se convierte en:

$$X(re^{iw}) = \sum_{n=-\infty}^{\infty} x(n)r^{-n}e^{-iwn} \quad (1.6)$$

Por tanto, de acuerdo con esta expresión la transformada Z se puede interpretar como la transformada de Fourier de $x(n)$ multiplicada por una serie exponencial. Para $r = 1$, es decir, $|z| = 1$, la transformada Z es igual a la transformada de Fourier de la serie considerada.

Teniendo en cuenta que la convergencia de la transformada de Fourier exige que la serie sea absolutamente sumable, según la expresión (1.6), la transformada Z convergerá para aquellos valores complejos cuyos módulos r verifiquen que:

$$\sum_{n=-\infty}^{\infty} |x(n)r^{-n}| < \infty \quad (1.7)$$

Una clase importante de transformadas Z son aquellas para las cuales $X(z)$ es una función racional. Las raíces del polinomio del numerador son aquellos valores de z para los cuales $X(z) = 0$ y se les llama ceros de $X(z)$. Los valores de z para los cuales $X(z)$ es infinito se llaman polos de $X(z)$. Los polos de $X(z)$ para valores finitos de z son las raíces del polinomio del denominador. Además los polos pueden darse en $z = 0$ y $z = \infty$. Para las transformadas Z racionales hay un buen número de relaciones importantes entre la localización de los polos de $X(z)$ y la región de convergencia de la transformada.

Las propiedades de la serie $x(n)$ determinan la región de convergencia de $X(z)$. Para ver cómo se relacionan la región de convergencia y la serie, es conveniente considerar algunos casos especiales:

Series de longitud finita

Supóngase que sólo un número finito de valores de la serie no son nulos, de forma que

$$X(z) = \sum_{n=n_1}^{n_2} x(n)z^{-n} \quad (1.8)$$

donde n_1 y n_2 son enteros finitos. La convergencia de esta expresión requiere simplemente que $|x(n)| < \infty$ para $n_1 \leq n \leq n_2$. Entonces z puede tomar

cualquier valor excepto $z = \infty$ si $n_1 < 0$ y $z = 0$ si $n_2 > 0$. Por tanto, las series de longitud finita tienen una región de convergencia que comprende al menos $0 < |z| < \infty$, y puede incluir o $z = 0$ o $z = \infty$.

Series no nulas por la derecha

Una serie no nula por la derecha es aquella para la cual $x(n) = 0$ para $n < n_1$. La transformada Z de una serie tal es

$$X(z) = \sum_{n=n_1}^{\infty} x(n)z^{-n} \quad (1.9)$$

La región de convergencia de esta serie es el exterior de un círculo. Para probar que esto es cierto, supóngase que la serie es absolutamente convergente para $z = z_1$, de forma que

$$\sum_{n=n_1}^{\infty} |x(n)z_1^{-n}| < \infty \quad (1.10)$$

Si se considera la serie $\sum_{n=n_1}^{\infty} |x(n)z_1^{-n}| < \infty$, se desprende que si $n_1 > 0$, entonces para $|z| > |z_1|$, cada término es más pequeño que en la serie de la expresión (1.10), y por tanto

$$\sum_{n=n_1}^{\infty} |x(n)z^{-n}| < \infty \quad \text{para } |z| > |z_1|$$

En el caso de que $n_1 < 0$, la serie se puede expresar como

$$\sum_{n=n_1}^{\infty} |x(n)z^{-n}| = \sum_{n=n_1}^{-1} |x(n)z^{-n}| + \sum_{n=0}^{\infty} |x(n)z^{-n}| \quad (1.11)$$

La primera serie del miembro de la derecha de (1.11) es finita para cualquier valor finito de z . La segunda serie por el argumento anterior converge para $|z| > |z_1|$. Por tanto si R_{x-} es el valor más pequeño de $|z|$ para el cual la serie de la ecuación (1.9) converge, entonces la serie converge para

$$R_{x-} < |z|$$

con la excepción de $z = \infty$ si $n_1 < 0$. Por tanto si la serie es causal, la transformada Z convergerá también para $z = \infty$.

Series no nulas por la izquierda

Una serie no nula por la izquierda es una para la cual $x(n) = 0$ para $n > n_1$. Cambiando el índice de la sumatoria de $n = -m$ se puede concluir que los

resultados para series no nulas por la derecha se pueden aplicar a este caso con n reemplazado por $-n$ y con z reemplazado por z^{-1} . La región de convergencia se puede demostrar que es interior a un círculo $|z| < R_{x+}$ excepto $z = 0$ si $n_2 > 0$. Si la transformada Z de la serie no nula por la izquierda converge para $z = 0$, entonces la serie es nula para $n \leq 0$.

Series no nulas por izquierda y por la derecha

Las series no nulas por la izquierda y por la derecha se extienden desde $n = -\infty$ a $n = +\infty$. En general se puede escribir

$$\begin{aligned} X(z) &= \sum_{n=-\infty}^{\infty} x(n)z^{-n} = \\ &= \sum_{n=0}^{\infty} x(n)z^{-n} + \sum_{n=-\infty}^{-1} x(n)z^{-n} \end{aligned} \quad (1.12)$$

La primera serie es no nula por la derecha y converge para $R_{x-} < |z|$; la segunda serie es no nula por la izquierda y converge para $|z| < R_{x+}$. Si $R_{x-} < R_{x+}$, hay una región común de convergencia de la forma

$$R_{x-} < |z| < R_{x+} \quad (1.13)$$

Si $R_{x-} > R_{x+}$ no hay región común de convergencia, y por tanto la serie (1.12) no converge.

1.2.1 Propiedades más útiles de la transformada Z

A continuación se enumeran las propiedades de la transformada Z que serán de utilidad en el desarrollo del resto de capítulos de esta Memoria.

Región de convergencia de transformadas Z racionales

Como se ha comentado previamente, para una serie con una transformada Z racional la región de convergencia no puede contener ningún polo y se puede comprobar que está acotada por polos o por cero o infinito. En general la región de convergencia es una región conexa, de forma que dada una distribución de polos en el plano complejo son distintas las opciones que hay para elegir la zona de convergencia y por tanto la serie de la que se originó la transformada Z racional (serie no nula por la izquierda, por la derecha o por ambos lados).

Linealidad

Es sencillo comprobar que

$$\mathcal{Z}[ax(n) + by(n)] = aX(z) + bY(z)$$

y la región de convergencia es al menos la intersección de las regiones de convergencia de cada transformada individual.

Traslación de una serie

Considérese la serie $x(n)$ tal que

$$\mathcal{Z}[x(n)] = X(z), \quad R_{x-} < |z| < R_{x+}$$

Para una serie cuyos valores sean $x(n + n_0)$, se tiene

$$\mathcal{Z}[x(n + n_0)] = z^{n_0} X(z) \quad R_{x-} < |z| < R_{x+}$$

Las regiones de convergencia de $x(n)$ y $x(n + n_0)$ son idénticas, con la posible excepción de $z = 0$ o $z = \infty$.

Convolución de series

Si $w(n)$ es la convolución de dos series $x(n)$ e $y(n)$, entonces la transformada Z de $w(n)$ es el producto de las transformadas Z de $x(n)$ e $y(n)$. La región de convergencia se construye como la intersección de las regiones constituyentes.

1.2.2 La función del sistema

Un sistema LTI se puede describir en términos de la transformada Z de la respuesta impulso. Con $x(n)$, $y(n)$ y $h(n)$ denotando la entrada, salida y la respuesta impulso respectivamente y $X(z)$, $Y(z)$ y $H(z)$ sus transformadas Z, se sigue de los apartados anteriores que dado que

$$y(n) = h(n) * x(n)$$

entonces

$$Y(z) = X(z)H(z) \tag{1.14}$$

Como ocurre con la transformada de Fourier, la relación entrada-salida para un sistema LTI se corresponde a la multiplicación de la transformada Z de la entrada y de la respuesta impulso.

La transformada Z de la respuesta impulso se suele llamar función del sistema. La función del sistema evaluada en el círculo unidad es la respuesta en frecuencia del sistema.

Ya se ha mostrado que una condición necesaria y suficiente para que un sistema sea estable es que su respuesta impulso sea absolutamente sumable. La región de convergencia de la transformada Z está definida por aquellos valores de z para los cuales $h(n)z^{-n}$ es absolutamente sumable. Por tanto, si la región de convergencia de la función de un sistema incluye el círculo unidad, el sistema es estable y viceversa. Lo que es más, se puede afirmar que para un sistema causal y estable, la región de convergencia incluirá el círculo unidad y la totalidad del plano complejo fuera del círculo unidad, incluido $z = \infty$. Cuando el sistema se puede describir por una ecuación en diferencias con coeficientes constantes del tipo

$$\sum_{k=0}^p a_k y(n-k) = \sum_{r=0}^q b_r x(n-r)$$

la función del sistema es un función racional de la forma

$$H(z) = \frac{\sum_{r=0}^q b_r z^{-r}}{\sum_{k=0}^p a_k z^{-k}} \quad (1.15)$$

Como se comprueba fácilmente tomando transformada Z en la ecuación en diferencias y aplicando las propiedades ya establecidas de la transformada.

La ecuación (1.15) no indica la región de convergencia de la función del sistema. Para la función del sistema de la ecuación (1.15) existe una variedad de elecciones para la región de convergencia consecuentes con los requisitos de que sean regiones anulares acotadas por, pero no conteniendo a, polos. Para un cociente dado de polinomios, cada posible elección para la región de convergencia conducirá a una respuesta impulso diferente. Si se supone que el sistema es estable, entonces se debe elegir la región anular que incluya el círculo unidad. Si se supone que el sistema es causal, entonces se debe elegir la región de convergencia que sea exterior a un círculo que pase por el polo de $H(z)$ que esté más alejado del origen. Si el sistema es también estable, entonces, por supuesto, todos los polos deben quedar dentro del círculo unidad y la región de convergencia incluirá al círculo unidad. Por este motivo es conveniente, cuando se describen las funciones del sistema en términos de un diagrama ceros y polos en el plano complejo, dibujar el círculo unidad para que quede claro si los polos caen dentro o fuera de él.

1.2.3 Condición de fase mínima

La transformada Z de una serie causal se puede recuperar de su parte real o imaginaria en el círculo unidad. A continuación se va a analizar qué condiciones

deben de darse para poder recuperarla a partir de su magnitud o de su fase: dada la magnitud recuperar la fase o viceversa.

Supóngase que

$$H(z) = |H(z)|e^{i \arg[H(z)]}$$

por lo que tomando logaritmo complejo

$$\hat{H}(z) = \log[H(z)] = \log |H(z)| + i \arg[H(z)]$$

Se puede considerar que $\hat{H}(z)$ es la transformada Z de una serie $\hat{h}(n)$. Por tanto, $\log |H(z)|$ y $\arg[H(z)]$ serán transformados Hilbert el uno del otro si y sólo si $\hat{h}(n)$ es real, causal y estable.

$\hat{H}(z)$ diverge en los ceros y polos de $H(z)$. Como se desea que $\hat{H}(z)$ sea la transformada Z de una serie causal, estable y real, debe tener sus polos dentro del círculo unidad. Por tanto, $H(z)$ debe tener sus polos y ceros dentro del círculo unidad.

Supuesto cierto lo anterior ($\hat{h}(n)$ real, estable y causal), se puede escribir que la relación entre la magnitud y la fase de $H(z)$ es:

$$\begin{aligned} \log |H(e^{iw})| &= \hat{h}(0) - \frac{1}{2\pi} P \int_{-\pi}^{\pi} \arg[H(e^{i\theta})] \cot \left(\frac{\theta - w}{2} \right) d\theta \\ \arg[H(e^{i\theta})] &= \frac{1}{2\pi} P \int_{-\pi}^{\pi} \log |H(e^{iw})| \cot \left(\frac{\theta - w}{2} \right) d\theta \end{aligned} \quad (1.16)$$

donde P indica valor principal de Cauchy. Téngase en cuenta que sin el conocimiento de $\hat{h}(0)$, $|H(e^{iw})|$ viene dado salvo por una constante multiplicativa por $\arg[H(e^{iw})]$. Esta expresión se verifica dado que $\log |H(e^{iw})|$ y $\arg[H(e^{i\theta})]$ son una pareja de transformadas Hilbert, condición que se conoce como de fase mínima.

Para que $H(z)$ verifique la condición de fase mínima, no puede tener ni ceros ni polos fuera del círculo unidad. Por tanto tiene inversa causal y estable.

Un sistema se dirá de fase mínima cuando su respuesta en frecuencia sea de fase mínima. De igual modo, una serie se dirá de fase mínima si su transformada de Fourier es de fase mínima. Es importante recalcar que un sistema puede ser causal de fase no mínima. Sin embargo, todos los sistemas estables de fase mínima son causales. Un sistema estable y causal tiene los polos dentro del círculo unidad, los ceros donde sea.

Por tanto, dada la magnitud de la función del sistema su respuesta en fase está perfectamente determinada si el sistema es de fase mínima. Como cualquier reflexión de los ceros con respecto al círculo unidad conserva la respuesta en frecuencia pero altera la respuesta en fase, si el sistema no es de fase mínima se

tiene un amplio rango de posibilidades de respuesta en frecuencia compatibles. Para mostrar esto último se procede de la siguiente manera: Cualquier sistema se puede poner como la cascada de un sistema de fase mínima y un sistema pasa todo $H_{pt}(e^{iw})$, donde por definición $|H_{pt}(e^{iw})| = 1 \forall w$. Un pasa todo de primer orden se puede expresar como

$$H_{pt}(z) = \frac{z^{-1} - a^*}{1 - az - 1}$$

apareciendo los ceros y polos en parejas inversa conjugadas.

Supóngase un sistema de fase no mínima con, por ejemplo, un cero fuera del círculo unidad en $z = 1/z_0$ $|z_0| < 1$, y con el resto de polos y ceros dentro del círculo unidad. Entonces $H(z) = H_1(z)(z^{-1} - z_0)$ con $H_1(z)$ de fase mínima. Equivalentemente se puede escribir:

$$\begin{aligned} H(z) &= H_1(z)(z^{-1} - z_0) \frac{1 - z_0^* z^{-1}}{1 - z_0^* z^{-1}} = \\ &= H_1(z)(1 - z_0^* z^{-1}) \frac{z^{-1} - z_0}{1 - z_0^* z^{-1}} = H_{\min}(z) \frac{z^{-1} - z_0}{1 - z_0^* z^{-1}} \end{aligned} \quad (1.17)$$

$H_{\min}(z)$ se diferencia de $H(z)$ en que tiene el cero reflejado dentro del círculo unidad $1/z_0 \rightarrow z_0^*$. Este ejemplo se puede generalizar para abarcar sistemas generales de fase no mínima con funciones del sistema racionales. Se puede concluir, por tanto, que cualquier función del sistema $H(z)$ racional, que corresponda a un sistema causal, se puede poner de la forma

$$H(z) = H_{\min}(z)H_{pt}(z)$$

donde $H_{\min}(z)$ es de fase mínima y $H_{pt}(z)$ es un pasa todo.

Análogamente, de un sistema de fase mínima se puede obtener otro de fase no mínima, reflejando ceros fuera del círculo unidad.

1.3 Momentos y cumulantes

Hasta ahora no se ha dicho nada acerca de la naturaleza de las variables que forman la serie discreta. Dada la naturaleza de los problemas a tratar en esta Memoria se supondrá que las variables son aleatorias y por tanto se hace necesario definir magnitudes que caractericen su estadística: éstas son sus momentos y cumulantes.

Dado un conjunto de n variables aleatorias reales $\{x_1, x_2, \dots, x_n\}$, sus

momentos conjuntos de orden $r = k_1 + k_2 + \dots + k_n$ vienen dados por

$$\begin{aligned} \text{Mom}[x_1^{k_1}, x_2^{k_2}, \dots, x_n^{k_n}] &\equiv \text{E}[x_1^{k_1} x_2^{k_2} \dots x_n^{k_n}] = \\ &= (-i)^r \frac{\partial^r \Phi(w_1, w_2, \dots, w_n)}{\partial w_1^{k_1} \partial w_2^{k_2} \dots \partial w_n^{k_n}} \Bigg|_{w_1=w_2=\dots=w_n=0} \end{aligned} \quad (1.18)$$

donde $\Phi(w_1, w_2, \dots, w_n) \equiv \text{E}[\exp(i(w_1 x_1 + w_2 x_2 + \dots + w_n x_n))]$ es su función característica conjunta. $\text{E}[\cdot]$ denota el operador valor esperado. Por ejemplo, para dos variables aleatorias $\{x_1, x_2\}$ se tienen los momentos de orden dos: $\text{Mom}[x_1, x_2] = \text{E}[x_1 x_2]$, $\text{Mom}[x_1^2] = \text{E}[x_1^2]$ y $\text{Mom}[x_2^2] = \text{E}[x_2^2]$.

Otra forma de la función característica conjunta se define como el logaritmo natural de $\Phi(w_1, w_2, \dots, w_n)$, es decir,

$$\tilde{\Psi}(w_1, w_2, \dots, w_n) \equiv \ln(\Phi(w_1, w_2, \dots, w_n)) \quad (1.19)$$

Los cumulantes conjuntos, también llamados semi-invariantes, de orden r , $\text{Cum}[x_1^{k_1}, x_2^{k_2}, \dots, x_n^{k_n}]$, del mismo conjunto de variables aleatorias, se definen como los coeficientes de la expansión en serie de Taylor de la segunda función característica alrededor del cero,

$$\text{Cum}[x_1^{k_1}, x_2^{k_2}, \dots, x_n^{k_n}] \equiv (-i)^r \frac{\partial^r \tilde{\Psi}(w_1, w_2, \dots, w_n)}{\partial w_1^{k_1} \partial w_2^{k_2} \dots \partial w_n^{k_n}} \Bigg|_{w_1=w_2=\dots=w_n=0} \quad (1.20)$$

Por tanto, los cumulantes conjuntos se pueden expresar en términos de momentos conjuntos de un conjunto de variables aleatorias. Por ejemplo, los momentos

$$\begin{aligned} m_1 &= \text{Mom}[x_1] = \text{E}[x_1] & m_2 &= \text{Mom}[x_1, x_1] = \text{E}[x_1^2] \\ m_3 &= \text{Mom}[x_1, x_1, x_1] = \text{E}[x_1^3] & m_4 &= \text{Mom}[x_1, x_1, x_1, x_1] = \text{E}[x_1^4] \end{aligned}$$

de la variable aleatoria $\{x_1\}$ están relacionados con sus cumulantes por

$$\begin{aligned} c_1 &= \text{Cum}[x_1] = m_1 \\ c_2 &= \text{Cum}[x_1, x_1] = m_2 - m_1^2 \\ c_3 &= \text{Cum}[x_1, x_1, x_1] = m_3 - 3m_2 m_1 + 2m_1^3 \\ c_4 &= \text{Cum}[x_1, x_1, x_1, x_1] = m_4 - 4m_3 m_1 - 3m_2^2 + 12m_2 m_1^2 - 6m_1^4 \end{aligned} \quad (1.21)$$

Para variables de media nula $c_2 = m_2$, $c_3 = m_3$ y $c_4 = m_4 - 3m_2^2$.

Relación entre momentos y cumulantes

La relación general entre los momentos de $\{x_1, x_2, \dots, x_n\}$ y los cumulantes conjuntos $\text{Cum}[x_1, x_2, \dots, x_n]$ de orden $r = n$ está dada por

$$\text{Cum}[x_1, x_2, \dots, x_n] = \sum (-1)^{p-1} (p-1)! \mathbb{E} \left[\prod_{i \in s_1} x_i \right] \mathbb{E} \left[\prod_{i \in s_2} x_i \right] \dots \mathbb{E} \left[\prod_{i \in s_p} x_i \right] \quad (1.22)$$

donde los productorios se extienden sobre todas las particiones (s_1, s_2, \dots, s_p) , $p = 1, 2, \dots, n$ del conjunto de enteros $(1, 2, \dots, n)$. Por ejemplo, el conjunto de enteros $(1, 2, 3)$ se puede particionar en

$$\begin{array}{lll} p = 1 & s_1 = \{1, 2, 3\} & \\ p = 2 & s_1 = \{1\} & s_2 = \{2, 3\} \\ & s_1 = \{2\} & s_2 = \{1, 3\} \\ & s_1 = \{3\} & s_2 = \{1, 2\} \\ p = 3 & s_1 = \{1\} & s_2 = \{2\} \quad s_3 = \{3\} \end{array}$$

y por tanto (1.22) se convierte en

$$\begin{aligned} \text{Cum}[x_1, x_2, x_3] = & \mathbb{E}[x_1 x_2 x_3] - \mathbb{E}[x_1] \mathbb{E}[x_2 x_3] - \mathbb{E}[x_2] \mathbb{E}[x_1 x_3] - \\ & - \mathbb{E}[x_3] \mathbb{E}[x_1 x_2] + 2 \mathbb{E}[x_1] \mathbb{E}[x_2] \mathbb{E}[x_3] \end{aligned} \quad (1.23)$$

Claramente la expresión (1.23) es idéntica a c_3 de (1.21) para $x_1 = x_2 = x_3$. Por otro lado, para un conjunto de 4 variables aleatorias de media nula, procediendo de igual manera, se obtiene:

$$\begin{aligned} \text{Cum}[x_1, x_2, x_3, x_4] = & \mathbb{E}[x_1 x_2 x_3 x_4] - \mathbb{E}[x_1 x_2] \mathbb{E}[x_3 x_4] - \\ & - \mathbb{E}[x_1 x_3] \mathbb{E}[x_2 x_4] - \mathbb{E}[x_1 x_4] \mathbb{E}[x_2 x_3] \end{aligned} \quad (1.24)$$

La relación (1.22) implica que en el cálculo de cumulantes conjuntos de orden r se requiere el conocimiento de todos los momentos hasta orden r .

1.3.1 Propiedades de momentos y cumulantes

Las propiedades de momentos y cumulantes se pueden resumir como sigue:

1. $\text{Mom}[a_1 x_1, a_2 x_2, \dots, a_n x_n] = a_1 \dots a_n \text{Mom}[x_1, \dots, x_n]$ y $\text{Cum}[a_1 x_1, a_2 x_2, \dots, a_n x_n] = a_1 \dots a_n \text{Cum}[x_1, \dots, x_n]$ donde (a_1, a_2, \dots, a_n) son constantes. Esto se obtiene directamente de (1.18) y (1.22)
2. Los momentos y cumulantes son funciones simétricas de sus argumentos, por ejemplo, $\text{Cum}[x_1, x_2, x_3] = \text{Cum}[x_2, x_1, x_3] = \text{Cum}[x_3, x_2, x_1]$ y así sucesivamente.

3. Si las variables aleatorias $\{x_1, x_2, \dots, x_n\}$ se pueden dividir en dos o más grupos que son estadísticamente independientes, su cumulante n -ésimo es idénticamente cero, es decir, $\text{Cum}[x_1, x_2, \dots, x_n] = 0$ mientras que en general $\text{Mom}[x_1, x_2, \dots, x_n] \neq 0$. Por ejemplo, si los grupos independientes son $\{x_1, x_2, \dots, x_p\}$ y $\{x_{p+1}, x_{p+2}, \dots, x_n\}$, su función característica conjunta es $\Phi(w_1, w_2, \dots, w_n) = \Phi(w_1, \dots, w_p)\Phi(w_{p+1}, \dots, w_n)$. Por otro lado, su segunda función característica conjunta es $\tilde{\Psi}(w_1, w_2, \dots, w_n) = \tilde{\Psi}(w_1, \dots, w_p) + \tilde{\Psi}(w_{p+1}, \dots, w_n)$. La prueba de esta propiedad se puede continuar fácilmente si se sustituye $\tilde{\Psi}(w_1, w_2, \dots, w_n)$ y $\Phi(w_1, w_2, \dots, w_n)$ en (1.20) y (1.18) respectivamente.
4. Si los conjuntos de variables aleatorias $\{x_1, x_2, \dots, x_n\}$ y $\{y_1, y_2, \dots, y_n\}$ son independientes, entonces, $\text{Cum}[x_1 + y_1, x_2 + y_2, \dots, x_n + y_n] = \text{Cum}[x_1, x_2, \dots, x_n] + \text{Cum}[y_1, y_2, \dots, y_n]$ donde en general esto mismo no se cumple para los momentos.
Sin embargo, para el conjunto de variables aleatorias $\{y_1, x_1, \dots, x_n\}$ se tiene que $\text{Cum}[x_1 + y_1, x_2, \dots, x_n] = \text{Cum}[x_1, x_2, \dots, x_n] + \text{Cum}[y_1, x_2, \dots, x_n]$ y $\text{Mom}[x_1 + y_1, x_2, \dots, x_n] = \text{Mom}[x_1, x_2, \dots, x_n] + \text{Mom}[y_1, x_2, \dots, x_n]$.
5. Si el conjunto de variables aleatorias es conjuntamente gaussiano, entonces toda la información sobre su distribución está contenida en los momentos de orden menor que dos. Por tanto, todos los momentos de orden mayor que dos no tienen información nueva que suministrar. Esto conduce al hecho de que todos los cumulantes conjuntos de orden mayor que dos son idénticamente nulos para vectores aleatorios gaussianos. En cierto sentido, los cumulantes de orden mayor que dos miden la no gaussianidad de una serie temporal.

Momentos y cumulantes de procesos estacionarios

Si $\{x(k)\}$ con $k = 0, \pm 1, \pm 2, \dots$ es un proceso real aleatorio y estacionario y sus momentos hasta orden n existen, entonces

$$\text{Mom}[x(k), x(k + \tau_1), \dots, x(k + \tau_{n-1})] = \text{E}[x(k)x(k + \tau_1)\dots x(k + \tau_{n-1})]$$

sólo dependerá de las diferencias temporales $\tau_1, \tau_2, \dots, \tau_{n-1}$ donde τ_i es entero $\forall i$. Por tanto, los momentos de un proceso aleatorio estacionario se escriben como:

$$m_n^x(\tau_1, \tau_2, \dots, \tau_{n-1}) \equiv \text{E}[x(k)x(k + \tau_1)\dots x(k + \tau_{n-1})] \quad (1.25)$$

De manera similar, los cumulantes de orden n -ésimo de $\{x(k)\}$ son funciones de dimensión $n - 1$ que se escriben en la forma:

$$c_n^x(\tau_1, \tau_2, \dots, \tau_{n-1}) \equiv \text{Cum}[x(k), x(k + \tau_1), \dots, x(k + \tau_{n-1})] \quad (1.26)$$

Combinando las expresiones (1.22), (1.25) y (1.26) se obtienen las siguientes relaciones entre las secuencias de momentos y cumulantes de $\{x(k)\}$, en particular para variables de media nula. Los cumulantes de orden 2 y 3 son idénticos a los momentos de orden 2 y 3 respectivamente. Sin embargo, para generar el cumulante de cuarto orden, se necesitan conocer los momentos de orden 4 y 2, en particular:

$$\begin{aligned} c_4^x(\tau_1, \tau_2, \tau_3) = & m_4^x(\tau_1, \tau_2, \tau_3) - m_2^x(\tau_1)m_2^x(\tau_3 - \tau_2) - \\ & - m_2^x(\tau_2)m_2^x(\tau_3 - \tau_1) - m_2^x(\tau_3)m_2^x(\tau_2 - \tau_1) \end{aligned} \quad (1.27)$$

Varianza, factor de asimetría y curtosis

Los siguientes cumulantes (supuestas variables de media nula) poseen nombres especiales

$$\begin{aligned} \gamma_2^x = E[x^2(k)] &= c_2^x(0) \quad \text{varianza} \\ \gamma_3^x = E[x^3(k)] &= c_3^x(0, 0) \quad \text{factor de asimetría} \\ \gamma_4^x = E[x^4(k)] - 3[\gamma_2^x]^2 &= c_4^x(0, 0, 0) \quad \text{curtosis} \end{aligned} \quad (1.28)$$

La curtosis normalizada se define como $\gamma_4^x/[\gamma_2^x]^2$.

Procesos de ruido blanco no gaussianos

Si $\{w(k)\}$ es un proceso estacionario no gaussiano con $E[w(k)] = 0$ y con una secuencia de cumulantes de orden n dada por

$$\begin{aligned} c_n^w(\tau_1, \tau_2, \dots, \tau_{n-1}) = & \text{Cum}[w(k), w(k + \tau_1), \dots, w(k + \tau_{n-1})] = \\ & = \gamma_n^w \delta(\tau_1, \tau_2, \dots, \tau_{n-1}) \end{aligned} \quad (1.29)$$

donde γ_n^w es una constante y $\delta(\tau_1, \tau_2, \dots, \tau_{n-1})$ es la función delta de Kronecker de dimensión $n - 1$, entonces se dice que $\{w(k)\}$ es blanco de orden n .

Procesos complejos

Si un proceso dado $\{x(k)\}$ es complejo, su secuencia de cumulantes de orden n tiene más de una definición dependiendo dónde se coloca el operador *conjugado* '*'. Por ejemplo, la secuencia de cumulantes de tercer orden se puede definir como:

$$\begin{aligned} c_3^{(1)}(\tau_1, \tau_2) &\equiv \text{Cum}[x(k), x(k + \tau_1), x(k + \tau_2)] \\ c_3^{(2)}(\tau_1, \tau_2) &\equiv \text{Cum}[x(k), x^*(k + \tau_1), x(k + \tau_2)] \\ c_3^{(3)}(\tau_1, \tau_2) &\equiv \text{Cum}[x(k), x^*(k + \tau_1), x^*(k + \tau_2)] \end{aligned} \quad (1.30)$$

etc. En general hay 2^n definiciones diferentes de cumulantes complejos.

Cumulantes de la salida de sistemas LTI

Supóngase que un sistema LTI con respuesta impulso $\{h(k)\}$ $k = -L_2, \dots, L_1$ está excitado por una serie compuesta por variables aleatorias independientes e idénticamente distribuidas (i.i.d) $\{x(k)\}$ $i = 1, 2, \dots$. Entonces se verifica la siguiente relación entre los cumulantes de la salida del sistema $\{y(k)\}$ y la serie i.i.d de entrada:

$$c_n^y(\tau_1, \dots, \tau_{n-1}) = \gamma_n^x \sum_{k=-L_2}^{L_1} h(k)h(k + \tau_1)\dots h(k + \tau_{n-1})$$

Esta expresión se obtiene fácilmente empleando la propiedad de linealidad de los cumulantes y constituye una generalización para la conocida relación para la estadística de segundo orden.

1.3.2 Sistemas LTI de fase no mínima

Lo que hace a los momentos y cumulantes de orden mayor que 2 útiles en la identificación de sistemas de fase no mínima es su habilidad para conservar información sobre la fase (excepto, claro está, por el término de fase lineal). Esta propiedad se muestra mediante un ejemplo sencillo. Considérese un sistema de respuesta impulso finita (FIR) de segundo orden, excitado por ruido blanco no gaussiano de media nula $\{w(k)\}$. Dependiendo de la localización de los ceros del sistema respecto al círculo unidad, pueden ocurrir los siguientes casos:

Sistema de fase mínima

En este caso la salida del sistema viene dada por

$$y_1(k) = w(k) - (a + b)w(k - 1) + abw(k - 2)$$

y por tanto la función del sistema es

$$H_1(z) = (1 - az^{-1})(1 - bz^{-1})$$

donde a y b son constantes con $0 < a < 1$ y $0 < b < 1$. Este es el caso en el que ambos ceros caen dentro del círculo unidad.

Sistemas de fase máxima

Ahora se tiene que

$$y_2(k) = w(k) - (a+b)w(k+1) + abw(k+2)$$

con su función del sistema correspondiente

$$H_2(z) = (1 - az)(1 - bz)$$

Ambos ceros se encuentran fuera del círculo unidad en $1/a$ y $1/b$.

Sistemas de fase no mínima o fase mixta

$$y_3(k) = -aw(k+1) + (1+ab)w(k) - bw(k-1)$$

o

$$H_3(z) = (1 - az)(1 - bz^{-1})$$

donde uno de los ceros cae dentro del círculo unidad b y el otro fuera $1/a$.

Las tres series de salida $\{y_1(k)\}$, $\{y_2(k)\}$ y $\{y_3(k)\}$ tienen idéntica secuencia de autocorrelación dadas por

$$\begin{aligned} c_2^y(0) &= 1 + a^2b^2 + (a+b)^2 \\ c_2^y(1) &= -(a+b)(1+ab) \\ c_2^y(2) &= ab \\ c_2^y(\tau) &= 0 \text{ para } \tau > 2 \end{aligned} \tag{1.31}$$

lo que implica que tienen idéntico espectro de potencia:

$$C_2^y(w) = |Y(z)|^2 = \gamma_2^w |H_1(z)|^2 = \gamma_2^w |H_2(z)|^2 = \gamma_2^w |H_3(z)|^2$$

para $z = \exp iw$. Los sistemas $H_1(z)$, $H_2(z)$ y $H_3(z)$ son por tanto espectralmente equivalentes.

Este es un resultado esperado porque el módulo cuadrado de la función de transferencia de un sistema LTI no diferencia entre un cero z_0 de su recíproco conjugado $1/z_0^*$. Por otro lado, las series de salida $\{y_1(k)\}$, $\{y_2(k)\}$ e $\{y_3(k)\}$ tienen diferentes cumulantes de orden n -ésimo. Las tablas 1.3 y 1.4 muestran los cumulantes de tercer orden de cada una de las series de salida.

	Fase mínima	Fase máxima
$H(z)$	$(1 - az^{-1})(1 - bz^{-1})$	$(1 - az)(1 - bz)$
$c_2^y(0, 0)$	$1 - (a+b)^2 + (ab)^3$	$a^3b^3 - (a+b)^3 + 1$

$c_2^y(1, 1)$	$(a + b)^2 - (a + b)a^2b^2$	$ab(a + b)^2 - (a + b)$
$c_2^y(2, 2)$	a^2b^2	ab
$c_2^y(1, 0)$	$-(a + b) + (a + b)^2ab$	$-a^2b^2(a + b) + (a + b)^2$
$c_2^y(2, 0)$	ab	a^2b^2
$c_2^y(2, 1)$	$-(a + b)ab$	$-ab(a + b)$

Tabla 1.3 *Cumulantes de tercer orden de sistemas espectralmente equivalentes. 1ª parte.*

	Fase no mínima I	Fase no mínima II
$H(z)$	$(1 - az)(1 - bz^{-1})$	$(1 - az^{-1})(1 - bz)$
$c_2^y(0, 0)$	$-a^3 + (1 + ab)^3 - b^3$	$-b^3 + (1 + ab)^3 - a^3$
$c_2^y(1, 1)$	$-a(1 + ab)^2 + (1 + ab)b^2$	$-b(1 + ab)^2 + (1 + ab)a^2$
$c_2^y(2, 2)$	$-ab^2$	$-ba^2$
$c_2^y(1, 0)$	$a^2(1 + ab) - (1 + ab)^2b$	$b^2(1 + ab) - (1 + ab)a$
$c_2^y(2, 0)$	$a^2(-b)$	$-b^2a$
$c_2^y(2, 1)$	$a(1 + ab)b$	$ab(1 + ab)$

Tabla 1.4 *Cumulantes de tercer orden de sistemas espectralmente equivalentes. 2ª parte.*

1.4 Identificación de sistemas

Identificar un sistema es determinar las propiedades del mismo que sean de interés para el problema concreto a tratar, ya sea la respuesta en fase, en frecuencia, etc. Existen innumerables métodos para llevar a cabo este cometido, tales como los métodos clásicos de estimación del espectro, métodos paramétricos con estadística de segundo orden, métodos paramétricos con estadística de alto orden, etc. En esta Memoria se van a emplear profusamente los métodos paramétricos con estadística de alto orden, por tanto se introducen a continuación.

1.4.1 Modelos paramétricos

Una de las técnicas más empleadas en la identificación de sistemas ha sido la construcción de un modelo LTI excitado mediante ruido blanco, a partir

de una realización de una señal aleatoria. Si el ruido de excitación se supone gaussiano, un sistema de fase no mínima se identificará como fase mínima cuando la entrada es inaccesible. Por tanto, la principal motivación detrás del hecho de usar modelos paramétricos excitados por ruido blanco gaussiano para la identificación de sistemas es recuperar tanto la respuesta en fase como la respuesta en frecuencia del sistema.

Considérese el proceso ARMA $\{x(k)\}$ real y estable definido por

$$\sum_{i=0}^p a(i)x(k-i) = \sum_{j=0}^q b(j)w(k-j) \quad (1.32)$$

donde $\{w(k)\}$ son variables aleatorias i.i.d, independientes de $\{x(m)\}$ para $m < k$, con media cero y

$$\text{Cum}[w(k), w(k+\tau_1), \dots, w(k+\tau_{n-1})] = \gamma_n^w \delta(\tau_1, \dots, \tau_{n-1}) \quad (1.33)$$

es decir, $\{w(k)\}$ es blanco de orden n . Nótese que tanto $\{w(k)\}$ como $\{x(k)\}$ son no gaussianos.

Para problemas clásicos de identificación de sistemas, se usa casi exclusivamente la estimación por mínimos cuadrados porque dan estimadores de máxima verosimilitud de los parámetros de sistemas excitados por ruido blanco gaussiano y porque las ecuaciones que se obtienen están en forma lineal conteniendo valores de la secuencia de autocorrelación. Sin embargo, los métodos de mínimos cuadrados no pueden identificar correctamente un sistema de fase no mínima a partir de su salida exclusivamente ya que se basan en medidas de la autocorrelación.

La identificabilidad de la respuesta en frecuencia y en fase de un sistema a partir de medidas de la salida solamente (identificación ciega), dependiendo de si $\{w(k)\}$ es gaussiano o no, se determina como sigue:

1. Si $\{w(k)\}$ es gaussiano y el sistema es de fase mínima, los métodos basados en la autocorrelación pueden determinar tanto la respuesta en fase como en frecuencia.
2. Si $\{w(k)\}$ es gaussiano y el sistema es de fase no mínima, ningún método puede identificar correctamente la respuesta en fase.
3. Si $\{w(k)\}$ es no gaussiano y el sistema es de fase no mínima, los métodos basados en la autocorrelación sólo pueden determinar correctamente la respuesta en frecuencia.
4. Si $\{w(k)\}$ es no gaussiano y el sistema es de fase no mínima, los métodos basados en estadística de alto orden pueden determinar tanto la respuesta en fase como en frecuencia, sin un conocimiento de la verdadera distribución de $\{w(k)\}$.

En el resto de este apartado se presentarán algunas técnicas de identificación de modelos paramétricos con estadística de alto orden, por lo que se supone que el sistema es de fase no mínima, aunque como se verá la estadística de alto orden también se puede utilizar por otros motivos (p.e. eliminación de ruido aditivo gaussiano).

1.4.2 Modelos de media móvil MA

La ecuación que define un proceso $\{x(k)\}$ de media móvil de orden q MA(q) se puede obtener de (1.32) fijando $p = 0$, esto es

$$x(k) = \sum_{i=0}^q b(i)w(k-i) \quad (1.34)$$

El problema de identificación en las condiciones establecidas necesita que los parámetros se estimen a partir de cumulantes de tercer orden o superiores. Por conveniencia se fija $b(0) = 1$. Algunas técnicas para estimar los coeficientes de un modelo MA son las siguientes.

Soluciones como expresiones cerradas

Considérese el modelo

$$y(k) = x(k) + n(k) \quad (1.35)$$

donde $x(k)$ viene dado por (1.34) y $n(k)$ es ruido gaussiano blanco, independiente de $w(k)$. Bajo la hipótesis de que $w(k)$ es ruido blanco de tercer orden, los cumulantes de tercer orden de $y(k)$ se pueden escribir como

$$c_3^y(\tau, \rho) = \gamma_3^w \sum_{i=0}^{\infty} b(i)b(i+\tau)b(i+\rho) \quad (1.36)$$

Evaluando (1.36) para $\tau = q$ y $\rho = k$ se obtiene

$$c_3^y(q, k) = \gamma_3^w b(k)b(q) \quad (1.37)$$

Y para $k = 0$, (1.37) se convierte en

$$c_3^y(q, 0) = \gamma_3^w b(q) \quad (1.38)$$

Combinando (1.37) y (1.38) y suponiendo que $\gamma_3^w \neq 0$ (es decir, que $w(k)$ no está distribuida simétricamente), se obtiene

$$b(k) = \frac{c_3^y(q, k)}{c_3^y(q, 0)} \quad (1.39)$$

La ecuación (1.39) relaciona la respuesta impulso del modelo MA con los cumulantes de tercer orden de la salida del sistema. Se conoce como la fórmula $c(q, k)$ y fue introducida por primera vez por Giannakis [Gia87]. La fórmula $c(q, k)$ proporciona una solución cerrada para el problema de la estimación de los parámetros MA, sin embargo, requiere conocimiento exacto del orden del modelo MA y dado que sólo usa dos cumulantes para la estimación de $b(k)$ no suaviza ni elimina los efectos del ruido aditivo en los estimadores de los cumulantes.

Métodos de optimización

Considérese la rebanada diagonal de los cumulantes de tercer orden del proceso $y(k)$ descrito en (1.35),

$$c_3^y(\tau, \tau) = \gamma_3^w \sum_{i=0}^q b(i)b^2(i + \tau), \quad \tau = -q, \dots, 0, \dots, q \quad (1.40)$$

Lii y Rosenblatt [LR82] sugirieron dos métodos para la estimación de los coeficientes MA basados en (1.40): una aproximación no lineal por mínimos cuadrados y una aproximación por programación lineal. Ambos métodos estiman los cumulantes de tercer orden $\hat{c}_3^y(\tau, \tau)$ y después prosiguen de la siguiente manera:

Aproximación no lineal por mínimos cuadrados Este método estima los parámetros $b(i)$ minimizando la función coste

$$\sum_{\tau=-q}^q \left[\hat{c}_3^y(\tau, \tau) - \gamma_3^w \sum_{i=0}^q b(i)b^2(i + \tau) \right]^2 \quad (1.41)$$

con respecto a las $q + 1$ incógnitas $b(i)$, $i = 1, \dots, q$ y γ_3^w donde $b(0) = 1$. A este método también se le denomina ajuste por cumulantes o cumulant matching.

Aproximación de programación lineal En esta aproximación se emplea primero un método basado en la secuencia de autocorrelación para estimar los coeficientes MA que reflejan adecuadamente la estructura de segundo orden de los datos, sea $\{b^{(2)}(k)\}$. La función de transferencia resultante para el modelo será

$$B(z) = \sum_{k=0}^q b^{(2)}(k)z^{-k} \quad (1.42)$$

con raíces r_j , $j = 1 \dots q$ tales que $|r_j| < 1$ (fase mínima).

Un estimador adecuado de la distribución de las raíces se puede obtener tomando el recíproco conjugado de un número apropiado de r_j s. El método

de programación lineal busca encontrar el conjunto de raíces y por tanto los coeficientes $\{b(k)\}$ que minimiza (1.41).

Los métodos que parten del sistema equivalente de fase mínima suelen recibir el calificativo de métodos SEMP (Spectrally Equivalent Minimum Phase) y una representación de ellos se analizará con detalle en la última parte de esta Memoria.

El método de Giannakis-Mendel GM

En 1989 Giannakis y Mendel [GM89] presentaron un método novedoso basado en estadística de segundo y tercer orden para la identificación de sistemas MA. Su método es conocido como el algoritmo GM.

Considérese el modelo (1.35), donde $w(k)$ es i.i.d, de media cero, no gaussiano y $n(k)$ es de media cero i.i.d con función densidad de probabilidad simétrica e independiente de $x(k)$. Se puede demostrar (véase el apartado 5.2) que se cumple:

$$c_2^x(\tau) + \sum_{i=1}^q b^2(i)c_2^x(\tau - i) = \epsilon \left[c_3^x(\tau, \tau) + \sum_{i=1}^q b(i)c_3^x(\tau - i, \tau - i) \right] \quad (1.43)$$

donde $\epsilon = \gamma_2^w / \gamma_3^w$. Haciendo variar $\tau = -q, \dots, 0, \dots, q$ las ecuaciones que resultan de (1.43) se pueden escribir en forma matricial como

$$\mathbf{R}\boldsymbol{\theta} = \mathbf{r} \quad (1.44)$$

donde $\boldsymbol{\theta} = [\epsilon b(1), \dots, \epsilon b(q), b^2(1), \dots, b^2(q)]^t$ y el resto se definen sin problema a partir de ésta.

La solución por mínimos cuadrados del sistema de ecuaciones sobredeterminado es

$$\boldsymbol{\theta}_{MC} = (\mathbf{R}^t \mathbf{R})^{-1} \mathbf{R}^t \mathbf{r}$$

Una forma adaptativa de resolverla, propuesta por Friedlander y Porat [FP89] se analiza en profundidad en la primera parte de esta Memoria.

1.4.3 Modelos autorregresivos AR

Se han propuesto varios métodos para la determinación de los coeficientes del modelo AR

$$\sum_{i=0}^p a(i)x(k-i) = w(k) \quad (1.45)$$

donde $\{w(k)\}$ es de media nula, no gaussiano con $\gamma_2^w \neq 0$ y $\gamma_3^w \neq 0$. Si el modelo AR es causal y estable (lo más normal) y por tanto de fase mínima, entonces cualquier método basado en estadística de segundo orden se puede emplear para calcular los coeficientes AR. La aplicación de estadística de alto orden se hace imprescindible para modelos AR no causales, que no se van a tratar en esta Memoria.

Sin embargo, el empleo de cumulantes para la identificación de sistemas AR en presencia de ruido aditivo gaussiano puede mejorar los estimadores de las técnicas que contengan estadística de segundo orden. Es por ello por lo que se emplearán en esta Memoria.

Método de la recursión de tercer orden TOR

Supóngase el modelo AR de orden p de la ecuación (1.45), donde como antes $w(k)$ está compuesto de variables i.i.d, no gaussianas, con media nula, varianza σ_w^2 y curtosis γ_w^3 no nula. Además $x(k)$ es independiente de $w(l)$ para $k < l$. Dado que $w(k)$ es estacionaria de tercer orden, $x(k)$ también lo será asumiendo que el AR sea estable. Para el modelo de (1.45) se puede comprobar que

$$c_3^x(-k, -l) + \sum_{i=1}^p a(i)c_3^x(i-k, i-l) = \gamma_w^3 \delta(k, l) \quad k, l > 0 \quad (1.46)$$

propuesto por Raghuveer y Nikias [RN85], donde $\delta(k, l)$ es la función impulso bidimensional. Partiendo de (1.46), a la que se le llamará ecuación de recursión de tercer orden (Third order recursion TOR), se sigue que los $2p+1$ valores de $c_3^x(k, l)$ a lo largo de la diagonal $k = l$ satisfacen la ecuación matricial

$$\mathbf{R}\mathbf{a} = \mathbf{r}$$

donde $\mathbf{a} \equiv [1, a(1), \dots, a(p)]^t$, $\mathbf{b} \equiv [\gamma_w^3, 0, \dots, 0]^t$ y la definición de la matriz \mathbf{R} está clara a partir de estas dos.

Ecuaciones de este tipo se pueden generalizar a cualquier orden de cumulantes, tal y como se lleva a cabo en el apartado 8.10.

Otros métodos

Algunos métodos específicos para la identificación de modelos AR no causales son los siguientes:

El método de Huzii [Huz81] Este método se apoya en el sistema de fase mínima equivalente, calculado a partir de la estadística de segundo orden. A

partir de él y llevando a cabo reflexiones en los ceros, construye sistemas candidatos. Entre los candidatos se elige como sistema verdadero aquel cuyo inverso (un MA) dé la salida más blanca al filtrar los datos $x(k)$. La blancura de la señal se comprueba mediante estadística de alto orden.

Soluciones basadas en cumulantes Fueron propuestas por Tugnait [Tug87], es una particularización del método de ajuste por cumulantes a los modelos AR no causales. En primer lugar se calcula el AR de fase mínima equivalente. Los distintos modelos obtenidos mediante reflexión de los ceros se someten a un criterio de error consistente en comparar el cumulante estimado experimentalmente con el teórico obtenido a partir de los coeficientes AR candidatos. El candidato que minimice la función coste es elegido como sistema verdadero.

Otros métodos con cumulantes se encuentran en [BRAC00].

1.4.4 Métodos para modelos ARMA

Este tipo de modelos no se van a emplear en esta Memoria por lo que en este capítulo sólo se mencionarán algunos métodos empleados en su identificación.

Estimación de los coeficientes ARMA a partir de la función del sistema

Dado un conjunto de datos $\{x(k)\}$ primero se estima su espectro de potencia. A continuación se calcula la respuesta en fase a partir de cualquiera de los métodos convencionales no paramétricos existentes. En tercer lugar, la respuesta impulso ARMA(p, q) se genera como una transformada inversa de la función de transferencia usando una transformada inversa de Fourier. Finalmente los parámetros del modelo se obtienen de la función de transferencia empleando unos aproximantes racionales de Padé [Lii82].

Series temporales residuales

Giannakis y Mendel [GM89] desarrollaron un método que estima los parámetros ARMA usando la autocorrelación y cumulantes de tercer o cuarto orden. La idea básica de su método es estimar los coeficientes AR primero y después formar una serie residual MA a partir de la cual los parámetros MA se pueden obtener por cualquiera de los métodos descritos anteriormente.

El algoritmo $c(q, k)$ doble

Definiendo las cantidades

$$a_3(i, j) = \sum_{k=0}^p a(k)a(k+i)a(k+j)$$

y

$$b_3(i, j) = \sum_{k=0}^q b(k)b(k+i)b(k+j)$$

Giannakis y Mendel mostraron que se verifica

$$\sum_{i=-p}^p \sum_{j=-p}^p c_3^x(m-i, n-j)a_3(i, j) = \begin{cases} 0 & (m, n) \notin S(q) \\ \gamma_w^3 b_3(m, n) & (m, n) \in S(q) \end{cases} \quad (1.47)$$

donde $S(q) = \{0 \leq m \leq q, n \leq m\}$ es la región no redundante de soporte de $b_3(m, n)$. Usando los cumulantes de orden 3 de $x(k)$ y la aproximación de mínimos cuadrados [GS90], los $a_3(m, n)$ se pueden determinar de (1.47)a. Entonces, usando los calculados $a_3(m, n)$ y los cumulantes de tercer orden de $x(k)$ se obtienen los $b_3(m, n)$ mediante (1.47)b.

Algoritmo de la rebanada Q

Es un algoritmo propuesto por Swami y Mendel [SM90] que se puede considerar como la generalización a modelos ARMA del método $c(q, k)$ para la identificación de modelos MA, de ahí su nombre. Los detalles se dan en el artículo referenciado.

1.5 Algoritmos adaptativos

En la gran mayoría de problemas que aparecen en el campo del tratamiento de señales se desconoce la estadística de la misma, con lo cual se hace necesario estimarla. Como normalmente las muestras van llegando paulatimamente conforme se realiza el muestreo, sería conveniente contar con algoritmos que con la llegada de nuevas muestras actualizaran de la manera más eficiente posible la información que se tiene sobre el problema. Este tipo de algoritmos reciben el nombre de algoritmos adaptativos.

Los algoritmos adaptativos se hacen imprescindibles en situaciones donde la estadística del problema cambie con el tiempo, ya que como su propio nombre indica son capaces de adaptarse a los cambios que sufran las condiciones del

problema mediante una conveniente ecuación de actualización de las incógnitas. Tras sucesivas iteraciones el algoritmo converge a la solución óptima si trabaja en situaciones estacionarias o lleva a cabo un rastreo de la evolución del problema en situaciones no estacionarias.

La formulación de la gran mayoría de algoritmos adaptativos se basa en la minimización de una función coste, cuyo mínimo es la solución óptima del problema. Para la búsqueda de ese mínimo se suele recurrir al método de máxima pendiente, basado en la sencilla idea de que para alcanzar el mínimo de una función basta con avanzar en la dirección del gradiente pero en sentido opuesto. Otros algoritmos adaptativos emplean relaciones algebraicas y matriciales para imponer la condición de mínimo. En este apartado se va a presentar un algoritmo de cada tipo, los cuales constituyen los algoritmos básicos de dos de las familias más importantes de algoritmos adaptativos: el cuadrado de media mínima (Least-Mean Square) y de mínimos cuadrados (Least Squares).

1.5.1 Algoritmos del cuadrado de media mínima

El objetivo primordial de estos algoritmos es el de estimar una variable aleatoria dada $d(n)$ a partir de la combinación lineal de p variables aleatorias $x(n), x(n-1), \dots, x(n-p+1)$ para cada instante n . La función coste a minimizar es el error cuadrático medio, donde este error es el error de estimación $e(n) = d(n) - \mathbf{w}^t \mathbf{x}(n)$, siendo \mathbf{w} el vector de coeficientes de la estimación lineal y $\mathbf{x}(n) = [x(n), x(n-1), \dots, x(n-p+1)]$ el vector de datos a tiempo n .

Derivando el error con respecto al vector de coeficientes se obtiene el gradiente, que igualado a cero constituye la condición de mínimo:

$$E[\mathbf{x}(n)e(n)] = 0$$

O equivalentemente, empleando la definición de error de estimación, el vector de coeficientes óptimos según este criterio es

$$\mathbf{w}_0 = \mathbf{R}^{-1} \mathbf{r}$$

siendo $\mathbf{R} = E[\mathbf{x}(n)\mathbf{x}^t(n)]$ la matriz de autocorrelación del vector de datos y $\mathbf{r} = E[\mathbf{x}(n)d(n)]$ el vector de correlación cruzada entre el vector de datos y la respuesta deseada.

A partir de estas magnitudes se puede alcanzar el mínimo de la función coste como sigue.

Algoritmo de máxima pendiente

Tal y como se comentó en párrafos precedentes si en la iteración $n - 1$ se dispone de una solución dada por $\mathbf{w}(n - 1)$, la siguiente recursión avanza en la dirección de máxima pendiente descendente,

$$\mathbf{w}(n) = \mathbf{w}(n - 1) + \mu \mathbf{E}[\mathbf{x}(n)e(n)]$$

donde μ es el tamaño del paso que se da en el sentido contrario al gradiente. Esta expresión es equivalente a

$$\mathbf{w}(n) = \mathbf{w}(n - 1) + \mu[\mathbf{r} - \mathbf{R}\mathbf{w}(n - 1)]$$

Esta expresión determinista requiere del conocimiento de la estadística del problema (\mathbf{R} y \mathbf{r}), requisito que casi nunca se cumple. El siguiente algoritmo salva este inconveniente.

El algoritmo LMS

Este algoritmo es la versión tipo gradiente estocástico del anterior. Su nombre se debe a que dada la imposibilidad de conocer el gradiente de la función coste, propone estimarlo mediante el valor instantáneo: $\hat{\nabla}(n) = \mathbf{x}(n)e(n)$ con lo que la ecuación de actualización del nuevo algoritmo queda:

$$\mathbf{w}(n) = \mathbf{w}(n - 1) + \mu \mathbf{x}(n)e(n)$$

Esta expresión constituye el corazón del algoritmo del cuadrado con media mínima o Least-Mean Square (LMS) [WH60]. Una descripción más detallada del mismo junto con su análisis de convergencia se presentan en el capítulo 7.

1.5.2 Algoritmos recursivos de mínimos cuadrados

El problema se plantea de forma parecida al del LMS pero en una situación determinista: se pretende aproximar una serie $d(n)$ a partir de la combinación lineal de p muestras de una serie de datos. La función coste a minimizar se define como:

$$\xi(n) = \sum_{i=1}^n \beta(i, n) |e(i, n)|^2$$

es decir, una combinación lineal de errores de estimación al cuadrado siendo $e(i, n) = d(i) - \mathbf{w}^t(n)\mathbf{x}(i)$ el error que se cometería con los datos a tiempo i si se emplease el estimador conocido en la iteración n -ésima. El coeficiente $\beta(i, n)$ es un factor de olvido que disminuye el peso de los errores a iteraciones pasadas y

potencia los errores actuales. Normalmente se suele emplear un factor de olvido exponencial dado por $\beta(n, i) = \lambda^{n-i}$ con $0 < \lambda < 1$.

Derivando de nuevo esta función coste con respecto al vector \mathbf{w} se obtiene que los coeficientes óptimos son aquellos que cumplen la siguiente ecuación matricial:

$$\mathbf{R}(n)\mathbf{w}(n) = \mathbf{r}(n) \quad (1.48)$$

donde

$$\mathbf{R}(n) = \sum_{i=1}^n \lambda^{n-i} \mathbf{x}(i)\mathbf{x}^t(i)$$

$$\mathbf{r}(n) = \sum_{i=1}^n \lambda^{n-i} \mathbf{x}(i)d(i)$$

que como se puede apreciar son la matriz de autocorrelación del vector de datos y el vector de correlaciones cruzadas entre el vector de datos y la respuesta deseada, en versión determinista.

Empleando el lema de inversión matricial (véase apéndice A) en la ecuación (1.48) se puede obtener una ecuación de actualización para las incógnitas:

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \mathbf{k}(n)e(n)$$

donde $e(n) = d(n) - \mathbf{w}^t(n-1)\mathbf{x}(n)$ es el error de estimación y $\mathbf{k}(n)$ es el vector de ganancia, definido como

$$\mathbf{k}(n) = \frac{\lambda^{-1}\mathbf{P}(n-1)\mathbf{x}(n)}{1 + \lambda^{-1}\mathbf{x}^t(n)\mathbf{P}(n-1)\mathbf{x}(n)}$$

donde $\mathbf{P}(n)$ es la inversa de la matriz de autocorrelación $\mathbf{R}(n)$.

Los detalles de este algoritmo así como un completo análisis de su convergencia se presentan en el capítulo 2. Se le denomina algoritmo recursivo de mínimos cuadrados RLS (Recursive Least Squares), y fue propuesto en [Pla50].

1.5.3 Algoritmos adaptativos y estadística de alto orden. Generalizaciones

Los algoritmos adaptativos suscitadamente presentados en los apartados anteriores son dos de los algoritmos adaptativos más extendidos y conocidos. A partir de ellos se han derivado innumerables versiones [BBCC88], [BM89], [KCM84], [MD95], [MT91], ya sea para mejorar su comportamiento (velocidad de convergencia, error en los estimadores), bien para reducir su carga computacional o bien para adaptarlos a nuevos problemas.

En lo que respecta a este último punto, los algoritmos adaptativos se han modificado para que puedan introducir estadística de alto orden y se beneficien de ella. En las partes I y II de esta memoria se presentan algunos miembros de las familias LMS y RLS respectivamente. En particular se enriquece la familia del algoritmo LMS con la obtención de nuevos algoritmos dentro del formalismo del principio de ortogonalidad.

1.6 Aplicación: Canales de comunicaciones

Una de las posibles aplicaciones que presentan los algoritmos de identificación de sistemas estudiados y propuestos en esta Memoria es el de ecualización de canales de comunicaciones. De todas maneras, no todas las simulaciones presentadas se podrían relacionar directamente con este problema.

1.6.1 Comunicación digital

Durante el proceso por el cual la información llega de un emisor a un receptor, ésta experimenta varias modificaciones [Hay88] para conseguir un mejor comportamiento del sistema (léase velocidad de transmisión, mayor fiabilidad, necesidad de una menor energía, etc.).

Actualmente ya no hay duda sobre las ventajas de utilización de formatos digitales sobre los formatos analógicos. Por tanto, se supone que la información a enviar está en dicha forma. En caso de no estarlo se debería usar un conversor analógico-digital. Entre estas ventajas se encuentra la mayor fiabilidad en la transmisión (se ve menos afectada por el ruido), la posibilidad de encriptar el mensaje (seguridad en las comunicaciones), etc.

1.6.2 Conversión Analógico-Digital

Por lo comentado anteriormente se entiende fácilmente que el primer paso a dar en el envío de cualquier información es su conversión a formato digital, en caso de estar en forma analógica.

Dicha conversión se compone de tres pasos (figura 1.2):

1. Muestreo: la función continua inicial se convierte en una serie de escalares, cuyos valores coinciden con el de la función en instantes de tiempo equiespaciados. El muestreo se lleva a cabo con una cierta frecuencia f_0 .

2. Cuantización: los escalares de la serie así obtenida se aproximan por su valor más cercano dentro de un conjunto discreto de valores.
3. Codificación: los valores muestreados y cuantizados sufren una primera codificación. Pasan de estar expresados en sistema decimal a binario.



Figura 1.2 *Proceso de conversión analógico-digital*

Una vez realizados estos procesos ya se está en condiciones de introducir la señal digital dentro del sistema de comunicaciones digital. Las partes de dicho sistema se detallan en la siguiente sección.

1.6.3 Componentes de un sistema de comunicación digital

La mayoría de las modificaciones que sufre la señal digital de entrada van encaminadas a mejorar el comportamiento del sistema y a volver a convertirla en analógica para su transmisión por el canal físico elegido (cable coaxial, fibra óptica, microondas, satélite, etc.).

La primera operación que se ejecuta es la codificación de la fuente (source coding) con el objeto de eliminar redundancias en la señal digital introducida, reduciendo así su longitud. Consúltese la figura 1.3.

A continuación sufre una segunda codificación (codificación del canal / channel coding) con el objeto de adecuar la serie obtenida en el paso anterior a las características del canal concreto. Aquí se incluye ya la elección del tipo de modulación elegida (PAM, PSK, ASK, FSK, etc.), que da como resultado una constelación de símbolos concreta.

Seguidamente la señal codificada pasa por el modulador con objeto de convertirla en una señal continua de nuevo (no tiene nada que ver con la señal analógica primera).

Esta señal continua pasa por el canal dando lugar a la señal continua de salida.

Con la señal continua de salida se alimenta el detector (demodulador) para obtener la constelación de símbolos de salida. Al conjunto de modulador, canal y detector se le denomina canal discreto, ya que tanto su entrada como salida son discretas.

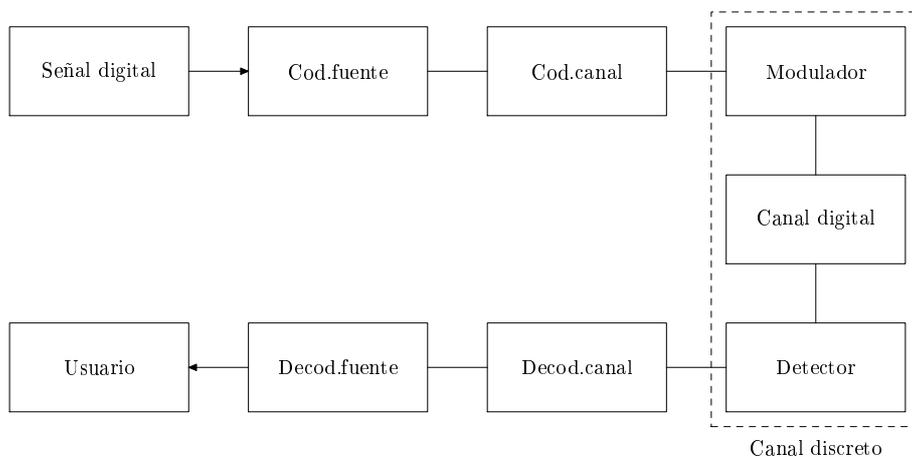


Figura 1.3 Componentes de un sistema de comunicaciones digital

Finalmente se invierten las codificaciones realizadas anteriormente. Primero se realiza la decodificación del canal y a continuación la decodificación de la fuente con lo que se obtiene la señal digital final para el usuario.

El objetivo es centrar el estudio en el canal discreto. Es decir, se supondrá una constelación de símbolos de entrada, que dependerá de la modulación que se considere, se filtrará por un canal discreto y se obtendrá una constelación de salida.

1.6.4 Formulación del problema

En la figura 1.4 [Hay94] se muestra el diagrama de bloques de un sistema de comunicaciones digital síncrono y baseband, sujeto a interferencia intersimbólica (la salida a tiempo n contiene nuestras pasadas de la entrada además de la entrada a tiempo n) y ruido aditivo gaussiano.

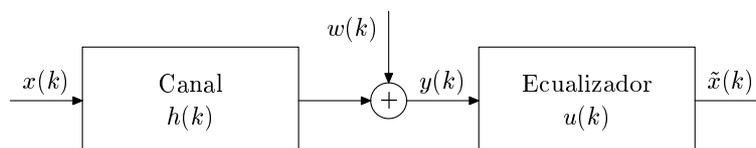


Figura 1.4 Modelo equivalente de un sistema de comunicaciones digital

Se asume que la serie recibida, tras ser demodulada, filtrada por un paso-baja y muestreada sincrónicamente (a una velocidad $1/T$) se puede escribir como:

$$y(k) = h(k) * x(k) + w(k) = \sum_{i=-L_2}^{L_1} h(i)x(k-i) + w(k) \quad (1.49)$$

La serie de datos de entrada $\{x(k)\}$ es, en general, un proceso complejo no gaussiano i.i.d. Usualmente, tiene una función de densidad de probabilidad simétrica y por tanto tiene media y factor de asimetría nulos. Por ejemplo, $\{x(k)\}$ podría ser la serie compleja equivalente baseband de una señal modulada por amplitud de cuadratura rectangular (rectangular quadrature amplitude modulated Q.A.M). El ruido aditivo $\{w(k)\}$ es de media nula, gaussiano, generalmente coloreado (no blanco) y estadísticamente independiente de la entrada. La respuesta impulso del canal $\{h(k)\}$ es, en general, de fase no mínima y da cuenta del filtro de transmisión, el canal y el filtro receptor. Además se asume que $\{h(k)\}$ cambia lentamente con el tiempo y se puede considerar constante para un número grande de muestras. Finalmente, la serie $\{y(k)\}$ es la entrada al ecualizador ciego.

El objetivo del ecualizador ciego con función de transferencia $U(z)$ es recuperar la serie de entrada $\{x(k)\}$ salvo posiblemente un retraso constante d y un desfase θ . Suponiendo que $\{\tilde{x}(k)\}$ es la salida del ecualizador, ésta se puede expresar como sigue en el dominio de la transformada Z:

$$\tilde{X}(z) = X(z)z^{-d}e^{i\theta} \quad (1.50)$$

Ignorando el ruido aditivo en la figura 1.4, la expresión (1.50) se puede reescribir como sigue

$$U(z)H(z)X(z) = X(z)z^{-d}e^{i\theta} \quad (1.51)$$

El retraso desconocido d no afecta a la recuperación del mensaje original, mientras que θ se puede eliminar ajustando adecuadamente la fase de la portadora antes o después de la ecualización. Por tanto, en lo que sigue, se supondrá que d y θ son nulos. Entonces, de (1.51) se obtiene

$$U(z) = \frac{1}{H(z)} \quad (1.52)$$

En consecuencia, el ecualizador debe primero identificar y aproximar las propiedades del canal inverso para cancelar la interferencia intersimbólica. Esto se conoce como el criterio de ecualización zero-forcing.

Es conveniente considerar lo siguiente:

1. Los métodos adaptativos clásicos de ecualización se basan casi exclusivamente en estadística de segundo orden, y por tanto fallan al intentar

ecualizar correctamente canales de fase no mínima cuando la señal de entrada es inaccesible. Esto se debe a la imposibilidad de la estadística de segundo orden para distinguir entre la información de fase mínima y fase máxima.

2. No hay solución al problema cuando la entrada es gaussiana y el canal es de fase no mínima.
3. En la mayoría de los enlaces de comunicaciones digitales, los datos transmitidos son series no gaussianas y los canales lineales distorsionadores son, en general, de fase no mínima.
4. Para conseguir la ecualización ciega de canales de fase no mínima, se debe emplear estadística de orden mayor que dos. Esto se debe a la capacidad de la estadística de alto orden para preservar el verdadero carácter de la fase de las señales. Ciertamente, todos los ecualizadores ciegos propuestos, indirectamente a través de transformaciones no lineales o directamente, utilizan estadística de alto orden.

Parte I

Algoritmos adaptativos de mínimos cuadrados con estadística de alto orden. Análisis y aplicación a identificación ciega

Introducción

En esta parte de la Memoria se van a analizar varios algoritmos representativos del grupo de algoritmos recursivos de mínimos cuadrados [GBT99]. Este grupo surge como resultado de minimizar una función coste definida como la distancia euclídea entre una serie de referencia y su estimador lineal para un número N de instantes temporales. Equivalentemente esta distancia puede entenderse como el error cometido en la estimación.

Al minimizar esta función coste se obtiene una condición de mínimo que se puede interpretar como imponer ortogonalidad entre el error de estimación y el vector de datos empleados en la estimación, utilizando un producto escalar particular. Esta equivalencia entre los algoritmos recursivos de mínimos cuadrados y el principio de ortogonalidad será muy útil para dotar a esta familia de un formalismo que permita aplicarlo a más problemas.

Estimar una serie de referencia (o deseada) mediante otra dada siguiendo un criterio de mínimos cuadrados, equivale, si las series son aleatorias, a recurrir a la estadística de segundo orden para resolver el problema. Procediendo de esta manera seguramente habrá información que se escapa al estar contenida en estadística de orden superior. Esta falta se puede solventar si dentro del formalismo del principio de ortogonalidad se sustituye el vector de datos por otro vector, definido convenientemente normalmente como función de éste, llamado de variable instrumental.

Aplicar cualquiera de estos algoritmos a la identificación ciega supone contar con algún tipo de ecuación (matricial) que relacione los coeficientes que definen al sistema con algún tipo de medida acerca de la salida, tal como sus momentos o cumulantes. Además tal ecuación debe poder expresarse como una condición de ortogonalidad.

Como es conocido, para llevar a cabo una identificación ciega completa es necesario conocer información acerca de la fase, que sólo se encuentra contenida en la estadística de alto orden, de manera que hay que recurrir a algoritmos capaces de manejar una variable instrumental. Más aún, si la ecuación matri-

cial asociada a la condición de ortogonalidad es sobredeterminada, el algoritmo deberá tener en cuenta además esta contingencia.

Todas las situaciones descritas están cubiertas en la actualidad por algún algoritmo concreto. El algoritmo base del cual se construyen todos los demás es el recursivo de mínimos cuadrados RLS (recursive least squares) capaz de resolver sistemas de ecuaciones cuadradas que involucren estadística de segundo orden. Si aparece estadística de alto orden hay que recurrir al algoritmo recursivo de variable instrumental RIV (recursive instrumental variable). Pero si además es sobredeterminado hay que emplear el algoritmo sobredeterminado recursivo de variable instrumental ORIV (overdetermined recursive instrumental variable).

Los objetivos primordiales de esta parte de la memoria son:

1. Presentar un análisis teórico de los 3 algoritmos descritos anteriormente con el fin de comprender mejor su funcionamiento.
2. Aplicar los dos últimos al problema de la identificación ciega mediante el uso de 2 sistemas de ecuaciones. El primero relaciona cumulantes de orden 2 y 3 con los coeficientes del sistema y el segundo lo hace con cumulantes de orden 3 y 4. Con esto se pretende discutir las ventajas de un método basado exclusivamente en estadística de alto orden.
3. Verificar que los resultados obtenidos mediante ORIV son mejores que los obtenidos mediante RIV, de forma que se muestre la ventaja de trabajar con sistemas sobredeterminados.

En el desarrollo de estos 3 puntos se ha aportado como contribución original de esta memoria lo siguiente:

1. Se ha completado el análisis de RIV realizado por Swami [Swa96] para sistemas no estacionarios y se han obtenido las condiciones óptimas de trabajo.
2. Se ha realizado un análisis completo (teórico y mediante simulaciones) de ORIV.
3. Se ha derivado en el dominio del tiempo una expresión general de la cual se puede obtener tanto la ecuación de GM que involucra cumulantes de orden 2 y 3 como su generalización a orden 3 y 4.
4. Se ha comparado teóricamente y mediante simulaciones RIV con ORIV.

El contenido de esta primera parte se ha estructurado de la siguiente manera. En el capítulo 2 se analiza (teóricamente) el algoritmo RLS para que sirva de toma de contacto de las técnicas a utilizar en los siguientes capítulos. El análisis

llevado a cabo por Swami se expone y se completa para ambientes no estacionarios en el capítulo 3. Paralelamente a este último capítulo, en el capítulo 4 se presenta un análisis completo de ORIV. Este análisis se contrasta mediante simulaciones involucrando estadística mixta y de alto orden en el capítulo 5, donde se muestra la superioridad de la segunda en ambientes ruidosos. Finalmente en el capítulo 6 se plantea la previsible superioridad de ORIV sobre RIV mediante análisis teórico y a través de las simulaciones realizadas.

Capítulo 2

El algoritmo recursivo de mínimos cuadrados RLS

Introducción

En este capítulo se pretende caracterizar lo mejor posible el algoritmo recursivo de mínimos cuadrados RLS. La principal propiedad de este algoritmo es que es adaptativo y que por tanto es capaz de proporcionar una solución actualizada a partir de la antigua con la llegada de un nuevo dato, con carga computacional pequeña. Conforme más datos se tengan mejor será el estimador obtenido, de forma que determinar cómo se produce la convergencia del estimador es de capital importancia. Cuando el parámetro verdadero evoluciona con el tiempo esta convergencia progresiva del estimador hacia él se convierte en un rastreo continuo.

Como caracterizar un estimador es dar al menos su sesgo y su varianza, para caracterizar a RLS hay que estudiar la convergencia en media y en media cuadrática tanto en ambientes estacionarios como no estacionarios.

La técnica recurrida para llevar a cabo este análisis es la técnica del promediado directo, que como su propio nombre indica consiste en sustituir, allí donde sea conveniente y apropiado, una variable aleatoria por su valor medio. Su uso concreto se detallará en el cuerpo de este capítulo.

La totalidad de este capítulo se basa en [Hay96] pero sin considerar el posible carácter complejo de las series. El algoritmo se obtiene en el apartado 2.1 y se analiza su convergencia en media y en media cuadrática para ambientes estacionarios en el apartado 2.2. Igualmente la capacidad de rastreo se analiza

en el apartado 2.3.

2.1 Motivación y obtención del algoritmo

Uno de los usos tradicionales y más directos de los algoritmos adaptativos es el de la estimación de una serie dada $\{d(n)\}$ mediante combinaciones lineales, definidas por un vector \mathbf{w}_0 , de otra serie conocida $\{x(n)\}$; los parámetros incógnita son los coeficientes de la combinación lineal y dado que esta operación se puede implementar fácilmente mediante un filtro transversal, a veces reciben el nombre de coeficientes o pesos del filtro.

Para series deterministas este objetivo se puede conseguir mediante la minimización de la siguiente función coste, tipo mínimos cuadrados:

$$\epsilon(n) = \sum_{i=1}^n \beta(n, i) |e(i; n)|^2 \quad (2.1)$$

donde $e(i; n) = d(i) - \mathbf{w}^t(n)\mathbf{x}(i)$ es el error cometido en la estimación con los datos conocidos a tiempo i pero introduciendo el vector de coeficientes estimado a tiempo n . Téngase en cuenta que $\mathbf{w}(n) = [w_1(n), \dots, w_q(n)]^t$ y $\mathbf{x}(n) = [x(n), x(n-1), \dots, x(n-q+1)]^t$. Asimismo, el factor $\beta(n, i)$ se conoce como factor peso ya que da distinta importancia a los errores cometidos para distintos tiempos; el más utilizado tiene la forma $\beta(n, i) = \lambda^{n-i}$ con $\lambda < 1$ y su función es la de considerar más aquellos errores cometidos a tiempos i más cercanos al actual n . De esta manera la expresión a minimizar es

$$\epsilon(n) = \sum_{i=1}^n \lambda^{n-i} |e(i; n)|^2 \quad (2.2)$$

y la ecuación que deben cumplir los coeficientes del filtro que la minimicen es:

$$\Phi(n)\mathbf{w}(n) = \mathbf{z}(n) \quad (2.3)$$

conocida como ecuación normal. En ella, las distintas magnitudes están definidas como:

$$\Phi(n) = \sum_{i=1}^n \lambda^{n-i} \mathbf{x}(i)\mathbf{x}^t(i) \quad (2.4)$$

$$\mathbf{z}(n) = \sum_{i=1}^n \lambda^{n-i} \mathbf{x}(i)d(i) \quad (2.5)$$

Nótese que las series conocidas se suponen nulas para tiempos $n = 0$ y anteriores, de ahí que las sumatorias empiecen todas por $i = 1$. Es decir, se ha asumido

cierta forma de preventaneado. La forma más directa de resolver la ecuación (2.3) es la de invertir la matriz $\Phi(n)$:

$$\mathbf{w}(n) = \Phi^{-1}(n)\mathbf{z}(n) \quad (2.6)$$

lo cual requeriría, con una inversión usual de matrices, del orden de q^3 operaciones. Obviamente, si la longitud del filtro aumenta, el número de operaciones aumenta más todavía. Si además se pretende resolver el sistema con la llegada de nuevos datos, $d(n)$ y $\mathbf{x}(n)$, la carga computacional puede llegar a ser insoportable.

El objetivo del algoritmo RLS (Recursive Least Squares) es el de calcular la expresión (2.6) de una manera recursiva y lo más eficientemente posible. El carácter recursivo se lo otorga el empleo de las ecuaciones de actualización de $\Phi(n)$ y $\mathbf{z}(n)$:

$$\Phi(n) = \lambda\Phi(n-1) + \mathbf{x}(n)\mathbf{x}^t(n) \quad (2.7)$$

$$\mathbf{z}(n) = \lambda\mathbf{z}(n-1) + \mathbf{x}(n)d(n) \quad (2.8)$$

La eficacia la consigue mediante la aplicación del lema de inversión matricial (véase apéndice A) a la expresión (2.7). Para ello se identifican las magnitudes empleadas tal y como sigue:

$$\mathbf{A} = \Phi(n) \quad (2.9)$$

$$\mathbf{B}^{-1} = \lambda\Phi(n-1) \quad (2.10)$$

$$\mathbf{C} = \mathbf{E} = \mathbf{x}(n) \quad (2.11)$$

$$\mathbf{D} = 1 \quad (2.12)$$

resultando, tras aplicar el lema:

$$\Phi^{-1}(n) = \lambda^{-1}\Phi^{-1}(n-1) - \frac{\lambda^{-2}\Phi^{-1}(n-1)\mathbf{x}(n)\mathbf{x}^t(n)\Phi^{-1}(n-1)}{1 + \lambda^{-1}\mathbf{x}^t(n)\Phi^{-1}(n-1)\mathbf{x}(n)} \quad (2.13)$$

Como se aprecia, una vez conocida la inversa de $\Phi(n-1)$, para calcular la inversa de $\Phi(n)$ la única inversión necesaria es la de un escalar.

Para simplificar la notación es conveniente definir:

$$\mathbf{k}(n) = \frac{\lambda^{-1}\Phi^{-1}(n-1)\mathbf{x}(n)}{1 + \lambda^{-1}\mathbf{x}^t(n)\Phi^{-1}(n-1)\mathbf{x}(n)} \quad (2.14)$$

que recibe el nombre de vector de ganancia por razones que se darán más adelante. Con esto la ecuación (2.13) se puede reescribir como:

$$\Phi^{-1}(n) = \lambda^{-1}\Phi^{-1}(n-1) - \lambda^{-1}\mathbf{k}(n)\mathbf{x}^t(n)\Phi^{-1}(n-1) \quad (2.15)$$

Para obtener la ecuación de actualización de los pesos se sustituye la expresión (2.8) en la ecuación (2.6) resultando:

$$\mathbf{w}(n) = \lambda\Phi^{-1}(n)\mathbf{z}(n-1) + \Phi^{-1}(n)\mathbf{x}(n)d(n) \quad (2.16)$$

Para continuar es conveniente hacer notar que

$$\mathbf{k}(n) = \Phi^{-1}(n)\mathbf{x}(n) \quad (2.17)$$

lo cual se comprueba reordenando la expresión (2.14) para dar

$$\mathbf{k}(n) = [\lambda^{-1}\Phi^{-1}(n-1) - \lambda^{-1}\mathbf{k}(n)\mathbf{x}^t(n)\Phi^{-1}(n-1)]\mathbf{x}(n) \quad (2.18)$$

y teniendo en cuenta que la expresión entre corchetes equivale a $\Phi^{-1}(n)$. Una vez demostrada esta igualdad se puede emplear en (2.16),

$$\mathbf{w}(n) = \lambda\Phi^{-1}(n)\mathbf{z}(n-1) + \mathbf{k}(n)d(n) \quad (2.19)$$

que se puede seguir modificando teniendo en cuenta la expresión de actualización (2.15) dando

$$\mathbf{w}(n) = \Phi^{-1}(n-1)\mathbf{z}(n-1) - \mathbf{k}(n)\mathbf{x}^t(n)\Phi^{-1}(n-1)\mathbf{z}(n-1) + \mathbf{k}(n)d(n) \quad (2.20)$$

Teniendo en cuenta ahora la igualdad (2.6) se puede escribir:

$$\mathbf{w}(n) = \mathbf{w}(n-1) - \mathbf{k}(n)\mathbf{x}^t(n)\mathbf{w}(n-1) + \mathbf{k}(n)d(n) \quad (2.21)$$

llegando finalmente a:

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \mathbf{k}(n)\xi(n) \quad (2.22)$$

donde $\xi(n)$ es el error de estimación a priori definido por

$$\xi(n) = d(n) - \mathbf{w}^t(n-1)\mathbf{x}(n) \quad (2.23)$$

Con las expresiones dadas hasta aquí se puede cerrar el ciclo recursivo de RLS; en la tabla 2.1 se muestran los pasos necesarios para llevar a cabo este algoritmo.

<u>Condiciones iniciales</u>	
$\Phi^{-1}(0) = \delta^{-1}\mathbf{I}$	$\delta = \text{constante positiva pequeña}$
$\mathbf{w}(0) = \mathbf{0}$	
<u>Cuerpo del algoritmo</u>	
$\mathbf{k}(n) = \frac{\lambda^{-1}\Phi^{-1}(n-1)\mathbf{x}(n)}{1 + \lambda^{-1}\mathbf{x}^t(n)\Phi^{-1}(n-1)\mathbf{x}(n)}$	
$\xi(n) = d(n) - \mathbf{w}^t(n-1)\mathbf{x}(n)$	
$\mathbf{w}(n) = \mathbf{w}(n-1) + \mathbf{k}(n)\xi(n)$	
$\Phi^{-1}(n) = \lambda^{-1}\Phi^{-1}(n-1) - \lambda^{-1}\mathbf{k}(n)\mathbf{x}^t(n)\Phi^{-1}(n-1)$	

Tabla 2.1 Pasos del algoritmo RLS

La inicialización indicada es la que se conoce como inicialización con ligadura suave, aunque también es posible inicializarlo de manera exacta.

2.2 Análisis de la convergencia en ambientes estacionarios

En este apartado se van a estudiar las propiedades del algoritmo RLS referentes a su convergencia, en particular se van a considerar la convergencia en media y en media cuadrática. Para llevar a cabo este análisis se va a suponer que la respuesta deseada $d(n)$ está generada mediante un modelo lineal de regresión múltiple determinado por los coeficientes \mathbf{w}_0 más cierto error de medición:

$$d(n) = \mathbf{w}_0^t \mathbf{x}(n) + e_0(n) \quad (2.24)$$

En lo que sigue se supondrá que el error de medición es un proceso blanco de media nula y varianza σ_0^2 . Asimismo, y por ahora, el vector \mathbf{w}_0 se supondrá constante lo que equivale a trabajar con sistemas estacionarios de forma tal que el factor de olvido pueda ser la unidad; de esta forma se tienen en cuenta todas las muestras en el promedio dado en las definiciones (2.4) y (2.5) y no sólo las más actuales como ocurriría si λ fuese menor que la unidad.

2.2.1 Convergencia en media

Teniendo en cuenta (2.24) en la definición de $z(n)$ dada en la ecuación (2.5) se obtiene

$$z(n) = \Phi(n) \mathbf{w}_0 + \sum_{i=1}^n \mathbf{x}(i) e_0(i) \quad (2.25)$$

por lo que la ecuación (2.6) arroja el siguiente valor para los coeficientes del filtro en este caso:

$$\mathbf{w}(n) = \mathbf{w}_0 + \Phi^{-1}(n) \sum_{i=1}^n \mathbf{x}(i) e_0(i) \quad (2.26)$$

Debido al carácter aleatorio¹ de las series involucradas y al modelo empleado para $d(n)$, la solución calculada mediante el algoritmo RLS presenta una parte determinista igual al vector constante utilizado como regresor más una componente aleatoria debida al ruido de medición. Es interesante resaltar que si no existiera este ruido la solución obtenida mediante el RLS sería la exacta, siempre

¹Cómo es posible tratar series aleatorias con un planteamiento del problema puramente determinista se detalla en el apartado 4.2.1 cuando se analiza el algoritmo ORIV.

y cuando se contase con un número de datos mayor que la longitud del regresor (para que existieran las inversas correspondientes).

En media el algoritmo tiende a:

$$\mathbf{E}[\mathbf{w}(n)] = \mathbf{w}_0 + \mathbf{E}\left[\Phi^{-1}(n) \sum_{i=1}^n \mathbf{x}(i)e_0(i)\right] \quad (2.27)$$

y teniendo en cuenta que $e_0(i)$ es independiente del vector de entrada $\mathbf{x}(i)$ para todo i el valor esperado del miembro de la derecha es nulo al ser el error de medición de media nula. Por lo que resulta:

$$\mathbf{E}[\mathbf{w}(n)] = \mathbf{w}_0 \quad n \geq q \quad (2.28)$$

Es decir, el algoritmo RLS converge en media siempre y cuando se cuente con más datos que incógnitas.

2.2.2 La desviación en media cuadrática

Para que la convergencia en media anterior sea significativa es necesario asegurarse de que también se produce convergencia en media cuadrática de los coeficientes del filtro. Es por tanto necesario estudiar cómo se comporta la cantidad $D(n) = \mathbf{E}[\Delta\mathbf{w}^t(n)\Delta\mathbf{w}(n)]$ donde $\Delta\mathbf{w}(n) = \mathbf{w}(n) - \mathbf{w}_0$ es el vector de error en los pesos. A $D(n)$ se le conoce como desviación en media cuadrática MSD (mean squared deviation).

Para comenzar se tiene en cuenta que el vector de error en los pesos, con la ayuda de (2.6) y (2.24) se puede reescribir como

$$\Delta\mathbf{w}(n) = \Phi^{-1}(n) \sum_{i=1}^n \mathbf{x}(i)e_0(i) \quad (2.29)$$

y con él se puede construir su matriz de autocorrelación $\mathbf{K}(n)$:

$$\mathbf{K}(n) = \mathbf{E} \left[\Phi^{-1}(n) \sum_{i=1}^n \sum_{j=1}^n \mathbf{x}(i)e_0(i)e_0(j)\mathbf{x}^t(j)\Phi^{-1}(n) \right] \quad (2.30)$$

Teniendo en cuenta de nuevo que $e_0(n)$ es independiente del resto de variables aleatorias que aquí aparecen y que proviene de un proceso de ruido blanco de varianza σ_0^2 , resulta:

$$\mathbf{K}(n) = \sigma_0^2 \mathbf{E} \left[\Phi^{-1}(n) \sum_{i=1}^n \mathbf{x}(i)\mathbf{x}^t(i)\Phi^{-1}(n) \right] \quad (2.31)$$

por lo que tras realizar la sumatoria indicada y simplificar se llega a:

$$\mathbf{K}(n) = \sigma_0^2 \mathbf{E} [\Phi^{-1}(n)] \quad (2.32)$$

Para calcular este valor esperado se recurre a nuevas hipótesis de independencia:

1. Los vectores de entrada $\mathbf{x}(1)$, $\mathbf{x}(2)$, ..., $\mathbf{x}(n)$ son independientes entre sí y están idénticamente distribuidos.
2. Los vectores de entrada $\mathbf{x}(1)$, $\mathbf{x}(2)$, ..., $\mathbf{x}(n)$ pertenecen a un proceso estocástico con una distribución gaussiana multivariada de media cero y matriz de correlación \mathbf{R} .

Según estas hipótesis la matriz Φ sigue una distribución de probabilidad tipo Wishart. Para este tipo de matrices se puede comprobar que:

$$\mathbf{E}[\Phi^{-1}(n)] = \frac{1}{n - q - 1} \mathbf{R}^{-1} \quad n > q + 1 \quad (2.33)$$

a partir de esto se llega sin dificultad a:

$$D(n) = \text{Tr}\{\mathbf{K}(n)\} = \frac{\sigma_0^2}{n - q - 1} \sum_{i=1}^q \frac{1}{\lambda_i} \quad (2.34)$$

donde λ_i son los autovalores de la matriz de autocorrelación \mathbf{R} . A partir de esta expresión se pueden hacer los siguientes comentarios:

1. El MSD viene determinado por el menor autovalor, por tanto, problemas mal condicionados pueden tener malas propiedades en convergencia.
2. El MSD decae linealmente con el tiempo n .

2.2.3 Curva de aprendizaje

Además de comprobar que el vector de los pesos del filtro converge tanto en media como en media cuadrática, es interesante asimismo estudiar cómo evoluciona la diferencia entre la señal deseada $d(n)$ y la señal estimada. Con objeto de simplificar las operaciones matemáticas que aparecerán a continuación, esta diferencia vendrá representada por el error a priori $\xi(n)$, que teniendo en cuenta el modelo regresivo utilizado se puede reescribir como

$$\begin{aligned} \xi(n) &= e_0(n) - [\mathbf{w}^t(n-1) - \mathbf{w}_0^t] \mathbf{x}(n) = \\ &= e_0(n) - \mathbf{\Delta} \mathbf{w}^t(n-1) \mathbf{x}(n) \end{aligned} \quad (2.35)$$

Con esto, la curva de aprendizaje, o el error cuadrático medio, definida como $J(n) = E[|\xi(n)|^2]$ toma la forma:

$$J(n) = E[|e_0(n)|^2] + E[\mathbf{x}^t(n)\Delta\mathbf{w}(n-1)\mathbf{\Delta}w^t(n-1)\mathbf{x}(n)] - 2E[\Delta\mathbf{w}^t(n-1)\mathbf{x}(n)e_0(n)] \quad (2.36)$$

El valor de cada uno de los tres sumandos del miembro de la derecha es:

1. El primer valor esperado no es más que la varianza del ruido de medición σ_0^2 .
2. El segundo de ellos se puede modificar, obteniéndose:

$$\begin{aligned} E[\mathbf{x}^t(n)\Delta\mathbf{w}(n-1)\mathbf{\Delta}w^t(n-1)\mathbf{x}(n)] &= \\ &= \text{Tr}\{E[\mathbf{x}(n)\mathbf{x}^t(n)]E[\Delta\mathbf{w}(n-1)\mathbf{\Delta}w^t(n-1)]\} = \\ &= \text{Tr}\{\mathbf{R}\mathbf{K}(n-1)\} \end{aligned} \quad (2.37)$$

3. Para determinar el valor del tercer sumando se recurre al hecho de que tanto $\mathbf{x}(n)$ como $e_0(n)$ son independientes de $\Delta\mathbf{w}(n-1)$ a la luz de las hipótesis de independencia ya enunciadas con anterioridad. Por tanto el valor esperado bajo estudio se descompone en dos:

$$E[\Delta\mathbf{w}^t(n-1)\mathbf{x}(n)e_0(n)] = E[\Delta\mathbf{w}^t(n-1)]E[\mathbf{x}(n)e_0(n)] \quad (2.38)$$

Teniendo en cuenta el principio de ortogonalidad se concluye que este valor esperado es cero.

Antes de continuar merece la pena hacer un comentario: El análisis detallado en este capítulo, como ya es conocido, se basa en lo expuesto por Haykin [Hay96]. Cuando estudiaba la convergencia en media supuso que $e_0(n)$ era independiente de $\mathbf{x}(i) \forall i$, sin embargo ahora afirma que $e_0(n)$ no es independiente de $\mathbf{x}(n)$ tal y como se puede apreciar si se reordena la expresión (2.24), por eso tiene que recurrir al principio de ortogonalidad. Aparte de esta contradicción (proveniente con toda probabilidad de usar fuentes distintas) también se puede destacar que mientras que en el estudio de la convergencia en media sólo emplea herramientas propias de mínimos cuadrados, aquí mezcla herramientas provenientes de la estimación de series aleatorias como es el principio de ortogonalidad.

Por tanto, teniendo en cuenta los puntos 1-3 anteriores se llega a:

$$J(n) = \sigma_0^2 + E[\mathbf{R}\mathbf{K}(n-1)] \quad (2.39)$$

y teniendo en cuenta las expresiones (2.32) y (2.33):

$$J(n) = \sigma_0^2 + \frac{q\sigma_0^2}{n-q-1} \quad (2.40)$$

A partir de esta expresión se pueden hacer los siguientes comentarios:

1. El MSE decae como $1/n$ a su valor en convergencia σ_0^2 . Este valor en convergencia es el mínimo valor posible J_{min} ya que proviene del error de medición y este no se puede eliminar. A la diferencia entre $J(n)$ y J_{min} se le conoce como error cuadrático medio de exceso $J_{exc}(n)$; pues bien, el algoritmo RLS no presenta MSE de exceso en convergencia. Nótese que otro algoritmo muy utilizado, el de mínima media cuadrática LMS (least mean squares) sí presenta este término de exceso.
2. La convergencia del MSE para el algoritmo RLS no depende de la dispersión de autovalores de la matriz \mathbf{R} .

2.3 Análisis de la convergencia en ambientes no estacionarios

En este apartado se supondrá que la respuesta deseada es la salida de un filtro transversal lineal cuyos coeficientes no son constantes. Bajo estas circunstancias el algoritmo adaptativo no sólo tiene que conseguir la convergencia del problema sino también tiene que ser capaz de seguir las variaciones temporales del filtro. Por tanto $d(n)$ viene ahora dada por:

$$d(n) = \mathbf{w}_0^t(n-1)\mathbf{x}(n) + e_0(n) \quad (2.41)$$

La forma en la que cambian los coeficientes viene determinada por un modelo de Markov de orden 1:

$$\mathbf{w}_0(n) = a\mathbf{w}_0(n-1) + \boldsymbol{\omega}(n) \quad (2.42)$$

donde a es un parámetro fijo del modelo y $\boldsymbol{\omega}(n)$ es el vector de ruido del proceso que se supone de media cero y matriz de correlación \mathbf{Q} . El parámetro a se supondrá lo suficientemente cercano a 1 y el ruido del proceso con varianza suficientemente pequeña como para que la evolución temporal del sistema sea lenta y pueda ser rastreada por el algoritmo adaptativo.

Como viene siendo habitual se aceptan unas hipótesis de independencia con el fin de hacer la matemática más tratable:

1. El vector de ruido del proceso $\boldsymbol{\omega}(n)$ es independiente tanto del vector de entrada $\mathbf{x}(n)$ como el ruido de medición $e_0(n)$.
2. El vector de entrada y el ruido de medición son independientes el uno del otro.
3. El ruido de medición es blanco de media cero y varianza σ_0^2 .

Hechas estas aclaraciones se puede proceder con el análisis propiamente dicho.

2.3.1 Capacidad de rastreo

La ecuación de partida es la que actualiza el vector de los pesos del filtro, reproducida aquí por conveniencia:

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \Phi^{-1}(n)\mathbf{x}(n)\xi(n) \quad (2.43)$$

Restando $\mathbf{w}_0(n)$ a ambos miembros se puede construir el vector de error en los pesos:

$$\begin{aligned} \Delta\mathbf{w}(n) = & [\mathbf{I} - \Phi^{-1}(n)\mathbf{x}(n)\mathbf{x}^t(n)]\Delta\mathbf{w}(n-1) + \Phi^{-1}(n)\mathbf{x}(n)e_0(n) + \\ & + (1-a)\mathbf{w}_0(n-1) - \boldsymbol{\omega}(n) \end{aligned} \quad (2.44)$$

Tal y como se comentó $a \sim 1$ de manera que el término $(1-a)\mathbf{w}_0(n-1)$ se puede despreciar, quedando la siguiente ecuación en diferencias que describe el comportamiento del modelo:

$$\begin{aligned} \Delta\mathbf{w}(n) = & [\mathbf{I} - \Phi^{-1}(n)\mathbf{x}(n)\mathbf{x}^t(n)]\Delta\mathbf{w}(n-1) + \Phi^{-1}(n)\mathbf{x}(n)e_0(n) - \\ & - \boldsymbol{\omega}(n) \end{aligned} \quad (2.45)$$

Antes de continuar es necesario encontrar una aproximación a la matriz $\Phi^{-1}(n)$ que haga posible proseguir con el análisis. La matriz $\Phi(n)$ está definida por la expresión (2.4) de manera que si las series $\mathbf{x}(i)$ son aleatorias también lo será ella. Como magnitud aleatoria se puede descomponer en una componente determinista, dada por su media, más una componente aleatoria que modela las desviaciones instantáneas respecto de la media:

$$\Phi(n) = \frac{\mathbf{R}}{1-\lambda} + \mathbf{R}^e(n) \quad (2.46)$$

donde la media $\mathbf{R}/(1-\lambda)$ se ha obtenido de aplicar el operador valor esperado a (2.4) (suponiendo que los $\mathbf{x}(i)$ son estacionarios) y $\mathbf{R}^e(n)$ constituye la componente aleatoria mencionada. La igualdad anterior se cumplirá más fácilmente conforme mayor sea n siempre y cuando se supongan series ergódicas². Llegará por tanto un momento en el que la componente aleatoria sea muy pequeña comparada con la determinista, por lo que la aproximación puede llegar un poco más lejos y suponer que

$$\Phi(n) \approx \frac{\mathbf{R}}{1-\lambda} \quad (2.47)$$

con lo que un valor aproximado de su inversa podría ser:

$$\Phi^{-1}(n) \approx (1-\lambda)\mathbf{R}^{-1} \quad (2.48)$$

Utilizando esta última aproximación en (2.45) queda

$$\begin{aligned} \Delta\mathbf{w}(n) = & [\mathbf{I} - (1-\lambda)\mathbf{R}^{-1}\mathbf{x}(n)\mathbf{x}^t(n)]\Delta\mathbf{w}(n-1) + \\ & + (1-\lambda)\mathbf{R}^{-1}\mathbf{x}(n)e_0(n) - \boldsymbol{\omega}(n) \end{aligned} \quad (2.49)$$

²El concepto de ergodicidad se discute en el apartado 5.3.1 en la página 98.

Para resolver esta ecuación en diferencias se va a recurrir al hecho de que como $\lambda \sim 1$ el factor $1 - \lambda$ es prácticamente nulo. Esto quiere decir que el coeficiente que multiplica a $\Delta \mathbf{w}(n-1)$ tiene una componente aleatoria muy pequeña y por tanto es lícito aplicar la técnica del promediado directo, ya que la solución a la ecuación en diferencias se comportará de forma muy similar a la solución de la siguiente ecuación:

$$\Delta \mathbf{w}(n) = \lambda \Delta \mathbf{w}(n-1) + (1 - \lambda) \mathbf{R}^{-1} \mathbf{x}(n) e_0(n) - \boldsymbol{\omega}(n) \quad (2.50)$$

donde el factor entre corchetes se ha sustituido por su media, de ahí el nombre de promediado directo. A partir de ella se puede calcular la correspondiente ecuación en diferencias para la matriz de correlación de $\Delta \mathbf{w}(n)$ utilizando las conocidas hipótesis de independencia:

$$\mathbf{K}(n) = \lambda^2 \mathbf{K}(n-1) + (1 - \lambda)^2 \sigma_0^2 \mathbf{R}^{-1} + \mathbf{Q} \quad (2.51)$$

Hay que tener en cuenta que cuando comienza expresamente la fase de rastreo es cuando la convergencia hacia el estado estacionario ha concluido. Para estudiar por tanto el MSD debido a la no estacionariedad del sistema es necesario resolver la ecuación anterior para n muy grandes. Una forma de conseguir esto es suponer $n \rightarrow \infty$ con lo que se obtiene

$$\mathbf{K}(\infty) = \frac{1 - \lambda}{2} \sigma_0^2 \mathbf{R}^{-1} + \frac{1}{2(1 - \lambda)} \mathbf{Q} \quad (2.52)$$

Aunque esta solución es estrictamente válida en el límite especificado, se puede suponer aceptable para n suficientemente grandes:

$$\mathbf{K}(n) = \frac{1 - \lambda}{2} \sigma_0^2 \mathbf{R}^{-1} + \frac{1}{2(1 - \lambda)} \mathbf{Q} \quad (2.53)$$

Como se recordará la MSD, $D(n)$, se obtenía sin más que calcular la traza de esta matriz:

$$D(n) = \frac{1 - \lambda}{2} \sigma_0^2 \text{Tr}\{\mathbf{R}^{-1}\} + \frac{1}{2(1 - \lambda)} \text{Tr}\{\mathbf{Q}\} \quad (2.54)$$

El primer sumando es el MSD causado porque el ruido de estimación no se puede eliminar en ambientes no estacionarios al trabajar con una memoria finita $\lambda < 1$. El segundo sumando se debe a la variación temporal de los coeficientes del filtro. Es de resaltar que mientras que la primera contribución es función de $(1 - \lambda)$ la segunda lo es de $1/(1 - \lambda)$, es decir, el ruido de la variación temporal de los coeficientes afecta mucho más que el ruido de medición. Como era de esperar si se intenta reducir la primera (haciendo $\lambda = 1$) la segunda diverge: son dos situaciones encontradas. La situación óptima de mínimo MSD se obtiene cuando el factor de olvido toma el valor,

$$\lambda_{opt} = 1 - \left(\frac{\sigma_0^2 \text{Tr}\{\mathbf{Q}\}}{\text{Tr}\{\mathbf{R}^{-1}\}} \right)^{1/2} \quad (2.55)$$

De forma similar se puede proceder para calcular el desajuste. Recuérdese que el desajuste $M(n)$ se definía como:

$$M(n) = \frac{J(n) - J_{min}}{J_{min}} \quad (2.56)$$

y ya que en este caso $J_{min} = \sigma_0^2$ resulta, utilizando (2.39):

$$M(n) = \frac{\text{Tr}\{\mathbf{R}\mathbf{K}(n-1)\}}{\sigma_0^2} \quad (2.57)$$

Por tanto, para calcular el desajuste, es necesario multiplicar la expresión (2.53) por la izquierda por \mathbf{R} :

$$\mathbf{R}\mathbf{K}(n-1) = \frac{1-\lambda}{2}\sigma_0^2\mathbf{I} + \frac{1}{2(1-\lambda)}\mathbf{R}\mathbf{Q} \quad (2.58)$$

Por lo que el desajuste queda:

$$M(n) = \frac{1-\lambda}{2}q + \frac{1}{2(1-\lambda)\sigma_0^2}\text{Tr}\{\mathbf{R}\mathbf{Q}\} \quad (2.59)$$

A la vista de esta ecuación se pueden comentar ciertos aspectos de cada sumando:

Primer sumando: Este es el desajuste debido al ruido de medición. Nótese que aumenta linealmente con el número de incógnitas q y con el factor $1-\lambda$.

Segundo sumando: Es el desajuste debido a la evolución temporal de los coeficientes del filtro. Está normalizado por el ruido de medición y por el factor $1-\lambda$.

Como se puede apreciar de nuevo, elegir el factor de olvido de manera que disminuya un sumando específico del desajuste hace automáticamente que aumente el otro, ya que dan cuenta de fenómenos opuestos. De nuevo es posible calcular el factor de olvido que produce el desajuste mínimo:

$$\lambda_{opt} = 1 - \left(\frac{\text{Tr}\{\mathbf{R}\mathbf{Q}\}}{\sigma_0^2 q} \right)^{1/2} \quad (2.60)$$

Lo que es más aún, el factor de olvido que hace mínimo el desajuste no coincide con el que hace mínimo el MSD. Una cosa es intentar identificar los coeficientes del filtro lo mejor posible y otra distinta es intentar que la diferencia entre la señal deseada y la estimada sea la menor posible.

Capítulo 3

El algoritmo recursivo de variable instrumental RIV

Introducción

La principal limitación de RLS es que su uso está restringido a situaciones donde la matriz del sistema de ecuaciones a resolver sea la matriz de autocorrelación de cierto vector. Esto reduce su aplicabilidad puesto que en algunas situaciones tales como en el problema de identificación ciega, surgen matrices que son la correlación cruzada entre dos vectores. En estos casos se aplica el algoritmo recursivo de variable instrumental RIV. Que es una generalización a estas situaciones del RLS.

En el apartado 3.1 se obtiene el algoritmo RIV y se analiza en ambientes estacionarios en el apartado 3.2 donde se estudia su convergencia en media y en media cuadrática. Un análisis análogo para ambientes no estacionarios se lleva a cabo en el apartado 3.3 donde además se complementa con un estudio acerca de las condiciones bajo las cuales el error cuadrático medio es óptimo en ambientes no estacionarios ruidosos.

Las aportaciones propias en este capítulo son las siguientes: en primer lugar el análisis de la convergencia de la desviación en media cuadrática en presencia de ruido en ambientes estacionarios, y no estacionarios sin ruido; en segundo lugar, conjugando todo lo desarrollado, se estudia el problema de la convergencia óptima.

3.1 Obtención del algoritmo

El principal objetivo que se persigue con los algoritmos que se están presentando en esta primera parte de la Memoria es el de resolver ciertas ecuaciones matriciales con una estructura bien definida. Por ahora no interesa cómo han surgido estas ecuaciones o qué problema resuelven.

El algoritmo RLS era capaz de obtener, de una manera eficiente, la solución de

$$\Phi(n)\mathbf{w}(n) = \mathbf{z}(n) \quad (3.1)$$

donde la matriz $\Phi(n)$ era la autocorrelación de cierto vector de datos y $\mathbf{z}(n)$ es la correlación entre ese vector de datos y cierta señal ‘deseada’. Esta estructura añadía mucha simetría a las magnitudes implicadas lo cual facilitaba la resolución algebraica del problema. Sin embargo, en otras situaciones puede que se necesite resolver un sistema parecido pero donde la matriz $\Phi(n)$ sea la correlación cruzada entre el vector de datos y un vector auxiliar y el vector $\mathbf{z}(n)$ sea la correlación entre el vector auxiliar y la respuesta deseada. Esta nueva situación, aun presentando muchas simetrías, requiere de una resolución distinta, que generalice la tarea del RLS. A tal fin se propone el algoritmo recursivo de variable instrumental RIV derivado a continuación (las bases de este capítulo se pueden encontrar en [Swa96]).

Según lo comentado, se definen

$$\Phi(n) = \sum_{i=1}^n \lambda^{n-i} \tilde{\mathbf{x}}(n) \mathbf{x}^t(n) \quad (3.2)$$

$$\mathbf{z}(n) = \sum_{i=1}^n \lambda^{n-i} \tilde{\mathbf{x}}(n) d(n) \quad (3.3)$$

donde los vectores se construyen de la siguiente manera: $\tilde{\mathbf{x}}(n) = [\tilde{x}(n), \tilde{x}(n-1), \dots, \tilde{x}(n-q+1)]^t$ y $\mathbf{x}(n) = [x(n), x(n-1), \dots, x(n-q+1)]^t$ que como se puede apreciar tienen la misma longitud y dan lugar a una matriz de correlación cruzada cuadrada. El objetivo ya conocido es, dadas las series $\{\tilde{x}(n)\}$, $\{x(n)\}$ y $\{d(n)\}$, obtener el vector $\mathbf{w}(n)$. De las expresiones (3.2) y (3.3) se obtienen directamente las ecuaciones de actualización correspondientes:

$$\Phi(n) = \lambda \Phi(n-1) + \tilde{\mathbf{x}}(n) \mathbf{x}^t(n) \quad (3.4)$$

$$\mathbf{z}(n) = \lambda \mathbf{z}(n-1) + \tilde{\mathbf{x}}(n) d(n) \quad (3.5)$$

Al igual que para el algoritmo RLS, para el RIV el lema de inversión matricial (apéndice A) juega un papel primordial. Se puede aplicar a la expresión (3.4)

llevando a cabo las siguientes identificaciones:

$$\begin{aligned} \mathbf{A} &= \Phi(n) \\ \mathbf{B}^{-1} &= \lambda\Phi(n-1) \\ \mathbf{C} &= \tilde{\mathbf{x}}(n) \\ \mathbf{D} &= 1 \\ \mathbf{E} &= \mathbf{x}(n) \end{aligned}$$

con lo que resulta:

$$\Phi^{-1}(n) = \lambda^{-1}\Phi^{-1}(n-1) - \lambda^{-2} \frac{\Phi^{-1}(n-1)\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)\Phi^{-1}(n-1)}{1 + \lambda^{-1}\mathbf{x}^t(n)\Phi^{-1}(n-1)\tilde{\mathbf{x}}(n)} \quad (3.6)$$

Esta última expresión será de ayuda para encontrar la ecuación de actualización de $\mathbf{w}(n)$. En primer lugar, de la ecuación (3.1) se obtiene sin más que:

$$\mathbf{w}(n) = \Phi^{-1}(n)\mathbf{z}(n) \quad (3.7)$$

donde se ha supuesto que la matriz $\Phi(n)$ es de rango completo. Si se sustituye la expresión (3.5) en esta última ecuación se obtiene:

$$\mathbf{w}(n) = \Phi^{-1}(n)\lambda\mathbf{z}(n-1) + \Phi^{-1}(n)\tilde{\mathbf{x}}(n)d(n) \quad (3.8)$$

Para continuar se sustituye el valor de $\Phi^{-1}(n)$, sólo en el primer sumando, por su equivalente dado en (3.6):

$$\begin{aligned} \mathbf{w}(n) &= \left[\lambda^{-1}\Phi^{-1}(n-1) - \lambda^{-2} \frac{\Phi^{-1}(n-1)\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)\Phi^{-1}(n-1)}{1 + \lambda^{-1}\mathbf{x}^t(n)\Phi^{-1}(n-1)\tilde{\mathbf{x}}(n)} \right] \\ &\quad \lambda\mathbf{z}(n-1) + \Phi^{-1}(n)\tilde{\mathbf{x}}(n)d(n) \end{aligned} \quad (3.9)$$

Realizando el producto indicado y teniendo en cuenta la ecuación (3.1) la expresión anterior se puede reescribir como sigue,

$$\begin{aligned} \mathbf{w}(n) &= \mathbf{w}(n-1) + \Phi^{-1}(n)\tilde{\mathbf{x}}(n)d(n) - \\ &\quad - \frac{\lambda^{-1}\Phi^{-1}(n-1)\tilde{\mathbf{x}}(n)}{1 + \lambda^{-1}\mathbf{x}^t(n)\Phi^{-1}(n-1)\tilde{\mathbf{x}}(n)} \mathbf{x}^t(n)\mathbf{w}(n-1) \end{aligned} \quad (3.10)$$

Para continuar sería conveniente encontrar una expresión alternativa para el producto $\Phi^{-1}(n)\tilde{\mathbf{x}}(n)$, la cual se obtiene multiplicando por la derecha ambos miembros de (3.6) por $\tilde{\mathbf{x}}(n)$:

$$\begin{aligned} \Phi^{-1}(n)\tilde{\mathbf{x}}(n) &= \lambda^{-1}\Phi^{-1}(n-1)\tilde{\mathbf{x}}(n) - \\ &\quad - \lambda^{-2} \frac{\Phi^{-1}(n-1)\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)\Phi^{-1}(n-1)\tilde{\mathbf{x}}(n)}{1 + \lambda^{-1}\mathbf{x}^t(n)\Phi^{-1}(n-1)\tilde{\mathbf{x}}(n)} \end{aligned} \quad (3.11)$$

y sacando factor común en el miembro de la derecha se puede escribir que

$$\Phi^{-1}(n)\tilde{\mathbf{x}}(n) = \lambda^{-1}\Phi^{-1}(n-1)\tilde{\mathbf{x}}(n) \left[1 - \frac{\lambda^{-1}\mathbf{x}^t(n)\Phi^{-1}(n-1)\tilde{\mathbf{x}}(n)}{1 + \lambda^{-1}\mathbf{x}^t(n)\Phi^{-1}(n-1)\tilde{\mathbf{x}}(n)} \right] \quad (3.12)$$

y una vez realizada la resta dentro del corchete se llega a la expresión buscada:

$$\Phi^{-1}(n)\tilde{\mathbf{x}}(n) = \frac{\lambda^{-1}\Phi^{-1}(n-1)\tilde{\mathbf{x}}(n)}{1 + \lambda^{-1}\mathbf{x}^t(n)\Phi^{-1}(n-1)\tilde{\mathbf{x}}(n)} \quad (3.13)$$

Teniendo en cuenta esta expresión se puede retomar la ecuación (3.10) para reescribirla tal y como sigue:

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \Phi^{-1}(n)\tilde{\mathbf{x}}(n)[d(n) - \mathbf{x}^t(n)\mathbf{w}(n-1)] \quad (3.14)$$

o en forma más compacta y familiar:

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \mathbf{k}(n)e(n) \quad (3.15)$$

en función de un vector de ganancia $\mathbf{k}(n)$ y del correspondiente error a priori, definidos de la siguiente manera:

$$\mathbf{k}(n) = \Phi^{-1}(n)\tilde{\mathbf{x}}(n) \quad (3.16)$$

$$e(n) = d(n) - \mathbf{x}^t(n)\mathbf{w}(n-1) \quad (3.17)$$

La expresión (3.16) indica que los dos miembros de la ecuación (3.13) no son más que dos definiciones equivalentes del vector de ganancia. Por tanto también se puede afirmar que

$$\mathbf{k}(n) = \frac{\lambda^{-1}\Phi^{-1}(n-1)\tilde{\mathbf{x}}(n)}{1 + \lambda^{-1}\mathbf{x}^t(n)\Phi^{-1}(n-1)\tilde{\mathbf{x}}(n)} \quad (3.18)$$

La expresión (3.18) permite reescribir la ecuación de actualización de la inversa de la matriz $\Phi(n)$, ecuación (3.6), de la siguiente manera,

$$\Phi^{-1}(n) = \lambda^{-1}\Phi^{-1}(n-1) - \lambda^{-1}\mathbf{k}(n)\mathbf{x}^t(n)\Phi^{-1}(n-1) \quad (3.19)$$

Con las expresiones dadas hasta aquí se puede construir un método recursivo para el cálculo del vector $\mathbf{w}(n)$ una vez conocido $\mathbf{w}(n-1)$ y con la llegada de nuevos datos $\tilde{\mathbf{x}}(n)$, $\mathbf{x}(n)$ y $d(n)$. Este método recursivo se conoce como RIV y se resume en la tabla 3.1.

<p><u>Condiciones iniciales</u></p> $\Phi^{-1}(0) = \delta^{-1}\mathbf{I} \quad \delta = \text{constante positiva pequeña}$ $\mathbf{w}(0) = \mathbf{0}$ <p><u>Cuerpo del algoritmo</u></p> $\mathbf{k}(n) = \frac{\lambda^{-1}\Phi^{-1}(n-1)\tilde{\mathbf{x}}(n)}{1 + \lambda^{-1}\mathbf{x}^t(n)\Phi^{-1}(n-1)\tilde{\mathbf{x}}(n)}$ $e(n) = d(n) - \mathbf{w}^t(n-1)\mathbf{x}(n)$ $\mathbf{w}(n) = \mathbf{w}(n-1) + \mathbf{k}(n)e(n)$ $\Phi^{-1}(n) = \lambda^{-1}\Phi^{-1}(n-1) - \lambda^{-1}\mathbf{k}(n)\mathbf{x}^t(n)\Phi^{-1}(n-1)$
--

Tabla 3.1 *Pasos del algoritmo RIV*

Como se puede apreciar en la tabla 3.1 la única diferencia con el algoritmo RLS, presentado en la tabla 2.1, es que en el cálculo del vector de ganancia $\mathbf{k}(n)$ aparece la variable instrumental $\tilde{\mathbf{x}}(n)$.

Una vez derivado el algoritmo RIV se procederá, en los próximos apartados, al estudio de su convergencia tanto en ambientes estacionarios como no estacionarios.

3.2 Convergencia en ambientes estacionarios

El primer paso importante en el análisis del algoritmo RIV es el de estudiar sus propiedades en ambientes estacionarios. Por estacionario se entiende que la estadística de las series $\{\tilde{\mathbf{x}}(n)\}$, $\{\mathbf{x}(n)\}$ y $\{d(n)\}$ no cambia con el tiempo. El análisis consiste en el estudio de la convergencia en media y en media cuadrática, tanto en presencia de ruido como sin él. Se comenzará con el estudio de la convergencia en media.

3.2.1 Convergencia en media. Efecto de las condiciones iniciales

Una cuestión importante es determinar a qué valor converge, en el proceso recursivo, la serie $\mathbf{w}(1)$, $\mathbf{w}(2)$, ..., $\mathbf{w}(n)$ y comprobar que ese valor es efectivamente el que se está buscando.

Obviamente la convergencia se obtiene en el límite $n \rightarrow \infty$ y es en este límite cuando hay que estudiar el comportamiento de la ecuación (3.7). Una de sus principales características es que tanto $\Phi(n)$ como $\mathbf{z}(n)$ están definidas como promedios temporales, de manera que cuanto más datos se conozcan, su valor se aproximará más al de su valor esperado, es decir, se pueden escribir como:

$$\Phi(n) = \frac{1}{1-\lambda} \mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}} + \mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}^e(n) \quad (3.20)$$

$$\mathbf{z}(n) = \frac{1}{1-\lambda} \mathbf{r}_{\tilde{\mathbf{x}}d} + \mathbf{r}_{\tilde{\mathbf{x}}d}^e(n) \quad (3.21)$$

donde $\mathbf{R}_{\tilde{x}x} = E[\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)]$ y $\mathbf{r}_{\tilde{x}d} = E[\tilde{\mathbf{x}}(n)d(n)]$. Los otros sumandos adicionales que aparecen $\mathbf{R}_{\tilde{x}x}^e(n)$ y $\mathbf{r}_{\tilde{x}d}^e(n)$ se pueden modelar como de media nula e independientes del resto de variables aleatorias involucradas y dan cuenta del ruido de estimación. Es de esperar que conforme el número de datos disponibles sea mayor, el promedio temporal funcionará mejor como estimador del operador valor esperado y por tanto los ruidos de estimación sean menores:

$$\|\mathbf{R}_{\tilde{x}x}^e(n)\| \ll \|\mathbf{R}_{\tilde{x}x}\| \quad (3.22)$$

A una matriz cuya descomposición en parte determinista y aleatoria cumpla esta condición se le denomina cuasideterminista. Lo mismo se le puede aplicar al vector $\mathbf{z}(n)$.

La inversa de la expresión (3.20) es:

$$\Phi^{-1}(n) = (1 - \lambda)[1 + (1 - \lambda)\mathbf{R}^{-1}\mathbf{R}^e(n)]^{-1}\mathbf{R}^{-1} \quad (3.23)$$

donde se han eliminado los subíndices, como se hará de aquí en adelante siempre y cuando no haya lugar a confusión. Teniendo en cuenta que $\Phi(n)$ es cuasideterminista:

$$\Phi^{-1}(n) = (1 - \lambda)[1 - (1 - \lambda)\mathbf{R}^{-1}\mathbf{R}^e(n)]\mathbf{R}^{-1} \quad (3.24)$$

Teniendo en cuenta las expresiones (3.24) y (3.21) en (3.7) se llega a:

$$\mathbf{w}(n) = (1 - \lambda)[1 - (1 - \lambda)\mathbf{R}^{-1}\mathbf{R}^e(n)]\mathbf{R}^{-1} \left[\frac{1}{1 - \lambda}\mathbf{r} + \mathbf{r}^e(n) \right] \quad (3.25)$$

realizando los productos indicados:

$$\begin{aligned} \mathbf{w}(n) = & \mathbf{R}^{-1}\mathbf{r} + (1 - \lambda)\mathbf{R}^{-1}\mathbf{r}^e(n) - (1 - \lambda)\mathbf{R}^{-1}\mathbf{R}^e(n)\mathbf{R}^{-1}\mathbf{r} - \\ & - (1 - \lambda)^2\mathbf{R}^{-1}\mathbf{R}^e(n)\mathbf{R}^{-1}\mathbf{r}^e(n) \end{aligned} \quad (3.26)$$

Como se aprecia en esta expresión, el estimador $\mathbf{w}(n)$ se comporta para n suficientemente grandes como el vector $\mathbf{R}^{-1}\mathbf{r}$ más otras contribuciones aleatorias de media nula. En definitiva, en media, se puede afirmar que

$$E[\mathbf{w}(n)] \underset{n \rightarrow \infty}{\sim} \mathbf{w}_0 \quad (3.27)$$

donde el vector \mathbf{w}_0 es la solución a la ecuación

$$\mathbf{R}\mathbf{w}_0 = \mathbf{r} \quad (3.28)$$

Como consecuencia de todo este comentario se puede afirmar que resolver la ecuación (3.1) es equivalente, bajo las condiciones de n grande y tomando valor medio, a resolver la ecuación (3.28). Veámos pues cómo afecta, a la convergencia de la serie $\{E[\mathbf{w}(n)]\}$ obtenida mediante RIV, las condiciones iniciales fijadas para este algoritmo.

Efecto de las condiciones iniciales

Se parte de la ecuación de actualización de los pesos $\mathbf{w}(n)$ dada en (3.14) y complementada por (3.17):

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \Phi^{-1}(n)\tilde{\mathbf{x}}(n)e(n) \quad (3.29)$$

El error a priori de (3.17) se puede reescribir como

$$\begin{aligned} e(n) &= d(n) - \mathbf{x}^t(n)\mathbf{w}(n-1) = \\ &= e_0(n) + \mathbf{x}^t(n)\mathbf{w}_0 - \mathbf{x}^t(n)\mathbf{w}(n-1) \end{aligned} \quad (3.30)$$

donde se ha supuesto que la respuesta deseada $d(n)$ se puede modelar como la salida de un filtro lineal dado por los coeficientes \mathbf{w}_0 , con entrada $\{x(n)\}$ y contaminado por ruido $e_0(n)$ de media nula. Esta hipótesis está fundamentada en el hecho de que la mayor parte de las aplicaciones de los algoritmos adaptativos consisten en estimar la respuesta $d(n)$ mediante $\{x(n)\}$ mediante regresión lineal. El término $e_0(n)$ constituiría de hecho una especie de ruido de medición, ya que impide la perfecta estimación de $d(n)$ mediante el modelo de regresión.

Por tanto, combinando (3.29) con (3.30), se obtiene

$$\Delta\mathbf{w}(n) = [\mathbf{I} - \Phi^{-1}(n)\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)]\Delta\mathbf{w}(n-1) + \Phi^{-1}(n)\tilde{\mathbf{x}}(n)e_0(n) \quad (3.31)$$

donde el vector de error en los pesos se ha definido como $\Delta\mathbf{w}(n) = \mathbf{w}(n) - \mathbf{w}_0$. Si se multiplica por la izquierda toda la expresión por $\Phi(n)$, teniendo en cuenta (3.4), queda:

$$\Phi(n)\Delta\mathbf{w}(n) = \lambda\Phi(n-1)\Delta\mathbf{w}(n-1) + \tilde{\mathbf{x}}(n)e_0(n) \quad (3.32)$$

que constituye una ecuación en diferencias para el factor $\Phi(n)\Delta\mathbf{w}(n)$ con un coeficiente constante e igual a λ y un término independiente aleatorio. Su solución es:

$$\Phi(n)\Delta\mathbf{w}(n) = \lambda^n\Phi(0)\Delta\mathbf{w}(0) + \sum_{i=1}^n \lambda^{n-i}\tilde{\mathbf{x}}(i)e_0(i) \quad (3.33)$$

o despejando para el vector de error en los pesos:

$$\Delta\mathbf{w}(n) = \lambda^n\Phi^{-1}(n)\Phi(0)\Delta\mathbf{w}(0) + \Phi^{-1}(n)\sum_{i=1}^n \lambda^{n-i}\tilde{\mathbf{x}}(i)e_0(i) \quad (3.34)$$

Para continuar se toma como válida la aproximación dada en (3.24) resultando

$$\begin{aligned} \Delta\mathbf{w}(n) &= \lambda^n(1-\lambda)[1 - (1-\lambda)\mathbf{R}^{-1}\mathbf{R}^e(n)]\mathbf{R}^{-1}\Phi(0)\Delta\mathbf{w}(0) + \\ &+ (1-\lambda)[1 - (1-\lambda)\lambda\mathbf{R}^{-1}\mathbf{R}^e(n)]\mathbf{R}^{-1}\sum_{i=1}^n \lambda^{n-i}\tilde{\mathbf{x}}(i)e_0(i) \end{aligned} \quad (3.35)$$

Si se toma valor esperado y se tiene en cuenta que:

1. La matriz $\mathbf{R}^e(n)$ es de media nula e independiente del resto de variables aleatorias.
2. El valor esperado $E[\tilde{\mathbf{x}}(n)e_0(n)]$ equivale a la diferencia $\mathbf{R}\mathbf{w}_0 - \mathbf{r}$ y por tanto es nulo. Esta afirmación se puede considerar como una expresión generalizada del principio de ortogonalidad.

se llega finalmente a

$$E[\Delta\mathbf{w}(n)] = \lambda^n(1 - \lambda)\mathbf{R}^{-1}\Phi(0)\Delta\mathbf{w}(0) \quad (3.36)$$

En primer lugar esta ecuación confirma que para $n \rightarrow \infty$ con $\lambda < 1$ y \mathbf{w}_0 fijo se cumple efectivamente que, en media, el algoritmo converge a \mathbf{w}_0 . Recuérdese que el factor de olvido desempeña su cometido en ambientes no estacionarios, en este caso ya que se trata con ambientes estacionarios debería ser la unidad y el término $(1 - \lambda)$ debería cambiarse por $1/n$ para mantener la normalización en (3.2) y (3.3). Con este cambio el efecto de las condiciones iniciales tiende a cero como $1/n$.

3.2.2 Convergencia en media cuadrática con ruido de estimación

Como es necesario en un análisis de este tipo, no basta con estudiar la convergencia en media sino que hay que comprobar que los estimadores tienen varianza finita, es decir, hay que estudiar la convergencia en media cuadrática. Además, esta vez también hay interés en comprobar cómo afecta el ruido de estimación al error cuadrático medio.

Para empezar se multiplican ambos miembros de la ecuación (3.32) por $\Phi^{-1}(n)$ y se va a considerar que $\Phi^{-1}(n)\Phi(n-1) \approx \mathbf{I}$. Esta aproximación es más cierta conforme mayor es n de forma que ambos promedio temporales se asemejan más a promedios en realizaciones. Con todo (3.32) queda:

$$\Delta\mathbf{w}(n) = \lambda\Delta\mathbf{w}(n-1) + \Phi^{-1}(n)\tilde{\mathbf{x}}(n)e_0(n) \quad (3.37)$$

El error en media cuadrática se suele definir como $q(n) = E[\Delta\mathbf{w}^t(n)\mathbf{Q}\Delta\mathbf{w}(n)]$ donde \mathbf{Q} es cualquier matriz definida positiva. Con la expresión (3.37) y aceptando como válida la aproximación de (3.24) se puede construir el siguiente producto:

$$\begin{aligned} \Delta\mathbf{w}^t(n)\mathbf{Q}\Delta\mathbf{w}(n) &= \lambda^2\Delta\mathbf{w}^t(n-1)\mathbf{Q}\Delta\mathbf{w}(n-1) + \\ &+ 2\lambda(1 - \lambda)\Delta\mathbf{w}^t(n-1)\mathbf{Q}[1 - (1 - \lambda)\mathbf{R}^{-1}\mathbf{R}^e(n)]\mathbf{R}^{-1}\tilde{\mathbf{x}}(n)e_0(n) + \\ &+ (1 - \lambda)^2e_0^2\tilde{\mathbf{x}}^t(n)\mathbf{R}^{-t}[1 - (1 - \lambda)\mathbf{R}^{-1}\mathbf{R}^e(n)]^t\mathbf{Q} \\ &[1 - (1 - \lambda)\mathbf{R}^{-1}\mathbf{R}^e(n)]\mathbf{R}^{-1}\tilde{\mathbf{x}}(n) \end{aligned} \quad (3.38)$$

Al tomar valor esperado hay que tener en cuenta lo siguiente respecto al miembro de la derecha:

Primer sumando: No es más que $\lambda^2 q(n-1)$.

Segundo sumando: Aquí hay que suponer que el vector de error en los pesos a tiempo $n-1$ es independiente de $\tilde{\mathbf{x}}(n)$ por lo que al tomar valor esperado este término es nulo por el principio de ortogonalidad.

Tercer sumando: Para este sumando hay que recordar la independencia de $e_0(n)$ y de $\mathbf{R}^e(n)$ del resto de variables aleatorias, además de que sus medias son nulas. Con esto se llega a:

$$\text{Tercer sumando} \xrightarrow{\text{E}[\cdot]} (1-\lambda)^2 \sigma_0^2 \beta_Q + (1-\lambda)^4 \sigma_0^2 V_e \quad (3.39)$$

donde se han definido las siguientes cantidades:

$$\beta_Q = \text{Tr}\{\mathbf{R}^{-t} \mathbf{Q} \mathbf{R}^{-1} \mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}\} \quad (3.40)$$

$$V_e = \text{Tr}\{\mathbf{R}^{-t} \text{E}[\mathbf{R}^{e t}(n) \mathbf{R}^{-t} \mathbf{Q} \mathbf{R}^{-1} \mathbf{R}^e(n)] \mathbf{R}^{-1} \mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}\} \quad (3.41)$$

Teniendo en cuenta estos tres puntos, el valor esperado de (3.38) se puede escribir como

$$q(n) = \lambda^2 q(n-1) + (1-\lambda)^2 \sigma_0^2 \beta_Q + (1-\lambda)^4 \sigma_0^2 V_e \quad (3.42)$$

Esta ecuación en diferencias con coeficientes constantes tiene un valor en convergencia dado por la siguiente expresión (suponiendo $\lambda \sim 1$):

$$q(\infty) = \frac{1-\lambda}{2} \sigma_0^2 \beta_Q + \frac{(1-\lambda)^3}{2} \sigma_0^2 V_e \quad (3.43)$$

Las contribuciones a la desviación en media cuadrática en convergencia se deben a dos fuentes. La primera de ella proviene del ruido de medición y está pesada por el término β_Q , que al depender de la variable instrumental se puede controlar su contribución con una elección adecuada de la misma. En ambientes estacionarios tiende a cero como $1/n$.

La segunda fuente se debe al ruido de estimación, aunque también depende de la variable instrumental. Para n grandes tiende a cero (en situaciones estacionarias) por dos motivos: porque depende de $1/n^3$ y porque conforme se disponga de más datos en la estimación el ruido debe disminuir.

Tal y como se acaba de comentar, para ambientes estacionarios, el MSD tiende a cero. Es sólo en situaciones no estacionarias, con $\lambda < 1$ donde hay cierta varianza en los estimadores de los pesos. Esto se debe a que para una longitud de la memoria finita los efectos de los ruidos de medición y estimación nunca se podrán eliminar.

3.2.3 Convergencia en media cuadrática con ruido de estimación II

En este apartado se van a repetir los cálculos del apartado anterior pero teniendo en cuenta otra aproximación distinta. Para empezar se retomará la ecuación (3.31) que se verifica exactamente:

$$\Delta \mathbf{w}(n) = [\mathbf{I} - \Phi^{-1}(n)\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)]\Delta \mathbf{w}(n-1) + \Phi^{-1}(n)\tilde{\mathbf{x}}(n)e_0(n) \quad (3.44)$$

Como el interés se centra en magnitudes en convergencia es lícito suponer n suficientemente grande como para aceptar la aproximación de (3.24), con lo que la expresión anterior queda

$$\begin{aligned} \Delta \mathbf{w}(n) = & [\mathbf{I} - (1-\lambda)[1 - (1-\lambda)\mathbf{R}^{-1}\mathbf{R}^e(n)]\mathbf{R}^{-1}\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)]\Delta \mathbf{w}(n-1) + \\ & + (1-\lambda)[1 - (1-\lambda)\mathbf{R}^{-1}\mathbf{R}^e(n)]\mathbf{R}^{-1}\tilde{\mathbf{x}}(n)e_0(n) \end{aligned} \quad (3.45)$$

que constituye una ecuación en diferencias de orden 1 con el coeficiente aleatorio. Este coeficiente aleatorio está formado por una parte determinista (un 1) más la parte aleatoria propiamente dicha afectada por el factor $(1-\lambda)$. Como normalmente el factor de olvido está próximo a uno, la oscilación de la parte aleatoria será muy pequeña con lo que se puede aproximar todo el coeficiente por su valor medio. Con esta aproximación es de esperar que la solución obtenida esté muy cercana a la de la ecuación original. Esta técnica se conoce con el nombre de promediado directo y ya apareció en el tema anterior.

Aplicando promediado directo a (3.45) se obtiene

$$\begin{aligned} \Delta \mathbf{w}(n) = & \lambda \Delta \mathbf{w}(n-1) + (1-\lambda)\mathbf{R}^{-1}\tilde{\mathbf{x}}(n)e_0(n) - \\ & - (1-\lambda)^2\mathbf{R}^{-1}\mathbf{R}^e(n)\mathbf{R}^{-1}\tilde{\mathbf{x}}(n)e_0(n) \end{aligned} \quad (3.46)$$

que es exactamente la misma expresión que se utilizó en el apartado 3.2.2 para obtener la expresión (3.38), por lo que la desviación en media cuadrática obtenida aquí coincidirá con la allí dada. En particular la expresión recursiva para el MSD es:

$$q(n) = \lambda^2 q(n-1) + (1-\lambda)^2 \sigma_0^2 \beta_Q + (1-\lambda)^4 \sigma_0^2 V_e \quad (3.47)$$

Recuérdese que las cantidades β_Q y V_e están definidas en (3.40) y (3.41). En convergencia toma el valor:

$$q(\infty) = \frac{1-\lambda}{2} \sigma_0^2 \beta_Q + \frac{(1-\lambda)^3}{2} \sigma_0^2 V_e \quad (3.48)$$

Por tanto hay que resaltar que a pesar de emplear aproximaciones distintas en el apartado 3.2.2 y en este, los resultados obtenidos son idénticos. Es decir, suponer que $\Phi^{-1}(n)\Phi(n-1) \approx \mathbf{I}$ equivale a aplicar promediado directo. Como es sabido la validez de ambas aproximaciones aumenta al aumentar n .

3.3 Convergencia en ambientes no estacionarios

Una de las principales características de los algoritmos adaptativos es su capacidad de seguir la evolución de los parámetros que identifica cuando estos cambian con el tiempo. En los apartados que siguen se va a suponer que el vector \mathbf{w}_0 ya no es constante sino que es una cierta función del tiempo $\mathbf{w}_0(n)$. Este cambio puede hacer, dependiendo de la aplicación concreta, que las series aleatorias involucradas tengan una estadística no estacionaria. En estos ambientes es cuando se hace imprescindible la existencia de un factor de olvido, que dé más importancia a las muestras presentes, de manera que los estimadores se vayan adaptando progresivamente a los valores actuales.

3.3.1 Capacidad de rastreo

A continuación se va a estudiar si el algoritmo RIV es capaz de seguir, en media, las variaciones que experimente el vector $\mathbf{w}_0(n)$. Para esto se parte de la ecuación de actualización de $\mathbf{w}(n)$ dada en (3.29) y reproducida aquí por conveniencia,

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \Phi^{-1}(n)\tilde{\mathbf{x}}(n)e(n) \quad (3.49)$$

Teniendo en cuenta que la respuesta deseada se expresa ahora como $d(n) = e_0(n) + \mathbf{x}^t(n)\mathbf{w}_0(n)$, el error a priori se puede reescribir como

$$e(n) = e_0(n) + \mathbf{x}^t(n)\mathbf{w}_0(n) - \mathbf{x}^t(n)\mathbf{w}(n-1) \quad (3.50)$$

, con lo que la expresión anterior se convierte en

$$\begin{aligned} \mathbf{w}(n) = [\mathbf{I} - \Phi^{-1}(n)\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)]\mathbf{w}(n-1) + \Phi^{-1}(n)\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)\mathbf{w}_0(n) + \\ + \Phi^{-1}(n)\tilde{\mathbf{x}}(n)e_0(n) \end{aligned} \quad (3.51)$$

La capacidad de rastreo es una propiedad que se hace patente cuando los efectos transitorios debido a las condiciones iniciales han desaparecido, por tanto, si se quiere estudiar será necesario suponer que n es muy grande. Además, para centrarnos en este aspecto de la capacidad de rastreo se ignorará el ruido de estimación. Con estas condiciones se puede dar un paso más y simplificar la expresión (3.24) a:

$$\Phi^{-1}(n) = (1 - \lambda)\mathbf{R}^{-1}(n) \quad (3.52)$$

Donde se ha incluido la posible dependencia temporal de \mathbf{R} al estar trabajando con series cuya estadística puede no ser estacionaria. Teniendo en cuenta esta aproximación en (3.51), al tomar valor esperado resulta

$$\mathbf{E}[\mathbf{w}(n)] = \lambda\mathbf{E}[\mathbf{w}(n-1)] + (1 - \lambda)\mathbf{w}_0(n) \quad (3.53)$$

donde se ha tenido en cuenta la independencia de $\mathbf{w}(n-1)$ con $\tilde{\mathbf{x}}(n)$ y con $\mathbf{x}(n)$ y que $\mathbf{E}[\tilde{\mathbf{x}}(n)e_0(n)] = 0$. La expresión (3.53) es una ecuación en diferencias

no homogénea con coeficientes constantes. Tiene una solución general de la homogénea asociada dada por

$$E[\mathbf{w}(n)] = \lambda^n E[\mathbf{w}(0)] \quad (3.54)$$

y una solución particular de la no homogénea que se puede obtener aplicando la transformada Z a ambos miembros de (3.53):

$$E[\mathbf{w}(Z)] = \lambda z^{-1} E[\mathbf{w}(Z)] + (1 - \lambda) \mathbf{w}_0(Z) \quad (3.55)$$

por lo que

$$E[\mathbf{w}(Z)] = \frac{1 - \lambda}{1 - \lambda z^{-1}} \mathbf{w}_0(Z) \quad (3.56)$$

o expresándola en el dominio del tiempo:

$$E[\mathbf{w}(n)] = (1 - \lambda) \sum_{i=0}^n \lambda^i \mathbf{w}_0(n - i) \quad (3.57)$$

Aunque para obtener una expresión concreta es necesario tener una expresión determinada para $\mathbf{w}_0(n)$, se puede afirmar que cuanto menor es λ (menos memoria) se produce menos contaminación en $E[\mathbf{w}(n)]$ debida a tiempos anteriores de $\mathbf{w}_0(n)$. Por tanto, una memoria corta es imprescindible para una buena capacidad de rastreo.

3.3.2 Convergencia en media cuadrática

En este apartado se va a suponer una forma funcional sencilla para el vector solución dada por

$$\mathbf{w}_0(n) = n\mathbf{p} \quad (3.58)$$

donde \mathbf{p} es un vector constante, que contiene las pendientes de cada componente de la variación lineal. Con esta expresión para $\mathbf{w}_0(n)$, el vector de error a priori se puede reescribir, siguiendo con (3.50), de la siguiente manera:

$$e(n) = e_0(n) - \mathbf{x}^t(n) \Delta \mathbf{w}(n - 1) + \mathbf{x}^t(n) \mathbf{p} \quad (3.59)$$

donde como es habitual el vector de error en los pesos se define como $\Delta \mathbf{w}(n) = \mathbf{w}(n) - \mathbf{w}_0(n)$. Con esta expresión para $e(n)$ la ecuación para la actualización de los pesos (3.29) se puede reescribir como:

$$\begin{aligned} \mathbf{w}(n) = & \mathbf{w}(n - 1) - \Phi^{-1}(n) \tilde{\mathbf{x}}(n) \mathbf{x}^t(n) \Delta \mathbf{w}(n - 1) + \\ & + \Phi^{-1}(n) \tilde{\mathbf{x}}(n) e_0(n) + \Phi^{-1}(n) \tilde{\mathbf{x}}(n) \mathbf{x}^t(n) \mathbf{p} \end{aligned} \quad (3.60)$$

Restando $\mathbf{w}_0(n)$ a cada miembro y teniendo en cuenta (3.58), se llega a:

$$\begin{aligned} \Delta \mathbf{w}(n) = & [\mathbf{I} - \Phi^{-1}(n) \tilde{\mathbf{x}}(n) \mathbf{x}^t(n)] \Delta \mathbf{w}(n - 1) - \\ & - [\mathbf{I} - \Phi^{-1}(n) \tilde{\mathbf{x}}(n) \mathbf{x}^t(n)] \mathbf{p} + \Phi^{-1}(n) \tilde{\mathbf{x}}(n) e_0(n) \end{aligned} \quad (3.61)$$

Como las condiciones de trabajo son las mismas que en el apartado anterior, es decir, interés exclusivo en el periodo tras la convergencia e ignorar el ruido de estimación, se puede volver a adoptar la aproximación (3.52) como válida y tras aplicar promediado directo la expresión anterior se convierte en:

$$\begin{aligned} \Delta \mathbf{w}(n) = & \lambda \Delta \mathbf{w}(n-1) - [\mathbf{I} - (1-\lambda)\mathbf{R}^{-1}(n)\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)]\mathbf{p} + \\ & + (1-\lambda)\mathbf{R}^{-1}(n)\tilde{\mathbf{x}}(n)e_0(n) \end{aligned} \quad (3.62)$$

Aunque en este apartado se estudia la convergencia en media cuadrática, es interesante estudiar, a partir de esta expresión, la convergencia en media, por lo que tomando valor esperado y considerando las hipótesis de independencia usuales junto con la ortogonalidad de $\tilde{\mathbf{x}}(n)$ y $e_0(n)$,

$$\mathbb{E}[\Delta \mathbf{w}(n)] = \lambda \mathbb{E}[\Delta \mathbf{w}(n-1)] - \lambda \mathbf{p} \quad (3.63)$$

dando un valor en convergencia de

$$\mathbb{E}[\Delta \mathbf{w}(\infty)] = \frac{-\lambda}{1-\lambda} \mathbf{p} \quad (3.64)$$

Como se aprecia el vector de error en los pesos no tiende a cero sino a un valor constante no nulo. Si se adopta este valor como válido no sólo en el límite $n \rightarrow \infty$ sino también para n grandes, se puede escribir:

$$\mathbb{E}[\mathbf{w}(n)] = \mathbf{w}_0(n) - \frac{\lambda}{1-\lambda} \mathbf{p} \quad (3.65)$$

o teniendo en cuenta (3.58)

$$\mathbb{E}[\mathbf{w}(n)] = \mathbf{w}_0 \left(n - \frac{\lambda}{1-\lambda} \right) \quad (3.66)$$

es decir, en media el algoritmo RIV sigue la evolución de $\mathbf{w}_0(n)$ pero con un tiempo de retraso τ dado por

$$\tau = \frac{\lambda}{1-\lambda} \quad (3.67)$$

Este tiempo de retraso puede ser idealmente cero si el factor de olvido fuese nulo, con una memoria de una sola muestra. Por el contrario, si la memoria fuese infinita ($\lambda = 1$) el tiempo de retraso también sería infinito, es decir, no tendría ninguna capacidad de rastreo.

Con el fin de retomar la convergencia en media cuadrática se construye, teniendo en cuenta la expresión (3.62), el siguiente producto escalar

$$\begin{aligned} \Delta \mathbf{w}^t(n) \Delta \mathbf{w}(n) = & \lambda^2 \Delta \mathbf{w}^t(n-1) \Delta \mathbf{w}(n-1) - \\ & - 2\lambda \Delta \mathbf{w}^t(n-1) [\mathbf{I} - (1-\lambda)\mathbf{R}^{-1}(n)\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)]\mathbf{p} + \\ & + 2\lambda(1-\lambda) \Delta \mathbf{w}^t(n-1) \mathbf{R}^{-1}(n)\tilde{\mathbf{x}}(n)e_0(n) + \\ & + \mathbf{p}^t [\mathbf{I} - (1-\lambda)\mathbf{R}^{-1}(n)\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)]^t [\mathbf{I} - (1-\lambda)\mathbf{R}^{-1}(n)\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)]\mathbf{p} - \\ & - 2(1-\lambda)\mathbf{p}^t [\mathbf{I} - (1-\lambda)\mathbf{R}^{-1}(n)\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)]^t \mathbf{R}^{-1}(n)\tilde{\mathbf{x}}(n)e_0(n) + \\ & + (1-\lambda)^2 e_0^2(n) \tilde{\mathbf{x}}^t(n) \mathbf{R}^{-t}(n) \mathbf{R}^{-1}(n) \tilde{\mathbf{x}}(n) \end{aligned} \quad (3.68)$$

Tomando valor esperado, teniendo en cuenta las hipótesis de independencia usuales, la expresión (3.65) y el principio de ortogonalidad, resulta:

$$D(n) = \lambda^2 D(n-1) + \left[2(1-\lambda)(1+\lambda) - 5 + \frac{2}{1-\lambda} \right] \mathbf{p}^t \mathbf{p} + (1-\lambda)^2 \mathbf{p}^t \mathbf{H}(n) \mathbf{p} + (1-\lambda)^2 \sigma_0^2 \beta'(n) \quad (3.69)$$

donde el factor entre corchetes se ha expresado como función de $(1-\lambda)$ y $(1+\lambda)$ para poder facilitar las simplificaciones venideras y donde se han definido:

$$D(n) = E[\Delta \mathbf{w}^t(n) \Delta \mathbf{w}(n)] \quad (3.70)$$

$$\mathbf{H}(n) = E[\mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \mathbf{R}^{-t}(n) \mathbf{R}^{-1}(n) \tilde{\mathbf{x}}(n) \mathbf{x}^t(n)] \quad (3.71)$$

$$\beta'(n) = \text{Tr}\{\mathbf{R}^{-t}(n) \mathbf{R}^{-1}(n) \mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}(n)\} \quad (3.72)$$

nótese que $\beta'(n)$ es el equivalente en ambientes no estacionarios de β_I (β_Q con $\mathbf{Q} = \mathbf{I}$). Igualmente se ha utilizado $D(n)$ en vez de $q(n)$ para simplificar la notación.

La solución a la ecuación en diferencias (3.69) es

$$D(n) = \lambda^{2n} D(0) + \left[2(1-\lambda)(1+\lambda) - 5 + \frac{2}{1-\lambda} \right] \sum_{i=0}^{n-1} \lambda^{2i} \mathbf{p}^t \mathbf{p} + (1-\lambda)^2 \sum_{i=0}^{n-1} \lambda^{2i} [\mathbf{p}^t \mathbf{H}(n-i) \mathbf{p} + \sigma_0^2 \beta'(n-i)] \quad (3.73)$$

que en convergencia toma el valor:

$$D(\infty) = \left[2(1-\lambda)(1+\lambda) - 5 + \frac{2}{1-\lambda} \right] \frac{1}{1-\lambda^2} \mathbf{p}^t \mathbf{p} + \frac{1-\lambda}{2} \mathbf{p}^t \mathbf{H} \mathbf{p} + \frac{1-\lambda}{2} \sigma_0^2 \beta' \quad (3.74)$$

donde

$$\mathbf{H} \equiv 2(1-\lambda) \lim_{n \rightarrow \infty} \sum_{i=0}^{n-1} \lambda^{2i} \mathbf{H}(n-i) \quad (3.75)$$

$$\beta' \equiv 2(1-\lambda) \lim_{n \rightarrow \infty} \sum_{i=0}^{n-1} \lambda^{2i} \beta'(n-i) \quad (3.76)$$

Se ha incluido el factor $2(1-\lambda)$ en estas definiciones para que cuando se trabaje en ambientes estacionarios se cumpla que $\beta_Q = \beta'$.

Teniendo en cuenta que $\lambda \sim 1$ la expresión (3.74) se puede escribir finalmente como:

$$D(\infty) = \frac{1}{(1-\lambda)^2} \mathbf{p}^t \mathbf{p} - \frac{5}{2} \frac{1}{(1-\lambda)} \mathbf{p}^t \mathbf{p} + 2 \mathbf{p}^t \mathbf{p} + \frac{1-\lambda}{2} \mathbf{p}^t \mathbf{H} \mathbf{p} + \frac{1-\lambda}{2} \sigma_0^2 \beta' \quad (3.77)$$

3.3.3 Convergencia óptima

Si en la expresión (3.77) se tiene también en cuenta el ruido de estimación, obtenido en la expresión (3.43), se obtiene una expresión para la media cuadrática en convergencia que incluye todos los factores estudiados hasta ahora:

$$D(\infty) = \frac{1}{(1-\lambda)^2} \mathbf{p}^t \mathbf{p} - \frac{5}{2} \frac{1}{(1-\lambda)} \mathbf{p}^t \mathbf{p} + 2 \mathbf{p}^t \mathbf{p} + \frac{1-\lambda}{2} \mathbf{p}^t \mathbf{H} \mathbf{p} + \frac{1-\lambda}{2} \sigma_0^2 \beta' + \frac{(1-\lambda)^3}{2} \sigma_0^2 V_e' \quad (3.78)$$

lo único a tener en cuenta es que V_e' se define como V_e de (3.41) haciendo $\mathbf{Q} = \mathbf{I}$ y dándole el mismo tratamiento que a β' en la ecuación (3.76), es decir, se considera su naturaleza no estacionaria y se introduce en la sumatoria.

Los términos más influyentes son los debidos a la no estacionariedad del sistema, destacando $\frac{1}{(1-\lambda)^2} \mathbf{p}^t \mathbf{p}$, que disminuye su valor conforme más pequeño es el factor de olvido (menor memoria implica mejor capacidad de rastreo). A continuación aparece el término debido al ruido de medición $\frac{(1-\lambda)}{2} \sigma_0^2 \beta'$, que disminuye para grandes memorias. Finalmente hay que considerar el ruido de estimación $\frac{(1-\lambda)^3}{2} \sigma_0^2 V_e'$, el que menos influye en condiciones normales.

Para encontrar el valor de λ que hace mínimo $D(\infty)$ se pueden considerar los términos más representativos de cada contribución, sin considerar ruido de estimación:

$$D(\infty) \simeq \frac{1}{(1-\lambda)^2} \mathbf{p}^t \mathbf{p} + \frac{1-\lambda}{2} \sigma_0^2 \beta' + \frac{1-\lambda}{2} \mathbf{p}^t \mathbf{H} \mathbf{p} \quad (3.79)$$

El último sumando se ha introducido porque es del mismo orden que el término del ruido de medición y por tanto 'compite' directamente con él. Haciendo la derivada correspondiente se obtiene que:

$$\lambda_{opt} = 1 - \left(\frac{4 \mathbf{p}^t \mathbf{p}}{\sigma_0^2 \beta' + \mathbf{p}^t \mathbf{H} \mathbf{p}} \right)^{1/3} \quad (3.80)$$

Es muy ilustrativo discutir los dos casos siguientes:

1. Cuando la variación lineal sea muy pequeña, $\mathbf{p} \rightarrow 0$ y se deduce que el factor de olvido óptimo en este caso es 1, es decir, memoria infinita.
2. Cuanto menor sea el ruido de medición o mayor sea la variación lineal, menor se hace a su vez λ_{opt} , de manera que se requiere menos memoria.

Capítulo 4

El algoritmo sobredeterminado recursivo de variable instrumental ORIV

Introducción

En muchas ocasiones no basta con poseer herramientas capaces de resolver de manera eficiente un sistema de ecuaciones típico de RIV sino que se hace necesario ampliar las posibilidades incluyendo el caso en el que la matriz sea sobredeterminada. Para tales circunstancias se diseñó el algoritmo sobredeterminado recursivo de variable instrumental ORIV.

Este algoritmo se obtiene en el apartado 4.1 y su análisis tanto en ambientes estacionarios como no estacionarios se presenta en los apartados 4.2 y 4.3 respectivamente. En ellos se estudia la convergencia en media, el error cuadrático medio, la influencia del ruido de estimación, etc. Todo este análisis se puede considerar condensado en el apartado 4.4 donde se deducen las condiciones óptimas para el rastreo. El capítulo finaliza mostrando, en el apartado 4.5, la equivalencia de las dos aproximaciones utilizadas en el análisis anterior: la técnica del promediado directo y el suponer que el algoritmo ha evolucionado suficientemente.

4.1 Obtención del algoritmo

Recuérdese que el algoritmo recursivo de variable instrumental RIV era capaz de resolver, de forma eficiente, el sistema de ecuaciones

$$\Phi(n)\mathbf{w}(n) = \mathbf{z}(n) \quad (4.1)$$

donde la matriz $\Phi(n)$ es cuadrada ya que las longitudes del vector de datos $\mathbf{x}(n)$ y el de variable instrumental $\tilde{\mathbf{x}}(n)$ eran idénticas. Recuérdese que

$$\Phi(n) = \langle \tilde{\mathbf{x}}(n), \mathbf{x}(n) \rangle = \sum_{i=1}^n \lambda^{n-i} \tilde{\mathbf{x}}(i) \mathbf{x}^t(i) \quad (4.2)$$

y que

$$\mathbf{z}(n) = \langle \tilde{\mathbf{x}}(n), d(n) \rangle = \sum_{i=1}^n \lambda^{n-i} \tilde{\mathbf{x}}(i) d(i) \quad (4.3)$$

Sin embargo, en muchos problemas es posible plantear sistemas con más ecuaciones que incógnitas, de forma que se cuente con más información. En estos casos la matriz $\Phi(n)$ tiene más filas que columnas y por tanto la solución de (4.1) no se obtiene mediante una simple inversión, ya que es necesario definir la solución como aquel vector que minimiza la distancia $\|\Phi(n)\mathbf{w}(n) - \mathbf{z}(n)\|^2$ dando como resultado

$$\mathbf{w}(n) = (\Phi^t(n)\Phi(n))^{-1} \Phi(n)^t \mathbf{z}(n) \quad (4.4)$$

El algoritmo sobredeterminado recursivo de variable instrumental ORIV [Fri84] es capaz de calcular la expresión (4.4), tal y como su propio nombre indica, de manera recursiva una vez conocido $\mathbf{w}(n-1)$ y los nuevos datos $\mathbf{x}(n)$, $\tilde{\mathbf{x}}(n)$ y $d(n)$; el proceso a seguir se detalla a continuación.

En primer lugar se hallan las ecuaciones de actualización de las magnitudes implicadas, obtenidas directamente de sus definiciones (4.2) y (4.3):

$$\Phi(n) = \lambda \Phi(n-1) + \tilde{\mathbf{x}}(n) \mathbf{x}^t(n) \quad (4.5)$$

$$\mathbf{z}(n) = \lambda \mathbf{z}(n-1) + \tilde{\mathbf{x}}(n) d(n) \quad (4.6)$$

Donde $\tilde{\mathbf{x}}(n) = [\tilde{x}(n), \tilde{x}(n-1), \dots, \tilde{x}(n-l+1)]$ y $\mathbf{x}(n) = [x(n), x(n-1), \dots, x(n-q+1)]$ con $l > q$ al ser un sistema sobredeterminado. Además, dado que se utilizará profusamente en los próximos apartados, es conveniente definir el producto

$$\Gamma(n) = \Phi^t(n)\Phi(n) \quad (4.7)$$

a partir del cual y teniendo en cuenta (4.5) se puede deducir fácilmente su ley de actualización

$$\Gamma(n) = \lambda^2 \Gamma(n-1) + \mathbf{X}(n) \Lambda^{-1}(n) \mathbf{X}^t(n) \quad (4.8)$$

donde se ha hecho uso de las siguientes definiciones,

$$\mathbf{X}(n) = [\Phi^t(n-1)\tilde{\mathbf{x}}(n) \quad \mathbf{x}(n)] \quad (4.9)$$

$$\mathbf{\Lambda}^{-1}(n) = \begin{pmatrix} 0 & \lambda \\ \lambda & \tilde{\mathbf{x}}^t(n)\tilde{\mathbf{x}}(n) \end{pmatrix} \quad (4.10)$$

Aplicando el lema de inversión matricial (ver apéndice A) a (4.8) se obtiene

$$\begin{aligned} \lambda^2 \mathbf{\Gamma}^{-1}(n) &= \mathbf{\Gamma}^{-1}(n-1) - \mathbf{\Gamma}^{-1}(n-1)\mathbf{X}(n)[\lambda^2 \mathbf{\Lambda}(n) + \\ &+ \mathbf{X}^t(n)\mathbf{\Gamma}^{-1}(n-1)\mathbf{X}(n)]^{-1} \mathbf{X}^t(n)\mathbf{\Gamma}^{-1}(n-1) \end{aligned} \quad (4.11)$$

donde se aprecia que es posible llevar a cabo la inversión de la matriz $\mathbf{\Gamma}(n)$, de dimensiones $q \times q$, invirtiendo sólo una matriz 2×2 . Para continuar con la derivación del algoritmo ORIV son necesarias dos relaciones adicionales:

1. Primera relación: teniendo en cuenta las ecuaciones de actualización (4.6) y (4.5) junto con $\Phi^t(n)\mathbf{z}(n) = \mathbf{\Gamma}(n)\mathbf{w}(n)$ (pequeña modificación de (4.4)) se obtiene sin dificultad que:

$$\begin{aligned} \Phi^t(n)\mathbf{z}(n) &= \lambda^2 \mathbf{\Gamma}(n-1)\mathbf{w}(n-1) + \lambda \Phi^t(n-1)\tilde{\mathbf{x}}(n)d(n) + \\ &+ \mathbf{x}(n)\tilde{\mathbf{x}}^t(n)\tilde{\mathbf{x}}(n)d(n) + \lambda \mathbf{x}(n)\tilde{\mathbf{x}}^t(n)\mathbf{z}(n-1) \end{aligned} \quad (4.12)$$

2. Segunda relación. Considerando la expresión (4.8) se obtiene, sin más que multiplicar por $\mathbf{\Gamma}^{-1}(n)$ y reagrupando términos, la siguiente relación:

$$\mathbf{\Gamma}^{-1}(n)\mathbf{\Gamma}(n-1) = \frac{1}{\lambda^2} [\mathbf{I} - \mathbf{\Gamma}^{-1}(n)\mathbf{X}(n)\mathbf{\Lambda}^{-1}(n)\mathbf{X}^t(n)] \quad (4.13)$$

Para obtener la ecuación de actualización deseada del vector incógnita $\mathbf{w}(n)$ se multiplica (4.11) con (4.12) y se tienen en cuenta (4.13) y (4.6):

$$\begin{aligned} \mathbf{w}(n) &= \mathbf{\Gamma}^{-1}(n)\Phi^t(n)\mathbf{z}(n) = \\ &= [\mathbf{I} - \mathbf{\Gamma}^{-1}(n)\mathbf{X}(n)\mathbf{\Lambda}^{-1}(n)\mathbf{X}^t(n)]\mathbf{w}(n-1) + \\ &+ \mathbf{\Gamma}^{-1}(n)[\lambda \Phi^t(n-1)\tilde{\mathbf{x}}(n)d(n) + \mathbf{x}(n)\tilde{\mathbf{x}}^t(n)\mathbf{z}(n)] \end{aligned} \quad (4.14)$$

O en forma más compacta, en analogía con el RLS y el RIV, se puede escribir de la siguiente manera

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \mathbf{K}(n)\boldsymbol{\alpha}(n) \quad (4.15)$$

donde las nuevas magnitudes están definidas como sigue

$$\mathbf{K}(n) = \mathbf{\Gamma}^{-1}(n)\mathbf{X}(n)\mathbf{\Lambda}^{-1}(n) \quad p \times 2 \quad (4.16)$$

$$\boldsymbol{\alpha}(n) = \mathbf{v}(n) - \mathbf{X}^t(n)\mathbf{w}(n-1) \quad 2 \times 1 \quad (4.17)$$

$$\mathbf{v}(n) = \begin{pmatrix} \tilde{\mathbf{x}}^t(n)\mathbf{z}(n-1) \\ d(n) \end{pmatrix} \quad (4.18)$$

Como es habitual a la matriz $\mathbf{K}(n)$ se le conoce como matriz de ganancia y al vector $\boldsymbol{\alpha}(n)$ como error a priori. Las ecuaciones presentadas hasta ahora son suficientes para establecer todo el proceso recursivo de ORIV y poder actualizar, con la llegada de nuevos datos, el estimador $\mathbf{w}(n)$.

Para hacer más compactas estas ecuaciones es conveniente notar la siguiente igualdad:

$$\mathbf{K}(n) = \mathbf{\Gamma}^{-1}(n-1)\mathbf{X}(n)[\lambda^2\mathbf{\Lambda}(n) + \mathbf{X}^t(n)\mathbf{\Gamma}^{-1}(n-1)\mathbf{X}(n)]^{-1} \quad (4.19)$$

demostrada a continuación. La ecuación (4.13) se puede reescribir en función de la matriz de ganancia, siendo directo que

$$\mathbf{\Gamma}^{-1}(n)\mathbf{\Gamma}(n-1) = \frac{1}{\lambda^2}[\mathbf{I} - \mathbf{K}(n)\mathbf{X}^t(n)] \quad (4.20)$$

y despejando $\mathbf{\Gamma}^{-1}(n)$ queda

$$\mathbf{\Gamma}^{-1}(n) = \frac{1}{\lambda^2}[\mathbf{I} - \mathbf{K}(n)\mathbf{X}^t(n)]\mathbf{\Gamma}^{-1}(n-1) \quad (4.21)$$

Teniendo esta expresión en cuenta en (4.16) y reagrupando términos

$$\begin{aligned} \mathbf{K}(n) \left[\mathbf{I} + \frac{1}{\lambda^2} \mathbf{X}^t(n)\mathbf{\Gamma}^{-1}(n-1)\mathbf{X}(n)\mathbf{\Lambda}^{-1}(n) \right] &= \\ &= \frac{1}{\lambda^2} \mathbf{\Gamma}^{-1}(n-1)\mathbf{X}(n)\mathbf{\Lambda}^{-1}(n) \end{aligned} \quad (4.22)$$

multiplicando ambos miembros por la derecha por $\lambda^2\mathbf{\Lambda}(n)$ y despejando la matriz de ganancia se obtiene la expresión buscada (4.19). Con esta nueva expresión para $\mathbf{K}(n)$, la ecuación de actualización de la matriz $\mathbf{\Gamma}^{-1}(n)$ (4.11), se puede reescribir de la siguiente manera:

$$\mathbf{\Gamma}^{-1}(n) = \frac{1}{\lambda^2}[\mathbf{\Gamma}^{-1}(n-1) - \mathbf{K}(n)\mathbf{X}^t(n)\mathbf{\Gamma}^{-1}(n-1)] \quad (4.23)$$

Teniendo en cuenta las expresiones derivadas en último lugar, en la tabla 4.1 se presenta el proceso recursivo paso a paso.

<p><u>Condiciones iniciales</u></p> $\boldsymbol{\Phi}^t(0) = \delta[\mathbf{I}_{q \times q} \mathbf{0}_{q \times l-q}]$ $\mathbf{\Gamma}(0) = \frac{1}{\delta^2} \mathbf{I}_{q \times q}$ <p>$\delta =$ escalar arbitrario</p> $\mathbf{w}(0) = \mathbf{0}$ $\mathbf{z}(0) = \mathbf{0}$
--

Cuerpo del algoritmo
$\Phi(n) = \lambda\Phi(n-1) + \tilde{\mathbf{x}}(n)\mathbf{x}^t(n)$
$\mathbf{X}(n) = [\Phi^t(n-1)\tilde{\mathbf{x}}(n) \quad \mathbf{x}(n)]$
$\lambda^2\mathbf{\Lambda}(n) = \begin{pmatrix} -\tilde{\mathbf{x}}^t(n)\tilde{\mathbf{x}}(n) & \lambda \\ \lambda & 0 \end{pmatrix}$
$\mathbf{K}(n) = \Gamma^{-1}(n-1)\mathbf{X}(n)[\lambda^2\mathbf{\Lambda}(n) + \mathbf{X}^t(n)\Gamma^{-1}(n-1)\mathbf{X}(n)]^{-1}$
$\Gamma^{-1}(n) = \frac{1}{\lambda^2}[\Gamma^{-1}(n-1) - \mathbf{K}(n)\mathbf{X}^t(n)\Gamma^{-1}(n-1)]$
$\mathbf{v}(n) = \begin{pmatrix} \tilde{\mathbf{x}}^t(n)z(n-1) \\ d(n) \end{pmatrix}$
$\boldsymbol{\alpha}(n) = \mathbf{v}(n) - \mathbf{X}^t(n)\mathbf{w}(n-1)$
$\mathbf{z}(n) = \lambda\mathbf{z}(n-1) + \tilde{\mathbf{x}}(n)d(n)$
$\mathbf{w}(n) = \mathbf{w}(n-1) + \mathbf{K}(n)\boldsymbol{\alpha}(n)$

Tabla 4.1 *Pasos del algoritmo ORIV*

La inicialización utilizada aquí es la conocida como ‘inicialización con ligadura suave’ (en la literatura inglesa aparece con el nombre de ‘soft constrained initialization’). De todas maneras es también posible inicializar el algoritmo de manera exacta. Para ello se toman r datos, con $r > q$ para evitar que la matriz $\Phi(r)$ sea de rango deficiente (sus columnas¹ no sean linealmente dependientes) y permitir la existencia de las inversas necesarias para el algoritmo, para construir dicha matriz junto con $\mathbf{z}(r)$. A partir de la primera se obtiene $\Gamma(r)$ y después $\mathbf{w}(r)$. Con estas magnitudes ya se puede comenzar el proceso recursivo indicado en la tabla 4.1.

4.2 Análisis de la convergencia para sistemas estacionarios

En los apartados que siguen a continuación se va a proceder al estudio de la convergencia de ORIV cuando trabaja con sistemas estacionarios. Más adelante se tratará el caso de sistemas no estacionarios. Como se ha comentado con anterioridad las propiedades a estudiar son la convergencia en media y en media cuadrática tanto con ruido de estimación como sin él. El análisis se ha enfocado de forma análoga al realizado acerca del RIV.

¹Recordar que esta matriz tiene más filas que columnas, por lo que sus filas nunca podrán ser linealmente independientes.

4.2.1 Convergencia en media

La posibilidad de resolver problemas que involucren series aleatorias mediante ecuaciones pensadas para situaciones deterministas, recurriendo a la ergodicidad, ya ha aparecido cuando se analizaron RLS y RIV. De nuevo el principal objetivo será resolver un sistema de ecuaciones tal y como este

$$\mathbf{R}_{\bar{x}x} \mathbf{w}_0 = \mathbf{r}_{\bar{d}x} \quad (4.24)$$

donde $\mathbf{R}_{\bar{x}x} = E[\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)]$ y $\mathbf{r}_{\bar{d}x} = E[\tilde{\mathbf{x}}(n)d(n)]$. Es por tanto necesario determinar bajo qué condiciones resolver (4.1) es equivalente a resolver (4.24).

Una opción posible es calcular (4.4) para n suficientemente grande utilizando distintas series numéricas y luego promediar los resultados obtenidos para esas distintas series. Esto es equivalente a dejar que ORIV evolucione suficientemente con el tiempo para varias series de datos y luego tomar valor esperado. Este procedimiento produce el siguiente efecto.

Como las series aleatorias con las que se trabaja se suponen ergódicas, los promedios temporales tienden a comportarse como promedios en realizaciones. Bajo esta hipótesis se puede aceptar que para $n \rightarrow \infty$ se verifica que:

$$\Phi(n) \sim \frac{\mathbf{R}_{\bar{x}x}}{1-\lambda} + \mathbf{R}_{\bar{x}x}^e(n) \quad (4.25)$$

$$\mathbf{z}(n) \sim \frac{\mathbf{r}_{\bar{x}d}}{1-\lambda} + \mathbf{r}_{\bar{x}d}^e(n) \quad (4.26)$$

Es decir, las magnitudes calculadas como promedios temporales se descomponen en un sumando determinista igual a su valor esperado más otro sumando en forma de ruido. Las componentes de ruido ($\mathbf{R}_{\bar{x}x}^e(n)$ y $\mathbf{r}_{\bar{x}d}^e(n)$) se pueden modelar convenientemente como de media nula e independientes del resto de variables aleatorias (ver comentarios relacionados en el apéndice C). Bajo estas condiciones la expresión (4.4) toma la forma

$$\begin{aligned} \mathbf{w}(n) \sim & [\mathbf{R}_{\bar{x}x}^t \mathbf{R}_{\bar{x}x} + (1-\lambda)\mathbf{R}_{\bar{x}x}^t \mathbf{R}_{\bar{x}x}^e(n) + (1-\lambda)\mathbf{R}_{\bar{x}x}^{et}(n)\mathbf{R}_{\bar{x}x} + \\ & + (1-\lambda)^2 \mathbf{R}^{et}(n)\mathbf{R}^e(n)]^{-1} [\mathbf{R}_{\bar{x}x}^t \mathbf{r}_{\bar{x}d} + (1-\lambda)\mathbf{R}_{\bar{x}x}^t \mathbf{r}_{\bar{x}d}^e(n) + \\ & + (1-\lambda)\mathbf{R}_{\bar{x}x}^{et}(n)\mathbf{r}_{\bar{x}d} + (1-\lambda)^2 \mathbf{R}^{et}(n)\mathbf{r}^e(n)] \quad (4.27) \end{aligned}$$

A partir de ahora se eliminarán los subíndices allí donde no haya lugar a confusión. Al ser λ del orden de la unidad, se tomará el factor $(1-\lambda)$ como parámetro perturbativo. Además, se considerará que la matriz $\Phi(n)$ es cuasideterminista para n grandes, es decir, $\|\mathbf{R}^e(n)\| \ll \|\mathbf{R}\|$ donde $\|\cdot\|$ es una norma conveniente para las matrices. Con esto, la inversa que aparece en el primer factor de (4.27) se puede aproximar quedando una expresión tal y como sigue:

$$\begin{aligned} \mathbf{w}(n) \sim & [\mathbf{I} - (1-\lambda)(\mathbf{R}^t \mathbf{R})^{-1} \mathbf{R}^t \mathbf{R}^e(n) - (1-\lambda)(\mathbf{R}^t \mathbf{R})^{-1} \mathbf{R}^{et}(n)\mathbf{R}] (\mathbf{R}^t \mathbf{R})^{-1} \\ & [\mathbf{R}^t \mathbf{r} + (1-\lambda)\mathbf{R}^t \mathbf{r}^e(n) + (1-\lambda)\mathbf{R}^{et}(n)\mathbf{r}] \quad (4.28) \end{aligned}$$

Trabajando a primer orden en $(1 - \lambda)$ y tomando valor esperado queda que

$$E[\mathbf{w}(n)] \sim (\mathbf{R}^t \mathbf{R})^{-1} \mathbf{R}^t \mathbf{r} = \mathbf{w}_0 \quad (4.29)$$

Por tanto se ha comprobado que si se resuelve (4.1) con series aleatorias, se deja evolucionar el algoritmo suficientemente y se toma valor esperado, es como resolver (4.24). Éste es un procedimiento práctico para resolver ecuaciones tales como (4.24).

Hay que recalcar una cuestión importante: $\mathbf{w}(n)$ cumple que $\langle \tilde{\mathbf{x}}(n), e(n) \rangle = 0 \forall n$ en mínimos cuadrados por construcción. Sin embargo, si se define $e_0(n) = d(n) - \mathbf{w}_0^t \mathbf{x}(n)$ se cumple que $\langle \tilde{\mathbf{x}}(n), e_0(n) \rangle \neq 0$ ya que hay que tomar valor esperado:

$$E[\tilde{\mathbf{x}}(n)e_0(n)] = 0 \quad (4.30)$$

Efecto de las condiciones iniciales

Una vez aclarado que, debido a la ergodicidad, es posible encontrar soluciones a problemas que involucran magnitudes aleatorias mediante ecuaciones destinadas a magnitudes deterministas, se procederá a comprobar cuál es el efecto de las condiciones iniciales cuando la ecuación (4.1) se resuelve mediante ORIV.

Partiendo de la ecuación (4.15), utilizando para la matriz de ganancia la expresión dada en (4.16), y restando \mathbf{w}_0 a ambos miembros queda

$$\Delta \mathbf{w}(n) = \Delta \mathbf{w}(n-1) + \mathbf{\Gamma}^{-1}(n) \mathbf{X}(n) \mathbf{\Lambda}^{-1}(n) \boldsymbol{\alpha}(n) \quad (4.31)$$

donde $\Delta \mathbf{w}(n) = \mathbf{w}(n) - \mathbf{w}_0$. Si se multiplica ahora por $\mathbf{\Gamma}(n)$ y se utiliza (4.8) se llega a

$$\begin{aligned} \mathbf{\Gamma}(n) \Delta \mathbf{w}(n) &= \lambda^2 \mathbf{\Gamma}(n-1) \Delta \mathbf{w}(n-1) + \\ &+ \lambda \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) [\mathbf{z}(n-1) - \mathbf{\Phi}(n-1) \mathbf{w}_0] + \mathbf{\Phi}^t(n) \tilde{\mathbf{x}}(n) e_0(n) \end{aligned} \quad (4.32)$$

una ecuación en diferencias cuya solución es

$$\begin{aligned} \mathbf{\Gamma}(n) \Delta \mathbf{w}(n) &= \lambda^{2n} \mathbf{\Gamma}(0) \Delta \mathbf{w}(0) + \sum_{j=1}^n \lambda^{2(n-j)} \mathbf{\Phi}^t(j) \tilde{\mathbf{x}}(j) e_0(j) + \\ &+ \sum_{i=1}^n \lambda \lambda^{2(n-i)} \mathbf{x}(i) \tilde{\mathbf{x}}^t(i) [\mathbf{z}(i-1) - \mathbf{\Phi}(i-1) \mathbf{w}_0] \end{aligned} \quad (4.33)$$

Para seguir con los cálculos es necesario hacer una hipótesis, que se puede enmarcar dentro de las conocidas como hipótesis de independencia HI (ver apéndice C):

III $\Phi(n)$ y $z(n)$ son independientes entre sí y del resto de variables aleatorias involucradas.

Tomando valor esperado, teniendo en cuenta el comentario después de la ecuación (4.29) y considerando III se llega a

$$\mathbb{E}[\Gamma(n)\Delta\mathbf{w}(n)] = \lambda^{2n}\Gamma(0)\Delta\mathbf{w}(0) \quad (4.34)$$

Para poder descomponer el valor esperado que en ella aparece se tiene en cuenta que lo que interesa es estudiar la convergencia, por lo que n se puede considerar suficientemente grande, lo cual permite utilizar la expresión (B.1) y afirmar que $\Gamma(n)$ es independiente de $\Delta\mathbf{w}(n)$. Siguiendo también a (B.1) es directo que

$$\mathbb{E}[\Gamma(n)] = (1 - \lambda)^{-2}\mathbf{R}^t\mathbf{R} + \mathbb{E}[\mathbf{R}^{et}(n)\mathbf{R}^e(n)] \quad (4.35)$$

y por tanto su inversa es

$$\begin{aligned} \mathbb{E}[\Gamma(n)]^{-1} &= \\ &= (1 - \lambda)^2[\mathbf{I} + (1 - \lambda)^2(\mathbf{R}^t\mathbf{R})^{-1}\mathbb{E}[\mathbf{R}^{et}(n)\mathbf{R}^e(n)]]^{-1}(\mathbf{R}^t\mathbf{R})^{-1} \approx \\ &\approx (1 - \lambda)^2(\mathbf{R}^t\mathbf{R})^{-1} \end{aligned} \quad (4.36)$$

donde de nuevo para hacer la aproximación se ha considerado λ del orden de la unidad y la matriz $\Phi(n)$ cuasideterminista. Con todas estas aproximaciones se puede retomar la ecuación (4.34) y escribirla como:

$$\mathbb{E}[\Delta\mathbf{w}(n)] = \lambda^{2n}(1 - \lambda)^2(\mathbf{R}^t\mathbf{R})^{-1}\Gamma(0)\Delta\mathbf{w}(0) \quad (4.37)$$

A la vista de este resultado se puede comentar lo siguiente:

1. La influencia de las condiciones iniciales, $\Gamma(0)$ y $\Delta\mathbf{w}(0)$, decae exponencialmente con λ^{2n} .
2. $\mathbb{E}[\mathbf{w}(n)] \rightarrow \mathbf{w}_0$ para n grandes, como ya se sabía.
3. Si $\lambda = 1$ hay que cambiar $(1 - \lambda)$ por $1/n$ para poder asegurar la correcta normalización de $\Phi(n)$ en (4.2) y de $z(n)$ en (4.3).
4. La expresión (4.33) se puede escribir de forma más compacta como

$$\Gamma(n)\Delta\mathbf{w}(n) = \lambda^{2n}\Gamma(0)\Delta\mathbf{w}(0) + \sum_{i=1}^n \lambda^{2(n-i)}\mathbf{X}(i)\Lambda^{-1}(i)\alpha_0(i) \quad (4.38)$$

donde

$$\alpha_0(n) = \mathbf{v}(n) - \mathbf{X}^t(n)\mathbf{w}_0 \quad (4.39)$$

, que se puede interpretar como la generalización del error de medición $e_0(n)$ para sistemas sobredeterminados. Por lo que (4.34) implica que

$$\mathbb{E}[\mathbf{X}(n)\Lambda^{-1}(n)\alpha_0(n)] = 0 \quad (4.40)$$

donde se ha tenido en cuenta que para sistemas estacionarios esta cantidad es a su vez estacionaria, es decir,

$$E[\mathbf{X}(i)\mathbf{\Lambda}^{-1}(i)\boldsymbol{\alpha}_0(i)] = cte \quad \forall i \quad (4.41)$$

La importancia de la ecuación (4.40) es que se puede considerar como la generalización de (4.30) para sistemas sobredeterminados.

4.2.2 Error cuadrático medio

En este apartado se hallan expresiones para estudiar la evolución de la magnitud $q(n) \equiv E[\Delta\mathbf{w}^t(n)\mathbf{Q}\Delta\mathbf{w}(n)]$ donde \mathbf{Q} es cualquier matriz definida positiva (métrica), es decir, se va a calcular la desviación en media cuadrática de los estimadores $\mathbf{w}(n)$ de los pesos. Para ello se comienza con la expresión (4.32) escrita en forma compacta:

$$\mathbf{\Gamma}(n)\Delta\mathbf{w}(n) = \lambda^2\mathbf{\Gamma}(n-1)\Delta\mathbf{w}(n-1) + \mathbf{X}(n)\mathbf{\Lambda}^{-1}(n)\boldsymbol{\alpha}_0(n) \quad (4.42)$$

Dado que se quiere calcular, en particular, el error cuadrático medio en convergencia, es lícito tomar n muy grande de manera que $\mathbf{\Gamma}^{-1}(n)\mathbf{\Gamma}(n-1) = \mathbf{I}$. De forma que si se despeja $\Delta\mathbf{w}(n)$ en la expresión anterior se puede escribir que aproximadamente se verifica

$$\Delta\mathbf{w}(n) = \lambda^2\Delta\mathbf{w}(n-1) + \mathbf{\Gamma}^{-1}(n)\mathbf{X}(n)\mathbf{\Lambda}^{-1}(n)\boldsymbol{\alpha}_0(n) \quad (4.43)$$

basándose en esta ecuación en diferencias se puede construir la siguiente igualdad

$$\begin{aligned} \Delta\mathbf{w}^t(n)\mathbf{Q}\Delta\mathbf{w}(n) &= \lambda^4\Delta\mathbf{w}^t(n-1)\mathbf{Q}\Delta\mathbf{w}(n-1) + \\ &+ 2\lambda^2\Delta\mathbf{w}^t(n-1)\mathbf{Q}\mathbf{\Gamma}^{-1}(n)\mathbf{X}(n)\mathbf{\Lambda}^{-1}(n)\boldsymbol{\alpha}_0(n) + \\ &+ [\mathbf{X}(n)\mathbf{\Lambda}^{-1}(n)\boldsymbol{\alpha}_0(n)]^t\hat{\boldsymbol{\beta}}_{\mathbf{Q}}(n)[\mathbf{X}(n)\mathbf{\Lambda}^{-1}(n)\boldsymbol{\alpha}_0(n)] \end{aligned} \quad (4.44)$$

donde $\hat{\boldsymbol{\beta}}_{\mathbf{Q}}(n) \equiv \mathbf{\Gamma}^{-t}(n)\mathbf{Q}\mathbf{\Gamma}^{-1}(n)$. Para continuar se calcula el valor esperado de cada sumando del miembro de la derecha:

Primer sumando: Éste no es más que $\lambda^4q(n-1)$.

Segundo sumando: Aquí hay que tener en cuenta la expresión (B.3), ya que no se quiere tener en cuenta ningún tipo de ruido de estimación, junto con otra hipótesis de independencia (ver apéndice C):

H12 $\mathbf{w}(n-1)$ es independiente de $\mathbf{X}(n)\mathbf{\Lambda}^{-1}(n)\boldsymbol{\alpha}_0(n)$.

Así que se puede afirmar que este sumando es nulo usando (4.40)

Tercer sumando: De nuevo se recurre a la simplificación para $\Gamma^{-1}(n)$ dada en (B.3) con lo que el valor esperado de este sumando es:

$$3^{\text{er}} \text{ sumando} \xrightarrow{\mathbb{E}[\cdot]} (1 - \lambda)^4 \text{Tr}\{\mathbf{V}_{X\alpha_0}\beta_Q\} \quad (4.45)$$

donde

$$\mathbf{V}_{X\alpha_0} = \mathbb{E}[\mathbf{X}(n)\mathbf{\Lambda}^{-1}(n)\boldsymbol{\alpha}_0(n)(\mathbf{X}(n)\mathbf{\Lambda}^{-1}(n)\boldsymbol{\alpha}_0(n))^t] \quad (4.46)$$

$$\beta_Q = (\mathbf{R}^t\mathbf{R})^{-t}\mathbf{Q}(\mathbf{R}^t\mathbf{R})^{-1} \quad (4.47)$$

Dado que el factor $\mathbf{X}(n)\mathbf{\Lambda}^{-1}(n)\boldsymbol{\alpha}_0(n)$ es de media nula, $\mathbf{V}_{X\alpha_0}$ se puede interpretar como su matriz de varianza.

Por tanto, teniendo en cuenta el valor de todos los sumandos, se llega a:

$$q(n) = \lambda^4 q(n-1) + (1 - \lambda)^4 \text{Tr}\{\mathbf{V}_{X\alpha_0}\beta_Q\} \quad (4.48)$$

válido para n suficientemente grande. Esta ecuación en diferencias tiene un valor en convergencia de:

$$q(\infty) = \frac{(1 - \lambda)^3}{(1 + \lambda^2)(1 + \lambda)} \text{Tr}\{\mathbf{V}_{X\alpha_0}\beta_Q\} \quad (4.49)$$

A partir de aquí se pueden hacer las siguientes consideraciones:

1. Para sistemas estacionarios, como es el caso que nos ocupa, el factor de olvido λ debe ser la unidad y por tanto $(1 - \lambda)$ hay que cambiarlo de nuevo por $1/n$. De esta forma $q(\infty) \rightarrow 0$.
2. Los dos factores que aparecen en (4.49) dependen de la variable instrumental $\hat{x}(n)$. La MSD en convergencia se podría reducir mediante una elección adecuada.
3. Los cálculos previos realizados tomando $(1 - \lambda)$ como parámetro perturbativo no son realmente válidos, porque $\mathbf{V}_{X\alpha_0}$ tiene dependencia en él como se verá en el siguiente apartado.
4. Combinando las expresiones (4.48) y (4.49) se puede obtener la siguiente ecuación:

$$q(n) = q(\infty) + \lambda^{4(n-n_0)} [q(n_0) - q(\infty)] \quad (4.50)$$

donde $n > n_0 \gg 1$. Esto significa que el efecto de las condiciones iniciales cae a cero como λ^{4n} . Además, la velocidad de convergencia sólo depende de λ y no de la dispersión de autovalores de \mathbf{R} , como es usual en otros algoritmos.

Sin embargo, a pesar del punto 4 anterior, no se elimina la posibilidad de que si el problema está mal condicionado, aparezcan problemas numéricos que acarreen problemas de estabilidad y una mala convergencia.

4.2.3 Error cuadrático medio en presencia de ruido

Cuando las series con las que se trabajen sean ruidosas, las componentes aleatorias $\mathbf{R}^e(n)$ y $\mathbf{r}^e(n)$ en (4.25) y (4.26) cobran cada vez más y más importancia de manera que hay que tenerlas en cuenta a la hora de hacer las aproximaciones en los cálculos anteriores. En esta sección se considera la influencia de estas componentes en los factores que aparecen en (4.48):

- i) $\beta_Q = E[\mathbf{\Gamma}^{-t}(n)\mathbf{Q}\mathbf{\Gamma}^{-1}(n)]$. Se utilizará la aproximación dada en (B.2). Con esta aproximación se aprecia que en el término que se estudia, la componente aleatoria aparece siempre ‘normalizada’ por \mathbf{R} , es decir, aparece siempre en la forma ‘término de \mathbf{R}^e/\mathbf{R} ’, el cual es despreciable comparado con los términos en los que \mathbf{R}^e aparezca sólo, simplemente porque $\mathbf{\Phi}(n)$ es cuasideterminista.
- ii) $\mathbf{V}_{X\alpha_0}$. Utilizando en la expresión (4.46) que define a $\mathbf{V}_{X\alpha_0}$ las expresiones (4.9), (4.10) y (4.39) que definen a sus componentes se puede observar que los términos en \mathbf{R}^e no están normalizados por \mathbf{R} aquí, por lo que no es posible por tanto omitir la componente aleatoria.

En definitiva, la influencia del ruido es mayor para $\mathbf{V}_{X\alpha_0}$ que para β_Q , de manera que el caso ii) hay que considerarlo en detalle.

Desarrollando según la definición de sus componentes se tiene que

$$\begin{aligned} \mathbf{X}(n)\mathbf{\Lambda}^{-1}(n)\boldsymbol{\alpha}_0(n) &= \\ &= \lambda\mathbf{x}(n)\tilde{\mathbf{x}}^t(n)[\mathbf{z}(n-1) - \mathbf{\Phi}(n-1)\mathbf{w}_0] + \mathbf{\Phi}^t(n)\tilde{\mathbf{x}}(n)e_0(n) \end{aligned} \quad (4.51)$$

considerando n suficientemente grande y teniendo en cuenta el comentario después de (4.29) resulta:

$$\begin{aligned} \mathbf{X}(n)\mathbf{\Lambda}^{-1}(n)\boldsymbol{\alpha}_0(n) &\underset{n \rightarrow \infty}{\sim} \\ &\sim \lambda\mathbf{x}(n)\tilde{\mathbf{x}}^t(n) \left[\frac{\mathbf{r}}{1-\lambda} + \mathbf{r}^e(n-1) - \frac{\mathbf{R}}{1-\lambda}\mathbf{w}_0 - \mathbf{R}^e(n-1)\mathbf{w}_0 \right] + \\ &\quad + \frac{\mathbf{R}^t}{1-\lambda}\tilde{\mathbf{x}}(n)e_0(n) + \mathbf{R}^{et}(n)\tilde{\mathbf{x}}(n)e_0(n) = \\ &= \lambda\mathbf{x}(n)\tilde{\mathbf{x}}^t(n)[\mathbf{r}^e(n-1) - \mathbf{R}^e(n-1)\mathbf{w}_0] + \frac{\mathbf{R}^t}{1-\lambda}\tilde{\mathbf{x}}(n)e_0(n) + \mathbf{R}^{et}(n)\tilde{\mathbf{x}}(n)e_0(n) \end{aligned} \quad (4.52)$$

El valor esperado de esta expresión es cero como es conocido para el caso general, aquí se ha demostrado para n grandes. Continuando con los cálculos se

construye el siguiente producto:

$$\begin{aligned}
\mathbf{X}(n)\mathbf{\Lambda}^{-1}(n)\boldsymbol{\alpha}_0(n)[\mathbf{X}(n)\mathbf{\Lambda}^{-1}(n)\boldsymbol{\alpha}_0(n)]^t &= \\
&= \{\lambda\mathbf{x}(n)\tilde{\mathbf{x}}^t(n)[\mathbf{r}^e(n-1) - \mathbf{R}^e(n-1)\mathbf{w}_0] + \\
&+ (1-\lambda)^{-1}\mathbf{R}^t\tilde{\mathbf{x}}(n)e_0(n) + \mathbf{R}^{et}(n)\tilde{\mathbf{x}}(n)e_0(n)\} \\
&\{\lambda\mathbf{x}(n)\tilde{\mathbf{x}}^t(n)[\mathbf{r}^e(n-1) - \mathbf{R}^e(n-1)\mathbf{w}_0] + \\
&+ (1-\lambda)^{-1}\mathbf{R}^t\tilde{\mathbf{x}}(n)e_0(n) + \mathbf{R}^{et}(n)\tilde{\mathbf{x}}(n)e_0(n)\}^t \quad (4.53)
\end{aligned}$$

Realizando los productos indicados y tomando valor esperado, teniendo en cuenta que $\mathbf{R}^e(n)$, $\mathbf{r}^e(n)$ y $e_0(n)$ son de media cero e independientes del resto de variables aleatorias, resulta:

$$\mathbf{V}_{X\alpha_0} \xrightarrow[n \rightarrow \infty]{} \lambda^2\mathbf{V}_X + \sigma_0^2 \left[\frac{\mathbf{A}}{(1-\lambda)^2} + \mathbf{V}_R \right] \quad (4.54)$$

donde

$$\mathbf{V}_X = \mathbb{E}[\mathbf{x}(n)\tilde{\mathbf{x}}^t(n)\{\mathbf{V}\mathbf{r}^e + \mathbb{E}[\mathbf{R}^e(n)\mathbf{w}_0\mathbf{w}_0^t\mathbf{R}^{et}(n)]\}\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)] \quad (4.55)$$

$$\mathbf{V}\mathbf{r}^e = \mathbb{E}[\mathbf{r}^e(n)\mathbf{r}^{et}(n)] \quad (4.56)$$

$$\sigma_0^2 = \mathbb{E}[e_0^2(n)] \quad (4.57)$$

$$\mathbf{V}_R = \mathbb{E}[\mathbf{R}^{et}(n)\mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}\mathbf{R}^e(n)] \quad (4.58)$$

$$\mathbf{A} = \mathbf{R}^t\mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}\mathbf{R} \quad (4.59)$$

De esta forma la expresión (4.49) se puede complementar con (4.54), teniendo en cuenta también que $\lambda \sim 1$:

$$\begin{aligned}
q(\infty) &= (1-\lambda)\sigma_0^2 \frac{\text{Tr}\{\boldsymbol{\beta}_Q\mathbf{A}\}}{4} + (1-\lambda)^3 \left[\sigma_0^2 \frac{\text{Tr}\{\boldsymbol{\beta}_Q\mathbf{V}_R\}}{4} + \frac{\text{Tr}\{\boldsymbol{\beta}_Q\mathbf{V}_X\}}{4} \right] \equiv \\
&\equiv q_0(\infty) + q_e(\infty) \quad (4.60)
\end{aligned}$$

donde $q_e(\infty)$ tiene toda la dependencia en los estimadores de \mathbf{R} y \mathbf{r} y por tanto aumentará su valor conforme disminuya la SNR; sin embargo $q_0(\infty)$ no dependerá de ella. Para situaciones estacionarias $q_0(\infty)$ decrece como $1/n$ y $q_e(\infty)$ como $1/n^3$. Esto es, el ruido de medición $e_0(n)$ es más importante que el ruido de estimación. Además, el término \mathbf{V}_X se ve afectado por el ruido aditivo gaussiano (se supone que la variable instrumental se diseña de manera que \mathbf{R} y \mathbf{r} no se vean).

Los resultados se han dado como una serie perturbativa en $(1-\lambda)$. Se deberían ir haciendo las mismas aproximaciones en todos los términos de (4.44) a la vez, para agrupar los sumandos que sean del mismo orden en el parámetro perturbativo. En el desarrollo que se ha seguido aquí se han tratado por separado la matriz $\boldsymbol{\Gamma}(n)$ primero y después $\mathbf{V}_{X\alpha_0}$ despreciando posibles interacciones

entre ambos factores. Sin embargo, cálculos más cuidadosos muestran que la contribución de cada término es del mismo orden que la aquí dada. Es decir, el ruido de medición contribuye como $(1 - \lambda)$, el ruido en los estimadores como $(1 - \lambda)^3$ y no aparece ningún término en $(1 - \lambda)^2$.

4.2.4 MSD en ambientes ruidosos II

En este apartado se considerará el efecto del ruido aditivo en la convergencia del MSD; es el análisis que se realizó en el apartado anterior pero utilizando otra aproximación. A partir de ahora la desviación en media cuadrática de los estimadores de los pesos se calculará mediante $D(n) = E[\Delta \mathbf{w}^t(n) \Delta \mathbf{w}(n)]$, es como $q(n)$ pero haciendo $\mathbf{Q} = \mathbf{I}$ por simplificar.

Combinando (4.15) y (4.16) se puede obtener la siguiente expresión:

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \mathbf{\Gamma}^{-1}(n) \mathbf{X}(n) \mathbf{\Lambda}^{-1}(n) \boldsymbol{\alpha}(n) \quad (4.61)$$

que desdoblado sus componentes según sus definiciones respectivas conduce a

$$\begin{aligned} \mathbf{w}(n) = \mathbf{w}(n-1) + \mathbf{\Gamma}^{-1}(n) \{ & \mathbf{\Phi}^t(n) \tilde{\mathbf{x}}(n) e_0(n) + \lambda \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) [z(n-1) - \\ & - \mathbf{\Phi}(n-1) \mathbf{w}(n-1)] + \mathbf{\Phi}^t(n) \tilde{\mathbf{x}}(n) \mathbf{x}^t(n) [\mathbf{w}_0 - \mathbf{w}(n-1)] \} \end{aligned} \quad (4.62)$$

con la cual se puede construir una ecuación en diferencias para el vector de error en los pesos $\Delta \mathbf{w}(n) = \mathbf{w}(n) - \mathbf{w}_0$,

$$\begin{aligned} \Delta \mathbf{w}(n) = & [\mathbf{I} - \lambda \mathbf{\Gamma}^{-1}(n) \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \mathbf{\Phi}(n-1) - \\ & - \mathbf{\Gamma}^{-1}(n) \mathbf{\Phi}^t(n) \tilde{\mathbf{x}}(n) \mathbf{x}^t(n)] \Delta \mathbf{w}(n-1) + \mathbf{\Gamma}^{-1}(n) \mathbf{\Phi}^t(n) \tilde{\mathbf{x}}(n) e_0(n) + \\ & + \lambda \mathbf{\Gamma}^{-1}(n) \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) z(n-1) - \lambda \mathbf{\Gamma}^{-1}(n) \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \mathbf{\Phi}(n-1) \mathbf{w}_0 \end{aligned} \quad (4.63)$$

Las aproximaciones a utilizar se fundamentan en las condiciones bajo las que se quieren obtener los resultados. Interesa el MSD en convergencia y en presencia de ruido de estimación, por tanto se supone n grande y se incluyen en el desarrollo los primeros términos que contengan información sobre ruido de estimación:

$$\begin{aligned} \mathbf{\Gamma}^{-1}(n) \approx & (1 - \lambda)^2 (\mathbf{R}^t \mathbf{R})^{-1} - \\ & - (1 - \lambda)^3 (\mathbf{R}^t \mathbf{R})^{-1} [\mathbf{R}^t \mathbf{R}^e(n) + \mathbf{R}^{et}(n) \mathbf{R}] (\mathbf{R}^t \mathbf{R})^{-1} \end{aligned} \quad (4.64)$$

$$\mathbf{\Phi}(n) \approx \frac{1}{1 - \lambda} \mathbf{R} + \mathbf{R}^e(n) \quad (4.65)$$

$$z(n) \approx \frac{1}{1 - \lambda} \mathbf{r} + \mathbf{r}^e(n) \quad (4.66)$$

Teniendo en consideración estas aproximaciones y aplicando promediado directo,

la ecuación (4.63) se convierte en:

$$\begin{aligned}
\Delta \mathbf{w}(n) = & \lambda^2 \Delta \mathbf{w}(n-1) + (1-\lambda)(\mathbf{R}^t \mathbf{R})^{-1} \mathbf{R}^t \tilde{\mathbf{x}}(n) e_0(n) + \\
& + (1-\lambda)^2 (\mathbf{R}^t \mathbf{R})^{-1} \mathbf{R}^{et}(n) \tilde{\mathbf{x}}(n) e_0(n) - \\
& - (1-\lambda)^2 (\mathbf{R}^t \mathbf{R})^{-1} [\mathbf{R}^t \mathbf{R}^e(n) + \mathbf{R}^{et}(n) \mathbf{R}] (\mathbf{R}^t \mathbf{R})^{-1} \mathbf{R}^t \tilde{\mathbf{x}}(n) e_0(n) - \\
& - (1-\lambda)^3 (\mathbf{R}^t \mathbf{R})^{-1} [\mathbf{R}^t \mathbf{R}^e(n) + \mathbf{R}^{et}(n) \mathbf{R}] (\mathbf{R}^t \mathbf{R})^{-1} \mathbf{R}^{et}(n) \tilde{\mathbf{x}}(n) e_0(n) - \\
& - (1-\lambda)^2 \lambda (\mathbf{R}^t \mathbf{R})^{-1} \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) [\mathbf{R}^e(n-1) \mathbf{w}_0 - \mathbf{r}^e(n-1)] + \\
& + (1-\lambda)^3 \lambda (\mathbf{R}^t \mathbf{R})^{-1} [\mathbf{R}^t \mathbf{R}^e(n) + \mathbf{R}^{et}(n) \mathbf{R}] (\mathbf{R}^t \mathbf{R})^{-1} \\
& \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) [\mathbf{R}^e(n-1) \mathbf{w}_0 - \mathbf{r}^e(n-1)]
\end{aligned} \tag{4.67}$$

construyendo a partir de esta expresión el producto escalar $\Delta \mathbf{w}^t(n) \Delta \mathbf{w}(n)$, tomando valor esperado y despreciando por el momento términos en $(1-\lambda)^4$ se obtiene:

$$D(n) = \lambda^4 D(n-1) + (1-\lambda)^2 \sigma_0^2 \tilde{\gamma} \tag{4.68}$$

donde sólo los efectos del error de medición se han incluido. Los términos en $(1-\lambda)^4$ son los que contienen información sobre el ruido de estimación y se agruparán bajo los términos V_e y ξ_e dando finalmente

$$D(n) = \lambda^4 D(n-1) + (1-\lambda)^2 \sigma_0^2 \tilde{\gamma} + (1-\lambda)^4 \sigma_0^2 V_e + (1-\lambda)^4 \lambda^2 \xi_e \tag{4.69}$$

donde las magnitudes involucradas se definen como

$$\tilde{\gamma} = \text{Tr}\{\mathbf{R}(\mathbf{R}^t \mathbf{R})^{-2} \mathbf{R}^t \mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}\} \tag{4.70}$$

$$\begin{aligned}
V_e = & \text{E}[\tilde{\mathbf{x}}^t(n) \mathbf{R}^e(n) (\mathbf{R}^t \mathbf{R})^{-2} \mathbf{R}^{et}(n) \tilde{\mathbf{x}}(n)] - \\
& - 2\text{E}[\tilde{\mathbf{x}}^t(n) \mathbf{R}^e(n) (\mathbf{R}^t \mathbf{R})^{-1} [\mathbf{R}^t \mathbf{R}^e(n) + \mathbf{R}^{et}(n) \mathbf{R}] (\mathbf{R}^t \mathbf{R})^{-2} \mathbf{R}^t \tilde{\mathbf{x}}(n)] - \\
& - 2\text{E}[\tilde{\mathbf{x}}^t(n) \mathbf{R}^e(n) (\mathbf{R}^t \mathbf{R})^{-2} [\mathbf{R}^t \mathbf{R}^e(n) + \mathbf{R}^{et}(n) \mathbf{R}] (\mathbf{R}^t \mathbf{R})^{-1} \mathbf{R}^t \tilde{\mathbf{x}}(n)] + \\
& + \text{E}[\tilde{\mathbf{x}}^t(n) \mathbf{R} (\mathbf{R}^t \mathbf{R})^{-1} [\mathbf{R}^t \mathbf{R}^e(n) + \mathbf{R}^{et}(n) \mathbf{R}] (\mathbf{R}^t \mathbf{R})^{-2} \\
& [\mathbf{R}^t \mathbf{R}^e(n) + \mathbf{R}^{et}(n) \mathbf{R}] (\mathbf{R}^t \mathbf{R})^{-1} \mathbf{R}^t \tilde{\mathbf{x}}(n)]
\end{aligned} \tag{4.71}$$

$$\begin{aligned}
\xi_e = & \text{E}[(\mathbf{w}_0^t \mathbf{R}^{et}(n-1) - \mathbf{r}^{et}(n-1)) \tilde{\mathbf{x}}(n) \mathbf{x}^t(n) (\mathbf{R}^t \mathbf{R})^{-2} \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \\
& (\mathbf{R}^e(n-1) \mathbf{w}_0 - \mathbf{r}^e(n-1))]
\end{aligned} \tag{4.72}$$

Si se considera que $\lambda \sim 1$ se obtiene un valor en convergencia de

$$D(\infty) = (1-\lambda) \sigma_0^2 \frac{\tilde{\gamma}}{4} + (1-\lambda)^3 \sigma_0^2 \frac{V_e}{4} + (1-\lambda)^3 \frac{\xi_e}{4} \tag{4.73}$$

en completa similitud con la expresión (4.60), donde se estudia la misma situación desde otra aproximación, si se tienen en cuenta las siguientes relaciones: $\tilde{\gamma} \leftrightarrow \text{Tr}\{\boldsymbol{\beta}_Q \mathbf{A}\}$, $V_e \leftrightarrow \text{Tr}\{\boldsymbol{\beta}_Q \mathbf{V}_R\}$ y $\xi_e \leftrightarrow \text{Tr}\{\boldsymbol{\beta}_Q \mathbf{V}_X\}$. Ambos grupos de magnitudes están definidos teniendo en cuenta los mismos contenidos pero tratados de distinta manera.

4.3 Análisis de la convergencia para sistemas no estacionarios

Al igual que se ha llevado a cabo hasta ahora un análisis suponiendo que los coeficientes del filtro \mathbf{w}_0 permanecían constantes en el tiempo, en los próximos apartados se llevará a cabo un estudio análogo pero sin esta restricción. A partir de ahora los coeficientes del filtro pueden evolucionar temporalmente según determina la serie $\mathbf{w}_0(n)$.

4.3.1 Convergencia en media

El primer punto importante es comprobar que la solución $\mathbf{w}(n)$ dada por ORIV en cada instante de tiempo sigue de hecho la evolución de los parámetros del filtro $\mathbf{w}_0(n)$. Esta propiedad es lo que se conoce como capacidad de rastreo.

Para estudiar la capacidad de rastreo se partirá de la ecuación (4.61) en la cual el vector de error a priori se puede escribir como:

$$\begin{aligned} \boldsymbol{\alpha}(n) &= \mathbf{v}(n) - \mathbf{X}^t(n)\mathbf{w}(n-1) = \\ &= \begin{pmatrix} \tilde{\mathbf{x}}^t(n)[\mathbf{z}(n-1) - \boldsymbol{\Phi}(n-1)\mathbf{w}(n-1)] \\ d(n) - \mathbf{x}^t(n)\mathbf{w}(n-1) \end{pmatrix} = \\ &= \begin{pmatrix} 0 \\ \mathbf{e}_0(n) \end{pmatrix} + \begin{pmatrix} \tilde{\mathbf{x}}^t(n)[\mathbf{z}(n-1) - \boldsymbol{\Phi}(n-1)\mathbf{w}(n-1)] \\ \mathbf{x}^t(n)[\mathbf{w}_0(n) - \mathbf{w}(n-1)] \end{pmatrix} \end{aligned} \quad (4.74)$$

Usando la expresión (4.74) conjuntamente con (4.9) y (4.10) tras un poco de álgebra se llega a:

$$\begin{aligned} \mathbf{X}(n)\boldsymbol{\Lambda}^{-1}(n)\boldsymbol{\alpha}(n) &= \\ &= \boldsymbol{\Phi}^t(n)\tilde{\mathbf{x}}(n)\mathbf{e}_0(n) + \lambda\mathbf{x}(n)\tilde{\mathbf{x}}^t(n)[\mathbf{z}(n-1) - \boldsymbol{\Phi}(n-1)\mathbf{w}(n-1)] + \\ &\quad + \boldsymbol{\Phi}^t(n)\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)[\mathbf{w}_0(n) - \mathbf{w}(n-1)] \end{aligned} \quad (4.75)$$

En este estudio particular sólo interesa el comportamiento de ORIV cuando los parámetros incógnita cambian según la serie $\{\mathbf{w}_0(n)\}$ (solución no estacionaria). Por tanto se van a desprestigiar el resto de contribuciones, como por ejemplo el ruido en los estimadores. Además, como sólo interesa la convergencia, se supondrán n grandes y es lícito suponer que $\boldsymbol{\Gamma}^{-1}(n) = (1 - \lambda)^2(\mathbf{R}^t(n)\mathbf{R}(n))^{-1}$, $\mathbf{z}(n) = (1 - \lambda)^{-1}\mathbf{r}(n)$ y $\boldsymbol{\Phi}(n) = (1 - \lambda)^{-1}\mathbf{R}(n)$. Nótese que ahora, debido a que los coeficientes del filtro varían con el tiempo, la estadística de los procesos involucrados puede ser también variante con el tiempo.

Calculando el valor esperado de (4.75) se obtiene:

$$\begin{aligned} \mathbf{E}[\mathbf{X}(n)\mathbf{\Lambda}^{-1}(n)\boldsymbol{\alpha}(n)] &= \frac{1}{1-\lambda}\mathbf{R}^t(n)\mathbf{R}(n)[\mathbf{w}_0(n) - \mathbf{E}[\mathbf{w}(n-1)]] + \\ &+ \frac{\lambda}{1-\lambda}[\mathbf{R}^t(n)\mathbf{r}(n-1) - \mathbf{R}^t(n)\mathbf{R}(n-1)\mathbf{E}[\mathbf{w}(n-1)]] \end{aligned} \quad (4.76)$$

Volviendo ahora a la expresión (4.61) y tomando valor esperado teniendo en cuenta la expresión (4.76) con $\mathbf{r}(n-1)$ reemplazado por $\mathbf{R}(n-1)\mathbf{w}_0(n-1)$ se llega a:

$$\begin{aligned} \mathbf{E}[\mathbf{w}(n)] &= \mathbf{E}[\mathbf{w}(n-1)] + \lambda(1-\lambda)[\mathbf{w}_0(n-1) - \mathbf{E}[\mathbf{w}(n-1)]] + \\ &+ (1-\lambda)(\mathbf{w}_0(n) - \mathbf{E}[\mathbf{w}(n-1)]) = \\ &= \lambda^2\mathbf{E}[\mathbf{w}(n-1)] + (1-\lambda)\mathbf{w}_0(n) + \lambda(1-\lambda)\mathbf{w}_0(n-1) \end{aligned} \quad (4.77)$$

donde se ha asumido que $[\mathbf{R}^t(n)\mathbf{R}(n)]^{-1}\mathbf{R}^t(n)\mathbf{R}(n-1) = \mathbf{I}$, o equivalentemente que el sistema evoluciona suficientemente lento como para poder hacer esta aproximación. La expresión (4.77) es una ecuación en diferencias que tiene como solución particular,

$$\mathbf{E}[\mathbf{w}(z)] = \mathbf{w}_0(z) - \lambda \frac{1-z^{-1}}{1-\lambda^2 z^{-1}} \mathbf{w}_0(z) \quad (4.78)$$

o equivalentemente

$$\mathbf{E}[\mathbf{w}(z)] = \mathbf{w}_0(z) - \lambda(1-z^{-1}) \sum_{i=0}^{\infty} \lambda^{2i} z^{-i} \mathbf{w}_0(z) \quad (4.79)$$

que se puede reescribir en función del tiempo como sigue,

$$\mathbf{E}[\mathbf{w}(n)] = (1-\lambda)\mathbf{w}_0(n) + \frac{1-\lambda^2}{\lambda} \sum_{i=1}^{\infty} \lambda^{2i} \mathbf{w}_0(n-i) \quad (4.80)$$

La solución general de la ecuación homogénea correspondiente a la ecuación (4.77) es $\mathbf{E}[\mathbf{w}(n)]_g = \lambda^{2n}\mathbf{E}[\mathbf{w}(n)]$ que tiende a cero para n grandes. El tiempo de convergencia para la solución de la parte homogénea se puede definir como sigue:

$$\mathbf{E}[\mathbf{w}(\tau)] = \frac{\mathbf{E}[\mathbf{w}(0)]}{e} \rightarrow 2\tau \ln \lambda = -1 \quad (4.81)$$

lo cual significa, asumiendo $\lambda \sim 1$, que

$$\tau = \frac{1}{2(1-\lambda)} \quad (4.82)$$

Cuanto mayor sea λ , más tiempo requiere la solución en converger a la solución particular. A su vez, la solución particular a tiempo n presenta menos contribuciones de tiempos anteriores de $\mathbf{w}_0(n)$ cuanto menor es λ ; esto significa que

un factor de olvido pequeño mejora las capacidades de rastreo. Como ejemplo supóngase que $\mathbf{w}_0(n) = n\mathbf{p}$ donde \mathbf{p} es un vector constante. En este caso la expresión (4.80) implica que $\mathbf{w}(n) = \mathbf{w}_0(n - \lambda/(1 - \lambda^2))$ lo cual quiere decir que la solución estacionaria está retrasada respecto de la solución verdadera un tiempo igual a $\lambda/(1 - \lambda^2)$. Téngase en cuenta que se han discutido dos aspectos distintos: uno es el tiempo que tarda la solución general de la homogénea en anularse y el otro es el tiempo de retraso que presenta la solución estacionaria respecto a la verdadera.

Como comprobación de (4.80), si se toma $\mathbf{w}_0(n) = \mathbf{w}_0$ se obtiene que $E[\mathbf{w}(n)] = \mathbf{w}_0$ como ya se dedujo en el análisis en ambientes estacionarios.

4.3.2 Desviación en media cuadrática en ambientes no estacionarios

En este apartado se supondrá, por razones operativas, que los parámetros del sistema tienen una dependencia temporal lineal, como en el ejemplo precedente:

$$\mathbf{w}_0(n) = n\mathbf{p} \quad (4.83)$$

Procediendo de igual manera que en el apartado anterior se llega fácilmente a que

$$\begin{aligned} \Delta\mathbf{w}(n) = & [\mathbf{I} - \lambda(1 - \lambda)(\mathbf{R}^t(n)\mathbf{R}(n))^{-1}\mathbf{x}(n)\tilde{\mathbf{x}}^t(n)\mathbf{R}(n-1) - \\ & - (1 - \lambda)(\mathbf{R}^t(n)\mathbf{R}(n))^{-1}\mathbf{R}^t(n)\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)]\Delta\mathbf{w}(n-1) + \\ & + (1 - \lambda)(\mathbf{R}^t(n)\mathbf{R}(n))^{-1}\mathbf{R}^t(n)\tilde{\mathbf{x}}(n)e_0(n) + \\ & + (1 - \lambda)(\mathbf{R}^t(n)\mathbf{R}(n))^{-1}\mathbf{R}^t(n)\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)\mathbf{p} - \mathbf{p} \end{aligned} \quad (4.84)$$

una ecuación en diferencias no homogénea con un coeficiente aleatorio. Debido a que $\lambda \sim 1$ las oscilaciones del coeficiente aleatorio son pequeñas y se puede considerar que se comporta como su valor medio. Es decir, volvemos a recurrir a la técnica del promediado directo, con lo que la ecuación anterior se convierte en

$$\begin{aligned} \Delta\mathbf{w}(n) = & \lambda^2\Delta\mathbf{w}(n-1) + (1 - \lambda)(\mathbf{R}^t(n)\mathbf{R}(n))^{-1}\mathbf{R}^t(n)\tilde{\mathbf{x}}(n)e_0(n) - \\ & - [\mathbf{I} - (1 - \lambda)(\mathbf{R}^t(n)\mathbf{R}(n))^{-1}\mathbf{R}^t(n)\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)]\mathbf{p} \end{aligned} \quad (4.85)$$

Construyendo el producto $\Delta\mathbf{w}^t(n)\Delta\mathbf{w}(n)$, tomando valor esperado y teniendo en cuenta que $e_0(n)$ es independiente del resto de variables aleatorias y de media cero, se llega fácilmente a que:

$$\begin{aligned} D(n) = & \lambda^4 D(n-1) - 2\lambda^2 E[\Delta\mathbf{w}^t(n-1)]\lambda\mathbf{p} + \mathbf{p}^t\mathbf{p} - 2(1 - \lambda)\mathbf{p}^t\mathbf{p} + \\ & + (1 - \lambda)^2 \sigma_0^2 \text{Tr}[\mathbf{R}(n)(\mathbf{R}^t(n)\mathbf{R}(n))^{-2}\mathbf{R}^t(n)\mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}(n)] + \\ & + (1 - \lambda)^2 \mathbf{p}^t E[\mathbf{x}(n)\tilde{\mathbf{x}}^t(n)\mathbf{R}(n)(\mathbf{R}^t(n)\mathbf{R}(n))^{-2}\mathbf{R}^t(n)\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)]\mathbf{p} \end{aligned} \quad (4.86)$$

Antes de continuar es conveniente hacer la siguiente consideración: tomando valor esperado en (4.85) se obtiene

$$\mathbb{E}[\Delta \mathbf{w}(n)] = \lambda^2 \mathbb{E}[\Delta \mathbf{w}(n-1)] - \lambda \mathbf{p} \quad (4.87)$$

y para su valor en convergencia

$$\mathbb{E}[\Delta \mathbf{w}(\infty)] = -\frac{\lambda}{1-\lambda^2} \mathbf{p} \quad (4.88)$$

Si esta expresión también se considera válida para n grandes (no sólo estrictamente en convergencia), se puede aceptar que

$$\mathbb{E}[\mathbf{w}(n)] = \mathbf{w}_0(n) - \frac{\lambda}{1-\lambda^2} \mathbf{p} \quad n \gg 1 \quad (4.89)$$

de forma que esta solución tiene un tiempo de retraso asociado igual a

$$\mathbb{E}[\mathbf{w}(n)] = \mathbf{w}_0(n - \tau') \quad \tau' = \frac{\lambda}{1-\lambda^2} \quad (4.90)$$

Este tiempo de retraso es el mismo obtenido después del comentario acerca de la ecuación (4.82) para este mismo ejemplo, como era de esperar.

Teniendo en cuenta la expresión (4.89), la ecuación (4.86) se puede reescribir como

$$\begin{aligned} D(n) = & \lambda^4 D(n-1) + \frac{2\lambda^4 - 2\lambda^3 + \lambda^2 + 2\lambda - 1}{1-\lambda^2} \mathbf{p}^t \mathbf{p} + \\ & + (1-\lambda)^2 \sigma_0^2 \tilde{\gamma}(n) + (1-\lambda)^2 \mathbf{p}^t \mathbf{H}(n) \mathbf{p} \end{aligned} \quad (4.91)$$

donde

$$\tilde{\gamma}(n) = \text{Tr}[\mathbf{R}(n)(\mathbf{R}(n)\mathbf{R}(n))^{-2}\mathbf{R}^t(n)\mathbf{R}_{\tilde{x}\tilde{x}}(n)] \quad (4.92)$$

$$\mathbf{H}(n) = \mathbb{E}[\mathbf{x}(n)\tilde{\mathbf{x}}^t(n)\mathbf{R}(n)(\mathbf{R}^t(n)\mathbf{R}(n))^{-2}\mathbf{R}^t(n)\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)] \quad (4.93)$$

La matriz \mathbf{H} y el escalar $\tilde{\gamma}$ dependen del índice temporal n porque el sistema no es estacionario. Recuérdese asimismo que esta ecuación sólo es válida para n grandes. La ecuación en diferencias (4.91) converge a

$$D(\infty) = \frac{1}{(1-\lambda)^2} \frac{\mathbf{p}^t \mathbf{p}}{4} + \frac{7}{2} \frac{\mathbf{p}^t \mathbf{p}}{4} + (1-\lambda) \sigma_0^2 \frac{\tilde{\gamma}}{4} + (1-\lambda) \frac{\mathbf{p}^t \mathbf{H} \mathbf{p}}{4} \quad (4.94)$$

donde se ha considerado que λ es muy próximo a uno y se han empleado las siguientes definiciones:

$$\mathbf{H} \equiv 4(1-\lambda) \lim_{n \rightarrow \infty} \sum_{i=0}^{n-1} \lambda^{4i} \mathbf{H}(n-i) \quad (4.95)$$

$$\tilde{\gamma} \equiv 4(1-\lambda) \lim_{n \rightarrow \infty} \sum_{i=0}^{n-1} \lambda^{4i} \tilde{\gamma}(n-i) \quad (4.96)$$

Los sumandos que intervienen en el valor en convergencia se han ordenado en importancia decreciente. Obsérvese que lo que más influye es la no estacionariedad del sistema en estudio y sin embargo el ruido de medición ocupa el tercer lugar. La única contribución que no se ha tenido en cuenta es el ruido de estimación de \mathbf{R} y \mathbf{r} que ya se estudió en los apartados 4.2.3 y 4.2.4. Estrictamente hablando sólo el análisis llevado a cabo en 4.2.4 es completamente paralelo, en cuanto a hipótesis, del realizado aquí ya que en el otro apartado se utilizó otro planteamiento. La equivalencia teórica y práctica entre ambos planteamientos se demuestra en el apartado 4.5.

4.4 Condiciones óptimas para el rastreo

Complementando los aspectos estudiados en la ecuación (4.73) con los de la ecuación (4.94) (ambos estudiados bajo las mismas hipótesis) se puede obtener una ecuación para el MSD en convergencia que englobe todos los factores estudiados hasta ahora:

$$D(\infty) = \frac{1}{(1-\lambda)^2} \frac{\mathbf{p}^t \mathbf{p}}{4} + \frac{7}{2} \frac{\mathbf{p}^t \mathbf{p}}{4} + (1-\lambda) \frac{\sigma_0^2 \tilde{\gamma}}{4} + (1-\lambda) \frac{\mathbf{p}^t \mathbf{H} \mathbf{p}}{4} + (1-\lambda)^3 \frac{\sigma_0^2 V_e + \xi_e}{4} \quad (4.97)$$

Donde a V_e y a ξ_e hay que darle un tratamiento análogo al dado a $\tilde{\gamma}$ en la expresión (4.96) (que se puede considerar como su definición en ambientes no estacionarios).

En [MB91] se obtiene un resultado análogo para RLS: el ruido de medición contribuye como $(1-\lambda)$ y el sumando debido a la no estacionariedad como $(1-\lambda)^{-2}$.

Una cuestión importante se plantea al estudiar esta expresión, ¿Cuál es el λ óptimo que da un $D(\infty)$ mínimo? Minimizando (4.97) con respecto a λ , suponiendo por sencillez que V_e y ξ_e son nulos, resulta

$$\lambda_{opt} = 1 - \left[\frac{2\mathbf{p}^t \mathbf{p}}{\sigma_0^2 \tilde{\gamma} + \mathbf{p}^t \mathbf{H} \mathbf{p}} \right]^{1/3} \quad (4.98)$$

A la vista de la cual es interesante discutir tres casos particulares:

1. Cuando la pendiente de la variación lineal del modelo no estacionario que se rastrea se hace muy grande, λ_{opt} decrece.
2. Por el contrario, como ya es conocido, si la pendiente se hace cada vez más pequeña el factor de olvido tiende a la unidad:

$$\lim_{p \rightarrow 0} \lambda_{opt} = 1 \quad (4.99)$$

3. Asimismo cuando el ruido de medición aumenta λ_{opt} también tiende a la unidad, porque se requiere una memoria larga para conseguir buenos estimadores.

La expresión (4.97) representa el valor en convergencia para el error cuadrático medio teniendo en cuenta las dos posibles fuentes de error. La primera, ecuación (4.73), incluye el error cuadrático medio debido a que la memoria finita ($\lambda \neq 1$) del algoritmo aumenta la varianza de los estimadores de las magnitudes implicadas en la resolución de (4.1). La segunda, ecuación (4.94), da cuenta del error cuadrático medio debido a la estimación de sistemas no estacionarios. Se sabe que (4.73) es el MSD en estado estacionario debido a que $\lambda \neq 1$. Si el factor de olvido fuese la unidad se tendría que $(1 - \lambda)$ pasa a ser $1/n$ y esta contribución se anularía para n grandes. Sin embargo (4.94) aparece al tener que rastrear sistemas no estacionarios (en particular en este caso sistemas con evolución temporal lineal) y lógicamente, para $(1 - \lambda) \rightarrow 1/n$ divergería, ya que si $\lambda = 1$ se elimina la posibilidad de rastreo. Como se ve, los puntos 1 y 3 anteriores presentan situaciones opuestas, de las cuales (4.98) sopesa ambas y da la óptima. Las implicaciones que acarrea cada caso se entenderán mejor a la vista de las simulaciones que se presentan más adelante en esta memoria.

4.5 Equivalencia de las aproximaciones utilizadas

En las deducciones realizadas en este capítulo se han empleado dos ecuaciones de partida distintas, en este apartado se comprueba que ambas aproximaciones son equivalentes bajo las hipótesis empleadas.

1ª aproximación

Se parte de (4.15) junto con la definición de $\mathbf{K}(n)$ dada en (4.16) que da lugar a

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \mathbf{\Gamma}^{-1}(n)\mathbf{X}(n)\mathbf{\Lambda}^{-1}(n)\boldsymbol{\alpha}(n) \quad (4.100)$$

Si se le resta \mathbf{w}_0 a ambos miembros queda

$$\Delta\mathbf{w}(n) = \Delta\mathbf{w}(n-1) + \mathbf{\Gamma}^{-1}(n)\mathbf{X}(n)\mathbf{\Lambda}^{-1}(n)\boldsymbol{\alpha}(n) \quad (4.101)$$

y si ahora se multiplican ambos miembros por $\mathbf{\Gamma}(n)$ y se usa (4.8) resulta:

$$\mathbf{\Gamma}(n)\Delta\mathbf{w}(n) = \lambda^2\mathbf{\Gamma}(n-1)\Delta\mathbf{w}(n-1) + \mathbf{X}(n)\mathbf{\Lambda}^{-1}(n)\boldsymbol{\alpha}(n) \quad (4.102)$$

donde como ya es conocido $\boldsymbol{\alpha}_0(n) = \mathbf{v}(n) - \mathbf{X}^t(n)\mathbf{w}_0$. Hasta aquí el cálculo ha sido exacto, es decir, sin aproximaciones. Si ahora se multiplica $\mathbf{\Gamma}^{-1}(n)$ a

ambos miembros y se supone que $\mathbf{\Gamma}^{-1}(n)\mathbf{\Gamma}(n-1) = \mathbf{I}$ para n grandes, se llega finalmente a

$$\Delta\mathbf{w}(n) = \lambda^2\Delta\mathbf{w}(n-1) + \mathbf{\Gamma}^{-1}(n)\mathbf{X}(n)\mathbf{\Lambda}^{-1}(n)\boldsymbol{\alpha}(n) \quad (4.103)$$

que es el punto de partida de los apartados 4.2.2 y 4.2.3.

2ª aproximación

Partiendo de la expresión (4.100) y teniendo en cuenta que

$$\begin{aligned} \boldsymbol{\alpha}(n) &= \mathbf{v}(n) - \mathbf{X}^t(n)\mathbf{w}(n-1) = \mathbf{v}(n) - \mathbf{X}^t(n)\mathbf{w}_0 - \mathbf{X}^t(n)\Delta\mathbf{w}(n-1) = \\ &= \boldsymbol{\alpha}_0(n) - \mathbf{X}^t(n)\Delta\mathbf{w}(n-1) \end{aligned} \quad (4.104)$$

se obtiene:

$$\begin{aligned} \mathbf{w}(n) &= \mathbf{w}(n-1) - \mathbf{\Gamma}^{-1}(n)\mathbf{X}(n)\mathbf{\Lambda}^{-1}(n)\mathbf{X}^t(n)\Delta\mathbf{w}(n-1) + \\ &+ \mathbf{\Gamma}^{-1}(n)\mathbf{X}(n)\mathbf{\Lambda}^{-1}(n)\boldsymbol{\alpha}_0(n) \end{aligned} \quad (4.105)$$

con lo que restando \mathbf{w}_0 a ambos miembros se llega a:

$$\begin{aligned} \Delta\mathbf{w}(n) &= [\mathbf{I} - \mathbf{\Gamma}^{-1}(n)\mathbf{X}(n)\mathbf{\Lambda}^{-1}(n)\mathbf{X}^t(n)]\Delta\mathbf{w}(n-1) + \\ &+ \mathbf{\Gamma}^{-1}(n)\mathbf{X}(n)\mathbf{\Lambda}^{-1}(n)\boldsymbol{\alpha}_0(n) \end{aligned} \quad (4.106)$$

Hasta aquí el cálculo es exacto. Para continuar se recurrirá a la técnica del promediado directo que consiste en sustituir el factor entre corchetes por su valor medio. Esta aproximación está justificada suponiendo que esta cantidad aleatoria se desvía poco de su valor medio, con lo que no se comete gran error. Además, si se supone que los componentes que integran este factor son estadísticamente independientes, lo cual es válido para n grande, es necesario saber cómo se comportan las siguientes cantidades:

$$\begin{aligned} \mathbf{X}(n)\mathbf{\Lambda}^{-1}(n)\mathbf{X}^t(n) &= \lambda\mathbf{x}(n)\tilde{\mathbf{x}}^t(n)\mathbf{\Phi}(n-1) + \mathbf{\Phi}^t(n)\tilde{\mathbf{x}}(n)\mathbf{x}^t(n) \xrightarrow{\mathbb{E}[\cdot]} \\ &\rightarrow \frac{1+\lambda}{1-\lambda}\mathbf{R}^t\mathbf{R} \end{aligned} \quad (4.107)$$

$$\mathbf{\Gamma}^{-1}(n) \xrightarrow{\mathbb{E}[\cdot]} (1-\lambda)^2(\mathbf{R}^t\mathbf{R})^{-1} \quad (4.108)$$

que teniéndolas en cuenta en (4.106) lo convierten en:

$$\Delta\mathbf{w}(n) = \lambda^2\Delta\mathbf{w}(n-1) + \mathbf{\Gamma}^{-1}(n)\mathbf{X}(n)\mathbf{\Lambda}^{-1}(n)\boldsymbol{\alpha}_0(n) \quad (4.109)$$

Las ecuaciones (4.103) y (4.109) son idénticas de manera que ambas aproximaciones son equivalentes. La segunda consiste en aplicar promediado directo

a (4.100), que es el método empleado en los apartados 4.2.4 y 4.3.1. Esto básicamente significa que suponer que $\Gamma^{-1}(n)\Gamma(n-1) = \mathbf{I}$ (lo cual se cumple mejor para n grandes) es equivalente a aplicar la técnica del promediado directo, lo cual es lógico atendiendo a la propiedad de ergodicidad de las series involucradas.

En consecuencia, aunque se han obtenido diferentes contribuciones a $D(n)$ mediante diferentes aproximaciones, ahora no es de extrañar que los resultados sean análogos.

Capítulo 5

Identificación ciega de sistemas MA

Introducción

Tal y como se ha ido exponiendo en los capítulos anteriores el algoritmo ORIV es el más completo y general de los tres presentados, es por ello por lo que se ha elegido para ser estudiado mediante simulaciones.

Se va a aplicar al problema de la identificación ciega de modelos de media móvil (MA), mediante el empleo de ecuaciones que relacionan los coeficientes del sistema con cumulantes de la salida del mismo.

Todos los aspectos estudiados teóricamente se van a analizar mediante simulaciones incidiendo asimismo en la influencia que el orden de los cumulantes empleados tiene en los resultados finales y en el proceso de la convergencia. Dos objetivos se persiguen por tanto en este capítulo:

1. Analizar el comportamiento del algoritmo ORIV de forma que complete el análisis teórico.
2. Estudiar las ventajas de emplear exclusivamente estadística de alto orden en la identificación ciega de sistemas MA.

Las hipótesis bajo las cuales se va a proceder a resolver el problema de la identificación ciega se exponen en el apartado 5.1, con ellas se pueden deducir ecuaciones que relacionan los coeficientes con los cumulantes de la salida, tarea que

se realizará en el apartado 5.2. La deducción se hará tanto en el dominio del tiempo como mediante transformada Z. Posteriormente se detallará cómo se puede emplear el algoritmo ORIV para resolver las ecuaciones propuestas en el apartado 5.3, donde se introducirá el concepto de ergodicidad. Antes de proceder a mostrar los resultados experimentales es necesario fijar conceptos tales como la relación señal-ruido, el error cuadrático medio, etc. Estos conceptos se definirán en el apartado 5.4. Los resultados y el análisis de las simulaciones llevadas a cabo en ambientes estacionarios ruidosos se exponen en el apartado 5.5, en el apartado 5.6 se mostrarán y analizarán los resultados obtenidos en ambientes no estacionarios, donde los coeficientes del sistema experimentan tanto una evolución temporal lineal como aleatoria. Finalmente en el apartado 5.7 se complementan los resultados anteriores trabajando en ambientes no estacionarios con ruido, además se discute el interesante fenómeno del seguimiento límite.

5.1 Hipótesis de partida para la identificación ciega

La necesidad de identificar y modelar sistemas mediante un conjunto de parámetros aparece al plantear infinidad de problemas tal y como se puso de manifiesto en la introducción general de esta memoria. En la mayoría de las veces es más que suficiente con modelar el sistema mediante una ley lineal, lo cual facilita el problema matemáticamente hablando enormemente. También es conocido de capítulos precedentes que un sistema lineal viene perfectamente definido por su respuesta impulso y por tanto el objetivo primordial de cualquier identificación es determinar ésta.

Para evitar problemas indeseables con el modelo obtenido en el proceso de identificación, tales como falta de estabilidad del mismo, es usual optar por un modelo con respuesta impulso finita de orden q , donde los parámetros que lo definen son $\{b_0, b_1, \dots, b_q\}$, es decir,

$$x(n) = \sum_{i=0}^q b_i w(n-i) \quad (5.1)$$

también conocido como un modelo de media móvil (moving average) MA(q), donde $\{w(n)\}$ es la secuencia de excitación del sistema. Esta secuencia es a veces accesible, al igual que la salida $\{x(n)\}$, con lo que la identificación se lleva a cabo sin grandes problemas.

Hay, sin embargo, situaciones más complejas en las que sólo la salida es conocida y con ella hay que llevar a cabo la identificación. Más aún, muchas veces ésta se encuentra contaminada con ruido aditivo $v(n)$ y es la salida ruidosa

$y(n)$ la que ha de emplearse. Téngase en cuenta que

$$y(n) = x(n) + v(n) \quad (5.2)$$

Como ya es sabido este nuevo problema recibe el nombre de identificación ciega y para resolverlo, de aquí en adelante, se tendrán en cuenta las siguientes suposiciones [Rui95]:

- A) La secuencia de excitación del modelo $\{w(n)\}$ no es accesible y es un proceso aleatorio no gaussiano de media cero, independiente e idénticamente distribuido con al menos $0 < |\gamma_{kw}| < \infty$ y $0 < |\gamma_{k-1w}| < \infty$ con $k > 3$, siendo γ_{kw} el cumulante (para intervalos temporales nulos) de orden k -ésimo de la variable aleatoria $w(n)$.
- B) El ruido aditivo $\{v(n)\}$ es un proceso gaussiano, blanco o coloreado, independiente de la señal de entrada $\{w(n)\}$, y en consecuencia de la salida $\{x(n)\}$.
- C) El sistema MA es causal e invariante en el tiempo, posiblemente de fase no mínima, con $b_q \neq 0$.

Empleando estas hipótesis, en el próximo apartado, se van a obtener ecuaciones que ligen los cumulantes de la salida del sistema con los parámetros que lo determinan. Cada ecuación que se derive tendrá sus propiedades particulares, la importancia de las cuales se pondrá de manifiesto en posteriores simulaciones.

5.2 Identificación ciega mediante cumulantes

Ya se ha puesto de manifiesto con anterioridad en esta memoria la idoneidad de usar cumulantes en vez de momentos. Sus características los hacen muy útiles a la hora de resolver ciertos problemas y para la identificación ciega no es una excepción (véase por ejemplo [CN90], [DL01], [HOS00], [JK92], [Kai99], [PF91]), siendo las más importantes las siguientes [Men91]:

1. Son operadores lineales, de manera que cualquier cumulante de la suma de dos conjuntos de variables aleatorias independientes equivale a la suma de ese mismo cumulante de cada una de las variables por separado:

$$C[x_1 + y_1, x_2 + y_2, \dots, x_n + y_n] = C[x_1, x_2, \dots, x_n] + C[y_1, y_2, \dots, y_n] \quad (5.3)$$

2. Los cumulantes de orden mayor que 2 de cualquier proceso gaussiano, de contenido espectral arbitrario, son nulos.

3. Los cumulantes llevan información sobre la fase del sistema de forma que pueden identificar sistemas de fase no nímima.

La principal consecuencia de los puntos 1 y 2 es que el cumulante de la salida contaminada $y(n)$ coincide con el cumulante de la salida sin contaminar $x(n)$ siempre y cuando el cumulante tenga orden mayor que dos:

$$c_{ky}(\tau_1, \dots, \tau_{k-1}) = c_{kx}(\tau_1, \dots, \tau_{k-1}) \quad (5.4)$$

nótese que las series se suponen estacionarias de orden k por lo que los cumulantes sólo dependen de las diferencias entre los tiempos estudiados, tal y como muestra la notación empleada.

Por tanto, para resolver el problema de la identificación ciega en la situación descrita en los últimos párrafos lo que queda es encontrar relaciones entre los cumulantes de la salida del sistema y los coeficientes del modelo MA. Existen diversas formas de encontrar dichas relaciones, en esta memoria se derivarán siguiendo dos caminos distintos: mediante el empleo de la transformada Z y mediante relaciones temporales entre cumulantes.

5.2.1 Obtención de ecuaciones mediante la transformada Z

En este apartado se van a encontrar relaciones entre cumulantes de orden k y $k-1$ -ésimo. La forma de relacionarlos va a ser mediante combinaciones lineales de los mismos donde los coeficientes son ciertas funciones de los parámetros del modelo. Para obtener tales relaciones se parte ([Rui95], [CRGM95]) de las siguientes secuencias:

$$s_1(\tau_1, \tau_2, \dots, \tau_k) = c_{kx}(\tau_1, \tau_2, \dots, \tau_{k-1})\delta(\tau_k) \quad (5.5)$$

$$s_2(\tau_1, \tau_2, \dots, \tau_k) = c_{k-1x}(\tau_1, \tau_2, \dots, \tau_{k-2})\delta(\tau_{k-1})\delta(\tau_k - \tau_{k-2}) \quad (5.6)$$

Tomando transformada Z a ambas igualdades resulta:

$$\begin{aligned} S_1(z_1, z_2, \dots, z_k) &= C_{kx}(z_1, z_2, \dots, z_{k-1}) = \\ &= \gamma_{kw} B(z_1)B(z_2)\dots B(z_{k-1})B((z_1 z_2 \dots z_{k-1})^{-1}) \end{aligned} \quad (5.7)$$

$$\begin{aligned} S_2(z_1, z_2, \dots, z_k) &= C_{k-1x}(z_1, z_2, \dots, z_{k-2} z_k) = \\ &= \gamma_{k-1w} B(z_1)B(z_2)\dots B(z_{k-3})B(z_k z_{k-2})B((z_1 z_2 \dots z_k)^{-1}) \end{aligned} \quad (5.8)$$

Donde se ha empleado la siguiente expresión debida a Brillinger y a Rosenblatt [BR67]:

$$c_{kx}(\tau_1, \tau_2, \dots, \tau_{k-1}) = \gamma_{kw} \sum_{i=0}^q b_i b_{i+\tau_1} \dots b_{i+\tau_{k-1}} \quad (5.9)$$

válida para el modelo MA(q) definido en la ecuación (5.1). Combinando a continuación las expresiones (5.7) y (5.8) se puede obtener que

$$\begin{aligned} & \gamma_{k-1w} B(z_k z_{k-2}) B((z_1 z_2 \dots z_k)^{-1}) C_{kx}(z_1, z_2, \dots, z_{k-1}) = \\ & = \gamma_{kw} B(z_{k-2}) B(z_{k-1}) B((z_1 z_2 \dots z_{k-1})^{-1}) C_{k-1x}(z_1, z_2, \dots, z_{k-2} z_k) \end{aligned} \quad (5.10)$$

Tomando la transformada Z inversa de orden k se obtiene la relación equivalente en el dominio del tiempo:

$$\begin{aligned} & \gamma_{k-1w} \sum_{i=0}^q b_i b_{\tau_k+i} c_{kx}(\tau_1 + i, \tau_2 + i, \dots, \tau_{k-2} - \tau_k, \tau_{k-1}) = \\ & = \gamma_{kw} \sum_{i=0}^q b_i b_{\tau_{k-2}-\tau_k+i} b_{\tau_{k-1}+i} c_{k-1x}(\tau_1 + i, \tau_2 + i, \dots, \tau_{k-3} - i, \tau_k) \end{aligned} \quad (5.11)$$

Esta ecuación general se particularizará a dos casos concretos que serán los empleados en las simulaciones posteriores. Como se puede apreciar, conociendo los cumulantes de la salida $x(n)$ del modelo MA se pueden determinar los coeficientes del modelo.

Particularización primera. Cumulantes de orden 2 y 3

Si en la expresión (5.11) se hace $k = 3$, $\tau_2 = 0$ y $\tau_1 = m$ se obtiene la siguiente relación:

$$\sum_{k=0}^q b_k^2 c_{2x}(m-k) = \frac{\sigma_w^2}{\gamma_{3w}} \sum_{k=0}^q b_k c_{3x}(m-k, m-k) \quad (5.12)$$

que tiene sentido cuando $-q \leq m \leq 2q$. Es una expresión que relaciona cumulantes de orden 2 y 3 con los parámetros del sistema, fue obtenida por primera vez por Giannakis y Mendel [GM89] y por este motivo se le suele conocer como ecuación GM.

Particularización segunda. Cumulantes de orden 3 y 4

Si en la expresión (5.11) se hace $k = 4$, $\tau_4 = \tau_3 = \tau_2 = 0$ y $\tau_1 = m$ se puede llegar, tras alguna modificación, a:

$$\sum_{k=0}^q b_k^3 c_{3x}(m-k, m-k) = \frac{\gamma_{3w}}{\gamma_{4w}} \sum_{k=0}^q b_k^2 c_{4x}(m-k, m-k, m-k) \quad (5.13)$$

que relaciona los cumulantes de orden 3 y 4 de la salida del sistema con los parámetros que lo definen. Esta ecuación fue propuesta de forma independiente en [Rui95] y [NKSK95].

5.2.2 Obtención de ecuaciones en el dominio del tiempo

También es posible obtener una expresión que generaliza la ecuaciones (5.12) y (5.13) en el dominio del tiempo. Para ello el punto de partida es la ecuación de Brillinger-Rosenblatt (5.9), gracias a la cual es posible escribir que

$$\sum_{i=0}^q b_i^k c_{kx}(0, (k..2), 0, i-m) = \gamma_{kw} \sum_{i,l}^q b_i^k b_l^{k-1} b_{l+i-m} \quad (5.14)$$

realizando sólomente la sumatoria en i y teniendo en cuenta de nuevo la expresión (5.9) queda:

$$\sum_{i=0}^q b_i^k c_{kx}(0, (k..2), 0, i-m) = \frac{\gamma_{kw}}{\gamma_{k+1w}} \sum_l^q b_l^{k-1} c_{k+1x}(0, (k..1), 0, l-m) \quad (5.15)$$

Renombrando l como i y teniendo en cuenta las simetrías de los cumulantes, se llega finalmente a la expresión deseada.

$$\sum_{i=0}^q b_i^k c_{kx}(m-i, (k..1), m-i) = \frac{\gamma_{kw}}{\gamma_{k+1w}} \sum_i^q b_i^{k-1} c_{k+1x}(m-i, (k), m-i) \quad (5.16)$$

A partir de ella es directo obtener la expresión (5.12) haciendo $k = 2$ y la expresión (5.13) haciendo $k = 3$ (ya derivada en el dominio del tiempo en [ARBC02a]).

5.2.3 Características de las ecuaciones obtenidas

Las características de las expresiones (5.12) y (5.13) se pueden deducir de las enunciadas para los cumulantes en la página 93. Básicamente se pueden resumir en dos: inmunidad al ruido gaussiano añadido y capacidad para determinar la fase del sistema. La inmunidad al ruido se refiere al hecho de que en (5.12) y (5.13) se pueden sustituir los cumulantes de $x(n)$ por los cumulantes de $y(n)$ tal y como sugiere la expresión (5.4), por lo que las expresiones a utilizar finalmente serán:

$$\sum_{k=0}^q b_k^2 c_{2y}(m-k) = \frac{\sigma_w^2}{\gamma_{3w}} \sum_{k=0}^q b_k c_{3y}(m-k, m-k) \quad (5.17)$$

$$\sum_{k=0}^q b_k^3 c_{3y}(m-k, m-k) = \frac{\gamma_{3w}}{\gamma_{4w}} \sum_{k=0}^q b_k^2 c_{4y}(m-k, m-k, m-k) \quad (5.18)$$

La importancia de este resultado radica en que incluso recurriendo a los cumulantes de la salida ruidosa $y(n)$ del sistema, se puede establecer una relación con

los coeficientes del sistema, de manera que es posible llevar a cabo la identificación sin sesgo. En realidad esto no es totalmente cierto ya que para $k = 2$ la expresión (5.4) no es válida y por tanto tampoco es totalmente cierta la ecuación (5.17); sin embargo, debido a que en muchas circunstancias el único dato conocido es la salida contaminada por ruido, se aceptará dicha ecuación (asumiendo sus limitaciones) y en las simulaciones posteriores se analizará cómo le afecta a su falta de inmunidad el ruido aditivo gaussiano.

Sin embargo ambas ecuaciones llevan información sobre la fase del sistema ya que incluyen cumulantes de orden mayor que dos. La ecuación GM incluye cumulantes de orden 3 y la expresión (5.18) incluye cumulantes de orden 3 y 4, por tanto serán perfectamente capaces de identificar sistemas de fase no mínima.

La autoría de la ecuación de GM parece no tener duda y en la literatura especializada siempre aparece con ese nombre debido a Giannakis y a Mendel [GM89]. Sin embargo la autoría de la ecuación (5.18) no se le puede adjudicar fácilmente a ningún autor porque se tiene conocimiento de su obtención tanto por Ruiz [Rui95] como por Y.J. Na et al. [NKSK95] mediante transformada Z, simultáneamente. La obtención de la ecuación con cumulantes de orden 3 y 4 se realizó en [ARBC02a] en el dominio del tiempo y en esta memoria se ha presentado en la ecuación (5.16) una expresión obtenida asimismo en el dominio del tiempo que generaliza tanto ésta como la de GM, ya que relaciona cumulantes de orden k y $k + 1$, $\forall k$.

Estas propiedades teóricas deben completarse con las propiedades de sus estimadores, ya que como es conocido los cumulantes necesarios hay que estimarlos ya que no suele ser habitual conocer perfectamente la estadística de las series involucradas. Este segundo conjunto de propiedades se discutirá más adelante en lo referente al análisis de los resultados de las simulaciones.

5.3 Resolución mediante el algoritmo ORIV

Una vez obtenidas las ecuaciones necesarias para identificar de forma ciega los sistemas MA, que son el culmen y resumen de un conjunto de hipótesis y decisiones sobre las herramientas a utilizar, es necesario dar un paso más y elegir la forma de resolver dichas ecuaciones.

Los métodos que trabajan con lotes de datos suelen dar resultados más exactos, pero son poco útiles en situaciones donde el sistema evolucione con el tiempo porque resolver las ecuaciones para cada instante de tiempo requiere gran carga computacional. Para ese tipo de situaciones se diseñaron los algoritmos adaptativos y con ellos se resolverán las ecuaciones presentadas anteriormente. En particular, de los tres algoritmos estudiado, RLS, RIV y ORIV se elegirá

este último por dos motivos:

1. Su formalismo permite el uso de una variable instrumental, necesaria como se verá un poco más adelante para resolver ecuaciones que involucren estadística de alto orden.
2. Permite tratar mayor cantidad de información ya que es capaz de resolver sistemas de ecuaciones sobredeterminados.

También se han desarrollado algoritmos ORIV rápidos [BSN00]. En los próximos apartados se detalla cómo es posible adaptar el algoritmo ORIV para resolver ecuaciones del tipo de (5.17) y (5.18), para ello es necesario conocer cómo se estiman los momentos y cumulantes recurriendo a la ergodicidad.

5.3.1 Ergodicidad

Tal y como se ha afirmado antes es necesario recurrir a estimadores de los cumulantes. Como es habitual estos se calcularán mediante promedios temporales porque la serie $y(n)$ se supondrá ergódica. Una serie de variables aleatorias se dirá que es ergódica cuando los promedios llevados a cabo en las distintas variables de la serie coincidan con promedios en realizaciones. Como ejemplo supóngase que se quiere calcular el siguiente valor esperado $E[s_1 s_2]$ donde s_1 y s_2 son dos variables aleatorias pertenecientes a una serie dada por $\{s_1, s_2, \dots, s_N\}$. Supóngase además que se dispone de una realización de dicha serie dada por $\{r_1, r_2, \dots, r_N\}$. Según la propiedad de ergodicidad, se puede aproximar

$$E[s_1 s_2] = \frac{1}{N-1} \sum_{i=2}^N r_i r_{i-1} \quad (5.19)$$

Para el caso más concreto que ahora nos ocupa, el índice utilizado en las series con las que se trabaja es el tiempo, por lo que la propiedad de ergodicidad se traduce ahora en afirmar que los promedios temporales igualan a promedios en realizaciones. De todas formas esta equivalencia se da completamente sólo en el caso de que se conozcan infinitas muestras temporales; como esto no es posible, en la práctica mediante promedios temporales no se consigue más que estimadores de los promedios en realizaciones. Es decir, sólo es posible estimar los momentos de las series aleatorias involucradas. De forma genérica, si se conocen L datos de la salida del sistema $\{y(1), y(2), \dots, y(L)\}$, los momentos de orden k se estiman mediante la siguiente expresión:

$$\hat{m}_{ky}(\tau_1, \tau_2, \dots, \tau_{k-1}) = \frac{1}{L} \sum_{i=0}^L y(i) y(i + \tau_1) y(i + \tau_2) \dots y(i + \tau_{k-1}) \quad (5.20)$$

El símbolo $\hat{\cdot}$ se ha incluido para indicar que es un estimador. A partir de los momentos se pueden estimar los cumulantes según su definición.

Téngase en cuenta que si una serie es ergódica entonces también es estacionaria (desde el punto de vista estadístico): si la serie no es estacionaria tiene poco sentido utilizar, por ejemplo, las parejas $y(1)y(2)$ y $y(14)y(15)$ para estimar el valor esperado $E[y_1y_2]$. De todas formas, catalogar a una serie como estacionaria depende del intervalo de observación. Si la estadística de una serie cambia poco con el tiempo y el intervalo no es muy grande, esta serie se puede suponer como estacionaria.

Estimación adaptativa de momentos

En el caso de procesos no estacionarios la argucia usada más comúnmente es la introducción de un factor de olvido (forgetting factor), notado en esta memoria por λ , que dé menos peso a las muestras más antiguas y mayor a las nuevas. Por ejemplo, para la varianza de la serie $\{y(n)\}$ se puede proponer:

$$\tilde{\sigma}^2(n) = \sum_{i=1}^n \lambda^{n-i} y^2(i) \quad (5.21)$$

En esta idea se basa también ORIV para estimar los cumulantes de tercer y cuarto orden, mediante las definiciones (4.2) y (4.3) teniendo en cuenta (5.31) y (5.37) según la ecuación a resolver.

Sería conveniente introducirle algún tipo de normalización¹ al estimador anterior. Se propone la siguiente: suponiendo que se trabaja en un entorno estacionario y tomando valor esperado resulta que (5.21) se convierte en:

$$E[\tilde{\sigma}^2(n)] = \sum_{i=1}^n \lambda^{n-i} \sigma_y^2 = \frac{1 - \lambda^n}{1 - \lambda} \sigma_y^2 \quad (5.22)$$

Por lo que parece natural definir el estimador de la siguiente manera:

$$\hat{\sigma}^2(n) = \frac{1 - \lambda}{1 - \lambda^n} \sum_{i=1}^n \lambda^{n-i} y^2(i) \quad (5.23)$$

Por su utilidad es conveniente derivar su ecuación de actualización, obtenida sin dificultad:

$$\hat{\sigma}^2(n+1) = \frac{1 - \lambda^n}{1 - \lambda^{n+1}} \lambda \hat{\sigma}^2(n) + \frac{1}{1 - \lambda^{n+1}} (1 - \lambda) y^2(n+1) \quad (5.24)$$

¹Esta normalización aparece en el cuerpo de ORIV y sólo es necesario introducirla para calcular la varianza en (5.37). En el resto de valores esperados no hace falta introducirla porque se trabaja con igualdades entre ellos y por tanto la normalización no influye.

Como el factor de olvido está comprendido entre 0 y 1, si se considera $n \gg 1$, la ecuación de actualización anterior se reduce a:

$$\hat{\sigma}^2(n+1) = \lambda \hat{\sigma}^2(n) + (1-\lambda)y^2(n+1) \quad (5.25)$$

En [FP89] se propone la utilización de (5.25), mientras que en esta Memoria se propone la actualización dada en (5.24) ya que es más exacta. Posteriormente se verá la diferencia de comportamiento entre ambos casos.

5.3.2 Aplicando ORIV a ecualización ciega

Una vez determinado cómo se estimarán los cumulantes se puede proceder a detallar cómo se pueden resolver las ecuaciones (5.17) y (5.18) mediante el algoritmo ORIV. Comencemos por la primera de ellas.

Para centrar ideas se puede afirmar que en la ecuación (5.17) aparecen elementos que se pueden englobar dentro de tres clases distintas:

1. Estadística de la salida del sistema, es decir, cumulantes diagonales de orden 2 y 3 de la serie $\{y(n)\}$.
2. Estadística de la excitación del modelo, en particular los cumulantes para intervalos temporales nulos de orden 2 y 3 de la serie $\{w(n)\}$.
3. Los parámetros que definen al sistema, que aparecen en diversas formas funcionales no lineales.

De estos tres grupos sólo se puede tener acceso directo al primero de ellos, ya que sólo se conoce la salida (contaminada) del sistema y con ella se pueden estimar los cumulantes correspondientes. De forma que los otros dos grupos constituyen las incógnitas del problema.

En principio el conjunto de las incógnitas es $\{\sigma_w^2, \gamma_{3w}, b_1, \dots, b_q\}$, recuérdese que $b_0 \equiv 1$. Pero si se quiere resolver mediante ORIV se está en la obligación de expresar la ecuación en forma de una ecuación matricial y por tanto debe ser lineal en las incógnitas. Una forma de conseguir esto es reescribir la ecuación (5.17) tal y como sigue:

$$\Phi(n)\mathbf{w}(n) = \mathbf{z}(n) \quad (5.26)$$

donde se define:

$$\Phi(n) = (\Phi_1(n) \quad | \quad \Phi_2(n)) \quad (5.27)$$

$$\begin{aligned}
\Phi_1(n) &= \begin{pmatrix} \hat{c}_3(-q; n) & 0 & \dots & 0 \\ \hat{c}_3(-q+1; n) & \hat{c}_3(-q; n) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ \hat{c}_3(q; n) & \hat{c}_3(q-1; n) & \dots & \hat{c}_3(0; n) \\ 0 & \ddots & & \\ \vdots & & \ddots & \\ 0 & \dots & \dots & \hat{c}_3(q; n) \end{pmatrix} & 3q+1 \times q+1 \\
\Phi_2(n) &= \begin{pmatrix} 0 & \dots & \dots & 0 \\ \hat{c}_2(-q; n) & \dots & \dots & 0 \\ \hat{c}_2(-q+1; n) & \hat{c}_2(-q; n) & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ \hat{c}_2(q; n) & \hat{c}_2(q-1; n) & \dots & \hat{c}_2(1; n) \\ \vdots & & & \\ 0 & \dots & \dots & \hat{c}_2(q-1; n) \end{pmatrix} & 3q+1 \times q
\end{aligned} \tag{5.28}$$

$$\mathbf{w}(n) = (\epsilon(n), \epsilon b_1(n), \dots, \epsilon b_q(n), -b_1^2(n), \dots, -b_q^2(n))^t \quad \epsilon = \frac{\sigma_w^2}{\gamma_{3w}} \tag{5.29}$$

$$\mathbf{z}(n) = (\hat{c}_2(-q; n), \hat{c}_2(-q+1; n), \dots, \hat{c}_2(q; n), 0, \dots, 0)^t \tag{5.30}$$

Donde la notación utilizada ha sido que $\hat{c}_k(m; n)$ es el estimador del cumulante diagonal $c_{ky}(m, m, \dots, {}^{(k-1)}, m)$ empleando n datos, o equivalentemente, a tiempo n . Como las magnitudes conocidas dependen de n también lo harán las incógnitas y así se ha indicado.

Hay que resaltar que las incógnitas son funciones no lineales de los parámetros del sistema, pero este es el precio que hay que pagar por expresarlas en forma matricial. Sería posible resolver las ecuaciones teniendo en cuenta estas dependencias, pero entonces la matemática se complicaría enormemente.

Para resolver el problema, por tanto, lo primero que se podría intentar es, una vez obtenidas n muestras de la serie de salida (contaminada) $y(n)$ calcular los cumulantes de segundo y tercer orden necesarios, con ellos construir la matriz $\Phi(n)$ y el vector $\mathbf{z}(n)$ y resolver la ecuación matricial correspondiente calculando la pseudoinversa. Sin embargo este método no hace uso de la estructura particular de la matriz ni del vector.

Una forma sencilla de estimar los cumulantes necesarios es definiendo las

siguientes cantidades:

$$\begin{aligned}\tilde{\mathbf{x}}(n) &= [y(n), \dots, y(n-3q)]^t \\ \mathbf{x}(n) &= [y^2(n-q), \dots, y^2(n-2q) | y(n-q-1), \dots, y(n-2q)]^t \\ d(n) &= y(n-q)\end{aligned}\quad (5.31)$$

y notando que, con ellas, la matriz $\Phi(n)$ definida en (4.2) coincide con la matriz $\Phi(n)$ aquí definida y que, análogamente, el vector $\mathbf{z}(n)$ definido en (4.3) coincide con el vector $\mathbf{z}(n)$ definido aquí. Esto implica que ORIV, dado en la tabla 4.1 y utilizando las definiciones de (5.31), es capaz de resolver la ecuación (5.26). Debido a las propiedades ya conocidas de ORIV la resolución implicará un ahorro computacional con respecto a una inversión matricial convencional, además de que con la llegada de una nueva muestra de la salida $y(n+1)$, no es necesario volver a estimar de nuevo la matriz y repetir el proceso sino que el mismo ORIV se encarga de actualizar los estimadores $\mathbf{w}(n)$. Una vez obtenidos los estimadores $\mathbf{w}(n)$ se pueden obtener los estimadores de los pesos del filtro utilizando la expresión (5.29).

También hay que resaltar que aunque la matriz $\Phi(n)$ tenga en teoría elementos que son nulos, cuando se estiman esos elementos no serán exactamente igual a cero puesto que tomarán el valor de un estimador de un cumulante que debe ser nulo. Este efecto tendrá influencias cada vez menores conforme más datos se conocen y aumenta la calidad de los estimadores.

Una discusión parecida se puede realizar para la resolución de la ecuación (5.18) que relaciona cumulantes de tercer y cuarto orden con los parámetros del sistema. Se puede escribir en forma de una ecuación matricial:

$$\Phi(n)\mathbf{w}(n) = \mathbf{z}(n) \quad (5.32)$$

donde se define:

$$\Phi(n) = (\Phi_1(n) \quad | \quad \Phi_2(n)) \quad (5.33)$$

$$\begin{aligned}
\Phi_1(n) &= \begin{pmatrix} \hat{c}_4(-q; n) & 0 & \dots & 0 \\ \hat{c}_4(-q+1; n) & \hat{c}_4(-q; n) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ \hat{c}_4(q; n) & \hat{c}_4(q-1; n) & \dots & \hat{c}_4(0; n) \\ 0 & \ddots & & \\ \vdots & & \ddots & \\ 0 & \dots & \dots & \hat{c}_4(q; n) \end{pmatrix} & 3q+1 \times q+1 \\
\Phi_2(n) &= \begin{pmatrix} 0 & \dots & \dots & 0 \\ \hat{c}_3(-q; n) & \dots & \dots & 0 \\ \hat{c}_3(-q+1; n) & \hat{c}_3(-q; n) & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ \hat{c}_3(q; n) & \hat{c}_3(q-1; n) & \dots & \hat{c}_3(1; n) \\ \vdots & & & \\ 0 & \dots & \dots & \hat{c}_3(q-1; n) \end{pmatrix} & 3q+1 \times q
\end{aligned} \tag{5.34}$$

$$\mathbf{w}(n) = (\epsilon(n), \epsilon b_1^2(n), \dots, \epsilon b_q^2(n), -b_1^3(n), \dots, -b_q^3(n))^t \quad \epsilon = \frac{\gamma_{3w}}{\gamma_{4w}} \tag{5.35}$$

$$\mathbf{z}(n) = (\hat{c}_3(-q; n), \hat{c}_3(-q+1; n), \dots, \hat{c}_3(q; n), 0, \dots, 0)^t \tag{5.36}$$

Y puede ser resuelta mediante ORIV si los vectores de entrada, instrumental y respuesta deseada se definen como:

$$\begin{aligned}
\tilde{\mathbf{x}}(n) &= [y(n), \dots, y(n-3q)]^t \\
\mathbf{x}(n) &= [y^3(n-q), \dots, y^3(n-2q)|y^2(n-q-1), \dots, y^2(n-2q)]^t - \\
&\quad - 3\hat{\sigma}^2(n)[y(n-q), \dots, y(n-2q)|0, \dots, 0]^t \\
d(n) &= y^2(n-q)
\end{aligned} \tag{5.37}$$

donde $\hat{\sigma}^2(n)$ es un estimador de la varianza de la serie de salida $y(n)$.

En definitiva, las ecuaciones (5.17) y (5.18) junto con el algoritmo ORIV constituyen dos métodos de identificación ciega y adaptativa de sistemas MA, de aquí en adelante se designarán por ORIV-C2C3 y ORIV-C3C4 respectivamente; propuestos, aunque no analizados, en [FP89] y [KJYS97]. De dichos métodos se conoce el comportamiento teórico de ORIV y el comportamiento teórico de los estimadores de los cumulantes que aparecen en las ecuaciones (5.17) y (5.18), lo cual da el comportamiento teórico global del método. Este comportamiento esperado se constatará mediante simulaciones [ARBC02b] en los próximos apartados.

5.4 Introducción al análisis experimental realizado

Una vez aclarados todos los aspectos teóricos del problema se hace necesario comprobarlos mediante simulaciones [RABC00]. Tanto el sistema a identificar como la salida del mismo y el resto de elementos se han generado mediante un ordenador personal, siguiendo las directrices especificadas en los siguientes apartados. Esta forma de actuar permite controlar perfectamente los parámetros del sistema de forma que se facilita en gran medida su análisis posterior.

5.4.1 Aspectos generales de los modelos a identificar

Los sistemas bajo estudio se modelan como sistemas MA(q) excitados mediante una serie $\{w(n)\}$ según la expresión que se reproduce a continuación por comodidad:

$$x(n) = \sum_{i=0}^q b_i w(n-i) \quad (5.38)$$

donde los parámetros $\{b_0, \dots, b_q\}$ son específicos de cada caso particular a estudiar y se determinarán en su momento. Asimismo, y como es conocido, en la situación más general posible la salida estará contaminada con ruido $v(n)$, de forma que los datos accesibles son

$$y(n) = x(n) + v(n) \quad (5.39)$$

Las características comunes a todos los casos a estudiar son:

1. El proceso $\{w(n)\}$ es i.i.d con FDP exponencial de media nula, varianza unidad, factor de asimetría 2 y curtosis 6.
2. El ruido aditivo $\{v(n)\}$ se obtiene filtrando ruido gaussiano blanco por un filtro ARMA definido por AR=[1 -2.2 1.77 -0.52] y MA=[1 -1.25].
3. Los coeficientes $\{b_0, \dots, b_q\}$ son todos reales con $b_0 = 1$.

5.4.2 Algunos conceptos útiles

Para fijar ideas y la notación, en las siguientes líneas se definen algunos conceptos que serán necesarios tanto para especificar completamente el sistema y el ambiente, como para estudiar el comportamiento de los métodos.

Relación señal ruido

La relación señal ruido o SNR (signal to noise ratio), se define como el cociente entre la potencia de la salida $x(n)$ del sistema lineal y la potencia del ruido $v(n)$ añadido, en decibelios:

$$\text{SNR} = 10 \log \frac{\sigma_x^2}{\sigma_v^2} \quad (5.40)$$

A menor SNR mayor es la importancia del ruido aditivo.

Error cuadrático medio

El error cuadrático medio MSE (mean square error) permite medir la bondad de los estimadores obtenidos mediante los métodos bajo estudio. En esta memoria se definirá como la desviación en media cuadrática de los vectores incógnita respecto a sus valores verdaderos, también incluyen un término por motivos de normalización:

$$\begin{aligned} \text{MSE}(n) &= \frac{\text{E}[(\mathbf{w}(n) - \mathbf{w}_0)^t (\mathbf{w}(n) - \mathbf{w}_0)]}{\mathbf{w}_0^t \mathbf{w}_0} \sim \\ &\sim \frac{1}{N_r} \sum_{r=1}^{N_r} \frac{(\mathbf{w}^{(r)}(n) - \mathbf{w}_0)^t (\mathbf{w}^{(r)}(n) - \mathbf{w}_0)}{\mathbf{w}_0^t \mathbf{w}_0} \end{aligned} \quad (5.41)$$

En este caso el valor esperado se ha sustituido por un promedio en las N_r realizaciones disponibles. Téngase en cuenta que el superíndice r indica la realización correspondiente.

Para los métodos ORIV-C3C4 y ORIV-C2C3 los vectores de incógnitas verdaderos son, respectivamente, los siguientes:

$$\mathbf{w}_0 = [\epsilon, \epsilon b_1^2, \dots, \epsilon b_q^2, -b_1^3, \dots, -b_q^3]^t \quad \epsilon = \frac{\gamma_{3w}}{\gamma_{4w}} \quad (5.42)$$

$$\mathbf{w}_0 = [\epsilon, \epsilon b_1, \dots, \epsilon b_q, -b_1^2, \dots, -b_q^2]^t \quad \epsilon = \frac{\sigma_w^2}{\gamma_{3w}} \quad (5.43)$$

Como se puede apreciar en las definiciones respectivas, el MSE coincide con $D(n)$ y $q(n)$ salvo normalización o reescalado de componentes. En lo que sigue no se distinguirán unos de otros.

Curva de aprendizaje

Como se aprecia en la definición (5.41) el error cuadrático medio depende del tiempo, ya que se dispone de un estimador para cada instante de tiempo. A

la representación del MSE en función de n se le conoce como curva de aprendizaje y es algo característico de los algoritmos adaptativos. Como se sabe de los capítulos respectivos al análisis teórico presenta una primera parte de convergencia hacia un valor estacionario (steady state) y luego una parte horizontal relativa a dicho valor de convergencia. El tiempo que tarda en converger al estado estacionario se conoce como tiempo de convergencia, aunque no debe confundirse con la capacidad que tiene un algoritmo adaptativo para seguir la evolución dinámica del sistema a identificar. La convergencia es una propiedad transitoria y el rastreo es estacionaria una vez acabada la primera. Un algoritmo puede tener un tiempo de convergencia muy pequeño pero rastrear muy mal los sistemas.

Tipos de estimadores de los filtros

El método ORIV-C2C3 da estimadores del vector (5.43) con lo que a partir de él se pueden obtener dos estimadores distintos de los pesos del filtro, dados respectivamente por:

$$\hat{b}_i^{(1)}(n) = \frac{w_{i+1}(n)}{w_1(n)} \quad i = 1 \dots q \quad (5.44)$$

$$\hat{b}_i^{(2)}(n) = \text{signo}\{\hat{b}_i^{(1)}(n)\} \sqrt{-w_{i+q+1}(n)} \quad i = 1 \dots q \quad (5.45)$$

De forma análoga se pueden obtener dos estimadores de los pesos del filtro mediante el método ORIV-C3C4 y recurriendo a (5.42):

$$\hat{b}_i^{(1)}(n) = \text{signo}\{\hat{b}_i^{(2)}(n)\} \sqrt{\frac{w_{i+1}(n)}{w_1(n)}} \quad i = 1 \dots q \quad (5.46)$$

$$\hat{b}_i^{(2)}(n) = \sqrt[3]{-w_{i+q+1}(n)} \quad i = 1 \dots q \quad (5.47)$$

5.5 Los métodos ORIV-C2C3 y ORIV-C3C4 en ambientes estacionarios ruidosos

En este apartado se van a llevar a cabo una serie de simulaciones con el fin de estudiar el comportamiento de los métodos ORIV-C2C3 y ORIV-C3C4 en ambientes estacionarios ruidosos. En estas simulaciones se supondrá que los pesos del filtro no cambian con el tiempo y que se cuenta con 5000 muestras (ruidosas) de la salida del mismo. Se han generado 1000 series distintas (realizaciones) y con ellas se ha llevado a cabo la identificación. Los resultados que se muestran

son promediados de ella, y también se ha calculado la desviación estándar de los mismos. Además se han estudiado diversas relaciones señal ruido: 20, 10, 5, 0, -5 y -10 dB.

En lo que respecta a los parámetros que quedan libres del algoritmo en sí se han elegido de la siguiente manera. Se ha optado por una inicialización exacta, de forma que las matrices y vectores iniciales se calculan con un tratamiento por lotes conocidos los $3q + 1$ primeros datos y el vector incógnita inicial se calcula mediante mínimos cuadrados a partir de ellos. Por otro lado, ya que se trabaja en un ambiente estacionario se ha fijado el factor de olvido como la unidad con el fin de minimizar los efectos del ruido de medición y estimación en la mayor medida posible.

5.5.1 Resultados de las simulaciones

Para estas condiciones de trabajo se han estudiado 5 modelos diferentes, tomados de [Nan94], que se detallan a continuación:

Modelo ER1

Como primer modelo se considera un modelo MA de fase mínima y parámetros MA = [1 - 0.8]. La media de los estimadores a tiempo $n = 5000$ se muestran en las tablas 5.1 y 5.2 para los métodos ORIV-C2C3 y ORIV-C3C4 junto con su desviación estándar. Aparecen tanto los estimadores tipo 1 como los tipo 2.

Asimismo, y dada la importancia de una curva de aprendizaje para un algoritmo adaptativo, en las gráficas 5.1 y 5.2 se representan para ORIV-C2C3 y ORIV-C3C4 respectivamente.

La equivalencia entre símbolo y SNR en las gráficas anteriores es la siguiente:

- ◇ SNR = 20 dB
- SNR = 10 dB
- X SNR = 5 dB
- + SNR = 0 dB
- * SNR = -5 dB
- SNR = -10 dB

SNR (dB)	Tipo 1		Tipo 2	
	b_1	var b_1	b_1	var b_1
20	-0.787	0.031	-0.799	0.023
10	-0.732	0.042	-0.784	0.028
5	-0.933	0.128	-0.580	0.043

0	0.437	139.606	0.000	0.663
-5	4.960	40.505	0.000	0.229
-10	2.747	27.990	0.202	0.967

Tabla 5.1 Media y varianza de los estimadores. ORIV-C2C3, modelo ER1

SNR (dB)	Tipo 1		Tipo 2	
	b_1	var b_1	b_1	var b_1
20	-0.798	0.023	-0.796	0.035
10	-0.797	0.030	-0.797	0.042
5	-0.796	0.041	-0.793	0.063
0	-0.789	0.113	-0.779	0.129
-5	-0.297	1.288	-0.504	0.571
-10	0.129	1.599	0.067	0.898

Tabla 5.2 Media y varianza de los estimadores. ORIV-C3C4, modelo ER1

Modelo ER2

Se analiza ahora el modelo MA = [1 - 1.25] que es de fase no mínima. La media y varianza de los 1000 estimadores obtenidos se presentan en las tablas 5.3 y 5.4 para los métodos ORIV-C2C3 y ORIV-C3C4 respectivamente. Además en las gráficas 5.3 y 5.4 se representan las curvas de aprendizaje respectivas de los métodos mencionados.

SNR (dB)	Tipo 1		Tipo 2	
	b_1	var b_1	b_1	var b_1
20	-1.271	0.047	-1.250	0.036
10	-1.621	0.115	-1.181	0.036
5	-11.014	325.968	0.436	0.618
0	0.390	0.191	0.000	0.068
-5	0.419	0.220	0.000	0.190
-10	1.519	14.154	0.157	0.888

Tabla 5.3 Media y varianza de los estimadores. ORIV-C2C3, modelo ER2

SNR (dB)	Tipo 1		Tipo 2	
	b_1	var b_1	b_1	var b_1
20	-1.250	0.035	-1.250	0.053
10	-1.250	0.048	-1.250	0.067
5	-1.270	0.074	-1.230	0.100
0	-1.450	1.230	-1.100	0.237
-5	-0.763	2.440	-0.480	0.648
-10	-0.017	2.160	0.034	0.815

Tabla 5.4 Media y varianza de los estimadores. *ORIV-C3C4*, modelo *ER2*

Modelo ER3

El modelo tercero se corresponde con un modelo MA de orden 2 dado por los parámetros MA = [1 -1.1314 0.6400] que es de fase mínima. Al igual que para modelos anteriores se presenta la media y varianza de los estimadores, en la tabla 5.5 para *ORIV-C2C3* y en la tabla 5.6 para el método *ORIV-C3C4*. Asimismo se representa la curva de aprendizaje en las gráficas 5.5 y 5.6 para estos mismos métodos.

SNR	Estimadores tipo 1				Estimadores tipo 2			
	b_1	b_2	var b_1	var b_2	b_1	b_2	var b_1	var b_2
20	-1.048	0.671	0.051	0.060	-1.126	0.626	0.036	0.057
10	-0.706	0.938	0.101	0.092	-1.021	0.381	0.051	0.182
5	-1.015	1.603	0.783	0.678	-0.701	0.012	0.089	0.224
0	-1.386	2.339	32.081	34.660	-0.089	0.008	0.382	0.773
-5	0.400	0.752	6.340	10.233	0.069	0.251	0.834	0.517
-10	-0.051	0.172	14.166	17.574	0.049	0.290	1.015	0.755

Tabla 5.5 Media y varianza de los estimadores. *ORIV-C2C3*, modelo *ER3*

SNR	Estimadores tipo 1				Estimadores tipo 2			
	b_1	b_2	var b_1	var b_2	b_1	b_2	var b_1	var b_2
20	-1.129	0.541	0.045	0.336	-1.135	0.563	0.058	0.286
10	-1.128	0.429	0.065	0.464	-1.136	0.510	0.079	0.418
5	-1.086	0.277	0.211	0.556	-1.116	0.519	0.149	0.539
0	-0.707	0.212	1.052	0.987	-0.813	0.730	0.494	0.559

-5	-0.249	0.368	2.436	2.171	-0.353	0.573	0.739	0.645
-10	-0.233	0.289	1.879	2.013	-0.525	0.258	0.708	0.771

Tabla 5.6 Media y varianza de los estimadores. *ORIV-C3C4, modelo ER3*

Modelo ER4

El modelo ER4 también se ocupa de un modelo MA de orden 2 con coeficientes $MA = [1 \quad -2.333 \quad 0.667]$ pero esta vez de fase no mínima. Para el método *ORIV-C2C3* se presentan los estimadores en la tabla 5.7 y para el método *ORIV-C3C4* en la tabla 5.8. Las curvas de aprendizaje asociadas se representan en las gráficas 5.7 y 5.8 para *ORIV-C2C3* y *ORIV-C3C4*.

SNR	Estimadores tipo 1				Estimadores tipo 2			
	b_1	b_2	var b_1	var b_2	b_1	b_2	var b_1	var b_2
20	-2.715	0.596	0.291	0.189	-2.218	0.719	0.185	0.
10	1.381	0.567	0.628	0.263	1.089	0.671	0.127	0.186
5	-0.013	0.208	0.222	0.263	0.039	0.222	0.461	0.448
0	-0.481	0.533	0.185	0.117	0.000	0.686	0.156	0.184
-5	-0.948	0.603	0.455	0.133	0.000	0.909	0.297	0.089
-10	-8.641	1.187	204.839	15.725	0.003	0.708	0.625	0.627

Tabla 5.7 Media y varianza de estimadores. *ORIV-C2C3, modelo ER4*

SNR	Estimadores tipo 1				Estimadores tipo 2			
	b_1	b_2	var b_1	var b_2	b_1	b_2	var b_1	var b_2
20	-1.926	-0.037	4.671	2.350	-0.952	0.125	1.335	0.938
10	-1.513	0.285	6.748	2.604	-0.805	0.268	1.296	0.934
5	-0.648	0.543	3.282	1.898	-0.661	0.397	1.112	0.873
0	-0.211	0.754	2.600	1.944	-0.600	0.402	0.871	0.853
-5	-0.090	0.700	2.110	1.253	-0.733	0.500	0.559	0.684
-10	-0.053	0.501	1.731	1.706	-0.518	0.398	0.733	0.688

Tabla 5.8 Media y varianza de estimadores. *ORIV-C3C4, modelo ER4*

Modelo ER5

El último modelo es un modelo MA de orden 3 de fase no mínima y con coeficientes MA = [1 -0.9 0.385 -0.771]. Las tablas 5.9 y 5.10 contienen los estimadores, tipo 1 y 2 respectivamente, del método ORIV-C2C3 y las tablas 5.11 y 5.12 las del método ORIV-C3C4. Por otro lado las curvas de aprendizaje se muestran en las gráficas 5.9 y 5.10.

SNR (dB)	b_1	b_2	b_3	var b_1	var b_2	var b_3
20	-0.912	0.301	-0.754	0.113	0.123	0.066
10	-1.368	-0.109	-0.669	0.172	0.175	0.083
5	-2.958	-0.957	-1.069	1.055	0.529	0.341
0	-2.238	-0.180	-1.008	144.330	38.776	42.881
-5	0.689	-0.672	-1.421	58.744	40.570	40.029
-10	-0.541	-0.126	-0.211	20.797	11.560	25.179

Tabla 5.9 Media y varianza de estimadores tipo 1. ORIV-C2C3, modelo ER5

SNR (dB)	b_1	b_2	b_3	var b_1	var b_2	var b_3
20	-0.900	0.271	-0.724	0.067	0.181	0.055
10	-0.814	0.000	-0.332	0.080	0.575	0.172
5	-0.731	0.000	-0.077	0.122	0.249	0.244
0	-0.117	0.000	-0.041	0.502	0.936	0.539
-5	0.000	0.003	-0.167	0.649	0.917	0.883
-10	-0.013	0.039	-0.132	0.913	0.851	0.901

Tabla 5.10 Media y varianza de estimadores tipo 2. ORIV-C2C3, modelo ER5

SNR (dB)	b_1	b_2	b_3	var b_1	var b_2	var b_3
20	-0.900	0.040	-0.761	0.078	0.414	0.112
10	-0.884	0.053	-0.750	0.106	0.419	0.187
5	-0.813	0.038	-0.629	0.224	0.471	0.457
0	-0.516	0.027	-0.184	0.533	0.726	0.812
-5	-0.192	0.044	-0.006	1.585	1.798	1.962
-10	-0.243	-0.067	0.253	1.448	1.724	1.574

Tabla 5.11 *Media y varianza de estimadores tipo 1. ORIV-C3C4, modelo ER5*

SNR (dB)	b_1	b_2	b_3	var b_1	var b_2	var b_3
20	-0.891	0.134	-0.750	0.075	0.442	0.105
10	-0.903	0.148	-0.730	0.091	0.494	0.155
5	-0.930	0.183	-0.624	0.135	0.597	0.346
0	-0.943	0.264	-0.256	0.305	0.717	0.631
-5	-0.516	0.165	-0.012	0.677	0.739	0.691
-10	-0.531	-0.146	0.227	0.665	0.736	0.710

Tabla 5.12 *Media y varianza de estimadores tipo 2. ORIV-C3C4, modelo ER5*

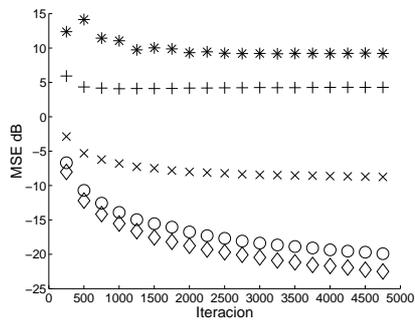


Figura 5.1 Modelo ER1.
ORIV-C2C3

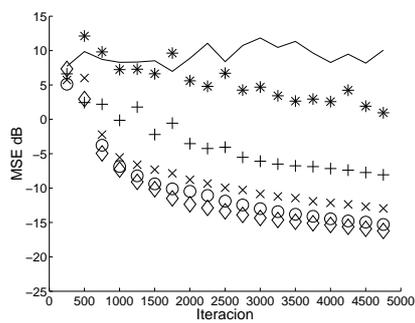


Figura 5.2 Modelo ER1.
ORIV-C3C4

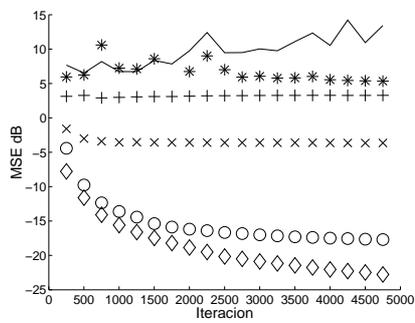


Figura 5.3 Modelo ER2.
ORIV-C2C3

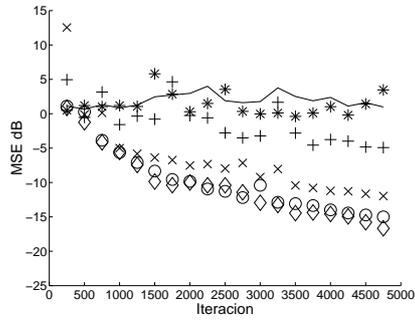


Figura 5.4 Modelo ER2.
ORIV-C3C4

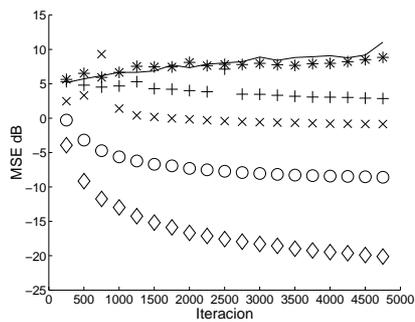


Figura 5.5 Modelo ER3.
ORIV-C2C3

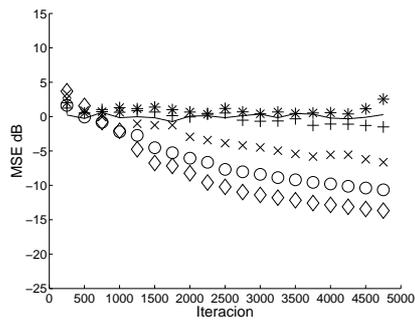


Figura 5.6 Modelo ER3.
ORIV-C3C4

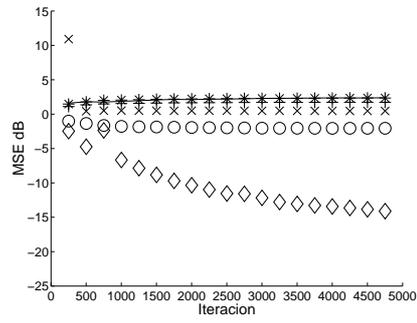


Figura 5.7 Modelo ER4.
ORIV-C2C3

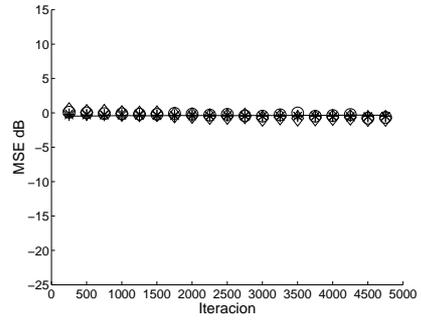


Figura 5.8 Modelo ER4.
ORIV-C3C4

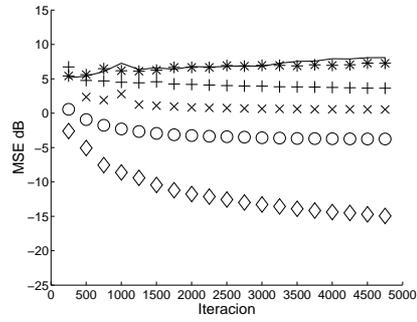


Figura 5.9 Modelo ER5.
ORIV-C2C3

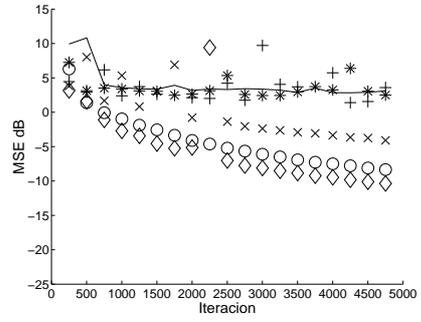


Figura 5.10 Modelo ER5.
ORIV-C3C4

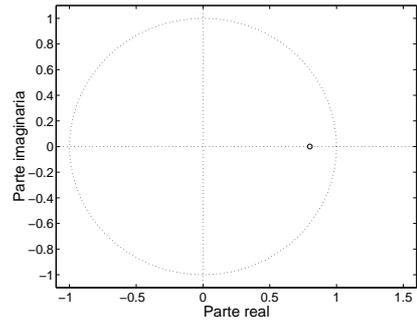


Figura 5.11 Diagrama cero-polo.
Modelo ER1

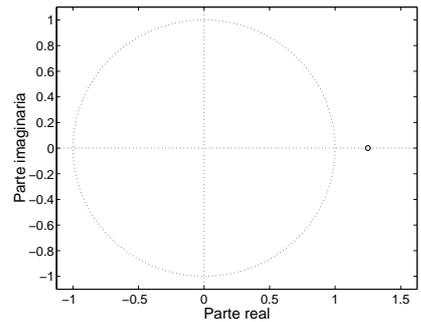


Figura 5.12 Diagrama cero-polo.
Modelo ER2

5.5.2 Análisis de las simulaciones

De los resultados numéricos y gráficos presentados en los apartados anteriores se pueden extraer consecuencias referentes a distintos puntos y aspectos del problema. Estos puntos se comentan a continuación.

Información sobre la fase

En las simulaciones precedentes se verifica un hecho que se esperaba desde el punto de vista teórico: tanto ORIV-C2C3 como ORIV-C3C4 identifican sin problemas sistemas de fase no mínima, véase por ejemplo el modelo ER2, ya que contienen estadística de orden 3 o superior. Todos los algoritmos basados en estadística de segundo orden hubieran confundido el modelo [1 - 1.25] con su equivalente de fase mínima [1 - 0.8], ya que dichos algoritmos no llevan información de la fase. Por equivalente se entiende aquel que tiene el mismo espectro de potencia pero distinta respuesta en fase. Se obtiene por inversión y complejo conjugando los ceros de la función de transferencia. Los ceros de la función de transferencia del modelo [1 - 0.8] se muestran en la figura 5.11 y para [1 - 1.25] en 5.12 mediante un pequeño círculo.

Análisis del estado estacionario

En general, los estimadores tipo 1 presentan mayor varianza que los tipo 2, ya que para obtenerlos hay que dividir por el estimador de ϵ que viene acompañado de su correspondiente error.

Otra causa del aumento de la varianza de los estimadores de los pesos es la disminución de la SNR. Teniendo en cuenta la descomposición utilizada en (4.25) y (4.26) para la matriz $\Phi(n)$ y el vector $z(n)$:

$$\begin{aligned}\Phi(n) &\sim \frac{\mathbf{R}_{\bar{x}x}}{1-\lambda} + \mathbf{R}_{\bar{x}x}^e(n) \\ z(n) &\sim \frac{\mathbf{r}_{\bar{x}d}}{1-\lambda} + \mathbf{r}_{\bar{x}d}^e(n)\end{aligned}$$

y que al aumentar el ruido aditivo los estimadores de esta matriz y este vector aumentan su varianza, resulta que las componentes $\mathbf{R}_{\bar{x}x}^e(n)$ y $\mathbf{r}_{\bar{x}d}^e(n)$ tienen más importancia. Traducándolo a la ecuación (4.60), implica que al disminuir la SNR, aumenta el $\text{MSE}(\infty)$. Recuérdese que aquí se identifica el $\text{MSE}(n)$ con el $q(n)$ y el $D(n)$ del capítulo 4 por mantener la notación empleada en la literatura [Hay96], [Swa96].

Esto implica, para ORIV-C3C4, que la inmunidad teórica que se preveía ante el ruido gaussiano no se da en la práctica. Esto es debido a que no se

dispone de los cumulantes verdaderos sino de estimadores de los mismos. Estos estimadores de los cumulantes, al menos para un número de datos finito, sí son sensibles al ruido gaussiano.

Es de resaltar que la influencia de la SNR en ORIV-C3C4 es mucho menor que para ORIV-C2C3, tanto en la varianza como en el sesgo. Inspecciónese por ejemplo las tablas 5.2 y 5.1. De la primera se desprende que ORIV-C3C4 presenta un resultado aceptable, en varianza y en sesgo, hasta los 0 dB. Sin embargo, de la segunda, se concluye que ORIV-C2C3 empieza a fallar ya casi a los 10 dB y falla claramente a los 5 dB. Este comportamiento era de esperar, ya que como se vio en el apartado 5.2.3 la ecuación (5.17) no es cierta si $\{x(n)\}$ está contaminada por ruido como es el caso presente. Por tanto, aunque ORIV-C3C4 no sea estrictamente inmune al ruido gaussiano, sí mejora ampliamente los resultados del algoritmo que incluye estadística de segundo orden. Estos comentarios extraídos del análisis de resultados para el caso 1 se pueden generalizar para todos los casos estudiados.

Cuando el nivel de ruido añadido es bajo, el valor del MSE al que converge ORIV-C2C3 es menor que para ORIV-C3C4. Mientras que el ruido no distorciona mucho las medidas realizadas, los estimadores de los cumulantes de tercer orden presentan menor varianza ($\mathbf{R}_{\hat{x}x}^e(n)$ y $\mathbf{r}_{\hat{x}d}^e(n)$) que los de cuarto. Esta varianza de los estimadores de los cumulantes se traduce en varianza en las incógnitas calculadas y por tanto en MSE. Cuando el nivel de ruido aumenta, aumenta la varianza de los cumulantes de cuarto, tercer y segundo orden, lo que pasa es que a la vez aumenta el sesgo de los estimadores de los cumulantes de segundo orden ya que estos no son teóricamente inmunes a él. En consecuencia, el algoritmo que se ve más afectado con SNR bajos es ORIV-C2C3 y es ORIV-C3C4 el que da mejores resultados. ¿Para qué SNR superará ORIV-C3C4 a ORIV-C2C3 dando un menor MSE estacionario? En el momento en que los efectos del ruido sobre los cumulantes de segundo orden, hagan que estos sean menos apropiados para utilizarlos en ecuaciones del tipo (5.17) y (5.18) que los de tercero y cuarto, debido al aumento tanto de varianza como de sesgo.

Como consecuencia de todo lo discutido en el párrafo anterior es de esperar que para SNR bajo, aunque empeoren los resultados de ORIV-C3C4, lo hagan en cuanto a varianza se refiere, siendo la media de los estimadores relativamente cercana al parámetro verdadero. Esto puede verse en la tabla 5.2. En ella se observa cómo la media de los estimadores casi no cambia al disminuir la SNR, y sin embargo la varianza sí aumenta de forma significativa. Si se estudia ahora la tabla 5.1 se puede observar que la degradación de los resultados es paralela en sesgo y varianza: media más desviada del valor real del parámetro y mayor varianza. Para niveles de ruido altísimos todo lo discutido anteriormente deja ya de tener validez ya que los resultados están excesivamente afectados por el nivel de ruido.

Otro hecho destacable y que es general para todos los algoritmos de iden-

tificación de sistemas es que los resultados empeoran conforme se aumenta el número de incógnitas a determinar. Una desventaja que presenta ORIV es que trata como independientes todas las componentes de $\mathbf{w}(n)$ (es decir, considera que no hay ninguna relación funcional entre las incógnitas del sistema). Esto no es cierto y para comprobarlo basta ver las definiciones de $\mathbf{w}(n)$ en cada caso (ecuaciones (5.42) y (5.43)). Para optimizar el resultado sería conveniente resolver el problema con alguna ligadura que tuviera en cuenta esta circunstancia; lo que ocurriría en este caso es que el sistema ya no sería lineal y su resolución es mucho más complicada.

Análisis de la etapa no estacionaria

A continuación se comentan las conclusiones que se pueden obtener a partir del estudio de la evolución temporal del MSE. Como se puede suponer, todas las diferencias que se indiquen entre ORIV-C2C3 y ORIV-C3C4 están íntimamente relacionadas con el orden de la estadística que usan. Así por ejemplo, en las figuras 5.1 y 5.2 se observa que ORIV-C2C3 converge antes, básicamente en 3000 iteraciones para SNR altos y en 1000 iteraciones para SNR bajos, y que ORIV-C3C4 prácticamente está empezando a converger en 5000 iteraciones. El motivo de este comportamiento está en que se necesitan muchos menos datos para estimar correctamente los cumulantes de segundo y tercer orden que los de tercero y cuarto. Cuanto mayor sea el orden de la estadística a estimar, mayor debe ser el número de datos $\{y(n)\}$ conocidos. También parece haber cierta relación entre el número de incógnitas y el tiempo de convergencia: si se aumenta el número de parámetros a estimar, el algoritmo va a necesitar más tiempo para converger.

Conclusiones generales

Para completar lo comentado anteriormente cabe resaltar también que las diferencias en el comportamiento de un mismo algoritmo al cambiar el SNR son menores para ORIV-C3C4 que para ORIV-C2C3. El método ORIV-C2C3 se muestra excesivamente sensible a la SNR dando resultados muy buenos para valores altos y muy malos para valores bajos; ORIV-C3C4 se mantiene en un término medio, mejorando a ORIV-C2C3 a SNR bajos.

En la práctica, y a la vista de las simulaciones, se puede afirmar que ORIV-C3C4 da un MSE estacionario mejor que ORIV-C2C3 para SNR de 5 dB o menores. Se ha comprobado que ORIV-C3C4 es fiable hasta en ambientes de 0 dB.

Problemas mal condicionados

El modelo ER4 merece un análisis más detallado. Compárese la figura 5.6 con la figura 5.8 y la figura 5.5 con la figura 5.7. En principio deberían dar resultados parecidos, ya que tienen todos los posibles parámetros, discutidos hasta ahora y que influyen en su comportamiento, similares. Tienen el mismo número de incógnitas y trabajan en ambientes con el mismo nivel de ruido. Sin embargo, los resultados del modelo ER4 son mucho peores. ORIV-C3C4 (en la figura 5.8) no converge en absoluto y ORIV-C2C3 (en la figura 5.7) funciona bastante peor que en una situación parecida (como en la figura 5.5).

El motivo de esta situación anómala es algo ya comentado brevemente con anterioridad: el mal condicionamiento de la matriz de cumulantes usada en (5.26) y en (5.32). Para este modelo concreto, $\mathbf{MA} = [1 \quad -2.333 \quad 0.667]$, tanto la matriz que hay que construir para resolver ORIV-C3C4 como ORIV-C2C3 tienen un número de condicionamiento muy bajo. La dispersión de autovalores² de estas matrices influye negativamente en su convergencia: si el cociente entre el autovalor mayor y el menor es muy grande (varios órdenes de magnitud), el tiempo de convergencia observado aumenta considerablemente.

El análisis realizado de la convergencia de $\text{MSE}(n)$ y contenido en la ecuación (4.50) no da indicios de una ralentización de la convergencia, en casos de mal condicionamiento. En principio sólo depende del factor de olvido. Lo que provoca la mala convergencia en estos casos es un problema de cálculo numérico: la inversión de una matriz mal condicionada puede llevar a inversas lejanas de la inversa verdadera. Esto puede llegar incluso a invalidar el análisis teórico llevado a cabo.

De todo lo anterior la idea fundamental que se obtiene es que al resolver sistemas mal condicionados con ORIV se espera una velocidad de convergencia bastante menor, pero además para asegurar la convergencia se deben utilizar algoritmos más robustos numéricamente hablando.

Una alternativa interesante es el algoritmo ORIV de raíz cuadrada [PF89]. Es más robusto porque asegura mejores propiedades numéricas de las matrices involucradas, no porque calcule mejor la inversa $(\mathbf{R}^t \mathbf{R})^{-1}$. Para el modelo ER4 los resultados reobtenidos mediante este algoritmo para la ecuación (5.18) (el equivalente a ORIV-C3C4) se presentan en la tabla 5.13 (convergencia en media y varianza) y la correspondiente curva de aprendizaje en 5.13.

	Estimador tipo 1				Estimador tipo 2			
modelo	b_1	b_2	var b_1	var b_2	b_1	b_2	var b_1	var b_2

²Sólo se puede hablar de autovalores de matrices cuadradas, aquí se refiere a los autovalores de las submatrices cuadradas, que se pueden obtener, convirtiendo el sistema en determinado en vez de sobredeterminado.

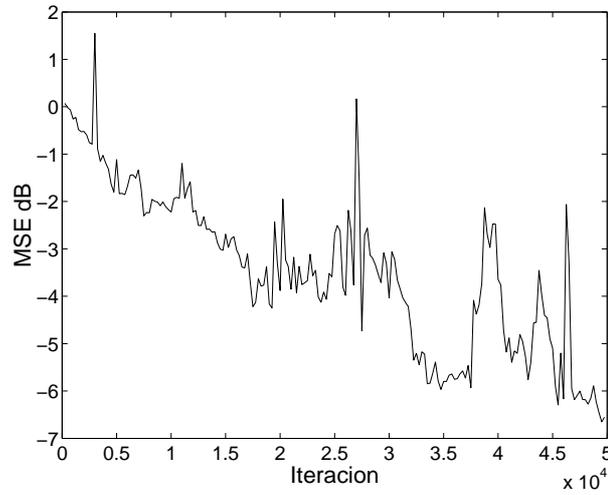


Figura 5.13 Modelo ER_4 resuelto mediante ORIV de raíz cuadrada

correcto	-2.333	0.667	0.000	0.000	-2.333	0.667	0.000	0.000
C3C4	-2.512	-0.044	0.927	0.760	-2.184	0.310	0.427	0.812

Tabla 5.13 Media y varianza de estimadores. ORIV-C3C4 de raíz cuadrada para el modelo ER_4

Esta simulación se ha llevado a cabo para 500 realizaciones y con $SNR = \infty$, es decir, sin ruido ya que lo que principalmente interesaba era comprobar si de verdad el algoritmo ORIV de raíz cuadrada solventaba el problema del mal condicionamiento.

Como se puede comprobar, en este tipo de situaciones, a medida que aumenta el tiempo aumenta la irregularidad y disminuye la suavidad en la evolución temporal del MSE. En cualquier instante de tiempo el resultado del algoritmo puede verse afectado por esta situación, es decir, por las características del problema: básicamente se está ante un problema numérico [Ste73].

También se puede destacar que para este caso se ha dejado que evolucione hasta la iteración 50000, 10 veces mayor que en las simulaciones anteriores. De este modo se ha conseguido que muestre cierta tendencia convergente, lo cual no se conseguía con el ORIV estándar. De todas maneras, como ya se mencionaba

antes, el tiempo de convergencia es muy grande, apenas llega a -7 dB de MSE en 50000 iteraciones.

5.6 Los métodos ORIV-C2C3 y ORIV-C3C4 en ambientes no estacionarios

En este apartado se va a analizar el comportamiento de los métodos ORIV-C2C3 y ORIV-C3C4 cuando los coeficientes del sistema a identificar cambian con el tiempo. Hay que resaltar el hecho de que estos métodos resuelven las ecuaciones (5.17) y (5.18) respectivamente y que dichas ecuaciones se derivaron para sistemas estacionarios. Si los sistemas no son estacionarios deberían calcularse los cumulantes involucrados para cada instante de tiempo y resolver el sistema para cada instante de tiempo. Precisamente esto último es lo que hace un algoritmo adaptativo como ORIV, resolver la ecuación para cada instante de tiempo teniendo en cuenta la última actualización de los datos, y por eso los métodos derivados de él tienen la propiedad de seguir las evoluciones temporales.

Para conseguir que esta capacidad de rastreo sea efectiva hay que realizar un pequeño ajuste en los métodos ORIV-C2C3 y ORIV-C3C4 con respecto a apartados anteriores: el factor de olvido λ debe ser estrictamente menor que la unidad. De esta manera los estimadores de los cumulantes se pueden ir adaptando a los cambios de los cumulantes verdaderos, tal y como se explicó en el apartado 5.3.1 dedicado a la ergodicidad.

En las simulaciones que se van a presentar a continuación se ha supuesto que el comportamiento dinámico del sistema puede venir descrito de dos maneras:

1. Evolución temporal lineal. La pendiente de dicha variación será clave para analizar los resultados.
2. Evolución temporal aleatoria mediante un proceso de Markov de orden 1.

A continuación se procede a la presentación de los resultados.

5.6.1 Resultados de las simulaciones con variación temporal lineal

Para estos análisis se han utilizado series de la salida contaminada del sistema $\{y(n)\}$ con 2000 muestras de longitud y para 1000 realizaciones distintas. La

ley lineal general que siguen los coeficientes del sistema es:

$$\mathbf{b}(n) = \mathbf{b}_0 + n\mathbf{p} \quad (5.48)$$

donde \mathbf{b}_0 es el vector de coeficientes del sistema en el estado inicial y \mathbf{p} es el vector cuyas componentes son las pendientes de la variación lineal de cada uno de los coeficientes. Nótese que como $b_0 \equiv 1 \quad \forall n$ la primera componente de \mathbf{p} es nula.

Esta ley empieza a aplicarse a partir de cierto instante temporal t_{cambio} y durante un intervalo cuya duración la determina l_{cambio} .

En estas circunstancias se han estudiado 3 modelos distintos.

Modelo NEL1

Se trata de un modelo de orden 2, con coeficientes iniciales $\mathbf{b}_0 = [1 \quad -1 \quad 0.75]$ y vector de pendiente $\mathbf{p} = [0 \quad -1/8000 \quad -1/8000]$. Además $t_{cambio} = 10$ y $l_{cambio} = 1989$. Finalmente como factor de olvido se ha elegido $\lambda = 0.9985$.

Para sistemas no estacionarios resulta ilustrativo representar la evolución de la media de los estimadores. Las figuras 5.14-5.17 muestran esta evolución junto con la evolución real para los estimadores tipo 1, coeficientes b_1 y b_2 , y para los estimadores tipo 2, coeficientes b_1 y b_2 , respectivamente, para el método ORIV-C2C3. Asimismo, y como es habitual, también se representa el MSE, en este caso en la figura 5.18.

Para el método ORIV-C3C4 se procede de forma análoga: la evolución de los coeficientes b_1 y b_2 obtenidos mediante los estimadores tipo 1 se presentan en las figuras 5.20 y 5.21, para los estimadores tipo 2 en las figuras 5.22 y 5.23. El MSE para este método se presenta en la figura 5.19.

Como otra medida del comportamiento, en la tabla 5.14 se muestran la media y la varianza de los estimadores de los dos tipos a tiempo $n = 2000$ (última iteración) y para los dos métodos.

	Estimadores tipo 1				Estimadores tipo 2			
	b_1	b_2	var b_1	var b_2	b_1	b_2	var b_1	var b_2
c	-1.249	0.999	0	0	-1.249	0.999	0	0
C2C3	-1.183	0.942	0.115	0.132	-1.184	0.933	0.078	0.085
C3C4	-1.178	0.929	0.091	0.093	-1.188	0.922	0.113	0.133

Tabla 5.14 Estimadores de la última iteración. Métodos ORIV-C2C3 y ORIV-C3C4, modelo NEL1. (c=correcto)

Modelo NEL2

Este modelo difiere del anterior sólomente en dos puntos: el factor de olvido vale ahora 0.9975 y el vector de las pendientes es $\mathbf{p} = [0 \quad -1/4000 \quad 1/4000]$.

Los resultados se presentan de forma similar al modelo NEL1. Para el método ORIV-C2C3 la evolución de los parámetros se puede observar en las figuras 5.26-5.29, para b_1 y b_2 , tipos 1 y 2 respectivamente. Las gráficas 5.30-5.33 recogen la misma información pero para el método ORIV-C3C4.

La curva de aprendizaje se recoge en la figura 5.24 para el método ORIV-C2C3 y en la figura 5.25 para ORIV-C3C4. La tabla 5.15 recoge toda la información de los estimadores en la última iteración.

	Estimadores tipo 1				Estimadores tipo 2			
	b_1	b_2	var b_1	var b_2	b_1	b_2	var b_1	var b_2
c	-1.498	1.248	0	0	-1.498	1.248	0	0
C2C3	-1.435	1.233	0.419	0.835	-1.421	1.167	0.142	0.159
C3C4	-1.394	1.160	0.471	0.485	-1.315	1.057	0.396	0.364

Tabla 5.15 *Estimadores de la última iteración. Métodos ORIV-C2C3 y ORIV-C3C4, modelo NEL2. (c=valores correctos).*

Modelo NEL3

De nuevo se considera el modelo NEL1 con una pequeña modificación. En este caso se utiliza la expresión (5.25) para actualizar la varianza, paso necesario para aplicar el método ORIV-C3C4 (no así el ORIV-C2C3). Como siempre, la evolución de los parámetros aparece en las figuras 5.34-5.37 y el MSE en la figura 5.38.

Para hacer la comparación más fácil entre el método ORIV-C3C4 empleando (5.24) y empleando (5.25) (modelo NEL1) en la siguiente tabla se muestran los estimadores de los parámetros del filtro, obtenidos en la iteración final:

	Estimadores tipo 1				Estimadores tipo 2			
	b_1	b_2	var b_1	var b_2	b_1	b_2	var b_1	var b_2
c	-1.249	0.999	0	0	-1.249	0.999	0	0
Con (5.24)	-1.178	0.929	0.091	0.093	-1.188	0.922	0.113	0.133
Con (5.25)	-1.156	0.909	0.282	0.272	-1.154	0.904	0.240	0.222

Tabla 5.16 *Comparación de los estimadores de la última iteración entre el método ORIV-C3C4 con la ecuación (5.24) y (5.25). (c=valores correctos).*

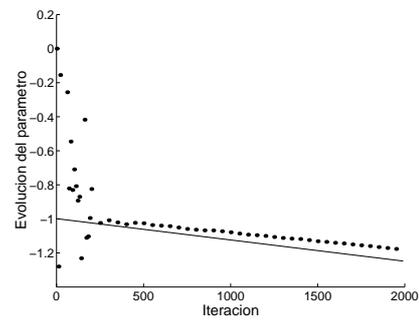


Figura 5.14 Modelo *NEL1*, estimador de b_1 tipo 1. *ORIV-C2C3*

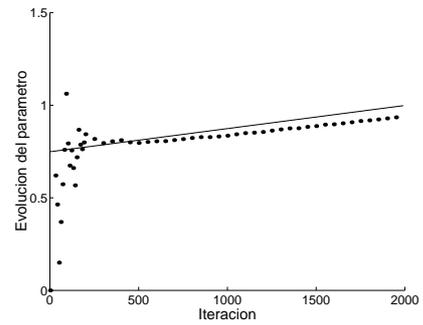


Figura 5.15 Modelo *NEL1*, estimador de b_2 tipo 1. *ORIV-C2C3*

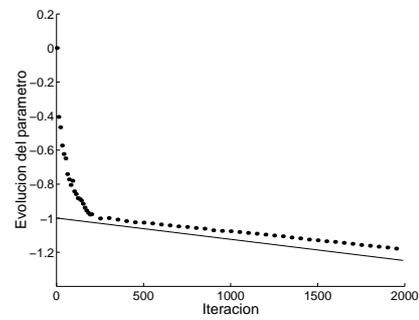


Figura 5.16 Modelo *NEL1*, estimador de b_1 tipo 2. *ORIV-C2C3*

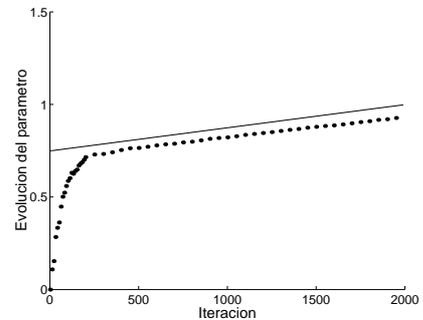


Figura 5.17 Modelo *NEL1*, estimador de b_2 tipo 2. *ORIV-C2C3*

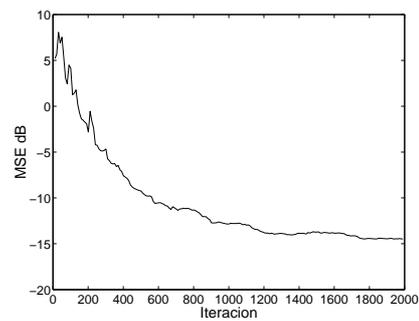


Figura 5.18 Modelo *NEL1*. Curva de aprendizaje de *ORIV-C2C3*

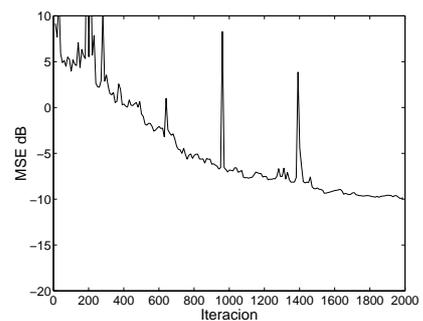


Figura 5.19 Modelo *NEL1*. Curva de aprendizaje de *ORIV-C3C4*

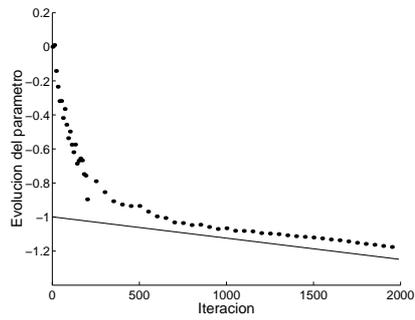


Figura 5.20 *Modelo NEL1, estimador de b_1 tipo 1. ORIV-C3C4*

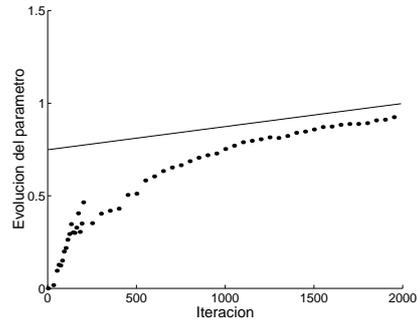


Figura 5.21 *Modelo NEL1, estimador de b_2 tipo 1. ORIV-C3C4*

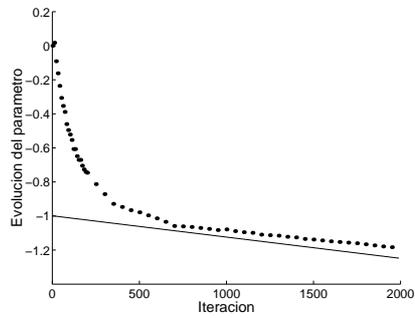


Figura 5.22 *Modelo NEL1, estimador de b_1 tipo 2. ORIV-C3C4*

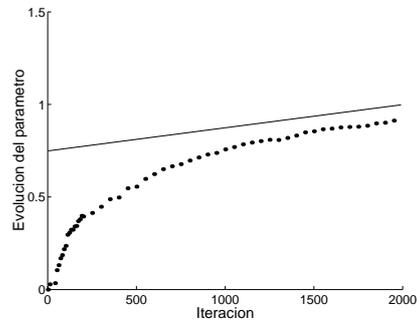


Figura 5.23 *Modelo NEL1, estimador de b_2 tipo 2. ORIV-C3C4*

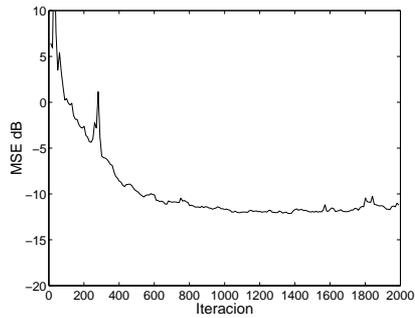


Figura 5.24 *Modelo NEL2. Curva de aprendizaje de ORIV-C2C3*

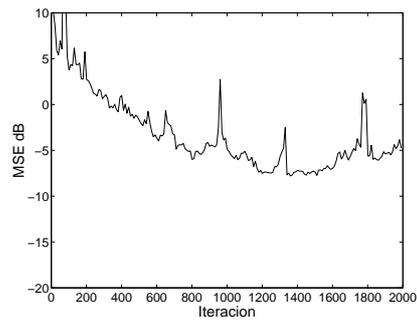


Figura 5.25 *Modelo NEL2. Curva de aprendizaje de ORIV-C3C4*

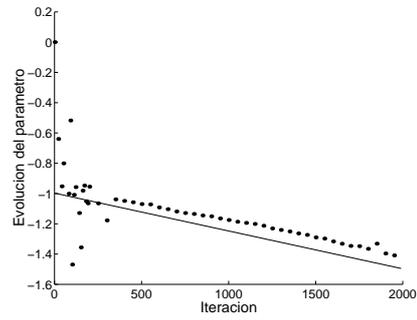


Figura 5.26 *Modelo NEL2, estimador de b_1 tipo 1. ORIV-C2C3*

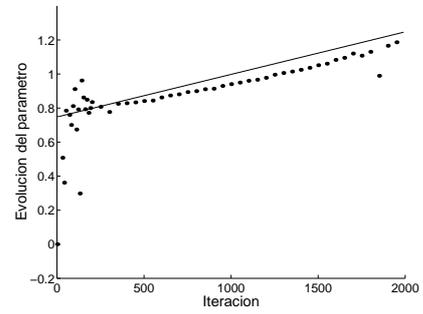


Figura 5.27 *Modelo NEL2, estimador de b_2 tipo 1. ORIV-C2C3*

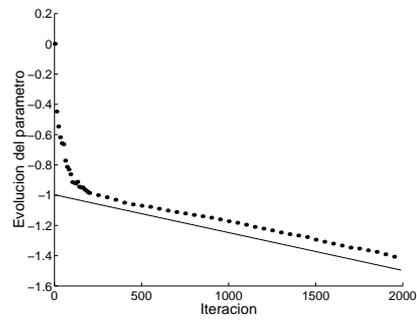


Figura 5.28 *Modelo NEL2, estimador de b_1 tipo 2. ORIV-C2C3*

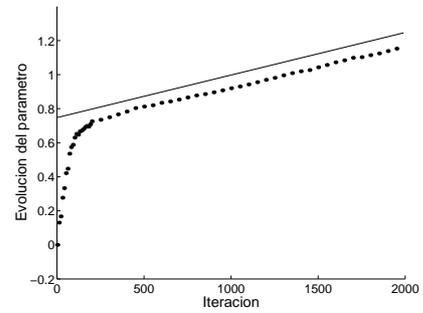


Figura 5.29 *Modelo NEL2, estimador de b_2 tipo 2. ORIV-C2C3*

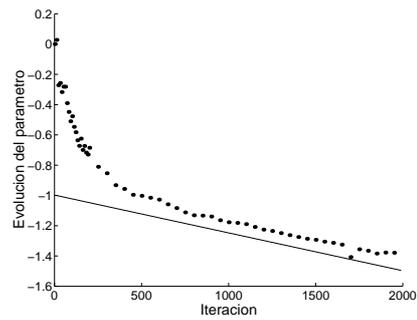


Figura 5.30 *Modelo NEL2, estimador de b_1 tipo 1. ORIV-C3C4*

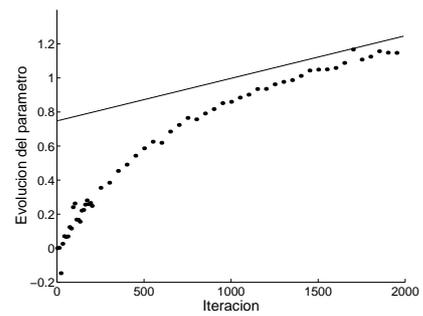


Figura 5.31 *Modelo NEL2, estimador de b_2 tipo 1. ORIV-C3C4*

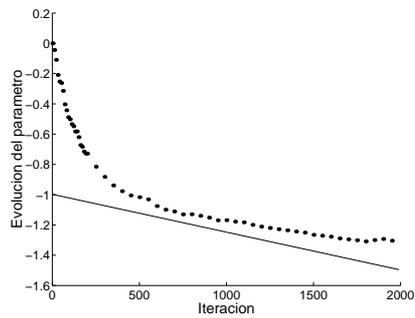


Figura 5.32 *Modelo NEL2, estimador de b_1 tipo 2. ORIV-C3C4*

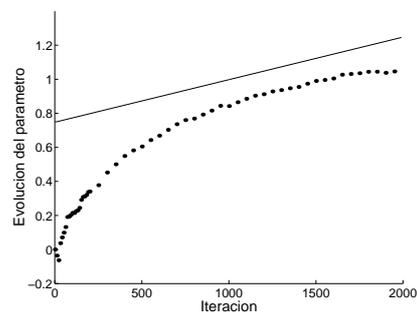


Figura 5.33 *Modelo NEL2, estimador de b_2 tipo 2. ORIV-C3C4*

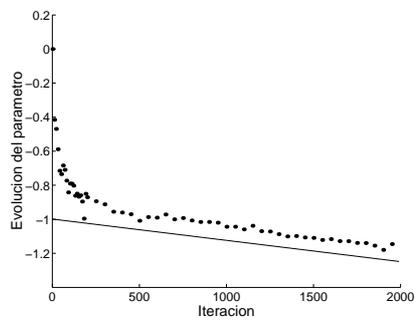


Figura 5.34 *Modelo NEL3, estimador de b_1 tipo 1. ORIV-C3C4*

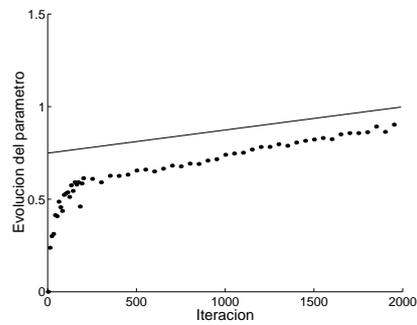


Figura 5.35 *Modelo NEL3, estimador de b_2 tipo 1. ORIV-C3C4*

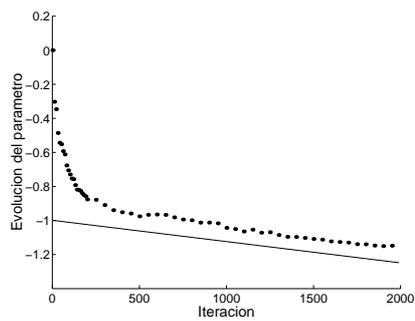


Figura 5.36 *Modelo NEL3, estimador de b_1 tipo 2. ORIV-C3C4*

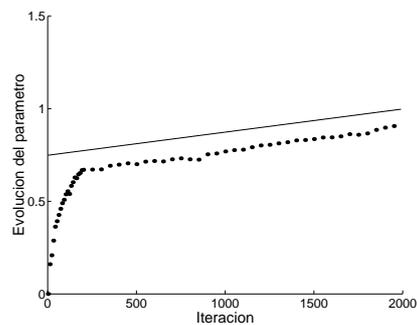


Figura 5.37 *Modelo NEL3, estimador de b_2 tipo 2. ORIV-C3C4*

5.6.2 Análisis de las simulaciones con evolución temporal lineal

Muchas de las características que son apreciables en estas simulaciones se pueden explicar con los comentarios realizados a tenor de las simulaciones llevadas a cabo en el apartado 5.5, por lo que aquí no se discutirán en detalle, sino que simplemente se indicarán.

Como ya se sabe de las simulaciones en ambientes estacionarios realizados en el apartado 5.5, los algoritmos adaptativos, incluso en esos ambientes, necesitan un tiempo de aprendizaje. Este tiempo de aprendizaje es el tiempo que tarda el MSE en llegar a su valor estacionario de convergencia. En ambientes no estacionarios, como es el caso de las simulaciones que se estudian ahora, este tiempo de aprendizaje se hace patente en las representaciones gráficas de la evolución temporal de los estimadores, véanse p.e. las figuras 5.14- 5.17. En estas gráficas se muestra, además de dicha evolución, la evolución temporal real de los parámetros que se han introducido en las simulaciones, que como se ha dicho es de tipo lineal. Los estimadores necesitan cierto tiempo para acercarse a la recta que marca la evolución real. El algoritmo va aprendiendo poco a poco cuáles son los valores verdaderos de los parámetros a cada instante de tiempo.

Hay que recordar la diferencia importantísima que existe entre convergencia en media y convergencia en media cuadrática. Como sus propios nombres indican, la primera pretende que la media para todas las realizaciones del estimador converja al parámetro verdadero. Es lo que se muestra en las gráficas de evolución temporal de estimadores, es decir evolución dinámica de la media de los estimadores. La segunda atañe a la convergencia de la varianza de los estimadores, supuesta como media el valor real del parámetro. Coincide con la definición de MSE, que se adopta en este trabajo, salvo un factor de normalización y está relacionada con las gráficas de la evolución temporal del MSE.

Según este último párrafo, no es de extrañar que el tiempo de convergencia obtenido por simple inspección de la figura 5.14 no coincida con el de la figura 5.18, ya que el tiempo de convergencia en media no es igual al de convergencia en media cuadrática. Esto se comprueba rápidamente examinando las ecuaciones (4.77) y (4.91), que dan como tiempo necesario para que se anule la solución de la homogénea los siguientes valores:

$$\begin{aligned}\tau_{med} &= \frac{1}{1-\lambda^2} \underset{\lambda \rightarrow 1}{\sim} \frac{1}{2(1-\lambda)} \\ \tau_{mse} &= \frac{1}{1-\lambda^4} \underset{\lambda \rightarrow 1}{\sim} \frac{1}{4(1-\lambda)}\end{aligned}\tag{5.49}$$

por lo que según esta expresión, debería converger antes el MSE que la media de los estimadores. Esto no ocurre así porque la solución particular de la no homogénea también tiene un tiempo de convergencia. Este tiempo de convergencia

dependerá de la dinámica particular del sistema en estudio.

Una vez dejado claro la existencia del tiempo de aprendizaje en ambientes no estacionarios, es de notar también la existencia de un tiempo de retraso o retardo. Si cierto parámetro del sistema ha evolucionado en cierto instante n hasta un valor c , el estimador, en media, llegará a él pasado un cierto intervalo Δn contado a partir de n . A Δn se le llama en este trabajo tiempo de retraso e indica que los estimadores siempre irán detrás, dinámicamente hablando, de los valores verdaderos de los parámetros. Este concepto ya apareció en el análisis teórico y se concretó en la ecuación (4.89) para variaciones temporales lineales. El tiempo de retraso se aprecia por ejemplo en las figuras 5.14- 5.17.

Se ha indicado anteriormente la existencia de un tiempo de convergencia. Dicho tiempo de convergencia tiene una definición clara en ambientes estacionarios, sin embargo, para modelos no estacionarios la idea no es clara ya que nunca llega a valer, el estimador correspondiente, un valor constante, sino que también continúa rastreando la evolución del sistema indefinidamente. Conviene por tanto matizar que por tiempo de convergencia se quiere indicar el tiempo necesario que tiene que transcurrir para que los estimadores logren un tiempo de retardo cte. Es decir, que la solución general de la ecuación homogénea asociada a la ecuación (4.77) sea prácticamente nula. En este apartado, donde la variación de los parámetros con el tiempo es lineal, el tiempo de convergencia experimental para los estimadores de los parámetros se puede determinar como aquel a partir del cual la evolución de dichos estimadores sea lineal con la misma pendiente que presentan los parámetros reales. Por ejemplo, de la figura 5.17 se podría decir que el tiempo de convergencia para ORIV-C2C3 en esta simulación es de 300 iteraciones. La ecuación (4.82) da un valor teórico, para $\lambda = 0.9985$, de 333 iteraciones.

La existencia del tiempo de retardo tiene fácil explicación dada la forma de funcionamiento del algoritmo ORIV. Éste estima para cada iteración los cumulantes necesarios para resolver las ecuaciones (5.17) y (5.18). En principio deberían ser los cumulantes a tiempo n , si es que estamos calculando los estimadores de los parámetros a tiempo n . Pero como esta estimación de cumulantes se lleva a cabo mediante promedios temporales, los estimadores de los cumulantes arrastran información de tiempos precedentes, aunque el factor de olvido intenta eliminarlos, con lo cual lo que se tiene es un estimador del cumulante pero a un tiempo ligeramente anterior a n .

También se observa que los estimadores tipo 1 tienen un comportamiento temporal más irregular a la vez que presentan mayor varianza.

Las conclusiones que se pueden obtener del estudio de la evolución del MSE (figuras 5.18 y 5.19, p.e.), son similares a los realizados en el caso de ambientes estacionarios. El método ORIV-C2C3 converge antes y a un valor de MSE estacionario más bajo que para ORIV-C3C4.

A continuación se discuten algunos de los factores que influyen en la capacidad de rastreo de los algoritmos. ORIV es mucho más sensible a los cambios del factor de olvido λ que a cualquier otra variable de la que dependa. La misión de λ es quitarle peso a las muestras antiguas en el promedio temporal que se hace para estimar los momentos y cumulantes. Por ejemplo su uso para la varianza es:

$$\hat{\sigma}^2(n) = \frac{1-\lambda}{1-\lambda^n} \sum_{i=1}^n \lambda^{n-i} y^2(i) \quad (5.50)$$

Como el valor de λ se encuentra comprendido entre 0 y 1, en el caso de que valga 1, todas las muestras tienen igual importancia en la suma y se dice que trabajamos con memoria infinita, cuanto menor sea su valor, sólo las muestras más recientes tendrán un papel importante en la suma y trabajaremos con memorias cada vez más cortas. Sería interesante cuantificar cómo de larga es la memoria o cuán larga es la longitud efectiva, entendiendo como tal el número de muestras pasadas que se considera que juegan un papel importante en el promedio. Para eso vamos recurrir a la expresión (5.25) que es más sencilla matemáticamente,

$$\hat{\sigma}^2(n+1) = \lambda \hat{\sigma}^2(n) + (1-\lambda)y^2(n+1) \quad (5.51)$$

que es una ecuación en diferencias no homogénea. Lo que se busca es la solución de la homogénea, que tiende a cero. Dicha solución es $\sigma^2(n) = \lambda^n \sigma^2(0) = \sigma^2(0) \exp(n \ln \lambda)$, si se supone que la longitud efectiva es aquella para la cual la solución de la homogénea ha caído hasta $1/e$ de su valor inicial, se tiene que

$$t_{effect} \ln \lambda = -1 \Rightarrow t_{effect} = \frac{-1}{\ln \lambda} = \frac{-1}{\ln(1 - (1 - \lambda))} \sim \frac{1}{1 - \lambda} \quad (5.52)$$

donde se ha supuesto que λ es cercano a la unidad, como ocurre en la mayoría de los casos. De la definición de longitud efectiva, se comprueba efectivamente que para factores de olvido iguales a la unidad, se tiene longitud infinita, es decir memoria infinita y que conforme más pequeño es, menor es la longitud efectiva.

Para obtener buenos estimadores, sin sesgo y poca varianza, se necesita aumentar el número de muestras que se van a promediar; es decir se necesita una longitud efectiva grande, cuanto más grande mejor, idealmente infinito. Sin embargo, si la serie aleatoria de la que se calculan los cumulantes no tiene una estadística estacionaria, para calcular los estimadores a cierto tiempo sólo se pueden tener en cuenta las muestras más cercanas al instante temporal considerado; lo que supone longitud efectiva finita. Cuanto mayor sea la variación de la estadística con el tiempo, menor longitud efectiva se puede utilizar para estimar los cumulantes a un tiempo dado.

Si se extrapolan las ideas anteriores al ejemplo que ahora se trata, cuanto mayor sea la pendiente de la variación lineal que se imponga a la evolución real de

los parámetros MA, se va a necesitar un λ más pequeño (menos memoria) para que se adapte antes el algoritmo y pueda seguir los cambios. En contrapartida, al tener menos memoria los estimadores de los momentos presentan mayor varianza y por tanto se tendrá un mayor MSE. Como muestra compárense los modelos NEL1 y NEL2:

Modelo NEL1: $\mathbf{p} \sim 1/8000 \lambda = 0.9985 t_{efect} = 667$

Modelo NEL2: $\mathbf{p} \sim 1/4000 \lambda = 0.9975 t_{efect} = 400$

La longitud efectiva en cada caso se obtuvo realizando varias simulaciones y eligiendo la que proporcionaba los resultados más satisfactorios. Experimentalmente, por tanto, se comprueba lo discutido anteriormente: el modelo NEL2 tiene mayor variación lineal por lo que necesita un λ menor y por tanto menor longitud efectiva. Ocurre lo contrario en el modelo NEL1. Y efectivamente se obtiene menor MSE en el modelo NEL1 que en el modelo NEL2. (Ver figuras 5.18 y 5.24, p.e.).

Ya se ha comentado también que es necesario cierta longitud efectiva para estimar bien los cumulantes. Si \mathbf{p} es muy grande se necesita, en teoría, un λ muy pequeño para poder seguir la variación temporal. Pero entonces tendría una longitud efectiva corta y por tanto malos estimadores que a la larga impiden seguir el comportamiento con el tiempo del sistema. Como muestra, en el modelo NEL2 se aprecia cierta divergencia tanto en la evolución de los parámetros (figuras 5.32 y 5.33) como en el comportamiento del MSE (figura 5.25).

Resumiendo se puede decir que existe un compromiso entre rastrear procesos que cambien rápidamente con el tiempo, que necesitan λ pequeña, y obtener buenos estimadores ya que se necesita $\lambda \sim 1$. Por tanto, está limitada la ‘velocidad’ de cambio del modelo MA que ORIV puede seguir. Como consecuencia de estas necesidades contrapuestas aparece un valor de λ que da un valor mínimo de MSE dada una velocidad de cambio, es decir, existe un λ_{opt} cuya expresión se obtuvo en el capítulo 4, ecuación (4.98).

Hasta ahora sólo se han descrito comportamientos generales de ORIV sin atender al sistema lineal de ecuaciones que resuelve. Pero a la vista de todo lo comentado hasta este punto es fácil deducir el porqué de la diferencia de comportamiento entre los métodos ORIV-C2C3 y ORIV-C3C4.

Ya se ha dicho que el método ORIV-C2C3 logra en menos tiempo un menor MSE estacionario. El motivo ya se comentó en el apartado 5.5. Además logra un menor tiempo de convergencia y un menor tiempo de retardo en el seguimiento de los parámetros MA, en igualdad de condiciones, como se comprueba observando las figuras 5.17 y 5.23. El motivo es análogo al dado para el caso estacionario. Aunque no se trabaje con memoria infinita, sino finita, siempre se tendrán mejores estimadores de tercer orden con ese número de datos (longitud

efectiva) que de cuarto y además, por la misma razón los estimadores de la tabla 5.14 presentan menor varianza para ORIV-C2C3 que para ORIV-C3C4.

¿Qué pasa si se aumenta la velocidad de la variación lineal (pendiente) y por tanto es necesario reducir la longitud efectiva? Si se reduce paulatinamente la memoria, llegará un momento en que no será lo suficientemente extensa como para conseguir estimadores de 4º orden de suficiente calidad, con lo que el algoritmo divergirá para el método ORIV-C3C4, siendo posible sin embargo que funcione para el método ORIV-C2C3. Este comportamiento puede verse por ejemplo en las figuras 5.24 y 5.25 relativas al modelo NEL2.

De todo lo discutido hasta aquí la superioridad de ORIV-C2C3 sobre ORIV-C3C4 es absoluta en todos los frentes: menor tiempo de convergencia, menor tiempo de retardo, menor MSE, menor varianza... Sin embargo no se debe olvidar que se está trabajando en ambientes libres de ruido y que ha sido en ambientes ruidosos donde ORIV-C3C4 ha mostrado su superioridad para sistemas estacionarios. Se estudiarán modelos no estacionarios en ambiente ruidoso en el apartado siguiente.

Todavía queda por comentar una pequeña cuestión que es ajena a la comparación entre ORIV-C2C3 y ORIV-C3C4 y que afecta en general al funcionamiento de ORIV. Es relativa al uso de las ecuaciones (5.24) ó (5.25) en la actualización de la varianza de $\{y(n)\}$, cálculo que hay que realizar sólo en el caso de que el sistema lineal que se desea resolver mediante ORIV involucre estadística de 4º orden, como ocurre si se trabaja con el método ORIV-C3C4. Las simulaciones presentadas hasta ahora han hecho uso de (5.24) por ser una expresión más exacta, sin embargo Friedlander [FP89] propuso el uso de (5.25). Para ver las diferencias de comportamiento se considera de nuevo el modelo NEL1 y se repiten los cálculos pero usando (5.25), dando lugar al modelo NEL3. En la tabla 5.16 se muestran los estimadores a tiempo 2000 en ambos casos; la ecuación de actualización utilizada en este trabajo se muestra superior tanto en sesgo como en varianza, por lo que también es de esperar que ocurra lo mismo en la evolución del MSE (ver figuras 5.19 y 5.38) como así ocurre. Por tanto, en general, los resultados son mejores utilizando la ecuación (5.24) que la ecuación (5.25).

5.6.3 Resultados de las simulaciones con variación temporal aleatoria

Otra posible forma de modelar la variación temporal de los coeficientes del filtro a identificar es mediante un proceso de Markov. Un proceso de Markov de orden 1 viene definido por la siguiente expresión:

$$\mathbf{b}(n) = a\mathbf{b}(n-1) + \mathbf{u}(n) \quad (5.53)$$

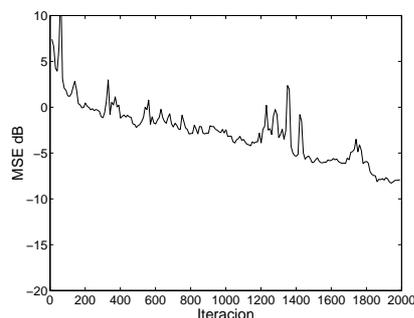


Figura 5.38 Modelo NEL3. Curva de aprendizaje de ORIV-C3C4

es decir, un proceso donde el estado a cierto tiempo n depende del estado a tiempo anterior $n - 1$ mediante un escalar multiplicativo a menor que la unidad y contiene cierto ruido expresado por el vector $\mathbf{u}(n)$.

A $\mathbf{u}(n)$ se le conoce como vector de ruido del proceso, de naturaleza aleatoria, que se tomará de media nula y matriz de autocorrelación múltiplo de la identidad. La función densidad de probabilidad de cada componente se supondrá constante y de varianza σ_u^2 .

Suponiendo una evolución de este tipo se han estudiado diversos modelos. La identificación de modelos de Markov empleando el algoritmo RLS aparece en [AJS99].

Modelo NEA1

Como primer modelo se ha intentado identificar un sistema que evoluciona según un proceso de Markov de orden 1 con $a = 0.9998$, $\sigma_u^2 = 0.04$ y condición inicial $\mathbf{b}(0) = [1 \ -1 \ 0.75]$. Para cada una de las 1000 realizaciones promediadas se disponía de 3000 muestras y el factor de olvido tomaba el valor 0.9985. Al igual que para la evolución temporal lineal, el cambio en los pesos del filtro empezaba a producirse a partir de un cierto t_{cambio} , fijado en este caso como $n = 10$.

Como la evolución de los parámetros verdaderos es aleatoria y ruidosa, de una representación gráfica no se podría obtener mucha información visual sobre el rastreo por parte de los parámetros estimados. Por este motivo la única información útil es la curva de aprendizaje, representada en la figura 5.39(a) para el método ORIV-C2C3 y en la figura 5.39(b) para el método ORIV-C3C4.

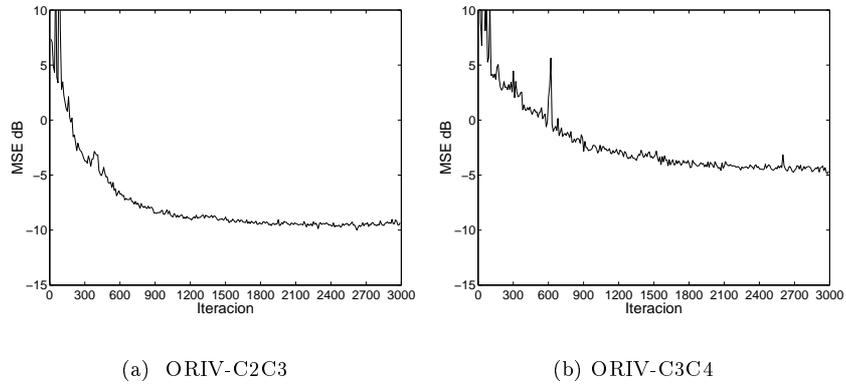


Figura 5.39 *Modelo NEA1. Curva de aprendizaje*

Modelo NEA2

Para este segundo modelo el único factor que cambia es λ que se fija ahora como 0.9975. La curva de aprendizaje para el método ORIV-C2C3 se presenta en la figura 5.40(a) y para ORIV-C3C3 en la figura 5.40(b).

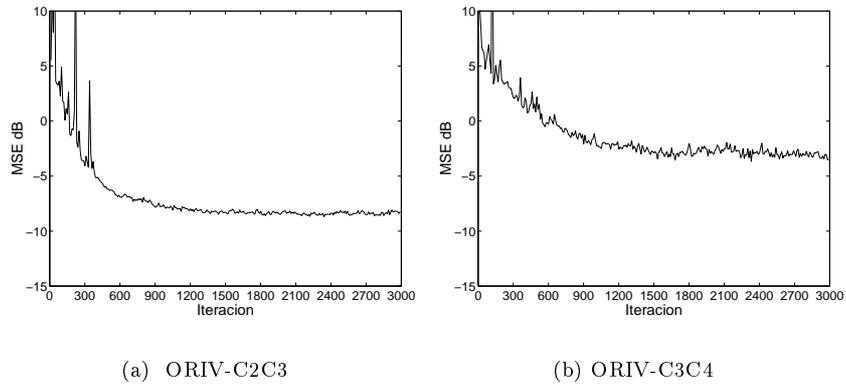


Figura 5.40 *Modelo NEA2. Curva de aprendizaje*

5.6.4 Análisis de las simulaciones con evolución temporal aleatoria

Con estas simulaciones se pone de manifiesto que los métodos ORIV-C2C3 y ORIV-C3C4 son capaces de seguir la evolución temporal aleatoria de un sistema lineal.

La mayoría de las conclusiones que se pueden obtener en este apartado, están ya incluidas en apartados anteriores, ya que presenta muchos elementos comunes. Por ejemplo, ORIV-C2C3 produce un menor MSE en convergencia que ORIV-C3C4 (ver figuras 5.39(a) y 5.39(b)). De la definición de MSE se sabe que es una medida tanto del sesgo de los estimadores como de la varianza de los mismos. En ambientes sin ruido, $\text{SNR} = +\infty$, como es este caso, ambos conjuntos de ecuaciones (5.17) y (5.18) producen estimadores no sesgados siempre que los estimadores de los cumulantes involucrados no lo sean. Si se supone que los cumulantes están correctamente estimados (aunque se trabaje con sistemas no estacionarios), el mayor MSE de ORIV-C3C4 se debe a una mayor varianza de los estimadores de los parámetros del sistema. El motivo se encuentra en el aumento de varianza que experimentan los estimadores de los cumulantes al ir subiendo el orden de la estadística, y basarse ORIV-C3C4 en estadística de orden superior a ORIV-C2C3. Pero además puede haber otro motivo para el aumento de MSE en ORIV-C3C4: puede que los cumulantes estimados a tiempo n estén sesgados ya que se necesita cierto intervalo temporal para poder realizar el promedio para la obtención de estos estimadores. Como también se ha comentado previamente, este intervalo temporal, o longitud efectiva, es necesario que sea mayor para estadística de cuarto orden que de tercer orden; una vez fijada la longitud efectiva, es decir, una vez fijada λ , ORIV-C2C3 se encuentra en ventaja frente a ORIV-C3C4. Estos dos se pueden considerar como los motivos principales por los que ORIV-C3C4 da peor MSE en convergencia.

En lo referente al tiempo de convergencia, cabe volver a recalcar que ORIV-C2C3 converge antes, ya que la estadística en la que se basa necesita mucho menos tiempo para ser estimada.

Finalmente se analiza la influencia del factor de olvido, λ , en el comportamiento del algoritmo. En el modelo NEA1 se trabaja con $\lambda = 0.9985$ ($t_{\text{effect}} = 666.67$), en el modelo NEA2 con $\lambda = 0.9975$ ($t_{\text{effect}} = 400$), lo que significa que el segundo es capaz de seguir más rápido las variaciones temporales pagando el precio de mayor varianza y sesgo. Esto se traduce en menor tiempo de convergencia, pero mayor MSE en convergencia. Además, en principio, el método que se va a ver más afectado por este cambio va a ser ORIV-C2C3 ya que necesita menor número de datos para estimar su estadística y si le imponemos que use más le estamos restando capacidad de rastrear el sistema, lo que implica un aumento de MSE en convergencia. Sin embargo, para ORIV-C3C4, aunque estime mejor los cumulantes de cuarto orden con 666 datos que con 400 influye

el hecho de que tiene menos capacidad de rastreo, por lo que son dos efectos que se contraponen y por tanto su comportamiento global no se ve muy afectado. En consecuencia, se pueden comparar las figuras 5.39(b) y 5.40(b), en la del modelo NEA1 se aprecia ligeramente un MSE menor y un tiempo de convergencia ligeramente superior al del modelo NEA2, pero sin grandes diferencias.

Por último, y comparándolo con el apartado 5.6.1 en el que la variación temporal era lineal, aún cabe un pequeño comentario. En los resultados de las simulaciones en sistemas con evolución temporal aleatoria no se aprecia ni la más mínima tendencia de divergencia como aparecía por ejemplo en el modelo NEL2 del apartado 5.6.1. Sin entrar en la discusión de si esto es debido a la menor variación del modelo aleatorio o no (quizás fuese conveniente definir un grado de no estacionariedad para ganar en objetividad³), lo que cabe destacar es la forma en la que varían los resultados para cada dinámica. Para dinámica lineal los resultados van aumentando su diferencia con el valor inicial conforme pasa el tiempo, si el algoritmo no puede seguirla cada vez dará peor resultado y acabará divergiendo. Sin embargo, en el caso aleatorio la estimación puede oscilar alrededor de un punto concreto, o dicho de otro modo, su evolución no tiene porqué llevarnos a un punto en el espacio de parámetros lejano al de partida, con lo que el algoritmo no tiene porqué divergir por este motivo.

5.7 Los métodos ORIV-C2C3 y ORIV-C3C4 en ambientes no estacionarios ruidosos

Ya se ha analizado el problema de la identificación ciega en dos ambientes distintos: uno incluía ruido a las muestras disponibles y el otro incluía la posibilidad de que los pesos del filtro evolucionaran con el tiempo. En esta sección se combinan ambos para analizar el problema en la situación más general posible, esto es, sistemas no estacionarios contaminados por ruido.

Es conocido del análisis teórico del capítulo 4 que la influencia del ruido se puede minimizar en los estimadores si el factor de olvido está muy próximo a uno. Sin embargo este requisito limita la capacidad de rastreo del algoritmo. Debido a esta oposición ‘eliminación del ruido’-‘capacidad de rastreo’ aparece el concepto de velocidad límite que se analizará mas adelante.

5.7.1 Resultados de las simulaciones

El ruido aditivo se va a suponer gaussiano coloreado, obtenido de igual manera que en el apartado 5.5 pasando ruido gaussiano blanco por un filtro ARMA

³Macchi lo hizo en [Mac95].

de coeficientes AR = $[1 \ -2.2 \ 1.77 \ -0.52]$ y MA = $[1 \ -1.25]$, con unos niveles de SNR de 20, 10, 5, 0, -5 y -10 dB. La evolución temporal se supondrá lineal, con 2000 muestras conocidas y un factor de olvido $\lambda = 0.9985$, aunque las características específicas de cada modelo estudiado se detallan a continuación. Los coeficientes del sistema empiezan a evolucionar a partir de $t_{cambio} = 10$.

Modelo NER1

El modelo MA de partida es $\mathbf{b}(0) = [1 \ -1 \ 0.75]$, que evoluciona en el tiempo según la pendiente $\mathbf{p} = [0 \ -1/8000 \ 1/8000]$ con $b_0 = 1 \ \forall n$. Los estimadores de los coeficientes del filtro en la iteración final para el método ORIV-C2C3 se muestran en la tabla 5.18, apareciendo tanto su media como su varianza para los distintos niveles de ruido estudiados. La misma información contiene la tabla 5.19 referente al método ORIV-C3C4. En la figura 5.41(a) se muestran las curvas de aprendizaje para ORIV-C2C3 y en 5.41(b) para ORIV-C3C4. El símbolo asociado a cada SNR es:

$$\begin{array}{ll} \diamond & = 20 \text{ dB} & + & = 0 \text{ dB} \\ \text{O} & = 10 \text{ dB} & * & = -5 \text{ dB} \\ \times & = 5 \text{ dB} & - & = -10 \text{ dB} \end{array}$$

SNR	Estimadores tipo 1				Estimadores tipo 2			
	b_1	b_2	var b_1	var b_2	b_1	b_2	var b_1	var b_2
c	-1.249	0.999	0	0	-1.249	0.999	0	0
20	-1.081	0.949	0.131	0.144	-1.179	0.931	0.085	0.091
10	-0.500	1.022	0.190	0.221	-1.137	0.895	0.175	0.135
5	0.497	0.811	26.106	38.722	-0.143	0.587	0.919	0.408
0	0.047	0.831	6.109	6.540	0.144	0.277	0.624	0.619
-5	12.430	-0.985	293.859	38.775	0.102	0.294	0.811	0.638
-10	-2.774	1.350	79.502	49.263	0.042	0.343	0.877	0.720

Tabla 5.18 Media y varianza de los estimadores (c =valores correctos), modelo NER1. Método ORIV-C2C3

SNR	Estimadores tipo 1				Estimadores tipo 2			
	b_1	b_2	var b_1	var b_2	b_1	b_2	var b_1	var b_2
c	-1.249	0.999	0	0	-1.249	0.999	0	0
20	-1.181	0.927	0.086	0.126	-1.183	0.918	0.121	0.156
10	-1.182	0.870	0.209	0.400	-1.135	0.846	0.227	0.333
5	-1.026	0.741	1.034	0.907	-0.844	0.701	0.580	0.502
0	-0.436	0.402	1.871	1.764	-0.410	0.423	0.730	0.689
-5	-0.253	0.160	1.998	1.850	-0.462	0.104	0.740	0.797

-10	-0.435	0.265	4.074	5.088	-0.602	0.179	0.710	0.791
-----	--------	-------	-------	-------	--------	-------	-------	-------

Tabla 5.19 Media y varianza de los estimadores (c =valores correctos), modelo *NER1*. Método *ORIV-C3C4*

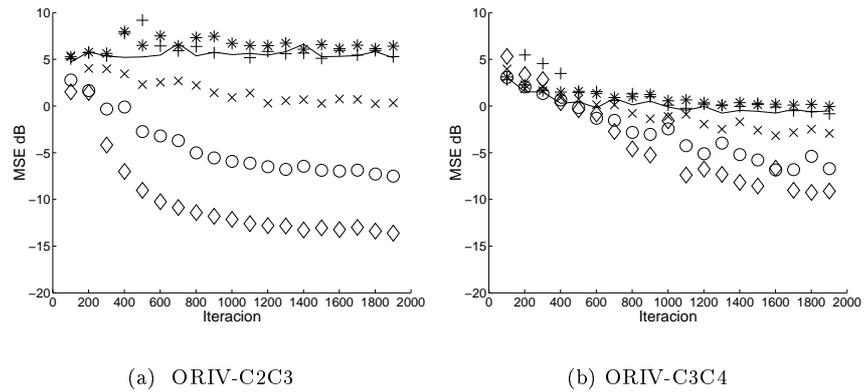


Figura 5.41 Modelo *NER1*. Curva de aprendizaje

Modelo NER2

Este segundo modelo es una $MA(3)$ con coeficientes iniciales $MA = [1 \ -0.9 \ 0.385 \ -0.771]$ y con pendiente de su variación lineal dada por $\mathbf{p} = [0 \ -1/8000 \ 1/8000 \ -1/8000]$. La media y varianza de los estimadores de las tablas 5.20 y 5.21 se obtuvieron con el método *ORIV-C2C3* para los estimadores tipo 1 y 2 respectivamente y los de las tablas 5.22 y 5.23 con *ORIV-C3C4*. Igualmente 5.42(a) muestra la curva de aprendizaje para *ORIV-C2C3* y 5.42(b) para *ORIV-C3C4*.

SNR	b_1	b_2	b_3	var b_1	var b_2	var b_3
c	-1.149	0.634	-1.020	0	0	0
20	-1.197	0.546	-0.924	1.718	2.627	1.406
10	-1.955	0.169	-0.580	1.951	2.337	0.826
5	-2.434	-1.038	-0.850	92.336	30.621	22.876
0	1.470	-1.492	0.890	62.523	44.972	18.504
-5	-1.010	0.404	-1.481	45.103	33.955	38.534

-10	18.662	-40.400	32.016	544.708	1258.038	1012.649
-----	--------	---------	--------	---------	----------	----------

Tabla 5.20 Media y varianza de los estimadores tipo 1 (c =valores correctos), modelo *NER2*. Método *ORIV-C2C3*

SNR	b_1	b_2	b_3	var b_1	var b_2	var b_3
c	-1.149	0.634	-1.020	0	0	0
20	-1.065	0.152	-0.835	0.198	0.554	0.192
10	-0.969	-0.035	-0.313	0.298	0.707	0.399
5	-0.703	-0.005	-0.191	0.548	0.919	0.548
0	-0.052	0.002	0.070	0.726	1.036	0.821
-5	0.005	0.026	-0.008	0.825	0.900	0.906
-10	0.009	0.104	0.050	0.857	0.796	0.817

Tabla 5.21 Media y varianza de los estimadores tipo 2 (c =valores correctos), modelo *NER2*. Método *ORIV-C2C3*

SNR	b_1	b_2	b_3	var b_1	var b_2	var b_3
c	-1.149	0.634	-1.020	0	0	0
20	-1.102	-0.074	-0.874	0.329	0.744	0.351
10	-1.054	-0.137	-0.763	0.673	0.846	0.645
5	-0.923	-0.078	-0.577	1.384	1.653	1.760
0	-0.662	0.053	-0.168	6.185	5.273	3.150
-5	-0.281	-0.118	0.216	2.660	2.959	2.712
-10	-0.328	-0.114	0.282	1.856	1.956	1.829

Tabla 5.22 Media y varianza de los estimadores tipo 1 (c =valores correctos), modelo *NER2*. Método *ORIV-C3C4*

SNR	b_1	b_2	b_3	var b_1	var b_2	var b_3
c	-1.149	0.634	-1.020	0	0	0
20	-1.020	0.125	-0.845	0.197	0.682	0.335
10	-1.002	0.086	-0.737	0.297	0.752	0.478
5	-0.927	0.078	-0.504	0.442	0.816	0.644
0	-0.508	0.014	-0.181	0.746	0.811	0.735

-5	-0.435	-0.289	0.073	0.688	0.707	0.753
-10	-0.658	-0.197	0.259	0.606	0.772	0.731

Tabla 5.23 Media y varianza de los estimadores tipo 2 (c =valores correctos), modelo *NER2*. Método *ORIV-C3C4*

Modelo *NER3*

Por último, este y el siguiente modelo estudian las ventajas de emplear la ecuación de actualización (5.24) frente a (5.25), para lo cual se retoman los modelos anteriores. En particular este modelo *NER3* es análogo al modelo *NER1* pero utilizando (5.25) al resolverlo mediante *ORIV-C3C4*.

La media y varianza de los estimadores se muestran en la tabla 5.24 y la curva de aprendizaje en la figura 5.43(a).

SNR	Estimadores tipo 1				Estimadores tipo 2			
	b_1	b_2	var b_1	var b_2	b_1	b_2	var b_1	var b_2
c	-1.2486	0.9986	0	0	-1.2486	0.9986	0	0
20	-1.157	0.896	0.275	0.294	-1.148	0.882	0.258	0.278
10	-1.099	0.893	0.498	0.392	-1.029	0.860	0.382	0.287
5	-1.044	0.806	0.947	0.831	-0.802	0.762	0.491	0.415
0	-0.242	0.469	1.592	1.424	-0.292	0.480	0.735	0.638
-5	-0.480	0.248	10.465	4.009	-0.371	0.099	0.735	0.768
-10	-0.179	0.311	2.705	2.259	-0.584	0.130	0.686	0.794

Tabla 5.24 Media y varianza de los estimadores (c =valores correctos), modelo *NER3*. Método *ORIV-C3C4*

Modelo *NER4*

Este último modelo retoma el modelo *NER2* pero actualiza la varianza de $\{y(n)\}$ necesaria para la definición (5.37) mediante (5.25). En las tablas 5.25 y 5.26 se recogen los estimadores y en la figura 5.43(b) la evolución temporal del MSE.

SNR	b_1	b_2	b_3	var b_1	var b_2	var b_3
c	-1.149	0.634	-1.020	0	0	0
20	-1.086	-0.082	-0.848	0.409	0.736	0.433
10	-1.077	-0.191	-0.737	0.562	0.846	0.627
5	-0.969	-0.158	-0.548	1.185	1.301	1.132
0	-0.529	-0.078	-0.316	2.348	1.943	2.130
-5	-0.101	-0.035	0.040	2.008	1.793	2.021
-10	-0.204	-0.050	0.270	2.069	2.084	2.081

Tabla 5.25 *Media y varianza de los estimadores tipo 1 (c =valores correctos), modelo NER_4 . Método ORIV-C3C4*

SNR	b_1	b_2	b_3	var b_1	var b_2	var b_3
c	-1.149	0.634	-1.020	0	0	0
20	-1.019	0.129	-0.840	0.231	0.684	0.357
10	-0.996	0.009	-0.738	0.285	0.754	0.486
5	-0.902	-0.053	-0.498	0.491	0.852	0.670
0	-0.512	-0.048	-0.274	0.758	0.845	0.712
-5	-0.314	-0.282	-0.015	0.734	0.734	0.736
-10	-0.664	-0.208	0.274	0.583	0.764	0.713

Tabla 5.26 *Media y varianza de los estimadores tipo 2 (c =valores correctos), modelo NER_4 . Método ORIV-C3C4*

5.7.2 Análisis de las simulaciones

La diferencia de las simulaciones llevadas a cabo en este apartado con respecto a las del apartado 5.6 está en la adición de ruido gaussiano coloreado. Por tanto, la discusión se limitará aquí a ver cómo influye este nuevo factor en los resultados, completando de esta manera los comentarios de dicho apartado. Antes de pasar a discutir las diferencias entre los métodos ORIV-C2C3 y ORIV-C3C4, se pasa a comentar las características que son comunes a ambos métodos.

Tanto la varianza como el sesgo de los estimadores de los parámetros del modelo MA, aumentan al disminuir la SNR, lo cual es lógico ya que cada vez la señal llega envuelta en mayor nivel de ruido, lo que hace más difícil la estima-

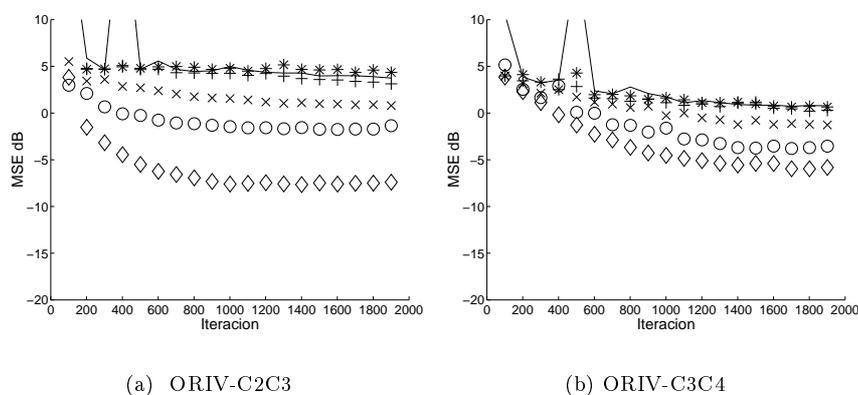


Figura 5.42 Modelo NER2. Curvas de aprendizaje

ción. Por el mismo motivo anterior el MSE de convergencia va aumentando al disminuir la SNR.

Como ya se ha comentado varias veces, y es algo directo de ver, el comportamiento de las ecuaciones (5.17) y (5.18) depende en gran medida de la capacidad de ORIV para estimar los cumulantes que cada sistema de ecuaciones necesita. El primero depende sólo de cumulantes de segundo y tercer orden, con lo cual, en presencia de ruido aditivo gaussiano, los de 2º orden serán sesgados. El sesgo será mayor conforme aumente el nivel de ruido, no así la varianza, ya que el cumulante de 2º orden simplemente verá una señal $\{y(n)\}$ compuesta de $\{x(n)\}$ y $\{v(n)\}$, por lo que el cumulante será la suma de los cumulantes de $\{x(n)\}$ y $\{v(n)\}$, al ser independientes. Cuanto mayor sea el ruido, mayor será el cumulante de $\{v(n)\}$ y por tanto, el cumulante de la suma más separado estará del cumulante de interés, es decir, el de $\{x(n)\}$. Esto viene a corroborar lo ya expresado con anterioridad, en el sentido de que el sistema de ecuaciones (5.17) se ve fuertemente afectado de ruido coloreado. La influencia de lo comentado sobre los resultados es clara: al aumentar el nivel de ruido, los estimadores de los cumulantes de 2º orden aumentan su sesgo, aumentando por tanto el sesgo de los estimadores de los parámetros, pero no su varianza. Además de depender de los cumulantes de 2º orden, la ecuación (5.17) también depende de cumulantes de tercer orden. Estos cumulantes son, teóricamente, inmunes al ruido gaussiano sea cual sea su contenido en frecuencia. En la práctica, sin embargo, cuando la SNR disminuye, los estimadores se ven afectados por una mayor varianza, no influyendo en principio en su sesgo. Si se une el aumento de sesgo causado por los cumulantes de 2º orden, con el aumento de varianza causado por los de tercero, se llega a la obtención de estimadores de parámetros cuyo sesgo y varianza aumentan al disminuir la SNR, tal y como se aprecia en

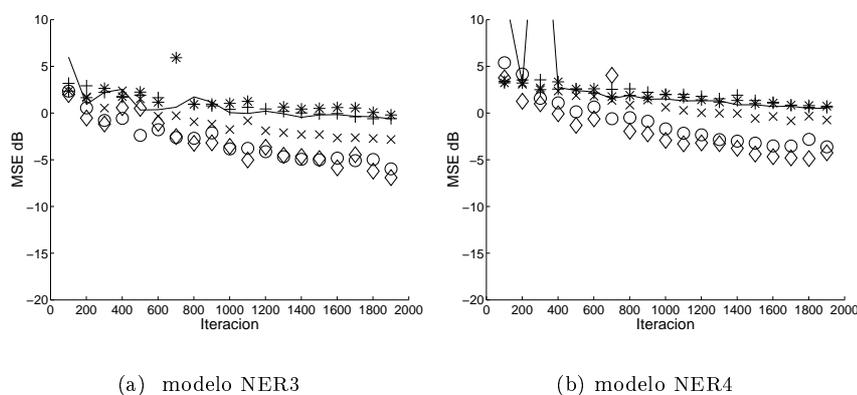


Figura 5.43 Curvas de aprendizaje de los modelos NER_3 y NER_4 para $ORIV-C3C4$

las tablas (ver p.e. tabla 5.18).

Por otro lado, la ecuación (5.18) depende de cumulantes de tercer y cuarto orden. Los estimadores de estos cumulantes, incluso en presencia de ruido coloreado, son no sesgados; la única influencia del ruido está en el aumento de la varianza. ¿Qué cabe esperar por tanto de sus resultados? Los parámetros estimados mediante la ecuación (5.18) deberían ser no sesgados y con una varianza que aumentase conforme aumente el nivel de ruido. Inspeccionando la tabla 5.19 se observa que esto es cierto hasta los 5 dB aproximadamente, de 0 dB hasta los -10 dB el sesgo se hace también importante. El motivo de la aparición del sesgo es que las 1000 realizaciones ya no son suficientes para obtener una buena media del estimador cuando el ruido es demasiado alto, es decir, la SNR es baja.

Todo lo comentado hasta ahora se trató en la apartado 5.5, en el contexto de modelos estacionarios en presencia de ruido. Para sistemas no estacionarios sigue siendo igualmente válido tal y como se ha visto aquí. La influencia de la variación temporal del sistema en la estimación está en un pequeño aumento del MSE debido al tiempo de retardo, que implica mayor sesgo en los estimadores para cada instante de tiempo. La influencia del tiempo de retardo y del factor de olvido en modelos estacionarios se vio en el apartado 5.6.1.

Las conclusiones que se pueden obtener de estas simulaciones serían un compendio de las extraídas en los apartados 5.5 y 5.6. Se puede afirmar que el método $ORIV-C3C4$, en general, mejora los resultados obtenidos por el método $ORIV-C2C3$ para $SNR < 10$ dB, es decir cuando el sesgo y la varianza de los

estimadores de los cumulantes de orden 2 y 3 es mayor que la varianza de los estimadores de orden 3 y 4.

En definitiva, los resultados obtenidos en este apartado concuerdan con los obtenidos en apartados anteriores, tanto en función del nivel de ruido coloreado contaminante como para la normalización utilizada en la estimación de la varianza de $\{y(n)\}$.

5.7.3 Resultado de las simulaciones relacionadas con el seguimiento límite

Se continúa con el estudio de sistemas no estacionarios y en particular aquellos que presentan evolución temporal lineal. En las simulaciones llevadas a cabo en esta sección se va a comprobar que para un factor de olvido dado el algoritmo no puede seguir variaciones en los parámetros generadas por una pendiente mayor que una cota dada. Este fenómeno se ha llamado seguimiento límite o velocidad límite en esta Memoria.

Los modelos estudiados se basan en un MA(2) con coeficientes iniciales MA = [1 -1 0.75] del que se conocen 5000 muestras ruidosas de su salida. La evolución temporal lineal tendrá efecto a partir de $t_{cambio} = 10$. Los detalles concretos de cada modelo son los siguientes:

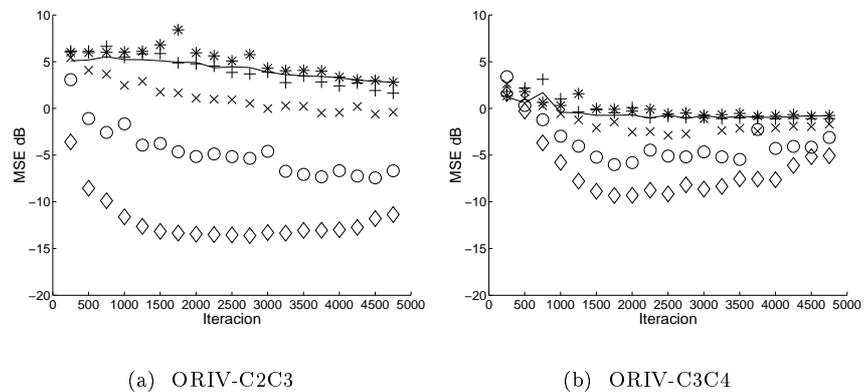


Figura 5.44 Curvas de aprendizaje del modelo SL1

Modelo SL1

Para este primer modelo el vector pendiente de la variación lineal es $\mathbf{p} = [0 - 1/8000 \ 1/8000]$, la media y varianza de los estimadores obtenidos en la última iteración por medio del método ORIV-C2C3 se presentan en la tabla 5.27 y por medio de ORIV-C3C4 en la tabla 5.28. Las curvas de aprendizaje respectivas se muestran en las figuras 5.44(a) y 5.44(b). Asimismo resulta ilustrativo mostrar la convergencia en media de los estimadores, para lo cual se han elegido los estimadores tipo 2 del método ORIV-C3C4 que aparecen en las figuras 5.46(a) y 5.46(b).

SNR	Estimadores tipo 1				Estimadores tipo 2			
	b_1	b_2	var b_1	var b_2	b_1	b_2	var b_1	var b_2
c	-1.624	1.374	0	0	-1.624	1.374	0	0
20	-1.424	1.315	0.164	0.227	-1.568	1.320	0.159	0.180
10	-0.694	1.747	0.540	4.295	-1.340	1.045	0.410	0.412
5	-3.681	-3.757	81.486	105.76	-0.039	0.308	0.891	0.634
0	-2.666	0.144	79.413	31.379	0.160	0.170	0.593	0.661
-5	1.050	1.507	28.379	24.850	0.128	0.254	0.739	0.631
-10	-1.007	0.254	11.263	13.111	0.048	0.320	0.859	0.731

Tabla 5.27 Media y varianza de los estimadores (c =valores correctos), modelo SL1. Método ORIV-C2C3

SNR	Estimadores tipo 1				Estimadores tipo 2			
	b_1	b_2	var b_1	var b_2	b_1	b_2	var b_1	var b_2
c	-1.624	1.374	0	0	-1.624	1.374	0	0
20	-1.541	1.340	0.756	0.758	-1.337	1.100	0.501	0.442
10	-1.455	1.307	1.691	1.628	-1.087	0.915	0.643	0.548
5	-1.112	1.100	2.041	1.901	-0.798	0.720	0.664	0.531
0	-0.562	0.585	2.499	2.459	-0.634	0.508	0.606	0.585
-5	-0.304	0.220	1.921	1.992	-0.562	0.227	0.658	0.735
-10	-0.223	0.207	1.462	1.538	-0.611	0.134	0.689	0.775

Tabla 5.28 Media y varianza de los estimadores (c =valores correctos), modelo SL1. Método ORIV-C3C4

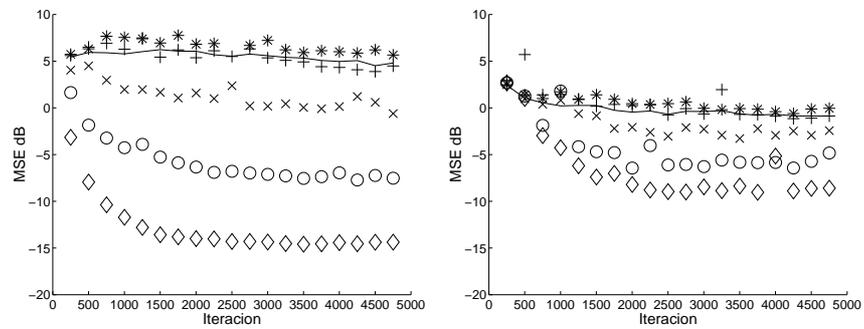
Modelo SL2

Para este modelo la pendiente se ha fijado en $\mathbf{p} = [0 - 1/16000 \ 1/16000]$ y los resultados obtenidos para él se presentan como sigue: la media y varianza de

los estimadores en la última iteración en las tablas 5.29 y 5.30 para los métodos ORIV-C2C3 y ORIV-C3C4 respectivamente; las curvas de aprendizaje en las figuras 5.45(a) y 5.45(b) para ORIV-C2C3 y ORIV-C3C4 respectivamente y finalmente la convergencia en media de los estimadores tipo 2 para ORIV-C3C4 en las figuras 5.47(a) y 5.47(b) para los coeficientes b_1 y b_2 respectivamente.

SNR	Estimadores tipo 1				Estimadores tipo 2			
	b_1	b_2	var b_1	var b_2	b_1	b_2	var b_1	var b_2
c	-1.312	1.062	0	0	-1.312	1.062	0	0
20	-1.160	1.030	0.126	0.142	-1.272	1.026	0.086	0.091
10	-0.544	1.093	0.196	0.254	-1.229	0.982	0.194	0.154
5	-0.005	1.336	1.732	4.592	-0.015	0.622	0.988	0.426
0	0.100	1.508	13.764	12.897	0.133	0.293	0.605	0.624
-5	0.476	-0.316	26.326	16.795	0.131	0.301	0.772	0.616
-10	-0.826	0.257	22.741	9.294	0.058	0.341	0.863	0.718

Tabla 5.29 Media y varianza de los estimadores (c =valores correctos), modelo SL2. Método ORIV-C2C3



(a) ORIV-C2C3

(b) ORIV-C3C4

Figura 5.45 Curvas de aprendizaje del modelo SL2. La tendencia divergente se ha perdido.

5.7.4 Análisis de los resultados obtenidos relacionados con el seguimiento límite

Estas simulaciones completan las presentadas en el apartado 5.7.1. En él, se simuló el mismo modelo que en el modelo SL1 que se trata ahora, pero sólo para una longitud de 2000 datos. Para ese número de iteraciones el algoritmo parecía que había convergido a sus valores estacionarios, y nada hacía pensar que pudiera empezar a divergir. Sin embargo, con el modelo SL1 se demuestra lo contrario: después de llegar a un mínimo en el MSE, los valores empiezan a crecer, en principio indefinidamente. Esta tendencia también se puede observar claramente en las figuras 5.46(a) y 5.46(b), donde se aprecia la divergencia de la media de los estimadores de los parámetros del modelo.

SNR	Estimadores tipo 1				Estimadores tipo 2			
	b_1	b_2	var b_1	var b_2	b_1	b_2	var b_1	var b_2
c	-1.312	1.062	0	0	-1.312	1.062	0	0
20	-1.274	1.028	0.185	0.160	-1.266	0.999	0.210	0.190
10	-1.243	1.010	0.458	0.399	-1.160	0.915	0.415	0.315
5	-1.061	0.849	1.584	1.704	-0.853	0.728	0.576	0.454
0	-0.555	0.533	4.054	4.292	-0.451	0.422	0.727	0.678
-5	-0.361	0.220	2.533	2.073	-0.496	0.116	0.708	0.765
-10	-0.276	0.249	2.258	2.875	-0.600	0.113	0.703	0.811

Tabla 5.30 Media y varianza de los estimadores (c =valores correctos), modelo SL2. Método ORIV-C3C4

Si se tienen en cuenta los comentarios de los apartados precedentes, se pueden clasificar en dos tipos los factores que afectan al comportamiento del algoritmo: los propios coeficientes que controlan el funcionamiento del algoritmo, y que en este caso sólo es el factor de olvido, y los coeficientes que determinan la dinámica del sistema lineal, que para evolución temporal lineal sólo depende de la pendiente elegida. Ya se conocen las restricciones propias impuestas sobre el factor de olvido. Tiene que ser lo suficientemente grande como para poder estimar convenientemente los cumulantes. Por tanto, una vez fijado este factor, queda estudiar hasta qué punto el algoritmo así definido es capaz de seguir la evolución del sistema. Pero también se conoce la respuesta; si el factor de olvido es grande, no podrá seguir cambios temporales muy bruscos, tal y como ocurre aquí. Se tiene por tanto que reducir la pendiente. Eso es lo que se ha hecho en el modelo SL2, donde no se aprecia divergencia en el MSE ni en los estimadores de los parámetros, corroborándose los comentarios anteriores.

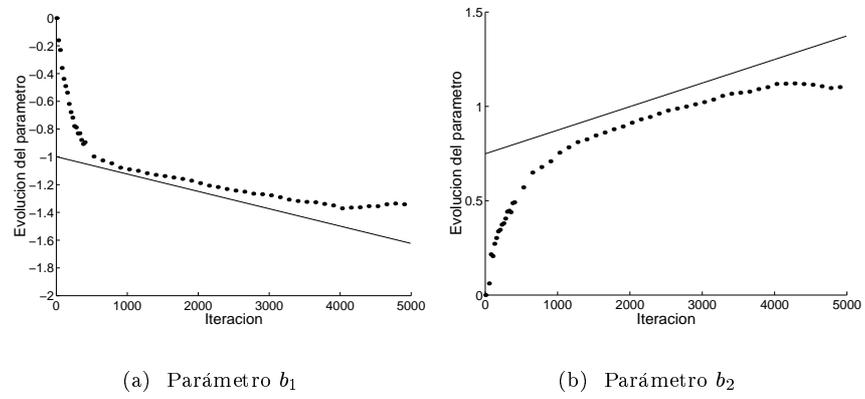


Figura 5.46 Evolución en media de los estimadores tipo 2 del modelo $SL1$ para $ORIV-C3C4$, $SNR=20$

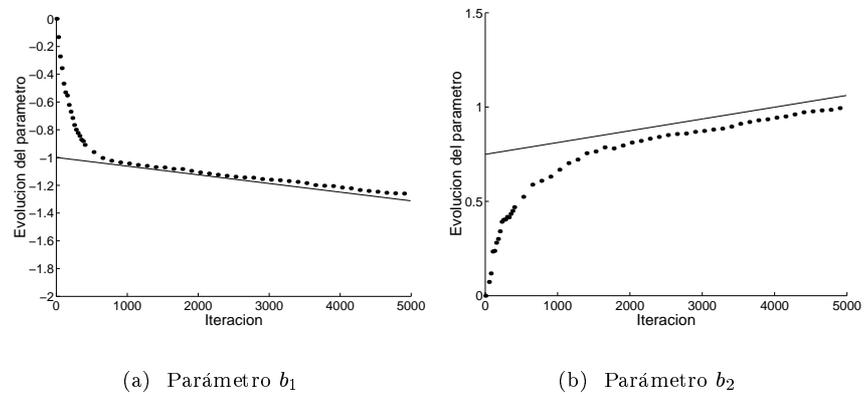


Figura 5.47 Evolución en media de los estimadores tipo 2 del modelo $SL2$ para $ORIV-C3C4$, $SNR=20$

Capítulo 6

Estudio comparativo de los algoritmos RIV y ORIV

Introducción

El estudio teórico sobre el comportamiento del algoritmo ORIV llevado a cabo en el capítulo 4 explicó satisfactoriamente el comportamiento observado en las simulaciones sobre este mismo algoritmo en el capítulo 5. Ya que ORIV es capaz de resolver sistemas de ecuaciones sobredeterminados se puede considerar como una generalización de RIV y en principio debe dar mejores resultados, por lo que debe constituir un tope superior en cuanto a comportamiento.

Este efecto que se espera desde un punto de vista lógico se confirma mediante el empleo de las expresiones analíticas obtenidas en los capítulos 3 y 4 y es confirmado mediante simulaciones tanto en ambientes estacionarios como no estacionarios. La mejora que supone el uso de ORIV frente a RIV requiere el pago de cierto precio: un aumento de la carga computacional. El principal objetivo de este capítulo es pues responder a la pregunta ¿Merece la pena el coste computacional que acarrea el uso de un algoritmo sobredeterminado?

La comparación teórica entre ambos algoritmos se lleva a cabo en el apartado 6.1 donde se analiza la diferencia tanto en ambientes estacionarios como no estacionarios, repasando todos y cada uno de los puntos analizados en capítulos anteriores: convergencia en media, en media cuadrática, contribución al MSD por ruido de medición, etc. La validez de dicho análisis teórico se lleva a cabo en el apartado 6.2 donde de nuevo se distingue entre ambientes estacionarios y no estacionarios y se analizan con detalle tanto la curva de aprendizaje como el

sesgo y varianza de los estimadores y la convergencia en media cuando proceda.

6.1 Comparación teórica

En este primer apartado se van a analizar las diferencias de comportamiento entre los algoritmos RIV y ORIV mediante las expresiones teóricas deducidas en capítulos anteriores. Para ello se intentará recuperar RIV a partir de ORIV para descubrir de manera más profunda las diferencias entre las expresiones obtenidas para uno y otro. Todos los aspectos estudiados apuntan a la misma conclusión, tal y como se describirá a continuación.

6.1.1 Cuadro comparativo

Para dar una visión general de las diferencias entre los algoritmos RIV y ORIV se ha querido mostrar, en la tabla 6.1 de forma confrontada, las expresiones obtenidas para uno y otro para cada uno de los aspectos estudiados: convergencia en media y en media cuadrática para ambientes estacionarios y no estacionarios, con ruido y sin él. Téngase en cuenta que hay magnitudes que comparten el nombre en el análisis de RIV y ORIV, por tanto dependiendo del algoritmo del que se trate tendrán una definición u otra. Tampoco se han indicado todas las definiciones en la tabla sino que desde aquí se remite a los capítulos 3 y 4 donde se estudiaron.

Se ha querido condensar todo el conocimiento que se tiene de ambos algoritmos. Para ello basta dar la evolución temporal de la media y la media cuadrática de los estimadores, en cada uno de los ambientes estudiados, para lo cual basta dividir en ambientes estacionarios y no estacionarios.

Las ecuaciones obtenidas en ambientes estacionarios eran ecuaciones en diferencias de primer orden con coeficientes no homogéneos constantes, con lo cual la solución viene perfectamente caracterizada por las condiciones iniciales $m(0)$, el coeficiente de la homogénea c y el valor en convergencia $m(\infty)$. En el apéndice D se muestra que para una ecuación en diferencias como la descrita:

$$m(n) = cm(n-1) + a \quad (6.1)$$

la solución viene dada por la siguiente expresión:

$$m(n) = m(\infty) + c^n [m(0) - m(\infty)] \quad (6.2)$$

De forma que conociendo la parte homogénea de la ecuación en diferencias y el valor en convergencia de la magnitud estudiada, se puede construir perfectamente la evolución temporal. Allí donde el valor en convergencia sea la suma

	Propiedad	RIV	ORIV
Estacionarios	Conv.media (Efecto cond.ini.)	$E[\Delta \mathbf{w}(n)] = \lambda^n (1 - \lambda) \mathbf{R}^{-1} \Phi(0) \Delta \mathbf{w}(0)$	$E[\Delta \mathbf{w}(n)] = \lambda^{2n} (1 - \lambda)^2 (\mathbf{R}^t \mathbf{R})^{-1} \Gamma(0) \Delta \mathbf{w}(0)$
	MSD ec.homogénea	$q(n) = \lambda^2 q(n - 1)$	$q(n) = \lambda^4 q(n - 1)$
	MSD(∞) medición I	$q(\infty) = \frac{1-\lambda}{2} \sigma_0^2 \beta_Q$	$q(\infty) = \frac{1-\lambda}{4} \sigma_0^2 \text{Tr}\{\beta_Q \mathbf{A}\}$
	MSD(∞) estimación I	$q(\infty) = \frac{(1-\lambda)^2}{2} \sigma_0^2 V_e$	$q(\infty) = \frac{(1-\lambda)^2}{4} [\sigma_0^2 \text{Tr}\{\beta_Q \mathbf{V}_R\} + \text{Tr}\{\beta_Q \mathbf{V}_X\}]$
	MSD(∞) medición II	$q(\infty) = \frac{1-\lambda}{2} \sigma_0^2 \beta_Q$	$q(\infty) = \frac{1-\lambda}{4} \sigma_0^2 \tilde{\gamma}$
	MSD(∞) estimación II	$q(\infty) = \frac{(1-\lambda)^2}{2} \sigma_0^2 V_e$	$q(\infty) = \frac{(1-\lambda)^2}{4} [\sigma_0^2 V_e + \xi_e]$
No estacionaria.	Conv.media(Sol. particular)	$E[\mathbf{w}(n)] = (1 - \lambda) \mathbf{w}_0(n) + (1 - \lambda) \sum_{i=1}^n \lambda^i \mathbf{w}_0(n - i)$	$E[\mathbf{w}(n)] = (1 - \lambda) \mathbf{w}_0(n) + \frac{(1-\lambda^2)}{\lambda} \sum_{i=1}^n \lambda^{2i} \mathbf{w}_0(n - i)$
	tiempo de retraso	$\tau_r = \lambda / (1 - \lambda)$	$\tau_r = \lambda / (1 - \lambda^2)$
	MSD(∞) rastreo	$q(\infty) = \frac{1}{(1-\lambda)^2} \mathbf{p}^t \mathbf{p} - \frac{1}{2} \frac{1}{1-\lambda} \mathbf{p}^t \mathbf{p} + 2 \mathbf{p}^t \mathbf{p} + \frac{1-\lambda}{2} \mathbf{p}^t \mathbf{H} \mathbf{p}$	$D(\infty) = \frac{1}{4(1-\lambda)^2} \mathbf{p}^t \mathbf{p} - \frac{1}{8} \mathbf{p}^t \mathbf{p} + \frac{1-\lambda}{4} \mathbf{p}^t \mathbf{H} \mathbf{p}$
	λ_{opt}	$1 - [4 \mathbf{p}^t \mathbf{p} / (\sigma_0^2 \beta_Q + \mathbf{p}^t \mathbf{H} \mathbf{p})]^{1/3}$	$1 - [2 \mathbf{p}^t \mathbf{p} / (\sigma_0^2 \tilde{\gamma} + \mathbf{p}^t \mathbf{H} \mathbf{p})]^{1/3}$

Tabla 6.1 Cuadro comparativo RIV-ORIV

de varias contribuciones, éstas se han especificado por separado (ejemplo: MSD debido al ruido de estimación o de medición).

En ambientes no estacionarios no es tan sencillo caracterizar la evolución de las magnitudes, ya que en este caso los coeficientes de la parte no homogénea dependen del tiempo y habría que conocerlos en todo los instantes. Es decir, en este caso la ecuación en diferencias sería:

$$m(n) = cm(n-1) + a(n) \quad (6.3)$$

y la solución ya no se puede escribir en la forma de la expresión (6.2). La solución a este tipo de ecuaciones se puede consultar igualmente en el apéndice D pero para el estudio a llevar a cabo en esta sección no es necesaria. En la tabla 6.1, en el caso de ambientes no estacionarios, sólo se muestran las magnitudes que van a ser de utilidad.

6.1.2 Análisis teórico de cada faceta. Ambientes estacionarios

Las diferencias entre RIV y ORIV en cuanto a evolución teórica de las magnitudes que determinan su comportamiento son:

Convergencia en media. Efecto de las condiciones iniciales

Al estudiar la convergencia en media se obtuvo para el algoritmo RIV la siguiente expresión:

$$E[\Delta \mathbf{w}(n)] = \lambda^n (1 - \lambda) \mathbf{R}^{-1} \Phi(0) \Delta \mathbf{w}(0) \quad (6.4)$$

y para ORIV:

$$E[\Delta \mathbf{w}(n)] = \lambda^{2n} (1 - \lambda)^2 (\mathbf{R}^t \mathbf{R})^{-1} \Gamma(0) \Delta \mathbf{w}(0) \quad (6.5)$$

Es conocido que tanto una como otra expresión tienden a cero en el límite $n \rightarrow \infty$ por tanto lo que importa es determinar cuál de los dos converge más rápido. El tiempo de convergencia, como ya es costumbre, se tomará como aquel necesario para que la magnitud estudiada, en este caso la media del vector de error, caiga a $1/e$ de su valor inicial. Operando como ya es habitual se obtiene, para RIV:

$$\tau = \frac{1}{1 - \lambda} \quad (6.6)$$

y para ORIV:

$$\tau = \frac{1}{2(1 - \lambda)} \quad (6.7)$$

Por tanto, el tiempo de convergencia para ORIV es la mitad que para RIV.

Ecuación homogénea asociada a la evolución del MSD

Como se comentó en el apéndice D la solución de una ecuación homogénea tiende a cero bajo cierta condición. Una vez que llega a cero, la parte transitoria de la evolución de la magnitud estudiada se termina y empieza el estado estacionario caracterizado por el valor en convergencia. Conocer por tanto la homogénea asociada es importante para determinar la velocidad con la que se alcanza el estado estacionario (steady state).

Para RIV es:

$$q(n) = \lambda^2 q(n-1) \quad (6.8)$$

y para ORIV:

$$q(n) = \lambda^4 q(n-1) \quad (6.9)$$

Al igual que antes se pueden calcular los tiempos de convergencia asociados: Para RIV:

$$\tau = \frac{1}{2(1-\lambda)} \quad (6.10)$$

Y para ORIV:

$$\tau = \frac{1}{4(1-\lambda)} \quad (6.11)$$

Estos tiempos indican el tiempo que tiene que pasar para que la solución de la homogénea sea menos importante que en las primeras iteraciones, de manera que la dinámica la dicta la solución particular de la no homogénea, que como su nombre indica es característica (particular) de cada algoritmo. Notar que ahora también el de ORIV es la mitad que el de RIV. A continuación se estudian los términos no homogéneos.

Término debido al ruido de medición en el MSD

La primera contribución al MSD que se estudiará es debida al ruido de medición. Recuérdese que esta contribución se obtuvo mediante dos aproximaciones diferentes en capítulos anteriores y que los resultados mediante ambas eran idénticas para RIV. Las expresiones obtenidas fueron: Para RIV (aproximaciones I y II, usadas en los apartados 3.2.2 y 3.2.3 respectivamente):

$$q(\infty) = \frac{1-\lambda}{2} \sigma_0^2 \beta_Q \quad (6.12)$$

para ORIV aproximación I (apartado 4.2.2):

$$q(\infty) = \frac{1-\lambda}{4} \sigma_0^2 \text{Tr}\{\beta_Q \mathbf{A}\} \quad (6.13)$$

y aproximación II (apartado 4.2.4):

$$q(\infty) = \frac{1-\lambda}{4} \sigma_0^2 \tilde{\gamma} \quad (6.14)$$

En primer lugar notar que recordando las siguientes definiciones:

$$\beta_Q = (\mathbf{R}^t \mathbf{R})^{-t} \mathbf{Q} (\mathbf{R}^t \mathbf{R})^{-1} \quad (6.15)$$

$$\mathbf{A} = \mathbf{R}^t \mathbf{R}_{\bar{x}\bar{x}} \mathbf{R} \quad (6.16)$$

$$\tilde{\gamma} = \text{Tr}\{\mathbf{R}(\mathbf{R}^t \mathbf{R})^{-2} \mathbf{R}^t \mathbf{R}_{\bar{x}\bar{x}}\} \quad (6.17)$$

se tiene que $\tilde{\gamma} = \text{Tr}\{\beta_Q \mathbf{A}\}$ con $\mathbf{Q} = \mathbf{I}$. Por tanto resulta que ambas aproximaciones son también idénticas para ORIV. Por tanto, para comparar RIV y ORIV se pueden comparar, por ejemplo, las expresiones (6.12) y (6.14).

ORIV por construcción es un algoritmo que generaliza a RIV para sistemas de ecuaciones sobredeterminados. Para ver mejor el motivo de la diferencia entre uno y otro es conveniente intentar reobtener la expresión (6.12) a partir de (6.14). En primer lugar si la matriz \mathbf{R} que aparece en el análisis de ORIV fuese cuadrada¹, su inversa existiría, y se podría escribir que:

$$\begin{aligned} \tilde{\gamma} &= \text{Tr}\{\mathbf{R}(\mathbf{R}^t \mathbf{R})^{-1} (\mathbf{R}^t \mathbf{R})^{-1} \mathbf{R}^t \mathbf{R}_{\bar{x}\bar{x}}\} = \\ &= \text{Tr}\{\mathbf{R}^{-t} \mathbf{R}^{-1} \mathbf{R}_{\bar{x}\bar{x}}\} \longleftrightarrow \beta_Q \text{ con } \mathbf{Q} = \mathbf{I} \end{aligned} \quad (6.18)$$

En este límite coincidirían las expresiones (6.12) y (6.14) salvo por un factor 1/2. Ver el origen de este factor necesita un proceso más elaborado:

Primero es necesario recordar que provienen de las expresiones (3.42) y (4.69) al calcular sus respectivos valores en convergencia y simplificar en los denominadores, para $\lambda \sim 1$, las expresiones:

$$\text{Para RIV: } 1 - \lambda^2 \sim 2(1 - \lambda)$$

$$\text{Para ORIV: } 1 - \lambda^4 = (1 + \lambda^2)(1 + \lambda)(1 - \lambda) \sim 4(1 - \lambda)$$

Por tanto, que aparezca un 2 o un 4 en el denominador depende de la ecuación homogénea asociada a las ecuaciones (3.42) y (4.69):

$$\text{Para RIV: } q(n) = \lambda^2 q(n-1)$$

$$\text{Para ORIV: } D(n) = \lambda^4 D(n-1)$$

¹La diferencia entre la matriz \mathbf{R} que aparece en RIV y la que aparece en ORIV no es más que las dimensiones, ya que ambas son la correlación cruzada de los mismos vectores salvo con el vector de variable instrumental con distinta longitud en cada caso; se empleará la misma notación para ambas salvo cuando haya lugar a confusión.

Segundo, esta forma de las ecuaciones en diferencias para $q(n)$ depende de la ecuación homogénea asociada a la evolución del vector de error en los pesos $\Delta \mathbf{w}(n)$:

Para RIV: $\Delta \mathbf{w}(n) = \lambda \Delta \mathbf{w}(n-1)$

Para ORIV: $\Delta \mathbf{w}(n) = \lambda^2 \Delta \mathbf{w}(n-1)$

que son consecuencia de aplicar promediado directo a las ecuaciones en diferencias originales para $\Delta \mathbf{w}(n)$, ecuación (3.31) para RIV y ecuación (4.63) para ORIV.

Tercero, estas ecuaciones en diferencias originales se obtuvieron partiendo de la ecuación de actualización del vector incógnita, ecuación (3.29) para RIV y ecuación (4.61) para ORIV y teniendo en cuenta que los respectivos errores se podían reescribir como sigue:

Para RIV: $e(n) = e_0(n) - \mathbf{x}^t(n)[\mathbf{w}(n-1) - \mathbf{w}_0]$

Para ORIV: $\boldsymbol{\alpha}(n) = \begin{pmatrix} 0 \\ e_0(n) \end{pmatrix} + \begin{pmatrix} \tilde{\mathbf{x}}^t(n)[\mathbf{z}(n-1) - \Phi(n-1)\mathbf{w}(n-1)] \\ \mathbf{x}^t(n)[\mathbf{w}_0 - \mathbf{w}(n-1)] \end{pmatrix}$

La diferencia entre ambos está en que para RIV se cumple exactamente que $\Phi(n)\mathbf{w}(n) = \mathbf{z}(n)$, pero no para ORIV en el que $\|\Phi(n)\mathbf{w}(n) - \mathbf{z}(n)\|$ es mínima, ya que se resuelve mediante mínimos cuadrados. Por tanto, la diferencia

$$\Phi(n)\mathbf{w}(n) - \mathbf{z}(n)$$

que aparece en $\boldsymbol{\alpha}(n)$ es una contribución más al error en ORIV.

Cuarto, para recuperar RIV a partir de ORIV hay que suponer que este nuevo error es cero, por lo que la expresión (4.63) se convierte en:

$$\begin{aligned} \Delta \mathbf{w}(n) = [\mathbf{I} - \Gamma^{-1}(n)\Phi^t(n)\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)]\Delta \mathbf{w}(n-1) + \\ + \Gamma^{-1}(n)\Phi^t(n)\tilde{\mathbf{x}}(n)e_0(n) \end{aligned} \quad (6.19)$$

y ahora suponer que la matriz $\Phi(n)$ es cuadrada y que por tanto $\Gamma^{-1}(n) = \Phi^{-1}(n)\Phi^{-t}(n)$, con lo que la expresión anterior queda como:

$$\begin{aligned} \Delta \mathbf{w}(n) = [\mathbf{I} - \Phi^{-1}(n)\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)]\Delta \mathbf{w}(n-1) + \\ \Phi^{-1}(n)\tilde{\mathbf{x}}(n)e_0(n) \end{aligned} \quad (6.20)$$

que efectivamente es la ecuación en diferencias de partida (3.31) para RIV.

Por tanto, que aparezca un λ^2 como coeficiente en la ecuación homogénea asociada en ORIV (tras aplicar promediado directo) y no un λ , es debido a

la componente adicional del error. Esto hace que converja más rápido (porque cuenta con más información) pero además que haya más términos no homogéneos.

Una vez aclarado el origen de ese factor $1/2$ extra hay que plantear la pregunta fundamental: ¿Cuál de las expresiones (6.12) y (6.14) da un $q(\infty)$ menor?. Como a grosso modo tanto β_Q como $\tilde{\gamma}$ son de la forma $\mathbf{R}_{\tilde{x}\tilde{x}}/\mathbf{R}^2$, la principal diferencia al tomar la traza va a ser la dimensión de las matrices. Supóngase que hay q incógnitas y l es la longitud del vector de variable instrumental para ORIV. La aproximación siguiente está basada en suponer que la traza se toma a matrices diagonales múltiplo de la identidad:

$$\beta_Q \sim qb \quad (6.21)$$

$$\tilde{\gamma} \sim lb \quad (6.22)$$

con lo que los valores en convergencia para el MSD se pueden escribir como:

$$\text{Para RIV: } q(\infty) = \frac{1-\lambda}{2}qb$$

$$\text{Para ORIV: } q(\infty) = \frac{1-\lambda}{4}lb$$

Nótese que la definición de β_Q para RIV está en la ecuación (6.18). A partir de los cuales se puede deducir, bajo esta hipótesis tan burda, que el MSD en convergencia para ORIV será menor que para RIV siempre y cuando $2q > l$, es decir, si el número de ecuaciones no dobla al número de incógnitas.

Término debido al ruido de estimación

Las expresiones que se obtuvieron para estos términos para cada algoritmo fueron:

Para RIV (aproximaciones I y II, usadas en los apartados 3.2.2 y 3.2.3 respectivamente):

$$q(\infty) = \frac{(1-\lambda)^3}{2}\sigma_0^2V_e$$

Para ORIV (aproximación I, usada en el apartado 4.2.3):

$$q(\infty) = \frac{(1-\lambda)^3}{4}\sigma_0^2V_e + \frac{(1-\lambda)^3}{4}\xi_e$$

Para ORIV (aproximación II, usada en el apartado 4.2.4):

$$q(\infty) = \frac{(1-\lambda)^3}{4}\sigma_0^2\text{Tr}\{\beta_Q\mathbf{V}_R\} + \frac{(1-\lambda)^3}{4}\text{Tr}\{\beta_Q\mathbf{V}_X\}$$

Debido al error de mínimos cuadrados aparecen, como ya se ha comentado, nuevos términos no homogéneos en la ecuación en diferencias para $\Delta \mathbf{w}(n)$ que acarrea nuevos términos no homogéneos en $q(n)$ y por tanto contribuciones nuevas en $q(\infty)$. Por esto, en principio ORIV debe tener más $q(\infty)$ causado por el ruido de estimación pero como la contribución al $q(\infty)$ total es del orden de $(1 - \lambda)^3$ puede que no influya mucho y se siga cumpliendo que el MSD en convergencia para ORIV es menor que para RIV.

La diferencia entre el factor $1/2$ y $1/4$ se explica como antes.

6.1.3 Análisis teórico de cada faceta. Ambientes no estacionarios

A continuación se extienden los comentarios realizados a ambientes no estacionarios.

Convergencia en media. Solución general de la homogénea

A diferencia de lo que ocurría para ambientes estacionarios la ecuación en diferencias que rige la convergencia en media es no homogénea. La ecuación homogénea asociada a cada algoritmo es:

$$\text{Para RIV: } E[\mathbf{w}(n)] = \lambda^n E[\mathbf{w}(0)]$$

$$\text{Para ORIV: } E[\mathbf{w}(n)] = \lambda^{2n} E[\mathbf{w}(0)]$$

Como se sabe la información que se puede obtener de aquí es el tiempo necesario para alcanzar el estado estacionario. Repitiendo cálculos similares a los realizados con anterioridad se obtienen los respectivos tiempos de convergencia:

$$\text{Para RIV: } \tau = \frac{1}{1-\lambda}$$

$$\text{Para ORIV: } \tau = \frac{1}{2(1-\lambda)}$$

De nuevo el menor tiempo de convergencia para ORIV se debe al error de mínimos cuadrados, que incorpora mayor información en el error a priori $\alpha(n)$ que en $e(n)$.

Convergencia en media. Solución particular de la no homogénea

Las expresiones de interés son, en este caso:

Para RIV: $E[\mathbf{w}(n)] = (1 - \lambda)\mathbf{w}_0(n) + (1 - \lambda) \sum_{i=1}^n \lambda^i \mathbf{w}_0(n - i)$

Para ORIV: $E[\mathbf{w}(n)] = (1 - \lambda)\mathbf{w}_0(n) + \frac{(1-\lambda^2)}{\lambda} \sum_{i=1}^n \lambda^{2i} \mathbf{w}_0(n - i)$

La contribución a la media de los estimadores por parte de $\mathbf{w}_0(n)$ es la misma para cada algoritmo, pero no así las contribuciones debidas a instantes anteriores $\mathbf{w}_0(n - i)$ $i > 0$. Una forma de comprobar para cuál de los dos esta segunda contribución (indeseada) es más importante es calculando el tiempo de retardo para alguna situación particular sencilla.

Este tiempo de retardo ya se calculó para el caso en el que las incógnitas verdaderas cambiaran con el tiempo según una evolución temporal lineal. Teniendo esto en cuenta las expresiones anteriores, se obtienen unos tiempos de retraso dados por:

Para RIV: $\tau = \frac{\lambda}{1-\lambda}$

Para ORIV: $\tau = \frac{\lambda}{1-\lambda^2}$

Como el tiempo de retraso para ORIV es menor que para RIV (siempre y cuando $\lambda < 1$), la influencia de pesos a tiempos anteriores al actual n son menores en el primero que en el segundo.

MSD en convergencia debido a la no estacionariedad

La última contribución que faltaba por analizar del $\text{MSD}(\infty)$ es la siguiente:

Para RIV: $q(\infty) = \frac{1}{(1-\lambda)^2} \mathbf{p}^t \mathbf{p} - \frac{5}{2} \frac{1}{1-\lambda} \mathbf{p}^t \mathbf{p} + 2 \mathbf{p}^t \mathbf{p} + \frac{1-\lambda}{2} \mathbf{p}^t \mathbf{H} \mathbf{p}$

Para ORIV: $D(\infty) = \frac{1}{4(1-\lambda)^2} \mathbf{p}^t \mathbf{p} - \frac{7}{8} \mathbf{p}^t \mathbf{p} + \frac{1-\lambda}{4} \mathbf{p}^t \mathbf{H} \mathbf{p}$

Recuérdese que la definición de la matriz \mathbf{H} difería en uno y otro algoritmo. Para RIV era de la forma:

$$\mathbf{H} = 2(1 - \lambda) \lim_{n \rightarrow \infty} \sum_{i=0}^{n-1} \lambda^{2i} \mathbf{H}(n - i) \quad (6.23)$$

donde

$$\mathbf{H}(n) = E[\mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \mathbf{R}^{-t}(n) \mathbf{R}^{-1}(n) \tilde{\mathbf{x}}(n) \mathbf{x}^t(n)] \quad (6.24)$$

y para ORIV era de la forma:

$$\mathbf{H} = 4(1 - \lambda) \lim_{n \rightarrow \infty} \sum_{i=0}^{n-1} \lambda^{4i} \mathbf{H}(n - i) \quad (6.25)$$

donde

$$\mathbf{H}(n) = E[\mathbf{x}(n)\tilde{\mathbf{x}}^t(n)\mathbf{R}(n)(\mathbf{R}^t(n)\mathbf{R}(n))^{-2}\mathbf{R}^t(n)\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)] \quad (6.26)$$

En primer lugar, para estudiar la equivalencia RIV-ORIV téngase en cuenta que si la matriz \mathbf{R} es cuadrada, la matriz $\mathbf{H}(n)$ coincide tanto en su versión RIV como ORIV. Para hacer coincidir las definiciones de \mathbf{H} es necesario recurrir a la regla de sustituir los factores 2 por 4 y los λ^2 por λ^4 al pasar de RIV a ORIV.

De esta manera se logra perfecta equivalencia para los términos que dependen de \mathbf{H} , el resto de términos proviene de una simplificación más laboriosa para $\lambda \sim 1$.

Aunque no es algo propio de la comparación RIV-ORIV, conviene hacer notar aquí que para incluir el resto de contribuciones a $\text{MSD}(\infty)$ (es decir, por ruido de medición y de estimación) en ambientes no estacionarios, hay que tener en cuenta la dependencia temporal de las magnitudes y darles un tratamiento similar al dado aquí a $\mathbf{H}(n)$ para reexpresarla como \mathbf{H} . Se pueden tener en cuenta las mismas expresiones que para ambientes estacionarios pero cambiando, por ejemplo, $\tilde{\gamma}(n)$ (definida por la expresión (4.92)) por

$$\tilde{\gamma} = 4(1 - \lambda) \lim_{n \rightarrow \infty} \sum_{i=0}^{n-1} \lambda^{4i} \tilde{\gamma}(n - i) \quad (6.27)$$

ya definida en (4.96). Nótese que hay que tener en cuenta si la magnitud está asociada a RIV o a ORIV para incluir los factores y exponentes adecuados.

Asimismo téngase en cuenta que las distintas contribuciones a $\text{MSD}(\infty)$ tienen la misma dependencia en $(1 - \lambda)$ para RIV que para ORIV, por lo que dado el factor 4 en el denominador de los sumandos para ORIV contra el factor 2 para RIV, es de esperar que el MSD debido a rastreo sea también menor para ORIV que para RIV.

Factor de olvido óptimo

Para terminar con el análisis teórico se analizará el factor de olvido óptimo para cada algoritmo:

$$\text{Para RIV: } \lambda_{opt} = 1 - \left(\frac{4\mathbf{p}^t \mathbf{p}}{\sigma_0^2 \beta_Q + \mathbf{p}^t \mathbf{H} \mathbf{p}} \right)^{1/3}$$

$$\text{Para ORIV: } \lambda_{opt} = 1 - \left(\frac{2\mathbf{p}^t \mathbf{p}}{\sigma_0^2 \tilde{\gamma} + \mathbf{p}^t \mathbf{H} \mathbf{p}} \right)^{1/3}$$

Por simplificar se supondrá de nuevo que $\beta_Q \sim qb$ y que $\tilde{\gamma} \sim lb$, y se discutirán dos casos:

1. Si el error de medición es muy grande:

$$\text{Para RIV: } \lambda_{opt} \sim 1 - \left(\frac{4\mathbf{p}^t \mathbf{p}}{\sigma_0^2 qb} \right)^{1/3} \quad (6.28)$$

$$\text{Para ORIV: } \lambda_{opt} \sim 1 - \left(\frac{2\mathbf{p}^t \mathbf{p}}{\sigma_0^2 lb} \right)^{1/3} \quad (6.29)$$

Y por tanto resulta que el factor de olvido óptimo para RIV es menor que para ORIV, es decir, necesita menos memoria para operar de forma que dé un MSD en convergencia mínimo.

2. Si la variación lineal es muy grande:

$$\text{Para RIV: } \lambda_{opt} \sim 1 - \left(\frac{4\mathbf{p}^t \mathbf{p}}{\mathbf{p}^t \mathbf{H} \mathbf{p}} \right)^{1/3} \quad (6.30)$$

$$\text{Para ORIV: } \lambda_{opt} \sim 1 - \left(\frac{2\mathbf{p}^t \mathbf{p}}{\mathbf{p}^t \mathbf{H} \mathbf{p}} \right)^{1/3} \quad (6.31)$$

Y también resulta que el factor de olvido óptimo para RIV es menor que para ORIV.

En conclusión, ORIV siempre necesita un λ mayor para trabajar en condiciones óptimas.

6.1.4 Conclusiones del análisis teórico

A la vista de todo lo discutido en este capítulo hasta ahora es de esperar que ORIV converja antes en media y en media cuadrática, con un valor de $q(\infty)$ compuesto de ruido de estimación y medición menor (aunque en particular el de estimación sea mayor.). Además presenta menor tiempo de retraso y menor $q(\infty)$ debido a la no estacionariedad.

En definitiva, para λ fijo ORIV es mejor. Sin embargo sería aún conveniente comparar el $q_{min}(\infty)$ de ambos algoritmos. Para lo cual es conveniente considerar las siguientes expresiones como representativas del $q(\infty)$ para cada algoritmo:

$$\text{Para RIV: } q(\infty) = \frac{1}{(1-\lambda)^2} \mathbf{p}^t \mathbf{p} + \frac{1-\lambda}{2} \mathbf{p}^t \mathbf{H} \mathbf{p} + \frac{1-\lambda}{2} \sigma_0^2 \beta_Q.$$

$$\text{Para ORIV: } q(\infty) = \frac{1}{4(1-\lambda)^2} \mathbf{p}^t \mathbf{p} + \frac{1-\lambda}{4} \mathbf{p}^t \mathbf{H} \mathbf{p} + \frac{1-\lambda}{4} \sigma_0^2 \beta_Q.$$

En ellas se ha eliminado la contribución del ruido de estimación y se han dejado la más influyente del ruido debido a rastreo junto con la debido al ruido de

medición, (y otra del mismo orden debido al ruido de rastreo por mantener el equilibrio en la aproximación).

Si en ellas se sustituye el valor de λ por el de λ_{opt} , se obtiene el MSD en convergencia mínimo:

$$\text{Para RIV: } q_{min}(\infty) = 2^{2/3} \frac{3}{4} (\sigma_0^2 \beta_Q + \mathbf{p}^t \mathbf{H} \mathbf{p})^{2/3} (\mathbf{p}^t \mathbf{p})^{1/3} .$$

$$\text{Para ORIV: } q_{min}(\infty) = 2^{1/3} \frac{3}{8} (\sigma_0^2 \tilde{\gamma} + \mathbf{p}^t \mathbf{H} \mathbf{p})^{2/3} (\mathbf{p}^t \mathbf{p})^{1/3}$$

Para estudiar las condiciones bajo las cuales el algoritmo ORIV dará un MSD en convergencia óptimo $q_{min}^{(O)}(\infty)$ menor que RIV, se supondrá que σ_0^2 es suficientemente grande y de nuevo que $\beta_Q \sim qb$ y que $\tilde{\gamma} \sim lb$. Con esto se tiene que:

$$\frac{q_{min}^{(R)}}{q_{min}^{(O)}} \sim 2 \cdot 2^{1/3} \left(\frac{q}{l} \right)^{2/3} \quad (6.32)$$

Y por tanto $q_{min}^{(O)}$ será menor que el equivalente para RIV si $2^2 q > l$, es decir, mientras que el número de ecuaciones sea menor que 4 veces el número de incógnitas.

Que $q_{min}(\infty)$ sea menor para ORIV que para RIV es consecuencia particular de lo que se afirmaba con anterioridad de que ORIV, a λ fijo, da menor $q(\infty)$ que RIV.

Todas las ventajas con las que cuenta ORIV sobre RIV se deben, o así se ha de suponer, a que cuenta con mayor información. Esta mayor información hace que converja antes, aunque con más ruido de estimación; por suerte este ruido de estimación es dos órdenes inferior que el ruido de medición y por tanto no afecta muy negativamente.

Como complemento, en el apéndice E se muestra que tras la eliminación de esa información adicional, lo que aquí se ha llamado error de mínimos cuadrados, y tras considerar sistemas con matrices cuadradas, el algoritmo RIV se puede recuperar del ORIV.

6.2 Comparación experimental

La comparación teórica llevada a cabo en el apartado anterior ha predicho un superior comportamiento de ORIV respecto al de RIV. Con el fin de comprobar experimentalmente este extremo, en este apartado se llevarán a cabo simulaciones para la identificación de sistemas tanto en ambientes estacionarios

como no estacionarios. Por un lado se ha construido de nuevo el método ORIV-C3C4 y por otro se ha procedido de forma análoga con RIV para construir el método RIV-C3C4. Recuérdese que el primero no es más que ORIV aplicado a la resolución de la ecuación (5.18).

6.2.1 Comparación en ambientes estacionarios

El sistema identificado es un MA de orden 2, denominado ERO, con coeficientes $MA = [1 \quad -1.1314 \quad 0.6400]$, bajo las mismas condiciones que en simulaciones anteriores: misma señal de excitación, misma forma de generar el ruido aditivo gaussiano, etc., salvo que ahora las series tienen 30000 muestras. Las relaciones señal-ruido estudiadas son 20, 10, 5, 0 y -5 dB.

Los resultados obtenidos se recogen en las gráficas de aprendizaje, en las figuras 6.1(b) y 6.1(a) para ORIV-C3C4 y RIV-C3C4 respectivamente, y en las tablas que muestran la media y la varianza de los estimadores obtenidos en la última iteración, tablas 6.3 y 6.2 respectivamente. De ellos se puede obtener la siguiente información:

Puntos a destacar a partir de la curva de aprendizaje

Como se aprecia claramente en las gráficas correspondientes, el método ORIV-C3C4 converge bastante más rápido que el RIV-C3C4 y además lo hace a valores más bajos de MSD. Esto corrobora lo deducido con las expresiones teóricas, aunque quizás la diferencia entre uno y otro método sea mayor de lo esperado: en este caso concreto ORIV-C3C4 cuenta con 2 ecuaciones más que RIV-C3C4 que cuenta con 5 ecuaciones, teniendo ambos 5 incógnitas. Estas 2 ecuaciones extra y, como se dijo en teoría, la forma de resolverlo por mínimos cuadrados mejoran enormemente los resultados.

También se aprecia como cuanto menor es la SNR mayor diferencia hay entre uno y otro método, siempre a favor de ORIV-C3C4. La MSD tan alta de RIV-C3C4, ¿será como consecuencia del sesgo o de la varianza de los estimadores?

Puntos a destacar a partir del sesgo y varianza de los estimadores

Como se puede apreciar de las tablas el gran MSD para RIV-C3C4 se debe al sesgo en el estimador del segundo parámetro y a la varianza de ambos parámetros, tanto en los estimadores tipo 1 como tipo 2. Aunque quizás lo que más destaque sea la gran varianza observada.

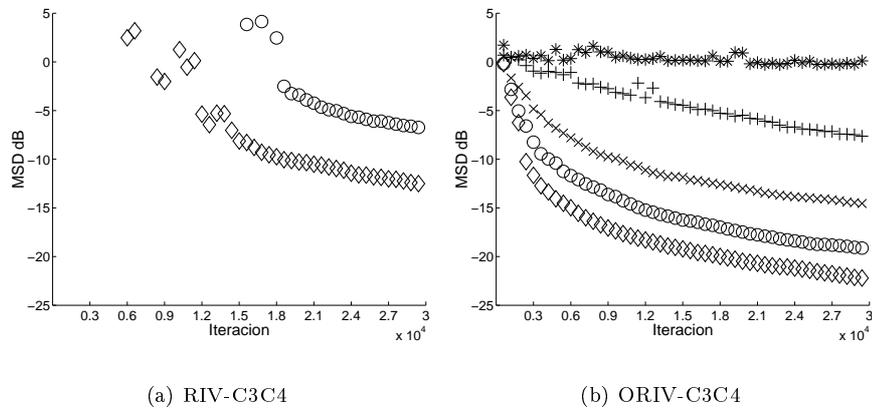


Figura 6.1 Curvas de aprendizaje. Modelo ERO. $\diamond = 20$ dB, $O = 10$ dB, $\times = 5$ dB, $+ = 0$ dB y $* = -5$ dB.

Según apuntaba el MSD, para 5 dB RIV-C3C4 ya no da ningún tipo de información.

SNR	Estimadores tipo 1				Estimadores tipo 2			
	b_1	b_2	var b_1	var b_2	b_1	b_2	var b_1	var b_2
20	-1.1267	0.4085	0.0689	0.4875	-1.1341	0.5167	0.0628	0.4024
10	-1.1278	0.1188	0.1292	0.6525	-1.1272	0.3698	0.1170	0.6631
5	-0.7720	0.2405	1.2512	1.2056	-0.9156	0.1579	0.7905	1.1239
0	-0.3877	0.0982	1.4087	1.4031	-0.9718	0.7356	1.2253	1.3947
-5	-0.0129	0.2192	1.4803	1.6341	-0.4786	0.9939	1.5628	1.1496

Tabla 6.2 Media y varianza de los estimadores de RIV-C3C4. Modelo ERO.

SNR	Estimadores tipo 1				Estimadores tipo 2			
	b_1	b_2	var b_1	var b_2	b_1	b_2	var b_1	var b_2
20	-1.1313	0.6384	0.0160	0.0278	-1.1328	0.6363	0.0229	0.0539
10	-1.1311	0.6303	0.0232	0.1037	-1.1330	0.6236	0.0329	0.1116
5	-1.1298	0.5173	0.0371	0.3640	-1.1315	0.5606	0.0566	0.3033
0	-1.0526	0.3200	0.2936	0.5443	-1.0836	0.5807	0.1475	0.4924
-5	-0.4523	0.2956	1.7727	1.6404	-0.6078	0.8320	0.5773	0.4734

Tabla 6.3 Media y varianza de los estimadores de ORIV-C3C4. Modelo ERO.

6.2.2 Comparación en ambientes no estacionarios

En este caso el sistema a identificar es también un MA de orden 2, denominado NERO, pero con parámetros iniciales $MA = [1 \quad -1 \quad 0.75]$, que sufre una evolución temporal lineal con pendiente dada por $\mathbf{p} = [0 \quad -1/16000 \quad 1/16000]$. La longitud de las series estudiadas es de 5000 datos no estando contaminadas de ruido aditivo. El resto de parámetros de la simulación se fijan de forma análoga a las ya realizadas con anterioridad.

La información obtenida ahora se presenta en las curvas de aprendizaje, figuras 6.7 y 6.6 para ORIV-C3C4 y RIV-C3C4 respectivamente, en las curvas de evolución de la media de los estimadores, figuras 6.8- 6.11 para ORIV-C3C4 y 6.2-6.5 para RIV-C3C4, y en la tabla 6.4 que recoge el sesgo y la varianza de los estimadores en la última iteración. De toda esta información se puede deducir lo siguiente:

Puntos a destacar de la curva de aprendizaje

Igual que sucedía en ambientes estacionarios, ORIV-C3C4 converge antes y a un MSD menor. El comportamiento para RIV-C3C4 es muy irregular y acaba con una clara tendencia divergente; como se comprobará al analizar la media y la varianza, esto se debe a la gran varianza de los estimadores. Hay que destacar que en general, en todas las representaciones gráficas, RIV-C3C4 presenta un comportamiento muy irregular y ruidoso.

Puntos a destacar de las gráficas de la evolución de la media

Tal y como se esperaba del análisis teórico ORIV-C3C4 converge al estado estacionario (steady state) antes que RIV-C3C4. Sin embargo ORIV-C3C4 tiene menor tiempo de retraso para los estimadores tipo 1 pero mayor para los tipo 2. En particular la media de los estimadores tipo 2 de RIV-C3C4 reproduce casi exactamente el comportamiento real de los parámetros.

No se aprecia ni la menor tendencia divergente por lo que el motivo de la divergencia del MSD debe ser un aumento progresivo de la varianza.

Puntos a destacar de las tablas de media y varianza

Como ya se ha comentado los estimadores no presentan sesgo apreciable, sin embargo la varianza de los estimadores tipo 1 es un tanto elevada y la del tipo 2 ya es excesiva para RIV. Esto último debe ser la causa del alto MSD.

	Estimadores tipo 1				Estimadores tipo 2			
	b_1	b_2	var b_1	var b_2	b_1	b_2	var b_1	var b_2
C	-1.3118	1.0618	0.0000	0.0000	-1.3118	1.0618	0.0000	0.0000
O	-1.2794	1.0297	0.0897	0.0987	-1.2882	1.0147	0.1441	0.1398
R	-1.2691	1.0030	0.1647	0.2406	-1.3062	1.0331	0.5025	0.4849

Tabla 6.4 Media y varianza de los estimadores de *ORIV-C3C4* y *RIV-C3C4*. Modelo *NERO*. (c = valores correctos de los coeficientes, O =*ORIV*, R =*RIV*)

En definitiva, los resultados experimentales corroboran los resultados teóricos en cuanto a tiempo de convergencia de la media y del MSD, asimismo confirman que *ORIV-C3C4* alcanza un menor MSD en todos los ambientes. Según el análisis teórico *RIV-C3C4* debía tener, asimismo, un tiempo de retraso mayor, este es el único punto no confirmado experimentalmente.

La falta de la información adicional en *RIV* hace que sus estimadores no sean de la calidad de los de *ORIV*, destacando sobre todo por su gran varianza.

Menos caricaturesca, aunque sigue los mismos patrones dados en los últimos comentarios, resulta la comparación *RIV-ORIV* con un modelo de orden 1. A modo ilustrativo se presenta la media y varianza de los estimadores, en la iteración 30000, de los coeficientes del modelo *ERO2* definido como $MA(2)=[1 \ -1.25]$, en la siguiente tabla:

	Método <i>RIV-C3C4</i>				Método <i>ORIV-C3C4</i>			
	Tipo 1		Tipo 2		Tipo 1		Tipo 2	
	b_1	var b_1	b_1	var b_1	b_1	var b_1	b_1	var b_1
SNR								
20	-1.250	0.014	-1.252	0.023	-1.251	0.014	-1.251	0.022
10	-1.251	0.018	-1.252	0.031	-1.251	0.018	-1.250	0.026
5	-1.250	0.028	-1.247	0.052	-1.253	0.026	-1.246	0.039
0	-1.258	0.063	-1.251	0.140	-1.278	0.063	-1.214	0.082
-5	-0.545	1.311	-0.581	1.758	-1.718	9.819	-0.878	0.309

Tabla 6.5 Media y varianza de los estimadores. Modelo *ERO2*.

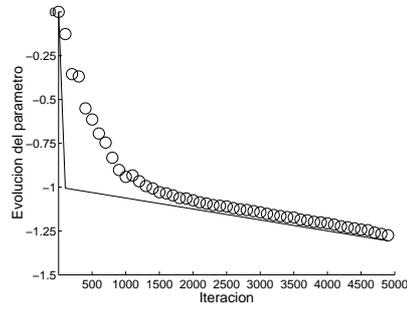


Figura 6.2 Evolución de b_1 tipo 1. Modelo NERO. RIV-C3C4

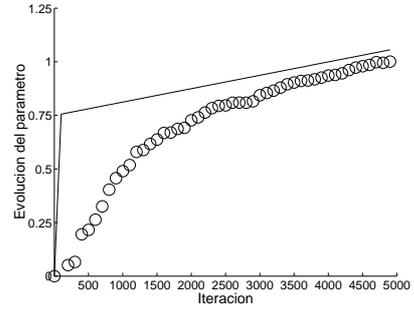


Figura 6.3 Evolución de b_2 tipo 1. Modelo NERO. RIV-C3C4

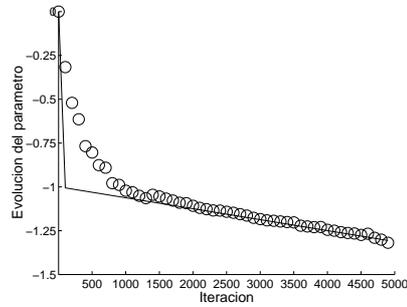


Figura 6.4 Evolución de b_1 tipo 2. Modelo NERO. RIV-C3C4

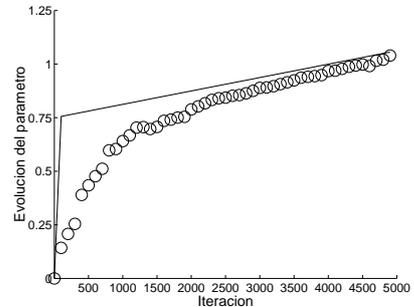


Figura 6.5 Evolución de b_2 tipo 2. Modelo NERO. RIV-C3C4

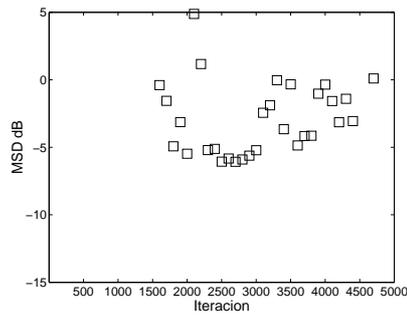


Figura 6.6 Curva de aprendizaje. Modelo NERO. RIV-C3C4

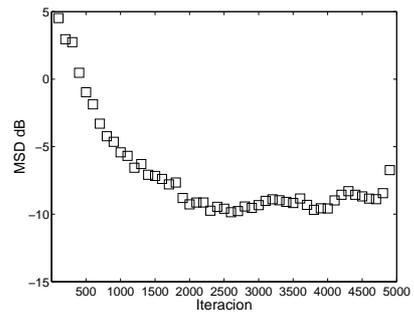


Figura 6.7 Curva de aprendizaje. Modelo NERO. ORIV-C3C4

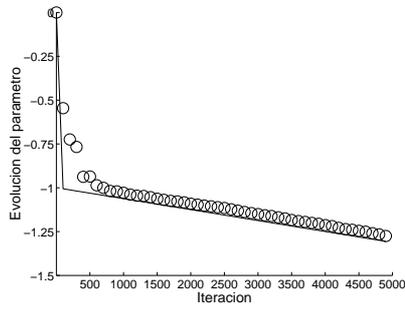


Figura 6.8 Evolución de b_1 tipo 1.
Modelo NERO. ORIV-C3C4

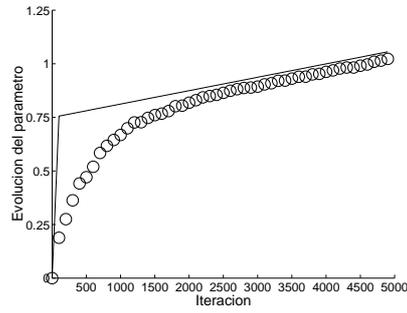


Figura 6.9 Evolución de b_2 tipo 1.
Modelo NERO. ORIV-C3C4

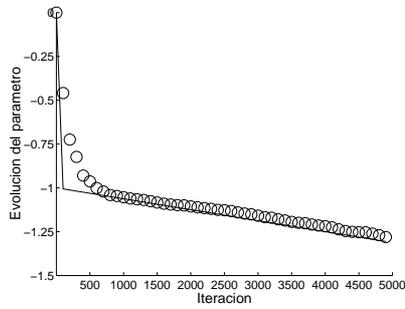


Figura 6.10 Evolución de b_1 tipo 2.
Modelo NERO. ORIV-C3C4

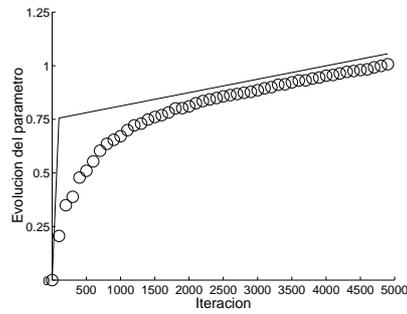


Figura 6.11 Evolución de b_2 tipo 2.
Modelo NERO. ORIV-C3C4

Parte II

Algoritmos adaptativos de mínimo error cuadrático medio con estadística de alto orden. Análisis y aplicación a identificación ciega

Introducción

Esta segunda parte de la Memoria comparte en su estructura muchos de los puntos de la primera parte. Ahora el problema de la estimación lineal se resuelve minimizando una función coste distinta, el error cuadrático medio. Debido a esta diferencia los algoritmos empleados para obtener la solución tienen unas propiedades distintas.

El mínimo error cuadrático medio se alcanza tras un proceso de búsqueda a lo largo de una trayectoria en el espacio de parámetros definida en cada punto por la dirección del gradiente y sentido opuesto. Dado que se desconoce la función densidad de probabilidad se hace necesario estimar el gradiente, normalmente adopta el valor que toma en la realización en la que se está trabajando (estimador instantáneo). De esta manera se construye el algoritmo de gradiente estocástico más difundido, el LMS o algoritmo del cuadrado de media mínima (Least-Mean Square).

El gradiente se anula una vez alcanzado el mínimo, esta condición tiene una naturaleza tal que permite ser interpretada como una condición de ortogonalidad entre el vector de datos empleados en la estimación y el error de estimación. A su vez esta condición se puede reexpresar como un sistema de ecuaciones lineales donde las incógnitas son los coeficientes de la estimación lineal.

El algoritmo LMS es capaz de imponer la condición de ortogonalidad cuando la matriz asociada es cuadrada. Además, dada la naturaleza del problema (estimación lineal con función de coste cuadrática) la estadística implicada es de segundo orden. La principal ventaja con respecto al RLS (que trabaja en idénticas condiciones) es su reducida carga computacional, aunque presenta un comportamiento algo más pobre. Por tanto, si se desean algoritmos adaptativos con baja carga computacional hay que recurrir a los de gradiente estocástico.

De igual manera que si se necesitaba introducir estadística de alto orden en el problema, el RLS debía ser sustituido por el RIV, en la misma circunstancia el LMS se debe sustituir por el LMS generalizado o GLMS. La principal desventaja que presenta este nuevo algoritmo es que si la matriz asociada no es definida,

no es capaz de converger y por tanto no permite resolver el problema.

Con estos antecedentes los objetivos perseguidos en esta parte de la Memoria son:

- i) Solventar el problema del algoritmo GLMS de manera que pueda converger con matrices no definidas.
- ii) Encontrar un algoritmo equivalente a ORIV tipo gradiente estocástico. Es decir, que trabaje bien en problemas con matrices sobredeterminadas que incluyan estadística de alto orden.

El cumplimiento de estos objetivos se fundamentó sólidamente en el análisis del algoritmo ORIV presentado en la primera parte. Así las aportaciones principales de esta parte son:

1. A partir de una condición de ortogonalidad derivada en el capítulo 4 se rederivó el algoritmo GLMS, dotándole de base, en contra de su obtención ad-hoc en [AAM96].
2. Se ha llevado a cabo un análisis riguroso del ya existente algoritmo GLMS en ambientes estacionarios, mediante el empleo de diversas aproximaciones.
3. El análisis teórico del GLMS se ha contrastado con resultados simulados, apreciándose muy buena concordancia.
4. A partir de la misma ecuación de ortogonalidad se construyen dos nuevos algoritmos, OGLMS1 y OGLMS3, como equivalentes del ORIV en la familia del gradiente estocástico.
5. Se analizan los algoritmos OGLMS1 y OGLMS3. De este análisis se desprende el mejor comportamiento del primero por lo que se elegirá para perfilar aún más sus propiedades.
6. Como consecuencia de ese resultado de depuración, el OGLMS1 genera el algoritmo AOGMS, punto final de la búsqueda. Este algoritmo cumple las propiedades requeridas de trabajar con matrices sobredeterminadas, con sus reducidas cuadradas no definidas e involucrando estadística de alto orden.
7. Por último se proponen tres nuevos algoritmos: CGLMS, A1GLMS y el PNGMS. El último es una versión normalizada del GLMS; los dos primeros son un intento sencillo de solventar el problema de matrices no definidas y donde haya más ecuaciones que incógnitas. Se incluye un análisis de todos ellos.

En definitiva, toda esta parte está orientada a encontrar un algoritmo robusto, tipo gradiente estocástico que trabaje en la situación más general posible. El resultado es el AOGMLS.

La presente parte se estructura de la siguiente manera. En el capítulo 7 se sientan las bases del problema a resolver presentando el algoritmo LMS y su análisis correspondiente con el fin de ir haciendo más familiares las hipótesis utilizadas. En el capítulo 8 se presenta el Principio de Ortogonalidad Sobredeterminado y Generalizado como fuente de los algoritmos GLMS, OGLMS1 y OGLMS3 y se analiza el primero de ellos con bastante profundidad ya que es un algoritmo existente en la literatura que carecía de un análisis riguroso. En el capítulo 9, tras un análisis teórico se decide proseguir mejorando el OGLMS1 en detrimento del OGLMS3. Con él, en el capítulo 10 se construye finalmente el algoritmo AOGMLS donde un apropiado análisis teórico confirma su convergencia, corroborada por resultados de simulaciones. Para concluir, en el capítulo 11 se presentan los algoritmos CGLMS, A1GLMS y PNGLS junto con un breve análisis que informe de su convergencia y corroborado mediante resultados de simulaciones.

Capítulo 7

El algoritmo del cuadrado de media mínima LMS

Introducción

Antes de comenzar con la presentación de algoritmos nuevos y con el análisis de otros ya existentes, es conveniente, para fijar un método de trabajo, incluir el algoritmo LMS junto con el estudio de su convergencia por ser temas profundamente tratados en la bibliografía y con sólidos fundamentos.

Este capítulo se encuentra estructurado en los siguientes apartados. En el apartado 7.1 se describe el problema de la estimación lineal de variables aleatorias y su solución óptima mediante un filtrado Wiener. A partir de esta base se construye el algoritmo LMS de forma natural gracias al método de máxima pendiente, en el apartado 7.2. Las técnicas e hipótesis que se suelen emplear en el análisis de este tipo de algoritmos están recogidas en los apartados 7.3 y 7.4 donde se estudia la convergencia en media y en media cuadrática del LMS respectivamente.

7.1 Estimación lineal de variables aleatorias. Filtrado Wiener

El problema de la estimación de variables aleatorias es uno de los más importantes dentro del campo de la estadística [Pap84]. Consiste en construir, a

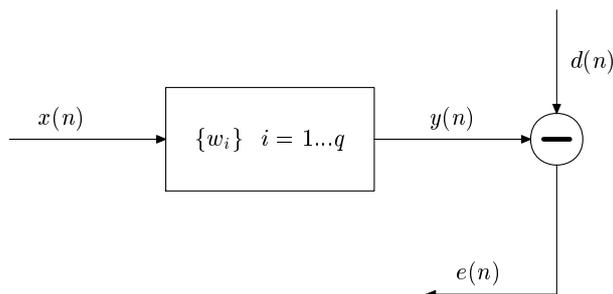


Figura 7.1 Estimación lineal de una respuesta deseada

partir de una variable conocida, otra de forma que se aproxime lo más posible a una variable dada. La nueva variable así construida se denomina estimador. Por ‘aproximarse’ se quiere indicar que se ajusta al criterio de selección establecido; éste puede ser el de menor error cuadrático medio, menor valor absoluto del error o simplemente menor media del error de estimación. En general, el problema se enuncia de la siguiente manera: Dada una variable aleatoria y , se busca el mejor estimador $\hat{y} = c(x)$, que minimice la función $f(e)$ del error $e = y - \hat{y}$. La variable aleatoria x es conocida y lo que queda por determinar es la función $c()$ una vez especificado el criterio de error (función $f()$).

En particular en este apartado se va a estudiar el caso de estimación lineal basada en un criterio de menor error cuadrático medio: dada una variable aleatoria y , se busca el estimador $\hat{y} = ax$ que minimice $E[(y - \hat{y})^2]$, siendo a un escalar. La condición de mínimo se alcanza para aquel valor de a que verifique $E[(y - \hat{y})x] = 0$, que es lo que se conoce como principio de ortogonalidad para estimación lineal de menor error cuadrático medio. Se puede interpretar diciendo que el error es ortogonal a los datos (variable conocida). Dos variables de media 0 se dicen ortogonales si no están correlacionadas.

A continuación se generaliza el problema de estimación al caso de contar como datos con un conjunto $\{x(n) \dots x(n-p)\}$ de variables aleatorias, pertenecientes a un proceso, para generar una señal de salida $d(n)$.

El principio de ortogonalidad se establece como una condición, deducida a partir de imponer cierto criterio de minimización, en el problema de estimación. Normalmente lo que se minimiza es el error cuadrático medio, tal y como ocurre en el filtrado Wiener, donde el error se define como la diferencia entre la salida del filtro wiener $y(n)$ y la señal de referencia a estimar $d(n)$ (véase la figura 7.1). El objetivo perseguido es encontrar un filtro w_0 que sea capaz de producir como salida la señal de referencia, teniendo como entrada la serie conocida $x(n)$.

Los coeficientes \mathbf{w}_0 del filtro se definen como aquellos que minimizan el MSE:

$$\text{MSE} = \text{E}[|d(n) - \mathbf{w}_0^t \mathbf{x}(n)|^2]$$

o equivalentemente, como aquellos que verifican la condición de ortogonalidad:

$$\text{E}[\mathbf{x}(n)(d(n) - \mathbf{w}_0^t \mathbf{x}(n))] = 0 \Leftrightarrow \text{E}[\mathbf{x}(n)e_0(n)] = 0 \quad (7.1)$$

Es sumamente interesante interpretar esta condición desde un punto de vista estadístico. Se puede suponer que la respuesta a estimar es $d(n) = \mathbf{w}_0^t \mathbf{x}(n) + e_0(n)$, donde $e_0(n)$ es ruido independiente del resto de variables aleatorias. Toda la información que pueda llevar $\mathbf{x}(n)$ sobre $d(n)$ está en \mathbf{w}_0 ; imponer (7.1) es extraer dicha información. Que dos variables aleatorias estén decorrelacionadas significa que no comparten información en el sentido de la predicción lineal: que una no se pueda estimar linealmente conociendo la otra. El principio de ortogonalidad lo que asegura por tanto es que en el error de estimación no quede ninguna información que pueda ser aportada por los datos conocidos $\mathbf{x}(n) = [x(n) \dots x(n-p)]^t$.

Una explicación cuantitativa podría ser la siguiente. De (7.1) se obtiene que

$$\mathbf{w}_0 = \text{E}[\mathbf{x}(n)\mathbf{x}^t(n)]^{-1} \text{E}[\mathbf{x}(n)d(n)] \quad (7.2)$$

Si $\text{E}[\mathbf{x}(n)d(n)] = 0$, es decir, si la señal a estimar y los datos con los que se cuenta no están correlacionados, si $\mathbf{x}(n)$ no contiene ninguna información acerca de $d(n)$, resulta que $\mathbf{w}_0 = 0$ y el mejor estimador de $d(n)$ es el 0 con un error de estimación $\text{E}[(d(n) - \mathbf{w}_0^t \mathbf{x}(n))^2] = \text{E}[d^2(n)] = \sigma_d^2$ que no dice nada desde un punto de vista de la estimación lineal en media cuadrática. Se sigue teniendo el mismo desconocimiento de $d(n)$.

El principio de ortogonalidad también posee una interpretación geométrica, recibiendo en este caso el nombre de teorema de proyección. Considérese el espacio vectorial de las variables aleatorias, junto con el producto escalar definido como el valor esperado:

$$\langle y, x \rangle = \text{E}[yx]$$

El estimador de $d(n)$ está contenido en el subespacio generado por $\{x(n) \dots x(n-p)\}$, minimizar el MSE significa minimizar la distancia entre d y su estimador: La distancia más corta se produce cuando $e(n)$ es ortogonal a dicho subespacio, es decir, cuando se verifica el principio de ortogonalidad. Como se aprecia en la figura 7.2 el estimador óptimo es la proyección de $d(n)$ sobre el subespacio, de ahí el nombre de teorema de proyección.

El criterio aquí dado para estimador óptimo está relacionado con el error cuadrático medio y por tanto con la estadística de segundo orden de $\mathbf{x}(n)$ y la correlación entre $\mathbf{x}(n)$ y $d(n)$. Si se eligiera la minimización de otra función de error que involucrase potencias de mayor orden, aparecería estadística de orden superior y por tanto la definición de ‘información’ sería distinta.

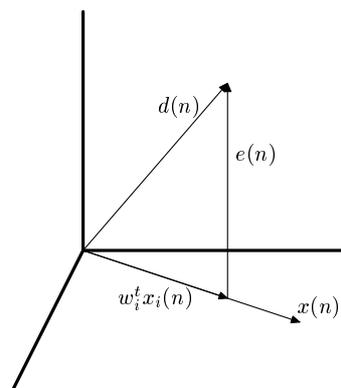


Figura 7.2 Representación gráfica del Teorema de Proyección

7.2 El algoritmo LMS

Una manera de acercarse al mínimo de una función es ir, una vez situado en un punto concreto, en la dirección contraria a la que marca el gradiente. Por tanto, si se quiere minimizar el MSE como función de los coeficientes del filtro, se puede proponer la siguiente ecuación para determinar el camino a seguir en la búsqueda del mínimo:

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \mu E[e(n)\mathbf{x}(n)]$$

recuérdese que, según lo visto en el apartado anterior, el gradiente del MSE en el punto $\mathbf{w}(n-1)$ se puede escribir como $\nabla = -E[e(n)\mathbf{x}(n)]$ donde $e(n) = d(n) - \mathbf{w}^t(n-1)\mathbf{x}(n)$. Teniendo en cuenta estas definiciones la ecuación de actualización anterior se puede reescribir como:

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \mu[\mathbf{r} - \mathbf{R}\mathbf{w}(n-1)] \quad (7.3)$$

donde \mathbf{r} es el vector de correlación entre la respuesta deseada y el vector de entrada y \mathbf{R} es la matriz de autocorrelación del vector de entrada.

El espacio en el que se avanza en la dirección del gradiente viene controlada por el factor μ y por tanto recibe el nombre de tamaño del paso. Una vez fijado un valor inicial $\mathbf{w}(0)$ y un tamaño del paso, la recursión anterior converge al vector \mathbf{w}_0 dado por la ecuación de Wiener (7.2). Este algoritmo recibe el nombre de máxima pendiente, ya que el gradiente indica en cada punto la dirección de mayor variación de una función. Para consultar las propiedades de la convergencia de este algoritmo véase por ejemplo [Hay96]. Lo que sí cabe destacar es que una vez ha convergido a \mathbf{w}_0 el valor del MSE toma su valor más bajo; este valor mínimo es $\text{MSE}_{\min} = E[e_0^2(n)]$.

La principal desventaja del algoritmo de máxima pendiente es que para poder aplicar la recursión dada en la expresión (7.3) es necesario conocer exactamente la estadística de $x(n)$ y de $d(n)$ para poder construir las correlaciones pertinentes. Este conocimiento no se tiene en la mayoría de las ocasiones y por tanto se hace preciso estimar dichas cantidades. Una posible opción es emplear un estimador instantáneo, de forma que la recursión quedase como

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \mu e(n)\mathbf{x}(n) \quad (7.4)$$

La principal diferencia es que este algoritmo cuenta con un gradiente estocástico y por eso se dice que pertenece a la familia de los algoritmos de gradiente estocástico, recibiendo el nombre de algoritmo del cuadrado de media mínima o Least-Mean-Square (LMS) [WH60].

En lo que resta de capítulo se va a proceder a analizar la convergencia del algoritmo LMS¹ en ambientes estacionarios. Una expresión que será muy útil en lo sucesivo se obtiene restando \mathbf{w}_0 a ambos miembros de (7.4) de manera que se puede escribir que

$$\Delta\mathbf{w}(n) = [\mathbf{I} - \mu\mathbf{x}(n)\mathbf{x}^t(n)]\Delta\mathbf{w}(n-1) + \mu e_0(n)\mathbf{x}(n) \quad (7.5)$$

donde el vector de error en los coeficientes del filtro se define como $\Delta\mathbf{w}(n) = \mathbf{w}(n) - \mathbf{w}_0$.

7.3 Convergencia de la media

Para que el algoritmo converja a \mathbf{w}_0 es necesario, aunque no suficiente, que $E[\Delta\mathbf{w}(n)] \rightarrow 0$ cuando $n \rightarrow \infty$. A este tipo de criterio se denomina en [Hay96] convergencia de la media, y coincide con lo que siempre se ha llamado como convergencia en media en la parte de esta Memoria relativa a los algoritmos RLS, RIV y ORIV. En [Hay96] la convergencia en media tiene que ver con la convergencia del valor esperado de la norma del vector de error: $E[\|\Delta\mathbf{w}(n)\|]$, criterio este suficiente para asegurar la convergencia.

Para estudiar la convergencia de la media se toma valor esperado a ambos miembros de la expresión (7.5), quedando:

$$E[\Delta\mathbf{w}(n)] = E[[\mathbf{I} - \mu\mathbf{x}(n)\mathbf{x}^t(n)]\Delta\mathbf{w}(n-1)] + 0 \quad (7.6)$$

El segundo sumando es nulo por el principio de ortogonalidad. Falta calcular el valor esperado $E[\mathbf{x}(n)\mathbf{x}^t(n)\Delta\mathbf{w}(n-1)]$ para lo cual se recurrirá a las hipótesis de independencia, ya enunciadas para los algoritmos tipo Least Squares y que para el LMS establecen que:

¹Múltiples artículos tratan acerca de la convergencia del LMS: [Bit83], [FW85], ...

Hipótesis de independencia

1. Los vectores de entrada $\mathbf{x}(1)$, $\mathbf{x}(2)$, ..., $\mathbf{x}(n)$ constituyen una serie de vectores estadísticamente independientes.
2. A tiempo n , el vector de entrada $\mathbf{x}(n)$ es estadísticamente independiente de todas las muestras previas de la respuesta deseada, esto es, de $d(1)$, $d(2)$, ..., $d(n-1)$.
3. A tiempo n , la respuesta deseada $d(n)$ depende del vector de entrada correspondiente $\mathbf{x}(n)$, pero es estadísticamente independiente de todas las muestras previas de la respuesta deseada.
4. Los procesos de entrada y de respuesta deseada, $\mathbf{x}(n)$ y $d(n)$, consisten en variables aleatorias distribuidas mutuamente de forma gaussiana para todo n .

A la vista de la ecuación (7.5) se puede afirmar que $\Delta\mathbf{w}(n)$ es función de $\{\mathbf{x}(n)$, $\mathbf{x}(n-1)$, ..., $\mathbf{x}(1)$, $d(n)$, $d(n-1)$, ..., $d(1)\}$ y de $\Delta\mathbf{w}(0)$. Teniendo en cuenta las hipótesis de independencia 1 y 2 resulta que el vector $\Delta\mathbf{w}(n)$ (y por tanto $\mathbf{w}(n)$) es independiente de $\mathbf{x}(n+1)$. En consecuencia, el valor esperado que se estaba calculando toma un valor de:

$$E[\mathbf{x}(n)\mathbf{x}^t(n)\Delta\mathbf{w}(n-1)] = E[\mathbf{x}(n)\mathbf{x}^t(n)]E[\Delta\mathbf{w}(n-1)]$$

Con este resultado la expresión (7.6) se puede escribir como una ecuación en diferencias

$$E[\Delta\mathbf{w}(n)] = [\mathbf{I} - \mu\mathbf{R}_{xx}]E[\Delta\mathbf{w}(n-1)] \quad (7.7)$$

El subíndice de \mathbf{R}_{xx} se eliminará siempre que no haya lugar a confusión. La solución de la ecuación (7.7) se puede obtener más fácilmente si se trabaja en la base en la que \mathbf{R} es diagonal, en ella se obtiene que la componente i -ésima del valor esperado del vector de error en los pesos evoluciona según la expresión:

$$E[\Delta w(n)]_i = [1 - \mu\lambda_i]E[\Delta w(n-1)]_i$$

y si se sustituye n veces en sí misma

$$E[\Delta w(n)]_i = [1 - \mu\lambda_i]^n E[\Delta w(0)]_i$$

y esta componente del vector tenderá a cero si

$$|1 - \mu\lambda_i| < 1$$

Recuérdese que como \mathbf{R} es una matriz definida positiva $\lambda_i > 0$ y por tanto la condición anterior es equivalente a

$$0 < \mu < \frac{2}{\lambda_i}$$

Como todas las componentes del vector deben tender a cero, esta condición debe cumplirse para cada uno de los autovalores de la matriz de autocorrelación del vector de entrada.

Bajo esta condición se asegura la convergencia de la media, como es conocido no es suficiente para afirmar que el algoritmo converge. Téngase en cuenta además que sólo se han empleado las 2 primeras hipótesis de independencia, el resto se emplearán en los próximos apartados donde su utilidad se hará más patente.

Estas hipótesis de independencia simplifican enormemente la matemática. Se han hecho algunos esfuerzos para poder relajarlas y llevar a cabo un análisis más riguroso, véase p.e. [Mac95], [BM91], [DM94], [Dou95], [RZ99] y [Sol97].

7.4 Convergencia en media cuadrática

Para asegurar que los estimadores de los pesos tienen una varianza finita se hace necesario comprobar que $E[\|\Delta\mathbf{w}(n)\|^2]$ no diverge, a esta cantidad se le suele notar por $D(n)$ y como se sabe se llama desviación del error (de los pesos) al cuadrado (squared error deviation o Mean-Square Deviation, MSD).

Como el algoritmo LMS surge como solución al problema de estimar la respuesta deseada $d(n)$ mediante la secuencia de datos $x(n)$, otro criterio útil y lleno de significado para la convergencia en media cuadrática es

$$J(n) = E[|e(n)|^2]$$

Más adelante se demostrará la equivalencia entre la convergencia de $J(n)$ y $D(n)$: si uno no diverge, el otro tampoco.

Sin embargo, y tal como se vio en el análisis de RIV o ORIV, $J(n)$ no tiene siempre sentido como aquí. Por tanto se estudiará primero la convergencia de $D(n)$ y después la de $J(n)$.

7.4.1 Convergencia de $D(n)$

Para construir una ecuación recursiva para $D(n)$ se tiene en cuenta que $D(n) = \text{Tr}\{E[\Delta\mathbf{w}(n)\Delta\mathbf{w}^t(n)]\}$ por lo que primero se estudiará la evolución dinámica de la matriz de correlación del error en los pesos $\mathbf{R}_\Delta(n) = E[\Delta\mathbf{w}(n)\Delta\mathbf{w}^t(n)]$.

Retomando la expresión (7.5),

$$\Delta \mathbf{w}(n) = [\mathbf{I} - \mu \mathbf{x}(n) \mathbf{x}^t(n)] \Delta \mathbf{w}(n-1) + \mu e_0(n) \mathbf{x}(n) \quad (7.8)$$

resulta que la evolución de $\Delta \mathbf{w}(n)$ viene regida por una ecuación en diferencias de orden 1 con coeficiente aleatorio. Resolver este tipo de ecuaciones es muy complejo por lo que se recurrirá a aplicar la técnica del promediado directo. Ésta consiste en suponer que si μ es muy pequeño la componente aleatoria va a ser despreciable y por tanto todo él se comportará como su valor medio.

Con todo esto se puede afirmar que la solución de (7.8) será muy parecida a la solución de:

$$\Delta \mathbf{w}(n) = [\mathbf{I} - \mu \mathbf{R}] \Delta \mathbf{w}(n-1) + \mu e_0(n) \mathbf{x}(n) \quad (7.9)$$

Estrictamente hablando estos vectores $\Delta \mathbf{w}(n)$ son distintos a los de la ecuación (7.8) pero se ha preferido mantener la notación por no complicarla.

A partir de (7.9) se construye el siguiente producto:

$$\begin{aligned} \Delta \mathbf{w}(n) \Delta \mathbf{w}^t(n) &= (\mathbf{I} - \mu \mathbf{R}) \Delta \mathbf{w}(n-1) \Delta \mathbf{w}^t(n-1) (\mathbf{I} - \mu \mathbf{R}) + \\ &\quad + \mu (\mathbf{I} - \mu \mathbf{R}) \Delta \mathbf{w}(n-1) \mathbf{x}^t(n) e_0(n) + \\ &\quad + \mu e_0(n) \mathbf{x}(n) \Delta \mathbf{w}^t(n-1) (\mathbf{I} - \mu \mathbf{R}) + \\ &\quad + \mu^2 e_0^2(n) \mathbf{x}(n) \mathbf{x}^t(n) \end{aligned} \quad (7.10)$$

Tomando ahora valor esperado queda:

$$\mathbf{R}_\Delta(n) = (\mathbf{I} - \mu \mathbf{R}) \mathbf{R}_\Delta(n-1) (\mathbf{I} - \mu \mathbf{R}) + \mu^2 \sigma_0^2 \mathbf{R} \quad (7.11)$$

Donde se ha tenido en cuenta lo siguiente:

- a) En el valor esperado del segundo sumando aparece el término $E[\Delta \mathbf{w}(n-1) \mathbf{x}^t(n) e_0(n)]$ donde $e_0(n) = d(n) - \mathbf{w}_0^t \mathbf{x}(n)$.

Tras las hipótesis de independencia se afirmó, usando la primera y la segunda, que $\Delta \mathbf{w}(n)$ es independiente de $\mathbf{x}(n+1)$; empleando también ahora la tercera se puede afirmar que también es independiente de $d(n+1)$. Uniendo ambas se puede concluir que $\Delta \mathbf{w}(n-1)$ es independiente de $e_0(n)$ y por tanto se tiene que

$$E[\Delta \mathbf{w}(n-1) \mathbf{x}^t(n) e_0(n)] = E[\Delta \mathbf{w}(n-1)] E[\mathbf{x}^t(n) e_0(n)] = 0$$

debido al principio de ortogonalidad.

Lo discutido en este punto también es aplicable al tercer sumando.

- b) El último sumando proviene del valor esperado

$$E[e_0(n) e_0(n) \mathbf{x}(n) \mathbf{x}^t(n)]$$

Teniendo en cuenta la cuarta hipótesis de independencia y la definición de $e_0(n)$ se puede aplicar el teorema de factorización gaussiana [Ree62], que junto con el principio de ortogonalidad dan como resultado el segundo sumando de (7.11).

Para resolver la ecuación (7.11) se trabajará de nuevo en la base en la que \mathbf{R} es diagonal. En esta base la matriz \mathbf{R}_Δ viene representada por \mathbf{R}'_Δ , pero como el operador traza es invariante bajo cambios de base resulta que $D(n) = \text{Tr}\{\mathbf{R}_\Delta(n)\} = \text{Tr}\{\mathbf{R}'_\Delta(n)\}$ y por tanto el interés se centra ahora en calcular los elementos diagonales de $\mathbf{R}'_\Delta(n)$:

$$\begin{aligned} R'_{\Delta ii}(n) &= (\delta_{il} - \mu\delta_{il}\lambda_i)R'_{\Delta lm}(n-1)(\delta_{mi} - \mu\delta_{mi}\lambda_m) + \mu\sigma_0^2\lambda_i = \\ &= (1 - \mu\lambda_i)^2 R'_{\Delta ii}(n-1) + \mu^2\sigma_0^2\lambda_i \end{aligned} \quad (7.12)$$

y como se puede observar los elementos diagonales se han desacoplado. La ecuación en diferencias obtenida se puede resolver fácilmente siguiendo las indicaciones del apéndice D.

$$R'_{\Delta ii}(n) = (1 - \mu\lambda_i)^{2n} [R'_{\Delta ii}(0) - R'_{\Delta ii}(\infty)] + R'_{\Delta ii}(\infty) \quad (7.13)$$

donde

$$R'_{\Delta ii}(\infty) = \frac{\mu}{2 - \mu\lambda_i} \sigma_0^2$$

En definitiva, el operador traza conduce a que

$$D(n) = \sum_{i=1}^q (1 - \mu\lambda_i)^{2n} [R'_{\Delta ii}(0) - R'_{\Delta ii}(\infty)] + \sum_{i=1}^q R'_{\Delta ii}(\infty)$$

Para que $D(n)$ no diverja es necesario que $|1 - \mu\lambda_i| < 1 \forall \lambda_i$, lo cual conduce a la siguiente condición

$$0 < \mu < \frac{2}{\lambda_i} \quad \forall \lambda_i$$

que coincide con lo exigible para que converja la media. Bajo esta condición $D(n) \rightarrow D(\infty)$ donde

$$D(\infty) = \sum_{i=1}^q \frac{\mu}{2 - \mu\lambda_i} \sigma_0^2 \quad (7.14)$$

7.4.2 Convergencia de $\mathbf{J}(n)$

Como se recordará $\mathbf{J}(n) = E[e^2(n)]$, es decir, la media del cuadrado del error de estimación. Esta expresión se puede reescribir en función del error de estimación $e_0(n)$ cometido si se utilizasen los filtros wiener óptimos, para ello

basta tener en cuenta que $d(n) = \mathbf{w}_0^t \mathbf{x}(n) + e_0(n)$ y que $e(n) = d(n) - \mathbf{w}^t(n-1)\mathbf{x}(n)$ con lo que se obtiene:

$$J(n) = E[(e_0(n) + \mathbf{w}_0^t \mathbf{x}(n) - \mathbf{w}^t(n-1)\mathbf{x}(n))^2] = E[(e_0(n) - \Delta \mathbf{w}^t(n-1)\mathbf{x}(n))^2]$$

Si se desarrolla el cuadrado y se tienen en cuenta las hipótesis de independencia usuales junto con el principio de ortogonalidad, esta expresión se simplifica a:

$$J(n) = J_{\min} + E[\mathbf{x}^t(n)\mathbf{w}(n-1)\mathbf{w}^t(n-1)\mathbf{x}(n)]$$

El segundo sumando que aquí aparece aún se puede expresar de otra manera:

$$\begin{aligned} E[\mathbf{x}^t(n)\mathbf{w}(n-1)\mathbf{w}^t(n-1)\mathbf{x}(n)] &= \\ \text{Tr}\{E[\mathbf{w}(n-1)\mathbf{w}^t(n-1)\mathbf{x}(n)\mathbf{x}^t(n)]\} &= \text{Tr}\{\mathbf{R}_\Delta(n-1)\mathbf{R}\} \end{aligned} \quad (7.15)$$

donde en la última igualdad se han vuelto a emplear las hipótesis de independencia.

Resulta por tanto que

$$J(n) = J_{\min} + \text{Tr}\{\mathbf{R}_\Delta(n-1)\mathbf{R}\}$$

Dado que tanto la matriz $\mathbf{R}_\Delta(n-1)$ como \mathbf{R} son definidas positivas, su producto también lo es y por tanto su traza es siempre positiva. Esto quiere decir que $J(n)$ dado por el algoritmo LMS está siempre por encima del obtenido mediante el filtrado Wiener J_{\min} . A esta cantidad extra se le denomina error cuadrático medio de exceso J_{exc} :

$$J_{\text{exc}} = J(n) - J_{\min} = \text{Tr}\{\mathbf{R}_\Delta(n-1)\mathbf{R}\}$$

Recordando de nuevo las propiedades de la traza,

$$J_{\text{exc}} = \text{Tr}\{\mathbf{R}'_\Delta(n-1)\mathbf{R}'\} = R'_{\Delta ii}(n-1)\lambda_i$$

que junto con la expresión (7.13) conduce a:

$$J_{\text{exc}}(n) = \sum_{i=1}^q \lambda_i (1 - \mu\lambda_i)^{2n} [R'_{\Delta ii}(0) - R'_{\Delta ii}(\infty)] + \sum_{i=1}^q \lambda_i R'_{\Delta ii}(\infty) \quad (7.16)$$

De la cual se puede concluir que el valor en convergencia del error cuadrático medio de exceso vale:

$$J_{\text{exc}}(\infty) = \sigma_0^2 \sum_{i=1}^q \frac{\mu\lambda_i}{2 - \mu\lambda_i} \equiv J_{\min} \sum_{i=1}^q \frac{\mu\lambda_i}{2 - \mu\lambda_i} \quad (7.17)$$

El desajuste

Se define como el cociente entre el error cuadrático medio de exceso en convergencia y el error cuadrático medio mínimo. Se designa por la letra \mathcal{M} :

$$\mathcal{M} = \frac{J_{\text{exc}}}{J_{\text{min}}}$$

Da una idea de cómo de cerca el comportamiento de LMS se encuentra del comportamiento óptimo. De la ecuación (7.17) es inmediato que:

$$\mathcal{M} = \sum_{i=1}^q \frac{\mu\lambda_i}{2 - \mu\lambda_i}$$

Comentario

Si el tamaño del paso μ es pequeño, el desajuste se puede aproximar por

$$\mathcal{M} \sim \frac{\mu}{2} \sum_{i=1}^q \lambda_i$$

Por otro lado, la parte que depende de n en (7.16) son exponenciales decrecientes de la forma $(1 - \mu\lambda_i)^{2n}$. Estas expresiones (como las del tipo (7.12)) tienen un tiempo de convergencia asociado dado por

$$\tau_i = \frac{1}{1 - (1 - \mu\lambda_i)^2} = \frac{1}{\mu\lambda_i} \frac{1}{2 - \mu\lambda_i}$$

finalmente el tiempo de convergencia se puede aproximar por

$$\tau_i \simeq \frac{1}{2\mu\lambda_i}$$

Por tanto, conseguir un \mathcal{M} bajo (es decir, trabajar con un μ pequeño), lleva asociado unos tiempos de convergencia grandes y viceversa. No se pueden conseguir ambos objetivos a la vez.

Capítulo 8

El Principio de Ortogonalidad Sobredeterminado y Generalizado. El algoritmo LMS generalizado

Introducción

En este capítulo se dará un paso adelante en la creación de varios algoritmos del tipo gradiente estocástico, sirviéndose de una ecuación derivada en el capítulo 4 y que ha recibido el nombre de Principio de Ortogonalidad Sobredeterminado y Generalizado.

El primer fruto de esta combinación máxima pendiente-principio de ortogonalidad es la reobtención del algoritmo LMS generalizado, GLMS. De esta manera el GLMS aparece como miembro de una familia de algoritmos que comparten el mismo formalismo. Los otros dos miembros de la familia son los algoritmos LMS sobredeterminados y generalizados 1 y 3, OGLMS1 y OGLMS3. Todos los miembros de esta familia pueden trabajar con estadística de alto orden (de ahí lo de ‘generalizados’) pero sólo los dos últimos son capaces de incluir matrices sobredeterminadas. Dado que el GLMS es un algoritmo existente se analizará primero, siendo esto contribución propia de esta Memoria. El resto de algoritmos se deja para el capítulo siguiente.

El Principio de Ortogonalidad Sobredeterminado y Generalizado se reproduce en el apartado 8.1 donde se propone un nuevo algoritmo llamado OGLMS. Este algoritmo, tras comprobar que converge en media en el apartado 8.2 da lugar a los tres algoritmos, ya comentados anteriormente, en el apartado 8.3. De ellos se elegirá el GLMS, apartado 8.4, para analizar su convergencia en media en el apartado 8.5. Para la convergencia en media cuadrática se emplean varias aproximaciones que se pueden agrupar en dos categorías: aquellas que trabajan a orden μ^2 (siendo μ el tamaño del paso que controla al algoritmo), en el apartado 8.6, y aquellas que trabajan a orden μ , en el apartado 8.7. Para tener una visión más general en el apartado 8.8 se unifican los resultados obtenidos con las diversas aproximaciones que culmina con una expresión general para el MSD en el apartado 8.9. Para comprobar las expresiones obtenidas se propone la aplicación del GLMS a un problema concreto en el apartado 8.10 y su implementación numérica ampliamente comentada en el apartado 8.11.

8.1 El Principio de Ortogonalidad Sobredeterminado y Generalizado

En el capítulo 7 se comentó que partiendo del principio de ortogonalidad para estimación lineal, gracias a que era consecuencia de imponer el mínimo del MSE (una función coste), se podía construir el algoritmo LMS. También como se comentó allí, si el razonamiento anterior se generaliza, dada cualquier condición de ortogonalidad entre un vector conocido y el error de estimación, se puede construir un algoritmo tipo gradiente estocástico que resuelva tal problema. Matemáticamente se puede decir que si la solución a cierto problema se consigue imponiendo que

$$E[\mathbf{f}(\mathbf{x}(n), e(n))] = 0$$

la siguiente recursión converge a la solución:

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \mu \mathbf{f}(\mathbf{x}(n), e(n)) \quad (8.1)$$

A continuación se sigue este procedimiento para obtener un nuevo algoritmo tipo LMS.

En el capítulo 4 se demostró (ecuación (4.40)) que el algoritmo ORIV impone la condición:

$$E[\mathbf{X}(n)\mathbf{\Lambda}^{-1}(n)\boldsymbol{\alpha}_0(n)] = 0 \quad (8.2)$$

que se puede considerar como un principio de ortogonalidad sobredeterminado que generaliza a (7.1). El carácter de sobredeterminado se lo confiere el problema del que proviene, ya que esta expresión se obtuvo en el análisis del algoritmo ORIV que resuelve sistemas de ecuaciones sobredeterminados. La matriz $\mathbf{X}(n)$ es la matriz de datos, el vector $\boldsymbol{\alpha}_0(n)$ el vector de error y la matriz $\mathbf{\Lambda}^{-1}(n)$ hace

de unión entre ambas magnitudes. El vector de error toma la forma conocida de una diferencia:

$$\boldsymbol{\alpha}_0(n) = \mathbf{v}(n) - \mathbf{X}^t(n)\mathbf{w}_0$$

donde las distintas magnitudes se definen de la siguiente manera:

$$\mathbf{X}(n) = [\boldsymbol{\Phi}^t(n-1)\tilde{\mathbf{x}}(n) \quad \mathbf{x}(n)]$$

$$\boldsymbol{\Lambda}^{-1}(n) = \begin{pmatrix} 0 & \lambda \\ \lambda & \tilde{\mathbf{x}}^t(n)\tilde{\mathbf{x}}(n) \end{pmatrix}$$

$$\mathbf{v}(n) = \begin{pmatrix} \tilde{\mathbf{x}}^t(n)\mathbf{z}(n-1) \\ d(n) \end{pmatrix}$$

que dependen del vector de datos $\mathbf{x}(n)$, vector de variable instrumental $\tilde{\mathbf{x}}(n)$, respuesta deseada $d(n)$, la matriz $\boldsymbol{\Phi}(n)$ de correlación entre $\mathbf{x}(n)$ y $\tilde{\mathbf{x}}(n)$ y el vector de correlación entre $\tilde{\mathbf{x}}(n)$ y $d(n)$, $\mathbf{z}(n)$. El significado de todas las magnitudes se detalla en el capítulo 4.

A tenor de lo discutido anteriormente se podría proponer una ecuación recursiva de actualización de los parámetros, que converge a la solución dada por ORIV, del tipo:

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \mu\mathbf{X}(n)\boldsymbol{\Lambda}^{-1}(n)\boldsymbol{\alpha}(n) \quad (8.3)$$

donde el vector de error que aquí aparece está calculado usando $\mathbf{w}(n-1)$. Dicha recursión resolvería el sistema de ecuaciones sobredeterminado:

$$\mathbf{E}[\boldsymbol{\Phi}(n)]\mathbf{w}_0(n) = \mathbf{E}[\mathbf{z}(n)] \quad (8.4)$$

de forma adaptativa. El planteamiento de dicho problema, junto con su resolución mediante ORIV se discutió en el capítulo 4 y ejemplos prácticos donde se resuelven ecuaciones tipo (8.4) mediante ORIV se presentaron en el capítulo 5.

El algoritmo resultante, que tiene como ecuación fundamental (8.3), se resume en la tabla 8.1. Se denominará algoritmo sobredeterminado y generalizado de mínima media cuadrática o en inglés, por mantener la nomenclatura existente, Overdetermined and Generalized Least-Mean-Square algorithm, OGLMS.

<p><u>Inicialización</u></p> $\mathbf{w}(0) = 0 \quad \boldsymbol{\Phi}(0) = 0 \quad \mathbf{z}(0) = 0$ <p>Con la llegada de $\mathbf{x}(n)$, $\tilde{\mathbf{x}}(n)$ y $d(n)$ hacer</p> <hr style="width: 80%; margin: auto;"/> <p style="text-align: center;">Construir los vectores $\mathbf{x}(n)$ y $\tilde{\mathbf{x}}(n)$</p> $\mathbf{X}(n) = [\boldsymbol{\Phi}^t(n-1)\tilde{\mathbf{x}}(n) \quad \mathbf{x}(n)]$ $\boldsymbol{\Lambda}^{-1}(n) = \begin{pmatrix} 0 & \lambda \\ \lambda & \tilde{\mathbf{x}}^t(n)\tilde{\mathbf{x}}(n) \end{pmatrix}$
--

$$\begin{aligned}
\mathbf{v}(n) &= \begin{pmatrix} \tilde{\mathbf{x}}^t(n)z(n-1) \\ d(n) \end{pmatrix} \\
\boldsymbol{\alpha}(n) &= \mathbf{v}(n) - \mathbf{X}^t(n)\mathbf{w}(n-1) \\
\mathbf{w}(n) &= \mathbf{w}(n-1) + \mu\mathbf{X}(n)\boldsymbol{\Lambda}^{-1}(n)\boldsymbol{\alpha}(n) \\
\boldsymbol{\Phi}(n) &= \lambda\boldsymbol{\Phi}(n-1) + \tilde{\mathbf{x}}(n)\mathbf{x}^t(n) \\
z(n) &= \lambda z(n-1) + \tilde{\mathbf{x}}(n)d(n)
\end{aligned}$$

Tabla 8.1 *Proceso iterativo del algoritmo OGLMS*

8.2 Convergencia en media de OGLMS

A continuación se comprobará que el algoritmo descrito converge en media a la solución de (8.4) con la ayuda del principio de ortogonalidad sobredeterminado. Para ello se parte de la ecuación (8.3) y se le restan a ambos miembros $\mathbf{w}_0(n)$:

$$\Delta\mathbf{w}(n) = [\mathbf{I} - \mu\mathbf{X}(n)\boldsymbol{\Lambda}^{-1}(n)\mathbf{X}^t(n)]\Delta\mathbf{w}(n-1) + \mu\mathbf{X}(n)\boldsymbol{\Lambda}^{-1}(n)\boldsymbol{\alpha}_0(n)$$

donde también se le ha sumado y restado al miembro de la derecha $\mu\mathbf{X}(n)\boldsymbol{\Lambda}^{-1}(n)\mathbf{X}^t(n)\mathbf{w}_0$. Si a continuación se toma valor esperado teniendo en cuenta las hipótesis de independencia usuales para ORIV dadas en el apéndice C, junto con (8.2), resulta:

$$\mathbb{E}[\Delta\mathbf{w}(n)] = [\mathbf{I} - \mu\mathbb{E}[\mathbf{X}(n)\boldsymbol{\Lambda}^{-1}(n)\mathbf{X}^t(n)]]\mathbb{E}[\Delta\mathbf{w}(n-1)] \quad (8.5)$$

Para calcular el valor esperado que aparece entre los corchetes se desarrolla primero el producto según la definición de sus componentes, resultando:

$$\mathbf{X}(n)\boldsymbol{\Lambda}^{-1}(n)\mathbf{X}^t(n) = \lambda\mathbf{x}(n)\tilde{\mathbf{x}}^t(n)\boldsymbol{\Phi}(n-1) + \boldsymbol{\Phi}^t(n)\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)$$

si ahora se toma valor esperado considerando las hipótesis de independencia usuales y que se trabaja en ambientes estacionarios ($\lambda = 1$), se llega a

$$\mathbb{E}[\Delta\mathbf{w}(n)] = [\mathbf{I} - 2\mu n\mathbf{R}^t\mathbf{R}]\mathbb{E}[\Delta\mathbf{w}(n-1)]$$

Por simplificar las expresiones es conveniente redefinir μ como μ/n de manera que desaparezca el n como factor en el coeficiente de la ecuación en diferencias. Otra opción es normalizar la matriz $\boldsymbol{\Phi}(n)$ si el sistema es estacionario. El siguiente paso es ver si esta ecuación en diferencias converge y a qué valor. Para desacoplar las ecuaciones para las distintas componentes de $\Delta\mathbf{w}(n)$ se trabaja en la base en la que $\mathbf{R}^t\mathbf{R}$ es diagonal, en ella:

$$\mathbb{E}[\Delta w_i(n)] = (1 - 2\mu\lambda_i)\mathbb{E}[\Delta w_i(n-1)]$$

Donde λ_i es el autovalor i -ésimo de $\mathbf{R}^t \mathbf{R}$. Si se sustituye esta ecuación n veces en sí misma, resulta:

$$\mathbb{E}[\Delta w_i(n)] = (1 - 2\mu\lambda_i)^n \mathbb{E}[\Delta w_i(0)]$$

Para que la media del vector de error en los pesos tienda a cero se debe cumplir que

$$|1 - 2\mu\lambda_i| < 1$$

de manera que debe verificarse que

$$0 < \mu < \frac{1}{\lambda_i} \quad (8.6)$$

donde se ha tenido en cuenta que los autovalores λ_i son siempre positivos. Si esta condición se cumple se tiene que

$$\mathbb{E}[\Delta \mathbf{w}(n)] \xrightarrow{n \rightarrow \infty} 0$$

lo que viene a decir que el algoritmo propuesto converge en media a los valores verdaderos de los pesos del filtro buscado, cuando n se hace suficientemente grande.

8.3 Algunas variantes de OGLMS

En vez de seguir con el estudio del comportamiento del algoritmo OGLMS, en este apartado se va a simplificar computacionalmente dando lugar a 4 versiones distintas del mismo. Una de las cuales será el algoritmo GLMS desarrollado en [AAM96], de ahí el nombre de OGLMS dado al algoritmo deducido en este capítulo.

Para realizar la simplificación se desarrolla el término de gradiente de (8.3), responsable de la convergencia del algoritmo:

$$\begin{aligned} \mathbf{X}(n)\mathbf{\Lambda}^{-1}(n)\boldsymbol{\alpha}(n) &= \lambda \mathbf{x}(n)\tilde{\mathbf{x}}^t(n)[z(n-1) - \mathbf{\Phi}(n-1)\mathbf{w}(n-1)] + \\ &+ \mathbf{\Phi}^t(n)[\tilde{\mathbf{x}}(n)d(n) - \tilde{\mathbf{x}}(n)\mathbf{x}^t(n)\mathbf{w}(n-1)] \end{aligned} \quad (8.7)$$

el miembro de la derecha de esta expresión está formado por dos sumandos, cada uno de los cuales vale cero al tomar valor esperado con n suficientemente grande cuando $\mathbf{w}(n) = \mathbf{w}_0$. Es decir, se podría usar solamente uno como término de gradiente, simplificando por tanto la recursión.

En consecuencia, se pueden proponer cuatro opciones distintas:

1ª Opción: Se toma como término de gradiente el primer sumando,

$$\nabla(n) = \lambda \mathbf{x}(n)\tilde{\mathbf{x}}^t(n)[z(n-1) - \mathbf{\Phi}(n-1)\mathbf{w}(n-1)]$$

2ª Opción: Como la parte que lleva información sobre la proximidad a la solución es la diferencia, se propone

$$\nabla(n) = z(n-1) - \Phi(n-1)\mathbf{w}(n-1)$$

La principal novedad respecto a las versiones anteriores es que ya no resuelve un problema sobredeterminado (por cuestión de homogeneidad de dimensiones en (8.1)), por lo que habrá que redimensionar las cantidades $\Phi(n)$ y $z(n)$ reduciendo el número de ecuaciones que tendrá el sistema de ecuaciones a resolver (8.4).

Esta expresión conduce a la formulación generalizada del problema de mínimos cuadrados cuando se introduce una variable instrumental. Es decir, si se tienen en cuenta las definiciones de $z(n)$ y $\Phi(n)$, se llega a:

$$\nabla = \sum_{i=0}^{n-1} \lambda^{n-i-1} \tilde{\mathbf{x}}(i) e(i, n-1)$$

que se puede considerar como el principio de ortogonalidad para series deterministas. Esta ecuación impone la ortogonalidad entre la variable instrumental y el error a tiempo i calculado con el estimador a tiempo $n-1$ dado por $e(i, n-1) = d(i) - \mathbf{x}^t(i)\mathbf{w}(n-1)$ [GBT99].

3ª Opción: Se toma como término de gradiente el segundo sumando,

$$\nabla(n) = \Phi^t(n) [\tilde{\mathbf{x}}(n)d(n) - \tilde{\mathbf{x}}(n)\mathbf{x}^t(n)\mathbf{w}(n-1)]$$

4ª Opción: Si simplemente se conserva la parte de la diferencia que lleva información suficiente,

$$\nabla(n) = \tilde{\mathbf{x}}(n)(d(n) - \mathbf{x}^t(n)\mathbf{w}(n-1))$$

De nuevo el problema deja de ser sobredeterminado y además se recupera el algoritmo GLMS. El principal problema de este algoritmo como se verá más adelante, tal y como le ocurre a la 2ª opción, es que no siempre converge, ya que necesita que la matriz \mathbf{R} sea definida (positiva o negativa). Las opciones 1ª y 3ª no sufren este problema, al igual que el algoritmo original OGLMS tal y como se observa en la condición (8.6).

Como se puede apreciar esta forma de actuar mediante la cual se construyen algoritmos estocásticos tipo máxima pendiente a partir de condiciones de ortogonalidad se ha mostrado muy fructífera. No sólo se reobtiene el algoritmo GLMS dándole un fundamento teórico sino que se proponen tres nuevos. Las propiedades de estos últimos se estudiarán en un tema posterior, ya que se ha preferido estudiar el algoritmo ya existente, el GLMS, para que sirva de punto de comparación, en primer lugar.

8.4 El algoritmo GLMS

En el resto del capítulo se va a estudiar [ARBC02c] el algoritmo generalizado de mínima media cuadrática, o Generalized Least-Mean-Square, GLMS. Se analizará su convergencia en media y en media cuadrática tanto mediante teoría como mediante simulaciones.

La ecuación fundamental de este algoritmo se presentaba en la opción 4ª anterior, que junto con los pasos detallados en la tabla 8.2 constituyen todo el proceso iterativo necesario.

<p><u>Inicialización</u></p> $\mathbf{w}(0) = 0$
<p>Con la llegada de $x(n)$, $\tilde{x}(n)$ y $d(n)$ hacer</p> <hr style="width: 80%; margin: auto;"/> <p>Construir los vectores $\mathbf{x}(n)$ y $\tilde{\mathbf{x}}(n)$</p> $e(n) = d(n) - \mathbf{w}^H(n-1)\mathbf{x}(n)$ $\mathbf{w}(n) = \mathbf{w}(n-1) + \mu\tilde{\mathbf{x}}(n)e^*(n)$

Tabla 8.2 Proceso iterativo del algoritmo GLMS

Como se puede apreciar requiere de una carga computacional muy baja por iteración y la única diferencia con el algoritmo LMS es la introducción de la variable instrumental $\tilde{x}(n)$ lo cual le permite trabajar con estadística de alto orden. El nombre GLMS le viene precisamente de generalizar al LMS. Además se va a considerar la posibilidad de que las magnitudes que intervienen en él sean complejas. Hay que tener en cuenta que dadas las simplificaciones realizadas en la condición de ortogonalidad generalizada y sobredeterminada (8.2), el principio de ortogonalidad para GLMS toma la forma:

$$E[\tilde{\mathbf{x}}(n)e^*(n)] = 0 \tag{8.8}$$

8.5 Convergencia de la media

En primer lugar se desea comprobar que el proceso iterativo presentado en la tabla 8.2 efectivamente converge en media a \mathbf{w}_0 . De la ecuación de actualización

de los pesos que allí aparece se sigue sin dificultad que:

$$\begin{aligned} \mathbf{w}(n) - \mathbf{w}_0 = & \mathbf{w}(n-1) - \mathbf{w}_0 + \mu \tilde{\mathbf{x}}(n) d^*(n) - \mu \tilde{\mathbf{x}}(n) \mathbf{x}^H(n) \mathbf{w}(n-1) + \\ & + \mu \tilde{\mathbf{x}}(n) \mathbf{x}^H(n) \mathbf{w}_0 - \mu \tilde{\mathbf{x}}(n) \mathbf{x}^H(n) \mathbf{w}_0 \end{aligned} \quad (8.9)$$

por lo que resulta

$$\Delta \mathbf{w}(n) = [\mathbf{I} - \mu \tilde{\mathbf{x}}(n) \mathbf{x}^H(n)] \Delta \mathbf{w}(n-1) + \mu \tilde{\mathbf{x}}(n) e_0^*(n) \quad (8.10)$$

siendo $\Delta \mathbf{w}(n) = \mathbf{w}(n) - \mathbf{w}_0$. Que constituye una ecuación en diferencias no homogénea con coeficientes aleatorios. Interesa ver qué forma tiene dicha ecuación en valor medio, para ello se hace uso de las hipótesis de independencia¹:

Hipótesis de independencia

1. Los vectores de entrada $\mathbf{x}(1)$, $\mathbf{x}(2)$, ..., $\mathbf{x}(n)$ constituyen una serie de vectores estadísticamente independientes.
2. Los vectores de variable instrumental $\tilde{\mathbf{x}}(1)$, $\tilde{\mathbf{x}}(2)$, ..., $\tilde{\mathbf{x}}(n)$ constituyen una serie de vectores estadísticamente independientes.
3. Los escalares de respuesta deseada $d(1)$, $d(2)$, ..., $d(n)$ constituyen una serie de escalares estadísticamente independientes.
4. Las series de vectores $\{\mathbf{x}(n)\}$ y $\{\tilde{\mathbf{x}}(n)\}$ son conjuntamente estacionarias.
5. El error de medición $e_0(n)$ es independiente del resto de variables a tiempos anteriores y de media nula².
6. Las variables $x(n)$, $\tilde{x}(n)$ y $d(n)$ son conjuntamente gaussianas.

A la vista de la ecuación (8.10) se desprende que $\mathbf{w}(n)$ sólo depende de muestras $\{d(n)\}$, $\{\mathbf{x}(n)\}$ y $\{\tilde{\mathbf{x}}(n)\}$ hasta tiempo n y que por tanto es independiente de $\{d(n+1)\}$, $\{\mathbf{x}(n+1)\}$ y $\{\tilde{\mathbf{x}}(n+1)\}$ por las 3 primeras hipótesis de independencia.

Por tanto, tomando valor esperado en (8.10) y teniendo en cuenta (8.8):

$$\mathbf{E}[\Delta \mathbf{w}(n)] = [\mathbf{I} - \mu \mathbf{R}] \mathbf{E}[\Delta \mathbf{w}(n-1)]$$

esta ecuación en diferencias homogénea y determinista tiene una solución que se calcula rápidamente si se sustituye n veces en sí misma:

$$\mathbf{E}[\Delta \mathbf{w}(n)] = [\mathbf{I} - \mu \mathbf{R}]^n \mathbf{E}[\Delta \mathbf{w}(0)]$$

¹Suponen una adaptación de las hipótesis de independencia para LMS y son análogas a las ya usadas en la parte de la Memoria relativa a ORIV.

²La interpretación final de $e_0(n)$ depende del problema concreto a resolver. Depende por tanto de la definición de $\mathbf{x}(n)$ y de $d(n)$.

si se trabaja en la base en la que \mathbf{R} es diagonal, las componentes del vector de error en los pesos $\Delta \mathbf{w}(n)$ se verán desacopladas en la expresión matricial anterior:

$$E[\Delta w_i(n)] = [1 - \mu \lambda_i]^n E[\Delta w_i(0)]$$

convirtiéndose en una expresión escalar. Lo que sería deseable es que dicha solución tendiese a 0 cuando n se hace muy grande, lo cual se conseguirá si $|1 - \mu \lambda_i| < 1 \quad \forall \lambda_i$ autovalor de \mathbf{R} . Esta condición sobre el tamaño del paso equivale a decir:

- i) Si $\Re\{\lambda_i\} > 0$ entonces se debe cumplir que $0 < \mu < 2\Re\{1/\lambda_i\}$.
- ii) Si $\Re\{\lambda_i\} < 0$ entonces se debe cumplir que $0 > \mu > 2\Re\{1/\lambda_i\}$.

Como para que todo el vector de error en los pesos converja a cero lo tienen que hacer cada una de sus componentes, es necesario que la matriz $\Re\{\mathbf{R}\}$ tenga signo definido. Así, si es definida positiva se necesita un $\mu > 0$ y si es definida negativa se necesita un $\mu < 0$.

Esta condición es importantísima e impone una restricción bastante fuerte sobre los sistemas que pueden ser resueltos mediante GLMS tal y como está formulado aquí. Pero además de imponer que converja la media, es necesario, que el vector de error en los pesos converja en media cuadrática: es necesario que los estimadores tengan una varianza finita para que sean útiles. En el siguiente apartado se verá qué restricciones impone esta segunda condición.

8.6 Convergencia en media cuadrática a orden μ^2

Como se comentó al analizar el algoritmo LMS, una forma de asegurar su convergencia era imponiendo que la cantidad $D(n) = E[\Delta \mathbf{w}^H(n) \Delta \mathbf{w}(n)]$, llamada desviación del error al cuadrado, que representa la varianza de los estimadores de los pesos, convergiera. Por tanto el interés se centra ahora en ver cómo evoluciona $D(n)$ con el tiempo y en particular calcular el valor al que converge.

La principal hipótesis utilizada hasta el momento, no enunciada expresamente pero que puede englobar al resto, ha sido la de suponer el tamaño del paso μ pequeño³. Por tanto, sería admisible que cualquier cálculo que se haga, se tomara como un desarrollo en serie en torno a $\mu = 0$.

Como se verá a continuación, en el estudio de la convergencia en media cuadrática, sólo aparecen términos en μ y μ^2 . A la vista del comentario anterior y

³Esto se verá más claro en los comentarios al final de capítulo.

para adquirir un mejor conocimiento del funcionamiento de GLMS, se trabajará a orden μ^2 en los tres siguientes apartados y a orden μ en el apartado 8.7.

8.6.1 Convergencia en media cuadrática. Promediado directo

Se parte de la ecuación (8.10) reproducida aquí por conveniencia:

$$\Delta \mathbf{w}(n) = [\mathbf{I} - \mu \tilde{\mathbf{x}}(n) \mathbf{x}^H(n)] \Delta \mathbf{w}(n-1) + \mu \tilde{\mathbf{x}}(n) e_0^*(n) \quad (8.11)$$

como ya se ha comentado constituye una ecuación en diferencias no homogénea con coeficiente aleatorio. Al tener los coeficientes no deterministas, su resolución se hace más complicada. Se puede obviar esta dificultad aplicando una simplificación conocida como método de promediado directo. Dicho método consiste, recuérdese, en suponer que la variación de la componente estocástica va a ser pequeña entorno a la media (en este caso equivale a decir que μ es pequeña), por lo que se puede sustituir por su valor medio. Con esto (8.11) se convierte en:

$$\Delta \mathbf{w}(n) = [\mathbf{I} - \mu \mathbf{R}] \Delta \mathbf{w}(n-1) + \mu \tilde{\mathbf{x}}(n) e_0^*(n)$$

Para calcular $\mathbf{D}(n)$ se construye, como siguiente paso, la matriz de correlación del error de los pesos:

$$\begin{aligned} \Delta \mathbf{w}(n) \Delta \mathbf{w}^H(n) &= [\mathbf{I} - \mu \mathbf{R}] \Delta \mathbf{w}(n-1) \Delta \mathbf{w}^H(n-1) [\mathbf{I} - \mu \mathbf{R}]^H + \\ &+ [\mathbf{I} - \mu \mathbf{R}] \Delta \mathbf{w}(n-1) \mu e_0(n) \tilde{\mathbf{x}}^H(n) + \mu \tilde{\mathbf{x}}(n) e_0^*(n) \Delta \mathbf{w}^H(n-1) [\mathbf{I} - \mu \mathbf{R}]^H + \\ &+ \mu^2 |e_0(n)|^2 \tilde{\mathbf{x}}(n) \tilde{\mathbf{x}}^H(n) \end{aligned} \quad (8.12)$$

Teniendo en cuenta las hipótesis de independencia y el principio de ortogonalidad (8.8), siguiendo un razonamiento paralelo al realizado en el análisis de LMS para obtener la ecuación (7.11), se llega sin dificultad a:

$$\mathbf{R}_\Delta(n) = [\mathbf{I} - \mu \mathbf{R}] \mathbf{R}_\Delta(n-1) [\mathbf{I} - \mu \mathbf{R}^H] + \mu^2 \sigma_0^2 \mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}} \quad (8.13)$$

donde

$$\sigma_0^2 = E[|e_0(n)|^2] \quad (8.14)$$

$$\mathbf{R}_\Delta(n) = E[\Delta \mathbf{w}(n) \Delta \mathbf{w}^H(n)] \quad (8.15)$$

$$\mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}} = E[\tilde{\mathbf{x}}(n) \tilde{\mathbf{x}}^H(n)] \quad (8.16)$$

La estadística de $\{\tilde{\mathbf{x}}(n)\}$ es estacionaria, pero no así la de $\{\Delta \mathbf{w}(n)\}$ y así se ha indicado en sus matrices de autocorrelación correspondientes, donde se muestra explícitamente su dependencia temporal.

Como se sabe, para la obtención del último sumando $\mu^2 \sigma_0^2 \mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}$ se ha aplicado la hipótesis de que las variables son conjuntamente gaussianas. En algunos

problemas concretos esto no sería necesario ya que $e_0(n)$ sería independiente del resto de variables, por ejemplo en un modelo AR con la SNR = ∞ .

La expresión (8.13) representa una ecuación en diferencias no homogénea, por lo cual no tendrá como solución $\mathbf{R}_{\Delta}(n) = 0$. Es decir, aunque el vector de error en los pesos tienda a cero en media, presenta pequeñas oscilaciones. Para el algoritmo LMS, Bucklew demostró que dichas oscilaciones eran de naturaleza gaussiana [BKS93].

Se procederá ahora a resolver (8.13). Trabajando en la base en la que \mathbf{R} es diagonal, los elementos diagonales de la matriz de autocorrelación del error en los pesos evolucionan según la siguiente expresión:

$$R_{\Delta ii}(n) = [1 - \mu\lambda_i][1 - \mu\lambda_i^*]R_{\Delta ii}(n-1) + \mu^2\sigma_0^2 R_{\bar{x}\bar{x}ii}$$

Valor en convergencia

Los parámetros más importantes de esta ecuación en diferencias se pueden calcular con la ayuda del apéndice D. Siguiendo las indicaciones allí dadas se llega a que el valor en convergencia vale:

$$R_{\Delta ii}(\infty) = \frac{\mu}{2\Re\{\lambda_i\} - \mu|\lambda_i|^2} \sigma_0^2 R_{\bar{x}\bar{x}ii}$$

A partir de aquí ya se puede calcular la magnitud de interés. Recordando que la traza de una matriz es invariante ante cambios de base, se cumple que:

$$\begin{aligned} D(\infty) &= \frac{\mu}{2} \sigma_0^2 \sum_{i=1}^p \frac{R_{\bar{x}\bar{x}ii}}{\Re\{\lambda_i\} - \frac{\mu}{2} |\lambda_i|^2} = \\ &= \frac{\mu}{2} \sigma_0^2 \text{Tr} \left\{ \frac{\mathbf{R}_{\bar{x}\bar{x}}}{\Re\{\mathbf{R}\} - \frac{\mu}{2} \mathbf{R}^H \mathbf{R}} \right\} \end{aligned} \quad (8.17)$$

Tiempos de convergencia

De forma análoga, siguiendo las indicaciones del apéndice D, se pueden calcular los tiempos de convergencia asociados a la parte homogénea de la ecuación en diferencias anterior o a toda ella. En este capítulo será de utilidad el segundo de ellos, cuya expresión viene dada en (D.6) y se reproduce aquí adaptada a esta situación por conveniencia:

$$\tau_i = \frac{\ln \left(\frac{R_{\Delta ii}(1)/e - R_{\Delta ii}(\infty)}{R_{\Delta ii}(1) - R_{\Delta ii}(\infty)} \right)}{\ln |1 - \mu\lambda_i|^2} + 1 \quad (8.18)$$

Esta es la constante de tiempo asociada al elemento diagonal $R_{\Delta ii}$, la constante de tiempo asociada a la magnitud $D(n)$, τ , se puede calcular como un promedio de ella. Obviamente para aquellos problemas en los que sólo haya una incógnita las expresiones para $D(\infty)$ y τ se simplifican enormemente:

$$D(\infty) = \frac{\mu}{2\Re\{\lambda\} - \mu|\lambda|^2} \sigma_0^2 \sigma_{\tilde{x}}^2 \quad (8.19)$$

$$\tau = \frac{\ln\left(\frac{D(1)/e - D(\infty)}{D(1) - D(\infty)}\right)}{\ln|1 - \mu\lambda|^2} + 1 \quad (8.20)$$

8.6.2 Convergencia en media cuadrática. Hipótesis de gaussianidad

Se parte de (8.11), que es exacta ya que no se ha hecho ninguna aproximación para llegar a ella, y se construye la matriz de autocorrelación del vector de error en los pesos \mathbf{R}_{Δ} . Primero el producto externo:

$$\begin{aligned} \Delta \mathbf{w}(n) \Delta \mathbf{w}^H(n) = & [\mathbf{I} - \mu \tilde{\mathbf{x}}(n) \mathbf{x}^H(n)] \Delta \mathbf{w}(n-1) \Delta \mathbf{w}^H(n-1) [\mathbf{I} - \mu \mathbf{x}(n) \tilde{\mathbf{x}}^H(n)] + \\ & + \mu^2 |e_0^2(n)|^2 \tilde{\mathbf{x}}(n) \tilde{\mathbf{x}}^H(n) + \\ & + \mu [\mathbf{I} - \mu \tilde{\mathbf{x}}(n) \mathbf{x}^H(n)] \Delta \mathbf{w}(n-1) \tilde{\mathbf{x}}^H(n) e_0(n) + \\ & + \mu e_0^*(n) \tilde{\mathbf{x}}(n) \Delta \mathbf{w}^H(n-1) [\mathbf{I} - \mu \mathbf{x}(n) \tilde{\mathbf{x}}^H(n)] \end{aligned} \quad (8.21)$$

y ahora se toma valor esperado, teniendo en cuenta las hipótesis de independencia. Según estas hipótesis, los dos últimos sumandos son cero. Esto se puede justificar desde dos puntos de vista:

1. Por ser $e_0(n)$ independiente del resto y de media nula, como en los modelos AR.
2. Como $\Delta \mathbf{w}(n-1)$ es independiente del resto, al tomar valor esperado y dejar evolucionar el algoritmo tiende a cero, ya que como se demostró con anterioridad $E[\Delta \mathbf{w}(n-1)] \rightarrow 0$

El resto de términos conduce a:

$$\begin{aligned} \mathbf{R}_{\Delta}(n) = & \mathbf{R}_{\Delta}(n-1) - \mu \mathbf{R} \mathbf{R}_{\Delta}(n-1) - \mu \mathbf{R}_{\Delta}(n-1) \mathbf{R}^H + \\ & + \mu^2 E[\tilde{\mathbf{x}}(n) \mathbf{x}^H(n) \Delta \mathbf{w}(n-1) \Delta \mathbf{w}^H(n-1) \mathbf{x}(n) \tilde{\mathbf{x}}^H(n)] + \\ & + \mu^2 \sigma_0^2 \mathbf{R}_{\tilde{x}\tilde{x}} \end{aligned} \quad (8.22)$$

Para calcular el valor esperado aquí indicado se supondrá que la variable aleatoria $[\tilde{\mathbf{x}}^H(n) | \mathbf{x}^H(n)]^H$ es conjuntamente gaussiana (última hipótesis de independencia). En general no lo será, pero puede suponer una primera aproximación

para cualquier otro tipo de dependencia. Con esta hipótesis, se tiene para el elemento ik -ésimo con $\Delta w(n-1)_j = \Delta_j$:

$$\begin{aligned} E[\tilde{x}_i x_j^* \Delta_j \Delta_l^* x_l \tilde{x}_k^*] &= E[\tilde{x}_i x_j^* x_l \tilde{x}_k^*] R_{\Delta jl} = \\ &= (E[\tilde{x}_i x_j^*] E[x_l \tilde{x}_k^*] + E[\tilde{x}_i \tilde{x}_k^*] E[x_j^* x_l]) R_{\Delta jl} = \\ &= (R_{ij} R_{lk}^H + R_{\bar{x}\bar{x}ik} R_{xjl}) R_{\Delta jl} \end{aligned} \quad (8.23)$$

Donde se ha aplicado el teorema de factorización gaussiana. Por tanto:

$$\begin{aligned} \mathbf{R}_{\Delta}(n) &= \mathbf{R}_{\Delta}(n-1) - \mu \mathbf{R} \mathbf{R}_{\Delta}(n-1) - \mu \mathbf{R}_{\Delta}(n-1) \mathbf{R}^H + \\ &+ \mu^2 \mathbf{R} \mathbf{R}_{\Delta}(n-1) \mathbf{R}^H + \mu^2 \text{Tr}\{\mathbf{R}_{xx} \mathbf{R}_{\Delta}(n-1)\} \mathbf{R}_{\bar{x}\bar{x}} + \\ &+ \mu^2 \sigma_0^2 \mathbf{R}_{\bar{x}\bar{x}} \end{aligned} \quad (8.24)$$

que supone la aparición de un nuevo término en relación con la aproximación de promediado directo. Este nuevo término impedirá que los elementos diagonales de $\mathbf{R}_{\Delta}(n)$ se desacoplen en la base en la que \mathbf{R} es diagonal y dificulta la obtención de una expresión compacta para la evolución de $D(n)$ y para $D(\infty)$.

Para el caso en el que se tenga una sola incógnita, se obtiene inmediatamente:

$$D(n) = [1 - 2\mu \Re\{\lambda\} + \mu^2 |\lambda|^2 + \mu^2 \sigma_x^2 \sigma_x^2] D(n-1) + \mu^2 \sigma_x^2 \sigma_0^2$$

de esta ecuación en diferencias se puede obtener la información habitual:

Valor en convergencia.

Recurriendo de nuevo a las expresiones del apéndice D y teniendo en cuenta la ecuación en diferencias anterior, se llega a:

$$D(\infty) = \frac{\mu \sigma_x^2}{2\Re\{\lambda\} - \mu |\lambda|^2 - \mu \sigma_x^2 \sigma_x^2} \sigma_0^2$$

Tiempo de convergencia.

Con este valor de $D(\infty)$ se construye la constante de tiempo correspondiente, empleando también las expresiones del apéndice D:

$$\tau = \frac{\ln\left(\frac{D(1)/e - D(\infty)}{D(1) - D(\infty)}\right)}{\ln(1 - 2\mu \Re\{\lambda\} + \mu^2 |\lambda|^2 + \mu^2 \sigma_x^2 \sigma_x^2)} + 1$$

8.6.3 Convergencia en media cuadrática. Tercera aproximación

En este apartado se obtendrá la evolución del MSD y su valor de convergencia sin tener en cuenta las técnicas anteriores de promediado directo o factorización gaussiana. El procedimiento es el mismo que el llevado a cabo en el apartado anterior y comparte los pasos hasta la ecuación (8.22). Una vez ahí, se calculará el valor esperado de forma exacta:

$$\mathbf{R}_\Delta(n) = \mathbf{R}_\Delta(n-1) - \mu \mathbf{R} \mathbf{R}_\Delta(n-1) - \mu \mathbf{R}_\Delta(n-1) \mathbf{R}^H + \mu^2 \mathbf{R}_\Delta(n-1) \Gamma + \mu^2 \sigma_0^2 \mathbf{R}_{\bar{x}\bar{x}} \quad (8.25)$$

donde se ha definido

$$\Gamma_{ij}^{kl} = E[\tilde{x}_i x_k^* x_l \tilde{x}_j^*]$$

y el siguiente producto

$$(\mathbf{R}_\Delta(n-1)\Gamma)_{ij} = \mathbf{R}_\Delta(n-1)_{kl} \Gamma_{ij}^{kl}$$

exceptuando las hipótesis de independencia usuales dadas en el apartado 8.5, esta aproximación es la exacta, porque no incorpora ninguna nueva hipótesis simplificatoria. El problema que conlleva es análogo al del enunciado anterior: los elementos diagonales de $\mathbf{R}_\Delta(n)$ no se desacoplan en la base en la que \mathbf{R} es diagonal, con lo que el cálculo de $D(n)$ se complica enormemente.

Obviamente el caso más sencillo es aquél en el que sólo se tuviera una sola incógnita. Entonces quedaría:

$$D(n) = (1 - 2\mu \Re\{\lambda\} + \mu^2 \Gamma) D(n-1) + \mu^2 \sigma_0^2 \sigma_{\bar{x}}^2$$

y la información que se obtiene de ella es:

Valor en convergencia.

$$D(\infty) = \frac{\mu \sigma_{\bar{x}}^2}{2\Re\{\lambda\} - \mu \Gamma} \sigma_0^2$$

Tiempo de convergencia

De igual modo como se procedía en los apartados anteriores, se obtiene que

$$\tau = \frac{\ln\left(\frac{D(1)/e - D(\infty)}{D(1) - D(\infty)}\right)}{\ln(1 - 2\mu \Re\{\lambda\} + \mu^2 \Gamma)} + 1$$

8.7 Convergencia en media cuadrática a orden μ

Las tres aproximaciones anteriores tienen en común los términos en μ del coeficiente de la ecuación en diferencias. Si se trabajase a orden μ , se tendría la misma ecuación dinámica en los tres casos. El otro término del mismo orden, $\mu^2\sigma_0^2\mathbf{R}_{\bar{x}}$, no se puede despreciar porque es muy importante en la dinámica de la evolución de \mathbf{R}_{Δ} ya que es el término independiente.

Por tanto, simplificando (8.13) se tiene

$$\mathbf{R}_{\Delta}(n) = \mathbf{R}_{\Delta}(n-1) - \mu\mathbf{R}\mathbf{R}_{\Delta}(n-1) - \mu\mathbf{R}_{\Delta}(n-1)\mathbf{R}^H + \mu^2\sigma_0^2\mathbf{R}_{\bar{x}\bar{x}}$$

De esta última ecuación se partirá ahora para discutir un par de aspectos más. Trabajando de nuevo en la base en la cual \mathbf{R} es diagonal:

$$R_{\Delta ii}(n) = R_{\Delta ii}(n-1) - \mu\lambda_i R_{\Delta ii}(n-1) - \mu\lambda_i^* R_{\Delta ii}(n-1) + \mu^2\sigma_0^2 R_{\bar{x}\bar{x}ii}$$

o equivalentemente

$$R_{\Delta ii}(n) = [1 - 2\mu\Re\{\lambda_i\}]R_{\Delta ii}(n-1) + \mu^2\sigma_0^2 R_{\bar{x}\bar{x}ii} \quad (8.26)$$

para que la solución de la parte homogénea de la ecuación anterior se anule (y no diverja) a tiempos grandes, es necesario imponer (ver apéndice D)

$$|1 - 2\mu\Re\{\lambda_i\}| < 1$$

que se traduce en la siguiente condición sobre el valor del tamaño del paso:

$$\text{i) Si } \Re\{\lambda_i\} > 0 \Rightarrow 0 < \mu < \frac{1}{\Re\{\lambda_i\}}$$

$$\text{ii) Si } \Re\{\lambda_i\} < 0 \Rightarrow 0 > \mu > \frac{1}{\Re\{\lambda_i\}}$$

Notar que la convergencia en media impone distintas condiciones sobre μ que la convergencia de la desviación del error al cuadrado.

Bajo estas condiciones la ecuación en diferencias anterior alcanza un valor en convergencia y un tiempo de convergencia dado por:

Valor en convergencia

Cada uno de los elementos de la diagonal de $\mathbf{R}_{\Delta}(n)$ converge a:

$$R_{\Delta ii}(\infty) = \mu \frac{\sigma_0^2}{2} \frac{R_{\bar{x}\bar{x}ii}}{\Re\{\lambda_i\}}$$

la suma de todos estos valores $\forall i$ es la desviación del error al cuadrado en convergencia:

$$D(\infty) = \mu \frac{\sigma_0^2}{2} \sum_{i=1}^p \frac{R_{\tilde{x}\tilde{x}ii}}{\Re\{\lambda_i\}}$$

como la traza no depende de la base en que se tome, si se deshace el cambio de base que diagonalizó a \mathbf{R} , la expresión anterior equivale a

$$D(\infty) = \mu \frac{\sigma_0^2}{2} \text{Tr}\{\mathbf{R}_{\tilde{x}\tilde{x}}(\Re\{\mathbf{R}\})^{-1}\} \quad (8.27)$$

Para el caso particular en el que $x(n) = \tilde{x}(n) \in \mathbb{R}$, como ocurre en el LMS con variables reales, se llega a

$$D(\infty) = \mu \frac{p}{2} \sigma_0^2$$

que viene a decir que la varianza de los estimadores de \mathbf{w}_0 en convergencia, aumentan con el ruido de medición y con el número de parámetros a estimar. Asimismo, indica que un tamaño del paso pequeño disminuye dicha varianza. Este resultado se puede obtener del análisis del LMS ya realizado en el tema anterior, haciendo μ pequeña en la expresión (7.14).

Tiempo de convergencia

El tiempo asociado a cada uno de los elementos de la diagonal de $\mathbf{R}_{\Delta}(n)$ es:

$$\tau_i = \frac{\ln\left(\frac{R_{\Delta ii}(1)/e - R_{\Delta ii}(\infty)}{R_{\Delta ii}(1) - R_{\Delta ii}(\infty)}\right)}{\ln(1 - 2\mu\Re\{\lambda_i\})} + 1$$

El tiempo de convergencia para $D(n)$ se podría obtener promediando los tiempos de la expresión anterior $\forall i$.

8.8 Unificación de resultados

Una vez que se han obtenido las expresiones que rigen el comportamiento dinámico de $D(n)$ (o en su caso el de $R_{\Delta ii}(n)$), y se ha caracterizado su solución, se van a presentar en este apartado de manera conjunta para poder llevar a cabo una comparación de manera más sencilla.

Lo que sigue se ha extraído del apéndice D. En las distintas aproximaciones utilizadas se ha llegado a una ecuación en diferencias de orden 1 no homogénea:

$$m(n) = cm(n-1) + a \quad (8.28)$$

cuya solución tiene una expresión explícita dada por

$$m(n) = c^n [m(0) - m(\infty)] + m(\infty) \quad (8.29)$$

con

$$m(\infty) = \frac{1}{1-c} a \quad (8.30)$$

y $m(0)$ es la condición inicial.

Además de $m(\infty)$, se suele tomar como parámetro importante de un algoritmo adaptativo su tiempo de convergencia. El tiempo de convergencia τ se define como el tiempo necesario para que $m(n)$ caiga a $1/e$ de su valor inicial:

$$c^\tau [m(0) - m(\infty)] + m(\infty) = \frac{m(0)}{e}$$

despejando directamente se obtiene:

$$\tau = \frac{\ln \left(\frac{m(0)/e - m(\infty)}{m(0) - m(\infty)} \right)}{\ln(c)} + 1 \quad (8.31)$$

normalmente el tiempo de convergencia es mucho mayor que 1, por lo que se puede despreciar el segundo sumando.

Por tanto, tal y como se ha hecho en los apartados anteriores, dada una ecuación en diferencias y por consiguiente los parámetros c y a correspondientes, se puede obtener la evolución de la magnitud estudiada mediante (8.29), el valor al que converge mediante (8.30) y el tiempo de convergencia correspondiente mediante (8.31). Como ya es conocido, el valor de c y a depende de la aproximación utilizada. La tabla 8.3 recoge los valores de los parámetros c y a obtenidos en las distintas aproximaciones empleadas.

	c	a
orden μ^2		
promediado directo	$1 - 2\mu\Re\{\lambda_i\} + \mu^2 \lambda_i ^2$	$\mu^2\sigma_0^2 R_{\bar{x}\bar{x}ii}$
factorización gaussiana	$1 - 2\mu\Re\{\lambda\} + \mu^2 \lambda ^2 + \mu^2\sigma_x^2\sigma_{\bar{x}}^2$	$\mu^2\sigma_0^2\sigma_{\bar{x}}^2$
aprox. exacta	$1 - 2\mu\Re\{\lambda\} + \mu^2\Gamma$	$\mu^2\sigma_0^2\sigma_{\bar{x}}^2$
orden μ		
aproximación a orden μ	$1 - 2\mu\Re\{\lambda_i\}$	$\mu^2\sigma_0^2 R_{\bar{x}\bar{x}}$

Tabla 8.3 *Parámetros característicos de las aproximaciones utilizadas en el análisis de la convergencia en media cuadrática*

Nótese que en la tabla 8.3 en aquellas aproximaciones donde se ha conseguido desacoplar la evolución temporal de cada elemento de la diagonal de \mathbf{R}_Δ , se ha notado expresamente el autovalor i -ésimo y el elemento diagonal de $\mathbf{R}_{\bar{x}\bar{x}}$ correspondiente. En el resto se ha considerado que sólo se tiene una incógnita.

A la vista de la tabla 8.3 se puede afirmar que $D(\infty)$ es proporcional a μ . Cuando tienda a cero, éste también tenderá a cero. Por tanto, se podrá despreciar respecto a $D(1)$ en (8.31) (suponiendo que ahora la magnitud $m(n)$ sea $D(n)$) y el denominador se podrá aproximar por $\ln(c) = c - 1$. En definitiva se tendría:

$$\tau = \frac{1}{1 - c} \quad (8.32)$$

esta expresión simplificada se puede tomar como válida cuando el tamaño del paso sea muy pequeño, es decir, cuando esté justificado trabajar a orden μ . En ese caso, se tendría que $R_{\Delta ii}$ tiene un tiempo de convergencia dado por

$$\tau_i = \frac{1}{2\mu\Re\{\lambda_i\}} \quad (8.33)$$

que tiene la forma típica de los tiempos de convergencia dada para otros algoritmos como el LMS. Y exáctamente como a él le pasaba, si se quiere conseguir con el GLMS un $D(\infty)$ pequeño, el tiempo de convergencia será grande.

8.9 Evolución temporal de la desviación del error al cuadrado

El principal objetivo de este apartado es caracterizar la evolución dinámica de $D(n)$ cuando se está en el caso de trabajar con más de una incógnita. Con lo visto en el apartado anterior sólo se estaba en disposición de dar esta evolución en el caso de tener una sola incógnita o como mucho de dar la evolución de los elementos diagonales de \mathbf{R}_Δ . Ya que son estos elementos diagonales los que sí verifican la ecuación (8.28). Cada uno de ellos convergerá a su correspondiente valor en el infinito y tendrá su propio tiempo de convergencia. La suma de todos ellos (por definición el MSD), tenderá a la suma de los valores en convergencia individuales y el tiempo de convergencia será un promedio de los tiempos de convergencia individuales.

Se trabajará, exclusivamente y a título de ejemplo, con las expresiones obtenidas trabajando a orden μ . Según se vio, el valor de convergencia para MSD venía dado por (8.27), y según la discusión del párrafo anterior junto con (8.33), se puede aproximar el tiempo de convergencia total por:

$$\tau = \frac{1}{2\mu} \frac{1}{p} \sum_{i=1}^p \frac{1}{\Re\{\lambda_i\}}$$

O lo que es lo mismo:

$$\tau = \frac{1}{2\mu} \frac{1}{p} \text{Tr}\{\Re\{\mathbf{R}\}^{-1}\}$$

Por regla general se suele tomar $\mathbf{w}(1) = \mathbf{0}$, por lo que $D(1) = \mathbf{w}_0^H \mathbf{w}_0$. Con estos datos y suponiendo una forma funcional tal y como (8.29), haciendo uso de (8.32), se llega a que la evolución temporal de $D(n)$ se rige por la siguiente expresión:

$$D(n) = D(\infty) + (D(1) - D(\infty))(1 - 1/\tau)^{n-1}$$

y sustituyendo los valores correspondientes:

$$\begin{aligned} D(n) = & \left[\mathbf{w}_0^H \mathbf{w}_0 - \mu \frac{\sigma_0^2}{2} \text{Tr}\{\mathbf{R}_{\bar{x}\bar{x}} (\Re\{\mathbf{R}\})^{-1}\} \right] \left[1 - 2\mu \frac{p}{\text{Tr}\{(\Re\{\mathbf{R}\})^{-1}\}} \right]^{n-1} + \\ & + \mu \frac{\sigma_0^2}{2} \text{Tr}\{\mathbf{R}_{\bar{x}\bar{x}} (\Re\{\mathbf{R}\})^{-1}\} \end{aligned} \quad (8.34)$$

recuérdese que sólo es válida para valores de μ pequeños. Un tratamiento análogo se puede hacer para el caso de la aproximación de promediado directo, donde las componentes de $\mathbf{R}_\Delta(n)$ también se desacoplan.

Con este procedimiento lo que se ha conseguido es obtener una expresión para la evolución dinámica del MSD para el caso de tener varias incógnitas (i.e. $p > 1$). El caso $p = 1$ se resuelve de forma directa. Quedan sin solución, sin embargo, la aproximación de factorización gaussiana y la aproximación exacta para el caso de $p > 1$.

8.10 Aplicación de GLMS

Como ejemplo de funcionamiento de este algoritmo se va a identificar un modelo AR conociendo sólo la salida del sistema. Se tomará como modelo AR complejo aquel que venga definido por la siguiente ecuación en diferencias:

$$x(n) = - \sum_{i=1}^p a_i^* x(n-i) + w(n) \quad (8.35)$$

Este sistema se caracterizará por su estadística de cuarto orden (i.e. cumulantes de cuarto orden). Los cumulantes de cuarto orden para señales complejas admiten distintas definiciones, aquí se ha optado⁴ por

$$C_{4x}(\tau_1, \tau_2, \tau_3) = \text{Cum}[x^*(n), x(n+\tau_1), x(n+\tau_2), x(n+\tau_3)]$$

⁴ Siguiendo la notación de [NP93] en su capítulo 2 (p.e ecuación (2.8)).

es decir, complejo conjugando sólo la primera variable aleatoria.

Se pretende, por tanto, obtener una ecuación que relacione los cumulantes de orden 4 de la salida con los coeficientes del modelo AR. Para ello considérese el cumulante $C_{4x}(-t, -t, -t)$ y sustitúyase la variable que está complejo conjugada por su expresión (8.35), es decir:

$$\begin{aligned} C_{4x}(-t, -t, -t) &= \text{Cum}[x^*(n), x(n-t), x(n-t), x(n-t)] = \\ &= -\text{Cum}\left[\sum_{i=1}^p a_i x^*(n-i) + w(n), x(n-t), x(n-t), x(n-t)\right] \end{aligned} \quad (8.36)$$

Utilizando primero la propiedad de linealidad de los cumulantes y después sabiendo que $w(n)$ es independiente de $x(n-t)$ (por tanto $C_{4x}[w(n), x(n-t), x(n-t), x(n-t)] = 0$), se obtiene

$$-\sum_{i=1}^p a_i C_{4x}(t-i) = C_{4x}(t) \quad t > 0 \quad (8.37)$$

donde

$$\begin{aligned} C_{4x}(t) &= C_{4x}(-t, -t, -t) = \text{mom}[x^*(n)x^3(n-t)] - \\ &\quad - 3\text{mom}[x^*(n)x(n-t)]\text{mom}[x^2(n-t)] \equiv \\ &\equiv \text{mom}[x^*(n)x^3(n-t)] - 3\sigma_{2x}^c \text{mom}[x^*(n)x(n-t)] \end{aligned} \quad (8.38)$$

Si se quiere resolver (8.37) mediante GLMS hay que construir un sistema de ecuaciones a partir de ella que se pueda poner en la forma de (8.8). Para ello basta con definir de la siguiente manera las variables involucradas:

$$\tilde{\mathbf{x}}(n) = [x^3(n-1), x^3(n-2), \dots, x^3(n-p)]^T - 3\sigma_{2x}^c [x(n-1), x(n-2), \dots, x(n-p)]^T \quad (8.39)$$

$$\mathbf{x}(n) = [x(n-1), x(n-2), \dots, x(n-p)]^T \quad (8.40)$$

$$d(n) = x(n) \quad (8.41)$$

donde $\sigma_{2x}^c = E[x(n)x(n)]$ se puede calcular aparte con los datos disponibles. Con estas definiciones se verifica que $-E[\tilde{\mathbf{x}}(n)\mathbf{x}^H(n)]\mathbf{a} = E[\tilde{\mathbf{x}}(n)d^*(n)]$.

Comparándolo con el desarrollo teórico de GLMS:

1. $\mathbf{w}_0 = -\mathbf{a}$ con $\mathbf{a} = [a_1, a_2, \dots, a_p]^T$.
2. $e_0(n) = d(n) - \mathbf{w}_0^H \mathbf{y}(n) = w(n)$.

Aquí se entiende un poco mejor porqué a $d(n)$ se le llama respuesta deseada, es decir, respuesta que se quiere obtener como combinación lineal de los elementos de $\mathbf{x}(n)$. Y porqué a $e_0(n)$ se le suele llamar error de medición.

8.11 Aplicación numérica. Ejemplo

Sea el modelo AR= $[1 \ 0.4 + 0.4i] = [1 \ a_1]$, por lo que sólo habrá que estimar un parámetro (aunque complejo), por tanto $p = 1$. El modelo se excitará mediante una serie, $w(n)$, i.i.d de variables aleatorias, donde cada una de ellas constituye una constelación 4-QAM de valores $[1+2i, 1-2i, -1+2i, -1-2i]$ de media nula.

En este caso la ecuación (8.37) se reduce a $-a_1 C_{4x}(0) = C_{4x}(1)$ y en consecuencia $\mathbf{R} = C_{4x}(0)$ es una matriz 1×1 , con autovalor $\lambda = C_{4x}(0) = 29.6887 - 3.0401i$. El resto de parámetros necesarios son:

1. $\sigma_x^2 = E[\tilde{x}(n)\tilde{x}^*(n)] = 908.391$
2. $\sigma_x^2 = E[x(n)x^*(n)] = 9.615$.
3. $\sigma_w^2 = \sigma_0^2 = 5$.
4. $\Gamma = E[x(n)\tilde{x}^*(n)\tilde{x}(n)x^*(n)] = 1.2e4$.

Este ejemplo se va a estudiar desde el punto de vista teórico con las 4 aproximaciones presentadas anteriormente. El principal objetivo es reproducir la evolución de MSD con el tiempo, para ello se utilizarán las expresiones (8.28), (8.29), (8.30) y la tabla 8.3. Como aplicación particular se intentará reproducir el tiempo de convergencia experimental mediante (8.31) y el valor en convergencia de MSD mediante (8.30).

8.11.1 Presentación de los resultados. Evolución de MSD con el tiempo

Mediante Promediado Directo (PD)

Por ser el primer caso a tratar se hará con un poco más de detalle. De (8.29), (8.30) y la tabla 8.3 resulta:

$$D(n) = D(\infty) + (D(0) - D(\infty))(1 - 2\mu\Re\{\lambda\} + \mu^2|\lambda|^2)^{n-1}$$

donde

$$D(\infty) = \frac{\mu\sigma_x^2}{2\Re\{\lambda\} - \mu|\lambda|^2}\sigma_0^2$$

recuérdese además que según la inicialización usada en el algoritmo (tabla 8.2), resulta que $D(0) = |-0.4 + 0.4i|^2$. Los valores numéricos de cada cantidad se han dado previamente. Con esta expresión se obtiene la evolución del MSD en

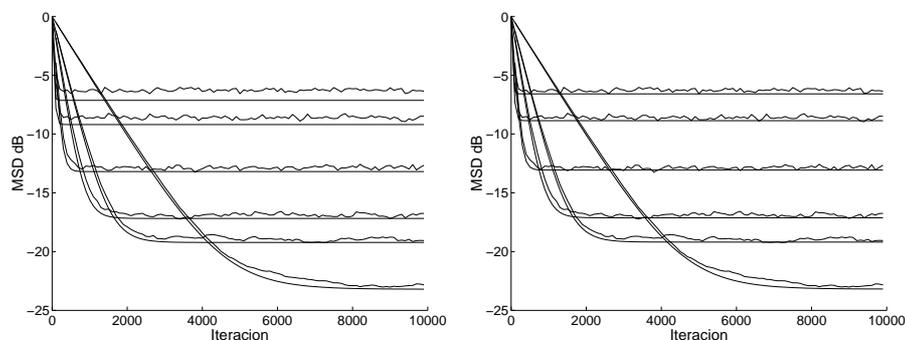
función del tamaño del paso μ . La comparación, para $\mu = 8e - 4$, $4e - 4$, $2e - 4$, $8e - 5$, $5e - 5$, y $2e - 5$, con las simulaciones realizadas con el ordenador se pueden observar en la figura 8.1(a) (las simuladas son las más irregulares).

Notar que en la gráfica se ha representado el MSD experimental definido y calculado de la siguiente manera:

$$\text{MSD}(n) = \frac{1}{n^{\circ}\text{reali.}} \sum_{r=1}^{n^{\circ}\text{reali.}} \frac{|\mathbf{w}^{(r)}(n) - \mathbf{w}_0|^2}{|\mathbf{w}_0|^2}$$

donde $\mathbf{w}^{(r)}(n)$ es el estimador obtenido en la iteración n -ésima de la realización r -ésima, por tanto, según las definiciones dadas aquí se cumple que coincide con $\text{MSD}(n) = D(n)/D(0)$, ya que se ha tomado $D(0) = |\mathbf{w}_0|^2$ (recordad la inicialización dada en la tabla 8.2). Donde $D(n)$ es la cantidad calculada teóricamente. En todas las simulaciones realizadas (y mientras no se diga lo contrario), el número de realizaciones ha sido de 1000 y la longitud de las series estudiadas de 10000 (como se aprecia en la gráfica).

El $\text{MSE}(n)$ ⁵ presenta una primera parte en la que se produce la convergencia, conforme va aprendiendo, en la que va disminuyendo hasta alcanzar su valor más bajo constante: su valor en convergencia, lo que se conoce por $D(\infty)$. Se observa que conforme va disminuyendo el tamaño del paso, tarda más en converger pero llega a un valor de $D(\infty)$ menor; los detalles específicos de este comportamiento se discutirán más adelante.



(a) Aproximación PD

(b) Aproximación FG

Figura 8.1 Evolución del MSD

⁵MSD=MSE a efectos prácticos y se emplearán indistintamente.

Mediante Factorización Gaussiana (FG)

Se vuelve a recurrir a las ecuaciones dadas en (8.29) y (8.30), pero utilizando los parámetros de la segunda línea de la tabla 8.3. Así se obtiene la evolución teórica del MSD, dada en la figura 8.1(b) junto con la evolución obtenida en las simulaciones. En un primera inspección parece muy similar a la figura 8.1(a) y así lo es. Las ‘pequeñas’ diferencias se estudiarán más adelante.

Mediante la aproximación exacta (AE)

Para trabajar con la última aproximación a orden μ^2 se utilizan los parámetros de la tercera línea de la tabla 8.3, junto con las ecuaciones dadas en (8.29) y (8.30). Las representaciones gráficas correspondientes a los distintos valores de μ y los resultados obtenidos en las simulaciones, se muestran en la figura 8.2(a).

Mediante la aproximación a orden μ (OMU)

Finalmente se utiliza la última línea de la tabla 8.3. De nuevo se representan conjuntamente las gráficas teóricas y experimentales, para facilitar las comparaciones, en la figura 8.2(b).

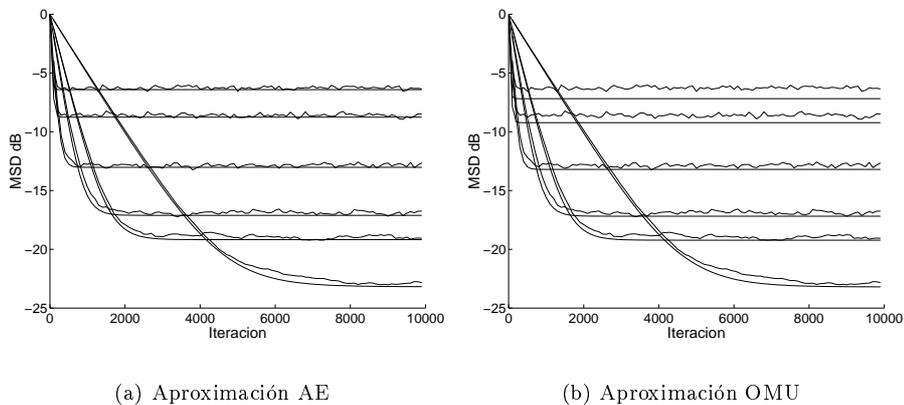


Figura 8.2 Evolución del MSD

8.11.2 Presentación de los resultados. Tiempo de convergencia

Mediante Promediado Directo

De igual manera se puede obtener el tiempo de convergencia teórico con las expresiones (8.31) y (8.30) y primera línea de la tabla 8.3. La comparación con el experimental se muestra en la figura 8.3(a) para distintos valores de μ . Éste se ha calculado tomando la iteración, en gráficas similares a las de las figuras 8.1(a) a 8.2(b) para varios valores del tamaño del paso, para la cual el MSE valía $-4.34 \text{ dB} = 10 * \log(1/e)$, es decir, $1/e$ veces su valor inicial (que es de 0 dB). Los resultados de la simulación se representan mediante cuadrados y los teóricos mediante una línea continua.

La tendencia es clara. Cuanto mayor es el tamaño del paso μ , mayor es la velocidad de convergencia. En principio la concordancia teoría-simulación es excelente, pero se verá más adelante que presenta ciertas peculiaridades que merecen ser comentadas.

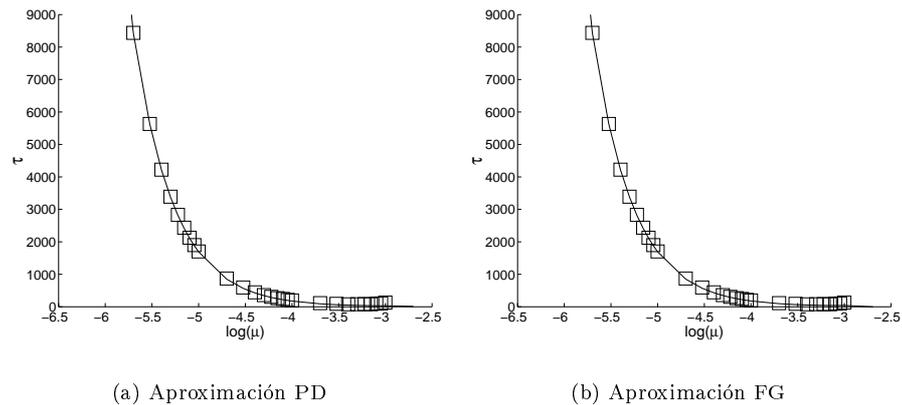


Figura 8.3 Variación del tiempo de convergencia con el tamaño del paso

Mediante FG, AE y OMU

Para calcular el tiempo de convergencia mediante el resto de aproximaciones se utilizan los parámetros especificados en la tabla 8.3, donde a cada aproximación le corresponde una línea distinta convenientemente indicada, y las

ecuaciones (8.31) y (8.30). Los resultados teóricos junto con los experimentales se muestran en las figuras 8.3(b) a 8.4(b).

Con este grado de detalle no se aprecian diferencias esenciales entre las aproximaciones utilizadas. Sólo quizás se entrevean para μ grande, aunque ya se comentará más adelante donde corresponda.

Como se aprecia en las gráficas, los datos experimentales se cortan para $\mu = 10^{-3}$, lo cual quiere decir que para μ mayores no llega nunca a los -4.34 dB que marcan que se ha llegado al tiempo de convergencia. En cuanto a los teóricos, se ha dibujado el tiempo de convergencia para todos los valores de μ que hacían que la expresión (8.31) fuese positiva (es decir, diera un tiempo de convergencia factible). Obviamente esta μ de corte depende de la aproximación utilizada. Más tarde se continuará comentando sobre esto.

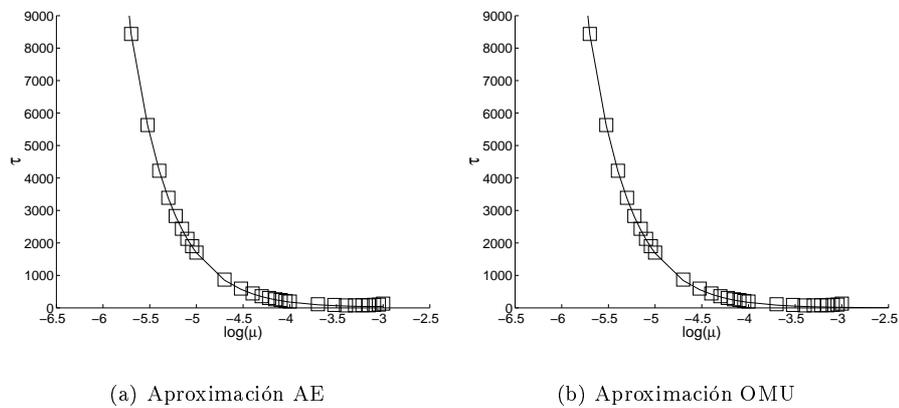


Figura 8.4 Variación del tiempo de convergencia con el tamaño del paso

8.11.3 Presentación de los resultados. MSD en convergencia

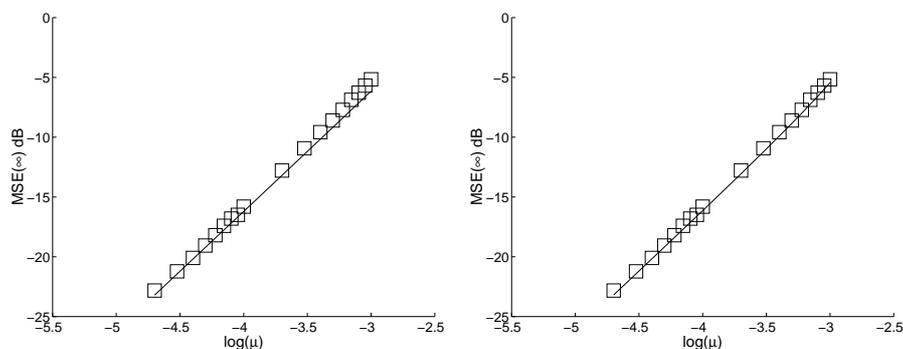
Mediante Promediado Directo

La ecuación (8.30) da el valor al que converge la recursión (8.28), que es el valor que alcanza el MSD cuando ha convergido, valor que interesa ahora para estudiarlo. Este valor depende de las variables c y a , que para el caso particular de promediado directo tienen su expresión en la primera línea de la tabla 8.3,

como ya se sabe de apartados anteriores. Esta vez sólo se han calculado los valores teóricos para valores de μ de los que se disponen resultados de las simulaciones.

De esta manera se calcula el valor teórico; el experimental (simulado) se ha obtenido promediando los últimos 500 datos de las gráficas experimentales para MSE como las que aparecen en las figuras 8.1(a) a 8.2(b) para un amplio rango de valores de μ . Ambos resultados, teórico y simulado, se muestran en la figura 8.5(a) en función de μ .

Como se puede observar en la gráfica los resultados de las simulaciones (representados por cuadrados) muestran una dependencia casi lineal entre $MSE(\infty)$ y el tamaño del paso. Las predicciones teóricas siguen muy de cerca esta tendencia, aunque presenta algunas discrepancias con las simulaciones que se discutirá en próximas secciones.



(a) Aproximación PD

(b) Aproximación FG

Figura 8.5 Variación del MSD en convergencia con el tamaño del paso

Mediante FG, AE y OMU

Utilizando de nuevo la ecuación (8.30), junto con el valor correspondiente de las variables c y a para cada aproximación a utilizar, se han obtenido las gráficas correspondientes de la variación de $MSE(\infty)$ en función de μ . Además se han vuelto a representar los resultados de las simulaciones para facilitar las comparaciones. Todo ello se muestra en las figuras 8.5(b) a 8.6(b).

El comportamiento general de los resultados teóricos coincide, a grandes

rasgos, con el descrito en el apartado anterior. Las diferencias que pudieran existir se comentarán más adelante.

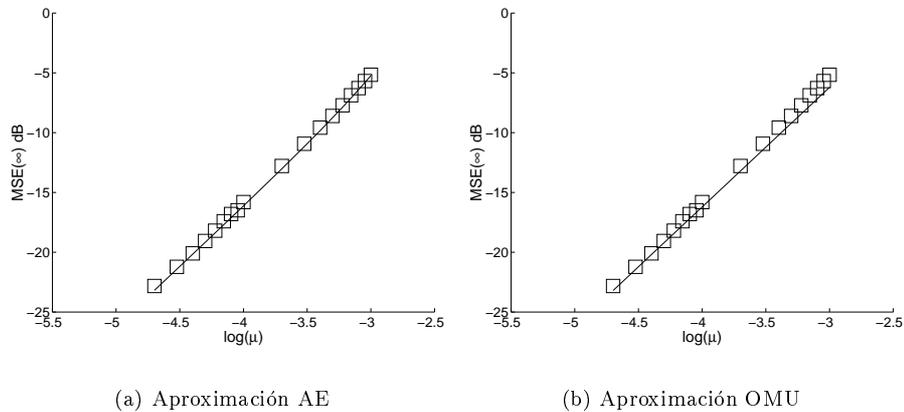


Figura 8.6 Variación del MSD en convergencia con el tamaño del paso

8.11.4 Comentario y discusión

Evolución dinámica de $D(n)$

En este apartado se discutirá cualitativamente cómo de bien se ajustan los resultados teóricos a los experimentales. La discusión cuantitativa se hará en los apartados siguientes mediante dos magnitudes que caracterizan el comportamiento de $D(n)$: una referente al comportamiento dinámico propiamente dicho (el tiempo de convergencia) y otra referente a la fase estacionaria (el $MSE(\infty)$).

En general y para todas las aproximaciones teóricas utilizadas la concordancia con los resultados simulados es bastante buena, según puede desprenderse de una inspección visual de las figuras 8.1 (a) a 8.2 (b). De todas maneras, y para poder hacer una discusión con un poco más de fundamento, sería conveniente fijarse con un poco más de detalle en aquellas zonas de $D(n)$ que pudieran ser interesantes. Para ello se representa gráficamente $D(n)$, para todas las aproximaciones utilizadas y los valores de μ especificados, en su zona de convergencia y las primeras iteraciones cuando ha convergido, figura 8.7 y figura 8.8. Nótese las diferencias de escala.

Lo primero que cabe destacar es que el ajuste teoría-simulaciones no es

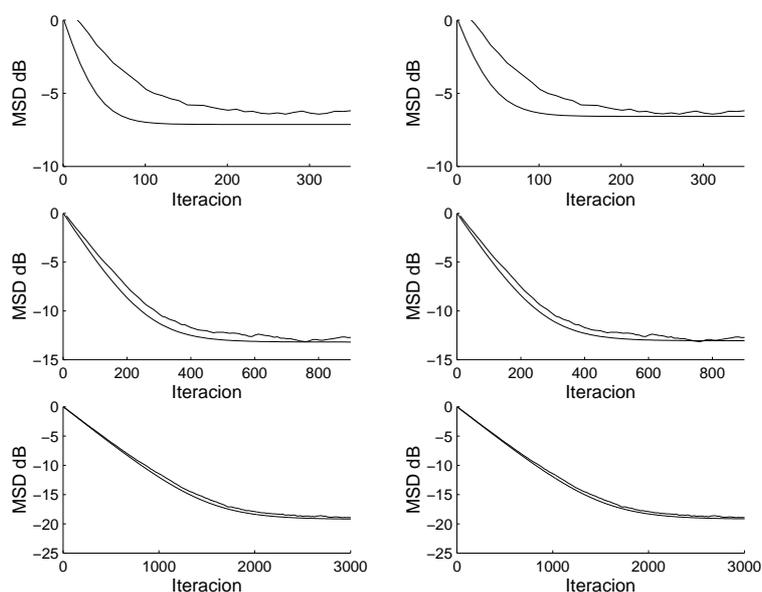


Figura 8.7 Detalles ampliados de la evolución del MSD. Aprox. PD(izq) y FG(dcha). Primera fila $\mu = 8e - 4$, segunda fila $\mu = 2e - 4$ y tercera fila $\mu = 5e - 5$.

tan bueno como podría pensarse antes y que ahora empiezan a notarse ciertos matices que cabe la pena destacar. En general, independientemente de la aproximación, la concordancia es mejor conforme disminuye el tamaño del paso, de manera que para $\mu = 5e - 5$ se podría decir que es casi perfecta para todas ellas. El motivo de tal comportamiento hay que buscarlo en los pasos comunes que se han dado para obtener las aproximaciones. Los pasos comunes se dan al principio, y el más fundamental de todos es la utilización de las hipótesis de independencia, apartado 8.6. Más concretamente interesa comentar la conclusión que se obtiene de ellas y que figuraba en el apartado 8.5, reproducida aquí por conveniencia:

‘A la vista de (8.10) se desprende que $w(n)$ sólo depende de muestras $\{d(n)\}$, $\{x(n)\}$ y $\{\tilde{x}(n)\}$ hasta tiempo n y que por tanto es independiente de $x(n + 1)$, $\tilde{x}(n + 1)$ y $d(n + 1)$ por las 3 primeras hipótesis de independencia’.

No cabe duda que esta consecuencia de las hipótesis de independencia se ha mostrado sumamente útil para simplificar matemáticamente todos los cálculos. Pero las 3 primeras hipótesis de independencia no siempre se cumplen exactamente porque siempre existirá cierta dependencia entre las variables involucradas en el problema y por tanto, las expresiones que se han obtenido basadas en

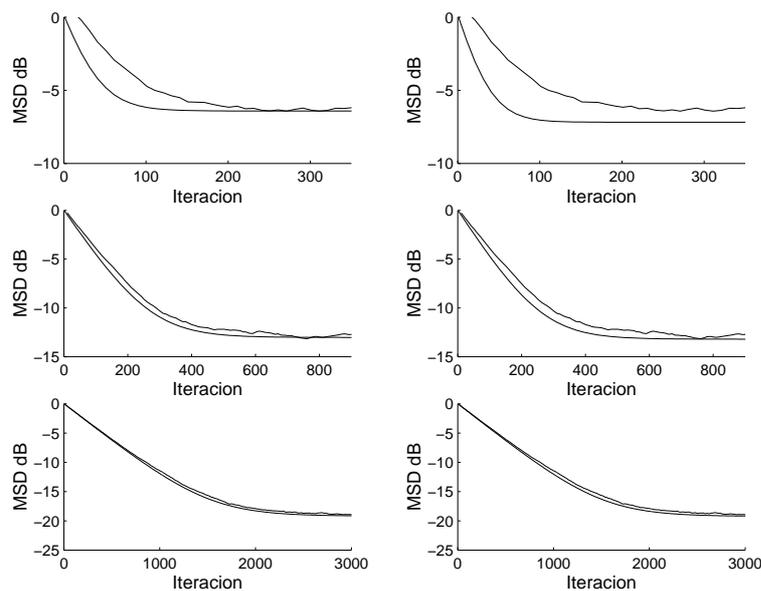


Figura 8.8 Detalles ampliados de la evolución del MSD. Aprox. $AE(izq)$ y $OMU(dcha)$. Primera fila $\mu = 8e - 4$, segunda fila $\mu = 2e - 4$ y tercera fila $\mu = 5e - 5$.

ellas no son plenamente exactas.

El mayor o menor grado de dependencia de las variables que intervienen en el problema no se puede modificar porque viene dado, el único parámetro que se tiene de libre elección es el tamaño del paso. Pues bien, si se hace suficientemente pequeño, gracias a la expresión (8.10) se puede conseguir ‘artificialmente’ que se cumpla mejor la independencia de $w(n)$ con el resto de variables aleatorias. Si se verifican mejor las hipótesis utilizadas en la obtención de las ecuaciones, mejores resultados darán. Por tanto, al disminuir el tamaño del paso mejor concordancia teoría-simulación se obtiene.

A la vista de la discusión anterior no cabe duda de que el parámetro μ va a ser crucial en la interpretación de todos los resultados que se obtengan. Si se han estudiado las expresiones teóricas como un desarrollo en torno a μ , es de esperar por tanto que aquellas que incluyan un mayor número de órdenes den un resultado más acertado. A continuación se analiza esto último: las aproximaciones a orden μ dan peor resultado que las de orden μ^2 . Que AE y FG dan mejor resultado que OMU no tiene discusión, pero a simple vista parece que PD y OMU dan igual resultado. Todos los resultados teóricos dan un MSE ligeramente menor que el verdadero (simulado), si el que da PD es mayor que

el que da OMU, el primero será una mejor aproximación. Para comprobar esto se ha representado la diferencia MSE(OMU) con MSE(PD), el resultado se muestra en la figura 8.11.4 para $\mu = 8e - 4$. Al ser la diferencia negativa, MSE(PD) es mayor y en consecuencia es mejor aproximación. En definitiva se ha comprobado que es más exacto trabajar con las aproximaciones a orden μ^2 , al menos en lo que a la evolución de $D(n)$ se refiere.

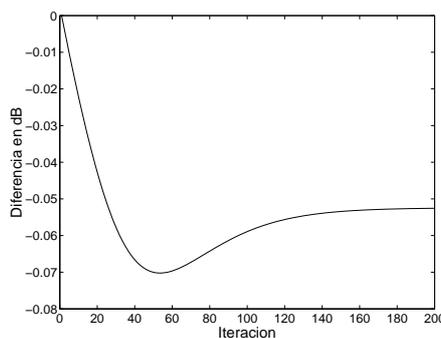


Figura 8.9 Diferencia entre el valor del MSD para OMU y PD en las primeras iteraciones

De todas maneras las diferencias entre las aproximaciones utilizadas disminuyen al disminuir el tamaño del paso como era previsible.

Se continúa ahora con el estudio del subgrupo de aproximaciones a orden μ^2 más específicamente: afortunadamente se obtienen mejores resultados conforme más elaborada es la aproximación utilizada. Las cosas salen como cabía esperar. Como primera aproximación se optó por sustituir parte de la componente estocástica por una determinista (su media), constituyendo el PD. El procedimiento anterior se mejoró suponiendo variables conjuntamente gaussianas, mejorando ostensiblemente los resultados, FG. Por último se calculan exactamente los valores esperados que aparecen y quedan como únicas hipótesis simplificadoras las hipótesis de independencia. A la vista de las gráficas se puede afirmar lo siguiente: FG supone una gran mejora respecto a PD, pero es un poco peor que AE, es decir, AE mejora menos a FG en proporción que FG a PD. Para μ pequeñas la diferencia desaparece y hay gran concordancia entre teoría y simulaciones. En definitiva: resultados lógicos y dentro de lo que cabía esperar.

Con todas las aproximaciones utilizadas se ha llegado a que el MSE tiene una forma funcional dada por,

$$D(n) = D(\infty) + (D(0) - D(\infty))c^{n-1} \quad (8.42)$$

con

$$D(\infty) = \frac{a}{1-c}$$

$D(0)$ es la condición inicial, independiente de la aproximación utilizada. c y a son los parámetros que determinan la función; su valor depende de la aproximación (ver tabla 8.3).

Por tanto, no hay libertad en la forma funcional sino sólo en los parámetros que la definen. Esta restricción en la forma viene impuesta por la mismísima ecuación que define la recursión de actualización de los pesos de la tabla 8.2, es decir, como ecuación en diferencias. El mayor o menor éxito que se tenga en la predicción de los datos de las simulaciones va a depender en la posibilidad de convertir (8.21) en una ecuación en diferencias una vez tomado valor esperado. Es decir, si (8.42) como ecuación en diferencias que es va a ser un buen modelo de (8.21). Para poder convertir (8.21) hacen falta dos cosas:

- a) Que el primer miembro de la parte derecha de (8.21) se pueda poner como un coeficiente que multiplica a $\mathbf{R}_\Delta(n-1)$.
- b) Que se puedan eliminar los dos últimos miembros de la parte derecha de (8.21).

La condición a) se puede cumplir exigiendo que $\Delta\mathbf{w}(n)$ sea independiente del resto de variables aleatorias, o equivalentemente que lo sea $\mathbf{w}(n)$. Sobre esto ya se ha hablado anteriormente.

En principio la condición b) es directa al trabajar en la identificación de modelos AR, ya que $e_0(n)$ coincide con el ruido de excitación del modelo AR que es independiente del resto de variables que aparecen y de media 0. Por tanto, al tomar valor esperado debería anularse como se indica en el comentario después de (8.21). Sin embargo, en la práctica, los valores esperados se implementan mediante promedios (es decir, se estima el valor esperado) y puede que no sea 0 exactamente. Si no es cero exactamente habrá que tener en cuenta el resto de factores que aparecen en estos miembros, en particular $\Delta\mathbf{w}(n)$, el error en los pesos. Este error va disminuyendo conforme se conocen más datos al avanzar las iteraciones, pero en los primeros pasos es comparativamente más grande, lo que contribuirá a engordar los miembros mencionados. En definitiva no se conseguirá una ecuación en diferencias. Idealmente, $\Delta\mathbf{w}(n)$ tiende a cero cuando n es suficientemente grande, con lo que en ese caso los miembros se anulan. El problema por tanto está en las primeras iteraciones, ya que se usa un estimador del valor esperado, por lo que en principio cuantas más iteraciones promedie mejor estimador tendré. Para comprobar esto último se ha llevado a cabo una simulación en la que se promedian 5000 realizaciones para $\mu = 8e - 4$ con la idea de que los datos simulados se acerquen más a los esperados teóricamente. La gráfica (no reproducida aquí) resultante como evolución de $D(n)$ sigue el mismo trazado que para el caso en el que se promediaron 1000 realizaciones, pero un poco más suave. No parece por tanto que los dos últimos miembros de (8.21) sean los culpables de la diferencia teoría-simulación.

En definitiva se puede concluir que las condiciones a) y b) reseñadas anteriormente se cumplirán peor para μ grande y en las primeras etapas de la convergencia. Este hecho se constata por mera inspección de las figuras 8.7 y 8.8.

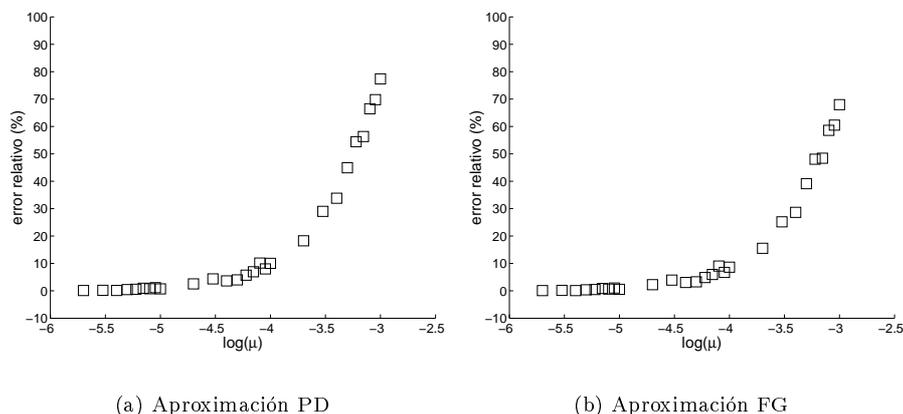


Figura 8.10 Error relativo cometido al reproducir el tiempo de convergencia

8.11.5 Tiempo de convergencia

En apartados anteriores, cuando se presentaron los resultados de las simulaciones correspondientes al tiempo de convergencia se afirmó que eran ajustados con bastante exactitud por las expresiones teóricas obtenidas. Ahora es el momento de cuantificar cómo de bien lo ajustan. No debe extrañar que conforme se aumenta el grado de detalle se descubran aspectos nuevos, pero procédase paso a paso. En primer lugar se mostrará el error relativo cometido en la estimación del tiempo de convergencia, para los valores de μ de los que se disponga datos de las simulaciones y para todas las aproximaciones, figuras 8.10(a) a 8.11(b). El error relativo se ha definido tomando como valor ‘verdadero’ de la magnitud el valor obtenido de las simulaciones, es decir:

$$\varepsilon_r = \frac{\tau(\text{teórico}) - \tau(\text{exper.})}{\tau(\text{exper.})} \times 100$$

El comportamiento del error relativo en los cuatro casos es análogo: error casi nulo para tamaños del paso pequeños, aumentando progresivamente, hasta el 70%, conforme aumenta el tamaño del paso con un comportamiento muy

ordenado. Esta tendencia no debe extrañar demasiado a la luz de los comentarios hechos en el apartado anterior: se esperan mejores resultados cuanto más pequeño sea μ .

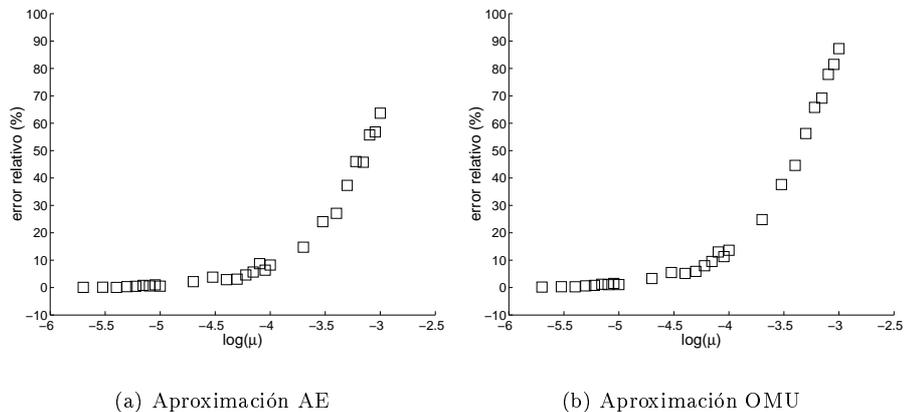


Figura 8.11 Error relativo cometido al reproducir el tiempo de convergencia

Para comparar las distintas aproximaciones se va a tomar como valor representativo de cada una su error relativo más alto. De esta manera se tienen los siguientes valores:

PD 78%.
FG 68%.
AE 64%.
OMU 86%.

Es decir, sigue la misma tendencia observada en la evolución de $D(n)$, los resultados mejoran conforme se mejora la aproximación teórica. Y además en la misma proporción: el paso cuantitativamente más grande se da entre PD y FG (que rebaja el error relativo máximo en un 10 %).

Se puede inspeccionar con más detalle la zona de interés mediante las figuras 8.12(a) a 8.13(b). Estas últimas figuras, además de confirmar el distanciamiento entre teoría y simulaciones al aumentar el tamaño del paso, son muy ricas en matices y merece la pena detenerse en ellas un poco más. Pero antes sería conveniente, en este apartado y para el tiempo de convergencia, volver a la pregunta ¿Por qué se obtiene un error tan grande para μ altas? Por dos motivos, el primero de los cuales es conocido y comentado previamente: cuanto mayor es μ peor se verifican las hipótesis de independencia, ya que, según la expresión de

actualización de los pesos de la tabla 8.2, más dependencia habrá entre $\mathbf{w}(n)$ y el resto de variables aleatorias involucradas. Por otro lado, τ es evidentemente una magnitud dinámica, en tanto que para su cálculo se necesita que las ecuaciones teóricas utilizadas se cumplan en los primeros pasos de convergencia. En estos pasos $\epsilon(n)$ ⁶ es grande y por tanto no se cumplen las hipótesis de independencia como se ha indicado antes. En resumen, como

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \mu \tilde{\mathbf{x}}(n) e^*(n) \quad (8.43)$$

para obtener una buena independencia el segundo sumando de (8.43) tiene que ser lo más pequeño posible. Esto se puede conseguir o haciendo μ pequeña o haciendo $\epsilon(n)$ pequeña (dejando evolucionar el algoritmo suficientemente de manera que empiece a aprender y disminuya el error).

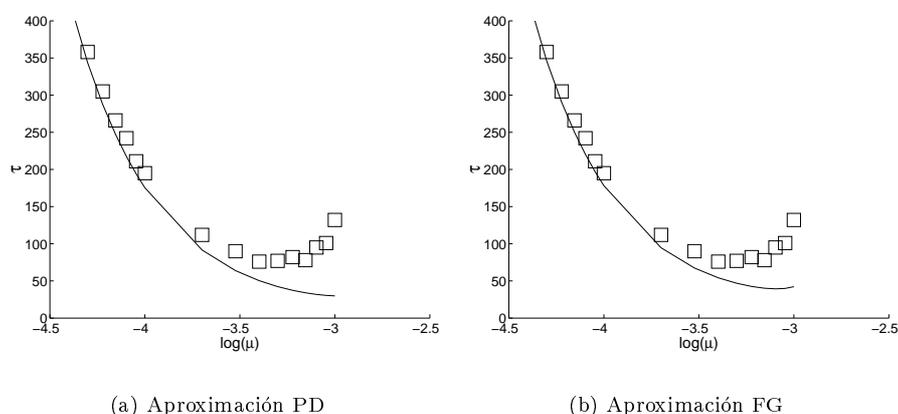


Figura 8.12 Zona de interés al reproducir el tiempo de convergencia

Resumiendo, en las primeras iteraciones donde hay que calcular τ , se presentan 2 factores que hacen que la independencia de $\mathbf{w}(n-1)$ con $\tilde{\mathbf{x}}(n)$ y $\mathbf{x}(n)$ no se cumplan: $\epsilon(n)$ es grande y (posiblemente) μ es grande. Y por tanto las expresiones para τ pueden no ser apropiadas.

Con todo lo comentado en este y en el anterior apartado, es bueno recalcar y unificar las conclusiones en una sola frase: Las hipótesis de independencia se cumplen mejor cuando μ es pequeño y n grande⁷.

A continuación se discutirán las características del tiempo de convergencia

⁶Es similar a decir que $\Delta\mathbf{w}(n)$ es grande, como se comentó en la página 217 a tenor de las condiciones a) y b).

⁷Sería conveniente dar algún tipo de intervalo dentro del cual esto se verifique.

que pueden apreciarse del estudio de las figuras 8.12(a) a 8.13(b): el μ de corte y el tiempo de convergencia mínimo.

Las aproximaciones a orden μ^2 no sólo mejoran a la de orden μ cuantitativamente sino también cualitativamente, ya que son capaces de dar respuesta y reproducir algunas características del comportamiento de τ en función de μ , como es el μ de corte. Primero se describirá qué se entiende por μ de corte: conforme se aumenta el tamaño del paso la velocidad de convergencia aumenta, es decir, disminuye el tiempo de convergencia tal y como puede observarse en el conjunto de figuras 8.3(a) a 8.4(b) y 8.12(a) a 8.13(b). Pero a la vez que esto ocurre el valor al que converge se va haciendo cada vez más grande (como se discutirá en el próximo apartado), con el riesgo de que llegue un momento en el que para cierto valor de μ $\text{MSE}(n)$ no alcance los -4.34 dB, es decir, no llegue al $1/e$ de su valor inicial y por tanto no se tenga definido el tiempo de convergencia correspondiente. Pues bien, a este valor del tamaño del paso se le denomina μ de corte, μ_c . Según las simulaciones para este ejemplo concreto se obtiene que $\mu_c = 1e - 3$. Sería muy deseable por tanto que las expresiones teóricas que se han obtenido fuesen capaces de reproducir este valor. Teóricamente se calcula como aquel valor de μ a partir del cual todos los μ mayores dan un tiempo de convergencia negativo o complejo (distinto de un real positivo). Precisamente lo que se ha dibujado (figuras 8.12(a) a 8.13(b)) para cada aproximación, para el rango de μ estudiado, son todos los tiempo de convergencia que son físicamente realizables. Se observa que mientras OMU acepta como válidos valores menores que $1e - 3$ para el tamaño del paso, PD, FG y AE no y además coinciden el valor de μ_c calculado con el simulado. También hay que resaltar que aunque se sea capaz de reproducir este valor, no coinciden el tiempo de convergencia teórico y simulado para μ_c , como ya se sabe de comentarios anteriores.

Otra peculiaridad de la función $\tau(\mu)$ merece ser descataada ahora. Se ha dicho con anterioridad que es una función decreciente, o al menos para hacer los comentarios nos ha bastado con eso. De todas maneras se comprueba por simple inspección de la figura 8.12(a) (p.e.) que hay un mínimo, es decir, no por mucho aumentar el tamaño del paso se va a conseguir que converja más rápido, hay un valor μ_m a partir del cual ese aumento se ve penalizado⁸. Sólo FG y AE son capaces de reproducir esta característica y dar un mínimo, además en el sitio correspondiente.

8.11.6 Error cuadrático medio en convergencia

Como ya es conocido, el algoritmo adaptativo irá dando, conforme conoce más datos (conforme aumentan las iteraciones) un MSE menor, hasta que llega un momento en el que deja de aprender (cuando ha convergido). Este MSE

⁸Otro fenómeno no lineal ya estudiado para LMS es la existencia de un mínimo en $\text{MSE}(\infty)$ como función de μ [RZ99].

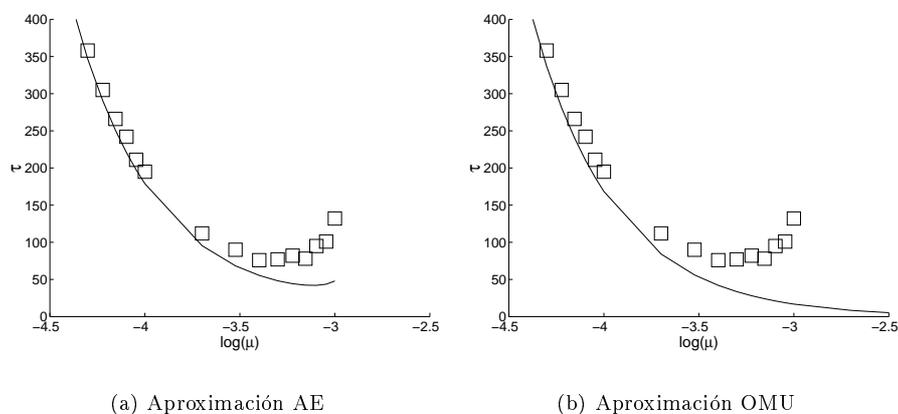


Figura 8.13 Zona de interés al reproducir el tiempo de convergencia

en convergencia, o $MSE(\infty)$ es lo que está representado en las figuras 8.5(a) a 8.6(b). Para cuantificar un poco qué aproximación es mejor se ha procedido, igual que en el apartado anterior, a calcular el error relativo de la misma manera. Dicho error se muestra en las figuras 8.14(a) a 8.15(b).

Lo primero que hay que hacer notar es la diferencia de escala entre estas figuras y las figuras 8.10(a) a 8.11(b), ya que el error relativo es mucho menor para el intervalo de μ estudiado. Esto en principio era de esperar ya que se está trabajando en una zona donde se supone que las hipótesis de independencia se verifican mejor, en la zona de convergencia donde por definición n es grande (idealmente infinito).

Los comentarios de los resultados en función de la aproximación utilizada son iguales a los del apartado anterior y una aproximación mejora a la otra en el mismo orden: OMU, PD, FG y AE.

Lo que más extraña es que el error relativo no dependa de una manera clara de μ . En principio se podría esperar que disminuyera al disminuir el tamaño del paso a tenor de todos los comentarios hechos hasta ahora. Es decir, se podría esperar un comportamiento parecido en forma al error relativo para el tiempo de convergencia, porque en definitiva ambas expresiones (la de τ y la de $D(\infty)$) se derivan de la misma expresión teórica.

Algunas posibles explicaciones:

- a) La función $D(n)$ simulada es irregular en convergencia y no mientras con-

verge, por lo tanto va a existir cierto error en datos experimentales para $D(\infty)$ y no para el tiempo de convergencia.

- b) $D(\infty)$, como su propia notación indica, es el valor de la función $D(n)$ cuando $n \rightarrow \infty$. Obviamente mediante simulaciones no ‘se alcanza el infinito nunca’, sólo es posible aproximarse a él. El problema es que el infinito a veces está cerca y otras veces está lejos: para situaciones donde μ sea grande, como convergerá rápido, el infinito está cerca (dentro del límite de 10000 iteraciones que contemplan las simulaciones hechas aquí), pero cuando sea pequeño estará más lejos (claramente más lejano que las 10000 iteraciones que marcan el límite del horizonte). De nuevo, por tanto, se tienen datos experimentales no todo lo exactos que deberían ser para poder poner un poco de orden en las caóticas gráficas de error relativo aquí mostradas.

A la vista de estas dos explicaciones parece que se está echando la culpa a los datos de las simulaciones, como si fueran los únicos responsables. No es que las expresiones teóricas obtenidas den resultados exactos, simplemente se espera ver alguna tendencia en los errores relativos, algo que no sucede. De todas maneras considérese que la tendencia tendría que ser muy sutil ya que los errores relativos caen en un intervalo con unas pocas unidades (entre el 1% y el 9%), de tal manera que casi se podría decir que es constante comparado con el rango que presenta para el tiempo de convergencia (del 1% al 90%). Quizás se esté pidiendo demasiado y esos efectos se vean sólo si se trabaja con series mucho más largas ($\geq 10^4$ ó 10^5 datos?) y más realizaciones que promediar (≥ 5000 quizás?), pero eso llevaría varios meses para un PC actual.

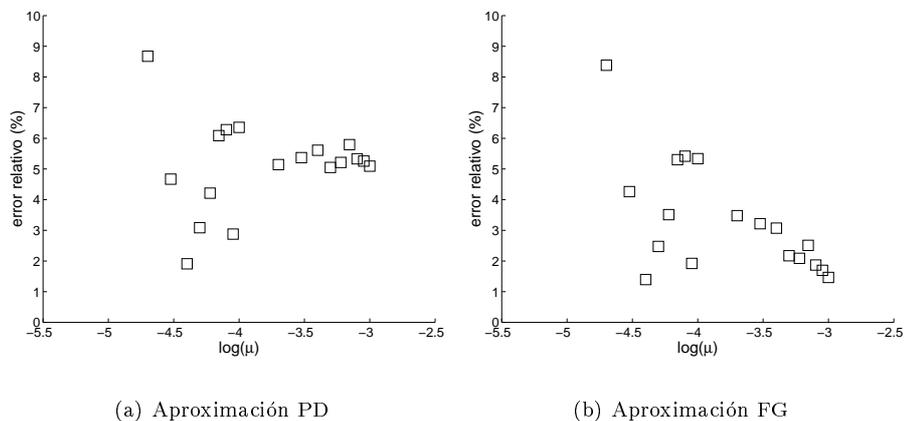


Figura 8.14 Error relativo al reproducir el MSD en convergencia

Si se concluye que el error relativo, para todas las aproximaciones, para $D(\infty)$ es constante, ¿Por qué lo es? Se podría aventurar que se contrarrestan dos efectos. Por un lado está la tendencia que se viene defendiendo de que si el tamaño del paso disminuye también disminuye el error. Y por otro está el nuevo efecto que se ha apuntado para el MSE en convergencia de que como le lleva mucho más a MSE converger efectivamente a μ pequeñas, no se cuenta con datos realmente fiables en ese rango (no queda más remedio que discutir en función de las simulaciones de las que se disponen).

Hasta ahora, en este apartado, sólo se han hecho comentarios generales que afectan a las 4 aproximaciones a la vez; simplemente se ha dicho que la tendencia es la conocida de apartados anteriores: cuanto más elaborada es la aproximación, mejor resultados da. Y aquí sí es verdad que conforme aumenta μ AE da mucho mejor resultado que OMU (por ejemplo y como era de esperar). Simplemente comentar un aspecto más. Los resultados de las simulaciones para $D(\infty)$ como función de μ , en un primera inspección, parece que siguen una ley lineal, pero no es así. Cuando aumenta el tamaño del paso hay un punto en el que la pendiente aumenta ligeramente, de manera que $D(\infty)$ decrece más rápidamente en función de μ . Pues bien, esto sólo es capaz de reproducirlo FG y AE como se puede comprobar en las figuras 8.5(a) a 8.6(b).

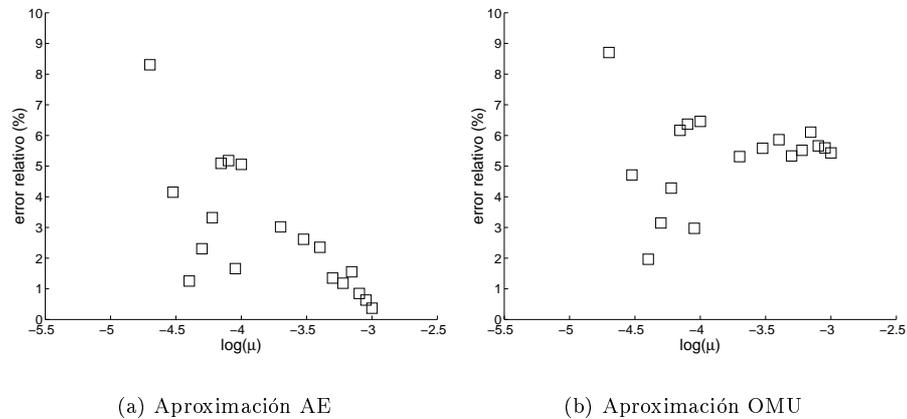


Figura 8.15 Error relativo al reproducir el MSD en convergencia

8.11.7 Comentario final

Se podría intentar sacar un poco de ventaja en la utilización del algoritmo GLMS en función de todo lo discutido en los apartados anteriores. Por ejemplo, ahora que se sabe que existe μ_m quizás se pretenda usarlo siempre de manera que el algoritmo converja lo más rápidamente posible. Esto, que supone una ventaja en cuanto a velocidad de convergencia, no lo es en cuanto a $D(\infty)$ ya que normalmente μ_m es un valor alto que hace que $D(\infty)$ también lo sea. En general:

$$D(\infty) = \frac{A\mu}{B - c\mu}$$
$$\tau \sim \frac{1}{\mu}$$

y ocurre lo siguiente:

- a) Si el tamaño del paso es grande, el tiempo de convergencia es pequeño y $D(\infty)$ grande.
- b) Si el tamaño del paso es pequeño, el tiempo de convergencia es grande y $D(\infty)$ pequeño.

Aparece por tanto la conocida oposición que se da en LMS. Hay que elegir, en función de la aplicación, entre τ bajo o $D(\infty)$ bajo, pero no pueden ser pequeños los dos a la vez.

Capítulo 9

Análisis de los algoritmos sobredeterminados y generalizados tipo LMS, OGLMS

Introducción

La principal motivación de este capítulo es comprobar que los algoritmos denominados en el capítulo 8 como OGLMS1 y OGLMS3 no sufren la fuerte limitación de GLMS de no poder trabajar con sistemas cuya matriz \mathbf{R}_{xx} asociada no sea definida. Una vez confirmado este punto se analizarán estos dos algoritmos y se elegirá el que mejor propiedades presente para profundizar en su estudio y seguir modificándolo para mejorar aún más sus propiedades. Se puede considerar por tanto éste como un capítulo de transición, que justifica por qué se ahonda en el estudio de uno y no otro algoritmo. Con lo cual se presentarán unos resultados teóricos sin simulaciones.

En este punto de la Memoria sería conveniente comentar la equivalencia entre los algoritmos tipo mínimos cuadrados y mínima media cuadrática. Por equivalencia se quiere indicar la relación que se puede establecer entre algoritmos de un tipo y otro pensados para resolver un problema concreto:

1. LMS y RLS. Ambos son capaces de imponer la condición de ortogonalidad entre el vector de datos y el error de estimación. La matriz asociada a

este problema es cuadrada y definida positiva al ser una matriz de auto-correlación.

2. GLMS y RIV. Esta vez la condición de ortogonalidad se establece entre el vector de variables instrumentales y el error de estimación. La matriz asociada sigue siendo cuadrada pero no tiene porqué ser definida, por tanto el algoritmo GLMS puede no funcionar correctamente.
3. OGLMS y ORIV. La situación es parecida a la anterior pero ahora se introducen más variables instrumentales que incógnitas tiene el problema. La matriz asociada por tanto es sobredeterminada.

Dentro de cada punto se pueden abrir más posibilidades versionando los algoritmos base, existiendo por ejemplo algoritmos LMS normalizados [Ber86], RLS rápidos [BM89] y [KCM84], los OGLMS1 y OGLMS3 a estudiar en este capítulo, etc.

El algoritmo que se estudia en primer lugar será el OGLMS1 en el apartado 9.1, donde se analizará su convergencia en media y en media cuadrática tanto en ambientes estacionarios como no estacionarios. El análisis del algoritmo OGLMS3 se deja para el apartado 9.2.

9.1 Análisis del algoritmo de gradiente estocástico basado en la 1ª opción, OGLMS1

Según se discutió en el capítulo 8, con ayuda de los resultados obtenidos en el análisis de OGLMS, la siguiente recursión resuelve el sistema (8.4):

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \frac{\mu}{n-1} \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) [z(n-1) - \Phi(n-1) \mathbf{w}(n-1)]$$

Esta expresión forma el núcleo de un algoritmo tipo gradiente estocástico, cuyos pasos se detallan en la tabla 9.1, que se construye a partir de ella tal y como se hizo en el capítulo anterior a partir del principio de ortogonalidad sobredeterminado.

<p><u>Condiciones iniciales</u></p> $\mathbf{w}(0) = 0 \quad \Phi(0) = 0 \quad z(0) = 0$
<p><u>Con la llegada de nuevos $x(n)$, $\tilde{x}(n)$ y $d(n)$ hacer</u></p> <p>Construir los vectores $\mathbf{x}(n)$ y $\tilde{\mathbf{x}}(n)$</p> $\mathbf{w}(n) = \mathbf{w}(n-1) + \frac{\mu}{n-1} \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) [z(n-1) - \Phi(n-1) \mathbf{w}(n-1)]$ $\Phi(n) = \lambda \Phi(n-1) + \tilde{\mathbf{x}}(n) \mathbf{x}^t(n)$

$$z(n) = \lambda z(n-1) + \tilde{\mathbf{x}}(n)d(n)$$

Tabla 9.1 *Proceso iterativo del algoritmo OGLMS1*

En los siguientes apartados se analizará este algoritmo, su convergencia tanto en media como en media cuadrática en ambientes estacionarios y no estacionarios.

9.1.1 Convergencia en media. Ambientes estacionarios

Siguiendo pasos análogos a los realizados para el análisis de OGLMS, la ecuación anterior se puede reescribir de la siguiente manera:

$$\begin{aligned} \Delta \mathbf{w}(n) = & \left[\mathbf{I} - \frac{\mu}{n-1} \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \Phi(n-1) \right] \Delta \mathbf{w}(n-1) + \\ & + \frac{\mu}{n-1} \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) [z(n-1) - \Phi(n-1) \mathbf{w}_0] \end{aligned} \quad (9.1)$$

Como lo que se va a estudiar son propiedades en convergencia se puede asumir en la expresión anterior que n es grande, de manera que sea válida la hipótesis de que $z(n)$ y $\Phi(n)$ son cuasideterministas. Es decir, recordando las ecuaciones (4.25) y (4.26):

$$\Phi(n) \sim n[\mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}} + \mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}^e(n)]$$

$$z(n) \sim n[\mathbf{r}_{d\tilde{\mathbf{x}}} + \mathbf{r}_{d\tilde{\mathbf{x}}}^e(n)]$$

donde las componentes de $\mathbf{R}^e(n)$ y $\mathbf{r}^e(n)$ son ruido blanco de media 0 e independientes del resto de variables aleatorias involucradas en el problema. Por lo que teniendo en cuenta $\mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}} \mathbf{w}_0 = \mathbf{r}_{d\tilde{\mathbf{x}}}$ se verifica que:

$$\begin{aligned} \Delta \mathbf{w}(n) = & \left[\mathbf{I} - \frac{\mu}{n-1} \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \Phi(n-1) \right] \Delta \mathbf{w}(n-1) + \\ & + \mu \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \mathbf{s}^e(n-1) \end{aligned} \quad (9.2)$$

se ha definido $\mathbf{s}^e(n) = \mathbf{r}^e(n) - \mathbf{R}^e(n) \mathbf{w}_0$ que contiene la información referente al ruido de estimación de las magnitudes \mathbf{r} y \mathbf{R} . Es de media nula e independiente del resto de variables aleatorias al serlo las variables a partir de las cuales se define.

Tomando valor esperado y teniendo en cuenta las hipótesis de independencia:

$$E[\Delta \mathbf{w}(n)] = [\mathbf{I} - \mu \mathbf{R}^t \mathbf{R}] E[\Delta \mathbf{w}(n-1)]$$

Si se trabaja en la base en la que $\mathbf{R}^t \mathbf{R}$ es diagonal se pueden desacoplar las componentes del vector de error en los pesos:

$$E[\Delta \mathbf{w}_i(n)] = [1 - \mu \lambda_i] E[\Delta \mathbf{w}_i(n-1)] \quad (9.3)$$

esta ecuación en diferencias tiende a cero si $|1 - \mu \lambda_i| < 1$, es decir, si

$$0 < \mu < \frac{\lambda_i}{2} \quad (9.4)$$

donde se ha tenido en cuenta que la matriz $\mathbf{R}^t \mathbf{R}$ es definida positiva¹ y por tanto sus autovalores son positivos.

Para ver mejor la influencia de los autovalores en la convergencia del algoritmo, es conveniente estudiar la solución de (9.3). Si se sustituye n veces en sí misma, se llega a:

$$E[\Delta \mathbf{w}_i(n)] = [1 - \mu \lambda_i]^n E[\Delta \mathbf{w}_i(0)]$$

cuanto más cercana a cero sea la diferencia $1 - \mu \lambda_i$ menos iteraciones serán necesarias para que la componente i -ésima del vector de error en los pesos sea cero. La velocidad de convergencia para una componente del vector se puede elevar con una adecuada elección del tamaño del paso μ . El problema está en que una buena elección para que una componente converja rápido puede ser una mala elección para otra componente (otro autovalor), incluso puede que no converja si no se cumple (9.4). Si el valor de todos los autovalores estuviera comprendido en un intervalo pequeño se podría elegir un valor del tamaño del paso que asegurara una rápida convergencia para todas las componentes. Si la dispersión de autovalores es grande, la disparidad en cuanto a velocidades de convergencia entre componentes será grande y la velocidad de convergencia total del vector vendrá dada por la velocidad de la componente que converja más despacio.

Este problema no se da en el algoritmo ORIV estudiado en el capítulo 4. Para ese algoritmo la convergencia en media se regía por la siguiente expresión (ecuación (4.37)):

$$E[\Delta \mathbf{w}(n)] = \lambda^{2n} (1 - \lambda)^2 (\mathbf{R}^t \mathbf{R})^{-1} \mathbf{\Gamma}(0) \Delta \mathbf{w}(0)$$

para adecuarla a ambientes estacionarios, donde no es necesario el factor de olvido, se hace $\lambda \rightarrow 1$ teniendo en cuenta que $(1 - \lambda) \rightarrow 1/n$:

$$E[\Delta \mathbf{w}(n)] = \frac{1}{n^2} (\mathbf{R}^t \mathbf{R})^{-1} \mathbf{\Gamma}(0) \Delta \mathbf{w}(0) \quad (9.5)$$

¹En verdad lo único que se puede asegurar es que es semidefinida positiva. Para que fuese definida positiva \mathbf{R} debería ser de rango completo y eso depende del problema concreto que se esté resolviendo.

En esta expresión se aprecia claramente que la velocidad de convergencia no depende de ninguna cantidad (p.e. autovalores de $\mathbf{R}^t \mathbf{R}$) ni de ningún parámetro libre del algoritmo (porque ORIV no tiene al ser ambientes estacionarios). Es decir, no se ve afectado por la dispersión de autovalores de $\mathbf{R}^t \mathbf{R}$. El vector de error en los pesos tiende a cero como $1/n^2$.

La forma en la que se ha conseguido reducir la carga computacional en los algoritmos presentados aquí tipo OGLMS, ha sido la de eludir el cálculo de la matriz $\mathbf{\Gamma}(n)$. Esta matriz consigue eliminar en el algoritmo ORIV cualquier dependencia en los autovalores como se ha puesto de manifiesto con las ecuaciones anteriores. Recuérdese, sin embargo, que en problemas mal condicionados ORIV no funcionaba todo lo bien que debía, no ya por un problema estructural, que se aprecia que no tiene (9.5), sino por un problema estrictamente numérico de inversión de matrices, es decir, un problema numérico en el cálculo de $\mathbf{\Gamma}(n)$. Sería muy interesante por tanto comparar el comportamiento de ORIV y OGLMS mediante simulaciones para problemas mal condicionados.

9.1.2 Convergencia en media cuadrática. Ambientes estacionarios

Se parte de la ecuación (9.2) reproducida aquí por comodidad:

$$\begin{aligned} \Delta \mathbf{w}(n) = & \left[\mathbf{I} - \frac{\mu}{n-1} \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \mathbf{\Phi}(n-1) \right] \Delta \mathbf{w}(n-1) + \\ & + \mu \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \mathbf{s}^e(n-1) \end{aligned} \quad (9.6)$$

Para simplificar los cálculos se va a despreciar el ruido de estimación $\mathbf{R}^e(n)$ respecto a \mathbf{R} en la descomposición de la matriz $\mathbf{\Phi}(n)$, aunque no así $\mathbf{s}^e(n-1)$ ya que lleva información importante en el término independiente. A continuación se construye el producto escalar del vector de error en los pesos:

$$\begin{aligned} \Delta \mathbf{w}^t(n) \Delta \mathbf{w}(n) = & \\ = & \Delta \mathbf{w}^t(n-1) [\mathbf{I} - \mu \mathbf{R}^t \tilde{\mathbf{x}}(n) \mathbf{x}^t(n)] [\mathbf{I} - \mu \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \mathbf{R}] \Delta \mathbf{w}(n-1) + \\ & + \mu^2 \mathbf{s}^{et}(n-1) \tilde{\mathbf{x}}(n) \mathbf{x}^t(n) \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \mathbf{s}^e(n-1) + \\ & + 2 \text{ términos de media } 0 \end{aligned} \quad (9.7)$$

y tomando valor esperado resulta²:

$$\begin{aligned} D(n) = \text{Tr}\{(\mathbf{I} - 2\mu \mathbf{R}^t \mathbf{R} + \mu^2 \mathbf{R}^t \mathbf{R}_{\tilde{\mathbf{x}}\mathbf{x}\tilde{\mathbf{x}}}\mathbf{R}) \mathbf{R}_{\Delta}(n-1)\} + \\ + \mu^2 \text{Tr}\{\mathbf{R}_{\tilde{\mathbf{x}}\mathbf{x}\tilde{\mathbf{x}}}\mathbf{R}_s(n)\} \end{aligned} \quad (9.8)$$

donde se han utilizado las siguientes definiciones:

$$D(n) = \text{E}[\Delta \mathbf{w}^t(n) \Delta \mathbf{w}(n)]$$

²Se ha tenido en cuenta que $\mathbf{v}^t \mathbf{M} \mathbf{v} = \text{Tr}\{\mathbf{M} \mathbf{v} \mathbf{v}^t\}$ como se comprueba fácilmente.

$$\begin{aligned}\mathbf{R} &= \text{E}[\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)] \\ \mathbf{R}_{\tilde{\mathbf{x}}\mathbf{x}\tilde{\mathbf{x}}} &= \text{E}[\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)\mathbf{x}(n)\tilde{\mathbf{x}}^t(n)] \\ \mathbf{R}_{\Delta} &= \text{E}[\Delta\mathbf{w}(n)\Delta\mathbf{w}^t(n)] \\ \mathbf{R}_s &= \text{E}[\mathbf{s}^e(n)\mathbf{s}^{e^t}(n)]\end{aligned}$$

La ecuación (9.8) no tiene una forma sencilla para su resolución, como las ecuaciones en diferencias de los casos anteriores. Se puede discutir algo sobre ella suponiendo que sólo hay una incógnita:

$$D(n) = (1 - 2\mu\lambda + \mu^2\lambda R_{\tilde{\mathbf{x}}\mathbf{x}\tilde{\mathbf{x}}})D(n-1) + \mu^2 R_{\tilde{\mathbf{x}}\mathbf{x}\tilde{\mathbf{x}}}\sigma_s^2(n) \quad (9.9)$$

Hay que tener cuidado porque aquí λ hace referencia al autovalor (al valor) de $\mathbf{R}^t\mathbf{R}$ y no al factor de olvido introducido en el producto escalar. A la vista de esta ecuación también llama la atención la influencia de dicho autovalor en la velocidad de convergencia del algoritmo en media cuadrática. Al igual que ocurría con la convergencia en media es de esperar que una aplicación con una gran dispersión de autovalores converja más lentamente. Igualmente, la influencia de la dispersión de autovalores es nula para el algoritmo ORIV en lo referente a su convergencia en media cuadrática. Véase por ejemplo la ecuación (4.50) donde el único parámetro que controla la velocidad de convergencia es el factor de olvido.

El valor al que converge la ecuación en diferencias (9.9) es

$$D(\infty) = \mu \frac{1}{\mu\lambda} \frac{\mu R_{\tilde{\mathbf{x}}\mathbf{x}\tilde{\mathbf{x}}}}{2 - \mu R_{\tilde{\mathbf{x}}\mathbf{x}\tilde{\mathbf{x}}}} \sigma_s^2(\infty)$$

del cual cabe destacar la proporcionalidad que existe entre $D(\infty)$ y:

- a) La varianza del ruido de estimación, $\sigma_s(\infty)$. Conforme se van conociendo más datos los estimadores van disminuyendo en varianza. Sin embargo la varianza no se hace cero nunca sino que tiende a un valor constante. Es una fuente de ruido para los estimadores de los pesos obvia: si las cantidades usadas en el cálculo de los pesos presentan ruido, también lo tendrán ellos.
- b) El tamaño del paso μ . Cuanto menor sea, menor error cuadrático medio en convergencia habrá.

Es posible reescribir esta última expresión de la siguiente manera:

$$D(\infty) = \mu \frac{R_{\tilde{\mathbf{x}}\mathbf{x}\tilde{\mathbf{x}}}}{\lambda} \frac{1}{2 - \mu R_{\tilde{\mathbf{x}}\mathbf{x}\tilde{\mathbf{x}}}} \sigma_s^2(\infty)$$

y como μ se suele elegir bastante pequeño:

$$D(\infty) = \mu \frac{1}{2} \frac{R_{\tilde{\mathbf{x}}\mathbf{x}\tilde{\mathbf{x}}}}{R^t R} \sigma_s^2(\infty) \quad (9.10)$$

donde se ha preferido poner λ como su forma equivalente $\mathbf{R}^t \mathbf{R}$. Esta última expresión es más fácil de interpretar. El término genuino de ruido (el ruido de estimación) está modificado por tres factores:

1. El factor $1/2$.
2. El tamaño del paso, ya discutido anteriormente.
3. El cociente $R_{\tilde{x}x\tilde{x}}/R^t R$ que contiene información acerca de la estadística del problema: la varianza de $\tilde{x}(n)x(n)$ dividida por la media al cuadrado.

Nótese como aspecto importantísimo que la única causa de error cuadrático medio es el ruido de estimación, y sin embargo no aparece el ruido de medición como es usual en casi todos los algoritmos adaptativos: por ejemplo el factor σ_0^2 en la ecuación (4.60). Esto puede que le beneficie cuando se compare con otros algoritmos.

Situaciones donde haya más de una incógnita

Para obtener una expresión válida en el caso de tener varias incógnitas se siguen los pasos dados en este apartado con anterioridad, pero en vez de construir el producto escalar dado en (9.7) se construye el producto externo $\Delta \mathbf{w}(n) \Delta \mathbf{w}^t(n)$ y tomando valor esperado resulta:

$$\begin{aligned} \mathbf{R}_\Delta(n) = & \mathbf{R}_\Delta(n-1) - \mu \mathbf{R}^t \mathbf{R} \mathbf{R}_\Delta(n-1) - \mu \mathbf{R}_\Delta(n-1) \mathbf{R}^t \mathbf{R} + \\ & + \mu^2 \mathbb{E}[\mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \mathbf{R} \mathbf{R}_\Delta(n-1) \mathbf{R}^t \tilde{\mathbf{x}}(n) \mathbf{x}^t(n)] + \mu^2 \mathbb{E}[\mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \mathbf{R}_s(n) \tilde{\mathbf{x}}(n) \mathbf{x}^t(n)] \end{aligned} \quad (9.11)$$

A continuación se procederá como en páginas precedentes para desacoplar las componentes diagonales de $\mathbf{R}_\Delta(n)$, es decir, se eliminará el término en μ^2 (no el independiente) y se trabajará en la base en la que $\mathbf{R}^t \mathbf{R}$ es diagonal,

$$D_i(n) = [1 - 2\mu\lambda_i] D_i(n-1) + \mu^2 R_{sx|ii}(n)$$

donde se ha definido la matriz³

$$\mathbf{R}_{sx}(n) = \mathbb{E}[\mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \mathbf{R}_s(n) \tilde{\mathbf{x}}(n) \mathbf{x}^t(n)]$$

suponiendo que se cuenta con un número suficiente de datos de forma que el ruido de estimación ya no dependa sustancialmente del tiempo, es decir, para un n suficientemente grande:

$$D_i(n) = [1 - 2\mu\lambda_i] D_i(n-1) + \mu^2 R_{sx|ii}(\infty)$$

³El carácter no estacionario a su estadística se lo da el ruido de estimación $\mathbf{s}^e(n)$.

por lo que lleva a un valor en convergencia

$$D_i(\infty) = \mu \frac{1}{2\lambda_i} R_{sx|ii}(\infty)$$

El error cuadrático medio total será la suma de todas las componentes (es la traza de la matriz \mathbf{R}_Δ):

$$D(\infty) = \mu \frac{1}{2} \sum_{i=1}^q \frac{1}{\lambda_i} R_{sx|ii}(\infty)$$

la sumatoria equivale a la traza del cociente entre la matriz \mathbf{R}_{sx} y $\mathbf{R}^t \mathbf{R}$ en la base en la que esta última es diagonal; pero como la traza no depende de base, coincide con la traza de dicho cociente en cualquier base:

$$D(\infty) = \mu \frac{1}{2} \text{Tr} \left\{ \frac{\mathbf{R}_{sx}(\infty)}{\mathbf{R}^t \mathbf{R}} \right\}$$

que es la expresión buscada.

9.1.3 Convergencia en media. Ambientes no estacionarios

Para trabajar en ambientes no estacionarios hay que realizar un pequeño cambio en la recursión que actualiza los pesos, ya que es necesario introducir el factor de olvido λ . Quedando por tanto de la siguiente manera, por motivos de normalización:

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \mu(1-\lambda) \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) [\mathbf{z}(n-1) - \Phi(n-1) \mathbf{w}(n-1)]$$

y tomando valor esperado,

$$\mathbf{E}[\mathbf{w}(n)] = \mathbf{E}[\mathbf{w}(n-1)] + \mu \mathbf{R}^t(n) [\mathbf{r}(n-1) - \mathbf{R}(n-1) \mathbf{E}[\mathbf{w}(n-1)]]$$

donde cabe destacar que al trabajar ahora en ambientes no estacionarios las cantidades \mathbf{r} y \mathbf{R} son función de n , como se indica explícitamente. Recordando ahora que por construcción

$$\mathbf{R}(n) \mathbf{w}_0(n) = \mathbf{r}(n)$$

es decir, la serie $\{\mathbf{w}_0(n)\}$ determina la evolución real de los parámetros del sistema, se puede poner que

$$\begin{aligned} \mathbf{E}[\mathbf{w}(n)] = & [\mathbf{I} - \mu \mathbf{R}^t(n) \mathbf{R}(n-1)] \mathbf{E}[\mathbf{w}(n-1)] + \\ & + \mu \mathbf{R}^t(n) \mathbf{R}(n-1) \mathbf{w}_0(n-1) \end{aligned} \quad (9.12)$$

si se consideran sistemas que no cambien mucho con el tiempo de manera que $\mathbf{R}^t(n) \mathbf{R}(n-1) = \mathbf{R}^t(n) \mathbf{R}(n)$ y se trabaja en la base en la que $\mathbf{R}^t(n) \mathbf{R}(n)$

sea diagonal, las componentes de $\mathbf{w}(n)$ se desacoplan en la ecuación anterior, quedando

$$E[\mathbf{w}_i(n)] = [1 - \mu\lambda_i(n)]E[\mathbf{w}_i(n-1)] + \mu\lambda_i(n)\mathbf{w}_{0i}(n-1) \quad (9.13)$$

donde por el mismo motivo anterior los autovalores dependen del instante temporal considerado. Ésta es la ecuación en diferencias buscada; es una ecuación en diferencias no homogénea con coeficientes no constantes. El término no homogéneo hace que la solución no tienda a cero, lo cual evita que los pesos en media tiendan al valor nulo, como debe ser. Debido a que los coeficientes no son constantes hallar la solución se complica enormemente y más al no conocer la expresión de los autovalores como función de n , de forma que la expresión para la solución que se halle deberá contener a todos $\lambda_i(1), \dots, \lambda_i(n)$. Este problema no aparecía en el algoritmo ORIV ya que debido a la existencia de la matriz $\mathbf{\Gamma}(n)$, desaparece la dependencia en los autovalores, quedando como único parámetro que controla la evolución de los pesos el factor de olvido λ (ver ecuación (4.77)). Como se puede apreciar a la vista de todos los comentarios realizados hasta ahora en cada uno de los análisis realizados, el comportamiento del algoritmo OGLMS se ve afectado de manera notable por los autovalores, bien ralentizando la convergencia en caso de problemas mal condicionados o bien complicando la resolución de las ecuaciones en diferencias que aparecen en dicho análisis. Aunque ORIV necesite más operaciones por iteración e involucre más cantidades intermedias, a la larga le beneficia porque en el análisis de su comportamiento acaban apareciendo ecuaciones más sencillas de resolver⁴ (ecuaciones en diferencias con coeficientes constantes) y su convergencia es independiente de cualquier variable externa al algoritmo (autovalores).

A continuación se procede a resolver la ecuación (9.13). En el apéndice D se resuelve una ecuación en diferencias con coeficientes no constantes. Aplicándolo a la ecuación de interés ahora, resulta, para n suficientemente grandes para eliminar la contribución de la solución de la homogénea:

$$E[\mathbf{w}_i(n)] = \mathbf{w}_{0i}(n) - \left(1 + \sum_{l=1}^{n-1} \prod_{k=l+1}^n [1 - \mu\lambda_i(k)] \right) \Delta\mathbf{w}_{0i}$$

donde se ha supuesto que la evolución de los pesos del sistema es lineal con el tiempo:

$$\mathbf{w}_0(n) = \mathbf{w}_0(n-1) + \Delta\mathbf{w}_0 \quad (9.14)$$

Por lo que se desprende que la media de los estimadores presenta un retraso dado por:

$$\tau_i = \lim_{n \rightarrow \infty} \left\{ 1 + \sum_{l=1}^{n-1} \prod_{k=l+1}^n [1 - \mu\lambda_i(k)] \right\}$$

⁴Se puede esperar intuitivamente que problemas bien planteados y resueltos de forma consistente permitan ser descritos mediante ecuaciones sencillas. Cuando la matemática se complica es que el formalismo empleado no es el más adecuado.

Como se aprecia el retraso varía según la componente i estudiada y es función del autovalor correspondiente. El valor exacto de los autovalores en función del tiempo se desconoce, por lo que no es posible dar una expresión más explícita del tiempo de retardo.

Para que las condiciones iniciales tiendan a cero, es decir, para que tienda a cero la solución de la homogénea, es necesario que se verifique (ver apéndice D) :

$$|1 - \mu\lambda_i(n)| < 1 \quad \forall i, n$$

O equivalentemente, para que se anule la solución de la homogénea para todas las componentes:

$$|1 - \mu \max_i \{\lambda_i(n)\}| < 1 \quad \forall n \quad (9.15)$$

Esta condición se verificará dependiendo del valor que se le dé al tamaño del paso y de en qué situación se trabaje:

- a) El sistema cambia poco con el tiempo. Bien porque el intervalo temporal estudiado sea pequeño o porque la velocidad de cambio sea pequeña. Entonces la condición se verifica fijado un μ constante en el tiempo $\forall n$ perteneciente al intervalo de interés.
- b) Si no se está en el caso anterior es necesario que el tamaño del paso cambie con el tiempo para que la condición (9.15) se verifique en todo el intervalo temporal durante el cual se realice el rastreo del sistema.

9.1.4 Ecuación para controlar la evolución de μ

Imponer (9.15) es equivalente a imponer,

$$0 < \mu < \frac{2}{\max_i \{\lambda_i(n)\}} \quad \forall n$$

por lo tanto hay que asegurarse que esta condición se cumple para todo instante de tiempo. Para ello se proponen dos métodos distintos explicados a continuación:

1ª Propuesta:

Como $\lambda_i(n) \leq \lambda_{\max}(n) \leq \text{Tr}\{\mathbf{R}^t(n)\mathbf{R}(n)\}$, al trasladarlo a los estimadores, resulta que $\hat{\lambda}_{\max}(n) \leq (1 - \lambda)^2 \text{Tr}\{\Phi^t(n)\Phi(n)\}$ por lo que se propone que el tamaño del paso debe verificar la condición:

$$0 < \mu(n) < \frac{2}{(1 - \lambda)^2 \text{Tr}\{\Phi^t(n)\Phi(n)\}}$$

El principal problema que presenta es que el cálculo del producto de matrices indicado requiere de lq^2 productos lo que aumenta la carga computacional del algoritmo. Recuérdese que la matriz $\Phi(n)$ es $l \times q$. Esto no interesa ya que una propiedad deseable de cualquier algoritmo adaptativo es que necesite pocas operaciones por iteración. De todas maneras, como al final lo que interesa es la traza del producto, sólo hace falta conocer los elementos diagonales. Para calcular los elementos diagonales se requieren lq productos, que es del mismo orden que la carga computacional requerida para hacer productos de matrices y vectores, como los que aparecen en el algoritmo, por lo que no aumenta la carga computacional en un nuevo orden.

También se puede estimar directamente la traza, sin tener que estimar previamente el producto mediante la estimación previa de las matrices. Para ello téngase en cuenta en primer lugar que se puede estimar directamente el valor esperado $\hat{E}[\Phi^t(n)\tilde{x}(n)x^t(n)]$ utilizando la recursión:

$$\hat{E}[\Phi^t(n)\tilde{x}(n)x^t(n)]|_n = \lambda\hat{E}[\Phi^t(n)\tilde{x}(n)x^t(n)]|_{n-1} + [\Phi^t(n)\tilde{x}(n)]x^t(n) \quad (9.16)$$

esta actualización requiere $lq + q^2$ productos y tiende a

$$\hat{E}[\Phi^t(n)\tilde{x}(n)x^t(n)] \xrightarrow{n \rightarrow \infty} \frac{1}{(1-\lambda)^2} R^t(n)R(n)$$

Por lo que la condición sobre el tamaño del paso quedaría:

$$0 < \mu(n) < \frac{2}{(1-\lambda)^2 \text{Tr}\{\hat{E}[\Phi^t(n)\tilde{x}(n)x^t(n)]|_n\}}$$

Aprovechando la forma de la recursión (9.16) se puede estimar, como se anunciaba anteriormente, directamente la traza buscada:

$$\hat{\text{Tr}}\{E[\Phi^t(n)\tilde{x}(n)x^t(n)]\}|_n = \lambda\hat{\text{Tr}}\{E[\Phi^t(n)\tilde{x}(n)x^t(n)]\}|_{n-1} + x^t(n)\Phi^t(n)\tilde{x}(n)$$

que requiere $lq + l$ productos. Análogamente al caso anterior la condición sobre el tamaño del paso sería:

$$0 < \mu(n) < \frac{2}{(1-\lambda)^2 \hat{\text{Tr}}\{E[\Phi^t(n)\tilde{x}(n)x^t(n)]\}|_n}$$

2ª Propuesta:

Sea \prec, \succ un producto escalar. Se verifica la desigualdad de Cauchy-Schwarz:

$$|\prec a, b \succ|^2 \leq \prec a, a \succ \prec b, b \succ$$

Si a y b son variables aleatorias, el valor esperado de su producto es un producto escalar. Aplicándolo a este problema se puede escribir que $R_{ij}(n) =$

$E[\tilde{\mathbf{x}}_i(n)\mathbf{x}_j(n)] \leq \sqrt{E[\tilde{\mathbf{x}}_i(n)\tilde{\mathbf{x}}_i(n)]E[\mathbf{x}_j(n)\mathbf{x}_j(n)]}$ con lo cual:

$$\begin{aligned} \text{Tr}\{\mathbf{R}^t(n)\mathbf{R}(n)\} &= \mathbf{R}_{ij}^t(n)\mathbf{R}_{ji}(n) = E[\mathbf{x}_i(n)\tilde{\mathbf{x}}_j(n)]E[\tilde{\mathbf{x}}_j(n)\mathbf{x}_i(n)] \leq \\ &\leq E[\mathbf{x}_i(n)\mathbf{x}_i(n)]E[\tilde{\mathbf{x}}_j(n)\tilde{\mathbf{x}}_j(n)] = \\ &= E[\mathbf{x}^t(n)\mathbf{x}(n)]E[\tilde{\mathbf{x}}^t(n)\tilde{\mathbf{x}}(n)] \end{aligned} \quad (9.17)$$

Para calcular los valores esperados indicados, como se está trabajando en ambientes no estacionarios, es preciso utilizar la recursión conocida:

$$\hat{E}[\mathbf{x}^t(n)\mathbf{x}(n)]|_n = \lambda\hat{E}[\mathbf{x}^t(n)\mathbf{x}(n)]|_{n-1} + \mathbf{x}^t(n)\mathbf{x}(n)$$

Por tanto, la condición sobre $\mu(n)$ es:

$$0 < \mu(n) < \frac{2}{(1-\lambda)^2\hat{E}[\mathbf{x}^t(n)\mathbf{x}(n)]|_n\hat{E}[\tilde{\mathbf{x}}^t(n)\tilde{\mathbf{x}}(n)]|_n}$$

El número de productos requerido por este método es $q+l+1$ y aunque es el más bajo de todos presenta el inconveniente de que no es inmune al ruido aditivo gaussiano cuando se utilice OGLMS para resolver ecuaciones que involucren cumulantes de alto orden, como se verá en las simulaciones.

Ejemplo: Aplicación práctica al caso de las ecuaciones de Giannakis-Mendel

Estas ecuaciones para la identificación ciega de sistemas MA se introdujeron en el capítulo 5, donde también se detalló cómo utilizar el algoritmo ORIV para resolverlas. Para resolverlas mediante OGLMS se definen las variables implicadas de forma idéntica. En particular:

$$\begin{aligned} \mathbf{x}(n) &= [y^2(n-q)\dots y^2(n-2q)|y(n-q-1)\dots y(n-2q)]^t \\ \tilde{\mathbf{x}}(n) &= [y(n)\dots y(n-3q)]^t \end{aligned}$$

Se propone controlar el valor del tamaño del paso mediante la 2ª propuesta. Para ello hay que calcular los siguientes valores esperados:

$$\begin{aligned} E[\mathbf{x}^t(n)\mathbf{x}(n)] &= (q+1)m_{4y}(0,0,0;n) + q\sigma_y^2(n) \\ E[\tilde{\mathbf{x}}^t(n)\tilde{\mathbf{x}}(n)] &= (3q+1)\sigma_y^2(n) \end{aligned} \quad (9.18)$$

Donde hay que tener en cuenta varios aspectos. El primero es que como se trabaja en ambientes no estacionarios la estadística cambia con el tiempo, así se ha indicado con el índice n en los momentos implicados. Aunque en los valores esperados que se han calculado debería aparecer los momentos desde tiempo n hasta $n-3q$, se ha supuesto que hay poca diferencia entre ellos y se han tomado igual al momento a tiempo n por simplicidad, tal y como se aprecia en (9.18).

A continuación se calcula cuál es la dependencia en el tiempo de estos momentos, para lo cual se va a suponer que los parámetros del sistema evolucionan de igual forma a lo especificado en la ecuación (9.14) para el cambio del vector incógnita⁵. O equivalentemente:

$$\mathbf{b}(n) = \mathbf{b}_0 + n\mathbf{p}$$

Por tanto:

$$\begin{aligned} \sigma_y^2(n) &= \sigma_w^2 \sum_{i=0}^q b_i^2(n) = \sigma_w^2 \sum_{i=0}^q (b_{0i} + np_i)^2 = \\ &= \sigma_w^2 \left(\sum_{i=0}^q b_{0i}^2 + 2n \sum_{i=0}^q b_{0i}p_i + n^2 \sum_{i=0}^q p_i^2 \right) = \sigma_y^2(0) + n\Delta\sigma_y^2(n) \end{aligned} \quad (9.19)$$

donde

$$\Delta\sigma_y^2(n) = \sigma_w^2 \sum_{i=0}^q (2b_{0i}p_i + p_i^2 n)$$

lo importante es resaltar que la varianza es un polinomio de segundo orden en np_i . Análogamente para el momento de orden 4:

$$\begin{aligned} m_{4y}(0, 0, 0; n) &= c_{4y}(0, 0, 0; n) + 3\sigma_y^2(n)\sigma_y^2(n) = \\ &= \gamma_{4w} \sum_{i=0}^q (b_{0i} + np_i)^4 + 3(\sigma_y^2(0) + n\Delta\sigma_y^2(n))^2 \end{aligned} \quad (9.20)$$

que resulta ser un polinomio de orden 4 en np_i . En definitiva se puede concluir que el producto $E[\mathbf{x}^t(n)\mathbf{x}(n)]E[\tilde{\mathbf{x}}^t(n)\tilde{\mathbf{x}}(n)]$ es un polinomio de orden 6 en np_i . Si el sistema evoluciona durante tiempo suficiente o la evolución es rápida, $np_i > 1$ y predomina el sumando con la potencia sexta, dicho sumando es $\sigma_w^2 \gamma_{4w} \sum_{i=1}^q (np_i)^2 \sum_{i=1}^q (np_i)^4$ por lo que la condición que debe verificar el tamaño del paso es:

$$0 < \mu(n) < \frac{2}{\sigma_w^2 \gamma_{4w} \sum_{i=0}^q (np_i)^4 \sum_{i=0}^q (np_i)^2} \quad (9.21)$$

Como aplicación numérica considérese un modelo MA de orden 1 con parámetros iniciales $[1 \quad -1.25]$ y una pendiente $\mathbf{p} = [1 \quad -1/10]$. En la figura 9.1 se compara el autovalor máximo de la matriz $\mathbf{R}^t \mathbf{R}$ para este ejemplo con dos cotas superiores estudiadas en estas últimas líneas: su traza y la cota dada en la 2ª propuesta y concretada en (9.21).

Como se puede apreciar la traza sobreestima claramente el valor del autovalor mayor, como se sabía de antemano. Aunque la cota dada por la segunda

⁵Como se sabe del capítulo 5, no coinciden los parámetros del sistema con el vector incógnita, sino que el segundo es una función de los primeros (véanse las expresiones (5.42) y (5.43)).

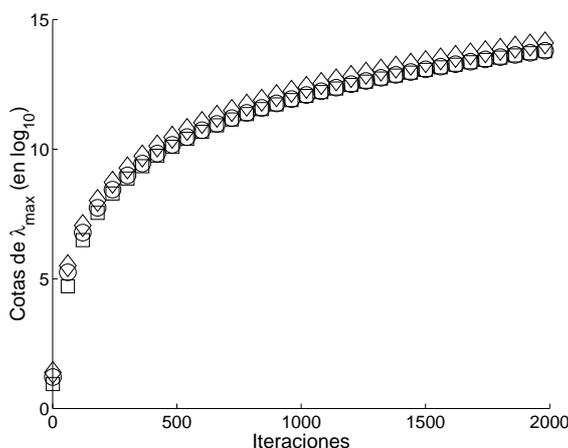


Figura 9.1 Cotas para el máximo autovalor: \blacktriangledown $Tr[\mathbf{R}^t \mathbf{R}]$; \circ Autovalor máximo; \square 2ª propuesta

propuesta debería tomar los mismos valores que la cota debida a la traza, no es así ya que se han eliminado términos y retenido solamente el que presentaba una mayor potencia en el índice temporal n . Por este motivo se encuentra bastante por debajo de él, justo coincidiendo con el valor real del autovalor. Esto es más cierto conforme mayor es el valor del producto np . Para este ejemplo se ha elegido un valor realmente alto para la pendiente de la variación lineal de los parámetros del sistema; tan alto que el algoritmo no podría, seguramente, seguir la evolución. Se han elegido estos valores como ilustración en este ejemplo.

Si se aplicase el mismo método para encontrar una cota al tamaño del paso, en el caso de que se resolviesen las ecuaciones basadas en cumulantes de tercer y cuarto orden mediante el algoritmo OGLMS, se podría esperar un resultado del tipo:

$$0 < \mu(n) < \frac{2}{\sigma_w^2 \gamma_{6w} \sum_{i=0}^q (np_i)^6 \sum_{i=0}^q (np_i)^2}$$

Llamando $K(n)$ a la cota superior para el tamaño del paso que establecen cada una de las aproximaciones anteriores, se puede asignar a dicho tamaño del paso el siguiente valor:

$$\mu(n) = cK(n)$$

donde c es una constante menor que la unidad, cuyo valor concreto dependerá del problema en cuestión y de los objetivos perseguidos (mayor velocidad, menor error cuadrático medio, etc.).

9.1.5 Convergencia en media. Ambientes no estacionarios II

Del modelado que se ha hecho anteriormente, de la variación de los parámetros del sistema como una función lineal en el tiempo, podría deducirse que tales parámetros se van separando cada vez más de sus valores iniciales, aumentando cada vez más su módulo. Sin embargo, sería más plausible suponer que dichos parámetros evolucionan aleatoriamente a partir de sus valores iniciales, describiendo un camino aleatorio en el espacio de parámetros. Una forma de modelar este comportamiento es suponer que siguen la evolución marcada por un proceso de Markov de orden 1:

$$\mathbf{w}_0(n) = a\mathbf{w}_0(n-1) + \mathbf{g}(n)$$

donde $\mathbf{g}(n)$ es el vector aleatorio del ruido del proceso con media nula y matriz de correlación \mathbf{Q} .

El análisis de la convergencia se lleva a cabo de manera análoga a como se realizó en los apartados anteriores. Se parte de la recursión para $\mathbf{w}(n)$:

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \mu(1-\lambda)\mathbf{x}(n)\tilde{\mathbf{x}}^t(n)[\mathbf{z}(n-1) - \Phi(n-1)\mathbf{w}(n-1)]$$

como se está interesado en el comportamiento cerca de y en convergencia, se pueden hacer las aproximaciones correspondientes a considerar n grande:

$$\begin{aligned} \mathbf{z}(n-1) &= \frac{1}{1-\lambda}\mathbf{r}(n-1) = \frac{1}{1-\lambda}\mathbf{R}(n-1)\mathbf{w}_0(n-1) \\ \Phi(n-1) &= \frac{1}{1-\lambda}\mathbf{R}(n-1) \end{aligned} \quad (9.22)$$

Hay que tener en cuenta que en este ambiente hay dos fuentes de aleatoriedad en las magnitudes. La primera es la usual debido a que la excitación del sistema es estocástica. La segunda es debida a que los parámetros del sistema evolucionan según un proceso de Markov de orden 1. Es por esto por lo que las cantidades $\mathbf{r}(n-1)$ y $\mathbf{R}(n-1)$ que aparecen en (9.22) son estocásticas (puede que sea más correcto poner $E[\mathbf{w}_0(n-1)]$ en (9.22)). Realizando estas sustituciones la ecuación anterior queda:

$$\mathbf{w}(n) = [\mathbf{I} - \mu\mathbf{x}(n)\tilde{\mathbf{x}}^t(n)\mathbf{R}(n-1)]\mathbf{w}(n-1) + \mu\mathbf{x}(n)\tilde{\mathbf{x}}^t(n)\mathbf{R}(n-1)\mathbf{w}_0(n-1) \quad (9.23)$$

tomando valor esperado y trabajando en la base en la que $\mathbf{R}^t(n)\mathbf{R}(n)$ es diagonal, la ecuación en diferencias para una componente queda:

$$E[\mathbf{w}_i(n)] = [1 - \mu\lambda_i(n)]E[\mathbf{w}_i(n-1)] + \mu\lambda_i(n)E[\mathbf{w}_{0i}(n-1)]$$

que tiene como solución:

$$\begin{aligned} E[\mathbf{w}_i(n)] = & E[\mathbf{w}_{0i}(0)] \prod_{j=1}^n (1 - \mu\lambda_i(j)) + \mu\lambda_i(n)E[\mathbf{w}_{0i}(n-1)] + \\ & + E[\mathbf{w}_{0i}(0)] \sum_{j=1}^{n-1} a^{j-1} \mu\lambda_i(j) \prod_{k=j+1}^n (1 - \mu\lambda_i(k)) \end{aligned} \quad (9.24)$$

Una condición suficiente para que las condiciones iniciales $E[\mathbf{w}_0(0)]$ se anulen para n grandes es que se verifique:

$$|1 - \mu\lambda_i(n)| < 1 \quad \forall n, i$$

condición ya conocida con anterioridad. En el segundo sumando, el factor

$$a^{j-1} \prod_{k=j+1}^n (1 - \mu\lambda_i(k))$$

atenúa la influencia de muestras pasadas. Por tanto, reescribiendo (9.24) para n grandes:

$$\begin{aligned} E[\mathbf{w}_i(n)] = & E[\mathbf{w}_{0i}(n-1)]\mu\lambda_i(n) + \\ & + E[\mathbf{w}_{0i}(0)] \sum_{j=1}^{n-1} a^{j-1} \mu\lambda_i(j) \prod_{k=j+1}^n (1 - \mu\lambda_i(k)) \end{aligned} \quad (9.25)$$

Es decir, el algoritmo en media converge al valor verdadero $E[\mathbf{w}_{0i}(n-1)]$ afectado por el factor $\mu\lambda_i(n)$ más un término responsable del retraso que presenta el algoritmo en el seguimiento de la evolución real. El factor que regula el retraso es

$$\sum_{j=1}^{n-1} a^{j-1} \mu\lambda_i(j) \prod_{k=j+1}^n (1 - \mu\lambda_i(k))$$

En el caso en que se escoja $\mu(n)\lambda_i(n) \sim 1 \quad \forall n$ se obtiene el mejor rastreo posible, ya que entonces se verificaría:

$$E[\mathbf{w}_i(n)] \simeq E[\mathbf{w}_{0i}(n-1)]$$

es decir, se elimina el retraso eligiendo el tamaño del paso lo mayor posible. Cuanto más rápido rastree menor retraso presenta. Además, en el caso en que el sistema sea estacionario⁶ $a = 1$, y este factor no aparece en la ecuación anterior como debe ser.

⁶Nótese que el vector de los pesos del filtro sigue afectado por una componente aleatoria, pero esta es de media nula y por tanto no influye en la convergencia en media.

9.1.6 Convergencia en media cuadrática en ambientes no estacionarios

Para llevar a cabo este análisis se parte de la ecuación (9.23),

$$\mathbf{w}(n) = [\mathbf{I} - \mu \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \mathbf{R}(n-1)] \mathbf{w}(n-1) + \mu \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \mathbf{R}(n-1) \mathbf{w}_0(n-1)$$

con la que es sencillo, suponiendo que los pesos del filtro evolucionan según un proceso de Markov de orden 1, llegar a:

$$\mathbf{w}(n) - \mathbf{w}_0(n) = [\mathbf{I} - \mu \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \mathbf{R}(n-1)] [\mathbf{w}(n-1) - \mathbf{w}_0(n-1)] - \mathbf{g}(n) + (1-a) \mathbf{w}_0(n-1) \quad (9.26)$$

Suponiendo que el coeficiente $a \sim 1$, que el cambio es pequeño en los pesos del filtro, el último sumando se puede despreciar. Con esta simplificación, si se construye el producto $\Delta \mathbf{w}(n) \Delta \mathbf{w}^t(n)$ se obtiene:

$$\begin{aligned} \Delta \mathbf{w}(n) \Delta \mathbf{w}^t(n) = & \mathbf{g}(n) \mathbf{g}^t(n) + \text{términos de media nula} + \\ & + [\mathbf{I} - \mu \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \mathbf{R}(n-1)] \Delta \mathbf{w}(n-1) \Delta \mathbf{w}^t(n-1) [\mathbf{I} - \mu \mathbf{R}^t(n-1) \tilde{\mathbf{x}}(n) \mathbf{x}^t(n)] \end{aligned} \quad (9.27)$$

tomando valor esperado,

$$\begin{aligned} \mathbf{R}_\Delta(n) = & \mathbf{R}_\Delta(n-1) + \mu^2 \mathbb{E}[\mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \mathbf{R}(n-1) \mathbf{R}_\Delta(n-1) \mathbf{R}^t(n-1) \tilde{\mathbf{x}}(n) \mathbf{x}^t(n)] - \\ & - \mu \mathbf{R}_\Delta(n-1) \mathbf{R}^t(n-1) \mathbf{R}(n) - \mu \mathbf{R}^t(n) \mathbf{R}(n-1) \mathbf{R}_\Delta(n-1) + \mathbf{Q} \end{aligned} \quad (9.28)$$

Por sencillez se tomará $n \approx n-1$, se trabajará en la base en la que $\mathbf{R}^t(n) \mathbf{R}(n)$ es diagonal y se despreciará el término en μ^2 . Entonces, llamando $D_i(n)$ a los elementos diagonales de $\mathbf{R}_\Delta(n)$, queda:

$$D_i(n) = [1 - 2\mu\lambda_i(n)] D_i(n-1) + Q_{ii}$$

Como en el resto de ecuaciones en diferencias estudiadas hasta ahora, si se desea que los efectos de las condiciones iniciales se anulen conforme evoluciona el algoritmo, es necesario imponer

$$|1 - 2\mu\lambda_i(n)| < 1 \quad \forall n, i$$

Esta condición se extrae fácilmente de la expresión general para $D_i(n)$:

$$D_i(n) = \prod_{j=1}^n [1 - 2\mu\lambda_i(j)] D_i(0) + \left[1 + \sum_{j=1}^{n-1} \prod_{k=j+1}^n [1 - 2\mu\lambda_i(k)] \right] Q_{ii}$$

y en convergencia, una vez eliminados los efectos de las condiciones iniciales,

$$D_i(\infty) = \left[1 + \sum_{j=1}^{\infty} \prod_{k=j+1}^{\infty} [1 - 2\mu\lambda_i(k)] \right] Q_{ii}$$

donde, como se aprecia, aparecen dos sumandos. La influencia de uno de ellos se puede minimizar con la adecuada elección del tamaño del paso, pero aún en el mejor caso quedará una contribución igual a Q_{ii} . El error cuadrático medio total será la suma de cada una de las componentes.

9.2 Algoritmo de gradiente estocástico basado en la 3ª opción, OGLMS3

Utilizando como término de gradiente el propuesto como 3ª opción en el capítulo anterior, la recursión para la actualización de los pesos queda de la siguiente manera:

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \frac{\mu}{n} \Phi^t(n) [\tilde{\mathbf{x}}(n)d(n) - \tilde{\mathbf{x}}(n)\mathbf{x}^t(n)\mathbf{w}(n-1)]$$

En los siguientes apartados se va a proceder a analizar el algoritmo resultante tanto en media como en media cuadrática. El objetivo primordial será compararlo con el basado en la 1ª opción para elegir el de mejor comportamiento.

9.2.1 Convergencia en media

Se supondrán ambientes estacionarios y libres de ruido. Partiendo de la ecuación anterior, sumando y restando los términos adecuados, de forma completamente análoga al análisis realizado para la 1ª opción, se llega a:

$$\begin{aligned} \Delta \mathbf{w}(n) = & \left[\mathbf{I} - \frac{\mu}{n} \Phi^t(n) \tilde{\mathbf{x}}(n) \mathbf{x}^t(n) \right] \Delta \mathbf{w}(n-1) + \\ & + \frac{\mu}{n} \Phi^t(n) \tilde{\mathbf{x}}(n) [d(n) - \mathbf{x}^t(n) \mathbf{w}_0] \end{aligned} \quad (9.29)$$

Como el interés se centra en estudiar la convergencia, es lógico aproximar la expresión (9.29) por su equivalente una vez supuestos n grandes. La única sustitución que se puede hacer es:

$$\Phi(n) = n\mathbf{R}$$

Una vez realizada dicha sustitución y recordando que

$$\mathbf{E}[\tilde{\mathbf{x}}(n)e_0(n)] = 0 \text{ con } e_0(n) = d(n) - \mathbf{x}^t(n)\mathbf{w}_0$$

que es la relación de ortogonalidad que se impone entre datos e incógnitas, ecuación a resolver durante todas estas páginas, se llega, tomando valor esperado a:

$$\mathbf{E}[\Delta \mathbf{w}(n)] = [\mathbf{I} - \mu \mathbf{R}^t \mathbf{R}] \mathbf{E}[\Delta \mathbf{w}(n-1)] \quad (9.30)$$

que es la misma ecuación en diferencias que se obtuvo en el análisis de la convergencia en media del algoritmo basado en la 1ª opción. Por tanto, en lo que respecta al comportamiento en media no hay diferencia entre una y otra opción.

9.2.2 Convergencia en media cuadrática. Ambientes estacionarios

Se parte de la ecuación deducida anteriormente (9.29) donde se supone ahora n grande

$$\Delta \mathbf{w}(n) = [\mathbf{I} - \mu \mathbf{R}^t \tilde{\mathbf{x}}(n) \mathbf{x}^t(n)] \Delta \mathbf{w}(n-1) + \mu \mathbf{R}^t \tilde{\mathbf{x}}(n) [d(n) - \mathbf{x}^t(n) \mathbf{w}_0]$$

y a partir de ella se construye la siguiente matriz

$$\begin{aligned} \Delta \mathbf{w}(n) \Delta \mathbf{w}^t(n) &= \mu^2 \mathbf{R}^t \tilde{\mathbf{x}}(n) \tilde{\mathbf{x}}^t(n) \mathbf{R} [d(n) - \mathbf{x}^t(n) \mathbf{w}_0]^2 + \\ &[\mathbf{I} - \mu \mathbf{R}^t \tilde{\mathbf{x}}(n) \mathbf{x}^t(n)] \Delta \mathbf{w}(n-1) \Delta \mathbf{w}^t(n-1) [\mathbf{I} - \mu \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \mathbf{R}] + \end{aligned} \quad (9.31)$$

+ dos términos cuya media tiende a 0

los términos cuya media tiende a cero son aquellos que contienen $E[\mathbf{w}(n-1)]$ al tomar valor esperado y aplicar hipótesis de independencia, ya que por (9.30) en condiciones de convergencia se anulan.

Tomando valor esperado:

$$\begin{aligned} \mathbf{R}_\Delta(n) &= \mathbf{R}_\Delta(n-1) - \mu \mathbf{R}^t \mathbf{R} \mathbf{R}_\Delta(n-1) - \mu \mathbf{R}_\Delta(n-1) \mathbf{R}^t \mathbf{R} + \\ &+ \mu^2 \mathbf{R}^t E[\tilde{\mathbf{x}}(n) \mathbf{x}^t(n) \mathbf{R}_\Delta(n-1) \mathbf{x}(n) \tilde{\mathbf{x}}^t(n)] \mathbf{R} + \end{aligned} \quad (9.32)$$

$$+ \mu^2 \mathbf{R}^t E[\tilde{\mathbf{x}}(n) [d(n) - \mathbf{x}^t(n) \mathbf{w}_0]^2 \tilde{\mathbf{x}}^t(n)] \mathbf{R}$$

en una primera aproximación se eliminará el término en μ^2 (no independiente) ya que se suele trabajar con valores del tamaño del paso pequeños, que a la vez facilita que se verifiquen las hipótesis de independencia (Apéndice C). La eliminación de ese término permite desacoplar las componentes diagonales de $\mathbf{R}_\Delta(n)$ en la base en la que $\mathbf{R}^t \mathbf{R}$ es diagonal y por tanto facilitan su resolución:

$$D_i(n) = (1 - 2\mu\lambda_i) D_i(n-1) + \mu^2 \mathbf{F}_{ii} \quad (9.33)$$

donde se ha definido la matriz \mathbf{F} :

$$\mathbf{F} = \mathbf{R}^t \mathbf{R}_{\tilde{\mathbf{x}}e} \mathbf{R}$$

$$\mathbf{R}_{\tilde{\mathbf{x}}e} = E[\tilde{\mathbf{x}}(n) [d(n) - \mathbf{x}^t(n) \mathbf{w}_0]^2 \tilde{\mathbf{x}}^t(n)]$$

A partir de la ecuación (9.33) se puede obtener la velocidad de convergencia y el valor en convergencia, tal y como se ha hecho en ocasiones anteriores. Con

el fin de comparar la convergencia en media cuadrática de las opciones 1ª y 3ª se va a analizar dicho aspecto de forma análoga a como se hizo para la primera. Es decir, se construye,

$$\begin{aligned} \Delta \mathbf{w}^t(n) \Delta \mathbf{w}(n) = & \mu^2 \tilde{\mathbf{x}}^t(n) \mathbf{R} \mathbf{R}^t \tilde{\mathbf{x}}(n) [d(n) - \mathbf{x}^t(n) \mathbf{w}_0]^2 + \\ \Delta \mathbf{w}^t(n-1) [\mathbf{I} - \mu \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \mathbf{R}] & [\mathbf{I} - \mu \mathbf{R}^t \tilde{\mathbf{x}}(n) \mathbf{x}^t(n)] \Delta \mathbf{w}(n-1) + \end{aligned} \quad (9.34)$$

+ dos términos cuya media tiende a 0

tomando valor esperado:

$$\begin{aligned} D(n) = & \text{Tr}\{[\mathbf{I} - 2\mu \mathbf{R}^t \mathbf{R} + \mu^2 \text{E}[\mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \mathbf{R} \mathbf{R}^t \tilde{\mathbf{x}}(n) \mathbf{x}^t(n)] \mathbf{R}\}_\Delta(n-1) + \\ & + \mu^2 \text{Tr}\{\mathbf{R} \mathbf{R}^t \text{E}[\tilde{\mathbf{x}}(n) [d(n) - \mathbf{x}^t(n) \mathbf{w}_0]^2 \tilde{\mathbf{x}}^t(n)]\} \end{aligned} \quad (9.35)$$

que se podía haber obtenido directamente de (9.32) tomando trazas a ambos miembros. Esta expresión es bien distinta a su análoga para la primera opción (9.8). Para discutir algo sobre ella supóngase de nuevo que sólo hay una incógnita:

$$D(n) = [1 - 2\mu\lambda + \mu^2 \lambda R_{\tilde{x}x\tilde{x}}] D(n-1) + \mu^2 \lambda \sigma_{\tilde{x}e}^2 \quad (9.36)$$

donde se han usado las siguientes definiciones:

$$\begin{aligned} R_{\tilde{x}x\tilde{x}} &= \text{E}[\tilde{x}(n)x(n)x(n)\tilde{x}(n)] \\ \sigma_{\tilde{x}e}^2 &= \text{E}[\tilde{x}(n)\tilde{x}(n)[d(n) - x(n)w_0]^2] \end{aligned}$$

Si se compara (9.36) con (9.9) se puede apreciar que ambas ecuaciones en diferencias van a presentar el mismo tiempo de convergencia, pero distinto valor en convergencia. En particular, ésta converge a:

$$D(\infty) = \frac{\mu}{2 - \mu R_{\tilde{x}x\tilde{x}}} \sigma_{\tilde{x}e}^2$$

y tomando el tamaño del paso suficientemente pequeño:

$$D(\infty) = \mu \frac{1}{2} \sigma_{\tilde{x}e}^2$$

Lo primero que llama la atención al compararla con (9.10) es que no contiene un término que dependa directamente de la estadística del problema (aunque para el caso de más de una incógnita sí que lo tiene). En segundo lugar llama la atención el origen del término de error:

1ª opción: sólo hay ruido de estimación de \mathbf{R} y \mathbf{r} contenido en σ_s .

3ª opción: aparece ruido debido a la ecuación de partida de resolución del problema $\sigma_{\tilde{x}e}$. Es decir, si se impone la condición

$$\text{E}[\tilde{\mathbf{x}}(n)e_0(n)] = 0$$

se resuelve el problema, pero su matriz de autocorrelación no es nula

$$E[\tilde{\mathbf{x}}(n)\tilde{\mathbf{x}}^t(n)e_0^2(n)] \neq \mathbf{0}$$

y dicha matriz de autocorrelación aparece en el análisis de la 3ª opción, pero no en el de la primera, ya que en esta última se evita realizando un promedio temporal.

Debido a que ya se contaba en la 3ª opción con esta fuente de error, no se ha incluido el error de estimación que sin duda aparece. Por tanto, esta opción presenta una fuente adicional de error que le llevará a un valor del error cuadrático medio mayor. Esta propiedad hará que de ahora en adelante se deje a un lado el algoritmo OGLMS3 para estudiar con más detalle el OGLMS1.

Capítulo 10

El algoritmo promediado y generalizado de mínima media cuadrática, AOGLMS. Obtención y análisis

Introducción

En el capítulo 8 se derivó un conjunto de algoritmos, basados en la idea del gradiente estocástico, a partir del principio de ortogonalidad sobredeterminado y generalizado, formulado a raíz del análisis del algoritmo ORIV. Dentro de este conjunto destaca el GLMS, rederivado de forma diferente a como se hace en la bibliografía [AAM96], y los algoritmos OGLMS1 y OGLMS3, que al contrario de lo que le ocurría a aquél son capaces de trabajar en situaciones generales donde la matriz asociada al problema sea sobredeterminada y no necesariamente definida. La principal desventaja que presentaban es que requerían del orden de lq operaciones por iteración, donde q es el número de incógnitas y l es la longitud del vector de variables instrumentales. Esta desventaja disminuye un poco el gran potencial que presenta en cuanto a espectro de aplicabilidad en problemas concretos.

El siguiente paso lógico sería intentar disminuir esa carga computacional, de manera que se asemeje más y más al tradicional algoritmo LMS. Lo ideal

sería conseguir un número de operaciones por iteración del orden de $l + q$, y eso es lo que se hace en este capítulo, proponiendo una ligera modificación a OGLMS1, que se ha comprobado analíticamente que tiene mejor comportamiento que OGLMS3. Hay que destacar que este algoritmo modificado se denominará Averaged Overdetermined Generalized Least Mean Square Algorithm, AOGLMS; el adjetivo de Averaged para este algoritmo proviene de que en su ecuación de actualización de los pesos aparece un promedio temporal que hace las veces de estimador de un valor esperado.

Como en todos los capítulos en los que se describe un algoritmo, en éste también se analizarán sus propiedades de convergencia. Hay que destacar que lo que más interesa es comprobar que de verdad el algoritmo diseñado posee las propiedades que se le exigieron cuando se le diseñó. Para AOGLMS eso significa comprobar que resuelve sistemas sobredeterminados cuya matriz de cumulantes no es definida, aparte de los exigibles a cualquier algoritmo basado en estadística de alto orden, principalmente su inmunidad al ruido aditivo gaussiano. Este análisis se va a llevar a cabo desde dos puntos de vista: el primero supone que el estimador del valor esperado que aparece en la ecuación de actualización tiene todas las propiedades de un valor esperado real, lo cual simplifica enormemente las cosas. El segundo punto de vista estudia con más detalle la interacción (dependencia o independencia) del estimador del valor esperado con el resto de variables aleatorias, como variable aleatoria a su vez que es. De todas formas no se espera reproducir exactamente los resultados experimentales, ya que en este capítulo no es la prioridad.

Este capítulo se puede considerar como aquel en el que se cumplen objetivos de la Memoria, ya que es el final de un proceso lógico-constructivo que se inició en el capítulo 4 con el análisis del algoritmo ORIV.

Este capítulo está dividido de la siguiente manera. En primer lugar se deriva el algoritmo AOGLMS en el apartado 10.1 para posteriormente proceder a su análisis: la convergencia en media con la aproximación extrema en el apartado 10.2, la convergencia en media cuadrática mediante promediado directo en el 10.3, mediante la segunda aproximación se analiza la convergencia en media y en media cuadrática en los apartados 10.4 y 10.5 respectivamente. Los fundamentos de la segunda aproximación se discuten en el apartado 10.6 y finalmente se completa el análisis de la convergencia en media con todo lo presentado en el apartado 10.7. Para concluir, la convergencia del algoritmo y su superioridad sobre GLMS se confirman mediante simulaciones en el apartado 10.8.

10.1 El algoritmo AOGGLMS

Como el problema a resolver es el mismo de los capítulos anteriores aquí no se volverá a enunciar. Además se respetará la notación utilizada hasta ahora. Por tanto, si se recuerda la ecuación de actualización para el algoritmo OGLMS1, reproducida aquí por conveniencia

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \frac{\mu}{n-1} \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) [\mathbf{z}(n-1) - \Phi(n-1) \mathbf{w}(n-1)] \quad (10.1)$$

se puede observar que la mayor parte de la carga computacional está concentrada en el cálculo de la matriz Φ y en el producto de esta matriz con el vector de incógnitas, ambas de orden lq . Si se quiere reducir el número de operaciones por iteración no hay más remedio que sustituir estas operaciones por otras equivalentes. Una posible opción es encontrar una ecuación de actualización simplificada para el error promediado $\bar{e}(n) = \mathbf{z}(n) - \Phi(n) \mathbf{w}(n)$, procediendo de la siguiente manera.

En primer lugar se emplean las ecuaciones de actualización de las magnitudes involucradas para reescribirla como:

$$\begin{aligned} \bar{e}(n) = & \lambda \mathbf{z}(n-1) + \tilde{\mathbf{x}}(n) d(n) - \\ & - [\lambda \Phi(n-1) + \tilde{\mathbf{x}}(n) \mathbf{x}^t(n)] [\mathbf{w}(n-1) + \mu \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \bar{e}(n-1)] \end{aligned} \quad (10.2)$$

reagrupando términos y recordando la expresión para el error a priori se llega a:

$$\bar{e}(n) = [\lambda - \Phi(n) \mu \mathbf{x}(n) \tilde{\mathbf{x}}^t(n)] \bar{e}(n-1) + \tilde{\mathbf{x}}(n) e(n)$$

Esta ecuación, que es la de actualización de $\bar{e}(n)$, quiere decir que $\bar{e}(n)$ es una versión filtrada, promediada, de $\tilde{\mathbf{x}}(n) e(n)$. El filtro es un AR de orden 1 con coeficiente matricial y aleatorio.

Si se quiere simplificar este proceso de actualización, esta matriz debe pasar a ser un escalar y parámetro libre del algoritmo:

$$\bar{e}(n) = f \bar{e}(n-1) + (1-f) \tilde{\mathbf{x}}(n) e(n)$$

lo cual quiere decir que el error promediado viene definido por

$$\bar{e}(n) = \frac{1-f^n}{1-f} \sum_{i=1}^n f^{n-i} \tilde{\mathbf{x}}(i) e(i) \quad (10.3)$$

o equivalentemente

$$\bar{e}(n) = \hat{\mathbb{E}}[\tilde{\mathbf{x}}(n) e(n)]$$

Por tanto se propone como ecuación fundamental del nuevo algoritmo

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \mu \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \bar{e}(n) \quad (10.4)$$

y al nuevo algoritmo se le llamará algoritmo promediado y sobredeterminado de mínima media cuadrática, AOGLMS (averaged and overdetermined generalized least mean square algorithm). Nótese que ya no es necesario dividir μ por n ya que $\bar{e}(n)$ se ha definido convenientemente normalizado. Además, allí donde sea conveniente por motivos de interpretación, en vez de escribir $\bar{e}(n)$ se escribirá su equivalente $\hat{E}[\tilde{\mathbf{x}}(n)e(n)]$.

El algoritmo AOGLMS se presenta en la tabla 10.1

<p>Inicialización</p> $\mathbf{w}(0) = \mathbf{0}$ $\bar{e}(0) = \mathbf{0}$ <p>Con la llegada de un nuevo dato: construir $\tilde{\mathbf{x}}(n)$, $\mathbf{x}(n)$ y $d(n)$ $e(n) = d(n) - \mathbf{w}^t(n-1)\mathbf{x}(n)$ $\bar{e}(n) = f\bar{e}(n-1) + (1-f)\tilde{\mathbf{x}}(n)e(n)$ $\mathbf{w}(n) = \mathbf{w}(n-1) + \mu\mathbf{x}(n)\tilde{\mathbf{x}}^t(n)\bar{e}(n)$</p>

Tabla 10.1 *Proceso iterativo del algoritmo AOGLMS*

Diferencia entre el algoritmo AOGLMS y OGLMS1

Sus diferencias se centran en sus respectivas ecuaciones de actualización. Lo que diferencia las ecuaciones (10.1) y (10.4) es el término entre corchetes de la primera y el estimador de la segunda. Para compararlas desarrollemos por un lado el estimador según su definición:

$$\begin{aligned} \hat{E}[\tilde{\mathbf{x}}(n)e^*(n)] &= \frac{1-f}{1-f^n} \sum_{i=1}^n f^{n-i} \tilde{\mathbf{x}}(i)e^*(i) = \\ &= \frac{1-f}{1-f^n} \sum_{i=1}^n f^{n-i} [\tilde{\mathbf{x}}(i)d^*(i) - \tilde{\mathbf{x}}(i)\mathbf{x}^t(i)\mathbf{w}(i-1)] \end{aligned} \quad (10.5)$$

y por el otro las cantidades $\mathbf{z}(n)$ y $\Phi(n)$ según las definiciones respectivas,

$$\mathbf{z}(n-1) - \Phi(n-1)\mathbf{w}(n-1) = \sum_{i=1}^{n-1} [\tilde{\mathbf{x}}(i)d^*(i) - \tilde{\mathbf{x}}(i)\mathbf{x}^t(i)\mathbf{w}(n-1)] \quad (10.6)$$

Se pueden descartar los siguientes puntos:

1. El límite superior de la suma se podría modificar sin problema para que no hubiera diferencia.
2. Para problemas donde las estadísticas conjuntas de $\tilde{x}(n)$ y $d(n)$ por un lado y $\tilde{x}(n)$ y $x(n)$ por otro sean estacionarias, las definiciones de $z(n)$ y $\Phi(n)$ no requieren, ni es conveniente, el uso de un factor de olvido. En situaciones donde estas estadísticas no sean estacionarias, sí es necesario un factor de olvido, lo que lleva asociado un factor de normalización $(1 - \lambda)/(1 - \lambda^n)$. Por tanto, en ambientes estacionarios existe una diferencia adicional entre (10.5) y (10.6) debido al factor de olvido y su normalización asociada. Esto no es así en ambientes no estacionarios.
3. El promedio temporal (10.5) incluye valores del vector de incógnitas para tiempos anteriores a $n - 1$, lo cual hace que la suma promedio arrastre información de iteraciones pasadas, de valores cada vez más alejados del valor de convergencia. Aunque el factor de olvido está ahí precisamente para intentar eliminar esta contribución no deseable, es de esperar que (10.1) tenga un mejor comportamiento que (10.4). Este es el precio a pagar por un algoritmo con un menor número de operaciones por iteración.

Siempre que se reduce la carga computacional partiendo de un algoritmo dado para dar lugar a algoritmos más sencillos, empeora alguna de las propiedades de dichos algoritmos. Si la dispersión de autovalores no afectaba la velocidad de convergencia de ORIV, al eliminar el cómputo de $\Gamma^{-1}(n)$, que era quien contrarrestaba el efecto de los autovalores, se obtiene un conjunto de algoritmos, OGLMS y variantes, sensibles a dicha dispersión. Y si en un paso más de simplificación se sustituye el cálculo de las matrices del algoritmo por promedios vectoriales, no hay más remedio, en aras de la simplicidad, que arrastrar valores del vector incógnita desde índices temporales muy anteriores ralentizando así la convergencia.

En un principio podría parecer muy sutil la diferencia entre AOGMLS y OGLMS1 pero ya se ha discutido cualitativamente para ilustrar su verdadera dimensión. Donde se apreciará cuantitativamente será en el análisis de la convergencia, no sólo en el resultado, sino en todo el desarrollo hasta llegar a él. Ahí se verá que los cálculos se complican enormemente, debido a la interacción del estimador del valor esperado, $\bar{e}(n)$, con el resto de variables aleatorias para todos los índices temporales.

Una vez constituido el algoritmo AOGMLS como un algoritmo distinto, por derecho propio y como fin de un proceso de búsqueda, se procederá a realizar el análisis de su convergencia. En primer lugar se llevará a cabo un análisis más sencillo debido a que las hipótesis a utilizar son más fuertes. Un análisis más detallado se llevará a cabo en subsiguientes apartados.

10.2 Convergencia en media. Aproximación extrema

Se supondrán dos hipótesis de partida:

1. La primera es aceptar que el estimador del valor esperado se comporta como un verdadero valor esperado.
2. La segunda es una hipótesis de independencia, aceptar que el vector de incógnitas (vector de los pesos) a tiempo n sólo depende de las variables involucradas hasta tiempo n .

Con esto se puede escribir:

$$\hat{E}[\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)\mathbf{w}(n-1)] = \hat{E}[\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)]\hat{E}[\mathbf{w}(n-1)]$$

Recordando la expresión (10.4) y haciendo uso de la definición de error a priori queda

$$\begin{aligned} \mathbf{w}(n) &= \mathbf{w}(n-1) + \mu\mathbf{x}(n)\tilde{\mathbf{x}}^t(n)\hat{E}[\tilde{\mathbf{x}}(n)e(n)] = \\ &= \mathbf{w}(n-1) - \mu\mathbf{x}(n)\tilde{\mathbf{x}}^t(n)\hat{E}[\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)]\hat{E}[\mathbf{w}(n-1)] + \mu\mathbf{x}(n)\tilde{\mathbf{x}}^t(n)\hat{E}[\tilde{\mathbf{x}}(n)d(n)] \end{aligned} \quad (10.7)$$

A continuación se construye el vector de error en los pesos como ya es habitual. Se resta a ambos miembros el vector \mathbf{w}_0 y al miembro derecho se le suma y se le resta $\mu\mathbf{x}(n)\tilde{\mathbf{x}}^t(n)\hat{E}[\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)]\hat{E}[\mathbf{w}_0]$ por lo que se llega a

$$\begin{aligned} \Delta\mathbf{w}(n) &= \Delta\mathbf{w}(n-1) - \mu\mathbf{x}(n)\tilde{\mathbf{x}}^t(n)\hat{E}[\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)]\hat{E}[\Delta\mathbf{w}(n-1)] + \\ &\quad + \mu\mathbf{x}(n)\tilde{\mathbf{x}}^t(n)\hat{E}[\tilde{\mathbf{x}}(n)e_0(n)] \end{aligned} \quad (10.8)$$

Siguiendo con la misma técnica desarrollada en el capítulo 4 para el análisis de ORIV, aquí se aproximarán los estimadores por una componente determinista, igual a su valor medio, más una componente estocástica, término de ruido de esta estimación concreta, que se supondrá de media cero e independiente del resto de variables aleatorias:

$$\begin{aligned} \hat{E}[\tilde{\mathbf{x}}(n)\mathbf{x}(n)^t] &= \mathbf{R} + \mathbf{R}^e(n) \\ \hat{E}[\tilde{\mathbf{x}}(n)e_0(n)] &= 0 + \mathbf{r}_z^e(n) \\ \hat{E}[\Delta\mathbf{w}(n-1)] &= \mathbf{E}[\Delta\mathbf{w}(n-1)] + \mathbf{w}^e(n) \end{aligned}$$

Con lo que la expresión anterior queda

$$\begin{aligned} \Delta\mathbf{w}(n) &= \Delta\mathbf{w}(n-1) - \mu\mathbf{x}(n)\tilde{\mathbf{x}}^t(n)(\mathbf{R} + \mathbf{R}^e(n))(\mathbf{E}[\Delta\mathbf{w}(n-1)] + \mathbf{w}^e(n)) + \\ &\quad + \mu\mathbf{x}(n)\tilde{\mathbf{x}}^t(n)\mathbf{r}_z^e(n) \end{aligned} \quad (10.9)$$

Si finalmente se toma valor esperado

$$E[\Delta \mathbf{w}(n)] = [\mathbf{I} - \mu \mathbf{R}^t \mathbf{R}] E[\Delta \mathbf{w}(n-1)]$$

Para estudiar cómo evoluciona esta ecuación en diferencias se trabajará en la base en la que $\mathbf{R}^t \mathbf{R}$ sea diagonal. En ella, como se comprueba fácilmente, las distintas componentes del vector de error en los pesos se desacoplan y se verifica para la componente i -ésima:

$$E[\Delta w_i(n)] = (1 - \mu \lambda_i) E[\Delta w_i(n-1)]$$

donde λ_i es el autovalor i -ésimo de $\mathbf{R}^t \mathbf{R}$. Esta es una ecuación en diferencias homogénea, por lo que su solución tenderá a cero siempre y cuando se verifique que:

$$|1 - \mu \lambda_i| < 1$$

o lo que equivale a decir que:

$$0 < \mu < \frac{2}{\lambda_i}$$

Si se quiere, como es el caso, que todas las componentes tiendan a cero, es decir, no tener sesgo en los estimadores de los pesos, es necesario imponer la condición anterior para todo autovalor de $\mathbf{R}^t \mathbf{R}$.

Por tanto, respecto a la convergencia en media, se podrían hacer los siguientes comentarios:

- La evolución de la media de los errores de los pesos no se ve afectada por la adición de ruido aditivo gaussiano, al depender exclusivamente de los autovalores de $\mathbf{R}^t \mathbf{R}$, siempre y cuando esta matriz \mathbf{R} sea una matriz de cumulantes de orden mayor que 2.
- No importa que la matriz \mathbf{R} no sea definida¹, ya que lo que importa son los autovalores de la matriz $\mathbf{R}^t \mathbf{R}$ que ya sí es definida.
- La convergencia se ve afectada por la dispersión de autovalores. En concreto, cuanto mayor sea el número de condicionamiento $\lambda_{\max}/\lambda_{\min}$ más lenta será su convergencia. En el capítulo 8 se comprobó que GLMS se ve afectado también por la dispersión de autovalores pero de la versión cuadrada de \mathbf{R} , lo cual implica que es un poco menos sensible a este factor. En cierto sentido es como si trabajara con la raíz cuadrada del número de condicionamiento para AOGMLS, porque AOGMLS trabaja con el producto de matrices \mathbf{R} .

¹Téngase en cuenta que ahora esta matriz no es cuadrada. En los capítulos dedicados a GLMS y variantes se trabaja con matrices cuadradas.

10.3 Convergencia en media cuadrática. Promediado directo

En este apartado se calculará la desviación en media cuadrática (Mean Square Deviation, MSD), o lo que es lo mismo, la convergencia en media cuadrática de los errores de los pesos. Como una primera aproximación se aplicará la técnica del promediado directo a la ecuación de partida, ecuación (10.9), quedando:

$$\begin{aligned} \Delta \mathbf{w}(n) = & \Delta \mathbf{w}(n-1) - \mu \mathbf{R}^t \mathbf{R} \mathbf{E}[\Delta \mathbf{w}(n-1)] - \\ & - \mu \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) (\mathbf{R} + \mathbf{R}^e(n)) \mathbf{w}^e(n) + \mu \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \mathbf{r}_z^e(n) \end{aligned} \quad (10.10)$$

y a partir de ella se puede construir el producto externo,

$$\begin{aligned} \Delta \mathbf{w}(n) \Delta \mathbf{w}^t(n) = & \Delta \mathbf{w}(n-1) \Delta \mathbf{w}^t(n-1) - \mu \Delta \mathbf{w}(n-1) \mathbf{E}[\Delta \mathbf{w}^t(n-1)] \mathbf{R}^t \mathbf{R} - \\ & - \mu \mathbf{R}^t \mathbf{R} \mathbf{E}[\Delta \mathbf{w}(n-1)] \Delta \mathbf{w}^t(n-1) + \\ & + \mu^2 \mathbf{R}^t \mathbf{R} \mathbf{E}[\Delta \mathbf{w}(n-1)] \mathbf{E}[\Delta \mathbf{w}^t(n-1)] \mathbf{R}^t \mathbf{R} + \\ & + \mu^2 \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) (\mathbf{R} + \mathbf{R}^e(n)) \mathbf{w}^e(n) \mathbf{w}^{et}(n) (\mathbf{R}^t + \mathbf{R}^{et}(n)) \tilde{\mathbf{x}}(n) \mathbf{x}^t(n) + \\ & + \mu^2 \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \mathbf{r}_z^e(n) \mathbf{r}_z^{et}(n) \tilde{\mathbf{x}}(n) \mathbf{x}^t(n) + \\ & + \text{términos de media nula} \end{aligned} \quad (10.11)$$

En todos los análisis realizados del MSD se ha tomado n suficientemente grande, ya que lo que interesan son las propiedades cerca de la convergencia del algoritmo. En esta circunstancia siempre se ha despreciado $\mathbf{E}[\Delta \mathbf{w}(n-1)]$ ya que el algoritmo converge en media. Si se supusiese ahora, al tomando valor esperado quedaría:

$$\mathbf{R}_\Delta(n) = \mathbf{R}_\Delta(n-1) + \mu^2 (\text{términos de ruido}) \quad (10.12)$$

Para que la solución de esta ecuación en diferencias no diverja para $n \rightarrow \infty$, no tiene que hacerlo en particular la solución de la homogénea. Por tanto, las raíces del polinomio característico tienen que ser menores o iguales que la unidad. En este caso el polinomio característico es $x - 1 = 0$ y tiene como única solución $x = 1$ y por tanto en principio no diverje. El problema está en que el término debido a ruido es escrictamente positivo y por tanto hace que la recursión anterior diverja claramente. Para verlo de otra manera, si se supone que el algoritmo ha converjido y se hace $\mathbf{R}_\Delta(n) = \mathbf{R}_\Delta(n-1) = \mathbf{R}_\Delta(\infty)$ en (10.12), se llega a la conclusión de que ‘términos de ruido’ = 0, es decir, no puede haber ningún término independiente.

Si se quiere imponer convergencia, manteniendo el término independiente, es necesario que la raíz del polinomio característico sea escrictamente menor que la unidad. Esto no se consigue ni introduciendo los términos de $\mathbf{E}[\Delta \mathbf{w}(n-1)]$, ya que lo que realmente habría que introducir son términos en $\mathbf{E}[\Delta \mathbf{w}(n-1) \Delta \mathbf{w}^t(n-1)]$.

Por tanto, en esta aproximación el algoritmo diverge en MSD. Esto no es problema de haber aplicado promediado directo, es consecuencia directa de dotar al estimador del valor esperado de las mismas propiedades que un auténtico valor esperado y por tanto aparece el término $E[\Delta \mathbf{w}(n-1)]$ fuente de todos los problemas.

10.4 Convergencia en media. Segunda aproximación

En este apartado se estudiará la convergencia en media del algoritmo considerando los estimadores como lo que son, una suma de variables aleatorias. Por comodidad se retomarán aquí las expresiones que van a ser útiles. En primer lugar, recuérdese que la ecuación en diferencias que servía como actualización de los pesos es (10.4):

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \mu \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \hat{E}[\tilde{\mathbf{x}}(n) e(n)] \quad (10.13)$$

y a partir de (10.5) se puede continuar deduciendo que:

$$\begin{aligned} \hat{E}[\tilde{\mathbf{x}}(n) e(n)] &= \frac{1-f}{1-f^n} \sum_{i=1}^n f^{n-i} \tilde{\mathbf{x}}(i) e(i) = \\ &= \frac{1-f}{1-f^n} \sum_{i=1}^n f^{n-i} [\tilde{\mathbf{x}}(i) d(i) - \tilde{\mathbf{x}}(i) \mathbf{x}^t(i) \mathbf{w}(i-1)] = \\ &= \hat{E}[\tilde{\mathbf{x}}(n) d(n)] - \hat{E}[\tilde{\mathbf{x}}(n) \mathbf{x}^t(n) \mathbf{w}(n-1)] \end{aligned} \quad (10.14)$$

que sustituida en (10.13), restando \mathbf{w}_0 a ambos miembros y sumando y restando $\hat{E}[\tilde{\mathbf{x}}(n) \mathbf{x}^t(n) \mathbf{w}_0]$ conduce a la ecuación de partida para el análisis subsiguiente (compárese con (10.8)):

$$\begin{aligned} \Delta \mathbf{w}(n) &= \Delta \mathbf{w}(n-1) - \mu \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \hat{E}[\tilde{\mathbf{x}}(n) \mathbf{x}^t(n) \Delta \mathbf{w}(n-1)] + \\ &\quad + \mu \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \hat{E}[\tilde{\mathbf{x}}(n) e_0(n)] \end{aligned} \quad (10.15)$$

Para empezar con el análisis de la convergencia en media se toma valor esperado sobre la expresión anterior. Se supondrá por ahora que el estimador del valor esperado, aunque es una variable aleatoria, es independiente del resto de variables aleatorias. Más adelante se discutirá lo acertado de esta hipótesis. Con ella se puede escribir:

$$\begin{aligned} E[\Delta \mathbf{w}(n)] &= E[\Delta \mathbf{w}(n-1)] - \mu E[\mathbf{x}(n) \tilde{\mathbf{x}}^t(n)] E[\hat{E}[\tilde{\mathbf{x}}(n) \mathbf{x}^t(n) \Delta \mathbf{w}(n-1)]] + \\ &\quad + \mu E[\mathbf{x}(n) \tilde{\mathbf{x}}^t(n)] E[\hat{E}[\tilde{\mathbf{x}}(n) e_0(n)]] \end{aligned} \quad (10.16)$$

Para continuar es necesario calcular los siguientes valores esperados:

$$\begin{aligned} \mathbb{E}[\hat{\mathbb{E}}[\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)\Delta\mathbf{w}(n-1)]] &= \frac{1-f}{1-f^n} \sum_{i=1}^n f^{n-i} \mathbb{E}[\tilde{\mathbf{x}}(i)\mathbf{x}^t(i)\Delta\mathbf{w}(i-1)] = \\ &= \frac{1-f}{1-f^n} \mathbf{R} \sum_{i=1}^n f^{n-i} \mathbb{E}[\Delta\mathbf{w}(i-1)] \end{aligned} \quad (10.17)$$

y

$$\mathbb{E}[\hat{\mathbb{E}}[\tilde{\mathbf{x}}(n)e_0(n)]] = \frac{1-f}{1-f^n} \sum_{i=1}^n f^{n-i} \mathbb{E}[\tilde{\mathbf{x}}(i)e_0(i)] = 0$$

Para obtener la primera de ellas se ha dado por válida la hipótesis de independencia que dice que $\mathbf{w}(n)$ sólo depende del resto de variables aleatorias hasta tiempo n , es decir, de $\tilde{\mathbf{x}}(1), \dots, \tilde{\mathbf{x}}(n), \mathbf{x}(1), \dots, \mathbf{x}(n), d(1), \dots, d(n)$. Para la segunda se ha utilizado el principio de ortogonalidad, fundamento del problema de estimación de una variable aleatoria mediante la combinación lineal de otras.

Con esto, (10.16) se convierte en:

$$\mathbb{E}[\Delta\mathbf{w}(n)] = \mathbb{E}[\Delta\mathbf{w}(n-1)] - \mu \mathbf{R}^t \mathbf{R} \frac{1-f}{1-f^n} \sum_{i=1}^n f^{n-i} \mathbb{E}[\Delta\mathbf{w}(i-1)]$$

es decir, en una ecuación en diferencias homogénea de orden n , donde n va aumentando en principio indefinidamente. Posteriormente se verá que el factor de olvido hace que el orden de esta ecuación en diferencias sea finito y pueda ser resuelto por los métodos tradicionales.

Al ser una ecuación en diferencias homogénea se tiene asegurado que su solución, en caso de que no diverja, tiende a cero. La homogeneidad se la ha conferido el hecho de haber supuesto que el estimador del valor esperado es independiente del resto de variables aleatorias y por tanto se pudo aplicar el principio de ortogonalidad.

La ecuación en diferencias anterior se desacopla para cada componente del vector de error en los pesos trabajando en la base en la que la matriz $\mathbf{R}^t \mathbf{R}$ es diagonal. Para cada componente vuelve a quedar otra ecuación en diferencias que se resuelve suponiendo una solución del tipo $\mathbb{E}[\Delta w_i(n)] = x_i^n$. Lo que conduce al polinomio característico siguiente cuyas raíces es necesario calcular:

$$\begin{aligned} x_i^n &= x_i^{n-1} - \mu \lambda_i \frac{1-f}{1-f^n} \sum_{k=0}^{n-1} f^{n-1-k} x_i^k = \\ &= x_i^{n-1} - \mu \lambda_i \frac{1-f}{1-f^n} \frac{x_i^n - f^n}{x_i - f} \end{aligned} \quad (10.18)$$

Como el factor de olvido, para realizar esta función, va a ser menor que la unidad, para n suficientemente grandes se pueden despreciar los términos en f^n

frente a 1:

$$x_i^n = x_i^{n-1} - \mu\lambda_i(1-f)\frac{x_i^n}{x_i-f}$$

Reagrupando términos se puede escribir:

$$x_i^{n-1} \left(x_i \left(1 + \mu\lambda_i \frac{1-f}{x_i-f} \right) - 1 \right) = 0$$

que tiene como soluciones la trivial por un lado $x_i = 0$ y los ceros de la siguiente expresión:

$$1 + \mu\lambda_i \frac{1-f}{x_i-f} = \frac{1}{x_i}$$

Esta expresión presenta las siguientes características:

- No depende del instante temporal n gracias al factor de olvido f , con lo que las raíces de esta expresión serán válidas para construir a partir de ellas la solución de la ecuación en diferencias para todo tiempo.
- En ella aparecen todos los factores de los que depende el sistema, el tamaño del paso, el factor de olvido y los autovalores de la matriz de autocorrelación.
- Es una ecuación de segundo orden:

$$x_i^2 + [\mu\lambda_i(1-f) - (1+f)]x_i + f = 0 \quad (10.19)$$

Esta última ecuación tiene como solución:

$$x_i = \frac{1+f - \mu\lambda_i(1-f) \pm \sqrt{[\mu\lambda_i(1-f) - (1+f)]^2 - 4f}}{2} \quad (10.20)$$

Como se puede apreciar, las raíces del polinomio característico son función del producto tamaño del paso-autovalor y del factor de olvido $x = x(\mu\lambda_i, f)$. Representando gráficamente esta función se pueden apreciar las zonas donde la solución de la ecuación en diferencias converge a cero (zonas donde el módulo de las raíces es menor que la unidad). Cuanto más cercana a la unidad sea la raíz, más lento convergerá y cuanto más cercana a 0 esté la raíz más rápido convergerá. Como son posibles dos valores para la raíz (10.20), en la figura 10.1 se ha representado la raíz con mayor módulo para cada par de valores $(\mu\lambda_i, f)$.

Como se puede apreciar en la figura, existe una región bastante amplia donde la raíz casi no depende del valor del factor $\mu\lambda_i$, siendo únicamente función del factor de olvido. Esta región se expande en la zona de f cercanos a la unidad. Para el caso particular $f \rightarrow 1$, la raíz tiende a su vez a 1 lo cual quiere decir que los algoritmos con memoria infinita² no convergen en media, ya que no pueden olvidar los valores iniciales.

²Recuérdese que la memoria o longitud efectiva del algoritmo se había definido en el capítulo 5 en la página 130 como $L = 1/(1-f)$.

Análogamente, si el tamaño del paso se hace muy pequeño se le está quitando al algoritmo capacidad de evolucionar, por lo que de nuevo se espera que la raíz tienda a uno, como así ocurre.

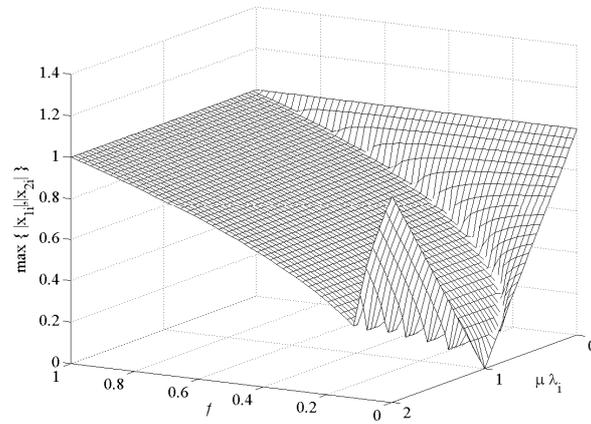


Figura 10.1 Raíz de mayor módulo

En el extremo opuesto, si se considera el límite de factor de olvido muy pequeño, cero por ejemplo, se llega a la expresión del algoritmo CGLMS que se estudiará en capítulos posteriores como modificación del GLMS, además se deducirá a partir de otras motivaciones. Como se verá en dicho capítulo, la ecuación de actualización de los pesos viene dada por:

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \mu \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \tilde{\mathbf{x}}(n) e(n)$$

Las características concretas de este nuevo algoritmo no interesan por ahora, salvo la convergencia en media. Se demostrará expresamente que la raíz (única), sin tener en cuenta la trivial, es:

$$x_i = 1 - \mu \lambda_i \quad (10.21)$$

esta expresión también se puede obtener como límite para factor de olvido nulo de (10.20). A partir de ella se puede llegar a una conclusión llamativa. Si se elige $\mu \lambda_i = 1$ resulta³ que $x_i = 0$, es decir, el algoritmo converge en un sólo paso!!. Esta consecuencia teórica no se puede llevar a la práctica ya que el tamaño del paso toma un valor fijo y por tanto no se puede imponer la condición anterior para todos los autovalores. Aún en el caso de que se tuviera un sólo autovalor, y el análisis hasta aquí realizado garantizara la convergencia en media, nada se sabe de la convergencia en media cuadrática. Además en el límite $f \rightarrow 0$ se

³Véase en la figura 10.1 el punto ($\mu \lambda_i = 1, f = 0$).

pierden muchas de las propiedades de estos algoritmos que son deseables, entre ellas la independencia al ruido, tal y como se verá más adelante.

En definitiva, AGLMS converge en media bajo ciertas condiciones dando estimadores no sesgados. El carácter no sesgado de los estimadores se lo ha concedido el hecho de considerar el estimador del valor esperado como independiente del resto de variables aleatorias. Suponer $f = 0$ elimina dicho estimador (algoritmo CGLMS), dando lugar a estimadores sesgados bajo ciertas condiciones.

10.5 Convergencia en media cuadrática. Segunda aproximación

Para calcular la desviación en media cuadrática hay que partir de la ecuación de actualización del error de los pesos, que por comodidad se reproduce aquí:

$$\begin{aligned} \Delta \mathbf{w}(n) = & \Delta \mathbf{w}(n-1) - \mu \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \hat{E}[\tilde{\mathbf{x}}(n) \mathbf{x}^t(n) \Delta \mathbf{w}(n-1)] + \\ & + \mu \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \hat{E}[\tilde{\mathbf{x}}(n) e_0(n)] \end{aligned} \quad (10.22)$$

Ya se ha podido comprobar en los apartados anteriores, que gran parte de la complejidad de los análisis que se van a llevar a cabo, está en ver qué tratamiento se le puede dar al estimador del valor esperado. Es conveniente que sea lo más sencillo posible para facilitar la matemática posterior, pero con la suficiente riqueza en contenido de información como para que dé resultados aceptables. En el apartado 10.3 se puso de manifiesto que no es posible aproximar el estimador por el valor verdadero (un auténtico valor esperado), ya que entonces aparecía el factor $E[\Delta \mathbf{w}(n-1)]$ que daba como resultado unas ecuaciones para el MSD divergentes. Por tanto, la posible (y real) dependencia del estimador de $E[\tilde{\mathbf{x}}(n) \mathbf{x}^t(n) \Delta \mathbf{w}(n-1)]$ con el resto de variables aleatorias es crítica en este análisis.

Lo que se propone a continuación es una primera aproximación para resolver este problema. Se va a tratar al estimador como lo que es: una suma de variables aleatorias y se va a hacer uso de las hipótesis de independencia. Por tanto, si

se estudian los términos conflictivos se puede escribir:

$$\begin{aligned}
\hat{E}[\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)\Delta\mathbf{w}(n-1)] &= \frac{1-f}{1-f^n} \sum_{i=1}^n f^{n-i} \tilde{\mathbf{x}}(i)\mathbf{x}^t(i)\Delta\mathbf{w}(i-1) = \\
&= \frac{1-f}{1-f^n} \left(\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)\Delta\mathbf{w}(n-1) + \sum_{i=1}^{n-1} f^{n-i} \tilde{\mathbf{x}}(n)\mathbf{x}^t(n)\Delta\mathbf{w}(i-1) \right) = \\
&= \frac{1-f}{1-f^n} \tilde{\mathbf{x}}(n)\mathbf{x}^t(n)\Delta\mathbf{w}(n-1) + \frac{1-f^{n-1}}{1-f^n} f \hat{E}[\tilde{\mathbf{x}}(n-1)\mathbf{x}^t(n-1)\Delta\mathbf{w}(n-2)]
\end{aligned} \tag{10.23}$$

donde se ha extraído la parte del estimador que más interesa, ya que es la que más dependencia presenta con el resto de miembros de la ecuación en diferencias. Una vez hecho esto se pueden volver a aplicar las hipótesis simplificatorias anteriores; en particular se va a tratar el estimador como un auténtico valor esperado y como se va a suponer n suficientemente grande, dado que el algoritmo converge en media, se tomará $E[\Delta\mathbf{w}(n-1)] = 0$. Es decir,

$$\begin{aligned}
\hat{E}[\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)\Delta\mathbf{w}(n-1)] &\underset{n \rightarrow \infty}{\sim} (1-f)\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)\Delta\mathbf{w}(n-1) + \\
&+ f\mathbf{R} \sum_{i=1}^{n-1} f^{n-1-i} E[\Delta\mathbf{w}(i-1)] = \\
&= (1-f)\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)\Delta\mathbf{w}(n-1)
\end{aligned} \tag{10.24}$$

Aunque parezca una aproximación muy radical, la decisión final sobre si es acertada o no la darán los resultados que con ella se deriven. La interpretación de esta aproximación es bastante clara: idealmente el estimador del valor esperado es igual al valor esperado de la variable aleatoria que se considere; dicho valor esperado, en este caso, en virtud de la hipótesis de independencia y de la convergencia en media del algoritmo es 0,

$$\hat{E}[\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)\Delta\mathbf{w}(n-1)] = E[\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)\Delta\mathbf{w}(n-1)] = \mathbf{R}E[\Delta\mathbf{w}(n-1)] \underset{n \rightarrow \infty}{\rightarrow} 0$$

pero el estimador, como tal, presenta una componente aleatoria que es la que representa el término dado en (10.24). Se podría discurrir de la misma manera para llegar al resultado análogo,

$$\hat{E}[\tilde{\mathbf{x}}(n)e_0(n)] \underset{n \rightarrow \infty}{\sim} (1-f)\tilde{\mathbf{x}}(n)e_0(n) \tag{10.25}$$

con el mismo significado de la aproximación anterior: su valor medio (cero en este caso también) más su correspondiente componente aleatoria por ser un estimador.

Por tanto, con lo discutido hasta ahora, se puede poner que la ecuación en diferencias que controla la evolución del vector de la desviación de los pesos es,

$$\begin{aligned}
\Delta\mathbf{w}(n) &= [\mathbf{I} - \mu(1-f)\mathbf{x}(n)\tilde{\mathbf{x}}^t(n)\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)]\Delta\mathbf{w}(n-1) + \\
&+ \mu(1-f)\mathbf{x}(n)\tilde{\mathbf{x}}^t(n)\tilde{\mathbf{x}}(n)e_0(n)
\end{aligned} \tag{10.26}$$

Se obtiene una ecuación análoga a la del CGLMS pero con el tamaño del paso multiplicado por el factor $(1 - f)$. A continuación se estudia cómo afecta esto al comportamiento de AOGMS. Con las hipótesis usuales de independencia se puede escribir inmediatamente,

$$\begin{aligned} \mathbf{R}_\Delta(n) = & \mathbf{R}_\Delta(n-1) - \mu(1-f)\mathbf{\Gamma}\mathbf{R}_\Delta(n-1) - \mu(1-f)\mathbf{R}_\Delta(n-1)\mathbf{\Gamma} + \\ & + \mu^2(1-f)^2\mathbf{E}[\boldsymbol{\psi}_1(n)\mathbf{R}_\Delta(n-1)\boldsymbol{\psi}_1(n)] + \mu^2(1-f)^2\mathbf{E}[e_o^2(n)\boldsymbol{\psi}_2(n)] \end{aligned} \quad (10.27)$$

con $\boldsymbol{\psi}_1(n) = \mathbf{x}(n)\tilde{\mathbf{x}}^t(n)\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)$; $\boldsymbol{\psi}_2(n) = \mathbf{x}(n)\tilde{\mathbf{x}}^t(n)\tilde{\mathbf{x}}(n)\tilde{\mathbf{x}}^t(n)\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)$ y $\mathbf{\Gamma} = \mathbf{E}[\boldsymbol{\psi}_1(n)]$. A la vista de esta nueva expresión se puede hacer una simplificación más, dado que el tamaño del paso suele ser pequeño y que el factor de olvido es cercano a la unidad, se puede tomar el producto $\mu(1-f)$ como parámetro perturbativo para desprestigiar el cuarto sumando del segundo miembro. El último sumando no se desprestigia porque cumple la función de término independiente, cuyo concurso es muy importante en la ecuación en diferencias ya que hace que la solución nula no sea una posible solución de ella.

Si además se trabaja en la base en la que $\mathbf{\Gamma}$ es diagonal⁴, las distintas componentes diagonales $D'_i(n)$ de $\mathbf{R}'_\Delta(n)$ se desacoplan, quedando para cada una de ellas,

$$D'_i(n) = [1 - 2\mu(1-f)\lambda_{\Gamma_i}]D'_i(n-1) + \mu^2(1-f)^2\Psi'_{eii}$$

donde λ_{Γ_i} es el autovalor i -ésimo de la matriz $\mathbf{\Gamma}$ y Ψ'_{eii} es el elemento (i,i) de la matriz $\Psi'_e = \mathbf{E}[e_o^2(n)\boldsymbol{\psi}_2(n)]$. Llegando a una ecuación en diferencias para escalares fácilmente resoluble, a partir de la cual se pueden estudiar aspectos tales como condiciones de convergencia, su tiempo de convergencia o el valor en convergencia.

- A) La solución general de la homogénea no divergerá siempre y cuando se cumpla que $|1 - 2\mu(1-f)\lambda_{\Gamma_i}| \leq 1$. Lo que equivale a decir que

$$0 < \mu(1-f) < \lambda_{\Gamma_i}$$

- B) El tiempo de convergencia, dentro de la aproximación utilizada de tamaño del paso pequeño, vale

$$\tau_i = \frac{1}{2\mu(1-f)\lambda_{\Gamma_i}}$$

- C) La componente diagonal i -ésima, es decir la (i,i) , converge al valor,

$$D'_i(\infty) = \frac{\mu(1-f)}{2} \frac{\Psi'_{eii}}{\lambda_{\Gamma_i}}$$

⁴La prima ' indica cantidades en una base distinta.

La intención de este apartado es calcular la desviación en media cuadrática, MSD, $D(n) = \text{Tr}\{\mathbf{R}_\Delta(n)\}$ y en particular su valor en convergencia y el tiempo que tarda. El valor en convergencia se puede calcular directamente a partir del apartado C) anterior:

$$\begin{aligned} D(\infty) &= \text{Tr}\{\mathbf{R}_\Delta(\infty)\} = \text{Tr}\{\mathbf{R}'_\Delta(\infty)\} = \sum_{i=1}^q D'_i(\infty) = \frac{\mu(1-f)}{2} \sum_{i=1}^q \frac{\Psi'_{eii}}{\lambda_{\Gamma_i}} = \\ &= \frac{\mu(1-f)}{2} \sum_{i=1}^q (\mathbf{\Gamma}'^{-1} \mathbf{\Psi}'_e)_{ii} = \frac{\mu(1-f)}{2} \text{Tr}\{\mathbf{\Gamma}'^{-1} \mathbf{\Psi}'_e\} = \\ &= \frac{\mu(1-f)}{2} \text{Tr}\{\mathbf{\Gamma}^{-1} \mathbf{\Psi}_e\} \end{aligned} \quad (10.28)$$

Y el tiempo de convergencia total se puede calcular como el tiempo de convergencia promedio de todas sus componentes:

$$\begin{aligned} \tau &= \frac{1}{2\mu(1-f)} \frac{1}{q} \sum_{i=1}^q \lambda_{\Gamma_i}^{-1} = \frac{1}{2\mu(1-f)} \frac{1}{q} \text{Tr}\{\mathbf{\Gamma}'^{-1}\} = \\ &= \frac{1}{2\mu(1-f)} \frac{1}{q} \text{Tr}\{\mathbf{\Gamma}^{-1}\} \end{aligned} \quad (10.29)$$

Como se puede apreciar a la vista de estos tres apartados y los resultados finales, el comportamiento en media cuadrática de AOGLMS depende, en referencia a los parámetros controlables externamente, del producto $\mu(1-f)$. A este parámetro se le puede dar el mismo papel que se le daba a μ solo, en algoritmos tipo LMS: al aumentar disminuye el tiempo de convergencia, pero en contra se obtiene un valor en convergencia mayor. Hay intereses contrapuestos entre un tiempo de convergencia pequeño y valor en convergencia pequeño, o se consigue una cosa u otra, pero no ambas: o algoritmos que evolucionen rápidamente o por el contrario algoritmos que aun yendo más despacio, den menos error de estimación.

10.6 Fundamentos de la segunda aproximación

Recordemos que la segunda aproximación se utilizó en el apartado anterior y se construyó extrayendo del término conflictivo, el estimador de $E[\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)e(n)]$, la parte que estadísticamente contenía términos más dependientes con el resto de variables que aparecen en la ecuación de actualización de los pesos, es decir, la información a tiempo n . Por tanto, los estimadores quedaron de la siguiente manera:

$$\begin{aligned} \hat{E}[\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)\Delta\mathbf{w}(n-1)] &\underset{n \rightarrow \infty}{\sim} (1-f)\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)\Delta\mathbf{w}(n-1) \\ \hat{E}[\tilde{\mathbf{x}}(n)e_0(n)] &\underset{n \rightarrow \infty}{\sim} (1-f)\tilde{\mathbf{x}}(n)e_0(n) \end{aligned} \quad (10.30)$$

En este apartado se pretende estudiar las consecuencias de usar estas aproximaciones. Es decir, se compararán las ecuaciones obtenidas sin ninguna simplificación con las del apartado anterior, basadas en las aproximaciones (10.30), para poder discutir bajo qué condiciones ambas ecuaciones serían idénticas: si se ha asumido implícitamente alguna otra hipótesis.

10.6.1 En el cálculo de $\mathbf{E}[\Delta \mathbf{w}(n)]$:

Si se tiene en cuenta la ecuación de actualización del vector de error en los pesos (10.15), junto con la definición para el estimador del valor esperado dada en (10.3), se llega a

$$\begin{aligned} \Delta \mathbf{w}(n) = & \left[\mathbf{I} - \mu \frac{1-f}{1-f^n} \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \tilde{\mathbf{x}}(n) \mathbf{x}^t(n) \right] \Delta \mathbf{w}(n-1) - \\ & - \mu \frac{1-f}{1-f^n} \sum_{i=1}^{n-1} f^{n-i} \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \tilde{\mathbf{x}}(i) \mathbf{x}^t(i) \Delta \mathbf{w}(i-1) + \\ & + \mu \frac{1-f}{1-f^n} \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \tilde{\mathbf{x}}(n) e_0(n) + \\ & + \mu \frac{1-f}{1-f^n} \sum_{i=1}^{n-1} f^{n-i} \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \tilde{\mathbf{x}}(i) e_0(i) \end{aligned} \quad (10.31)$$

Normalmente siempre se han aceptado como válidas las hipótesis de independencia, aunque sólo sea por cuestiones prácticas, aquí no se hará una excepción. Las utilizadas aquí son:

1. La serie de vectores $\{\mathbf{x}(m)\}$ y $\{\tilde{\mathbf{x}}(m)\}$ son independientes e idénticamente distribuidos.
2. La variable aleatoria $\Delta \mathbf{w}(n)$ sólo depende de los elementos de las series $\{\mathbf{x}(m)\}$ y $\{\tilde{\mathbf{x}}(m)\}$ hasta tiempo n .

Con las que se llega, tomando valor esperado a

$$\begin{aligned} \mathbf{E}[\Delta \mathbf{w}(n)] = & \left[\mathbf{I} - \mu \frac{1-f}{1-f^n} \mathbf{\Gamma} \right] \mathbf{E}[\Delta \mathbf{w}(n-1)] - \\ & - \mu \frac{1-f}{1-f^n} \mathbf{R}^t \mathbf{R} \sum_{i=1}^{n-1} f^{n-i} \mathbf{E}[\Delta \mathbf{w}(i-1)] + \\ & + \mu \frac{1-f}{1-f^n} \mathbf{E}[\mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \tilde{\mathbf{x}}(n) e_0(n)] \end{aligned} \quad (10.32)$$

Por otro lado, si se hubiesen utilizado las aproximaciones de (10.30) se habría llegado a:

$$E[\Delta \mathbf{w}(n)] = [\mathbf{I} - \mu(1-f)\mathbf{\Gamma}]E[\Delta \mathbf{w}(n-1)] - \mu(1-f)E[\mathbf{x}(n)\tilde{\mathbf{x}}^t(n)\tilde{\mathbf{x}}(n)e_0(n)] \quad (10.33)$$

En consecuencia, las aproximaciones (10.30) eliminan la sumatoria de (10.32) y mantienen el término independiente de la ecuación en diferencias, lo que lleva a resaltar las dos cuestiones siguientes:

- i) Asumir (10.30) implica eliminar una parte importante de la dinámica de la ecuación en diferencias para la convergencia en media al suprimir los términos del vector de error en los pesos para tiempos anteriores a $n-1$. La existencia de estos términos hace que se modifiquen aspectos de la convergencia en media como las condiciones bajo las cuales se produce la misma, como puede apreciarse en el estudio llevado a cabo en el apartado 10.4 y resumido en la figura 10.1. Básicamente esta aproximación supone suprimir la memoria del estimador del valor esperado, algo que puede parecer paradójico porque lo que se pretende con este algoritmo es precisamente introducir tal memoria. De todas formas no hay que perder de vista que esta aproximación se dedujo para utilizarla en el análisis de la convergencia en media cuadrática, para el cual lo que importa principalmente son las posibles fuentes de ruido, las oscilaciones de las variables aleatorias respecto de su media. Esto sí se tiene en cuenta, para n suficientemente grandes en (10.30) y en consecuencia dio resultados aceptables para la convergencia en media cuadrática (confirma la convergencia). En definitiva, (10.30) no es aceptable para estudiar la dinámica de la convergencia en media, pero se presenta lo suficientemente válido para la media cuadrática (al menos es la primera aproximación que da convergencia en media cuadrática).
- ii) Tanto en la aproximación más realista (10.32) como en (10.33) aparece un término independiente que conduce irremediablemente a un sesgo en los estimadores de los pesos. De todas maneras, en condiciones de convergencia usuales, con el factor de olvido cercano a la unidad, el sesgo será bastante pequeño.

La necesidad de que el factor de olvido sea cercano a la unidad para reducir el sesgo es un punto a favor para conservar el término de la sumatoria en (10.32), responsable de la evolución del algoritmo. Si en (10.33) se hiciese $f \sim 1$, el algoritmo convergería muy lento en media, la existencia de más términos anteriores al tiempo de estudio en la ecuación en convergencia, favorecen una dinámica más activa.

Sería deseable que al ser un algoritmo basado en estadística de alto orden fuese inmune al ruido aditivo gaussiano, para ello es necesario que todas las magnitudes que aparecen en (10.32) lo fuesen. La matriz \mathbf{R} está formada

por cumulantes de orden mayor que 2, sin embargo la matriz $\mathbf{\Gamma}$ y el vector $E[\mathbf{x}(n)\tilde{\mathbf{x}}^t(n)\tilde{\mathbf{x}}(n)e_0(n)]$ no, por lo que estos últimos se verán afectados por el ruido. Que $\mathbf{\Gamma}$ dependa del ruido afecta al tiempo de convergencia que se puede corregir mediante una elección adecuada del tamaño del paso, pero el vector mencionado afecta directamente al sesgo del estimador del vector de los pesos, cosa nada deseable; la ventaja de este algoritmo es que el factor de olvido atenúa este efecto.

De esta discusión anterior se puede deducir que el estimador del valor esperado cumple su función:

1. Permite la resolución de problemas sobredeterminados.
2. Tiene una carga computacional por iteración baja.
3. No es estrictamente inmune al ruido aditivo, pero es capaz de reducir su efecto al máximo.
4. La inmunidad asintótica al ruido no le hace ir infinitamente lento en cuanto a convergencia, es decir, tiene un tiempo de convergencia finito.

Recuérdese no obstante que se ha demostrado que en el límite $f \sim 1$ el tiempo de convergencia tiende a infinito (la raíz del polinomio característico tiende a 1, ver figura 10.1). Es decir, la inmunidad total al ruido se consigue a cambio de una convergencia infinitamente lenta, en completo paralelismo con la familia de algoritmos de máxima pendiente aunque el mecanismo que consigue este efecto sea distinto: conceptualmente no es lo mismo el tamaño del paso que el factor de olvido.

Las condiciones normales de trabajo se sitúan en un término medio: entre un factor de olvido nulo (CGLMS) y un factor de olvido que tiende a la unidad (situación impráctica). Se tendrá un tiempo de convergencia aceptable a cambio de un pequeño sesgo.

En definitiva, si se aceptase (10.33) en vez de (10.32) no se estarían incluyendo hipótesis nuevas sino que se estarían simplificando las expresiones por motivos de sencillez matemática, pero esas simplificaciones llevarían a ecuaciones donde se pierde una gran cantidad de información que es fácilmente extraíble como se vio en el apartado 10.4.

10.6.2 En el cálculo de $E[\Delta\mathbf{w}(n)\Delta\mathbf{w}^t(n)]$.

Se parte de la ecuación en diferencias para la actualización del vector de error en los pesos, planteada según se recoge en (10.31) y reproducida aquí por

conveniencia

$$\begin{aligned}
\Delta \mathbf{w}(n) &= \boxed{1} + \boxed{2} + \boxed{3} + \boxed{4} \\
\boxed{1} &= [\mathbf{I} - \mu \frac{1-f}{1-f^n} \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \tilde{\mathbf{x}}(n) \mathbf{x}^t(n)] \Delta \mathbf{w}(n-1) \\
\boxed{2} &= -\mu \frac{1-f^{n-1}}{1-f^n} \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) f \hat{\mathbf{E}}[\tilde{\mathbf{x}}(n-1) \mathbf{x}^t(n-1) \Delta \mathbf{w}(n-2)] \\
\boxed{3} &= \mu \frac{1-f}{1-f^n} \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \tilde{\mathbf{x}}(n) e_0(n) \\
\boxed{4} &= \mu \frac{1-f^{n-1}}{1-f^n} \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) f \hat{\mathbf{E}}[\tilde{\mathbf{x}}(n-1) e_0(n-1)]
\end{aligned} \tag{10.34}$$

A continuación se construirá el producto $\Delta \mathbf{w}(n) \Delta \mathbf{w}^t(n)$ discutiendo término a término las hipótesis que hay que realizar para, una vez tomado valor esperado, obtener (10.27) reproducida a continuación:

$$\begin{aligned}
\mathbf{R}_\Delta(n) &= \mathbf{R}_\Delta(n-1) - \mu(1-f) \mathbf{\Gamma} \mathbf{R}_\Delta(n-1) - \mu(1-f) \mathbf{R}_\Delta(n-1) \mathbf{\Gamma} + \\
&\quad + \mu^2(1-f)^2 \mathbf{E}[\boldsymbol{\psi}_1(n) \mathbf{R}_\Delta(n-1) \boldsymbol{\psi}_1(n)] + \mu^2(1-f)^2 \mathbf{E}[e_o^2(n) \boldsymbol{\psi}_2(n)]
\end{aligned} \tag{10.35}$$

Término $\boxed{1} \boxed{1}^t$:

Una vez realizado el producto indicado y tomando valor esperado, da lugar a los términos que contienen a $\mathbf{R}_\Delta(n-1)$ en (10.35).

Término $\boxed{1} \boxed{2}^t$:

Si además de suponer que $\{\tilde{\mathbf{x}}(n)\}$ y $\{\mathbf{x}(n)\}$ son i.i.d se supone que también lo es $\{\mathbf{w}(n)\}$, al tomar valor esperado queda una combinación de $\mathbf{E}[\Delta \mathbf{w}(i)]$ con $i = 0 \dots n-2$. La estrategia usual es reducir el campo de trabajo al caso de n grandes, estudiando valores en convergencia, de manera que este valor esperado tiende a cero. Eso es posible para n grandes, pero no para índices i que también se mueven en el intervalo inicial de tiempos, pero estos sumandos están afectados por el factor de olvido f^{n-1-i} , que consigue precisamente eso, que se puedan olvidar (despreciar).

Nótese que $\Delta \mathbf{w}(n-1)$ depende⁵ de $\tilde{\mathbf{x}}(i)$ y de $\mathbf{x}(i)$, por lo que no se puede decir sin más que se tiene el factor $\mathbf{E}[\Delta \mathbf{w}(n-1)]$ que tiende a cero para n grandes. Hay que recurrir al razonamiento del párrafo anterior.

⁵ En esta dependencia el tamaño del paso μ juega el papel de un factor de olvido (véase (10.4)).

Por tanto este término no contribuye.

Término $\boxed{1 \mid 3}$:

Ahora sí se puede recurrir al hecho de que $\mathbf{w}(n-1)$ es independiente del resto de variables aleatorias que aparecen en este término, y que al tomar valor esperado se obtendrá el factor $E[\Delta \mathbf{w}(n-1)]$, que tiende a cero para n grandes.

Por tanto este término tampoco contribuye.

Término $\boxed{1 \mid 4}$:

Tiende a cero por dos motivos: por el mismo que en el caso anterior y porque el estimador de $E[\tilde{\mathbf{x}}(n-1)e_0(n-1)]$ es independiente del resto y al tomar valor esperado vale cero idénticamente.

Por tanto no contribuye.

Término $\boxed{2 \mid 2}$:

En principio no saldría cero sino

$$\boxed{2 \mid 2}^t = f^2 \mu^2 E[\mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \mathbf{M} \tilde{\mathbf{x}}(n) \mathbf{x}^t(n)] \quad (10.36)$$

donde la matriz \mathbf{M} viene definida por la siguiente expresión

$$\mathbf{M} = E[\hat{E}[\tilde{\mathbf{x}}(n-1) \mathbf{x}^t(n-1) \Delta \mathbf{w}(n-2)] \hat{E}[\Delta \mathbf{w}^t(n-2) \mathbf{x}(n-1) \tilde{\mathbf{x}}^t(n-1)]]$$

si se utiliza la definición de estimador de valor esperado para desarrollar esta expresión, se puede escribir lo siguiente

$$\begin{aligned} \mathbf{M} &= \left(\frac{1-f}{1-f^{n-1}} \right)^2 \sum_{ij} f^{n-1-i} f^{n-1-j} E[\tilde{\mathbf{x}}(i) \mathbf{x}^t(i) \Delta \mathbf{w}(i-1) \\ &\quad \Delta \mathbf{w}^t(j-1) \mathbf{x}(j) \tilde{\mathbf{x}}^t(j)] = \quad (10.37) \\ &= \left(\frac{1-f}{1-f^{n-1}} \right)^2 \sum_{i=1}^{n-1} f^{2(n-1-i)} E[\tilde{\mathbf{x}}(i) \mathbf{x}^t(i) \mathbf{R}_\Delta(i-1) \mathbf{x}(i) \tilde{\mathbf{x}}^t(i)] \end{aligned}$$

donde se ha empleado que la serie $\{\Delta \mathbf{w}(n)\}$ es i.i.d y el hecho de que aunque $\mathbf{w}(n)$ dependa de las variables $\mathbf{x}(k)$ y $\tilde{\mathbf{x}}(k)$ para $k = 1 \dots n$, esta dependencia va disminuyendo conforme más se alejan temporalmente la primera de las últimas,

debido a que el tamaño del paso es pequeño (ver de nuevo nota a pie de página 5). Lo importante de las expresiones (10.36) y (10.37) está en que conforman el equivalente al término de la sumatoria en (10.32) para la convergencia en media: incluye información de la incógnita en la ecuación en diferencias a tiempos pasados, convirtiéndola en una ecuación en diferencias de orden n . Este término es consecuencia directa de la no independencia del valor esperado con el resto de variables aleatorias (incluido él mismo).

Volviendo a tener en cuenta la definición de estimador de valor esperado, la ecuación anterior se puede reescribir como

$$\mathbf{M} = (1 - f)\mathbf{E}[\hat{\mathbf{E}}[\tilde{\mathbf{x}}(n-1)\mathbf{x}^t(n-1)\mathbf{R}_\Delta(n-2)\mathbf{x}(n-1)\tilde{\mathbf{x}}(n-1)]|_{f^2}]$$

Obviamente, si se desea recuperar (10.35) es necesario despreciar este término. Es un término en μ^2 , con lo cual hay que respetarlo al igual que se han respetado otros del mismo orden. Bien es verdad que la influencia de los $\mathbf{R}_\Delta(n-i)$ $i > 1$ va decayendo a cero con $f^{2(n-i)}$ pero como hay $n \rightarrow \infty$ términos el efecto es apreciable (finito).

En principio este término contribuye.

Término $\boxed{2|3}$:

Aquí sí se puede suponer, gracias a las hipótesis de independencia, que el estimador del valor esperado es independiente del resto. Por tanto, al tomar valor esperado aparecerá el factor $\mathbf{E}[\Delta\mathbf{w}(n)]$ que tiende a cero para n grandes.

Por tanto este término no contribuye.

Término $\boxed{2|4}$:

En principio los dos estimadores que aparecen en este factor son dependientes; su valor esperado es

$$\begin{aligned} & \mathbf{E}[\hat{\mathbf{E}}[\tilde{\mathbf{x}}(n-1)\mathbf{x}^t(n-1)\Delta\mathbf{w}(n-2)]\hat{\mathbf{E}}[\tilde{\mathbf{x}}^t(n-1)\mathbf{e}_0(n-1)]] = \\ & = \left(\frac{1-f}{1-f^{n-1}}\right)^2 \sum_{ij} f^{n-1-i} f^{n-1-j} \mathbf{E}[\tilde{\mathbf{x}}(i)\mathbf{x}^t(i)\Delta\mathbf{w}(i-1)\tilde{\mathbf{x}}^t(j)\mathbf{e}_0(j)] \end{aligned} \quad (10.38)$$

Para analizar este término se fija, por ejemplo, un valor de i arbitrario i_0 y a partir de él se puede afirmar que:

1. Los términos con $j > i_0$ son cero pues en ellos el factor $\tilde{\mathbf{x}}(j)e_0(j)$ es independiente del resto y su valor esperado es nulo en virtud del principio de ortogonalidad.
2. Los términos con $j = i_0$ son nulos pues en ellos los $\Delta \mathbf{w}(i_0 - 1)$ son independientes y su valor esperado tiende a cero conforme n tiende a infinito.
3. Los términos con $j < i_0$ toman el siguiente valor:

$$\begin{aligned} & \mathbf{E}[\hat{\mathbf{E}}[\tilde{\mathbf{x}}(n-1)\mathbf{x}^t(n-1)\Delta \mathbf{w}(n-2)]\hat{\mathbf{E}}[\tilde{\mathbf{x}}^t(n-1)e_0(n-1)]] = \\ & = \left(\frac{1-f}{1-f^{n-1}} \right)^2 \sum_i \sum_{j < i} f^{n-1-i} f^{n-1-j} \mathbf{RE}[\Delta \mathbf{w}(i-1)\tilde{\mathbf{x}}^t(j)e_0(j)] \end{aligned} \quad (10.39)$$

Recuérdese que se ha mantenido este término por la dependencia que $\Delta \mathbf{w}(i-1)$ presenta en las variables aleatorias $\tilde{\mathbf{x}}(k)$ y $\mathbf{x}(k)$ para $k = 1 \dots i-1$. Pero también se comentó con anterioridad que la dependencia va decayendo conforme aumenta la separación temporal entre ellas, por tanto la expresión anterior se aproximará de la siguiente manera

$$\begin{aligned} & \mathbf{E}[\hat{\mathbf{E}}[\tilde{\mathbf{x}}(n-1)\mathbf{x}^t(n-1)\Delta \mathbf{w}(n-2)]\hat{\mathbf{E}}[\tilde{\mathbf{x}}^t(n-1)e_0(n-1)]] = \\ & = \left(\frac{1-f}{1-f^{n-1}} \right)^2 f \sum_{i=1}^{n-1} f^{2(n-1-i)} \mathbf{RE}[\Delta \mathbf{w}(i-1)\tilde{\mathbf{x}}^t(i-1)e_0(i-1)] \end{aligned} \quad (10.40)$$

esta sumatoria se puede reescribir como sigue

$$\begin{aligned} & \mathbf{E}[\hat{\mathbf{E}}[\tilde{\mathbf{x}}(n-1)\mathbf{x}^t(n-1)\Delta \mathbf{w}(n-2)]\hat{\mathbf{E}}[\tilde{\mathbf{x}}^t(n-1)e_0(n-1)]] = \\ & = \left(\frac{1-f}{1-f^{n-1}} \right)^2 \frac{1-f^{2(n-1)}}{1-f^2} f \mathbf{RE}[\hat{\mathbf{E}}[\Delta \mathbf{w}(n-2)\tilde{\mathbf{x}}^t(n-2)e_0(n-2)]|_{f^2}] \end{aligned} \quad (10.41)$$

donde se ha indicado explícitamente que este estimador del valor esperado utiliza como factor de olvido f^2 . Para n suficientemente grandes, condición bajo la cual se ha obtenido la expresión anterior, queda

$$\begin{aligned} & \mathbf{E}[\hat{\mathbf{E}}[\tilde{\mathbf{x}}(n-1)\mathbf{x}^t(n-1)\Delta \mathbf{w}(n-2)]\hat{\mathbf{E}}[\tilde{\mathbf{x}}^t(n-1)e_0(n-1)]] = \\ & = \frac{1-f}{1+f} f \mathbf{RE}[\hat{\mathbf{E}}[\Delta \mathbf{w}(n-2)\tilde{\mathbf{x}}^t(n-2)e_0(n-2)]|_{f^2}] \approx \\ & \approx \frac{1}{2}(1-f) f \mathbf{RE}[\hat{\mathbf{E}}[\Delta \mathbf{w}(n-2)\tilde{\mathbf{x}}^t(n-2)e_0(n-2)]|_{f^2}] \end{aligned} \quad (10.42)$$

En el último paso se ha considerado $f \sim 1$. Con este resultado se puede afirmar que el término $\boxed{2 \boxed{4}^t}$ es de orden $\mu^2 f^3(1-f)$, es decir, le ocurre algo parecido a lo

que le ocurre al término $\boxed{2 \mid 2}^t$ donde al anularse varios sumandos en el producto de dos estimadores de valores esperados aparece un factor global $(1 - f)$.

Por tanto este término también contribuye.

Término $\boxed{3 \mid 3}$:

Este término también contribuye, siendo su contribución,

$$\boxed{3 \mid 3}^t = \mu^2(1 - f)^2 \mathbb{E}[\mathbf{x}(n)\tilde{\mathbf{x}}^t(n)\tilde{\mathbf{x}}(n)\tilde{\mathbf{x}}^t(n)\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)e_0^2(n)]$$

Término $\boxed{3 \mid 4}$:

En este caso el estimador del valor esperado es independiente del resto por las hipótesis de independencia y su valor esperado es nulo por el principio de ortogonalidad.

Por tanto este término no contribuye.

Término $\boxed{4 \mid 4}$:

La parte más conflictiva de este término es

$$\begin{aligned} \mathbb{E}[\hat{\mathbb{E}}[\tilde{\mathbf{x}}(n-1)e_0(n-1)]\hat{\mathbb{E}}[\tilde{\mathbf{x}}^t(n-1)e_0(n-1)]] &= \\ &= \left(\frac{1-f}{1-f^{n-1}}\right)^2 \sum_{ij} f^{n-1-i}f^{n-1-j} \mathbb{E}[\tilde{\mathbf{x}}(i)\tilde{\mathbf{x}}^t(j)e_0(i)e_0(j)] = \\ &= \left(\frac{1-f}{1-f^{n-1}}\right)^2 \sum_i f^{2(n-1-i)} \mathbb{E}[\tilde{\mathbf{x}}(i)\tilde{\mathbf{x}}^t(i)e_0^2(i)] \end{aligned} \quad (10.43)$$

donde este valor esperado no depende del tiempo al ser las variables involucradas conjuntamente estacionarias⁶, y se le denominará $\mathbf{R}_{\tilde{\mathbf{x}}e}$. Suponiendo de nuevo n suficientemente grande y f cercano a la unidad, resulta

$$\begin{aligned} \mathbb{E}[\hat{\mathbb{E}}[\tilde{\mathbf{x}}(n-1)e_0(n-1)]\hat{\mathbb{E}}[\tilde{\mathbf{x}}^t(n-1)e_0(n-1)]] &= \\ &= \frac{1}{2}(1-f)\mathbf{R}_{\tilde{\mathbf{x}}e} \end{aligned} \quad (10.44)$$

⁶Esto no siempre es así ya que, por ejemplo, en problemas de identificación ciega de sistemas no estacionarios, estas variables no son conjuntamente estacionarias.

con lo que el término $\boxed{4 \mid 4}^t$ completo quedaría

$$\boxed{4 \mid 4}^t = \frac{1}{2} \mu^2 f^2 (1-f) \mathbb{E}[\mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \mathbf{R}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}} \tilde{\mathbf{x}}(n) \mathbf{x}^t(n)]$$

Este término, por tanto, también contribuye.

Suma de las contribuciones de todos los términos

Con todo lo comentado hasta aquí, la ecuación que rige la evolución de $\mathbf{R}_\Delta(n)$ quedaría:

$$\begin{aligned} \mathbf{R}_\Delta(n) = & \mathbf{R}_\Delta(n-1) - \mu(1-f)\Gamma \mathbf{R}_\Delta(n-1) - \mu(1-f)\mathbf{R}_\Delta(n-1)\Gamma + \\ & + \mu^2(1-f)^2 \mathbb{E}[\psi_1(n) \mathbf{R}_\Delta(n-1) \psi_1(n)] + \\ & + \mu^2 f^2 (1-f) \mathbb{E}[\mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \hat{\mathbb{E}}[\tilde{\mathbf{x}}(n-1) \mathbf{x}^t(n-1) \\ & \quad \mathbf{R}_\Delta(n-2) \mathbf{x}(n-1) \tilde{\mathbf{x}}^t(n-1)]|_{f^2} \tilde{\mathbf{x}}(n) \mathbf{x}^t(n)] - \\ & - \mu^2 f^3 \frac{1-f}{2} \mathbb{E}[\mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \mathbf{R} \hat{\mathbb{E}}[\Delta \mathbf{w}(n-2) \tilde{\mathbf{x}}^t(n-2) \\ & \quad e_0(n-2)]|_{f^2} \tilde{\mathbf{x}}(n) \mathbf{x}^t(n)] + \\ & + \mu^2(1-f)^2 \mathbb{E}[\psi_2(n) e_0^2(n)] + \\ & + \mu^2 f^2 \frac{1-f}{2} \mathbb{E}[\mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \mathbf{R}_{\tilde{\mathbf{x}}e} \tilde{\mathbf{x}}(n) \mathbf{x}^t(n)] \end{aligned} \quad (10.45)$$

Las principales diferencias con (10.35) son:

1. En la práctica se convierte en una ecuación en diferencias de orden n , aunque la influencia de $\mathbf{R}_\Delta(n-i)$ cae como f^{2i} . (En teoría, si el estimador se comportase como un verdadero valor esperado, es de orden 2).
2. Aparecen términos en $\hat{\mathbb{E}}[\Delta \mathbf{w}(n-2) \tilde{\mathbf{x}}^t(n-2) e_0(n-2)]$ los cuales desdibujan la expresión que ya no es una ecuación en diferencias. Los sumandos de este $\hat{\mathbb{E}}$ también van cayendo como f^{2i} .
3. Aparece un término independiente nuevo, de orden $(1-f)$ y por tanto más importante que el de (10.35).

Por tanto, en (10.35) obtenida con la segunda aproximación, se han eliminado todos los términos debidos al rastro que deja el $\hat{\mathbb{E}}[\]$ y ha conservado el primer sumando de $\hat{\mathbb{E}}[\]$. De todas maneras su propósito era ser una primera aproximación que diera convergencia para $\mathbf{R}_\Delta(n)$ después del primer intento fallido.

Comentario

Todo este análisis, más o menos minucioso, se debe a que es la primera vez que aparece en el estudio de un algoritmo adaptativo un factor como el estimador de un valor esperado. Debido a este motivo, se hace necesario encontrar algún tipo de aproximación que agilice los cálculos y que, a la vez, dé resultados con la suficiente información como para describir el comportamiento del algoritmo.

En capítulos anteriores, al analizar por ejemplo el GLMS, se pudo recurrir sin más a muchas técnicas utilizadas extensamente en la literatura para algoritmos de la misma familia (entiéndase el LMS) tales como las hipótesis de independencia o el promediado directo. En este capítulo hay que establecer las bases de las nuevas técnicas a utilizar en análisis de algoritmos, que como el AOGMLS, integren términos del estilo del $\hat{E}[\cdot]$.

10.7 Convergencia en media utilizando la segunda aproximación

Se parte de la ecuación de actualización del vector de error en los pesos (10.15), la cual establecía que

$$\begin{aligned} \Delta \mathbf{w}(n) = & \Delta \mathbf{w}(n-1) - \mu \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \hat{E}[\tilde{\mathbf{x}}(n) \mathbf{x}^t(n) \Delta \mathbf{w}(n-1)] + \\ & + \mu \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \hat{E}[\tilde{\mathbf{x}}(n) e_0(n)] \end{aligned} \quad (10.46)$$

que utilizando las aproximaciones propuestas en (10.24) y (10.25) conduce a

$$\begin{aligned} \Delta \mathbf{w}(n) = & \Delta \mathbf{w}(n-1) - \mu(1-\lambda) \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \tilde{\mathbf{x}}(n) \mathbf{x}^t(n) \Delta \mathbf{w}(n-1) + \\ & + \mu(1-f) \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \tilde{\mathbf{x}}(n) e_0(n) \end{aligned} \quad (10.47)$$

y tomando valor esperado se llega a la ecuación buscada:

$$\begin{aligned} E[\Delta \mathbf{w}(n)] = & [\mathbf{I} - \mu(1-f)\mathbf{\Gamma}] E[\Delta \mathbf{w}(n-1)] + \\ & + \mu(1-f) E[\mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \tilde{\mathbf{x}}(n) e_0(n)] \end{aligned} \quad (10.48)$$

En esta expresión se puede apreciar un poco mejor lo que significa hacer esta aproximación. Por un lado elimina la influencia de los vectores de error en los pesos a nuestras pasadas, que ya se ha comprobado que influye bastante en la convergencia del algoritmo en media como pone de manifiesto la figura 10.1. Sin embargo, parece acertada su sustitución en el segundo miembro, puesto que al estar evaluando un valor esperado (estimador) de unas variables conjuntamente estacionarias⁷, se puede afirmar que dicho estimador se acerca al valor verdadero

⁷Aquí está la clave. Si $\hat{E}[a]$ es independiente del resto y 'a' es estacionaria, entonces se puede afirmar que $\hat{E}[a] = E[a]$ con bastante aproximación.

conforme aumenta el número de datos. Sólo permanece el primer sumando de este estimador que es claramente dependiente de las variables aleatorias que acompaña, pero corregido por el factor $(1 - f)$.

A la vista de lo cual se puede concluir, teniendo en cuenta el comentario del párrafo anterior, que la expresión más exacta para la convergencia en media viene dada por:

$$\begin{aligned} E[\Delta \mathbf{w}(n)] &= [\mathbf{I} - \mu(1 - f)\mathbf{\Gamma}]E[\Delta \mathbf{w}(n - 1)] + \\ &+ \mu(1 - f)\mathbf{R}^t \mathbf{R} \sum_{i=1}^{n-1} f^{n-i} E[\Delta \mathbf{w}(i - 1)] + \mu(1 - f)E[\mathbf{x}(n)\tilde{\mathbf{x}}^t(n)\tilde{\mathbf{x}}(n)e_0(n)] \end{aligned} \quad (10.49)$$

que recoge las dos propiedades comentadas anteriormente:

- a) Incluye los vectores de error en los pesos para todos los tiempos que influyen en la ecuación.
- b) Aparecen ‘los restos’ del término independiente. Aunque no son deseables ya que producen un sesgo en los estimadores.

El sesgo en los estimadores viene dado por $E[\Delta \mathbf{w}(\infty)]$. Este valor se puede obtener de (10.49), como es habitual, suponiendo que se ha llegado a la convergencia y despejando:

$$E[\Delta \mathbf{w}(\infty)] = (1 - f) \frac{\mathbf{I}}{(1 - f)\mathbf{\Gamma} - \mathbf{R}^t \mathbf{R}} E[\mathbf{x}(n)\tilde{\mathbf{x}}^t(n)\tilde{\mathbf{x}}(n)e_0(n)]$$

Como se sabía el sesgo es un infinitésimo de orden $(1 - f)$ y no depende de μ .

Para que la solución converja efectivamente a esta solución estacionaria es necesario que las raíces del polinomio característico sean menores que 1. El problema es que aquí las componentes no se desacoplan en la base en la que $\mathbf{R}^t \mathbf{R}$ es diagonal. Para tener una idea del intervalo de valores μ , λ_{Γ} y $\lambda_{\mathbf{R}}$ en el cual se produce la convergencia se estudiará el caso de una única incógnita. Para esta situación la ecuación homogénea asociada toma la forma:

$$E[\Delta w(n)] = [1 - \mu(1 - f)\lambda_{\Gamma}]E[\Delta w(n - 1)] - \mu(1 - f) \sum_{i=1}^{n-1} \lambda_{\mathbf{R}} f^{n-i} E[\Delta w(n - i)]$$

El polinomio característico se construye suponiendo que la solución es de la forma $E[\Delta w(n)] = x^n$:

$$x^n = [1 - \mu(1 - f)\lambda_{\Gamma}]x^{n-1} - \mu(1 - f)\lambda_{\mathbf{R}}f \frac{x^{n-1} - f^{n-1}}{x - f}$$

Y dado que $f < 1$ para n suficientemente grandes se puede reescribir como:

$$x^{n-1} \left[x - 1 + \mu(1-f)\lambda_\Gamma + \mu \frac{(1-f)f\lambda_R}{x-f} \right] = 0$$

que aparte de la solución trivial $x = 0$ tiene como soluciones las raíces de:

$$x^2 + [\mu(1-f)\lambda_\Gamma - (1+f)]x + f[1 - \mu(1-f)(\lambda_\Gamma - \lambda_R)] = 0$$

donde haciendo $\lambda_\Gamma = \lambda_R$ se recupera la expresión dada en (10.19) como era de esperar.

Las soluciones de esta ecuación cuadrática son:

$$x = \frac{1+f - \mu(1-f)\lambda_\Gamma}{2} \pm \frac{\sqrt{(1+f - \mu(1-f)\lambda_\Gamma)^2 - 4f[1 - \mu(1-f)(\lambda_\Gamma - \lambda_R)]}}{2} \quad (10.50)$$

ahora, al introducir un nuevo parámetro en el análisis, x es función de $x = x(\mu\lambda_R, \mu\lambda_\Gamma, f)$ y por tanto no es posible realizar una representación gráfica como la de la figura 10.1. Sin embargo sí es posible discutir algunos aspectos:

- a) Cuando $f \rightarrow 1$ se sigue que $x \rightarrow 1$ y no hay convergencia.
- b) Cuando el factor de olvido tiende a cero, se cumple que $x = 1 - \mu\lambda_\Gamma$ (aparte de la solución trivial). Esta solución se encontrará en el análisis del CGLMS como se adelantó en la página 260 aunque esta de aquí sea la expresión correcta al aparecer λ_Γ y no λ_R , allí sólo se quiso presentar el resultado para poder ampliar el comentario.

10.8 Simulaciones

Una vez estudiado el algoritmo AONGLMS y comprobado teóricamente que converge, es el momento de verificarlo mediante simulaciones. A tal fin se han ideado tres conjuntos de simulaciones: en el primero se intenta corroborar que efectivamente el AONGLMS cumple el principal requisito que se le exigió cuando se diseñó, es decir, que funcione cuando la matriz asociada no sea definida; en el segundo se presenta una idea del comportamiento del algoritmo en la identificación de modelos de diversos órdenes y con distintos niveles de ruido; finalmente se detalla una aplicación de AONGLMS en la ecualización ciega de una constelación 4-Q.A.M transmitida por un modelo AR. Es conveniente destacar en este punto que no se pretende hacer un estudio mediante simulaciones pormenorizado, como se hizo para el algoritmo ORIV, sino simplemente mostrar por encima las propiedades del nuevo algoritmo. El principal trabajo en esta parte de la Memoria ha sido su deducción teórica.

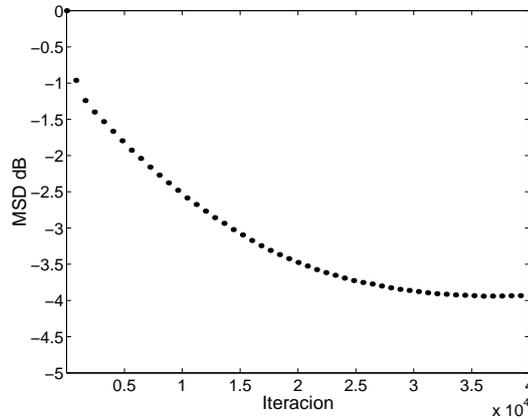


Figura 10.2 *Convergencia de AOGMLS con una matriz no definida*

10.8.1 Comportamiento de AOGMLS cuando la matriz asociada no es definida

Se considera un modelo AR de orden 2 con coeficientes complejos $[1 - 1 + 0.9i \ 0.04 - 0.44i]$ que se pretende identificar mediante las ecuaciones normales de orden 4, tal y como se describieron en el apartado 8.10. El proceso de excitación es un 4-Q.A.M formado por variables i.i.d con distribución de probabilidad constante y con posibles valores $\{1 + 2i, 1 - 2i, -1 + 2i, -1 - 2i\}$. El algoritmo AOGMLS se implementará con el mismo número de ecuaciones que de incógnitas, por lo que la matriz asociada a este problema concreto tiene como autovalores $\lambda_1 = -105 - 99i$ y $\lambda_2 = 8.75 - 6.2i$ y por tanto su parte real no es definida. Contará con series de 40000 datos cada una y se llevarán a cabo 200 realizaciones. Los parámetros libres del algoritmo se han fijado como $\mu = 8 \cdot 10^{-10}$ y $f = 0.99$.

Según el análisis teórico llevado a cabo, el algoritmo GLMS no es capaz de converger en estas circunstancias. El comportamiento del AOGMLS se muestra en la figura 10.2, en ella se puede apreciar que la velocidad de convergencia es muy lenta pero que de hecho converge. Sin embargo, tal y como se esperaba el GLMS no puede converger. Para mostrar esto último se han elegido dos factores de olvido muy pequeños de distinto signo, como se observa en las gráficas 10.3(a) y 10.3(b) diverge en ambos casos. Al ser el autovalor de mayor peso (el de mayor valor absoluto) el de signo negativo, para un tamaño del paso negativo la componente asociada a este autovalor empieza a converger, sin embargo, en cuanto las iteraciones van aumentando, la componente asociada al autovalor positivo diverge claramente con lo que el resultado neto es una divergencia

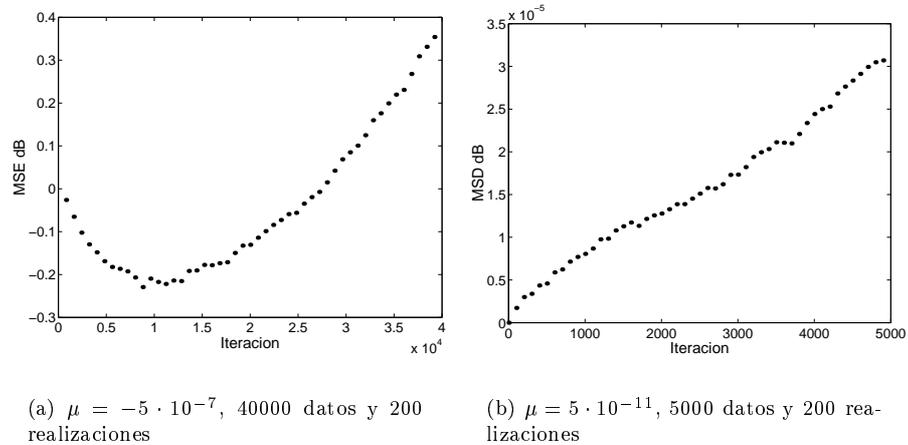


Figura 10.3 Constatación de la no convergencia de GLMS con \mathbf{R} no definida

general. Para este caso se han tenido que emplear un mayor número de datos y un tamaño del paso ligeramente mayor para poder apreciarlo claramente; si se hubiesen empleado sólo 5000 datos con $\mu = 5 \cdot 10^{-11}$ se habría observado una convergencia (causada por la componente asociada al mayor autovalor) muy lenta, ya que la fase divergente empezaría mucho más allá de los 5000 datos.

Una vez constatada la convergencia del AGLMS se puede hacer que converja a valores más bajos siempre y cuando se disminuya el tamaño del paso. Si esto se llevara a cabo el número de iteraciones necesarias sería desorbitante.

El gran número de iteraciones necesarias es debido a la dispersión de autovalores de la matriz $\mathbf{R}^t \mathbf{R}$. Como se verá más adelante en situaciones no patológicas el número de iteraciones necesarias disminuye hasta valores más razonables.

10.8.2 Identificación ciega de modelos MA en ambientes ruidosos mediante AGLMS

El segundo grupo de simulaciones, que se van a llevar a cabo para dar una idea del comportamiento del algoritmo AGLMS, consiste en identificar sistemas MA mediante la ecuación de Giannankis-Mendel. Este mismo problema se trató en el análisis del algoritmo ORIV y aquí se volverán a emplear los mismos modelos. En la tabla 10.2 se presentan los coeficientes de los modelos empleados y la dispersión de la matriz asociada, teniendo en cuenta los siguientes puntos:

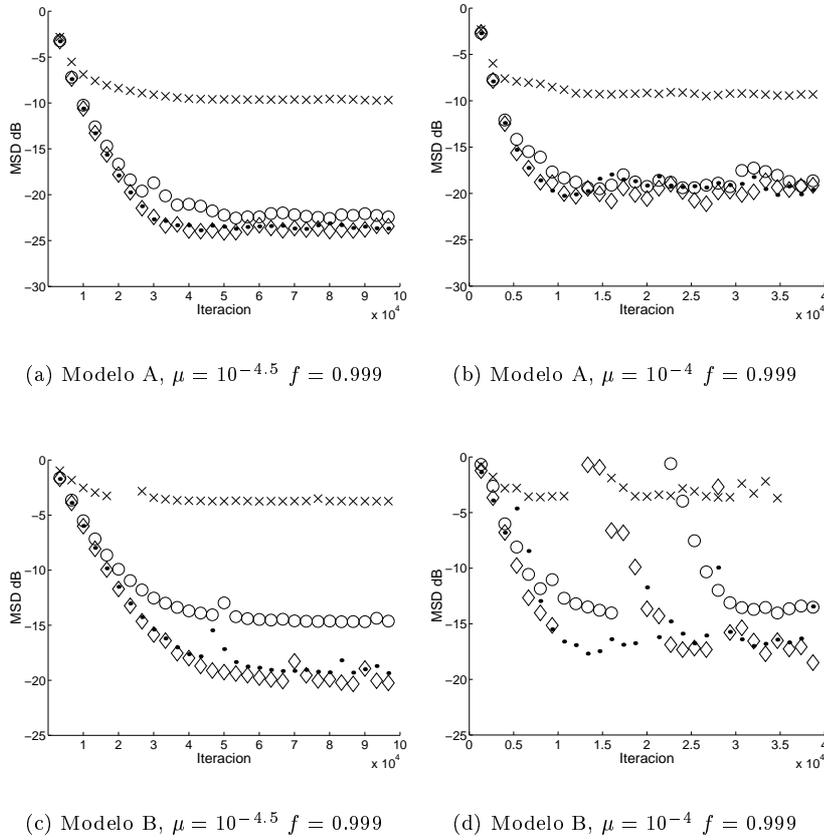


Figura 10.4 Evolución del MSD para los modelos A y B

1. Se va a implementar la modalidad de AOG LMS sobredeterminado, de forma que se puedan emplear las $3q+1$ ecuaciones, que propone la ecuación GM, para identificar las $2q+1$ incógnitas⁸ del problema.
2. Para resolver el problema basta definir las variables $\tilde{\mathbf{x}}(n)$, $\mathbf{x}(n)$ y $d(n)$ que aparecen en la tabla 10.1 tal y como se definieron para el algoritmo ORIV en el apartado 5.3.
3. De esta manera el comportamiento del algoritmo viene determinado por la dispersión de autovalores de la matriz $\mathbf{R}^t \mathbf{R}$ donde $\mathbf{R} = E[\tilde{\mathbf{x}}(n) \mathbf{x}^t(n)]$.

⁸Véase el apartado 5.2.

Modelo	Coefficientes	Dispersión
A	[1 - 0.8]	3.2211
B	[1 - 1.25]	13.3117
C	[1 - 1.1314 0.6400]	41.3750
D	[1 - 2.33 0.6400]	3048
E	[1 - 0.9 0.385 - 0.771]	279

Tabla 10.2 Modelos MA identificados mediante AOGLMS

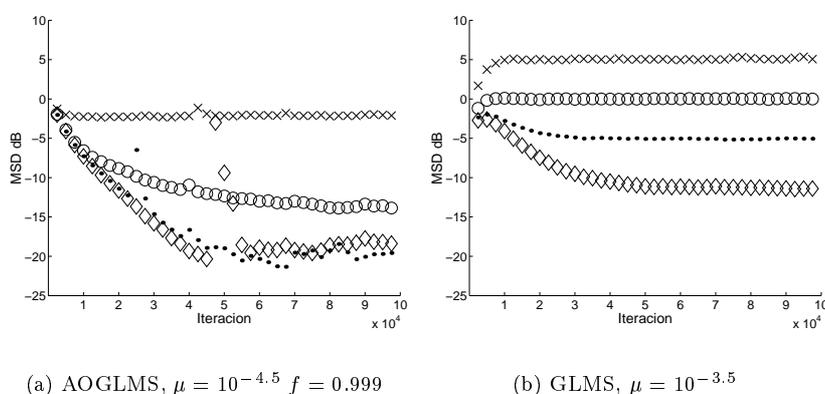


Figura 10.5 Evolución del MSD del modelo C

Los modelos MA se excitan mediante procesos i.i.d formados por variables con una función densidad de probabilidad exponencial por un lado, de media 0, varianza 1, factor de asimetría 2 y curtosis 6. Se dispone de series de 40000 y 100000 muestras ruidosas y los resultados se promedian sobre 200 realizaciones. El ruido añadido se ha generado filtrando ruido blanco gaussiano a través de un filtro ARMA con coeficientes MA = [1 - 1.25] y AR = [1 - 2.2 1.77 - 0.52] y los niveles de ruido estudiados han sido 20, 15, 10, 5 y 0 dB (aunque para 0 dB no converge nunca).

Los resultados que se muestran a continuación se han elegido como los mejores dentro de un proceso de búsqueda: para cada modelo y SNR se hacía variar el tamaño del paso según la función $\mu = 10^{ex}$ donde $ex = -6, -5.5, -5, \dots, -3.5$ y el factor de olvido f tomaba los valores $f = 0.9, 0.99$ y 0.999 . En la figura 10.4(a) se muestra la evolución del MSD para el modelo A y los niveles de ruido indicados, cuando $\mu = 10^{-4.5}$ y $f = 0.999$. La lectura de la gráfica

debe hacerse atendiendo a las siguientes asociaciones:

- ◇...20 dB
- ...15 dB
- o...10 dB
- × ...5 dB

Análogamente, en la figura 10.4(b) se representan los resultados para el modelo A pero con $\mu = 10^{-4}$ y $f = 0.999$. En ambas figuras (nótese la diferencia de escalas) se pueden apreciar aspectos comunes a los algoritmos tipo máxima pendiente y a aquellos que trabajan con estadística de alto orden, como por ejemplo que cuanto menor es el tamaño del paso tienden a un valor de $D(\infty)$ menor pero tardan más en converger; a su vez son menos sensibles a la SNR que otros algoritmos del mismo tipo pero basados en estadística de segundo orden. De todas maneras el principal objetivo de estas gráficas es mostrar que AOG LMS de hecho converge para estos modelos, aspecto importantísimo puesto que GLMS no es capaz de identificar estos modelos como se discutirá más adelante.

La evolución del MSD para el modelo B y para los mismos pares de valores de los parámetros μ y f que para el modelo A se presentan en las figuras 10.4(c) y 10.4(d). Al ser su dispersión de autovalores mayor, el comportamiento observado es peor. En particular para $\mu = 10^{-4}$ es más inestable.

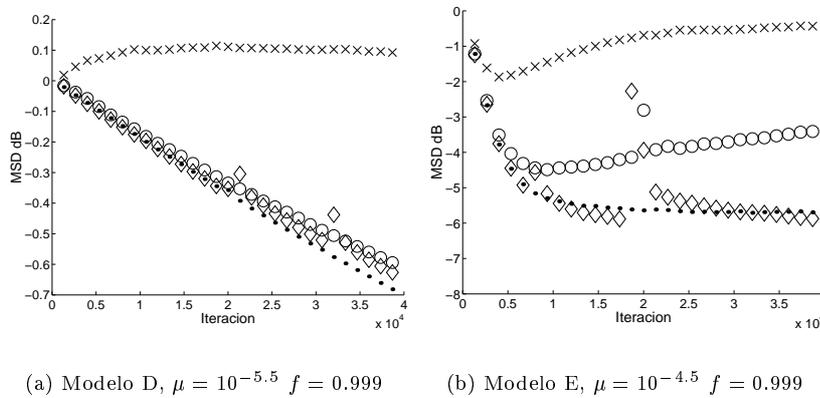


Figura 10.6 Evolución del MSD para los modelos D y E

Para el modelo C las simulaciones dan resultados aceptables para $\mu = 10^{-4.5}$ y $f = 0.999$ y se detallan en la figura 10.5(a). Tal y como ya se sabía del análisis de ORIV y de los comentarios precedentes, al tener el modelo D una

dispersión tan elevada los algoritmos tienen grandes dificultades en alcanzar la convergencia, la evolución para $\mu = 10^{-5.5}$ y $f = 0.999$ se presenta en la figura 10.6(a); como se observa el tiempo de convergencia es enorme. Finalmente se presentan los resultados para el modelo E con $\mu = 10^{-4.5}$ y $f = 0.999$ en la figura 10.6(b), y dado que su dispersión no es muy elevada da resultados con una desviación en media cuadrática aceptable, téngase en cuenta que se están estimando 7 parámetros.

Se quiere comparar el comportamiento descrito del AOGLEMS con el de GLMS en la misma situación, en la identificación ciega de sistemas MA mediante la ecuación de GM. Lo primero que hay que hacer notar es que GLMS no puede tratar con sistemas sobredeterminados, por lo que la longitud de la variable instrumental hay que acortarla hasta igualar la del vector de datos $\mathbf{x}(n)$, así se consigue tener el mismo número de ecuaciones que de incógnitas. Por otro lado, la convergencia del algoritmo viene determinada por los autovalores de la matriz $\mathbf{R} = E[\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)]$, cuadrada por tanto en este caso. La dispersión de la parte real de los autovalores para cada modelo se detalla a continuación:

modelo A	...	-0.8611
modelo B	...	-1.0213
modelo C	...	20.9897
modelo D	...	29.6193
modelo E	...	-2.4734

Para aquellos modelos cuya dispersión sea negativa no es posible la convergencia mediante GLMS, ya que indican la existencia de autovalores positivos y negativos, por lo que \mathbf{R} no es definida. Si los autovalores son complejos, para que haya convergencia es necesario que sus partes reales tengan el mismo signo: los modelos C y D cumplen este requisito. Sin embargo como el modelo D tiene un autovalor de valor $0.2652 + 0.4062i$ y su parte real es muy pequeña, la convergencia se hace muy difícil, ya que al trabajar con estimadores de la matriz \mathbf{R} , puede que a veces este autovalor presente un estimador con la parte real negativa que impediría la convergencia. La evolución del MSD para el modelo C se presenta en la figura 10.5(b) junto con la evolución de la MSD del mismo modelo obtenida mediante el algoritmo AOGLEMS (figura 10.5(a)). No se busca una comparación rigurosa, sino una mera ilustración; ambos algoritmos tardan unas 40000 iteraciones en alcanzar el estado estacionario para una SNR=20 dB, sin embargo, para este caso AOGLEMS mejora en 10 dB el MSD(∞) de GLMS, además de mejorar por varios dB para otras SNRs.

10.8.3 Aplicando el algoritmo AOGLMS a ecualización ciega

El último grupo de simulaciones presenta una aplicación del AOGLMS a la ecualización de un proceso 4-Q.A.M que ha sido transmitido (filtrado) por un modelo AR= $[1 - 0.4 + 0.4i]$. La ecualización se llevará a cabo en dos fases, en primer lugar se identifica el modelo AR siguiendo el mismo método que en las simulaciones del apartado 10.8.1, y en segundo lugar se filtra la señal de salida por un filtro MA, de coeficientes los estimados para el AR en el primer paso. Se supone que el ambiente no es ruidoso.

La longitud de la señal es de 10000 muestras y tras promediar 200 realizaciones el estimador del coeficiente del modelo AR es $-0.3961 + 0.3953i$ con una desviación estandar de 0.0349, habiendo fijado $\mu = 5\Delta 10^{-7}$ y $f = 0.99$. Para ilustrar gráficamente el problema en la figura 10.7(a) se presentan 50 muestras de la salida del canal para una realización arbitraria y en la figura 10.7(b) la misma salida ecualizada; como se puede apreciar se recupera claramente la estructura de la constelación 4-Q.A.M aunque con un pequeño sesgo.

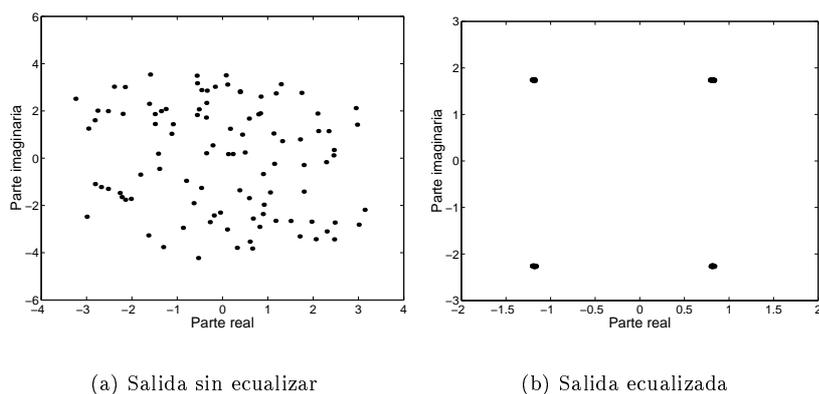


Figura 10.7 *Ecualización ciega mediante AOGLMS*

10.8.4 Conclusión

Tal y como se ha podido constatar mediante simulaciones, el algoritmo AOGLMS tiene un comportamiento mejor al de GLMS con la ventaja de que es capaz de funcionar en situaciones donde el segundo no puede. En definitiva, se

ha establecido la superioridad del nuevo algoritmo con respecto al ya existente, a falta de llevar a cabo un análisis más riguroso.

Capítulo 11

Algunas versiones directas del algoritmo GLMS

Introducción

El principal objetivo del último capítulo de esta parte de la Memoria es presentar tres nuevas versiones del algoritmo GLMS. Estas versiones, en cierta manera, completan el cuadro formal que se ha ido creando desde el comienzo de la Memoria con el algoritmo ORIV.

Sólo se pretende motivar un poco la creación de estos algoritmos y añadir un mínimo análisis que certifique el correcto funcionamiento de los mismos tanto en media como en media cuadrática. La convergencia se corroborará mediante simulaciones.

El primer algoritmo que se propone intenta, mediante la correcta elección de la función coste, solventar el problema del GLMS de no converger con matrices no definidas. Aparece sin embargo un pequeño sesgo en los estimadores obtenidos con este nuevo algoritmo. Para evitar este problema se propone un segundo algoritmo completamente ad-hoc. El tercer algoritmo se sale un poco de la línea anterior ya que lo que constituye es una versión del GLMS normalizado.

El algoritmo CGLMS se presenta en el apartado 11.1 donde además se incluye un análisis de su convergencia en media y en media cuadrática. Análogo tratamiento reciben los algoritmos A1GLMS y PNLMS en los apartados 11.2 y 11.3 respectivamente. Para confirmar su convergencia en el apartado 11.4 se muestran los resultados de algunas simulaciones. Para concluir, en el apartado 11.5, se presenta un cuadro resumen con los algoritmos de las partes primera y

segunda de la Memoria.

11.1 El algoritmo GLMS cuadrático, CGLMS

En la mayoría de los problemas de estimación lineal se suele buscar como solución aquella que minimice el error cuadrático medio $MSE = E[e^2(n)]$. Sin embargo en el problema central de esta Memoria lo que se pretende es imponer que $E[\tilde{\mathbf{x}}(n)e(n)] = 0$, por lo que una posible manera de conseguirlo es minimizando la siguiente cantidad $\|E[\tilde{\mathbf{x}}(n)e(n)]\|^2$.

No se debe perder de vista que lo que se busca es un algoritmo del tipo gradiente estocástico y, como tal, prescindirá de cualquier tipo de operador ‘valor esperado’ en su formulación. Por tanto, en vez de minimizar el criterio anterior, se minimizará la siguiente función de coste estocástica:

$$\hat{f}(n) = \|\tilde{\mathbf{x}}(n)e(n)\|^2$$

Dado que $e(n) = d(n) - \mathbf{x}^t(n)\mathbf{w}(n-1)$ se cumple que

$$\hat{\nabla}_i(n) = \frac{\partial \hat{f}(n)}{\partial w_i(n-1)} = -2x_i(n)\tilde{\mathbf{x}}^t(n)\tilde{\mathbf{x}}(n)e(n)$$

Recordando además la estructura general de los algoritmos de gradiente estocástico, se propone como ecuación central del nuevo algoritmo la siguiente:

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \mu \mathbf{x}(n)\tilde{\mathbf{x}}^t(n)\tilde{\mathbf{x}}(n)e(n) \quad (11.1)$$

Esta ecuación es la base del algoritmo GLMS cuadrático o CGLMS, descrito en su totalidad en la tabla 11.1.

<p>Inicialización: $\mathbf{w}(0) = \mathbf{0}$</p> <p>Con la llegada de un nuevo dato: construir $\mathbf{x}(n)$, $\tilde{\mathbf{x}}(n)$ y $d(n)$ $e(n) = d(n) - \mathbf{x}^t(n)\mathbf{w}(n-1)$ $\mathbf{w}(n) = \mathbf{w}(n-1) + \mu \mathbf{x}(n)\tilde{\mathbf{x}}^t(n)\tilde{\mathbf{x}}(n)e(n)$</p>
--

Tabla 11.1 *Proceso iterativo del algoritmo CGLMS*

Como se puede comprobar la única diferencia con el algoritmo GLMS se encuentra en el término del gradiente que en este caso está multiplicado por la matriz $\mathbf{x}(n)\tilde{\mathbf{x}}^t(n)$. La ventaja de incluir esta matriz es que de esta manera el algoritmo es capaz de resolver problemas sobredeterminados, aunque la verdadera motivación de elegir esta función coste (y por tanto incluir esta matriz extra) es conseguir que el algoritmo converja aunque la matriz \mathbf{R} asociada no sea definida. Esto se ha buscado haciendo que la función $\hat{f}(n)$ sea cuadrática de forma que el gradiente siempre apunta a un mínimo estable. Otra cuestión a discutir es comprobar que ese mínimo de verdad es la solución buscada.

11.1.1 Convergencia en media

De nuevo se supondrá que $d(n) = \mathbf{w}_0^t \mathbf{x}(n) + e_0(n)$. Recuerdese que la solución \mathbf{w}_0 es tal que

$$E[\tilde{\mathbf{x}}(n)e_0(n)] = 0 \quad (11.2)$$

Partiendo de la ecuación (11.1) se puede escribir que

$$\begin{aligned} \Delta \mathbf{w}(n) = & [\mathbf{I} - \mu \mathbf{x}(n)\tilde{\mathbf{x}}^t(n)\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)]\Delta \mathbf{w}(n-1) + \\ & + \mu \mathbf{x}(n)\tilde{\mathbf{x}}^t(n)\tilde{\mathbf{x}}(n)e_0(n) \end{aligned} \quad (11.3)$$

Por ahora sólo se asumirá la siguiente hipótesis de independencia:

HI1 El vector $\mathbf{w}(n)$ sólo depende de las variables $\mathbf{x}(i)$, $\tilde{\mathbf{x}}(i)$ y $d(i)$ hasta tiempo n , es decir, $i = 1 \dots n$.

Con esto resulta que:

$$E[\Delta \mathbf{w}(n)] = [\mathbf{I} - \mu \mathbf{\Gamma}]E[\Delta \mathbf{w}(n-1)] + \mu E[\mathbf{x}(n)\tilde{\mathbf{x}}^t(n)\tilde{\mathbf{x}}(n)e_0(n)]$$

Donde $\mathbf{\Gamma} = E[\boldsymbol{\psi}_1(n)]$ siendo $\boldsymbol{\psi}_1(n) = \mathbf{x}(n)\tilde{\mathbf{x}}^t(n)\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)$. Como se puede apreciar, esta ecuación no arroja convergencia en media del algoritmo debido a la permanencia del término independiente. Sólo con la ecuación (11.2) no es suficiente, se hace necesario introducir alguna condición o hipótesis más. Esta hipótesis consiste en suponer que:

HI2 La variable $e_0(n)$ es independiente del resto de procesos aleatorios y es de media nula.

Es una hipótesis estrictamente cierta en algunos problemas, como identificación ciega de modelos AR y contiene a la condición (11.2) como caso particular.

Con esta nueva hipótesis se obtiene el resultado deseado:

$$\mathbf{E}[\Delta \mathbf{w}(n)] = [\mathbf{I} - \mu \mathbf{\Gamma}] \mathbf{E}[\Delta \mathbf{w}(n-1)]$$

Trabajando en la base en la que $\mathbf{\Gamma}$ es diagonal,

$$\mathbf{E}[\Delta w_i(n)] = [1 - \mu \lambda_{\Gamma_i}] \mathbf{E}[\Delta w_i(n-1)]$$

esta recursión tenderá a cero siempre y cuando:

$$|1 - \mu \lambda_{\Gamma_i}| < 1 \quad \Rightarrow \quad 0 < \mu < \frac{2}{\lambda_{\Gamma_i}} \quad (11.4)$$

Como se puede apreciar la matriz $\mathbf{\Gamma}$ es definida positiva por lo que $\lambda_{\Gamma_i} > 0$ y por tanto se puede encontrar un valor de μ que verifique (11.4) $\forall \lambda_{\Gamma_i}$. No le ocurre lo mismo que al algoritmo GLMS; sin embargo, $\mathbf{\Gamma}$ se ve afectada por ruido aditivo gaussiano y \mathbf{R} no.

11.1.2 Convergencia en media cuadrática

A partir de la expresión (11.3) se puede construir el producto externo,

$$\begin{aligned} \Delta \mathbf{w}(n) \Delta \mathbf{w}^t(n) &= [\mathbf{I} - \mu \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \tilde{\mathbf{x}}(n) \mathbf{x}^t(n)] \Delta \mathbf{w}(n-1) \\ &\quad \Delta \mathbf{w}^t(n-1) [\mathbf{I} - \mu \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \tilde{\mathbf{x}}(n) \mathbf{x}^t(n)] + \\ &\quad + \mu^2 \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \tilde{\mathbf{x}}(n) \tilde{\mathbf{x}}^t(n) \tilde{\mathbf{x}}(n) \mathbf{x}^t(n) e_0^2(n) + \\ &\quad + 2[\mathbf{I} - \mu \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \tilde{\mathbf{x}}(n) \mathbf{x}(n)] \Delta \mathbf{w}(n-1) \tilde{\mathbf{x}}^t(n) \tilde{\mathbf{x}}(n) \mathbf{x}^t(n) e_0(n) \end{aligned} \quad (11.5)$$

Debido a H11 y a la convergencia en media del algoritmo, el 3^{er} sumando del miembro de la derecha tiende a 0 una vez tomado valor esperado. El resto de sumandos conduce a:

$$\begin{aligned} \mathbf{R}_{\Delta}(n) &= \mathbf{R}_{\Delta}(n-1) - \mu \mathbf{\Gamma} \mathbf{R}_{\Delta}(n-1) - \mu \mathbf{R}_{\Delta}(n-1) \mathbf{\Gamma} + \\ &\quad + \mu^2 \mathbf{E}[\psi_1(n) \mathbf{R}_{\Delta}(n-1) \psi_1(n)] + \mu^2 \sigma_0^2 \chi \end{aligned} \quad (11.6)$$

donde $\chi = \mathbf{E}[\mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \tilde{\mathbf{x}}(n) \tilde{\mathbf{x}}^t(n) \tilde{\mathbf{x}}(n) \mathbf{x}^t(n)]$.

Para proseguir con el análisis se tomará $\mu \sim 0$ y se despreciará el coeficiente de $\mathbf{R}_{\Delta}(n-1)$ en μ^2 . Bajo esta condición y trabajando en la base en la que $\mathbf{\Gamma}$ es diagonal:

$$D_i(n) = (1 - 2\mu \lambda_{\Gamma_i}) D_i(n-1) + \mu^2 \sigma_0^2 \chi_{ii}$$

siendo $D_i(n) = R_{\Delta ii}(n)$.

Como siempre, a partir de esta expresión se obtendrá información acerca del tiempo de convergencia y del valor en convergencia. Para el tiempo de convergencia resulta:

$$\frac{D_i(\tau_i)}{D_i(0)} = \frac{1}{e} \Rightarrow \tau_i = \frac{-1}{\ln(1 - 2\mu\lambda_{\Gamma_i})} \underset{\mu \sim 0}{\approx} \frac{1}{2\mu\lambda_{\Gamma_i}}$$

Si éste es el tiempo de convergencia para la componente i -ésima, se puede definir uno total como promedio de ellos:

$$\tau_i = \frac{1}{q} \sum_i \tau_i = \frac{1}{2\mu q} \text{Tr}\{\mathbf{\Gamma}^{-1}\}$$

donde q es el número de incógnitas.

En lo que respecta al valor en convergencia:

$$D_i(\infty) = \frac{\mu}{2} \sigma_0^2 \frac{\chi_{ii}}{\lambda_{\Gamma_i}}$$

y dado que $\text{MSD} = \text{Tr}\{\mathbf{R}_\Delta\}$ resulta:

$$\text{MSD}(\infty) = \frac{\mu}{2} \sigma_0^2 \text{Tr}\{\mathbf{\Gamma}^{-1} \boldsymbol{\chi}\}$$

11.2 El algoritmo A1GLMS

Este algoritmo se propone, completamente ad-hoc partiendo de CGLMS, para eliminar el sesgo que CGLMS presenta bajo HI1. La solución se basa en aceptar otra hipótesis de independencia ya conocida y menos restrictiva que HI2:

HI3 Los vectores $\mathbf{x}(n)$ y $\tilde{\mathbf{x}}(n-i)$ son independientes para $i \neq 0$, además de ser i.i.d.

Bajo esta nueva hipótesis se propone como ecuación de actualización de los pesos:

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \mu \mathbf{x}(n+1) \tilde{\mathbf{x}}^t(n+1) \tilde{\mathbf{x}}(n) e(n) \quad (11.7)$$

nótese que la diferencia estriba en que la matriz $\mathbf{x}(n+1) \tilde{\mathbf{x}}^t(n+1)$ está definida a tiempo $n+1$ y no a tiempo n . La influencia de esta modificación quedará clara en el análisis subsiguiente. El algoritmo que se construye a partir de esta ecuación recibe el nombre de 'CGLMS aumentando en 1 el instante temporal' o A1GLMS. Los pasos requeridos por el algoritmo A1GLMS se detallan en la tabla 11.2.

<p>Inicialización: $\mathbf{w}(0) = \mathbf{0}$</p> <p>Con la llegada de un nuevo dato: construir $\mathbf{x}(n+1)$, $\tilde{\mathbf{x}}(n+1)$ y $d(n+1)$ $e(n) = d(n) - \mathbf{x}^t(n)\mathbf{w}(n-1)$ $\mathbf{w}(n) = \mathbf{w}(n-1) + \mu\mathbf{x}(n+1)\tilde{\mathbf{x}}^t(n+1)\tilde{\mathbf{x}}(n)e(n)$</p>

Tabla 11.2 *Proceso iterativo del algoritmo A1GLMS*

11.2.1 Convergencia en media

Siguiendo el mismo procedimiento empleado muchas veces ya en esta Memoria, se llega, partiendo de la ecuación (11.7), a la siguiente expresión para el vector de error en los pesos:

$$\begin{aligned} \Delta\mathbf{w}(n) = & [\mathbf{I} - \mu\mathbf{x}(n+1)\tilde{\mathbf{x}}^t(n+1)\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)]\Delta\mathbf{w}(n-1) + \\ & + \mu\mathbf{x}(n+1)\tilde{\mathbf{x}}^t(n+1)\tilde{\mathbf{x}}(n)e_0(n) \end{aligned} \quad (11.8)$$

Tomando valor esperado, teniendo en cuenta HI1 y HI3, se obtiene,

$$\mathbf{E}[\Delta\mathbf{w}(n)] = [\mathbf{I} - \mu\mathbf{R}^t\mathbf{R}]\mathbf{E}[\Delta\mathbf{w}(n-1)] + \mu\mathbf{R}^t\mathbf{E}[\tilde{\mathbf{x}}(n)e_0(n)]$$

Pero por el principio de ortogonalidad el segundo sumando del miembro de la derecha es nulo, con lo cual se puede escribir que

$$\mathbf{E}[\Delta\mathbf{w}(n)] = [\mathbf{I} - \mu\mathbf{R}^t\mathbf{R}]\mathbf{E}[\Delta\mathbf{w}(n-1)]$$

Como siempre, para estudiar esta ecuación, si se trabaja en la base en la que $\mathbf{R}^t\mathbf{R}$ es diagonal, se deduce que converge a cero la solución si

$$0 < \mu < \frac{2}{\lambda_i} \quad \forall \lambda_i$$

Es decir, en esta situación los estimadores serán no sesgados.

11.2.2 Análisis de la convergencia en media cuadrática

Para seguir los pasos habituales en primer lugar se construye el producto

$$\begin{aligned}
\Delta \mathbf{w}(n) \Delta \mathbf{w}^t(n) &= [\mathbf{I} - \mu \mathbf{x}(n+1) \tilde{\mathbf{x}}^t(n+1) \tilde{\mathbf{x}}(n) \mathbf{x}^t(n)] \Delta \mathbf{w}(n-1) \\
&\quad \Delta \mathbf{w}^t(n-1) [\mathbf{I} - \mu \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \tilde{\mathbf{x}}(n+1) \mathbf{x}^t(n+1)] + \\
&\quad + \mu [\mathbf{I} - \mu \mathbf{x}(n+1) \tilde{\mathbf{x}}^t(n+1) \tilde{\mathbf{x}}(n) \mathbf{x}^t(n)] \Delta \mathbf{w}(n-1) \\
&\quad \tilde{\mathbf{x}}^t(n) \tilde{\mathbf{x}}(n+1) \mathbf{x}^t(n+1) e_0(n) + \\
&\quad + ()^t + \\
&\quad + \mu^2 \mathbf{x}(n+1) \tilde{\mathbf{x}}^t(n+1) \tilde{\mathbf{x}}(n) \tilde{\mathbf{x}}^t(n) \tilde{\mathbf{x}}(n+1) \mathbf{x}^t(n+1) e_0^2(n)
\end{aligned} \tag{11.9}$$

Considerando que se trabaja en condiciones bajo las cuales se produce convergencia en media, asumiendo que $n \rightarrow \infty$ y bajo H11 y H13 el segundo y el tercer sumando de la expresión se anulan una vez tomado valor esperado. El resto de términos dan:

$$\begin{aligned}
\mathbf{R}_\Delta(n) &= \mathbf{R}_\Delta(n-1) - \mu \mathbf{R}^t \mathbf{R} \mathbf{R}_\Delta(n-1) - \mu \mathbf{R}_\Delta(n-1) \mathbf{R}^t \mathbf{R} + \\
&\quad + \mu^2 \mathbb{E}[\mathbf{x}(n+1) \tilde{\mathbf{x}}^t(n+1) \tilde{\mathbf{x}}(n) \mathbf{x}^t(n) \mathbf{R}_\Delta(n-1) \\
&\quad \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \tilde{\mathbf{x}}(n+1) \mathbf{x}^t(n+1)] + \\
&\quad + \mu^2 \mathbf{N}
\end{aligned} \tag{11.10}$$

donde $\mathbf{N} = \mathbb{E}[\mathbf{x}(n+1) \tilde{\mathbf{x}}^t(n+1) \mathbb{E}[\tilde{\mathbf{x}}(n) \tilde{\mathbf{x}}^t(n) e_0^2(n)] \tilde{\mathbf{x}}(n+1) \mathbf{x}^t(n+1)]$. Se obtendrá información de esta ecuación trabajando en la base en la que la matriz $\mathbf{R}^t \mathbf{R}$ es diagonal. Además se despreciará el coeficiente en μ^2 de $\mathbf{R}_\Delta(n-1)$:

$$D_i(n) = (1 - 2\mu\lambda_i) D_i(n-1) + \mu^2 N_{ii} \tag{11.11}$$

Análogamente al algoritmo CGLMS, de (11.11) se obtiene un tiempo de convergencia para el MSD dado por:

$$\tau = \frac{1}{\mu} \frac{1}{2q} \text{Tr}\{(\mathbf{R}^t \mathbf{R})^{-1}\}$$

y un valor para el error cuadrático medio en convergencia de:

$$\text{MSD}(\infty) = \mu \frac{1}{2} \text{Tr}\{(\mathbf{R}^t \mathbf{R})^{-1} \mathbf{N}\}$$

11.3 El algoritmo PNLMS

Este algoritmo se basa en el GLMS estudiado con anterioridad. Por conveniencia se retomarán algunas de sus expresiones. La ecuación de actualización de los pesos del filtro venía dada por:

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \mu \tilde{\mathbf{x}}(n) e^*(n) \tag{11.12}$$

donde $e(n) = d(n) - \mathbf{w}^H(n-1)\mathbf{x}(n)$ se denomina error a priori, $\mathbf{w}(n)$ es el estimador de \mathbf{w}_0 a tiempo n y el factor μ (real) recibe el nombre de tamaño del paso (step size).

Se puede definir un nuevo error de la siguiente manera $e_p(n) = d(n) - \mathbf{w}^H(n)\mathbf{x}(n)$, denominado error a posteriori, ya que utiliza el estimador de \mathbf{w}_0 a tiempo n , en vez de a tiempo $n-1$. Es de esperar por tanto, que lleve más información que el error a priori, ya que cuenta con los datos obtenidos a tiempo n . La ecuación de actualización (11.12) se verá beneficiada por el uso de este error y, en definitiva, se propone esta nueva expresión:

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \mu \tilde{\mathbf{x}}(n) e_p^*(n) \quad (11.13)$$

Para que esta ecuación sea útil, sería conveniente encontrar una expresión que permita calcular el error a posteriori sin utilizar (ya que no se dispone de él) $\mathbf{w}(n)$. De (11.13) se obtiene fácilmente:

$$d(n) - \mathbf{w}^H(n)\mathbf{x}(n) = d(n) - \mathbf{w}^H(n-1)\mathbf{x}(n) - \mu e_p(n) \tilde{\mathbf{x}}^H(n)\mathbf{x}(n)$$

y de aquí es directo que

$$e_p(n) = \frac{e(n)}{1 + \mu \tilde{\mathbf{x}}^H(n)\mathbf{x}(n)}$$

Por lo que finalmente se obtiene que

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \frac{\mu}{1 + \mu \tilde{\mathbf{x}}^H(n)\mathbf{x}(n)} \tilde{\mathbf{x}}(n) e^*(n) \quad (11.14)$$

Esta recursión para la actualización de los pesos del filtro adaptativo, tiene el aspecto de las utilizadas en los conocidos como algoritmos normalizados tipo LMS. En este caso, se estaría trabajando, por tanto, con una versión de un algoritmo normalizado para GLMS, basado en el uso del error a posteriori. Se denomina por tanto Posteriori Normalized Generalized Least Mean Square Algorithm (PNGLMS). Los pasos del proceso iterativo que define se detallan en la tabla 11.3.

<p>Inicialización: $\mathbf{w}(0) = \mathbf{0}$</p> <p>Con la llegada de un nuevo dato: construir $\mathbf{x}(n)$, $\tilde{\mathbf{x}}(n)$ y $d(n)$ $e(n) = d(n) - \mathbf{w}^H(n-1)\mathbf{x}(n)$ $\mathbf{w}(n) = \mathbf{w}(n-1) + \frac{\mu}{1 + \mu \tilde{\mathbf{x}}^H(n)\mathbf{x}(n)} \tilde{\mathbf{x}}(n) e^*(n)$</p>
--

Tabla 11.3 Proceso iterativo del algoritmo PNGLMS

La motivación que dió lugar al algoritmo LMS normalizado es minimizar la distancia entre $\mathbf{w}(n-1)$ y $\mathbf{w}(n)$ en cada iteración, sujetos a una cierta condición. La misión del denominador en (11.14) sería contrarrestar las posibles oscilaciones puntuales tanto de $\tilde{\mathbf{x}}(n)$ como de $\mathbf{x}(n)$; así, si cambia mucho $\tilde{\mathbf{x}}(n)$ en un instante dado, no tiene porqué hacerlo $\mathbf{w}(n)$.

11.3.1 Convergencia en media.

Partiendo de (11.14) y sumando y restando a cada miembro los términos adecuados se obtiene sin dificultad,

$$\begin{aligned} \Delta \mathbf{w}(n) = & \left[\mathbf{I} - \frac{\mu}{1 + \mu \mathbf{x}^H(n) \tilde{\mathbf{x}}(n)} \tilde{\mathbf{x}}(n) \mathbf{x}^H(n) \right] \Delta \mathbf{w}(n-1) + \\ & + \frac{\mu}{1 + \mu \mathbf{x}^H(n) \tilde{\mathbf{x}}(n)} \tilde{\mathbf{x}}(n) e_0^*(n) \end{aligned} \quad (11.15)$$

donde $e_0(n) = d(n) - \mathbf{w}_0^H \mathbf{x}(n)$.

Antes de continuar se recuerdan las hipótesis de independencia usuales, reproducidas aquí por conveniencia:

- a) Los vectores de entrada $\mathbf{x}(1)$, $\mathbf{x}(2)$, ..., $\mathbf{x}(n)$ constituyen una serie de vectores estadísticamente independientes.
- b) Los vectores de variable instrumental $\tilde{\mathbf{x}}(1)$, $\tilde{\mathbf{x}}(2)$, ..., $\tilde{\mathbf{x}}(n)$ constituyen una serie de vectores estadísticamente independientes.
- c) Los escalares de respuesta deseada $d(1)$, $d(2)$, ..., $d(n)$ constituyen una serie de escalares estadísticamente independientes.
- d) Las series de vectores $\{\mathbf{x}(n)\}$ y $\{\tilde{\mathbf{x}}(n)\}$ son conjuntamente estacionarias.

A la vista de (11.14) se desprende que $\mathbf{w}(n)$ sólo depende de muestras $\{d(n)\}$, $\{\mathbf{x}(n)\}$ y $\{\tilde{\mathbf{x}}(n)\}$ hasta tiempo n y que por tanto es independiente de $\mathbf{x}(n+1)$, $\tilde{\mathbf{x}}(n+1)$ y $d(n+1)$ por las 3 primeras hipótesis de independencia.

Tomando valor esperado en (11.15)

$$\begin{aligned} E[\Delta \mathbf{w}(n)] = & \left[\mathbf{I} - E \left[\frac{\mu}{1 + \mu \mathbf{x}^H(n) \tilde{\mathbf{x}}(n)} \tilde{\mathbf{x}}(n) \mathbf{x}^H(n) \right] \right] E[\Delta \mathbf{w}(n-1)] + \\ & + E \left[\frac{\mu}{1 + \mu \mathbf{x}^H(n) \tilde{\mathbf{x}}(n)} \tilde{\mathbf{x}}(n) e_0^*(n) \right] \end{aligned} \quad (11.16)$$

Por su utilidad es conveniente considerar el siguiente comentario: Se desea calcular el valor esperado del producto de dos variables aleatorias \mathcal{A} y \mathcal{B} , una de

las cuales, por ejemplo \mathcal{A} , varía muy lentamente respecto de la otra. Se puede suponer por tanto que se mantiene en torno a su valor medio mientras la otra cambia. Es decir, se puede aceptar que son independientes:

$$E[\mathcal{A}\mathcal{B}] = E[\mathcal{A}]E[\mathcal{B}]$$

Suponiendo que μ es muy pequeña y tomando como base el razonamiento anterior, los valores esperados de (11.16) se pueden aproximar por

$$\begin{aligned} E\left[\frac{\mu}{1 + \mu\mathbf{x}^H(n)\tilde{\mathbf{x}}(n)}\tilde{\mathbf{x}}(n)\mathbf{x}^H(n)\right] &\approx \mu E\left[\frac{1}{1 + \mu\mathbf{x}^H(n)\tilde{\mathbf{x}}(n)}\right] E[\tilde{\mathbf{x}}(n)\mathbf{x}^H(n)] \equiv \\ &\equiv \mu\sigma_{x\tilde{x}}(\mu)\mathbf{R} \end{aligned} \quad (11.17)$$

$$E\left[\frac{\mu}{1 + \mu\mathbf{x}^H(n)\tilde{\mathbf{x}}(n)}\tilde{\mathbf{x}}(n)e_0^*(n)\right] \approx \mu E\left[\frac{1}{1 + \mu\mathbf{x}^H(n)\tilde{\mathbf{x}}(n)}\right] E[\tilde{\mathbf{x}}(n)e_0^*(n)] = 0 \quad (11.18)$$

Llegando por tanto a

$$E[\Delta\mathbf{w}(n)] = [\mathbf{I} - \mu\sigma_{x\tilde{x}}(\mu)\mathbf{R}]E[\Delta\mathbf{w}(n-1)] \quad (11.19)$$

El cálculo exacto de $\sigma_{y\tilde{x}}(\mu)$ es muy laborioso, y además, aunque se llevara a cabo, al estar suponiendo un tamaño de paso pequeño, sólo interesarían los primeros términos de su desarrollo en torno a $\mu \sim 0$, ya que sólo estos serían los significativos. Se hará por tanto el correspondiente desarrollo en serie de Taylor:

$$\begin{aligned} \sigma_{x\tilde{x}}(\mu) &= E\left[\frac{1}{1 + \mu\mathbf{x}^H(n)\tilde{\mathbf{x}}(n)}\right] = E[1 - \mu\mathbf{x}^H(n)\tilde{\mathbf{x}}(n)] + \theta(\mu^2) = \\ &= 1 - \mu\text{Tr}\{\mathbf{R}^H\} + \theta(\mu^2) \end{aligned} \quad (11.20)$$

Si se cortase por el primer término del desarrollo, es decir, por el 1, se estaría en el caso del GLMS. Para ver lo que aporta novedoso el algoritmo PNLMS hay que ir un paso más allá y cortar a orden μ ; entonces (11.19) se convierte en:

$$E[\Delta\mathbf{w}(n)] = [\mathbf{I} - \mu\mathbf{R} + \mu^2\text{Tr}\{\mathbf{R}^H\}\mathbf{R}]E[\Delta\mathbf{w}(n-1)]$$

El análisis de la convergencia de esta ecuación es el establecido para este tipo de casos, como se hizo por ejemplo con el GLMS. Trabajando en la base en la que \mathbf{R} es diagonal se desacoplan las distintas componentes del vector de error de los pesos. Por ejemplo, para la componente i -ésima se tendría:

$$E[\Delta\mathbf{w}(n)]_i = [1 - \mu\lambda_i + \mu^2\text{Tr}\{\mathbf{R}^H\}\lambda_i]E[\Delta\mathbf{w}(n-1)]_i$$

La solución de la ecuación en diferencias anterior convergerá a cero si se verifica que

$$|1 - \mu\lambda_i + \mu^2\text{Tr}\{\mathbf{R}^H\}\lambda_i|^2 < 1$$

lo que equivale a imponer que, a orden μ^2 ,

$$-1 < \underset{b)}{1 - 2\mu\Re(\lambda_i) + \mu^2|\lambda_i|^2 + 2\mu^2\Re(\lambda_i\text{Tr}\{\mathbf{R}^H\})} < \underset{a)}{1}$$

A continuación se estudia qué restricciones imponen las desigualdades anteriores:

Análisis de la desigualdad en a)

Esta desigualdad es equivalente a:

$$\mu(-2\Re(\lambda_i) + \mu|\lambda_i|^2 + 2\mu\Re(\lambda_i\text{Tr}\{\mathbf{R}^H\})) < 0$$

o expresado de otra manera $\mu f(\mu) < 0$ donde $f(\mu) = -2\Re(\lambda_i) + \mu|\lambda_i|^2 + 2\mu\Re(\lambda_i\text{Tr}\{\mathbf{R}^H\})$. Es decir, μ y $f(\mu)$ deben tener distinto signo, lo cual indica que el intervalo válido para μ es aquel para el cual la representación de $f(\mu)$ en función de μ se encuentra en el 2º o 4º cuadrante. La función $f(\mu)$ tiene como propiedades características:

$$\text{cero en } \mu_0 = \frac{2\Re(\lambda_i)}{|\lambda_i|^2 + 2\Re(\lambda_i\text{Tr}\{\mathbf{R}^H\})}$$

$$\text{pendiente } m = |\lambda_i|^2 + 2\Re(\lambda_i\text{Tr}\{\mathbf{R}^H\})$$

Es necesario distinguir 4 casos dependiendo de los signos del cero y de la pendiente:

- i) $m > 0$ y $\mu_0 > 0$: En este caso f se mantiene en el 4º cuadrante para valores de μ comprendidos entre 0 y μ_0 .
- ii) $m < 0$ y $\mu_0 > 0$: son válidas dos situaciones, bien para $\mu < 0$ o para $\mu > \mu_0$.
- iii) $m > 0$ y $\mu_0 < 0$: f se mantiene en el 2º cuadrante para $\mu_0 < \mu < 0$.
- iv) $m < 0$ y $\mu_0 < 0$: vuelven a ser válidas dos situaciones, para $\mu > 0$ y $\mu < \mu_0$.

Análisis de la desigualdad en b)

Imponer la desigualdad b) es lo mismo a imponer que:

$$0 < 2 - 2\mu\Re(\lambda_i) + \mu^2|\lambda_i|^2 + 2\mu^2\Re(\lambda_i\text{Tr}\{\mathbf{R}^H\}) \equiv f_2(\mu)$$

Es necesario por tanto que esta nueva función $f_2(\mu)$ sea positiva.

Para saber el intervalo de valores de μ para el cual $f_2(\mu)$ es positiva, es conveniente calcular sus ceros:

$$\mu_{1,2} = \frac{\Re(\lambda_i) \pm \sqrt{\Delta}}{|\lambda_i|^2 + 2\Re(\lambda_i \text{Tr}\{\mathbf{R}^H\})}$$

donde $\Delta \equiv \Re^2(\lambda_i) - 2[|\lambda_i|^2 + 2\Re(\lambda_i \text{Tr}\{\mathbf{R}^H\})]$. A partir de aquí se pueden discutir los siguientes casos:

1. $\Delta < 0$. Lo cual implica dos cosas: primero, que los ceros de la función son complejos y por tanto no corta a los ejes; y segundo, que $0 < \Re^2(\lambda_i) < 2[|\lambda_i|^2 + 2\Re(\lambda_i \text{Tr}\{\mathbf{R}^H\})]$ y por tanto:

$$\lim_{\mu \rightarrow \infty} f_2(\mu) = +\infty$$

Uniendo ambas se llega a la conclusión de que $f_2(\mu)$ es positiva $\forall \mu$.

2. $\Delta > 0$. Siguiendo un razonamiento análogo al caso anterior se puede afirmar que: primero, los ceros son reales y por tanto la función corta a los ejes en dos puntos; y segundo, que dependiendo del signo de $g = |\lambda_i|^2 + 2\Re(\lambda_i \text{Tr}\{\mathbf{R}^H\})$ el límite de la función será $+\infty$ ó $-\infty$ y por tanto:

- (a) Si $\text{signo}(g) > 0$ $f_2(\mu)$ será positiva en los intervalos $\mu < \mu_1$ y $\mu > \mu_2$
- (b) Si $\text{signo}(g) < 0$ $f_2(\mu)$ será positiva en el intervalo $\mu_1 < \mu < \mu_2$

En definitiva, el intervalo de valores para μ válido será la intersección de los intervalos que se obtengan de a) y b) más la imposición de que sea pequeño¹.

Caso particular

Supongase que \mathbf{R} es una matriz 1×1 , con autovalor λ , y $\Re(\lambda) > 0$.

¿Qué impone a)? que $m = 3|\lambda|^2 > 0$. Estando por tanto en el caso i), lo cual implica que debe verificarse que:

$$0 < \mu < \frac{2}{3}\Re\left(\frac{1}{\lambda}\right) \quad (11.21)$$

¿Qué impone b)? que $\Delta = -5\Re^2(\lambda) - 6\Im^2(\lambda) < 0$. Por tanto, todo valor del tamaño del paso es válido.

Uniendo las restricciones impuestas por a) y b), se llega a que el tamaño del paso debe cumplir la condición especificada en (11.21).

¹Todo el análisis se basa directa o indirectamente en que μ sea muy pequeño, ya sea al despreciar términos, ya al aplicar hipótesis de independencia, etc.

11.3.2 Convergencia en media cuadrática

Se parte de la ecuación (11.15), reproducida aquí por conveniencia.

$$\begin{aligned} \Delta \mathbf{w}(n) = & \left[\mathbf{I} - \frac{\mu}{1 + \mu \mathbf{x}^H(n) \tilde{\mathbf{x}}(n)} \tilde{\mathbf{x}}(n) \mathbf{x}^H(n) \right] \Delta \mathbf{w}(n-1) + \\ & + \frac{\mu}{1 + \mu \mathbf{x}^H(n) \tilde{\mathbf{x}}(n)} \tilde{\mathbf{x}}(n) e_0^*(n) \end{aligned} \quad (11.22)$$

Como μ es pequeña se aplicará la técnica del promediado directo y la aproximación descrita en (11.17), por lo que la expresión anterior se modifica para obtener,

$$\Delta \mathbf{w}(n) = [\mathbf{I} - \mu \sigma_{x\tilde{x}}(\mu) \mathbf{R}] \Delta \mathbf{w}(n-1) + \frac{\mu}{1 + \mu \mathbf{x}^H(n) \tilde{\mathbf{x}}(n)} \tilde{\mathbf{x}}(n) e_0^*(n)$$

Ésta será la ecuación de partida para construir el MSD. Si se construye la matriz $\mathbf{w}(n) \mathbf{w}^H(n)$, se toma valor esperado y se tienen en cuenta las hipótesis de independencia, junto con (11.18), resulta,

$$\mathbf{R}_\Delta(n) = [\mathbf{I} - \mu \sigma_{x\tilde{x}}(\mu) \mathbf{R}] \mathbf{R}_\Delta(n-1) [\mathbf{I} - \mu \sigma_{x\tilde{x}}^*(\mu) \mathbf{R}^H] + \mu^2 \sigma_{2x\tilde{x}}(\mu) \mathbf{R}_{e\tilde{x}}$$

donde se han definido las siguientes cantidades:

$$\begin{aligned} \mathbf{R}_{e\tilde{x}} &= \mathbb{E}[\tilde{\mathbf{x}}(n) \tilde{\mathbf{x}}^H(n) | e_0(n)|^2] \\ \sigma_{2x\tilde{x}} &= \mathbb{E} \left[\left| \frac{1}{1 + \mu \mathbf{x}^H(n) \tilde{\mathbf{x}}(n)} \right|^2 \right] \end{aligned}$$

A la vista de los resultados obtenidos para PNLMS, cabe resaltar que se pueden obtener los correspondientes a GLMS haciendo $(1 + \mu \mathbf{x}^H(n) \tilde{\mathbf{x}}(n))^{-1} \sim 1$.

Para estudiar las condiciones de convergencia de la solución y en su caso a qué converge, se trabajará en la base en la que \mathbf{R} es diagonal. Por lo que la ecuación que rige la evolución dinámica de los elementos diagonales de la matriz de autocorrelación del vector de error en los pesos es:

$$D_i(n) = |1 - \mu \sigma_{x\tilde{x}}(\mu) \lambda_i|^2 D_i(n-1) + \mu^2 \sigma_{e\tilde{x}}^2 \sigma_{2x\tilde{x}}(\mu) \quad (11.23)$$

De esta manera se obtienen las mismas condiciones para la convergencia en media cuadrática que las necesarias para la convergencia en media.

Suponiendo convergencia en (11.23), es decir, $D_i(n) = D_i(n-1) = D_i(\infty)$, resulta,

$$D_i(\infty) = \frac{\mu^2}{1 - |1 - \mu \sigma_{x\tilde{x}}(\mu) \lambda_i|^2} \sigma_{2x\tilde{x}}(\mu) \sigma_{e\tilde{x}}^2 \quad (11.24)$$

Para encontrar una expresión un poco más operativa se tiene, exactamente igual que antes, que al ser μ pequeña, basta con la aproximación $(1 +$

$\mu \mathbf{x}^H(n) \tilde{\mathbf{x}}(n))^{-1} = 1 - \mu \mathbf{x}^H(n) \tilde{\mathbf{x}}(n)$. Así $\sigma_{x\tilde{x}}(\mu)$ toma la forma descrita en (11.20) y para la nueva cantidad queda, en el orden más bajo en μ :

$$\sigma_{2x\tilde{x}}(\mu) \sim \mathbb{E}[|1 - \mu \mathbf{x}^H(n) \tilde{\mathbf{x}}(n)|^2] \sim 1 - 2\mu \Re(\text{Tr}\{\mathbf{R}^H\})$$

Con estas aproximaciones la expresión (11.24) queda:

$$D_i(\infty) = \frac{\mu}{2\Re(\lambda_i) - \mu[2\Re(\lambda_i \text{Tr}\{\mathbf{R}^H\}) + |\lambda_i|^2]} \sigma_{e\tilde{x}}^2 [1 - 2\mu \Re(\text{Tr}\{\mathbf{R}^H\})]$$

obviamente es una corrección a orden μ del resultado para GLMS. Sin embargo en el análisis realizado para el GLMS se solían despreciar los términos en μ del denominador por ser muy pequeños. Aquí no se hace porque en ellos radica la diferencia entre GLMS y PNGLMS.

Tiempos de convergencia

Sería conveniente tener una estimación del tiempo que se necesita para que la solución de la homogénea de (11.23) tienda a 0. Procediendo como se ha hecho con el resto de ecuaciones en diferencias de esta Memoria, se puede afirmar que (véase apéndice D),

$$\tau_i = \frac{-1}{\ln(|1 - \mu \sigma_{x\tilde{x}}(\mu) \lambda_i|^2)} \underset{\mu \rightarrow 0}{\sim} \tau_i = \frac{1}{1 - |\lambda_i|^2} \quad (11.25)$$

A la vista de esta expresión, (11.24) se puede reformular para escribirse como

$$D_i(\infty) = \tau_i \mu^2 \sigma_{2x\tilde{x}}(\mu) \sigma_{e\tilde{x}}^2 \quad (11.26)$$

Esta expresión puede parecer poco lógica, ya que viene a decir que cuanto mayor sea el tiempo de convergencia, mayor será el MSD en convergencia. Es decir, que cuanto más rápido converja, menor varianza tendrán los estimadores de los pesos. Esto va en contra de los resultados del análisis de todos los algoritmos adaptativos.

Sin embargo, no se puede perder de vista que el único parámetro libre que controla el algoritmo es el tamaño del paso μ . Por tanto, como el tiempo de convergencia τ_i depende de él también, el factor clave en (11.26) es el producto $\tau_i \mu^2$. Y este producto se comporta linealmente con el tamaño del paso. Básicamente se puede afirmar que

$$D_i(\infty) \sim \mu \quad (11.27)$$

Retómese el razonamiento anterior a la vista de este resultado. Si aumenta el tiempo de convergencia es porque disminuye el tamaño del paso (ver (11.25)) y por tanto, disminuye el MSD en convergencia (ver (11.27)). Este resultado sí es el que cabía esperar, conocido el comportamiento de algoritmos como el LMS, GLMS y variantes.

11.4 Simulaciones

Con las simulaciones que se van a presentar en este apartado se pretende solamente justificar las versiones creadas del algoritmo GLMS, un estudio más riguroso, tanto teórico como experimental, se deja para posteriores trabajos de investigación.

11.4.1 Algoritmo CGLMS

Para mostrar la convergencia de este algoritmo, junto con su propiedad de convergencia incluso cuando la matriz \mathbf{R} asociada es no definida, se ha elegido probarlo en la identificación ciega del mismo modelo AR(2) del apartado 10.8.1 con el mismo proceso de excitación 4-Q.A.M.

En esta ocasión se dispone de 30000 muestras y se promedian los resultados de 200 realizaciones. El tamaño del paso de CGLMS se ha tomado como $\mu = 5 \cdot 10^{-11}$. La evolución dinámica del MSD se muestra en la gráfica 11.1. Como se aprecia la convergencia es más que satisfactoria. Mirando con detalle a

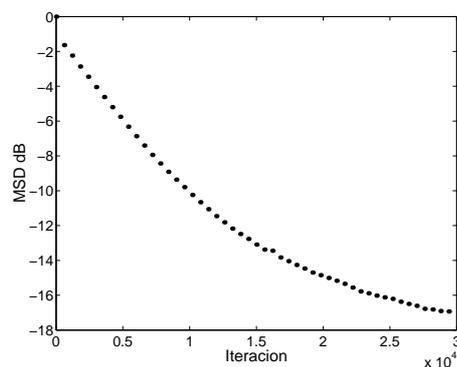


Figura 11.1 Convergencia de CGLMS con \mathbf{R} no definidas

la ecuación de actualización del algoritmo CGLMS en la tabla 11.1 se puede advertir que este algoritmo es una especie de LMS con tamaño del paso variable² e igual a $\mu \tilde{\mathbf{x}}^t(n) \tilde{\mathbf{x}}(n)$; a la vista de lo cual no es de extrañar su buen comportamiento. Sería sumamente interesante probar este algoritmo en ambientes ruidosos para determinar en qué medida aumenta el sesgo de sus estimadores.

²Otras versiones con μ variable se pueden encontrar en [AB01], [GWK99] y [HTL00], entre otros.

11.4.2 Algoritmo A1GLMS

La principal motivación de este algoritmo fue poder tratar situaciones donde la matriz asociada no fuera definida y se planteó con la hipótesis de independencia como base. Tras un análisis mediante simulaciones bastante extenso no se ha podido conseguir que este algoritmo converja más allá de los -1 dB con 40000 muestras para el conocido sistema AR= $[1 \ -1 + 0.9i \ 0.04 \ -0.44i]$. Por tanto este algoritmo no tiene comportamiento aceptable. La principal consecuencia de este análisis es clara: aunque es conocido que la hipótesis que afirma que las series de vectores $\tilde{\mathbf{x}}(n)$ y $\mathbf{x}(n)$ son i.i.d no es cierta, permite obtener expresiones analíticas sencillas para la evolución de las magnitudes de un algoritmo (tales como MSD(n)), pero no son útiles a la hora de formular un nuevo algoritmo basado en ellas.

11.4.3 Algoritmo PNGLMS

La motivación de este algoritmo no fue más que intentar mejorar el comportamiento de GLMS con la inclusión del error a posteriori en su ecuación de actualización. Para determinar en primer lugar si PNGLMS de hecho converge y en segundo lugar si mejora los resultados de GLMS, se ha procedido a la identificación del modelo AR(1)= $[1 \ -0.4 + 0.4i]$, excitado por la ya anteriormente mencionada constelación 4-Q.A.M., mediante ambos algoritmos.

En esta ocasión se emplearon series de 10000 muestras de longitud y se promediaron 200 realizaciones; tanto para PNGLMS como para GLMS se tomó $\mu = 3 \cdot 10^{-5}$. La evolución dinámica del MSD se detalla en las figuras 11.2(a) y 11.2(b) respectivamente.

Al ser PNGLMS una corrección en orden μ a GLMS el resultado no puede ser muy diferente por lo que no se puede apreciar en las figuras anteriores, lo que se podría esperar es una pequeña mejora. Para constatar esta mejora en la figura 11.3 se representa la evolución de la diferencia de los MSD de ambos algoritmos en la fase de convergencia. Como era de esperar PNGLMS consigue un MSD ligeramente menor al trabajar con el error a posteriori.

11.5 Diagrama resumen

Para concluir este capítulo y esta segunda parte sería conveniente recapitular lo hecho en toda esta segunda parte y cómo se ha llegado a derivar el algoritmo A1GLMS y las versiones directas del GLMS. En la figura 11.4 se muestra todo el proceso seguido, que comenzó en primer lugar con el análisis de la convergencia

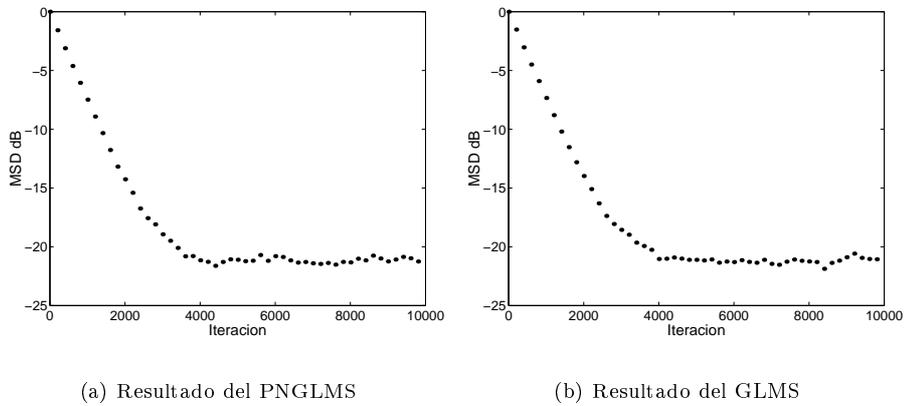


Figura 11.2 Comparación PNGLMS y GLMS con el mismo tamaño del paso

de ORIV en la primera parte de la Memoria. Este análisis permitió enunciar el Principio de Ortogonalidad Sobredeterminado y Generalizado, mediante el cual y la idea de los algoritmos de máxima pendiente, se pudo proponer el algoritmo OGLMS. Descomponiendo el vector gradiente utilizado en la ecuación de actualización de los pesos para este algoritmo se podían derivar dos nuevos algoritmos, el OGLMS1 y el OGLMS3. El primero de ellos resultó tener mejores propiedades de convergencia y fue elegido para ir siendo mejorado, llegando a la formulación del algoritmo AOGGLMS; el segundo de ellos se empleó para, en un segundo paso de simplificación, recuperar el ya existente algoritmo GLMS que encuentra cabida natural en este esquema formal.

El algoritmo GLMS es el primer algoritmo de todos los mencionados hasta ahora que no es capaz de converger cuando la matriz \mathbf{R} asociada al problema que está resolviendo no es definida. Sin embargo, esta desventaja se puede salvar introduciendo modificaciones a GLMS para proponer las versiones conocidas como CGLMS y A1GLMS, que además también son capaces de converger cuando la matriz es sobredeterminada. En general, los algoritmos sobredeterminados no sufren esta limitación.

No obstante, la motivación del algoritmo PNGLMS, versión del GLMS, fue la de mejorar su convergencia trabajando con el error a posteriori (en la fase estacionaria el error a priori y a posteriori coinciden) y por tanto no es sobredeterminado.

Finalmente, el ciclo presentado en los párrafos anteriores se puede cerrar al tener en cuenta la siguiente consideración: el algoritmo CGLMS se puede

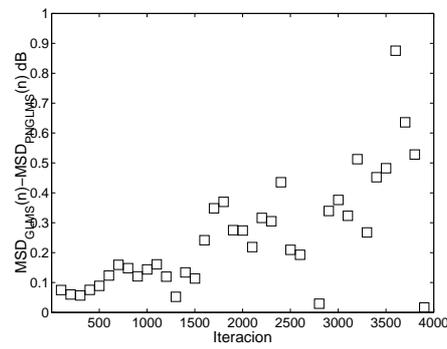


Figura 11.3 Comparación PNLMS y GLMS en la fase de convergencia

considerar como una aproximación estocástica extrema del algoritmo AONGLMS. En efecto, si en la ecuación de actualización de los pesos del segundo se sustituye el operador valor esperado por el estimador instantáneo, se recupera el primero.

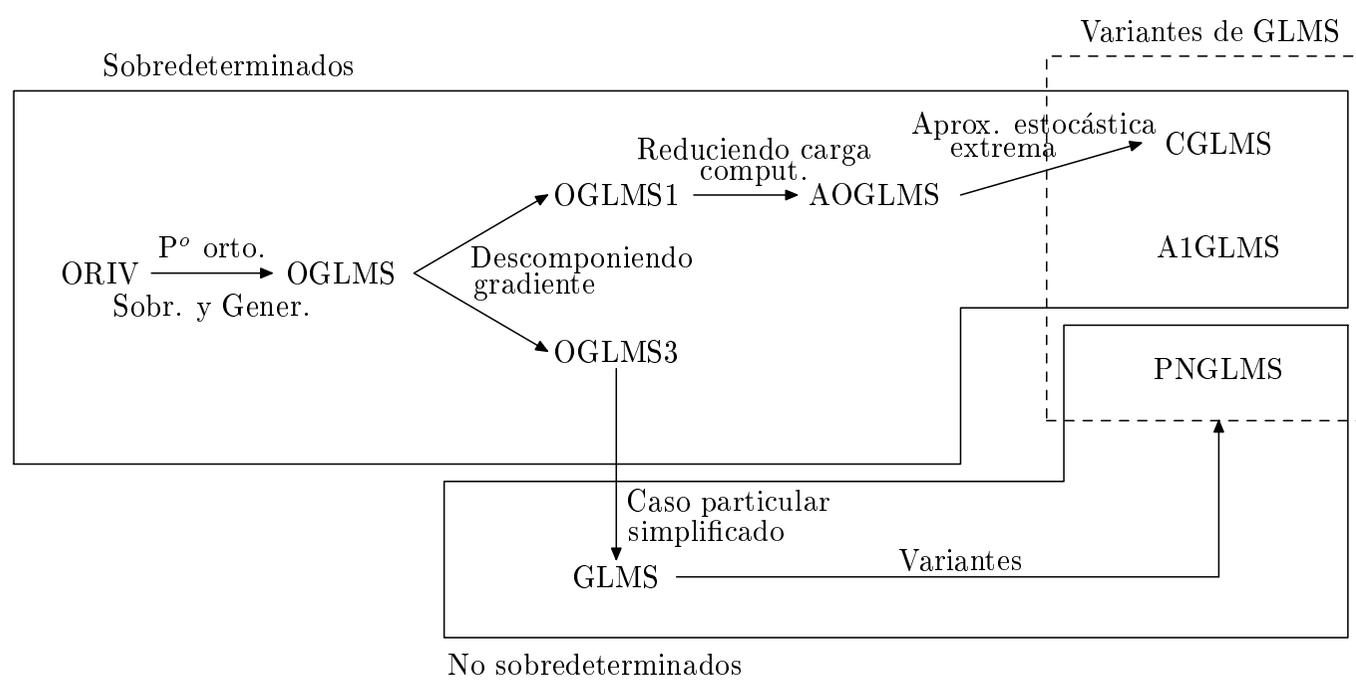


Figura 11.4 Diagrama resumen de la relación entre los algoritmos estudiados en la segunda parte

Parte III

Identificación ciega con estadística de segundo orden. Métodos de elección de la fase

Introducción

En esta parte se continúa con el estudio del problema de la identificación ciega de sistemas iniciada en partes anteriores. La principal diferencia con ellas es que aquí buena parte de la resolución utiliza estadística de segundo orden y por tanto no se basa exclusivamente en estadística de alto orden. Como es conocido, de esta manera se requerirá completar la técnica que identifica al sistema de forma que conste al menos de dos pasos. Si en un primer paso se calculan los coeficientes del sistema, empleando estadística de segundo orden de su salida, se obtendrán como solución sistemas que compartan el espectro de potencia con el sistema problema, pero no se tiene la certeza de que compartan también la respuesta en fase porque la secuencia de autocorrelación no contiene información sobre la misma. De hecho se puede comprobar que de la estadística de segundo orden se pueden obtener todos los sistemas espectralmente equivalentes del sistema a identificar.

Como segundo paso queda por tanto elegir el sistema con la fase correcta, para lo cual hay que recurrir a usar otras herramientas, como por ejemplo el uso de criterios basados en la optimización de funciones, estadística de alto orden, etc.

Obviamente, el método así descrito, que consta de 2 pasos, sufre de la desventaja de tener que soportar más carga computacional principalmente por dos motivos:

1. Tiene que calcular todos los sistemas espectralmente equivalentes a pesar de sólo interesar uno.
2. Precisamente por esto tiene que llevar a cabo un proceso de búsqueda dentro del conjunto espectralmente equivalente para seleccionar al sistema correcto.

A este tipo de métodos se le conoce en la literatura como métodos basados en el sistema espectralmente equivalente de fase mínima (Spectrally Equivalent

Mimumum Phase SEMP), porque a partir del SEMP calculan todo el conjunto espectralmente equivalente y después llevan a cabo el proceso de búsqueda.

Son varios los objetivos que se persiguen en esta parte de la Memoria:

1. Obtener expresiones que relacionen explícitamente los coeficientes del sistema lineal con la estadística de segundo orden de su salida.
2. Analizar las propiedades de estas expresiones: valores que toman las soluciones, sensibilidad a errores de estimación, etc.
3. Basándose en el estudio de los puntos anteriores derivar un nuevo método que evite el proceso de búsqueda y que por tanto tenga menor complejidad computacional.
4. Estudiar diversos criterios para la elección del sistema con la fase correcta, justificando adecuadamente los resultados obtenidos con ellos en las simulaciones.

Las expresiones y el análisis para los sistemas de primer y segundo orden se detallan en los capítulos 12 y 13 respectivamente. En el capítulo 14 se propone un nuevo método, inspirado en el estudio de los capítulos anteriores, que calcula los coeficientes del sistema usando principalmente estadística de segundo orden, pero complementada mínimamente con estadística de orden superior. Por este motivo se la ha denominado método de Identificación Ciega mediante estadística de Segundo orden y Mínima de Alto ICSMA. En este mismo capítulo 14 se propone una solución iterativa basada en ICSMA para sistemas de tercer orden.

Finalmente, en el capítulo 15, se describen métodos de elección de la fase ya descritos en la bibliografía pero no estudiados con cierta profundidad. A tal fin se definen magnitudes nuevas que permiten justificar teóricamente el comportamiento observado en las simulaciones: Mínima Diferencia Relativa MDR, Diferencia Relativa Promedio DRP, etc. Esta justificación también se lleva a cabo clasificando a los modelos estudiados según conceptos como calidad, riesgo, etc.

Capítulo 12

Sistemas de primer orden

Introducción

La estructura de este primer capítulo de la parte III se puede considerar como un modelo a seguir para el resto, ya que presenta en un sistema de primer orden, donde la matemática es sencilla, el planteamiento básico utilizado. Primero se hallan expresiones para los coeficientes del filtro en función de la secuencia de autocorrelación de la salida. Como ya es conocido las soluciones obtenidas son espectralmente equivalentes al sistema problema. Debido a que estas expresiones se utilizarán con estimadores de la secuencia de autocorrelación, se hace recomendable llevar a cabo un estudio de la propagación de los errores de estimación, y ver en qué situaciones esta propagación puede llevar a un aumento del error nada despreciable. Este análisis permite hablar de la sensibilidad asociada a un sistema dado, factor fundamental en el capítulo 15 que presenta los resultados de las simulaciones.

En el apartado 1 se expone el problema a resolver más detalladamente; en el apartado 2 se presenta la relación entre la estadística de segundo orden y los coeficientes del sistema; para las expresiones derivadas en este apartado, se estudia la sensibilidad a pequeñas variaciones de la secuencia de autocorrelación en el apartado 3. Finalmente, en el apartado 4, para caracterizar situaciones donde el error propagado sea muy elevado, se presenta la localización de los ceros para problemas mal condicionados.

12.1 Exposición del problema

Considérese un sistema lineal definido por su respuesta impulso, que se supondrá finita y real $\{h(n)\}$, excitado mediante un proceso aleatorio, independiente e idénticamente distribuido i.i.d $\{w(n)\}$ de estadística conocida con varianza σ_w^2 y curtosis γ_{4w} no nula. La salida del filtro se puede escribir como

$$x(n) = \sum_{i=0}^q h(i)w(n-i) \quad (12.1)$$

donde q es el orden del modelo. En general la salida puede estar contaminada por ruido aditivo gaussiano blanco $\{v(n)\}$ de varianza conocida. De esta manera la salida total del sistema es

$$y(n) = x(n) + v(n) \quad (12.2)$$

El problema es determinar los coeficientes del filtro mediante medidas solamente de la salida del sistema bajo las hipótesis anteriores. En primera instancia sólo se utilizará la información contenida en la estadística de segundo orden. En lo que sigue la secuencia de autocorrelación $r_{xx}(m) = E[x(n)x(n+m)]$ de la salida no contaminada se considera conocida. Bajo la hipótesis de i.i.d de la entrada queda

$$r_m = \sigma_w^2 \sum_{i=0}^q h_i h_{i+m} \quad (12.3)$$

donde por simplificar la notación $h_i = h(i)$ y $r_m = r_{xx}(m)$.

12.2 Relación entre estadística de segundo orden y coeficientes del sistema

Sea el sistema de primer orden definido por el vector de coeficientes $\mathbf{h} = [h_0 \ h_1]$, de manera que la relación entre la secuencia de autocorrelación de la salida y estos coeficientes es

$$r_0 = \sigma_w^2 (h_0^2 + h_1^2) \quad (12.4)$$

$$r_1 = \sigma_w^2 h_0 h_1 \quad (12.5)$$

En lo que sigue se supondrá que la secuencia de autocorrelación ha sido normalizada por la varianza de $\{w(n)\}$, o equivalentemente que se trabaja con varianza unidad. No es difícil demostrar que los coeficientes del filtro satisfacen las siguientes ecuaciones:

$$\hat{h}_0^2 = \frac{r_0 \pm \sqrt{r_0^2 - 4r_1^2}}{2} \quad (12.6)$$

$$\hat{h}_1 = \frac{r_1}{\hat{h}_0} \quad (12.7)$$

Así que hay 4 soluciones diferentes al problema: 2 provienen de la opción \pm en la expresión (12.6) para h_0^2 y 2 más debido a la libertad existente en el signo una vez se toma raíz cuadrada a h_0^2 . Basicamente hay 2 soluciones, las otras 2 se obtienen cambiando el signo a las dos primeras. Si la señal de entrada es simétrica, no hay manera de distinguir entre un filtro y su opuesto correspondiente, porque sus salidas son estadísticamente equivalentes, de manera que sólo se considerarán valores positivos para h_0 . Como ya es sabido esta variedad de soluciones se debe a que se trabaja con estadística de segundo orden, por ese motivo se le ha puesto sombrero a las coeficientes calculados a partir de ella, para indicar que su valor no es único.

Denotando por $h_0^+(h_0^-)$ el resultado de evaluar la expresión (12.6) tomando el signo $+(-)$, y $h_1^+(h_1^-)$ el resultado de (12.7) usando $h_0^+(h_0^-)$, es sencillo demostrar que $h_1^+h_1^- = r_1$, de manera que

$$h_1^- = \frac{r_1}{h_1^+} = r_1 \frac{h_0^+}{r_1} = h_0^+ \tag{12.8}$$

y análogamente

$$h_1^+ = h_0^- \tag{12.9}$$

Esta propiedad permite resolver el sistema y afirmar que las soluciones son las que se muestran en la tabla 12.1.

	1 ^{er} coeficiente	2 ^o coeficiente
1 ^a opción	h_0	h_1
2 ^a opción	h_1	h_0

Tabla 12.1 *Sistemas espectralmente equivalentes para un sistema de primer orden*

Esta tabla muestra que sólo es necesario calcular un conjunto de los posibles valores de (12.6) - (12.7) , el otro conjunto se obtiene de ellos. Una aproximación parecida se utilizará para sistemas de orden superior para simplificar el problema. La estructura de la tabla 12.1 también se podría haber obtenido reflejando el cero del filtro de la primera opción con respecto al círculo unidad, manteniendo la misma potencia de salida. Como se esperaba las soluciones a la ecuaciones (12.4)- (12.5) son todos sistemas espectralmente equivalentes ya que todos comparten el mismo espectro de potencia (estadística de segundo orden), pero tienen diferente respuesta en fase porque sus estadísticas de alto orden no son las mismas.

12.2.1 Características de la solución

Usando las definiciones de r_0 y r_1 , dadas en (12.4) y (12.5), en la expresión para \hat{h}_0 , dada en (12.6), resulta

$$\hat{h}_0^2 = \frac{h_1^2 + h_0^2 \pm \sqrt{(h_0^2 + h_1^2)^2 - 4h_0^2 h_1^2}}{2} = \frac{h_1^2 + h_0^2 \pm \sqrt{(h_0^2 - h_1^2)^2}}{2} \quad (12.10)$$

donde, volvemos a incidir, el sombrero en h_0 indica que puede ser cualquiera de los estimadores dentro de las opciones válidas para él. De acuerdo con esta expresión el radicando será cercano a cero si $h_0^2 \sim h_1^2$ y un estimador de esta cantidad puede fácilmente ser negativo, de manera que el coeficiente del filtro puede ser complejo incluso si el valor verdadero es un número real. Evitar esta situación por este motivo parece un argumento débil. Es más instructivo afirmar que la degeneración aumenta, ya que tienden a coincidir dos soluciones, y es más difícil distinguir entre sistemas (soluciones).

Utilizando la expresión (12.10) se puede comprobar de manera alternativa que las soluciones para (12.4)-(12.5) son efectivamente las que se muestran en la tabla 12.1. Está claro que dependiendo del signo elegido en (12.10) los posibles valores para \hat{h}_0^2 son h_0^2 y h_1^2 . Aquí de nuevo se pone de manifiesto la imposibilidad de determinar los signos de los coeficientes del filtro. A partir de los valores para \hat{h}_0 los valores para \hat{h}_1 son h_1 y h_0 , confirmando lo afirmado.

12.3 Sensibilidad de las soluciones a pequeñas variaciones en la secuencia de autocorrelación

En aplicaciones reales la secuencia de autocorrelación tiene que ser estimada a partir de la salida del canal. Cuantas más muestras estén disponibles, mejor serán los estimadores, pero en cualquier caso, habrá algunas diferencias entre los valores estimados y teóricos. En este apartado se estudiará la influencia sobre la solución de esta (pequeña) diferencia. Se llevará a cabo mediante derivadas parciales de las expresiones (12.6) y (12.7).

12.3.1 Sensibilidad del coeficiente h_0

Debido a que este coeficiente depende explícitamente de las variables r_0 y r_1 se hace necesario estudiar por separado la sensibilidad proveniente de cada una de ellas.

Variaciones respecto a r_0

La derivada parcial de \hat{h}_0 con respecto a r_0 es

$$\frac{\partial \hat{h}_0}{\partial r_0} = \frac{\sqrt{2}}{4} \left[1 \pm \frac{r_0}{\sqrt{r_0^2 - 4r_1^2}} \right] \frac{1}{(r_0 \pm \sqrt{r_0^2 - 4r_1^2})^{1/2}} \quad (12.11)$$

Si esta derivada parcial tomase un valor elevado en alguna situación, implicaría que un pequeño error en r_0 daría un gran error en \hat{h}_0 , provocando una amplificación del mismo. Estas situaciones críticas pueden pasar si:

Caso a) $r_0^2 \approx 4r_1^2$, que hace que el primer factor de (12.11) diverja. ¿Qué condiciones imponen sobre los coeficientes del filtro? Usando de nuevo las relaciones dadas en (12.4)-(12.5), se sigue que esta condición equivale a $(h_0^2 + h_1^2)^2 \approx 4h_0^2h_1^2$ o lo que es lo mismo $(h_0^2 - h_1^2)^2 \approx 0$ que conduce a $|h_0| \approx |h_1|$. Bajo esta condición la ecuación (12.11) toma la forma

$$\frac{\partial \hat{h}_0}{\partial r_0} \sim \frac{1}{4} \left[1 \pm \frac{r_0}{\epsilon} \right] \frac{\sqrt{2}}{(r_0 \pm \epsilon)^{1/2}} \approx \frac{1}{4} \left[1 \pm \frac{r_0}{\epsilon} \right] \frac{\sqrt{2}}{\sqrt{r_0}} \quad (12.12)$$

donde $\epsilon \sim 0$. La única forma de evitar que el primer factor diverja es asignando a r_0 un valor cercano a cero, pero este arreglo también provoca que el segundo factor diverja, de manera que la divergencia es inevitable. Además, si r_0 es cercano a cero deben serlo también h_0 o h_1 , que bajo la condición anterior implica que ambos deben ser cercanos a cero a la vez, de manera que no queda sistema distinto del trivial [0 0].

Caso b) La segunda situación que hay que considerar es aquella en la que $r_0 \sim \pm \sqrt{r_0^2 - 4r_1^2}$, es decir, cuando \hat{h}_0 está cercano a cero (ver ecuación (12.6)). Esto se puede dar bien cuando h_0 o h_1 son prácticamente nulos (ver tabla 12.1), lo cual significa a su vez que r_1 es también casi nulo, situación efectivamente bajo la cual se verifica la primera condición dada al comienzo de este caso b. Bajo esta condición la ecuación (12.11) toma la siguiente forma

$$\frac{\partial \hat{h}_0}{\partial r_0} \underset{r_1 \sim 0}{\sim} \frac{1}{4} \left[1 \pm \left(1 + 2\frac{r_1^2}{r_0^2} \right) \right] \frac{\sqrt{2}}{\sqrt{r_0} \sqrt{1 \pm (1 - 2(r_1/r_0)^2)}} \quad (12.13)$$

donde se ha empleado el desarrollo en serie de Taylor de $\sqrt{1+x}$ y de $(1+x)^{-1}$.

Caso b.1) Elijiendo el signo + la expresión anterior se reduce a

$$\frac{\partial \hat{h}_0}{\partial r_0} \sim \frac{1}{2} \frac{1}{\sqrt{r_0}} \quad (12.14)$$

Que es finito siempre y cuando r_0 no esté próximo a cero, cosa que sólo ocurre cuando ambos coeficientes son cercanos a cero que vuelve a conducir a una situación trivial.

Caso b.2) Eligiendo el signo $-$ la expresión (12.13), queda

$$\frac{\partial \hat{h}_0}{\partial r_0} = \frac{-1}{2} \frac{r_1}{r_0^{3/2}} \quad (12.15)$$

que como se puede apreciar de nuevo no presenta ningún problema de divergencia a menos que r_0 sea muy pequeño que vuelve a conducir a una situación trivial.

Variaciones respecto a r_1

En este apartado la magnitud que interesa es la derivada parcial de \hat{h}_0 con respecto a r_1 , que tiene la siguiente expresión:

$$\frac{\partial \hat{h}_0}{\partial r_1} = \mp \frac{r_1}{\sqrt{r_0^2 - 4r_1^2}} \frac{\sqrt{2}}{\left(r_0 \pm \sqrt{r_0^2 - 4r_1^2}\right)^{1/2}} \quad (12.16)$$

A continuación se detallan las situaciones en las que se puede dar divergencia de la expresión previa.

Caso a) $r_0^2 \approx 4r_1^2$, usando de nuevo el mismo procedimiento que para el caso a anterior nos lleva a que esta condición se puede dar si $|h_0| \approx |h_1|$ y conduce a que la ecuación (12.16) se pueda reescribir como

$$\frac{\partial \hat{h}_0}{\partial r_1} \sim \mp \frac{r_1}{\epsilon} \frac{\sqrt{2}}{(r_0 \pm \epsilon)^{1/2}} \approx \mp \frac{r_1}{\epsilon} \frac{\sqrt{2}}{\sqrt{r_0}} \quad (12.17)$$

donde de nuevo $\epsilon \sim 0$. Para evitar la divergencia la única opción es ajustar r_1 cercano a cero, pero bajo la condición dada para este caso también implicaría que r_0 tiene que ser cercano a cero y se llegaría a un caso trivial.

Caso b) Es como el caso b anterior, si $r_1 \sim 0$ el segundo factor diverge pero el primero tiende a cero, con lo que este caso necesita un análisis en profundidad. Bajo esta condición y utilizando expansiones en serie de Taylor, la expresión (12.16) se convierte en:

$$\frac{\partial \hat{h}_0}{\partial r_1} \underset{r_1 \sim 0}{\sim} \mp \frac{r_1}{r_0 - 2r_1^2/r_0} \frac{\sqrt{2}}{\left(r_0 \pm (r_0 - 2r_1^2/r_0)\right)^{1/2}} \quad (12.18)$$

De nuevo aparecen dos situaciones dependiendo del signo elegido:

Caso b.1) Si se elige el signo + en el denominador del segundo factor, la ecuación anterior se convierte en

$$\frac{\partial \hat{h}_0}{\partial r_1} \propto \frac{r_1}{r_0 - 2r_1^2/r_0} \frac{1}{\sqrt{r_0 - r_1^2/r_0}} \quad (12.19)$$

no habiendo divergencia por muy pequeño que sea el coeficiente r_1 .

Caso b.2) Si se elige el signo menos en el denominador del segundo factor y tras hacer un poco de álgebra se llega a:

$$\frac{\partial \hat{h}_0}{\partial r_1} \propto \frac{\sqrt{r_0}}{r_0 - 2r_1^2/r_0} \quad (12.20)$$

no presentando tampoco divergencia aunque r_1 sea nulo.

12.3.2 Sensibilidad del coeficiente h_1

El mismo estudio que se ha llevado a cabo para el coeficiente h_0 se va a llevar a cabo en este apartado para el coeficiente h_1 . De nuevo son de interés las derivadas respecto de las variables de las que depende: r_0 y r_1 .

Variaciones respecto a r_0

Teniendo en cuenta la expresión (12.7), la derivada de \hat{h}_1 con relación a r_0 es

$$\frac{\partial \hat{h}_1}{\partial r_0} = -\frac{r_1 \partial \hat{h}_0 / \partial r_0}{\hat{h}_0^2} \quad (12.21)$$

Esta derivada parcial puede diverger si se da alguna de las condiciones siguientes:

Caso a) cuando lo haga $\partial \hat{h}_0 / \partial r_0$, esto es, cuando $h_0^2 \sim h_1^2$. En este caso, usando la definición de r_1 dada en la ecuación (12.5), se llega a

$$\frac{\partial \hat{h}_1}{\partial r_0} \propto \frac{\partial \hat{h}_0}{\partial r_0} \quad (12.22)$$

de manera que la divergencia ocurre bajo esta condición.

Cabo b) Cuando $\hat{h}_0 \sim 0$. En este caso $\partial\hat{h}_0/\partial r_0$ es finita y $r_1 \sim 0$. Si de nuevo se usa la ecuación (12.5) para r_1 se puede llegar a

$$\frac{\partial\hat{h}_1}{\partial r_0} = -\frac{h_1 h_0 \partial\hat{h}_0/\partial r_0}{\hat{h}_0^2} \quad (12.23)$$

Es claro que \hat{h}_0 será cero si alguno de sus posibles opciones es cero, es decir, si h_0 o h_1 son cero. Esto haría que se cancelase uno de ellos en el numerador con un \hat{h}_0 en el denominador, quedando todavía otro que provocaría la divergencia para este caso.

Variaciones con respecto a r_1

Teniendo en cuenta la expresión (12.7) la derivada de \hat{h}_1 con respecto a r_1 es

$$\frac{\partial\hat{h}_1}{\partial r_1} = \frac{\hat{h}_0 - r_1 \partial\hat{h}_0/\partial r_1}{\hat{h}_0^2} \quad (12.24)$$

Para esta expresión las posibles causas de divergencia son:

Caso a) Cuando diverja $\partial\hat{h}_0/\partial r_1$, esto es, cuando $h_0^2 \sim h_1^2$. Si se quiere evitar esta divergencia habría que hacer r_1 lo más cercano a cero, pero bajo esta condición implicaría que ambos coeficientes del filtro deberían ser cero y se llegaría a un filtro trivial. De manera que aquí la divergencia es inevitable.

Caso b) Cuando $\hat{h}_0 \sim 0$. Usando de nuevo el mismo argumento que en el caso b anterior, se puede concluir que la divergencia bajo esta condición tampoco se puede evitar.

Para resumir, en la tabla 12.2 se han recogido todas las situaciones en las que un pequeño error en la estimación de la secuencia de autocorrelación puede dar lugar a grandes desviaciones de los coeficientes estimados del filtro respecto a los verdaderos. Estas situaciones se describen en función de los valores reales de los coeficientes del filtro tal y como se ha hecho en el análisis llevado a cabo en páginas anteriores.

∂/∂	h_0	h_1
r_0	$h_0^2 \sim h_1^2$	$h_0^2 \sim h_1^2$ $\hat{h}_0 \sim 0$
r_1	$h_0^2 \sim h_1^2$	$h_0^2 \sim h_1^2$ $\hat{h}_0 \sim 0$

Tabla 12.2 *Situaciones de alta sensibilidad para la estimación*

Nótese que \hat{h}_0 se ha incluido en la tabla en vez de h_0 porque es este estimador el que juega un papel en la estimación de h_1 usando la ecuación (12.5). De esta forma, la condición sobre \hat{h}_0 se divide en dos para los coeficientes del filtro, es decir, $h_0 \sim 0$ o $h_1 \sim 0$, véase la tabla 12.1.

12.4 Localización de los ceros para problemas mal condicionados

Las condiciones enunciadas en la tabla 12.2 tienen una interpretación práctica. Que $\hat{h}_0 \sim 0$ significa que el sistema que se suponía tenía dos coeficientes tiene en realidad sólo uno, por lo que la salida es sólo la entrada multiplicada por un factor constante y carece de interés. La segunda condición es más rica en información, porque $h_0^2 \sim h_1^2$ significa que el cero del sistema está muy cerca del círculo unidad, así que es más difícil distinguir entre los dos sistemas espectralmente equivalentes (fase mínima o máxima).

Capítulo 13

Sistemas de segundo orden

Introducción

Este capítulo es una generalización del capítulo anterior a sistemas de segundo orden, los pasos y análisis dados en él se generalizan aquí por lo que lo único que representa es una pequeña complejidad matemática añadida. Una vez que se obtienen las soluciones para sistemas de este tipo, se hace necesario estudiar su estructura con detalle, ya que será fundamental para proponer el método de identificación ciega mediante estadística de segundo orden y mínima de alto en capítulos posteriores. Igualmente importante es determinar los sistemas que se pueden considerar como sensibles en este caso.

De nuevo en el apartado 1 se presenta la relación entre la estadística de segundo orden y los coeficientes del sistema, además se discute esta solución y se estudian las características estructurales. En el apartado 2 se comprueba que las soluciones obtenidas en el apartado anterior constituyen todo el conjunto de sistemas espectralmente equivalentes. La sensibilidad de las soluciones a pequeñas variaciones de la secuencia de autocorrelación se estudia en el apartado 3, el porqué de tales situaciones de sensibilidad se presenta en el apartado 4. Finalmente la sensibilidad de la localización de los ceros a pequeñas variaciones de los coeficientes del filtro se da en el apartado 5.

13.1 Relación entre estadística de segundo orden y coeficientes del sistema

Supóngase ahora que el canal de interés es un sistema lineal, invariante temporal, de segundo orden, definido por su respuesta impulso $\mathbf{h} = [h_0 \ h_1 \ h_2]$. La ecuación que relaciona estos coeficientes con la estadística de segundo orden de la salida es

$$r_m = \sigma_w^2 \sum_{i=0}^2 h_i h_{i+m} \quad (13.1)$$

o de forma explícita para cada valor de m

$$r_0 = \sigma_w^2 (h_0^2 + h_1^2 + h_2^2) \quad (13.2)$$

$$r_1 = \sigma_w^2 (h_0 h_1 + h_1 h_2) \quad (13.3)$$

$$r_2 = \sigma_w^2 (h_0 h_2) \quad (13.4)$$

Recuérdese que $r_m = r_{xx}(m)$. De nuevo, y tal y como se hiciera para los sistemas de orden 1, se supondrá que el proceso de entrada tiene varianza unidad (o que la secuencia de autocorrelación está normalizada por la varianza de la entrada). El problema es calcular los coeficientes del filtro una vez conocida la secuencia de autocorrelación.

Los coeficientes se pueden calcular como sigue. Partiendo de (13.4) es directo de ver que

$$\hat{h}_2 = \frac{r_2}{\hat{h}_0} \quad (13.5)$$

que junto con (13.3) conduce a

$$\hat{h}_1 = \frac{r_1}{\hat{h}_0 + r_2/\hat{h}_0} \quad (13.6)$$

Como se aprecia, para calcular \hat{h}_1 y \hat{h}_2 se necesita \hat{h}_0 . Éste se puede calcular usando (13.5) y (13.6) en (13.2) para dar

$$r_0 = \hat{h}_0^2 + \frac{r_1^2 \hat{h}_0^2}{(\hat{h}_0^2 + r_2)^2} + \frac{r_2^2}{\hat{h}_0^2} \quad (13.7)$$

Calculando h_0 a partir de (13.7) y sustituyendo su valor en (13.5) y (13.6) todo el sistema puede llegar a ser identificado. El problema es que discutir las posibles soluciones utilizando (13.5)-(13.7) no es nada sencillo ni directo, de manera que sería más satisfactoria otra expresión para los coeficientes del filtro. Para tal fin se procede de la siguiente manera. Considérese el cuadrado de (13.3)

$$r_1^2 = h_0^2 h_1^2 + h_1^2 h_2^2 + 2h_1^2 h_0 h_2 = h_1^2 (r_0 - h_1^2 + 2r_2) \quad (13.8)$$

donde se han empleado (13.2) y (13.4). Despejando para h_1 da

$$\hat{h}_1^2 = r_2 + r_0/2 \pm \sqrt{(r_2 + r_0/2)^2 - r_1^2} \tag{13.9}$$

una vez que \hat{h}_1 es conocida, la expresión (13.6) se puede usar para calcular \hat{h}_0 , resultando

$$\hat{h}_0 = \frac{r_1 \pm \sqrt{r_1^2 - 4\hat{h}_1^2 r_2}}{2\hat{h}_1} \tag{13.10}$$

y finalmente, (13.5) se puede emplear para obtener \hat{h}_2 a partir del valor de \hat{h}_0 .

13.1.1 Discusión de la solución

Para proceder al análisis de la solución obtenida para un sistema de segundo orden, se utilizará el conjunto de expresiones derivado en segundo lugar (13.9)-(13.10) junto con (13.5). De la ecuación (13.9) se pueden obtener 2 valores diferentes de h_1^2 que generarán, debido a la raíz cuadrada, dos valores cada uno, diferenciando en el signo. Estos cuatro valores se notarán por h_{11} , h_{12} , $-h_{11}$ y $-h_{12}$ correspondiendo a la elección de + o - en (13.9) y a la elección de + o - una vez que se toma la raíz cuadrada para calcular h_1 a partir de h_1^2 .

De ellos se pueden calcular 8 valores diferentes para h_0 usando (13.10), dos por cada valor previo de h_1 . Pero debido a la estructura de (13.10), 4 son los opuestos de los otros 4 tal y como se muestra en la tabla 13.1.

Valores de h_1	Valores obtenidos para h_0
h_{11}	h_{01} y h_{02}
h_{12}	h_{03} y h_{04}
$-h_{11}$	$-h_{01}$ y $-h_{02}$
$-h_{12}$	$-h_{03}$ y $-h_{04}$

Tabla 13.1 Valores de h_0 obtenidos de los de h_1

Análogamente, de cada valor de h_0 se obtiene, usando (13.5), un valor para h_2 donde de nuevo la mitad de ellos son los opuestos de la otra mitad. De esta discusión se puede deducir que las soluciones al sistema de ecuaciones (13.2)-(13.4) son 8 canales diferentes, donde la mitad de ellos iguala a la otra mitad salvo en el signo global. Tal y como se comentó en el capítulo anterior, si la entrada presenta distribución de probabilidad simétrica no hay manera de distinguir entre un sistema dado y su opuesto, así que sólo las cuatro primeras

soluciones (las ‘positivas’) se considerarán de aquí en adelante. Debido a la posibilidad que presenta cada coeficiente calculado a partir de la secuencia de autocorrelación de tomar más de un valor, se ha incluido el ‘sombrero’ en las cantidades anteriores. En la figura 13.1 se esquematiza lo comentado en este párrafo.



Figura 13.1 *Diversidad de la solución. Caminos posibles para construir un sistema*

13.1.2 Características de la solución

Como se comentó previamente todas las magnitudes involucradas se suponen reales. En este apartado se intentará encontrar expresiones analíticas explícitas para las soluciones en función de los coeficientes del filtro, nótese que ahora se conocen en función de la secuencia de autocorrelación. El estudio que se va a llevar a cabo será de vital importancia en capítulos posteriores para proponer un nuevo método para la identificación ciega de sistemas.

Posibles valores para \hat{h}_1 Usando las relaciones entre la secuencia de autocorrelación y los coeficientes del filtro dados en (13.2)- (13.4), la expresión (13.9) se puede escribir como

$$\hat{h}_1^2 = \frac{1}{2} \left[(h_0 + h_1)^2 + h_1^2 \pm \sqrt{((h_0 + h_2)^2 - h_1^2)^2} \right] \quad (13.11)$$

de la cual se puede obtener bastante información. Dos puntos merecen ser destacados:

1. Supóngase que para el sistema en estudio se verifica que $(h_0 + h_2)^2 \sim h_1^2$, es decir, que el radicando que aparece en la expresión anterior sea muy cercano a cero. Cuando las magnitudes tienen que ser estimadas, es posible que este radicando tome un valor negativo dando lugar a valores complejos para \hat{h}_1 cuando se esperan reales. Esta situación es mejor evitarla.
2. Si se realizan las operaciones indicadas en (13.11) se obtienen dos solucio-

nes

$$\hat{h}_1^2 = h_1^2 \tag{13.12}$$

$$\hat{h}_1^2 = (h_0 + h_2)^2 \tag{13.13}$$

su significado y el porqué de su valor se analizará más adelante.

Posibles valores para \hat{h}_0 Usando de nuevo las relaciones dadas en (13.2)-(13.4), el estimador del coeficiente h_0 se puede escribir como sigue a partir de (13.10)

$$\hat{h}_0 = \frac{h_0h_1 + h_1h_2 \pm \sqrt{(h_0h_1 - h_1h_2)^2 - 4\hat{h}_1^2h_0h_2}}{2\hat{h}_1} \tag{13.14}$$

Como se aprecia, en esta última expresión, aparece \hat{h}_1 que puede tomar dos valores distintos dados en (13.12) y (13.13). Para poder evaluarla totalmente hay que sustituir ambos.

Si se sustituye por el primero de ellos resulta

$$\hat{h}_0 = \frac{h_0h_1 + h_1h_2 \pm \sqrt{(h_0 - h_2)^2h_1^2}}{2h_1} \tag{13.15}$$

donde de nuevo se pueden resaltar dos puntos:

1. De nuevo hay más opciones de obtener un valor complejo si el radicando es cercano a cero. Esto ocurre si $h_0 \sim h_2$. Hay que hacer notar que si esta situación se da hay un efecto más nocivo que el simple hecho de obtener valores complejos. Este efecto consiste en la eliminación de opciones posibles para el candidato que se esté estudiando, en este caso \hat{h}_0 , ya que si el radicando es casi nulo, las dos soluciones posibles tienden a degenerar en una sola y es mucho más difícil distinguirlas.
2. Los valores posibles de \hat{h}_0 para este valor de \hat{h}_1 son

$$\hat{h}_{01} = h_0 \tag{13.16}$$

$$\hat{h}_{02} = h_2 \tag{13.17}$$

Si se sustituye por el segundo resulta

$$\hat{h}_0 = \frac{h_1 \pm \sqrt{h_1^2 - h_0h_2}}{2} \tag{13.18}$$

En esta ocasión los puntos a resaltar son los siguientes:

1. La condición sobre los coeficientes del filtro que hacen más probable un valor complejo o una degeneración en las soluciones es $h_1 = 2\sqrt{h_0 h_2}$
2. Los nuevos valores para \hat{h}_0 son

$$\hat{h}_{03} = \frac{h_1 + \sqrt{h_1^2 - h_0 h_2}}{2} \quad (13.19)$$

$$\hat{h}_{04} = \frac{h_1 - \sqrt{h_1^2 - h_0 h_2}}{2} \quad (13.20)$$

Posibles valores para \hat{h}_2 Usando la expresión (13.5) se pueden obtener los cuatro valores para \hat{h}_2 a partir de los cuatro valores de \hat{h}_0 , estos son respectivamente:

$$\hat{h}_{21} = h_2 \quad (13.21)$$

$$\hat{h}_{22} = h_0 \quad (13.22)$$

$$\hat{h}_{23} = \frac{h_1 - \sqrt{h_1^2 - h_0 h_2}}{2} \quad (13.23)$$

$$\hat{h}_{24} = \frac{h_1 + \sqrt{h_1^2 - h_0 h_2}}{2} \quad (13.24)$$

Para resumir este apartado, en la tabla 13.2 se recogen los cuatro sistemas espectralmente equivalentes de uno dado, definido por $\mathbf{h} = [h_0 \ h_1 \ h_2]$.

	\hat{h}_0	\hat{h}_1	\hat{h}_2
1ª Opción	h_0	h_1	h_2
2ª Opción	h_2	h_1	h_0
3ª Opción	$\frac{h_1 - \sqrt{h_1^2 - h_0 h_2}}{2}$	$h_0 + h_2$	$\frac{h_1 + \sqrt{h_1^2 - h_0 h_2}}{2}$
4ª Opción	$\frac{h_1 + \sqrt{h_1^2 - h_0 h_2}}{2}$	$h_0 + h_2$	$\frac{h_1 - \sqrt{h_1^2 - h_0 h_2}}{2}$

Tabla 13.2 *Sistemas espectralmente equivalentes de un sistema de segundo orden*

13.2 Sistemas espectralmente equivalentes

En este apartado se van a calcular expresiones analíticas explícitas de los coeficientes de los sistemas espectralmente equivalentes de uno dado. Se comprobará que coinciden con las expresiones expuestas en la tabla 13.2. Como es sabido, especificado un sistema lineal se puede construir otro espectralmente

equivalente reflejando uno de sus ceros respecto al círculo unidad. Suponiendo que los ceros son a y b , la función de transferencia del sistema se puede escribir como

$$H(z) = (1 - az^{-1})(1 - bz^{-1}) = 1 - (a + b)z^{-1} + abz^{-2} \quad (13.25)$$

o en un caso más general, donde no esté normalizado su primer coeficiente,

$$H(z) = h_0 - h_0(a + b)z^{-1} + h_0abz^{-2} \equiv h_0 + h_1z^{-1} + h_2z^{-2} \quad (13.26)$$

Obviamente, los ceros en función de los coeficientes del sistema son

$$a = \frac{-h_1 + \sqrt{h_1^2 - 4h_0h_2}}{2h_0} \quad (13.27)$$

$$b = \frac{-h_1 - \sqrt{h_1^2 - 4h_0h_2}}{2h_0} \quad (13.28)$$

nótese que el papel de a y b es intercambiable.

Invirtiendo el primer cero a , la ecuación (13.26) conduce al siguiente valor de h_1 :

$$h_1' = -h_0 \left(\frac{2h_0}{-h_1 + \sqrt{h_1^2 - 4h_0h_2}} + \frac{-h_1 - \sqrt{h_1^2 - 4h_0h_2}}{2h_0} \right) \quad (13.29)$$

realizando la suma indicada el resultado es

$$h_1' = \frac{-2h_0(h_0 + h_2)}{-h_1 + \sqrt{h_1^2 - 4h_0h_2}} \quad (13.30)$$

y para el coeficiente h_2

$$h_2' = h_0 \left(\frac{2h_0}{-h_1 + \sqrt{h_1^2 - 4h_0h_2}} \cdot \frac{-h_1 - \sqrt{h_1^2 - 4h_0h_2}}{2h_0} \right) \quad (13.31)$$

que tras simplificar queda como

$$h_2' = h_0 \frac{-h_1 - \sqrt{h_1^2 - 4h_0h_2}}{-h_1 + \sqrt{h_1^2 - 4h_0h_2}} \quad (13.32)$$

Al invertir uno de los ceros se ha obtenido un sistema espectralmente equivalente. Se pueden obtener más sistemas de este tipo si se invierte el otro cero o si se invierten ambos a la vez. El resultado de todos estos procesos se muestra en la tabla 13.3.

Se refleja...	1 ^{er} Coeficiente	2 ^o Coeficiente	3 ^{er} coeficiente
ningún cero	h_0	h_1	h_2

1 ^{er} cero	h_0	$\frac{-2h_0(h_0+h_2)}{-h_1+\sqrt{h_1^2-4h_0h_2}}$	$h_0 \frac{-h_1-\sqrt{h_1^2-4h_0h_2}}{-h_1+\sqrt{h_1^2-4h_0h_2}}$
2 ^o cero	h_0	$\frac{-2h_0(h_0+h_2)}{-h_1-\sqrt{h_1^2-4h_0h_2}}$	$h_0 \frac{-h_1+\sqrt{h_1^2-4h_0h_2}}{-h_1-\sqrt{h_1^2-4h_0h_2}}$
ambos ceros	h_0	$h_0 \frac{h_1}{h_2}$	$h_0 \frac{h_0}{h_2}$

Tabla 13.3 *Sistemas resultantes en el proceso de inversión de ceros*

En una primera inspección las tablas 13.2 y 13.3 parecen tener muy poco en común, cuando se espera que sean idénticas. Es cierto que tomado un sistema en la tabla 13.2 otro sistema con los mismos ceros se puede encontrar en la tabla 13.3, la única diferencia es pues un problema de normalización: todos los sistemas en ambas tablas deberían tener la misma ganancia. Los sistemas de la tabla 13.2 ya cumplen esta condición porque satisfacen la ecuación (13.2), pero no es el caso para los de la tabla 13.3.

Sea n el factor de normalización requerido y considérese el último sistema de esta última tabla, resultado de invertir ambos ceros. Si ha de tener la misma ganancia que el sistema original es necesario que se verifique:

$$n^2 h_0^2 + n^2 h_0^2 \frac{h_1^2}{h_2^2} + n^2 h_0^2 \frac{h_0^2}{h_2^2} = h_0^2 + h_1^2 + h_2^2 \quad (13.33)$$

que se puede conseguir fijando $n = h_2/h_0$. Aplicando este factor a este sistema se obtiene que el sistema equivalente es $\mathbf{h} = [h_2 \ h_1 \ h_0]$, que es el mismo que el dado en la opción segunda de la tabla 13.2.

De forma análoga se pueden encontrar factores de normalización para el resto de sistemas, resultando

$$n_2 = \frac{h_1 - \sqrt{h_1^2 - 4h_0h_2}}{2h_0} \quad (13.34)$$

$$n_3 = \frac{h_1 + \sqrt{h_1^2 - 4h_0h_2}}{2h_0} \quad (13.35)$$

para el segundo y tercer sistema de la tabla 13.3 respectivamente. Después de esta normalización todos los sistemas de ambas tablas coinciden como era de esperar. Esto significa que las soluciones al sistema de ecuaciones (13.2)-(13.4) están relacionadas mediante la reflexión de ceros respecto al círculo unidad, como es conocido de la teoría de filtros.

13.3 Sensibilidad de las soluciones

De manera análoga al estudio que se llevó a cabo para sistemas de primer orden, en esta sección se analizarán los efectos que pequeños errores en la estimación de la secuencia de autocorrelación pueden tener sobre los estimadores de los coeficientes del filtro calculados mediante (13.9), (13.10) y (13.5).

13.3.1 Sensibilidad del coeficiente h_1

Como es sabido, lo que interesa calcular ahora son las derivadas parciales de los coeficientes del filtro en función de la secuencia de autocorrelación. Como en las expresiones que se han adoptado como solución se debe calcular h_1 en primera instancia, es por él por el que se empezará el análisis.

Variaciones respecto a r_0

Por conveniencia se reproduce en la siguiente línea la ecuación de partida

$$\hat{h}_1^2 = r_2 + r_0/2 \pm \sqrt{(r_2 + r_0/2)^2 - r_1^2} \quad (13.36)$$

cuya derivada parcial con respecto al coeficiente r_0 es

$$\frac{\partial \hat{h}_1}{\partial r_0} = \pm \frac{1}{4} \frac{\left(\sqrt{(r_2 + r_0/2)^2 - r_1^2} \pm (r_2 + r_0/2) \right)^{1/2}}{\sqrt{(r_2 + r_0/2)^2 - r_1^2}} \quad (13.37)$$

Esta derivada puede tomar un valor elevado o diverger en una única situación:

Cuando $(r_2 + r_0/2)^2 \sim r_1^2$, que trasladado a una condición sobre los coeficientes del filtro significa que $(h_0 + h_2)^2 \sim h_1^2$. Bajo esta condición la ecuación (13.37) toma la siguiente forma:

$$\frac{\partial \hat{h}_1}{\partial r_0} \sim \frac{(\epsilon \pm (r_2 + r_0/2))^{1/2}}{\epsilon} \quad (13.38)$$

con $\epsilon \sim 0$. Como se puede apreciar esta divergencia es inevitable.

Variaciones con respecto a r_1

La derivada parcial de interés en este apartado es

$$\frac{\partial \hat{h}_1}{\partial r_1} = \mp \frac{1}{2} \frac{r_1}{\sqrt{(r_2 + r_0/2)^2 - r_1^2} \left(r_2 + r_0/2 \pm \sqrt{(r_2 + r_0/2)^2 - r_1^2} \right)^{1/2}} \quad (13.39)$$

Y las situaciones asociadas a ella en las que puede tomar valores elevados son

Caso a) Como el caso a anterior $(r_2 + r_0/2)^2 \sim r_1^2$, que trasladado a los coeficientes del sistema implica que $(h_0 + h_2)^2 \sim h_1^2$. Bajo esta condición la ecuación anterior se convierte en

$$\frac{\partial \hat{h}_1}{\partial r_1} = \mp \frac{1}{2} \frac{r_1}{\epsilon (r_2 + r_0/2 \pm \epsilon)^{1/2}} \quad (13.40)$$

Los efectos de tener un factor muy pequeño en el denominador se pueden contrarrestar imponiendo $r_1 \sim 0$, lo que cual significa que $h_1 \sim 0$ o $h_0 + h_2 \sim 0$. Pero esta nueva condición tiene que ser consistente con la anterior así que h_1 y $h_0 + h_2$ tienen que estar próximas a cero a la vez. Con esta condición total, el factor $r_2 + r_0/2 = 1/2[(h_0 + h_2)^2 + h_1^2]$ en el denominador es también cero con lo que la divergencia no se puede evitar.

Caso b) De igual manera es posible que (13.39) diverja si el segundo factor que aparece en el denominador se acerca a cero, es decir, cuando ocurre que $r_2 + r_0/2 \pm \sqrt{(r_2 + r_0/2)^2 - r_1^2} \sim 0$. Teniendo en cuenta la expresión (13.36), quiere decir que $\hat{h}_1 \sim 0$, que se puede cumplir bien si $h_1 \sim 0$ o si $h_0 + h_2 \sim 0$, en definitiva, si $r_1 \sim 0$. Si r_1 es muy pequeño la expresión (13.39) se puede aproximar por

$$\frac{\partial \hat{h}_1}{\partial r_1} = \pm \frac{1}{2} \frac{r_1}{(r_2 + r_0/2 - r_1^2/(r_2 + r_0/2))} \cdot \frac{1}{(r_2 + r_0/2 \pm (r_2 + r_0/2 - r_1^2/(r_2 + r_0/2)))^{1/2}} \quad (13.41)$$

Dependiendo del signo elegido en el denominador pueden surgir dos casos diferentes:

Caso b.1) Tomando la primera opción en los signos queda

$$\frac{\partial \hat{h}_1}{\partial r_1} = + \frac{1}{2} \frac{r_1}{(r_2 + r_0/2 - r_1^2/(r_2 + r_0/2)) (2r_2 + r_0 - r_1^2/(r_2 + r_0/2))^{1/2}} \quad (13.42)$$

que no diverge salvo que $r_2 + r_0/2 = ((h_0 + h_2)^2 + h_1^2)/2 \sim 0$, que implica que las dos condiciones relacionadas con este caso b tienen que darse a la vez $h_1 \sim 0$ y $h_0 + h_2 \sim 0$. Como se aprecia esto es un situación particular del caso a, que como se sabe provoca divergencia ineludible.

Caso b.2) Tomando la segunda opción en los signos:

$$\frac{\partial \hat{h}_1}{\partial r_1} = \frac{1}{2} \frac{r_1}{(r_2 + r_0/2 - r_1^2/(r_2 + r_0/2)) (r_1^2/(r_2 + r_0/2))^{1/2}} \quad (13.43)$$

y simplificando un poco queda

$$\frac{\partial \hat{h}_1}{\partial r_1} = \frac{1}{2} \frac{\sqrt{r_2 + r_0/2}}{(r_2 + r_0/2 - r_1^2/(r_2 + r_0/2))} \quad (13.44)$$

Recuérdese que r_1 es muy pequeño aquí. La única situación que puede provocar la divergencia de esta expresión es la misma que para el subcaso b.1 y de hecho provoca una divergencia ineludible como es conocido. Nótese que el caso b es una situación particular contenida en el caso a, por lo que no se añade información nueva al análisis.

Variación con respecto a r_2

La derivada parcial de \hat{h}_1 con respecto a r_2 es

$$\frac{\partial \hat{h}_1}{\partial r_2} = \pm \frac{1}{2} \frac{\left(\sqrt{(r_2 + r_0/2)^2 - r_1^2} \pm (r_2 + r_0/2) \right)^{1/2}}{\sqrt{(r_2 + r_0/2)^2 - r_1^2}} \quad (13.45)$$

que es idéntica, salvo un factor $1/2$, a $\frac{\partial \hat{h}_1}{\partial r_0}$ de forma que la discusión consiguiente es idéntica y no se repetirá aquí.

13.3.2 Sensibilidad del coeficiente h_0

La expresión relacionando el candidato \hat{h}_0 para el coeficiente h_0 , con la secuencia de autocorrelación y el candidato \hat{h}_1 , se reescribe aquí por conveniencia:

$$\hat{h}_0 = \frac{r_1 \pm \sqrt{r_1^2 - 4\hat{h}_1^2 r_2}}{2\hat{h}_1} \quad (13.46)$$

En los siguientes apartados se analizan las derivadas parciales de esta expresión respecto de las variables de las que depende.

Variación con respecto a r_0

La derivada parcial de \hat{h}_0 con respecto a r_0 es

$$\frac{\partial \hat{h}_0}{\partial r_0} = - \left[\frac{2(\partial \hat{h}_1^2 / \partial r_0) r_2}{\sqrt{r_1^2 - 4\hat{h}_1^2 r_2}} \hat{h}_1 + \frac{\partial \hat{h}_1}{\partial r_0} (r_1 \pm \sqrt{r_1^2 - 4\hat{h}_1^2 r_2}) \right] \frac{1}{2\hat{h}_1^2} \quad (13.47)$$

Las posibles situaciones en las que esta expresión puede diverger son:

Caso a) $r_1^2 - 4\hat{h}_1^2 r_2 \sim 0$, de manera que el primer sumando del corchete toma un valor muy grande. Según el valor que tome el candidato \hat{h}_1 la expresión anterior se puede escribir de dos maneras en función de los coeficientes del filtro:

Caso a.1) Fijando $\hat{h}_1 = h_1$ la expresión anterior se convierte en $h_1^2(h_0 + h_2)^2 - 4h_1^2 h_0 h_2 \sim 0$ o equivalentemente $h_1^2(h_0 - h_2)^2 \sim 0$. Esto se puede conseguir con $h_1 \sim 0$ o $h_0 \sim h_2$. El siguiente paso en el análisis es comprobar si esta situación de divergencia se puede contrarrestar con condiciones adicionales sobre los coeficientes del filtro. Bajo las condiciones enunciadas hasta ahora la ecuación (13.47) se convierte en

$$\frac{\partial \hat{h}_0}{\partial r_0} = - \left[\frac{2(\partial \hat{h}_1^2 / \partial r_0) r_2}{\epsilon} h_1 + \frac{\partial \hat{h}_1}{\partial r_0} (r_1 \pm \epsilon) \right] \frac{1}{2h_1^2} \quad (13.48)$$

Como lo que interesa es evitar la divergencia es suficiente estudiar el primer sumando de esta expresión:

$$\frac{\partial \hat{h}_0}{\partial r_0} \sim - \frac{(\partial \hat{h}_1^2 / \partial r_0) r_2}{\epsilon} \frac{1}{h_1} \quad (13.49)$$

Si $r_2 \sim 0$ ambos ceros se podrían anular. Para hacer $r_2 \sim 0$ bien h_0 o bien h_2 tienen que estar cercanos a cero. Ahora bien, si se ha llegado a (13.49) porque $h_1 \sim 0$ el h_1 del denominador refuerza la divergencia. Por otro lado, si se ha llegado a (13.49) porque $h_0 \sim h_2$, imponiendo aún más que h_0 o h_2 sea zero lleva a que h_0 y h_2 estén ambos cercanos a cero, dejando el sistema con un sólo parámetro (un sistema de orden nulo). Por tanto, salvo para el caso extremo patológico comentado, en el caso a.1 la divergencia no se puede evitar.

Caso a.2) Si el candidato se elige como $\hat{h}_1 = h_0 + h_2$, la condición establecida en el caso a se convierte en $h_1^2(h_0 + h_2)^2 - 4(h_0 + h_2)^2 h_0 h_2 \sim 0$ o equivalentemente $(h_1^2 - 4h_0 h_2)(h_0 + h_2)^2 \sim 0$. Esto se puede conseguir bien

imponiendo que $h_0 \sim -h_2$ o bien que $h_1 = 2\sqrt{h_0 h_2}$. Sea como fuere, bajo esta condición la ecuación (13.47) se convierte en una similar a (13.48)

$$\frac{\partial \hat{h}_0}{\partial r_0} = - \left[\frac{2(\partial \hat{h}_1^2 / \partial r_0) r_2}{\epsilon} (h_0 + h_2) + \frac{\partial \hat{h}_1}{\partial r_0} (r_1 \pm \epsilon) \right] \frac{1}{2(h_0 + h_2)^2} \quad (13.50)$$

y de nuevo la única forma de evitar la divergencia es haciendo finito el primer factor:

$$\frac{\partial \hat{h}_0}{\partial r_0} \sim - \frac{(\partial \hat{h}_1^2 / \partial r_0) r_2}{\epsilon} \frac{1}{h_0 + h_2} \quad (13.51)$$

El cero en el denominador se podría evitar haciendo $r_2 \sim 0$, pero en este caso significaría que h_0 o h_2 son cero, llevando a un sistema de orden 1 carente de interés aquí. En definitiva, para el caso a.2 también hay divergencia.

Caso b) La ecuación (13.47) puede diverger también si $\partial \hat{h}_1^2 / \partial r_0$ diverge, esto es, cuando $\hat{h}_1^2 \sim (h_0 + h_2)^2$. Para que esto no ocurra se podría intentar fijar $r_2 \sim 0$ (relativo al primer factor) y $r_1 \sim \pm \sqrt{r_1^2 - 4\hat{h}_1^2 r_2}$ (relativo al segundo factor). Nótese que no se puede hacer $\hat{h}_1 \sim 0$ para evitar la divergencia del primer factor porque el término común reforzaría la divergencia. La primera de estas últimas condiciones impone que $h_0 \sim 0$ o $h_2 \sim 0$ y la segunda impone que $\hat{h}_1^2 r_2 \sim 0$, que a su vez se divide en $r_2 \sim 0$, como la condición anterior, y $\hat{h}_1 \sim 0$. Como se ha comentado no se puede fijar $\hat{h}_1 \sim 0$, porque lleva a divergencia igualmente, de forma que hay que hacer $h_0 \sim 0$ o $h_2 \sim 0$ lo que conduciría a un sistema de primer orden, no de segundo, y se debería aplicar el análisis que para ese tipo de sistemas se desarrolló en el capítulo anterior. Por tanto, en este caso b la divergencia es inevitable, como es lógico, ya que si \hat{h}_1 se emplea para calcular \hat{h}_0 , cuando \hat{h}_1 esté mal estimado también lo estará \hat{h}_0 .

Variaciones con respecto a r_1

La derivada parcial de \hat{h}_0 con respecto a r_1 es

$$\begin{aligned} \frac{\partial \hat{h}_0}{\partial r_1} = & \frac{1}{2\hat{h}_1} \left(1 \pm \frac{r_1 - 2(\partial \hat{h}_1^2 / \partial r_1) r_2}{\sqrt{r_1^2 - 4\hat{h}_1^2 r_2}} \right) - \\ & - \frac{\partial \hat{h}_1}{\partial r_1} \left(r_1 \pm \sqrt{r_1^2 - 4\hat{h}_1^2 r_2} \right) \frac{1}{2\hat{h}_1^2} \end{aligned} \quad (13.52)$$

Las condiciones bajo las cuales esta expresión puede tomar valores elevados son:

Caso a) $r_1^2 - 4\hat{h}_1^2 r_2 \sim 0$, como en el caso a anterior, que hace que esta expresión se convierta en

$$\frac{\partial \hat{h}_0}{\partial r_1} = \left[\left(1 \pm \frac{r_1 - 2(\partial \hat{h}_1^2 / \partial r_1) r_2}{\epsilon} \right) \hat{h}_1 - \frac{\partial \hat{h}_1}{\partial r_1} (r_1 \pm \epsilon) \right] \frac{1}{2\hat{h}_1^2} \quad (13.53)$$

Como se puede observar la divergencia ocurre en el primer sumando y es él el que hay que estudiar:

$$\frac{\partial \hat{h}_0}{\partial r_1} \sim \frac{r_1 - 2(\partial \hat{h}_1^2 / \partial r_1) r_2}{\epsilon} \frac{1}{2\hat{h}_1} \quad (13.54)$$

Si se quiere evitar la divergencia hay que fijar a la vez dos valores:

1. $r_1 \sim 0$, es decir, que $h_1 \sim 0$ o que $h_0 + h_2 \sim 0$
2. $r_2 \sim 0$, es decir, que $h_0 \sim 0$ o que $h_2 \sim 0$.

Para poder cumplir el segundo punto es necesario que el sistema, que se suponía de orden 2, sea de orden 1, por lo que se convierte desde el punto de vista de sistemas de segundo orden en una situación trivial. Por tanto la divergencia en este caso a no se puede evitar. Por último precisar que a esta situación se puede llegar por dos vías:

Caso a.1) $h_1 \sim 0$ o $h_0 \sim h_2$.

Caso a.2) $h_0 \sim -h_2$ o $h_1 = 2\sqrt{h_0 h_2}$.

Caso b) Que diverja $\partial \hat{h}_1^2 / \partial r_1$ o $\partial \hat{h}_1 / \partial r_1$, es decir, que $h_1^2 \sim (h_0 + h_2)^2$, y es completamente análogo al caso b anterior con una discusión idéntica: la única forma de evitar la divergencia de (13.52) bajo esta condición es fijar $r_2 \sim 0$ para el primer factor y $\hat{h}_1^2 r_2 \sim 0$ para el segundo factor. Por tanto, para contener la divergencia del primer factor es necesario reducir el sistema a uno de primer orden, situación trivial.

Variación con respecto a r_2

La derivada parcial de \hat{h}_0 con respecto a r_2 es

$$\frac{\partial \hat{h}_0}{\partial r_2} = \left[\mp \frac{r_2 \partial \hat{h}_1^2 / \partial r_2 + \hat{h}_1^2}{\sqrt{r_1^2 - 4\hat{h}_1^2 r_2}} \hat{h}_1 - \frac{\partial \hat{h}_1}{\partial r_2} \left(\frac{r_1 \pm \sqrt{r_1^2 - 4\hat{h}_1^2 r_2}}{2} \right) \right] \frac{1}{\hat{h}_1^2} \quad (13.55)$$

Las posibles causas de divergencia de esta expresión son:

Caso a) $r_1^2 - 4\hat{h}_1^2 r_2 \sim 0$ como ya es conocido de los casos a anteriores. Bajo esta condición la ecuación (13.55) se convierte en:

$$\frac{\partial \hat{h}_0}{\partial r_2} = \left[\mp \frac{r_2 \partial \hat{h}_1^2 / \partial r_2 + \hat{h}_1^2}{\epsilon} \hat{h}_1 - \frac{\partial \hat{h}_1}{\partial r_2} \left(\frac{r_1 \pm \epsilon}{2} \right) \right] \frac{1}{\hat{h}_1^2} \quad (13.56)$$

E igualmente el término dominante en este límite es el primero:

$$\frac{\partial \hat{h}_0}{\partial r_2} \sim \mp \frac{r_2 \partial \hat{h}_1^2 / \partial r_2 + \hat{h}_1^2}{\epsilon} \frac{1}{\hat{h}_1} \quad (13.57)$$

Para contrarrestar el cero en el denominador, r_2 y \hat{h}_1^2 tienen que estar próximos a cero. Pero nada de esto se puede conseguir porque el primero haría que el sistema fuese de primer orden y el segundo reforzaría el cero en el denominador (ver el factor común). De manera que la divergencia es ineludible. Recuérdese que a esta situación se puede llegar por dos caminos:

Caso a.1) $h_1 \sim 0$ o $h_0 \sim h_2$.

Caso a.2) $h_0 \sim -h_2$ o $h_1 = 2\sqrt{h_0 h_2}$.

Caso b) Si divergen $\partial \hat{h}_1^2 / \partial r_2$ o $\partial \hat{h}_1 / \partial r_2$. Para evitarlo, entre otras acciones habría que hacer $r_2 \sim 0$ y se obtendría de nuevo un sistema de primer orden. Por tanto, en esta situación tenemos divergencia.

Caso c) Si $\hat{h}_1 \sim 0$. Bajo esta condición la ecuación (13.55) queda de la siguiente forma:

$$\frac{\partial \hat{h}_0}{\partial r_2} = \left[\mp \frac{r_2 \partial \hat{h}_1^2 / \partial r_2 + \epsilon^2}{\epsilon} \epsilon - \frac{\partial \hat{h}_1}{\partial r_2} \left(\frac{\epsilon \pm \epsilon}{2} \right) \right] \frac{1}{\epsilon^2} \quad (13.58)$$

donde se ha tenido en cuenta que si \hat{h}_1 es un infinitésimo, r_1 es un infinitésimo del mismo orden. El interés se centra en evitar la divergencia si es posible. Considérese por separado el primer sumando:

$$\frac{\partial \hat{h}_0}{\partial r_2} \sim \mp \frac{r_2 \partial \hat{h}_1^2 / \partial r_2}{\epsilon^2} \quad (13.59)$$

Como es sabido no se puede contrarrestar el cero del denominador haciendo $r_2 \sim 0$ porque lleva a un sistema de primer orden. Y en este límite de $\hat{h}_1 \sim 0$ la parcial que aparece no se hace nula (ver ecuación (13.37)) con lo que la divergencia es inevitable.

13.3.3 Sensibilidad del coeficiente h_2

Para concluir el análisis que se está llevando a cabo, de la sensibilidad de los estimadores de los coeficientes del filtro a errores en la secuencia de autocorrelación, se estudiará el coeficiente h_2 . Este coeficiente es el último que se calcula tal y como se ha propuesto la solución del problema. Recuérdese que la expresión de interés en este apartado es:

$$\hat{h}_2 = \frac{r_2}{\hat{h}_0} \quad (13.60)$$

Variación con respecto a r_0

Teniendo en cuenta la expresión anterior se deduce que la derivada parcial de \hat{h}_2 con respecto a r_0 es:

$$\frac{\partial \hat{h}_2}{\partial r_0} = -r_2 \frac{\partial \hat{h}_0 / \partial r_0}{\hat{h}_0^2} \quad (13.61)$$

Las posibles situaciones en las que puede aparecer divergencia son:

Caso a) Que \hat{h}_0 esté cercano a cero. Se puede conseguir bien haciendo h_0 o h_2 cercanos a cero; de cualquier manera se obtiene un sistema de primer orden y es el análisis de dichos sistemas el que hay que considerar.

Caso b) Que $\partial \hat{h}_0 / \partial r_0$ diverja. En este caso sólo se podría evitar haciendo $r_2 \sim 0$ pero de nuevo se llegaría a un sistema de primer orden y por tanto la divergencia es inevitable.

Variación con respecto a r_1

Se obtiene una expresión análoga a la última:

$$\frac{\partial \hat{h}_2}{\partial r_1} = -r_2 \frac{\partial \hat{h}_0 / \partial r_1}{\hat{h}_0^2} \quad (13.62)$$

y por tanto aparecen los mismos casos para discutir con las mismas conclusiones.

Variación con respecto a r_2

En este caso la derivada parcial de interés es

$$\frac{\partial \hat{h}_2}{\partial r_0} = \frac{\hat{h}_0 - r_2 \partial \hat{h}_0 / \partial r_0}{\hat{h}_0^2} \tag{13.63}$$

y las causas de divergencia asociadas son:

Caso a) Si \hat{h}_0 tiende a cero. La divergencia es claramente inevitable en este caso.

Caso b) Si $\partial \hat{h}_0 / \partial r_2$ diverge, toda la expresión diverge.

En la tabla 13.4 se recogen todas las situaciones de alta sensibilidad estudiadas en los casos anteriores. Por sensibilidad se entiende que los errores de estimación en la secuencia de autocorrelación se pueden magnificar al calcular los coeficientes de los filtros.

∂/∂	h_0	h_1	h_2
r_0	$h_1 \sim 0$ $h_1 \sim 2\sqrt{h_0 h_2}$ $h_2 \sim \pm h_0$ $(h_0 + h_2)^2 \sim h_1^2$	$(h_0 + h_2)^2 \sim h_1^2$	$h_0 \sim 0$ $h_2 \sim 0$ $h_1 \sim 0$ $h_1 = 2\sqrt{h_0 h_2}$ $h_2 \sim \pm h_0$ $(h_0 + h_2)^2 \sim h_1^2$
r_1	$h_1 \sim 0$ $h_1 \sim 2\sqrt{h_0 h_2}$ $h_2 \sim \pm h_0$ $(h_0 + h_2)^2 \sim h_1^2$	$(h_0 + h_2)^2 \sim h_1^2$	$h_0 \sim 0$ $h_2 \sim 0$ $h_1 \sim 0$ $h_1 = 2\sqrt{h_0 h_2}$ $h_2 \sim \pm h_0$ $(h_0 + h_2)^2 \sim h_1^2$
r_2	$h_1 \sim 0$ $h_1 \sim 2\sqrt{h_0 h_2}$ $h_2 \sim \pm h_0$ $(h_0 + h_2)^2 \sim h_1^2$	$(h_0 + h_2)^2 \sim h_1^2$	$h_0 \sim 0$ $h_2 \sim 0$ $h_1 \sim 0$ $h_1 = 2\sqrt{h_0 h_2}$ $h_2 \sim \pm h_0$ $(h_0 + h_2)^2 \sim h_1^2$

Tabla 13.4 Resumen de las situaciones de alta sensibilidad

13.4 Interpretación de las situaciones de sensibilidad

La tabla 13.4 contiene todas las situaciones posibles en las que se puede dar divergencia para cualquiera de los valores de los candidatos. Esto significa que algunas situaciones afectan sólo cuando $\hat{h}_1 = h_1$ y otras cuando $\hat{h}_1 = h_0 + h_2$. Quizás sea una buena idea clasificar las situaciones de divergencia atendiendo a qué valor concreto tome cada candidato.

Nótese también que las condiciones bajo las que se produce divergencia en \hat{h}_0 contienen a aquellas para \hat{h}_1 y las condiciones de divergencia para \hat{h}_2 contienen a las de \hat{h}_0 .

13.4.1 Interpretación para \hat{h}_1

Un sistema está perfectamente caracterizado, salvo un factor de escala, por la localización de sus ceros. Puede ser instructivo traducir las condiciones anteriores sobre los coeficientes a condiciones sobre la localización de los ceros. Para ello recuérdese que si a y b son los ceros del sistema, sus coeficientes están relacionados con ellos de la siguiente manera:

$$h_0 = h_0 \quad (13.64)$$

$$h_1 = -h_0(a + b) \quad (13.65)$$

$$h_2 = h_0ab \quad (13.66)$$

donde h_0 hace el papel del factor de normalización. Teniendo en cuenta estas relaciones, la condición bajo la cual es posible una mala estimación de \hat{h}_1 se reescribe como:

$$(h_0 + h_2)^2 \sim h_1^2 \Rightarrow (1 + ab)^2 \sim (a + b)^2 \quad (13.67)$$

Tomando raíz cuadrada se obtienen dos condiciones sobre los ceros:

1. $1 + ab \sim a + b$ lo cual implica que $a \sim 1$ y b puede tomar cualquier valor.
2. O considerando que los miembros de (13.67) tienen distinto signo $1 + ab \sim -(a + b)$ que lleva a $b \sim -1$ con a tomando cualquier valor.

De manera que esta condición llevada a los ceros significa que ninguno de ellos debería estar cercano a ± 1 . Lo cual significa que si los ceros son reales no pueden estar cerca del círculo unidad. Si los ceros fuese complejos el único caso que se puede discutir aquí es cuando son una pareja complejo conjugada ya que sólo así los coeficientes del filtro serían reales. Recuérdese que las ecuaciones se han

13.4. INTERPRETACIÓN DE LAS SITUACIONES DE SENSIBILIDAD 337

derivado suponiendo todas las magnitudes reales. Para este caso la condición quedaría:

$$(1 + |a|^2) \sim (2\Re\{a\})^2 \Rightarrow (1 \mp a_r)^2 \sim -a_i^2 \quad (13.68)$$

Donde $\Re\{a\}$ indica la parte real de a , siendo $a = a_r + ia_i$ complejo. La única solución posible es haciendo $a_i = 0$ y $a_r = \pm 1$. Es decir, que el cero sea real con la misma restricción anterior. Esto parece indicar que si es complejo no tiene porqué estar lejos del círculo unidad, lo cual va un poco contra lo que se esperaba. Cuando un cero está cerca del círculo unidad más difícil es, al estimarlo, decidir si está dentro o fuera de él. Téngase en cuenta sin embargo la siguiente observación: un sistema de segundo orden cuyos ceros sean complejos conjugados no puede ser de fase mixta (o están los dos dentro o están fuera simplemente por el hecho de ser conjugados). Según esta observación debe ser más sencillo identificar los sistemas porque elimina posibilidades, al haber sólo dos candidatos, es por tanto menos sensible.

13.4.2 Interpretación para \hat{h}_0

Recuérdese que este estimador se calcula utilizando \hat{h}_1 y por tanto sus causas de divergencia dependen de si el valor de \hat{h}_1 es h_1 o su equivalente $h_0 + h_2$.

Caso 1) Si se ha estimado como h_1 . Hay dos posibles situaciones en las que aparece la divergencia:

1. $h_1 \sim 0$. De acuerdo con la ecuación (13.65) esto significa que los ceros son el opuesto uno de otro $a \sim -b$.
2. $h_0 \sim h_2$. De nuevo utilizando la ecuación (13.65) significa que los ceros son el inverso uno de otro $a \sim 1/b$.

Caso 2) Si se ha estimado como $h_0 + h_2$. También hay dos posibles condiciones bajo las cuales puede aparecer la divergencia:

1. $h_0 + h_2 \sim 0$, o equivalentemente $h_0 \sim -h_2$. Que impone que los ceros no deben ser inversos y opuestos $a \sim -1/b$.
2. $h_1^2 \sim 4h_0h_2$ que trasladado a los ceros implica que no deben ser iguales.

Los dos subcasos del caso 1 junto con los dos subcasos del caso 2 dan un total de 4 situaciones de sensibilidad que se corresponden con los cuatro valores que puede tomar \hat{h}_0 : por cada grado de libertad aparece una posible divergencia.

Por esta misma razón, no siempre que alguna de estas situaciones se cumple tiene que dar necesariamente malos estimadores; bajo una cierta condición se obtiene un mal estimador si esta condición se ha derivado para el caso particular en el que el candidato toma ese valor. Por ejemplo, si el valor real de \hat{h}_1 es h_1 , ninguno de los subcasos de 2 pueden afectar al resultado. Lo que es más, para este valor particular de \hat{h}_1 se pueden obtener dos valores concretos de \hat{h}_0 , cada uno de los cuales se verá afectado sólo por un subcaso de 1. En cierto sentido, cada camino (ver diagrama 13.1) seguido para construir la solución exacta se verá afectado por un conjunto de condiciones. Este conjunto no afectará a la solución si el camino seguido es otro. Por camino se entiende una combinación dada de $[\hat{h}_0 \hat{h}_1 \hat{h}_2]$.

En general se requiere conocer todo el conjunto de sistemas espectralmente equivalentes, de manera que si alguna condición de sensibilidad se verifica se puede estar seguro que afectará a alguno de los sistemas equivalentes y por tanto a los métodos que se basen en ellos.

Aparte de estas posibles causas de divergencia, hay que añadir un caso adicional debido al hecho de que \hat{h}_0 se calcula a partir de \hat{h}_1 . Si este último está mal estimado también lo estará el primero, de manera que la condición para la mala estimación de \hat{h}_1 hay que tenerla en cuenta aquí.

13.4.3 Interpretación para \hat{h}_2

La causa propia de divergencia para \hat{h}_2 es que $\hat{h}_0 \sim 0$, que puede suceder si $h_0 \sim 0$ o si $h_2 \sim 0$. Pero esta situación conduce a un sistema de primer orden y se ha estudiado usando el análisis de ese tipo de sistemas. De nuevo \hat{h}_2 puede estar mal estimado si lo está \hat{h}_0 , de forma que las condiciones de alta sensibilidad de este último hay que incluirlas en las del primero.

13.4.4 Relación entre sensibilidad y degeneración de las soluciones

La situación más favorable que se podría encontrar es aquella en las que los 4 sistemas obtenidos a partir de la estadística de segundo orden fuesen lo más distintos posible. De esta manera las medidas que se hagan sobre ellos permitirían diferenciarlos fácilmente y compararlos mejor respecto a uno de referencia, por ejemplo. El otro caso extremo es encontrarse con que todos los sistemas son idénticos, dando a entender que la solución al problema, incluso con este tipo de estadística, es única, siendo por tanto una situación favorable.

El problema aparece cuando los distintos sistemas obtenidos se asemejan

entre sí. En esta circunstancia las posibles operaciones y modificaciones que se le realicen a los filtros candidatos con el fin de analizarlos no darán resultados distintivos, llevando a confundirlos. Esto se verá más claro en el capítulo 15 cuando se analiza la fase de estos sistemas.

De lo discutido en estos párrafos queda claro que es conveniente estudiar las situaciones en las que se está próximo a la degeneración en las soluciones. La degeneración se produce cuando soluciones que se suponen toman distinto valor no lo hacen. Estas situaciones son, atendiendo al coeficiente calculado, las siguientes:

1. La expresión para \hat{h}_1 se da en la ecuación (13.9). Habrá degeneración si $(r_2 + r_0/2)^2 \sim r_1^2$ o como ya es sabido cuando $(h_0 + h_2)^2 \sim h_1^2$
2. La expresión para \hat{h}_0 se da en la ecuación (13.10). Siguiendo los pasos dados en análisis anteriores, esta condición implica que $h_1^2(h_0 - h_2)^2 \sim 0$, o equivalentemente que $h_1 \sim 0$ o $h_0 \sim h_2$.

Como se aprecia, las condiciones para la degeneración de las soluciones están contenidas en las condiciones bajo las cuales la sensibilidad de las mismas es alta. La sensibilidad impide obtener buenos estimadores de los candidatos y la degeneración impide distinguirlos correctamente; ambas enturbian la resolución del problema. Sus orígenes son distintos, pero las situaciones bajo las que se dan son coincidentes. Por tanto es conveniente rehuir de tales situaciones en lo posible.

13.5 Sensibilidad de la localización de los ceros a pequeñas variaciones de los coeficientes del filtro

Puede darse la situación en la que a pesar de estar razonablemente bien estimados los coeficientes del filtro, sus ceros varíen significativamente de los del sistema verdadero. Esto puede llegar a confundir la fase del sistema identificando un sistema de fase mínima como de máxima o situaciones similares. En este apartado se llevará a cabo un análisis análogo al de apartados precedentes pero relativo a la posición de los ceros en función de los coeficientes del sistema.

Sea la transformada Z de un sistema de segundo orden

$$H(z) = h_0 + h_1 z^{-1} + h_2 z^{-2} \quad (13.69)$$

de manera que sus ceros satisfacen la siguiente expresión:

$$z_0 = \frac{-h_1 \pm \sqrt{h_1^2 - 4h_0h_2}}{2h_0} \quad (13.70)$$

De nuevo en el análisis son necesarias las derivadas parciales de la magnitud cuya sensibilidad se desea estudiar respecto de las variables de las que depende.

Variaciones respecto de h_0

La derivada parcial de z_0 respecto de h_0 es

$$\frac{\partial z_0}{\partial h_0} = \left[\frac{\mp 2h_2h_0}{\sqrt{h_1^2 - 4h_0h_2}} + h_1 \mp \sqrt{h_1^2 - 4h_0h_2} \right] \frac{1}{4h_0^2} \quad (13.71)$$

que lleva asociada unas posibles causas de divergencia:

Caso a) $h_1^2 \sim 4h_0h_2$, que significa que los ceros están muy próximos.

Caso b) $h_0 \sim 0$, que conduce a un sistema de orden 1.

Variaciones respecto de h_1

La derivada parcial para este caso es:

$$\frac{\partial z_0}{\partial h_1} = \frac{1}{2h_0} \left[-1 \pm \frac{h_1}{\sqrt{h_1^2 - 4h_0h_2}} \right] \quad (13.72)$$

que lleva a las mismas condiciones de sensibilidad anteriores.

Variaciones respecto de h_2

Finalmente para el coeficiente h_2 la derivada es

$$\frac{\partial z_0}{\partial h_2} = \frac{\pm 1}{\sqrt{h_1^2 - 4h_0h_2}} \quad (13.73)$$

que toma valores elevados si $h_1^2 \sim 4h_0h_2$, esto es, si ambos ceros están juntos.

Para resumir, la localización de los ceros es más sensible a pequeñas variaciones en los coeficientes del filtro cuando $h_1^2 \sim 4h_0h_2$.

Capítulo 14

El método de identificación ciega de sistemas usando estadística de segundo orden y mínima de alto

Introducción

En este capítulo se utilizará el análisis de la estructura de las soluciones de los capítulos anteriores para proponer un nuevo método de identificación ciega inspirado en el sistema SEMP. Entre sus principales ventajas se encuentran:

1. Elimina el proceso de búsqueda del sistema con la fase correcta, inherente a todos los métodos existentes que comparten esta filosofía.
2. Se basa casi exclusivamente en estadística de segundo orden, dejando en manos de la estadística de orden superior una simple decisión binaria.
3. Presenta propiedades típicas de técnicas basadas en estadística de segundo orden como pequeño número de muestras requerido, menor varianza, etc. y las ventajas de la de orden superior como información acerca de la fase.

La forma de construir este nuevo método se ilustrará en primer lugar para un sistema de primer orden donde se expondrá el fundamento. El principal requisito es contar con expresiones explícitas para los coeficientes del filtro en función de

la secuencia de autocorrelación. Por este motivo también se puede extender a sistemas de segundo orden sin el mayor problema. Las trabas comienzan cuando no se dispone de tales expresiones tal y como ocurre para sistemas de tercer orden; para ellos es necesario buscar algunas modificaciones al método de manera que no sean necesarias las expresiones explícitas sino que basten relaciones implícitas resueltas mediante un proceso iterativo.

En el apartado 1 se ilustra el método para sistemas de primer orden y para sistemas de segundo en el apartado 2. Finalmente se propone una aproximación a la solución de los sistemas de tercer orden en el apartado 3.

14.1 Ilustrando el método para sistemas de primer orden

En los capítulos precedentes se obtuvieron expresiones explícitas de los coeficientes de un sistema lineal en función de la secuencia de autocorrelación de la salida del mismo. Asimismo se comprobó que no existe solución única, sino todo un conjunto que contiene todos los sistemas espectralmente equivalentes, EE, del sistema dado. Normalmente en ese conjunto de sistemas EE se lleva a cabo una búsqueda para elegir el que tiene la fase correcta.

Sería muy útil poder evitar este proceso de búsqueda con el fin de ahorrar operaciones y tiempo. En este capítulo se propone un nuevo método [AMC02] que consigue este objetivo. Se ilustrará identificando un sistema de primer orden.

Recuérdese que las expresiones para los coeficientes de un filtro de primer orden en función de la secuencia de autocorrelación son:

$$\hat{h}_0^2 = \left[r_0 \pm \sqrt{r_0^2 - 4r_1^2} \right] / 2 \quad (14.1)$$

$$\hat{h}_1 = \frac{r_1}{\hat{h}_0} \quad (14.2)$$

Si en (14.1) se sustituye la secuencia de autocorrelación por sus valores en función de los coeficientes del filtro (ver ecuaciones (12.4) y (12.5)) resulta

$$\hat{h}_0^2 = \left[h_0^2 + h_1^2 \pm \sqrt{(h_0^2 - h_1^2)^2} \right] / 2 \quad (14.3)$$

De aquí es fácil de ver que los dos posibles valores para el estimador del primer coeficiente del filtro son h_0 y h_1 , dependiendo de la elección de signo que se haga justo antes de la raíz cuadrada. De forma que la elección correcta de este signo puede conducir directamente a la respuesta deseada h_0 . La elección se puede hacer de la siguiente manera:

a) Si $h_0^2 > h_1^2$, la expresión (14.3) se convierte, una vez tomada la raíz cuadrada, en

$$\hat{h}_0^2 = [h_0^2 + h_1^2 \pm (h_0^2 - h_1^2)] / 2 \quad (14.4)$$

y a partir de aquí pueden aparecer dos situaciones

a.1) Eligiendo el signo + se obtiene $\hat{h}_0 = h_0$.

a.2) Eligiendo el signo - se obtiene $\hat{h}_0 = h_1$.

b) Si $h_0^2 < h_1^2$ la expresión (14.3) se convierte ahora en

$$\hat{h}_0^2 = [h_0^2 + h_1^2 \pm (h_1^2 - h_0^2)] / 2 \quad (14.5)$$

b.1) Eligiendo el signo + queda $\hat{h}_0 = h_1$.

b.2) Eligiendo el signo - queda $\hat{h}_0 = h_0$.

Considerando todas las situaciones descritas aquí, la solución deseada se puede obtener siempre si la expresión (14.1) se sustituye por

$$\hat{h}_0^2 = \left[r_0 + s \sqrt{r_0^2 - 4r_1^2} \right] / 2 \quad (14.6)$$

donde

$$s = \text{signo}[h_0^2 - h_1^2] \quad (14.7)$$

Para resolver el problema por completo ya sólo queda calcular el signo s . Resulta claro que conocer este signo es equivalente a conocer la fase del canal: si $|h_0| > |h_1|$ el canal es de fase mínima y es de fase máxima si $|h_1| > |h_0|$. De manera que no se puede estimar usando estadística de segundo orden de la salida porque es ciega a fase. Una opción posible es el uso de cumulantes porque proporcionan una forma sencilla de realizar esta tarea.

Primero hay que notar que s se puede escribir equivalentemente como:

$$s = \text{signo}[1 - \bar{h}_1^2] \quad (14.8)$$

donde $\bar{h}_1 \equiv h_1/h_0$. Con la ayuda de la expresión de Brillinger-Rosenblatt [BR67] es directo comprobar que:

$$\bar{h}_1 = \frac{c_{4x}(0, 1, 1)}{c_{4x}(0, 0, 1)} \quad (14.9)$$

Se deberían hacer algunos comentarios acerca de (14.9):

1. Se han elegido los cumulantes de cuarto orden para que la expresión se pueda aplicar incluso en el caso de señales de entrada con distribución de probabilidad simétricas, que producen señales de salida simétricas.
2. Esta expresión se puede considerar como un caso particular de las fórmulas de la rebanada q (q -slice) para coeficientes normalizados.
3. Se pueden generalizar para cualquier orden de cumulante y filtros de cualquier orden.

Una vez que \hat{h}_0 se ha estimado correctamente, la expresión (14.2) da el valor correcto para \hat{h}_1 .

Nótese que la información necesaria obtenida de la estadística de alto orden EAO es simplemente un signo, así que no se espera que el método sufra de las conocidas desventajas de EAO: necesidad de longitudes de datos grandes, gran varianza en sus estimadores, etc. Además, la única condición que se necesita para que este método se pueda aplicar es que la señal de entrada sea blanca. No se necesita ninguna información adicional sobre su estadística, salvo la varianza. Debido a las características del método descrito se le denominará método de Identificación Ciega de sistemas mediante estadística de Segundo orden y Mínima de Alto ICSMA.

14.2 Aplicación de ICSMA a modelos de segundo orden

La técnica que se acaba de describir se puede generalizar a filtros de cualquier orden siempre y cuando se disponga de expresiones explícitas de los coeficientes en función de la secuencia de autocorrelación. Para ilustrar este punto se aplicará este método a sistemas de segundo orden.

Recuérdese que las soluciones para los coeficientes del filtro son

$$\hat{h}_1^2 = r_2 + r_0/2 \pm \sqrt{(r_2 + r_0/2)^2 - r_1^2} \quad (14.10)$$

$$\hat{h}_0 = \frac{r_1 \pm \sqrt{r_1^2 - 4\hat{h}_1^2 r_2}}{2\hat{h}_1} \quad (14.11)$$

$$\hat{h}_2 = \frac{r_2}{\hat{h}_0} \quad (14.12)$$

De nuevo la elección de los signos en (14.10) y (14.11) es crucial para llegar al filtro verdadero. Para ver qué signo es el que se va a escoger es conveniente

reescribir (14.10) en función de los coeficientes verdaderos:

$$\hat{h}_1^2 = \left[(h_0 + h_2)^2 + h_1^2 \pm \sqrt{(h_1^2 - (h_0 + h_2)^2)^2} \right] / 2 \quad (14.13)$$

en la que pueden aparecer dos situaciones:

a) Si $h_1^2 > (h_0 + h_2)^2$ la expresión (14.13) se puede escribir como

$$\hat{h}_1^2 = \frac{1}{2} [(h_0 + h_2)^2 + h_1^2 \pm (h_1^2 - (h_0 + h_2)^2)] \quad (14.14)$$

y dependiendo del signo elegido pueden suceder dos posibilidades:

a.1) Si se elige el signo +, el resultado es $\hat{h}_1 = h_1$.

a.2) Por otro lado, si se elige el -, queda $\hat{h}_1 = h_0 + h_2$

b) Pero si $h_1^2 < (h_0 + h_2)^2$ la expresión (14.13) se convierte en

$$\hat{h}_1^2 = \frac{1}{2} [(h_0 + h_2)^2 + h_1^2 \pm ((h_0 + h_2)^2 - h_1^2)] \quad (14.15)$$

y las dos posibilidades son:

b.1) Si se elige el +, $\hat{h}_1 = h_0 + h_2$

b.2) Si se elige el -, $\hat{h}_1 = h_1$

Considerando todas estas situaciones, el valor verdadero del segundo coeficiente del filtro se puede obtener si la expresión (14.10) se modifica a:

$$\hat{h}_1^2 = r_2 + r_0/2 + s_1 \sqrt{(r_2 + r_0/2)^2 - r_1^2} \quad (14.16)$$

donde

$$s_1 = \text{signo}[h_1^2 - (h_0 + h_2)^2] \quad (14.17)$$

A partir de \hat{h}_1 el siguiente coeficiente \hat{h}_0 se puede calcular usando (14.11). Para este fin (14.11) se escribe en función de los verdaderos coeficientes del filtro asumiendo que \hat{h}_1 ha sido correctamente estimado y por tanto es igual a h_1 ,

$$\hat{h}_0 = \frac{h_1(h_0 + h_2) \pm \sqrt{h_1^2(h_0 - h_2)^2}}{2h_1} \quad (14.18)$$

Y siguiendo la técnica conocida,

a) Si $h_1(h_0 - h_2) > 0$ la expresión (14.18) se puede reescribir como

$$\hat{h}_0 = \frac{h_1(h_0 + h_2) \pm h_1(h_0 - h_2)}{2h_1} \quad (14.19)$$

a.1) Usando el signo +, el estimador toma el valor $\hat{h}_0 = h_0$.

a.2) Con el signo -, $\hat{h}_0 = h_2$.

b) Si $h_1(h_0 - h_2) < 0$ la expresión (14.18) se convierte en

$$\hat{h}_0 = \frac{h_1(h_0 + h_2) \pm h_1(h_2 - h_0)}{2h_1} \quad (14.20)$$

b.1) Eligiendo el signo + resulta $\hat{h}_0 = h_2$.

b.2) Eligiendo el signo - resulta $\hat{h}_0 = h_0$.

Considerando todas estas posibles situaciones, la expresión (14.11) se puede modificar para dar siempre el valor del verdadero h_0 ,

$$\hat{h}_0 = \frac{r_1 + s_2 \sqrt{r_1^2 - 4\hat{h}_1^2 r_2}}{2\hat{h}_1} \quad (14.21)$$

donde

$$s_2 = \text{signo}[h_1(h_0 - h_2)] \quad (14.22)$$

Finalmente, a partir de \hat{h}_0 el último coeficiente \hat{h}_2 se puede estimar usando (14.12).

Todavía queda el problema de la estimación de los signos s_1 y s_2 . Una aproximación sencilla consiste en escribir ambos en función de los coeficientes normalizados,

$$s_1 = \text{signo} [\bar{h}_1^2 - (1 + \bar{h}_2)^2] \quad (14.23)$$

$$s_2 = \text{signo} [\bar{h}_1(1 - \bar{h}_2)] \quad (14.24)$$

donde los coeficientes normalizados están definidos como $\bar{h}_i \equiv h_i/h_0$ y se pueden calcular usando la siguiente expresión:

$$\bar{h}_i = \frac{c_{4x}(0, i, 2)}{c_{4x}(0, 0, 2)} \quad (14.25)$$

como se puede comprobar fácilmente mediante la expresión de Brillinger-Rosenblatt.

14.3 Sistemas de tercer orden

El problema para órdenes inferiores ya es conocido con anterioridad, se trata de obtener los coeficientes del filtro $\mathbf{h} = [h_0 \ h_1 \ h_2 \ h_3]$ a partir de la secuencia de autocorrelación de la salida, usando, en este caso, el siguiente sistema de ecuaciones:

$$r_0 = h_0^2 + h_1^2 + h_2^2 + h_3^2 \quad (14.26)$$

$$r_1 = h_0h_1 + h_1h_2 + h_2h_3 \quad (14.27)$$

$$r_2 = h_0h_2 + h_1h_3 \quad (14.28)$$

$$r_3 = h_0h_3 \quad (14.29)$$

Donde como siempre se ha tomado $\sigma_w^2 = 1$ o la secuencia de autocorrelación se ha normalizado. Las ecuaciones son no lineales y la solución no es tan directa como para sistemas de órdenes inferiores. Se sabe que hay 8 soluciones diferentes correspondiente a las 2^3 posibles combinaciones de los tres ceros de \mathbf{h} cuando se reflejan con respecto al círculo unidad, de manera que si se encuentra una expresión analítica y explícita para alguno de los coeficientes puede ser la solución para una ecuación de orden octavo. Si se recuerda la solución para un sistema de 2º orden, el coeficiente h_1 era la solución a una ecuación cuadrática y no la solución a una de cuarto orden porque 2 de las soluciones espectralmente equivalentes compartían el mismo h_1 , esto es, había cierta simetría que simplificaba el problema. Con el fin de encontrar ecuaciones más manejables aquí, es decir, ecuaciones de menor orden, sería útil estudiar la estructura del conjunto de soluciones espectralmente equivalentes en busca de alguna de tales simetrías.

Con el uso de un programa de simulación se puede calcular el conjunto espectralmente equivalente reflejando los ceros de un sistema dado y se pueden estudiar los coeficientes. La única simetría que se encontró fue que el producto h_1h_2 aparece por parejas, esto es, hay cuatro valores diferentes para el producto h_1h_2 , así que quizás sea una buena idea el definir una nueva variable $\alpha = h_1h_2$ y despejarla en el sistema. Para mantener la homogeneidad de las incógnitas se definen más variables; una elección obvia es llamar $r_3 = h_0h_3$ y el resto se escogen de manera aleatoria como $\beta = h_1h_0$ y $\gamma = h_0h_2$. Las ecuaciones (14.26)-(14.29) se convierten en:

$$r_2 = \gamma + \frac{\alpha r_3}{\gamma} \quad (14.30)$$

$$r_1 = \beta + \alpha + \frac{\alpha r_3}{\beta} \quad (14.31)$$

$$r_0 = \frac{\beta\gamma}{\alpha} + \frac{\beta\alpha}{\gamma} + \frac{\alpha\gamma}{\beta} + \frac{\alpha r_3^2}{\beta\gamma} \quad (14.32)$$

La ecuación (14.32) asegura que el filtro tiene la potencia correcta bajo la estructura definida por (14.30) y (14.31). Este sistema consta de tres incógnitas α , β y γ y de tres ecuaciones (14.30)-(14.32).

Partiendo de la ecuación (14.30) es fácil comprobar que

$$\gamma = \frac{1}{2} \left(r_2 \pm \sqrt{r_2^2 - 4\alpha r_3} \right) \quad (14.33)$$

y de (14.31)

$$\beta = \frac{1}{2} \left(r_1 - \alpha \pm \sqrt{(r_1 - \alpha)^2 - 4\alpha r_3} \right) \quad (14.34)$$

Si α fuera conocida, el problema estaría resuelto. El siguiente paso lógico es sustituir γ y β por sus expresiones como función de α ((14.33) y (14.34)) en (14.32) y despejar α . Desgraciadamente la ecuación resultante es demasiado complicada para encontrar soluciones analíticas.

14.3.1 Una solución iterativa basada en ICSMA

La ecuación (14.32) se puede reescribir como una ecuación cuadrática para α como sigue:

$$\left[\frac{r_3^2}{\beta\gamma} + \frac{\beta}{\gamma} + \frac{\gamma}{\beta} \right] \alpha^2 - r_0 \alpha + \beta\gamma = 0 \quad (14.35)$$

con su solución

$$\alpha = \frac{1}{2 \left[\frac{r_3^2}{\beta\gamma} + \frac{\beta}{\gamma} + \frac{\gamma}{\beta} \right]} \left[r_0 \pm \sqrt{r_0^2 - 4\beta\gamma \left(\frac{r_3^2}{\beta\gamma} + \frac{\beta}{\gamma} + \frac{\gamma}{\beta} \right)} \right] \quad (14.36)$$

Siguiendo un procedimiento similar al seguido para sistemas de órdenes inferiores, el signo correcto en (14.33), (14.34) y (14.36) se puede calcular de la siguiente manera:

$$s_1 = \text{signo}[\bar{h}_2 - \bar{h}_1 \bar{h}_3] \quad (14.37)$$

$$s_2 = \text{signo}[\bar{h}_1 - \bar{h}_2 \bar{h}_3] \quad (14.38)$$

$$s_3 = \text{signo}[(\bar{h}_1^2 + \bar{h}_2^2 + \bar{h}_3^2) - 1] \quad (14.39)$$

respectivamente y las expresiones reescritas como

$$\gamma = \frac{1}{2} \left(r_2 + s_1 \sqrt{r_2^2 - 4\alpha r_3} \right) \quad (14.40)$$

$$\beta = \frac{1}{2} \left(r_1 - \alpha + s_2 \sqrt{(r_1 - \alpha)^2 - 4\alpha r_3} \right) \quad (14.41)$$

$$\alpha = \frac{1}{2 \left[\frac{r_3^2}{\beta\gamma} + \frac{\beta}{\gamma} + \frac{\gamma}{\beta} \right]} \left[r_0 + s_3 \sqrt{r_0^2 - 4\beta\gamma \left(\frac{r_3^2}{\beta\gamma} + \frac{\beta}{\gamma} + \frac{\gamma}{\beta} \right)} \right] \quad (14.42)$$

Estas ecuaciones estilo ICSMA se van a resolver mediante un proceso iterativo. El proceso comienza con la estimación de los signos (14.37)-(14.39) mediante

el empleo de la expresión para los coeficientes normalizados en función de los cumulantes de cuarto orden. Una vez que se conocen estos signos el proceso iterativo se inicializa fijando un valor inicial para α , usando, por ejemplo:

$$\alpha = \frac{c_{4x}(1, 2, 3)}{\gamma_4 r_3} \quad (14.43)$$

Este valor de α se utiliza en (14.40) y (14.41) para obtener valores iniciales de γ y β , que a su vez se usan para calcular un nuevo valor de α via (14.42) y entonces el ciclo se repite. La parte más importante del proceso es la inicialización de α , malos estimadores iniciales pueden llevar a que el proceso converja a unos coeficientes distintos de los del filtro original.

Comprobación práctica

Para comprobar el funcionamiento del proceso iterativo presentado, se ha llevado a cabo una simulación en una situación favorable. En primer lugar se han supuesto conocidos los coeficientes verdaderos para el cálculo de s_1 , s_2 y s_3 y la secuencia de autocorrelación. A continuación se da un valor inicial a α ad hoc cerca del valor verdadero (normalmente con un error relativo menor del 10%). El modelo estudiado fue un MA=[2.000 9.2000 12.7700 5.4672], el verdadero valor de α es 115.1854 y es inicializado como 126. En la figura 14.1 se muestra la evolución de α y su verdadero valor. Después de 2000 iteraciones los coeficientes del filtro estimados se ajustan perfectamente a los verdaderos.

Esta simulación verifica que teóricamente funciona la técnica iterativa propuesta. Desafortunadamente no es el caso cuando se emplean estimadores en el proceso de inicialización, aunque se necesitan más simulaciones para comprobar este punto.

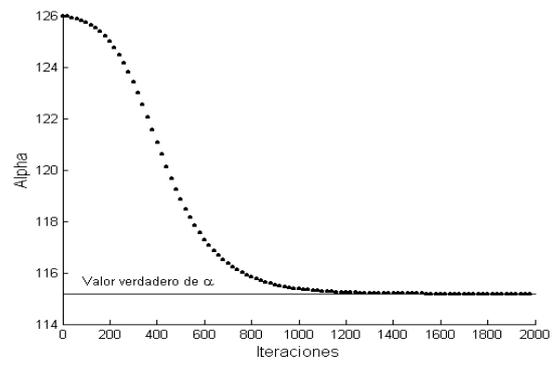


Figura 14.1 Evolución de α en el método ICSMA iterativo

Capítulo 15

Análisis de los métodos de elección de la fase

Introducción

En este capítulo se presentan métodos de elección de la fase, en este caso se aplican al conjunto de sistemas espectralmente equivalentes, y algunas herramientas y observaciones útiles para su análisis. Los métodos se pueden clasificar entre basados en el ecualizador y no basados en el ecualizador, pero casi todos ellos se basan en el criterio de ajuste por cumulantes, salvo el nuevo método propuesto y otro basado en el criterio del módulo constante. En total se analizan 10 métodos diferentes aplicados a 12 modelos distintos.

En el apartado 1 se comentan algunos detalles acerca de la estimación de la secuencia de autocorrelación para justificar los resultados experimentales; con este mismo fin también se define una nueva magnitud, la mínima diferencia relativa en el apartado 2. En el apartado 3 se describen los métodos estudiados y se presenta un primer análisis esperado de ellos en el apartado 4. Los modelos estudiados, junto con el análisis de los resultados obtenidos, se presentan en los apartados 5 y 6 respectivamente.

15.1 Acerca de la estimación de la secuencia de autocorrelación

Recuérdese que el sistema lineal \mathbf{h} a identificar relacionaba la entrada y la salida según la siguiente expresión:

$$x(n) = \sum_{i=0}^q h_i w(n-i) \quad (15.1)$$

donde el proceso $w(n)$ es una serie de variables aleatorias independientes e idénticamente distribuidas. La salida de los sistemas está contaminada por ruido gaussiano blanco $\{v(n)\}$, independiente de la entrada y de varianza σ_v^2 conocida. De manera que las muestras disponibles tienen la forma:

$$y(n) = x(n) + v(n) \quad (15.2)$$

El problema que tratamos en este apartado es el de estudiar los factores que afectan la estimación de la secuencia de autocorrelación de $\{x(n)\}$ dada en (15.1), suponiendo conocida la varianza del ruido y teniendo sólo acceso a muestras de $\{y(n)\}$ en (15.2). Dicha secuencia se estimará, como es habitual, como un promedio temporal: dadas N muestras, $y(1), \dots, y(N)$, el estimador de $r_{yy}(m) = E[y(n)y(n+m)]$ es

$$\hat{r}_{yy}(m) = \frac{1}{N-m} \sum_{i=1}^{N-m} y(i)y(i+m) \quad (15.3)$$

Usando (15.1) y (15.2) la expresión anterior se puede expresar como

$$\begin{aligned} \hat{r}_{yy}(m) = \frac{1}{N-m} \sum_{i=0}^{N-m} \left[\sum_{k,l=1}^q b_k b_l w(i-k)w(i+m-l) + \right. \\ \left. + \sum_{k=0}^q b_k w(i-k)v(i+m) + \right. \\ \left. + \sum_{l=0}^q b_l w(i+m-l)v(i) + v(i)v(i+m) \right] \end{aligned} \quad (15.4)$$

que se puede reescribir en función de los estimadores de la autocorrelación

$$\begin{aligned} \hat{r}_{yy}(m) = \sum_{k,l=0}^q b_k b_l \hat{r}_{ww}(m+k-l) + \\ + \sum_{k=0}^q b_k [\hat{r}_{wv}(m+k) + \hat{r}_{wv}(m-k)] + \hat{r}_{vv}(m) \end{aligned} \quad (15.5)$$

Los estimadores de la autocorrelación del miembro de la derecha tienen el valor esperado siguiente

$$E[\hat{r}_{ww}(m)] = \sigma_w^2 \delta(m) \quad (15.6a)$$

$$E[\hat{r}_{wv}(m)] = 0 \quad (15.6b)$$

$$E[\hat{r}_{vv}(m)] = \sigma_v^2 \delta(m) \quad (15.6c)$$

y ya que son no sesgados, estos son los valores que estamos buscando. Donde aparecen las varianzas conocidas de los procesos involucrados. Usando (15.6a)-(15.6c), el valor esperado de (15.5) (o su comportamiento asintótico) es

$$r_{yy}(m) = \sigma_w^2 \sum_k^q b_k b_{k+m} + \sigma_v^2 \delta(m) \quad (15.7)$$

o equivalentemente

$$r_{yy}(m) = r_{xx}(m) + \sigma_v^2 \delta(m) \quad (15.8)$$

de manera que sabiendo la potencia del ruido, la secuencia de autocorrelación de $x(n)$ se puede obtener de la de $y(n)$:

$$r_{xx}(m) = r_{yy}(m) - \sigma_v^2 \delta(m) \quad (15.9)$$

Si en vez de utilizar el valor real de la autocorrelación de $y(n)$ en (15.9) se utiliza el estimador de (15.4), diferentes causas pueden hacer que esté lejos de su valor verdadero. Éstas se pueden obtener mediante una inspección de (15.5):

1. Si \hat{r}_{ww} se diferencia mucho de su valor esperado (15.6a).
2. Si \hat{r}_{wv} se desvía mucho de 0, su valor esperado en (15.6b).
3. Si \hat{r}_{vv} está lejos de su valor esperado (15.6c).

Suponiendo que los únicos parámetros libres son los coeficientes del filtro y la SNR —es decir, la potencia del ruido—, los puntos anteriores afectan más al valor de (15.5) si,

1. Cuanto mayor sean los coeficientes del filtro, mayor será la influencia del segundo sumando en el miembro de la derecha en (15.5).
2. Cuando menor sea la SNR, mayor será la desviación de \hat{r}_{wv} de 0.

15.2 Una magnitud nueva: Mínima Diferencia Relativa

Como es conocido, la estadística de alto orden lleva información sobre la fase de un proceso aleatorio dado. En muchas ocasiones se utilizan los cumulantes como herramienta para identificar a un sistema lineal totalmente, tanto su respuesta en magnitud como su respuesta en fase; piénsese por ejemplo en la técnica conocida como ajuste de cumulantes (cumulant matching).

Si se tiene un conjunto de sistemas lineales como posibles candidatos del sistema real que se quiere identificar, se puede elegir como candidato óptimo, o candidato válido aquel cuyos cumulantes se acerquen más a los cumulantes estimados con las muestras medidas a la salida del sistema. Piénsese por el momento que se va a utilizar un solo cumulante, en particular de la diagonal principal, para llevar a cabo la elección. Sea por ejemplo $c_{4x}(m, m, m)$ el cumulante en cuestión, notado como $c_{4x}(m)$ por simplicidad. Si los distintos cumulantes, cada uno correspondiente a los distintos sistemas candidatos, están muy cerca unos de otros, será más difícil elegir entre ellos el que más se acerca al cumulante verdadero. En las próximas líneas se expresa matemáticamente en una nueva magnitud esta idea sencilla.

Sea $\Upsilon(m) = \{c_{4x}^{(1)}(m), \dots, c_{4x}^{(P)}(m)\}$ el conjunto formado por los P cumulantes correspondientes a los P sistemas candidatos, ordenados de menor a mayor. A continuación se coge el primer elemento y se le subtrae al siguiente, dividiendo por el primero por motivos de normalización. De esta manera se calcula la primera diferencia relativa. Para calcular la segunda diferencia relativa se sigue el proceso anterior con el segundo elemento del conjunto. Esto se repite hasta calcular las $P - 1$ diferencias relativas. La mínima diferencia relativa [AMGR02] para este cumulante $\text{MRD}(c_{4x}(m))$ es la menor de todas ellas y da una idea del grado de dificultad en el proceso de selección del sistema candidato verdadero basándose en ajuste por cumulantes. Cuanto más pequeño sea su valor, más difícil será diferenciar entre los posibles candidatos aquel cuyo cumulante se acerque más al cumulante verdadero. Si se usan todos los cumulantes diagonales en el proceso de selección, la MRD se calcula como el promedio de $\text{MRD}(c_{4x}(m))$ para todos los valores de m para los cuales el cumulante no sea nulo.

15.2.1 MRD para sistemas espectralmente equivalentes

Una situación donde se suele utilizar la técnica de ajuste por cumulantes para la identificación de sistemas, es en el proceso de elegir el sistema que tiene la fase correcta entre todos los sistemas que comparten el mismo espectro de potencia. En capítulos anteriores se han obtenido todos los sistemas espec-

tralmente equivalentes de uno dado, para sistemas de orden 1 y 2. Éstos se obtuvieron mediante la reflexión de cada uno de los ceros respecto al círculo unidad. El mismo resultado se obtuvo al intentar identificar el sistema a partir de la secuencia de autocorrelación de la salida: no había solución única, como ya es sabido. Por tanto, el proceso de identificación ciega se debe completar con alguna técnica que elija la fase correcta entre todos los posibles candidatos. Una opción posible es el uso de ajuste por cumulantes tal y como se ha explicado en párrafos precedentes.

Dadas las simetrías de cumulantes por un lado y sistemas espectralmente equivalentes por otro, se cumple la siguiente

Proposición 15.2.1 *La mínima diferencia relativa (MDR) de un conjunto de sistemas espectralmente equivalentes de segundo orden presenta las siguientes propiedades:*

1. $MDR(c_{4x}(m)) = MDR(c_{4x}(-m))$
2. $MDR(c_{4x}(0)) = 0$

Demos.- La propiedad 1 se demuestra fácilmente teniendo en cuenta que dado un sistema definido por sus coeficientes $\mathbf{h} = [h_0 \ h_1 \ h_2]$, se puede obtener otro espectralmente equivalente con coeficientes $\mathbf{h}' = [h_2 \ h_1 \ h_0]$, estando por tanto agrupados por parejas según este criterio. Por otro lado, teniendo en cuenta la expresión para los cumulantes de la salida de un sistema lineal dada por Brillinger y Rossenblat [BR67]

$$c_{kx}(\tau_1, \dots, \tau_{k-1}) = \gamma_{kw} \sum_{i=0}^q h_i h_{i+\tau_1} h_{i+\tau_{k-1}} \quad (15.10)$$

se tiene que

$$c_{4x}^{(\mathbf{h})}(m) = c_{4x}^{(\mathbf{h}')}(m) \quad (15.11)$$

por lo que $\Upsilon(m) = \Upsilon(-m)$ y por tanto la mínima diferencia relativa obtenida de cada conjunto es la misma. No hace falta recurrir a la expresión (15.10) para obtener (15.11), basta recurrir al hecho de que bajo estadística estacionaria, cambiar los intervalos temporales m , en momentos y cumulantes, por sus opuestos $-m$ equivale a invertir el orden de los coeficientes que definen el filtro.

La propiedad 2 se obtiene particularizando (15.10) al caso $k = 4$ y $\tau_1 = \dots = \tau_3 = 0$

$$c_{4x}(0, 0, 0) = \sum_{i=0}^q h_i^4 \quad (15.12)$$

con lo que es directo comprobar que $c_{4x}^{(\mathbf{h})}(0) = c_{4x}^{(\mathbf{h}')}(0)$ y por tanto $MRD(c_{4x}(0)) = 0$. Que también se puede ver como un caso especial de (15.11) para $m=0$.

Predicciones teóricas de la MDR

La proposición 15.2.1 tiene consecuencias importantes en lo que respecta a la elección de la fase adecuada en un conjunto de sistemas espectralmente equivalentes. Éstas consecuencias son dos:

Consecuencia 1: Es indistinto emplear el cumulante $c_{4x}(m)$ o el $c_{4x}(-m)$ ya que es de esperar que la dificultad encontrada para identificar el sistema sea la misma con ambos.

Consecuencia 2: Técnicas que utilicen exclusivamente la curtosis darán resultados erróneos siempre, ya que al menos hay dos sistemas espectralmente equivalentes con la misma curtosis y por tanto no es posible diferenciarlos atendiendo a ella.

15.3 Descripción de los métodos estudiados

En capítulos anteriores se ha tratado el problema de la identificación de sistemas utilizando estadística de segundo orden de la salida del mismo. Como ya es conocido la solución al problema planteado de esta manera no es única, sino que se obtiene todo el conjunto de sistemas espectralmente equivalentes al sistema problema. Como también se ha ido insinuando anteriormente, una segunda parte necesaria para la completa identificación del sistema es la determinación de su fase correcta, es decir, elegir dentro del conjunto de sistemas espectralmente equivalentes aquel cuya fase coincida con la fase del sistema problema.

Esta elección se puede llevar a cabo utilizando una gran variedad de técnicas; las utilizadas en este trabajo las podemos clasificar dentro de dos categorías generales: basadas en el ecualizador y no basadas en el ecualizador o directas. Todas estas técnicas, como ya se ha indicado, deciden la fase a posteriori. La herramienta utilizada para llevar a cabo dicha elección debe estar basada en estadística de alto orden, ya que la estadística de segundo orden es ciega a fase según reflejan los capítulos anteriores. En definitiva, estas técnicas utilizan estadística de alto orden a posteriori. Hay, sin embargo, la posibilidad de extraer la información necesaria mediante estadística de alto orden antes de calcular el conjunto de sistemas espectralmente equivalentes (antes de aplicar la estadística de segundo orden), tal y como se demostró en el capítulo 14. A este último tipo de técnicas se le llamarán técnicas con estadística de alto orden a priori, estando incluido en él el método de Identificación Ciega basado en estadística de Segundo orden con Mínima de Alto (ICSMA). Las técnicas con estadística de alto orden a posteriori se describen a continuación, el método a priori se derivó en el capítulo 14.

15.3.1 Técnicas basadas en el ecualizador

Este tipo de técnicas calculan el ecualizador de cada uno de los sistemas candidatos estimados, y junto con la señal de salida del sistema, obtienen la señal de entrada al canal ecualizada. Una de las principales ventajas que presentan estos métodos es que a la par que se lleva a cabo la identificación, también se lleva a cabo la ecualización. Obviamente vienen penalizados por un mayor número de operaciones y el riesgo de que los errores se vayan acumulando en el proceso de calcular el ecualizador. Como la distribución de la señal de entrada es conocida, ya sólo resta ver qué candidato da la señal ecualizada más cercana a la verdadera. La forma de ver la cercanía entre ambas señales depende del criterio concreto a aplicar.

Criterio del módulo constante (MC)

Consiste en imponer que las muestras de la señal ecualizada estén lo más cerca posible de tener módulo fijo. Está inspirado en el conocido algoritmo de módulo constante de Godard [God74], uno de los pioneros en el ámbito de la identificación y ecualización ciega. Resulta lógico pensar que este criterio funcionará mejor si de hecho la señal primigenia tenía módulo constante, como sucede en el caso aquí analizado ya que se utiliza una secuencia binaria $\{+1, -1\}$. Si no fuera así, la constante R a la que debería acercarse el módulo de la señal ecualizada se obtendría de minimizar

$$f(R) = E[|w_{eq}(n)|^2 - R]^2 \quad (15.13)$$

resultando $R = E[|w(n)|^4]/E[|w(n)|^2]$. Obviamente, para la secuencia binaria anterior $R = 1$.

Criterio de curtosis próxima (CP)

Además de tener en cuenta una propiedad de inspiración geométrica como es el módulo, también es posible considerar criterios puramente estadísticos. Como la estadística de la señal de entrada es conocida, una opción válida es elegir como candidato perfecto aquél que presente esta misma distribución. Una forma de comprobar la similitud de distribuciones es mediante el empleo de sus momentos y cumulantes, ya que el pleno conocimiento de estos a todos los órdenes la caracterizan completamente. Como por cuestiones prácticas no es factible usar todos, hay que elegir aquellos que se utilizarán para esta tarea. En primer lugar se emplearán los cumulantes, en vez de los momentos, por los motivos ya comentados en varios puntos de esta Memoria. El orden de los mismos es otro punto importante, ya que a mayor orden más dificultosa es su estimación. Lo mejor sería emplear los cumulantes de tercer orden, pero al ser de orden

impar son nulos cuando las variables aleatorias involucradas tienen una función de densidad de probabilidad par, como es el caso aquí estudiado. Por tanto hay que recurrir a los cumulantes de cuarto orden. Al ser la señal de entrada iid, el único cumulante de cuarto orden no nulo es la curtosis y será éste el que se empleará para elegir al candidato con la fase correcta: la curtosis estimada para cada una de las señales ecualizadas con los diferentes candidatos se comparará con la curtosis verdadera conocida de la señal de entrada real. El candidato cuyo ecualizador dé la señal con la curtosis más cercana a la verdadera será elegido.

Criterio de la curtosis próxima modificado (CPM)

El criterio anterior se ve fortalecido imponiendo además que los cumulantes diagonales (distintos de la principal—curtosis—) sean lo más próximos a cero. En particular se ha tomado $c_{4w_{eq}}(1, 1, 1)$.

15.3.2 Métodos directos

Este tipo de métodos utilizan directamente la salida del sistema $y(n)$ para determinar cuál de los sistemas candidatos tiene la fase correcta. Presentan la ventaja principal de que son más sencillos y por tanto requieren un menor número de operaciones por lo que se espera que sean menos sensibles a la acumulación de errores.

Ajuste por cumulantes (AC)

Este método ya se ha esbozado en el apartado 15.2 que trataba sobre la MDR y en el subapartado 15.3.1 anterior. En este caso se comparan las estadísticas estimadas y teórica de la salida del sistema por medio de los cumulantes diagonales de cuarto orden. Sólo se utilizan las secciones diagonales porque se ha informado en la literatura que el resultado no mejora respecto al así obtenido si se incluyeran todos los cumulantes de cuarto orden posibles [ZMM93]. Al igual que en el subapartado 15.3.1 los cumulantes estimados se obtienen de los momentos estimados. Como es habitual el estimador del operador valor esperado se implementa mediante un promedio temporal. Los cumulantes teóricos se calculan mediante (15.10), conocida la curtosis de la señal de entrada y los coeficientes de los distintos filtros candidatos. De esta manera el sistema elegido \mathbf{h}^* es aquel que cumple

$$\mathbf{h}^* = \min_{\mathbf{h} \in \mathcal{S}} \sum_{m=-2}^2 \left| \sum_{i=1}^q \gamma_{4w} h_i h_{i+m}^3 - \hat{c}_{4x}(m, m, m) \right|^2 \quad (15.14)$$

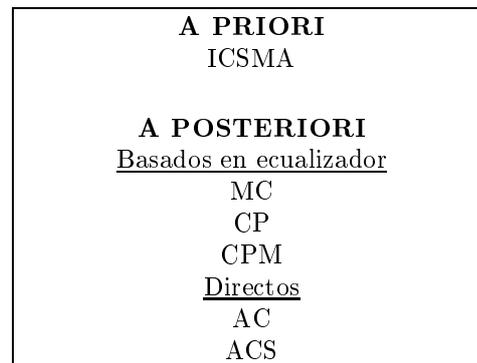


Figura 15.1 *Métodos de elección de fase*

donde ζ es el conjunto de sistemas espectralmente equivalentes y el sombrero indica que la magnitud ha sido estimada.

Ajuste por cumulantes, versión simplificada (ACS)

En el apartado anterior se emplean todos los cumulantes diagonales para decidir el sistema con la fase adecuada. En esta sección se propone el uso de un único cumulante diagonal para llevar a cabo dicha identificación. Según el intervalo temporal m considerado en $c_{4x}(m)$ el método se llamará ACS- m .

15.4 El comportamiento esperado de los métodos

Una vez descritos los métodos que se van a emplear para complementar la identificación de los sistemas, es posible discutir a priori cuál es el comportamiento que van a presentar. Para tener una visión más general de los distintos tipos de técnicas, en la figura 15.1 se muestran por categorías. Por el momento, dos ideas básicas serán suficientes para hacer una primera clasificación atendiendo a la calidad del resultado esperado:

1. Cantidad de información utilizada.
2. Sencillez del método, traducida en bajo coste computacional.

Según la idea 2 los métodos basados en el ecualizador son los que darán los peores resultados, ya que como se comentó al presentarlos los posibles errores originados

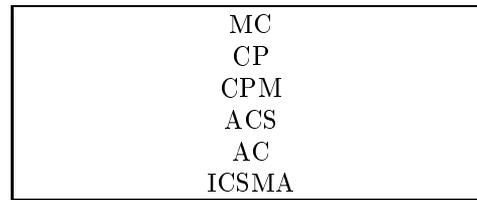


Figura 15.2 *Métodos de elección de fase ordenados de peor a mejor comportamiento (de arriba a abajo)*

en la estimación de los coeficientes del filtro pueden verse incrementados durante el proceso del cálculo del filtro inverso (utilizado aquí como ecualizador). Dentro de ese grupo, y atendiendo a la idea 1, es de esperar que CPM mejore los resultados de CP ya que utiliza más cumulantes para hacer la elección. Además, dada la naturaleza aleatoria de la señal de entrada, es más conveniente confiar en las técnicas que usan medidas estadísticas a las que usan otras, como por ejemplo criterios geométricos; por tanto MC dará los peores resultados.

El resto de métodos, tanto los que utilizan estadística de alto orden a priori como los a posteriori directos están a la par en lo que respecta a la sencillez del método, con lo que las posibles diferencias surgirán bajo el punto 1. No cabe duda que es más lógico esperar mejores resultados de AC que de ACS ya que es más fácil comparar estadísticas conforme más cumulantes se conozcan. Por tanto, y por ahora, el mejor método que se presenta es AC. El único método a priori utilizado ICSMA, cuenta con la información justa y necesaria de alto orden antes de calcular, en última instancia, los coeficientes usando estadística de segundo orden; estima, por tanto, más información que el resto de métodos y es tan sencillo o más que AC y ACS, con lo que su comportamiento debe ser, al menos, tan bueno como AC.

La figura 15.2 recoge los métodos estudiados de peor a mejor comportamiento esperado. La justificación final de si las ideas 1 y 2 son suficientes como para explicar la actuación de las técnicas empleadas la darán las simulaciones presentadas en el próximo apartado. Sin embargo, dada la gran cantidad de factores¹ que intervienen en todo el proceso, así como que cada técnica consta de varios pasos, es de esperar una gran riqueza en cuanto a fenómenos observados en las simulaciones. Tal variedad no es probable que se pueda explicar tomando como base las ideas anteriores. ¿Será posible obtener mejores resultados utilizando sólo un cumulante diagonal que todos ellos?

¹estimación de estadística de segundo y alto orden, sensibilidad de las expresiones para los coeficientes del filtro, etc.

15.5 Descripción del entorno y de los modelos utilizados

Se han llevado a cabo simulaciones donde se estudian 12 filtros de segundo orden. Cada uno viene definido por un vector de coeficientes $\mathbf{h} = [h_0 \ h_1 \ h_2]$ de acuerdo con la expresión (15.1) donde el proceso $w(n)$ sigue una secuencia binaria $\{+1, -1\}$ siendo cada valor equiprobable. La estadística del proceso de entrada es la siguiente

$$r_{ww}(m_1) = \sigma_w^2 \delta(m_1) \quad (15.15a)$$

$$c_{3w}(m_1, m_2) = 0 \quad (15.15b)$$

$$c_{4w}(m_1, m_2, m_3) = \gamma_{4w} \delta(m_1) \delta(m_2) \delta(m_3) \quad (15.15c)$$

donde $c_{kw}(m_1, \dots, m_{k-1})$ denota los cumulantes de orden k -ésimo del proceso $\{w(n)\}$, $\sigma_w^2 = 1$ y $\gamma_{4w} = -2$. La tabla 15.1 recoge los coeficientes \mathbf{h} de los modelos estudiados junto con la localización de los ceros z_1 y z_2 de cada uno.

	h_0	h_1	h_2	z_1	z_2
modelo 1	1	1.3	0.4	-0.8	-0.5
modelo 2	1	0.05	-0.455	-0.7	0.65
modelo 3	1	1.5	0.54	-0.9	-0.6
modelo 4	1	-0.05	-0.765	0.9	-0.85
modelo 5	1	1.35	0.455	-0.7	-0.65
modelo 6	1	1.85	0.855	-0.9	-0.95
modelo 7	1	2.5	1	-0.5	-2
modelo 8	1	-0.5	-1.04	-0.8	1.3
modelo 9	1	2	0.99	-0.9	-1.1
modelo 10	1	2.9	1.8	-0.9	-2
modelo 11	1	3.8	3.45	-2.3	-1.5
modelo 12	1	-0.1	-1.32	-1.1	1.2

Tabla 15.1 Coeficientes de los modelos estudiados y correspondiente localización de los ceros.

La varianza del ruido aditivo gaussiano se reescala atendiendo a la relación señal ruido (SNR) deseada, donde la SNR (en dB) se define como

$$\text{SNR} = 10 \log \frac{\sigma_x^2}{\sigma_v^2} \quad (15.16)$$

Aquí los niveles de ruido considerados serán -20, -15, -10, -5, 0, 5, 10, 15, 20 y 25 dB.

15.5.1 Características de los modelos

Los modelos presentados en la tabla 15.1 se han elegido ex-profeso para cubrir la mayor cantidad de situaciones interesantes posibles. Ya es conocido de análisis llevados a cabo en el capítulo 13 que la localización de los ceros del sistema influye de manera concluyente en la calidad de las soluciones y atendiendo a ella se han clasificado como sensibles y no sensibles. Por tanto, por situaciones interesantes se quiere expresar aquellas que cubren todo el rango de relaciones entre los ceros del sistema. Para cada modelo las relaciones son las siguientes:

Modelo 1 Es un modelo de fase mínima, no sensible ya que los ceros no están relacionados de manera especial.

Modelo 2 Es un modelo de fase mínima, sensible ya que sus ceros son opuestos.

Modelo 3 Es un modelo de fase mínima, sensible porque uno de sus ceros está cerca del círculo unidad.

Modelo 4 Es un modelo de fase mínima, sensible porque sus ceros son el opuesto el uno del otro y están cerca del círculo unidad.

Modelo 5 Es un modelo de fase mínima, sensible porque sus ceros están cerca el uno del otro.

Modelo 6 Es un modelo de fase mínima y es sensible porque ambos ceros están cerca del círculo unidad.

Modelo 7 Es un modelo de fase mixta, sensible porque los ceros son el inverso el uno del otro.

Modelo 8 Es un modelo de fase mixta, sensible porque los ceros son opuestos e inversos.

Modelo 9 Es un modelo de fase mixta, sensible porque los ceros son cercanos al círculo unidad y están próximos el uno del otro.

Modelo 10 Es un modelo de fase mixta, sensible porque uno de sus ceros se encuentra cerca del círculo unidad.

Modelo 11 Es un modelo de fase máxima, no sensible.

Modelo 12 Es un modelo de fase máxima, sensible porque ambos ceros están cerca del círculo unidad.

15.6 Resultados de las simulaciones

Las simulaciones llevadas a cabo en este bloque de la Memoria cubren una ingente cantidad de modelos, ya definidos en la sección 15.5, estudiados mediante 10 técnicas distintas, ICSMA, CP, CPM, MC, AC y las 5 modalidades de ACS_m. Para poder llevar a cabo un análisis de los resultados con la profundidad necesaria es muy conveniente proceder a presentar el comportamiento de las simulaciones al menos en dos grupos y así poder prestar atención más fácilmente a los detalles. El primer grupo estará compuesto por los métodos AC y ACS_m aplicados a la identificación de los 12 modelos definidos. La composición del segundo grupo se detallará más adelante cuando se pueda justificar a la vista de lo expuesto acerca del primer grupo.

15.6.1 Primer grupo. Métodos AC y ACS_m

La necesidad de tratar todas las técnicas que utilizan ajuste por cumulantes (a la salida del canal en estudio) formando un grupo aparte es doble. En primer lugar está claro que forman una categoría por sí mismas ya que se basan en la misma idea. En segundo lugar hay que destacar que aunque la idea de utilizar ajustes por cumulantes no es ni mucho menos nueva [BL95], [CWM97], [CL99], [Swa97], [Tug95], [ZMM93], [ZV98], no se ha llevado a cabo —que se tenga conocimiento de ello— ningún análisis acerca de la idoneidad de usar un cumulante u otro, o todos, en el ajuste. Dicho análisis se desarrollará con el empleo del concepto de MDR y se verá refrendado experimentalmente con los resultados aportados.

El primer punto a tratar antes de proceder al análisis concreto de los resultados es determinar qué factores influyen en ellos. Ya que los métodos empleados constan de varios pasos, parece inevitable que haya que tratar cada uno por separado. Los distintos pasos son:

1. Estimar la secuencia de autocorrelación de la salida del sistema.
2. Calcular los coeficientes de los sistemas espectralmente equivalentes al sistema problema.
3. Decidir la fase empleando las técnicas AC y ACS_m.

El comportamiento que tenga el método en cada uno de los pasos afectará en un aspecto concreto a los resultados obtenidos. Los dos primeros pasos afectarán a la capacidad de obtener buenos estimadores del sistema problema, el tercero afectará a qué técnica en particular dará los mejores resultados. Es decir, los primeros dan la posibilidad de obtener buenos estimadores y el tercero dice qué

metodo va a ser el que de verdad dé esos estimadores. El tercer punto además es capaz de decir no sólo quién va ser el mejor método sino de ordenarlos atendiendo a su comportamiento.

Estimación de la secuencia de autocorrelación

El primer paso tiene que ver exclusivamente con la exactitud con la que se pueda estimar la secuencia de autocorrelación de la salida de un sistema dado. Según se puso de manifiesto en la sección 15.1 depende principalmente de dos factores:

1. La SNR o equivalentemente el nivel de ruido aditivo. Cuanto mayor sea la potencia del ruido más difícil será la estimación, es decir, mayor varianza tendrán los estimadores.
2. Cuanto mayor sean los coeficientes del filtro a identificar mayor será la varianza de los estimadores.

Aún hay un tercer factor que interviene en la calidad de la estimación, el número de datos. Es directo de ver que al aumentar el número de muestras conocidas mejorarán los estimadores de la parte derecha de (15.5) acercándose a su valor ideal, mejorando por tanto el estimador de $\hat{r}_{xx}(m)$. Todos los modelos están sujetos a los mismos valores de la SNR y al número de datos, por lo que lo único que es propio de cada modelo es el factor 2 y en función de él se hará la discusión.

El límite a partir del cual se considerará un coeficiente como alto se ha deducido experimentalmente, siendo por tanto un dato empírico. Este límite está en torno a 1.5 y en función de él se pueden clasificar atendiendo a las siguientes categorías:

Modelos de riesgo bajo: todos sus coeficientes son bastante menores que el límite.

En esta memoria los modelos 2, 4 y 8 pertenecen a esta categoría.

Modelos de riesgo medio: alguno de sus coeficientes se encuentra en el entorno del límite, como los modelos 1, 3, 5 y 12.

Modelos de riesgo alto: todos sus coeficientes son mayores que el límite. En esta Memoria los modelos 6, 7, 9, 10 y 11 pertenecen a esta categoría.

Hay que tener en cuenta que los modelos estudiados aquí están normalizados de manera que el primer coeficiente siempre es la unidad. Son el resto de coeficientes los que se utilizan en la lista anterior para determinar la categoría a la que pertenecen.

El término ‘riesgo’ aparece por los motivos que se van a dar en el próximo apartado.

Cálculo de los sistemas espectralmente equivalentes

Una vez obtenida la secuencia de autocorrelación se procede a calcular los coeficientes de los filtros de los sistemas que son espectralmente equivalentes al sistema a identificar. Para ello se utilizan las expresiones (13.5), (13.9) y (13.10). Dos factores intervienen aquí de forma decisiva: cómo de buenos sean los estimadores de la secuencia de autocorrelación y cómo de sensibles son estas expresiones para el modelo en particular.

Si el modelo no es sensible dará buenos estimadores de los coeficientes aún cuando no lo sean tanto los de la secuencia de autocorrelación. Sin embargo, un modelo sensible depende mucho de la calidad de estos últimos, si son buenos el resultado será bueno, si no lo son, el resultado tampoco lo será. Recordar que el factor que controlaba la calidad de los estimadores de la secuencia de autocorrelación es el tamaño de los coeficientes del filtro. La tabla 15.2 resume esta idea e indica los modelos que reúnen las características especificadas.

	Riesgo bajo	Riesgo medio	Riesgo alto
No sensible (modelos)	Buena (-)	Buena (1)	Buena (11)
Sensibles (modelos)	Buena (2, 4 y 8)	Regular (3, 5 y 12)	Mala (6, 7, 9 y 10)

Tabla 15.2 *Calidad de los estimadores según la sensibilidad y el riesgo*

Un modelo con bajo riesgo tendrá asociados estimadores de los coeficientes de buena calidad mientras que los de alto riesgo tienen más riesgo de que sea mala, de ahí su calificativo. Modelos con buena calidad asociada permiten a las técnicas de elección de fase llevar a cabo su misión con plena libertad, siendo ya tarea de esta última el elegirla adecuadamente. Si la técnica funciona adecuadamente la MSD final será muy pequeña. Esto se puede comprobar observando por ejemplo los modelos 2, 4 y 8 trabajando con 1000 muestras en figuras 15.12, 15.20 y 15.36 donde las técnicas que mejor funcionan dan MSD rondando e incluso mejorando los -20 dB, cuando lo normal es que estén por los -15 dB en el resto de modelos.

Los modelos con mala calidad asociada impiden la correcta elección de la fase, con lo que todos los métodos dan igual de malos resultados, existiendo muy

poca dispersión entre ellos. Veáanse por ejemplo las figuras 15.28, 15.32, 15.40 y 15.44 correspondientes a los modelos 6, 7, 9 y 10 donde las MSD rondan los -10 dB.

Para modelos con calidad regular (modelos 3, 5 y 12 en figuras 15.16, 15.24 y 15.52) los MSD se mueven en la región intermedia de los -15 dB y la dispersión entre los métodos es un poco mayor.

Hasta ahora todo lo discutido permite decir con qué modelo se pueden obtener los mejores resultados atendiendo a características tales como la sensibilidad y el riesgo, que son factores propios de cada uno. La influencia de factores externos tales como la SNR y el número de datos es clara y conocida. Tanto si aumenta la SNR como el número de muestras disponibles, se tendrán mejores estimadores y la MSD será más pequeña.

Idoneidad de AC y ACS_m al modelo identificado

El tercer paso de los métodos estudiados consiste en elegir la fase correcta entre todos los sistemas espectralmente equivalentes, mediante el uso de técnicas de ajuste por cumulantes. En los comentarios realizados en relación a los pasos precedentes se ha justificado porqué se obtienen mejores resultados para unos modelos, pero no se ha dicho nada sobre qué técnica es la que consigue esos resultados óptimos. Este punto se cubrirá en este apartado.

Ya es conocido que si el número de datos es muy pequeño los estimadores de los sistemas espectralmente equivalentes tendrán mucho sesgo y varianza con lo que las técnicas de elección de fase no podrán trabajar adecuadamente. Conforme aumenta el número de datos los estimadores van mejorando y las diversas técnicas pueden ir mostrando sus virtudes y sus defectos. Parece por tanto importante hacer la discusión en función del número de datos.

Para **100 muestras** el comportamiento observado es muy similar para todas las técnicas, debido al hecho ya comentado de que los estimadores de los coeficientes de los sistemas espectralmente equivalentes son muy pobres, y no permiten que éstas funcionen adecuadamente. Veáanse las figuras 15.7, 15.11, 15.15, 15.19, 15.23, 15.27, 15.31, 15.35, 15.39, 15.43, 15.47 y 15.51 correspondientes a todos los modelos para este número de muestras.

Para **1000 muestras** el comportamiento es mucho más rico en matices y se puede observar cómo según los modelos que se estudien las técnicas que dan mejores resultados son unas u otras. Para un modelo dado, la magnitud que controla cómo se ordenan las diversas técnicas en función del MSD es la MDR. Tal y como se explicó cuando se definió, la MDR da una idea de la dificultad que presenta un determinado cumulante para distinguir los sistemas dentro del

conjunto espectralmente equivalente. Cuanto menor es, más dificultad encuentra ese determinado cumulante para elegir el sistema con la fase correcta. Por tanto, es de esperar que, dado un sistema para identificar, el cumulante que tenga la MRD más grande asociada a ese sistema dé los mejores resultados. En general, van ordenados de menor a mayor MSD según va decreciendo la MDR. Este comportamiento tipo se ve matizado atendiendo a cómo de diferentes sean los valores de la MDR y si superan cierto valor umbral:

1. Si la MDR para todos los cumulantes utilizados (incluida AC) es menor que cierto valor límite, ninguna técnica debe funcionar bien porque le resulta muy difícil distinguir entre los sistemas. Este valor límite o umbral se ha determinado experimentalmente y resulta ser 0.1. Un ejemplo claro es el modelo 9; según se aprecia en la figura 15.3 donde se muestra la inversa de la MDR (en dB) para los distintos modelos estudiados, el modelo 9 presenta valores superiores a los 10 dB (valor umbral) para todas las técnicas y en consecuencia la MSD es grande según muestra la figura 15.40. Se ha optado por representar la inversa de la MDR para que haya una relación más directa: a menor inversa de la MRD menor MSD.
2. Si algunos cumulantes presentan MDR por debajo del valor umbral y otros por encima, está claro que los segundos darán un MSD inferior a los primeros. Se puede considerar como una situación típica para aplicar MRD. Véase por ejemplo la figura 15.36 correspondiente al modelo 8 donde se aprecia que AC, AC₋₁ y AC₁ presentan mejores resultados que AC₋₂, AC₂ y AC₀, en clara concordancia con lo esperado al observar la figura 15.3 de la inversa de MDR donde el primer grupo de cumulantes tiene valores por debajo de 10 dB y el segundo por encima.
Este comportamiento tipo también lo presenta el modelo 11 en la figura 15.48.
3. Si todos los cumulantes tiene una MDR mayor que 0.1 todos están en disposición de dar buenos resultados. En este caso la manera en que se ordenan no está dictada por la MDR sino por otros posibles factores como la estimación de los cumulantes. En esta situación están los modelos 2, 4 y 12. Según la figura 15.3 en todos ellos las técnicas AC₋₁ y AC₁ deberían dar los mejores resultados seguidas de AC y finalmente AC₋₂ y AC₂, pero al ser $h_1 \sim 0$ la estimación de $c_{4x}(1, 1, 1)$ y $c_{4x}(-1, -1, -1)$ es más sensible y por tanto las técnicas que los usan dan los peores resultados según se aprecia en las figuras 15.12, 15.20 y 15.52.

La figura 15.3 donde se muestran los valores de la inversa de la MDR para los distintos modelos merece unos comentarios ex profeso. Estos valores se han calculado suponiendo que se conocían los verdaderos coeficientes del filtro a identificar. A partir de ellos se calcularon todos los sistemas espectralmente equivalentes usando las expresiones de la tabla 13.2. Una vez conocido el conjunto espectralmente equivalente, para calcular la MDR para un cumulante se

calculaba dicho cumulante para todos los sistemas del conjunto y se procedía tal y como explica la definición de la MDR.

Además en esta misma figura se puede comprobar cómo se verifican las propiedades de la MDR expuestas en la proposición 15.2.1, ya que las parejas $c_{4x}(1, 1, 1) - c_{4x}(-1, -1, -1)$ y $c_{4x}(2, 2, 2) - c_{4x}(-2, -2, -2)$ presentan cada una el mismo valor. Por otro lado, ya que $\text{MDR}(c_{4x}(0, 0, 0))=0$ su inversa es ∞ y por tanto no aparece en la gráfica. Algo parecido le ocurre al modelo 7 donde al tener la simetría $h_0 = h_2$ la MDR para todos los cumulantes es idénticamente cero.

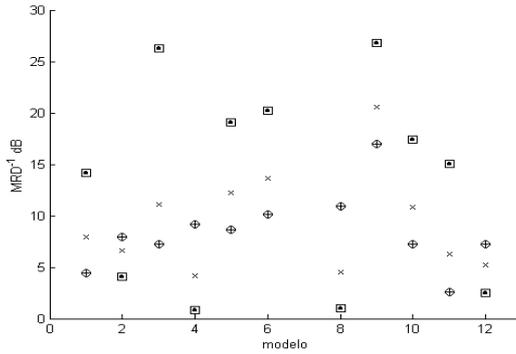


Figura 15.3 MDR para los modelos estudiados
 o ACS_2; □ ACS_1; ◇ ACS_0; ● ACS_1; + ACS_2; x AC

Para **10000 muestras** cabe destacar el hecho de que hay algunos modelos que siguen mejor el comportamiento dictado por la MDR, y que no lo hacían para 1000 datos. El comportamiento dictado por la MDR se puede resumir en los 3 puntos dados anteriormente más el hecho de que las parejas de técnicas ya conocidas deben dar resultados iguales. Los modelos que cumplen lo aquí comentado son el 1, 3, 5 y 10 representados respectivamente en las figuras 15.9, 15.17, 15.25 y 15.45.

A la vista de lo expuesto hasta ahora para el comportamiento de las distintas técnicas en función del número de datos, sería posible extrapolar una pauta de comportamiento aplicable a todos los modelos:

Número de datos pequeño $0 - N_1$: Todos los métodos dan resultados insatisfactorios.

Número de datos intermedio $N_1 - N_2$: El comportamiento está gobernado por la MDR.

Número de datos grande $N_2 \rightarrow \infty$: Todos los métodos se comportan igual de bien. Comportamiento asintótico.

El principal problema es que N_1 y N_2 dependen del modelo tal y como se ha puesto de manifiesto anteriormente. Obviamente la calidad de los estimadores, tanto de los sistemas espectralmente equivalentes como de los cumulantes empleados en la identificación, son pieza fundamental en los intervalos descritos. Si la MDR es grande y muy diferente para los distintos cumulantes, el comportamiento gobernado por ella es de esperar que se dé para número de datos pequeño. El factor clave es el cociente 'error de estimación'/MDR. Sin embargo estimar N_1 y N_2 parece una tarea muy compleja.

Modelos que siguen el comportamiento aquí descrito en el rango de muestras estudiado son el 4 y el 10, para ello ver las series de figuras 15.19-15.22 y 15.43-15.46.

Finalmente, las simulaciones realizadas con **100000 muestras** presentan, como era de esperar, un comportamiento próximo al asintótico para muchos modelos, tales como el 2, 4, 8 y 12, y casi el 10. Para ver los detalles consultar las figuras 15.14, 15.22, 15.38 y 15.54 respectivamente.

El modelo 7 y la Diferencia Relativa Promedio (DRP): Como ya se apuntó con anterioridad el modelo 7 presenta ciertas peculiaridades consecuencia de que $h_0 = h_2$. La primera de ellas es que el conjunto de sistemas espectralmente equivalentes disminuye el número de sus elementos (ver tabla 13.2), o dicho de otro modo, uno de sus elementos tiene multiplicidad dos, en particular el sistema que estamos buscando. No importa por tanto que $MDR(ACS_0)=0$ tal y como se ha calculado aquí. De hecho el método ACS_0 funciona para este modelo. La definición de MDR, aun siendo muy sencilla, se ha mostrado muy eficaz para describir los resultados obtenidos y eso la justifica en último extremo. Considérese sin embargo los casos hipotéticos siguientes:

Caso 1: Sea un conjunto de 3 sistemas espectralmente equivalentes, cuyos cumulantes para cierta diagonal toman los valores 1, 2 y 4 respectivamente. La MDR en este caso es 1.

caso 2: Lo mismo que en el caso 1 pero con valores 1, 3 y 4. La MDR es 1/3.

Si en ambos casos el sistema correcto es el primero (el que tiene un cumulante igual a 1), será más fácil de identificar en el caso 2, porque se diferencia mejor del resto. Sin embargo, la MDR es menor aquí que el primer caso. Se puede afirmar que la MDR se pone en el peor de las circunstancias suponiendo que el cumulante del sistema a identificar se encuentra en la zona más densa (más cercano al resto de cumulantes). Quizás, en vez de con el mínimo se debería definir con el promedio: Diferencia Relativa Promedio (DRP). Es una exigencia teórica, puesto que MDR funciona estupendamente.

En la figura 15.4 se muestra la DRP para los modelos estudiados. Una desventaja de DRP es que no es cero para ACS_0, incluso para algunos modelos como el 2 toma valores relativamente altos que pueden llegar a la falsa conclusión de que pudiesen funcionar. Sin embargo se ha comentado previamente que esta técnica no funciona salvo para modelos con $h_0 = h_2$ como el modelo 7.

El orden en que DRP ordena las técnicas coincide con el de MDR, sólo difieren en los valores concretos que toma. El valor límite para DRP no se puede 'definir' experimentalmente con tanta facilidad como para MDR porque no coincide para los distintos modelos, un valor aproximado es 0.18.

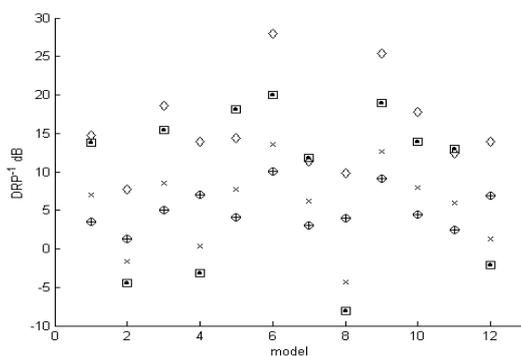


Figura 15.4 DRP para los modelos estudiados
 o ACS_-2; □ ACS_-1; ◇ ACS_0; ● ACS_1; + ACS_2; x AC

En cuanto a la ventaja obvia que se espera conseguir al definir la DRP, la obtención de información acerca de modelos con simetrías como el 7, se puede afirmar lo siguiente. La magnitud DRP no refleja que la técnica ACS_0 cuente con ventaja respecto al resto debido a la doble multiplicidad del sistema a identificar, de hecho es la técnica que mejor funciona, junto con AC que siempre se beneficia de sus técnicas derivadas. Una vez aclarado este punto, tal y como prevee DRP, las técnicas ACS_1 y ACS_-1 se comportan peor que ACS_2 y ACS_-2. Este comportamiento se observa para número de muestras alto, ver figura 15.33 para 10000 muestras.

En estos comentarios se ha puesto de manifiesto la gran cantidad de factores que intervienen a la hora de explicar el resultado de las simulaciones. Se puede resumir gráficamente tal y como muestra la figura 15.5. En ella MSD_m denota la MSD mínima que se puede conseguir para un determinado modelo, depende de la calidad de los estimadores según se detalla en la tabla 15.2.

Según ponen de manifiesto los valores de MDR para AC, los resultados dados por esta técnica por regla general ni son los mejores ni los peores, teniendo un

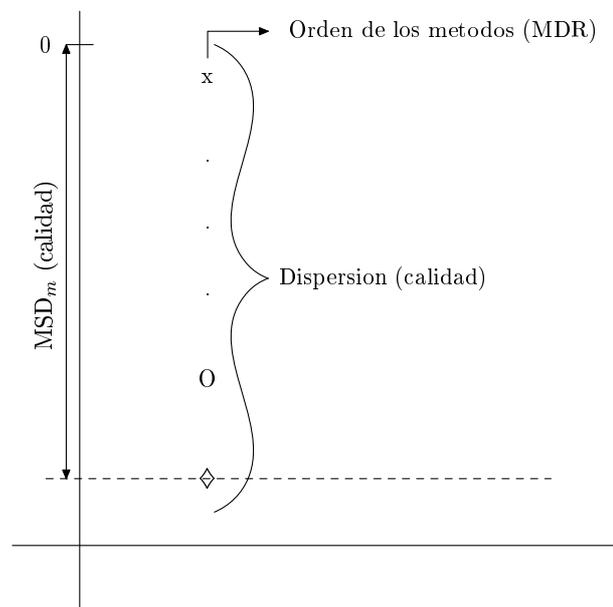


Figura 15.5 Resumen gráfico de la justificación de las simulaciones

comportamiento medio. Es extraño, sin embargo, que técnicas que utilizan menos información que ella, como las ACS_m puedan dar mejores resultados. Un ejemplo claro es el modelo 3 en las figuras 15.17 y 15.18.

15.6.2 Segundo grupo. Comparación general

Los métodos mostrados en este segundo grupo intentan dar una visión general de todos los estudiados mediante simulaciones. Con el fin de no cargar demasiado las gráficas se dibujarán sólo los métodos típicos que se puedan considerar como representativos de una clase o subconjunto. Distintos criterios se han seguido para elegir cada representante:

1. Ya es conocido mediante análisis teórico y comprobación experimental que los métodos ACS_m y ACS_{-m} tienden a comportarse de manera similar. Para cada pareja se ha elegido como representante en cada situación ACS_{Rm} aquel que dé el mejor resultado.
2. El método ACS₀ no es capaz de elegir la fase correcta salvo que se dé una simetría tal y como la que presenta el modelo 7, por tanto sólo para él se considerará dicho método.

3. El método AC ha sido empleado con asiduidad en la literatura especializada y dada su importancia se tendrá en cuenta en este segundo grupo de simulaciones.
4. La técnica ICSMA forma por sí sola una categoría y por tanto hay que considerarla.
5. Experimentalmente se ha comprobado que todas las técnicas basadas en el ecualizador se comportan igual. Un par de ejemplos característicos pueden observarse en las gráficas de la figura 15.55, para el resto de situaciones y modelos la tendencia es la misma. El método CPM ha sido elegido como representante de esta clase, denominada a partir de ahora ECU.

Respecto al punto 5 es posible hacer un comentario interesante. Sólo en SNR cercanas a 0 dB se ha observado una pequeña desviación en cuanto al comportamiento señalado, ya que entonces el método MC funciona ligeramente peor que sus relacionados CP y CPM (ver figura 15.56) . Se puede interpretar que para estas situaciones críticas da mejor resultado hacer medidas puramente estadísticas que basadas exclusivamente en la geometría de la señal ecualizada.

En definitiva son 5 los métodos a analizar y comparar en este apartado: ACS_R1, ACS_R2, AC, ICSMA y ECU. Ellos darán una idea del comportamiento de los 10 métodos descritos en el apartado 15.3. En las figuras 15.57-15.104 se muestran todos los resultados obtenidos para los 12 modelos estudiados, SNR desde -20 a 25 dB y número de muestras 100, 1000, 10000 y 100000.

Aspectos generales

Todo lo relativo a los métodos AC y ACS está comentado extensa y exhaustivamente en el apartado 15.6.1 anterior y no se volverá a tratar aquí. Aparecen otra vez en este apartado simplemente por motivos de comparación y poder dar una visión general con respecto a las otras técnicas empleadas. Cabe destacar que los métodos ECU y ICSMA no parecen tener tanta variedad de comportamiento como los primeros.

Recuérdese asimismo las pautas seguidas generalmente en lo que respecta a SNR y número de muestras. A mayor SNR y más número de muestras la MSD disminuye. Este comportamiento también se ha observado aquí.

Dado que los métodos empleados comparten los primeros pasos, a saber, estimación de la secuencia de autocorrelación y estimación a partir de ella de los coeficientes de los filtros espectralmente equivalentes, aspectos ya realizados del análisis de los métodos AC y ACS también son aplicables aquí. Es decir, la MSD mínima alcanzable va a depender de la calidad asociada a cada modelo, dada en la tabla 15.2, siendo el método particular utilizado el responsable de

alcanzar esa cota mínima de MSD si funciona adecuadamente. Los aspectos concretos de las técnicas ECU e ICSMA se detallan en los siguientes apartados. Notar que aunque ICSMA no necesite el cálculo de los sistemas espectralmente equivalentes, las expresiones que utiliza se basan directamente en ellas por lo que aspectos como la sensibilidad del modelo le afectan de lleno, y por tanto conceptos derivados de ella como es la calidad asociada también le son aplicables.

Análisis de la técnica ICSMA

Por regla general este método se comporta de manera muy satisfactoria, estando cerca o igualando la MSD dada por el mejor de todos los métodos para cada caso. Para un número de muestras intermedio se comporta mejor que AC, pero nunca supera al mejor de los ACS. Es posible que se den algunas situaciones críticas en las que este método tenga problemas para funcionar de manera adecuada. Recuértese que el núcleo de ICSMA consistía en hacer dos decisiones binarias, elegir un signo en función de medidas basadas en HOS. Resulta lógico pensar que estimar dicho signo va a ser más difícil conforme se acerque más a cero. Los signos a estimar son:

Signo primero: Decide el valor de \hat{h}_1 . Esta variable toma el valor correcto, es decir, h_1 si el signo $s_1 = \text{signo}[h_1^2 - (h_0 + h_2)^2]$ está bien estimado. La cantidad entre corchetes se acerca a cero si $h_1^2 \sim (h_0 + h_2)^2$, siendo una situación de peligro ‘relativo’. Es un peligro relativo porque si no se estima bien el signo, \hat{h}_1 toma el valor $h_0 + h_2$, que es la otra opción para su sistema espectralmente equivalente (ver tabla 13.2), pero si $h_1^2 \sim (h_0 + h_2)^2$ no hay mucha diferencia entre un valor y otro.

Signo segundo: Se corresponde con la elección del valor de \hat{h}_0 , el signo asociado es $s_2 = \text{signo}[h_1(h_0 - h_2)]$. Una vez que \hat{h}_1 ha sido bien estimado esta segunda decisión escoge h_0 o h_2 como estimador de \hat{h}_0 . Aquí se puede alcanzar una situación crítica si $h_1 \sim 0$ o $h_0 \sim h_2$. En el segundo caso es una relación de peligro relativo tal y como sucede para el signo primero. El primer caso sí es una situación real crítica y puede afectar seriamente el comportamiento de este método.

Modelos que cumplan las condiciones anteriores son, para el signo primero el 1, 3, 5, 6, 9, 10 y 12 (figuras 15.59, 15.67, 15.75, 15.79, 15.91, 15.95 y 15.103 respectivamente); para el signo segundo caso 1 el 6, 7 y 9 (figuras 15.79, 15.83 y 15.91) y para el caso 2 el 2, 4 y 12 (figuras 15.63, 15.71 y 15.103). Como se puede observar en las gráficas, sólo para los modelos 2, 4 y 12 ICSMA funciona realmente mal tal y como se comentó en los dos puntos anteriores.

El modelo 2 es un caso típico de lo expuesto, sin embargo en el modelo 4 no se nota tanto el efecto de la situación crítica puesto que la dispersión en

los métodos es más pequeña, de todas maneras resulta claro que ICSMA es uno de los peores métodos para este modelo. Finalmente para el modelo 12 el efecto es menor todavía, esto se debe principalmente a que h_1 toma ya valores ligeramente superiores a los de los modelos anteriores.

El comportamiento observado para ICSMA se explica con estas sencillas reglas salvo para el modelo 11 donde da peores resultados de lo esperado. En principio no debería tener problemas ya que h_1 no es cercano a cero. Una posible explicación radica en el hecho de que este modelo es de alto riesgo y por tanto la estadística que haya que estimar relacionada con él sufrirá de mayores varianzas. Esta varianza influye negativamente en el signo (estimado mediante estadística de cuarto orden), pudiendo llegar incluso a no estimarlo correctamente.

Técnicas basadas en el ecualizador

Este conjunto de técnicas, representadas en esta sección por ECU, se diferencian del resto principalmente en que requieren un paso adicional antes de llevar a cabo lo que es realmente la elección de la fase correcta. Una vez calculados los sistemas espectralmente equivalentes es necesario calcular la inversa de cada uno de ellos. Lo que parece claro es que esta nueva operación incrementará el riesgo de que los errores se vayan acumulando con respecto a los pasos anteriores – de manera que estos métodos son los que dan resultados más insatisfactorios– y que estas técnicas sean, por tanto, muy sensibles a la calidad asociada a cada modelo. De hecho, modelos con buena calidad se identifican satisfactoriamente mediante ECU, y los modelos de calidad regular o mala se identifican mal mediante esta técnica; esto se puede comprobar observando las figuras 15.57-15.104 correspondientes a este apartado y comprobando la calidad asociada a cada modelo dada en la tabla 15.2.

Hay, sin embargo, 3 modelos que se salen de esta tendencia general. Los modelos 1 y 11 dan resultados insatisfactorios aun siendo de calidad buena y el modelo 12, de calidad regular, da resultados satisfactorios. El comportamiento inesperado de este último se puede explicar aduciendo que, en verdad, los módulos de sus coeficientes no son tan grandes, y por tanto se le puede considerar casi en la frontera de los métodos con calidad buena asociada.

La explicación para el comportamiento en los modelos 1 y 11 es un poco más elaborada y necesita tratar cómo se han calculado los ecualizadores de los sistemas dados. Los ecualizadores se han calculado como simples sistemas inversos, con una parte no causal asociada a los ceros de módulo mayor que 1 del sistema original y otra causal con el resto de los ceros; los detalles del método se encuentran en [ZMM93] pero básicamente se intenta que en N puntos del círculo unidad la distancia entre el sistema inverso calculado y el sistema inverso real sea la menor posible en el sentido de mínimos cuadrados. Esto es necesario porque

normalmente la inversa de un filtro de respuesta finita es de respuesta infinita, la forma de implementarlo aquí es truncando el filtro de respuesta infinita a partir del punto en el que los coeficientes sean menores que cierto parámetro de calidad. Los N puntos estudiados del círculo unidad se eligen al azar de forma aleatoria. Pues bien, se puede comprobar experimentalmente que para los modelos 1 y 11 la inversa así calculada no elimina la interferencia intersimbólica IIS, es decir no ecualiza al canal a identificar, todo lo bien que sería deseable. Empíricamente se ha encontrado que si la IIS es mayor que unos -10 dB para un cierto modelo, los métodos ECU no funcionarán adecuadamente. Para el 1 y el 11 su IIS se sitúa entorno a ese límite por lo que no se espera que den resultados satisfactorios. Nótese que para los modelos donde estos métodos sí funcionan la IIS ronda los -30 dB. En la figura 15.6 se muestra la IIS para todos los modelos estudiados. Esta IIS se ha calculado de la siguiente manera: supuesto conocido el canal a identificar se calcula su inversa mediante el método anteriormente descrito y se componen ambos filtros. Del filtro resultante se calcula su coeficiente de módulo mayor y se le sustrae. La sumatoria de los coeficientes restantes al cuadrado constituye la IIS. Estos valores se han obtenido suponiendo que se conocen exactamente los parámetros del modelo, si se estimasen dichos parámetros la IIS aumentará debido al ruido de estimación con lo que se reforzaría la explicación aquí presentada.

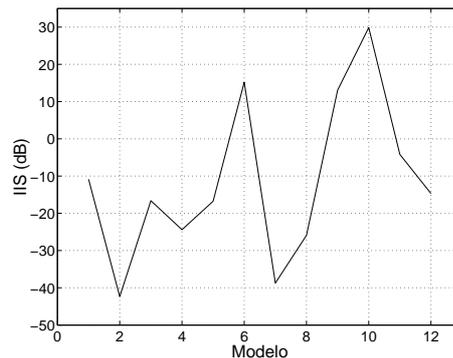


Figura 15.6 Interferencia intersimbólica IIS para los modelos estudiados

Acerca del modelo 7: Para este modelo, además de los métodos anteriores, se ha representado el método ACS₀ que como se sabe no sólo no funciona para él, sino que presenta ventaja respecto a otros métodos, siendo de hecho uno de los que mejor comportamiento presenta. Para este modelo ECU da muy buenos resultados (tan bueno como el mejor de los métodos) cuando no debería ser así ya que es de calidad mala asociada. Esto es debido a que la degeneración que presenta en el conjunto EE le beneficia ya que tiene doble opción de elegir correctamente el sistema que da la mejor señal ecualizada, ya no interviene sólo la calidad.

Respecto al método ICSMA cabe decir que sigue los comportamientos generales dados anteriormente. La degeneración no le beneficia más que en la elección del segundo signo, en la del primero se ve afectado por ser un modelo de riesgo alto y por tanto se comporta mejor conforme aumenta el número de muestras, como es habitual en él.

Gráficas del primer grupo. Métodos AC y ACS

Leyenda: O Método ACS₋₂; cuadrado Método ACS₋₁; ◇ Método ACS₀; · Método ACS₁; + Método ACS₂ y x Método AC.

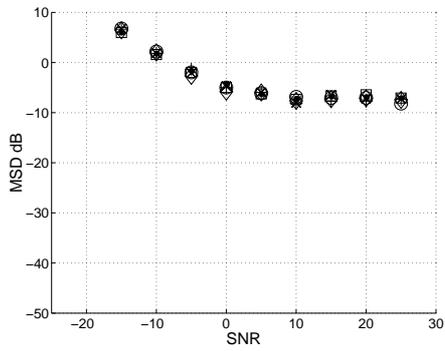


Fig. 15.7 Modelo 1. 100 datos

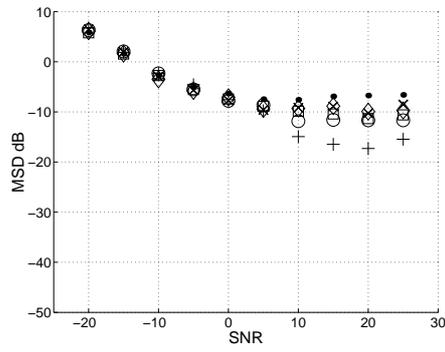


Fig. 15.8 Modelo 1. 1000 datos

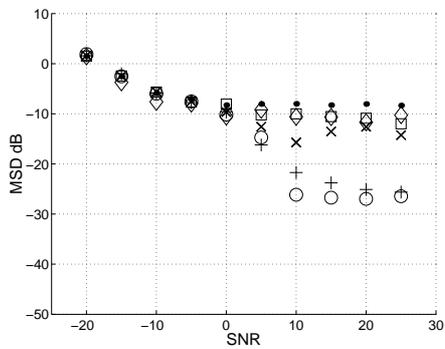


Fig. 15.9 Modelo 1. 10000 datos

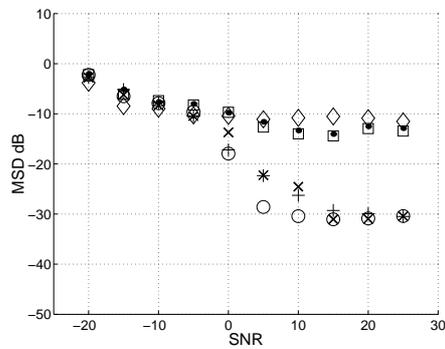


Fig. 15.10 Modelo 1. 100000 datos

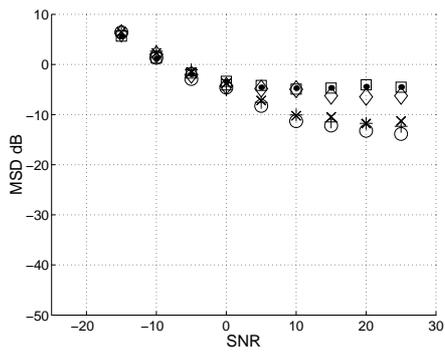


Fig. 15.11 Modelo 2. 100 datos

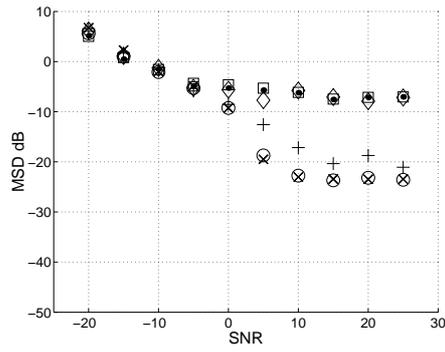


Fig. 15.12 Modelo 2. 1000 datos

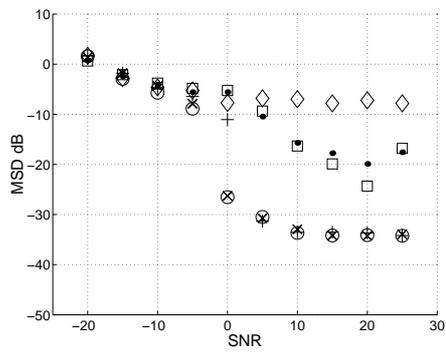


Fig. 15.13 Modelo 2. 10000 datos

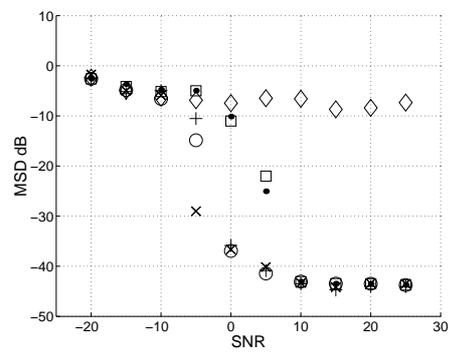


Fig. 15.14 Modelo 2. 100000 datos

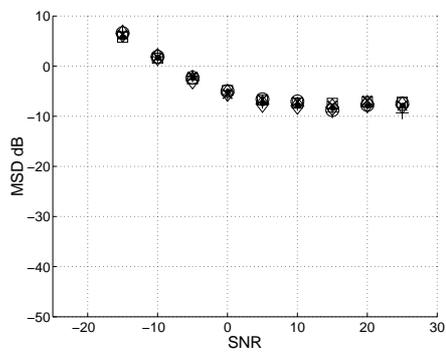


Fig. 15.15 Modelo 3. 100 datos

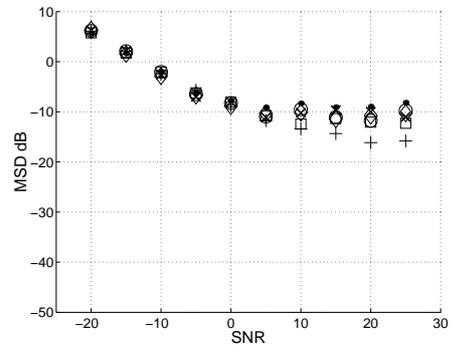


Fig. 15.16 Modelo 3. 1000 datos

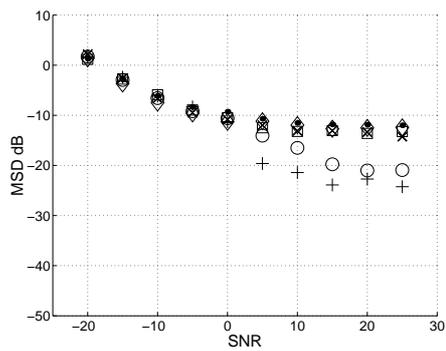


Fig. 15.17 Modelo 3. 10000 datos

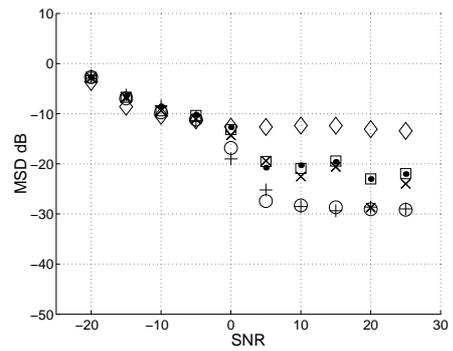


Fig. 15.18 Modelo 3. 100000 datos

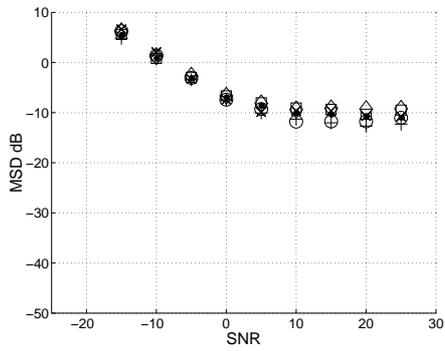


Fig. 15.19 Modelo 4. 100 datos

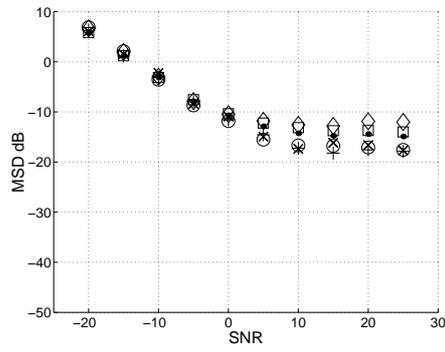


Fig. 15.20 Modelo 4. 1000 datos

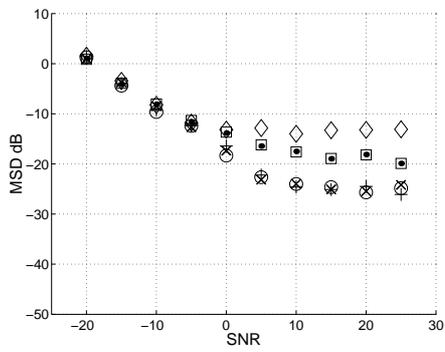


Fig. 15.21 Modelo 4. 10000 datos

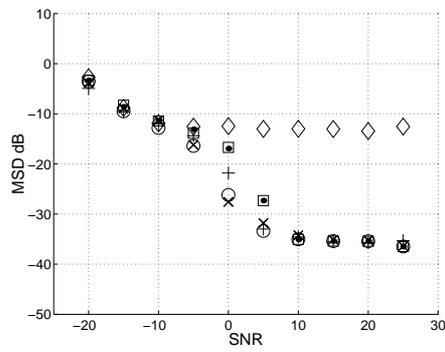


Fig. 15.22 Modelo 4. 100000 datos

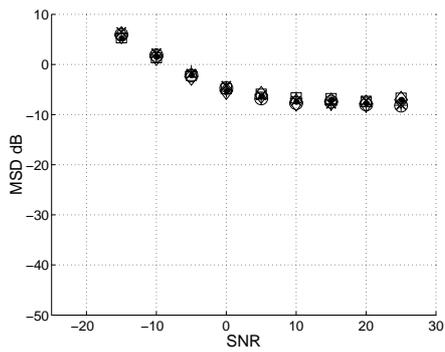


Fig. 15.23 Modelo 5. 100 datos

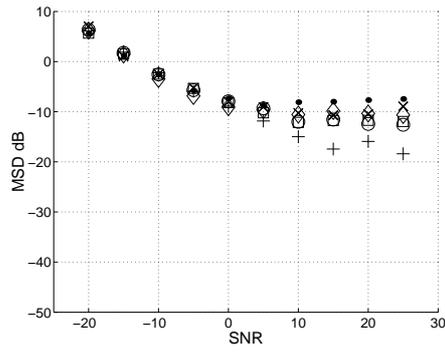


Fig. 15.24 Modelo 5. 1000 datos

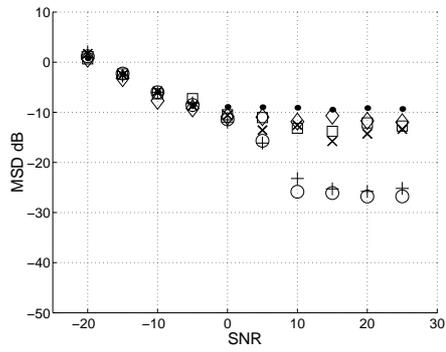


Fig. 15.25 Modelo 5. 10000 datos

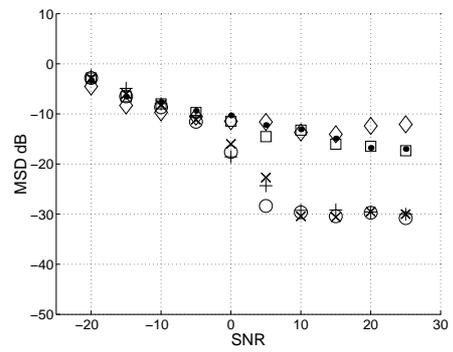


Fig. 15.26 Modelo 5. 100000 datos

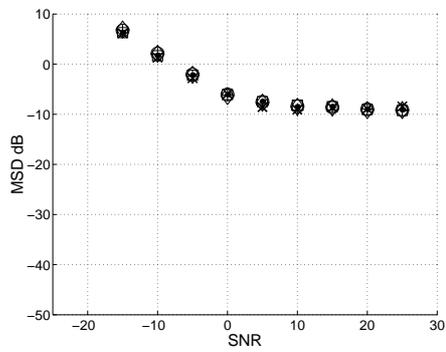


Fig. 15.27 Modelo 6. 100 datos

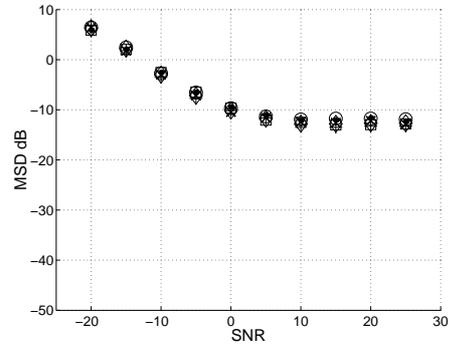


Fig. 15.28 Modelo 6. 1000 datos

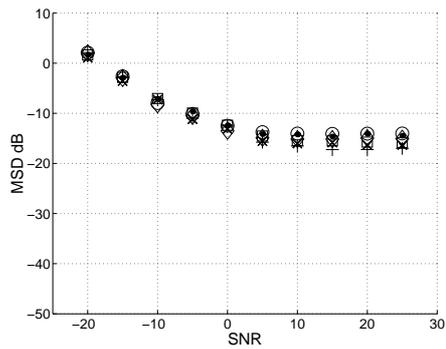


Fig. 15.29 Modelo 6. 10000 datos

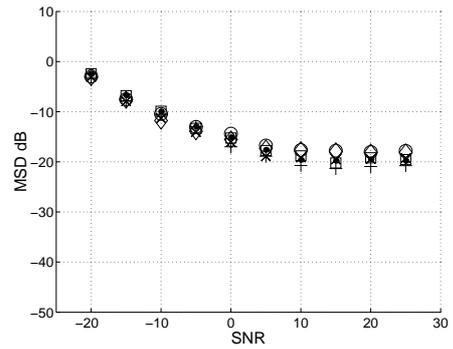


Fig. 15.30 Modelo 6. 100000 datos

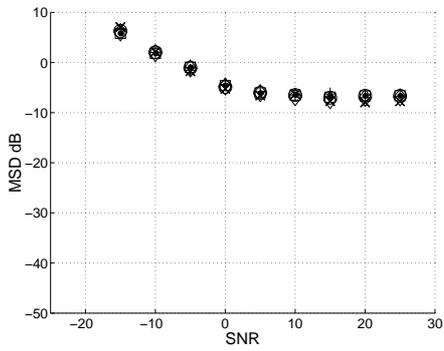


Fig. 15.31 Modelo 7. 100 datos

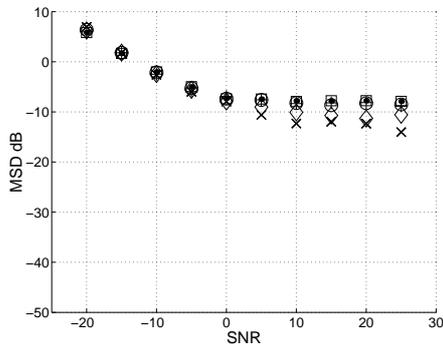


Fig. 15.32 Modelo 7. 1000 datos

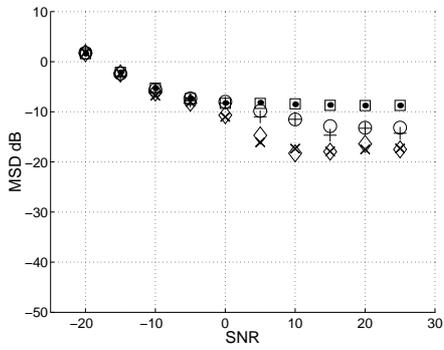


Fig. 15.33 Modelo 7. 10000 datos

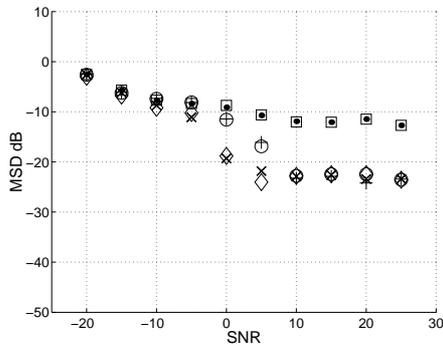


Fig. 15.34 Modelo 7. 100000 datos

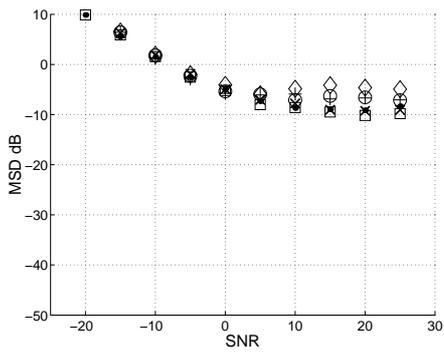


Fig. 15.35 Modelo 8. 100 datos

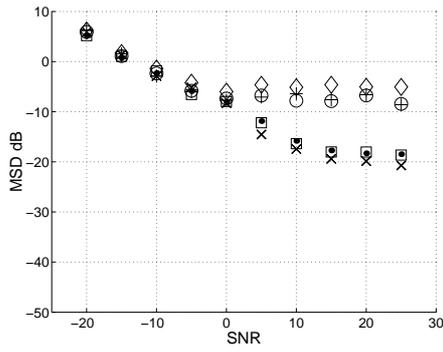


Fig. 15.36 Modelo 8. 1000 datos

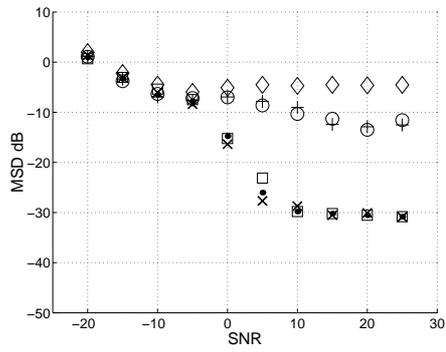


Fig. 15.37 Modelo 8. 10000 datos

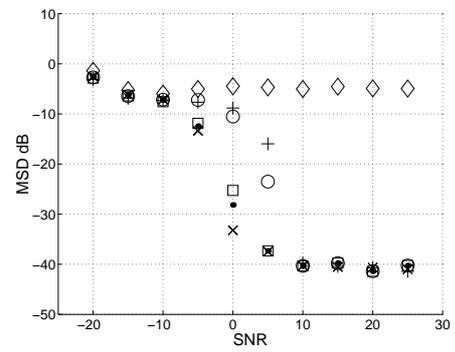


Fig. 15.38 Modelo 8. 100000 datos

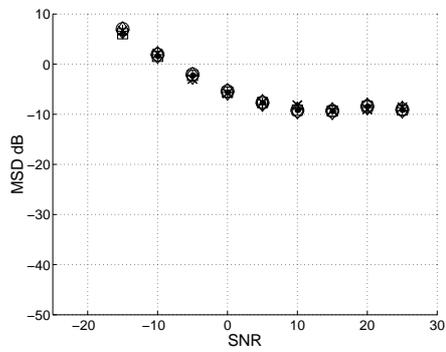


Fig. 15.39 Modelo 9. 100 datos

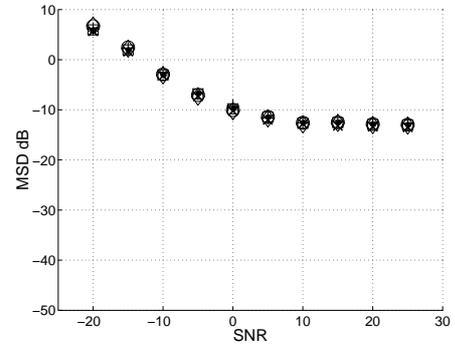


Fig. 15.40 Modelo 9. 1000 datos

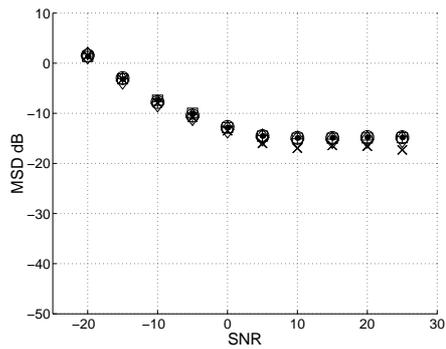


Fig. 15.41 Modelo 9. 10000 datos

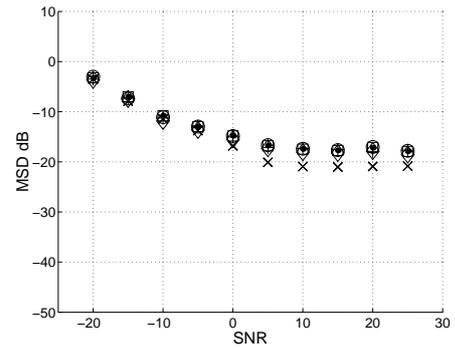


Fig. 15.42 Modelo 9. 100000 datos

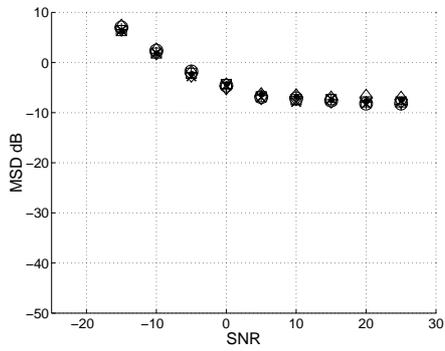


Fig. 15.43 *Modelo 10. 100 datos*

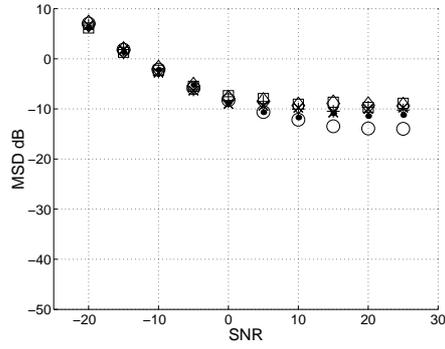


Fig. 15.44 *Modelo 10. 1000 datos*

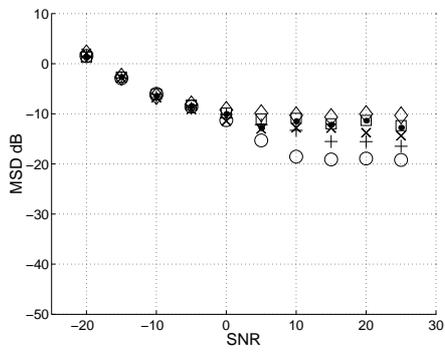


Fig. 15.45 *Modelo 10. 10000 datos*

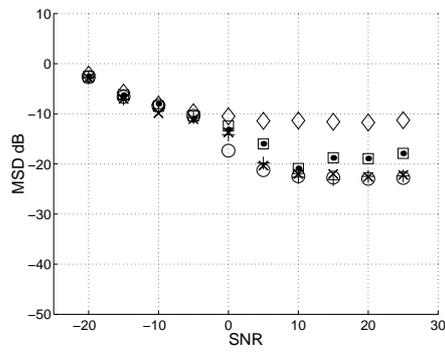


Fig. 15.46 *Modelo 10. 100000 dat.*

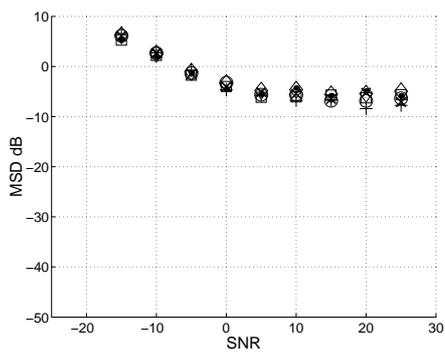


Fig. 15.47 *Modelo 11. 100 datos*

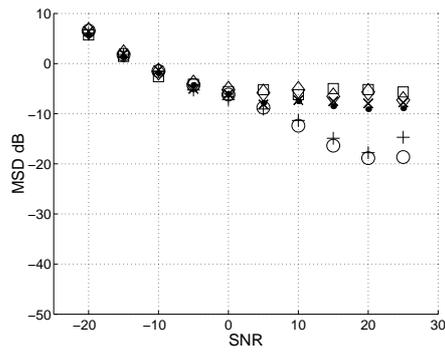


Fig. 15.48 *Modelo 11. 1000 datos*

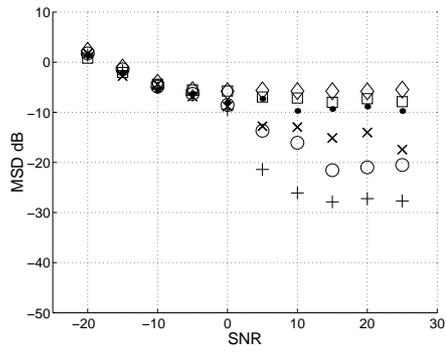


Fig. 15.49 Modelo 11. 10000 datos

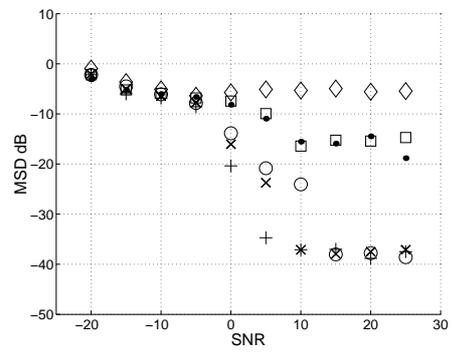


Fig. 15.50 Modelo 11. 100000 dat.

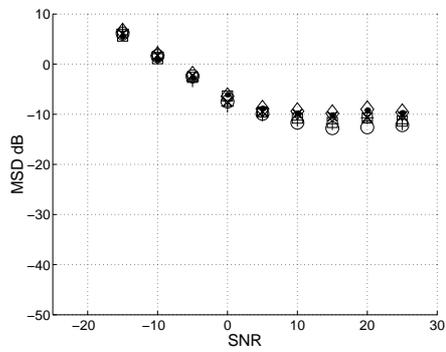


Fig. 15.51 Modelo 12. 100 datos

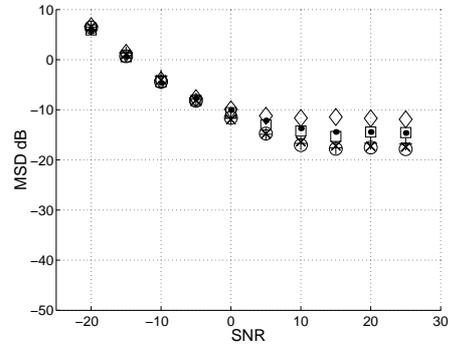


Fig. 15.52 Modelo 12. 1000 datos

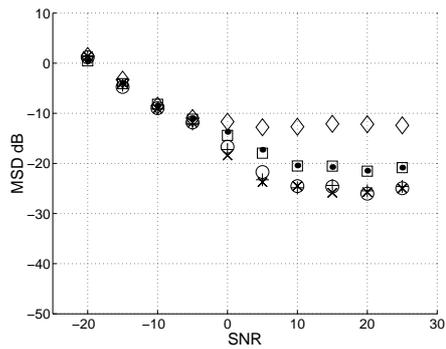


Fig. 15.53 Modelo 12. 10000 datos

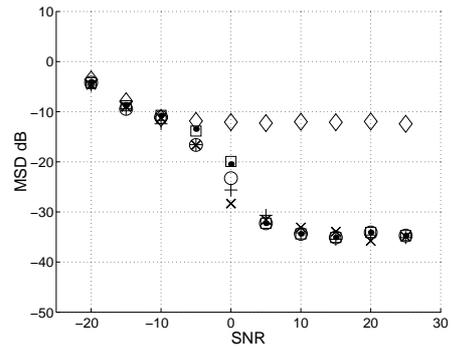


Fig. 15.54 Modelo 12. 100000 dat.

Gráficas de los métodos basados en el ecualizador

Leyenda: ○ Método del módulo constante MC; □ Método de la curtosis próxima CP; ◇ Método de la curtosis próxima modificada CPM.

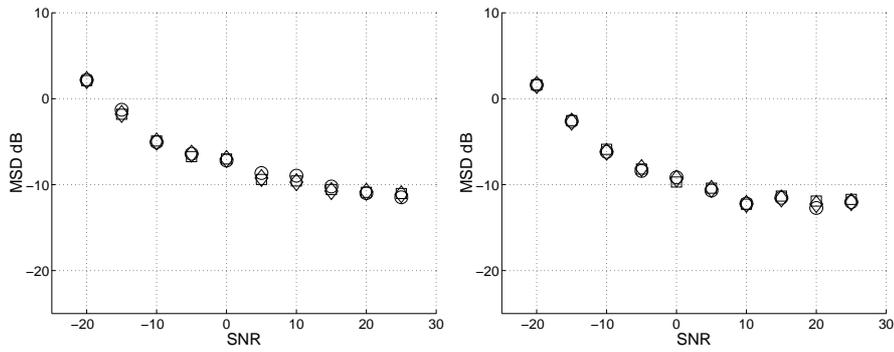


Fig. 15.55 Modelos 1(izdq) y 3(dcha). 10000 datos

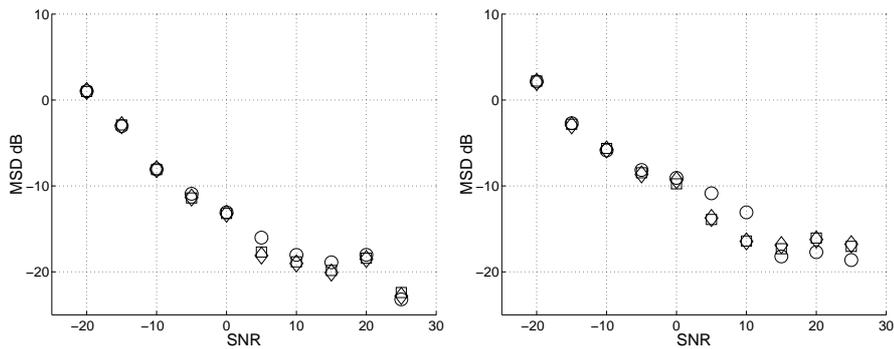


Fig. 15.56 Modelos 4(izqd) y 7(dcha). 10000 datos

Gráficas del 2º grupo. Comparación general

Leyenda: ○ Representante de los métodos ACS₂ y ACS₋₂ ACR₂; □ Representante de los métodos ACS₁ y ACS₋₁ ACR₁; x Método de ajuste por cumulantes AC; * Método ICSMA; ◇ Representante de los métodos basados en el ecualizador ECU.

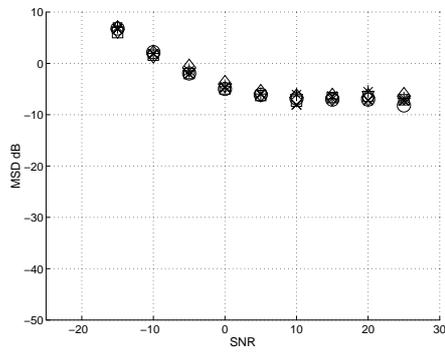


Fig. 15.57 Modelo 1. 100 datos

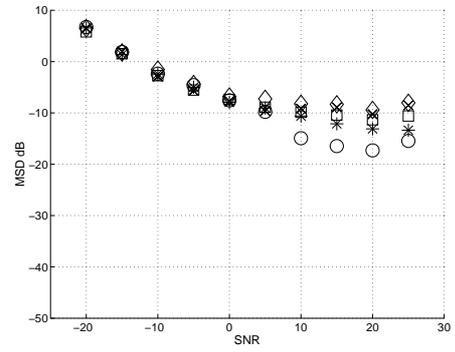


Fig. 15.58 Modelo 1. 1000 datos

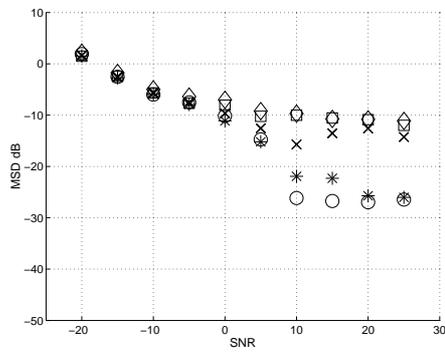


Fig. 15.59 Modelo 1. 10000 datos

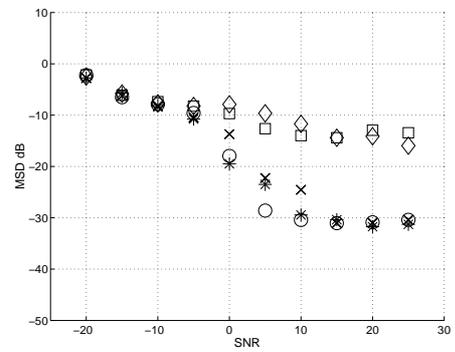


Fig. 15.60 Modelo 1. 100000 datos

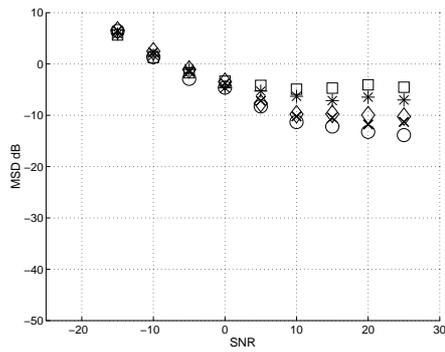


Fig. 15.61 Modelo 2. 100 datos

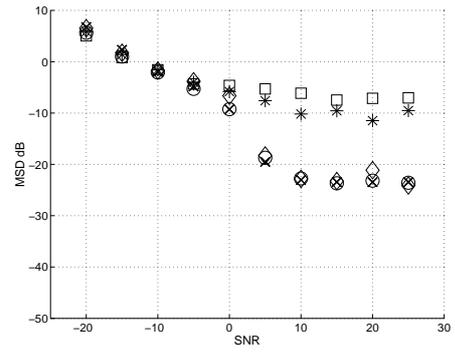


Fig. 15.62 Modelo 2. 1000 datos

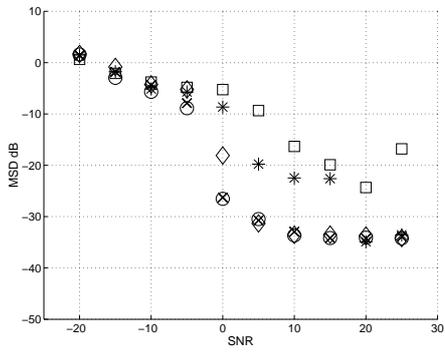


Fig. 15.63 Modelo 2. 10000 datos

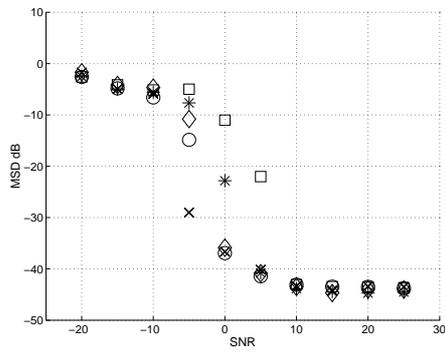


Fig. 15.64 Modelo 2. 100000 datos

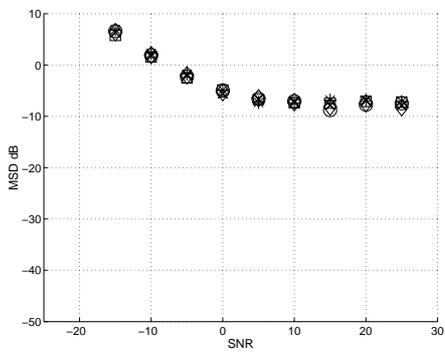


Fig. 15.65 Modelo 3. 100 datos

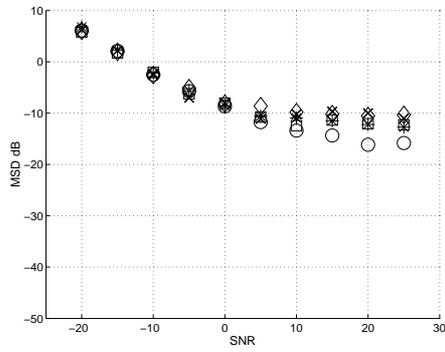


Fig. 15.66 Modelo 3. 1000 datos

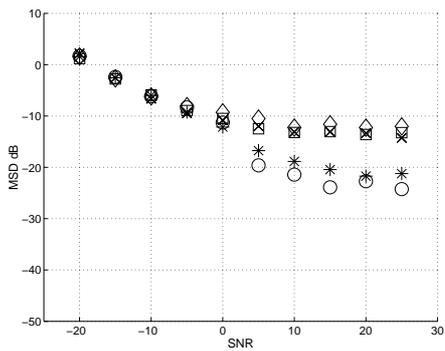


Fig. 15.67 Modelo 3. 10000 datos

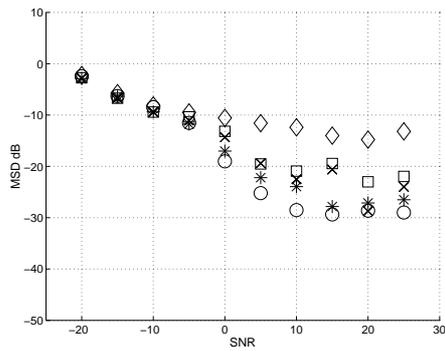


Fig. 15.68 Modelo 3. 100000 datos

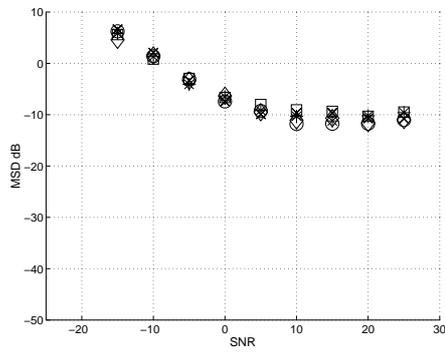


Fig. 15.69 Modelo 4. 100 datos

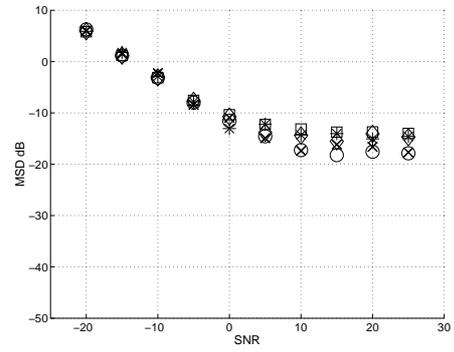


Fig. 15.70 Modelo 4. 1000 datos

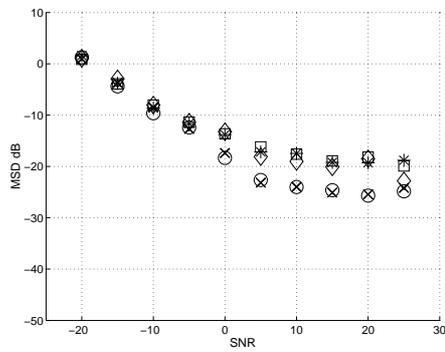


Fig. 15.71 Modelo 4. 10000 datos

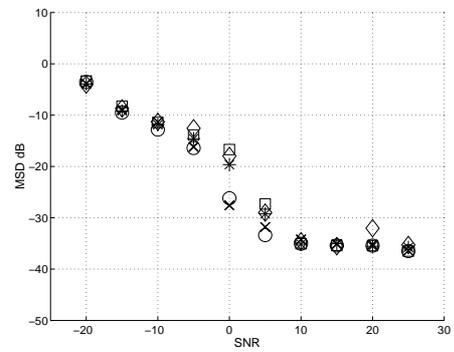


Fig. 15.72 Modelo 4. 100000 datos

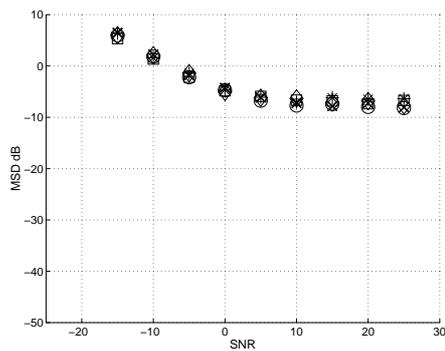


Fig. 15.73 Modelo 5. 100 datos

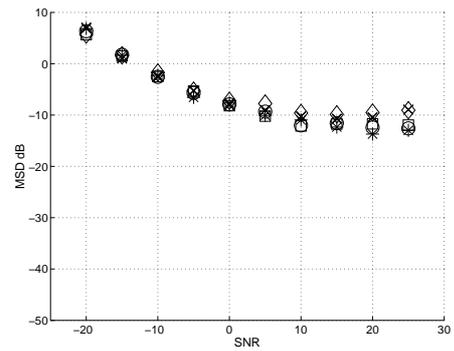


Fig. 15.74 Modelo 5. 1000 datos

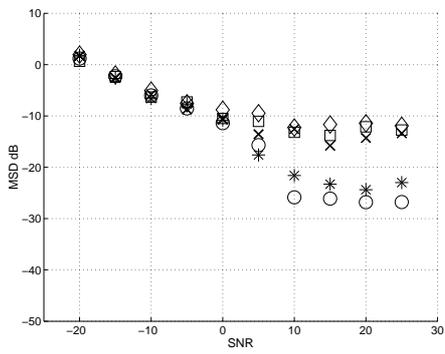


Fig. 15.75 Modelo 5. 10000 datos

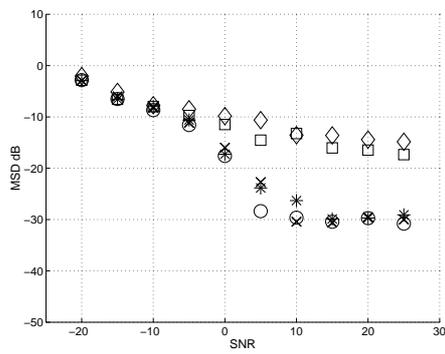


Fig. 15.76 Modelo 5. 100000 datos

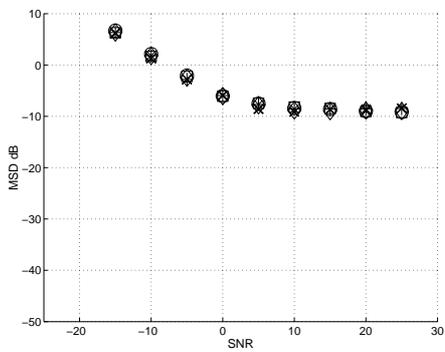


Fig. 15.77 Modelo 6. 100 datos

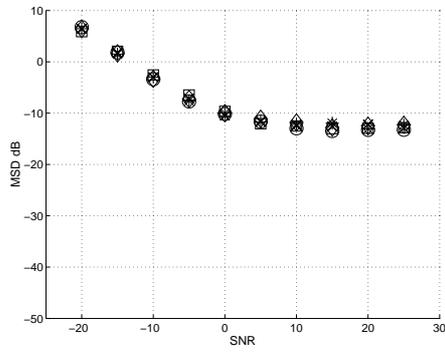


Fig. 15.78 Modelo 6. 1000 datos

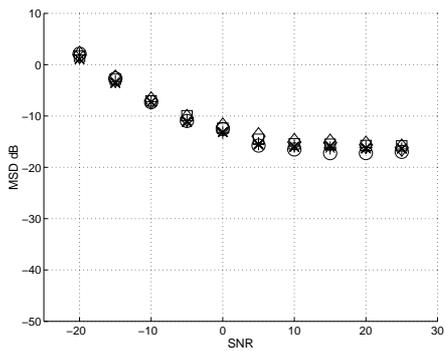


Fig. 15.79 Modelo 6. 10000 datos

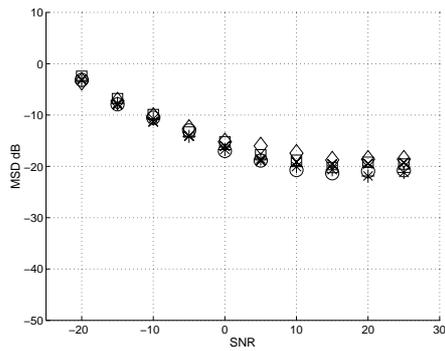


Fig. 15.80 Modelo 6. 100000 datos

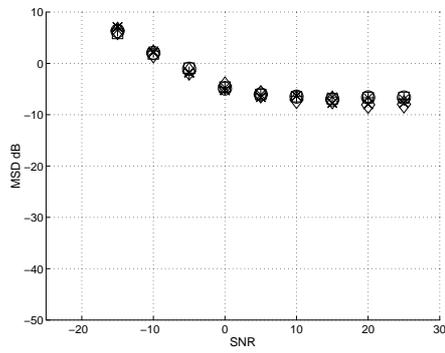


Fig. 15.81 Modelo 7. 100 datos

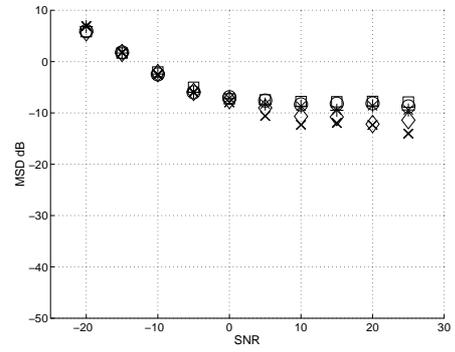


Fig. 15.82 Modelo 7. 1000 datos

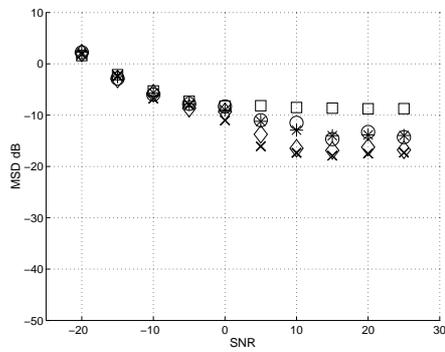


Fig. 15.83 Modelo 7. 10000 datos

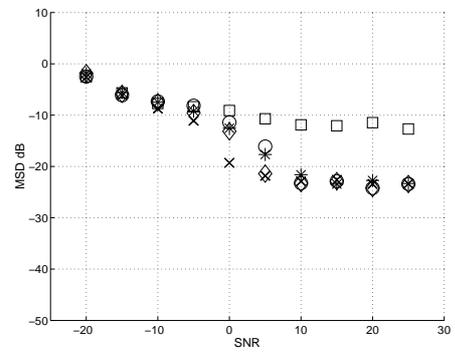


Fig. 15.84 Modelo 7. 100000 datos

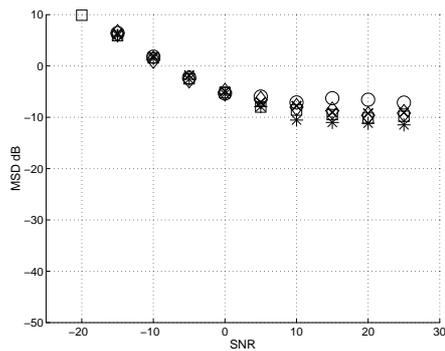


Fig. 15.85 Modelo 8. 100 datos

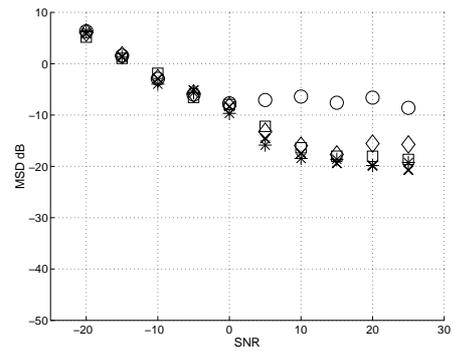


Fig. 15.86 Modelo 8. 1000 datos

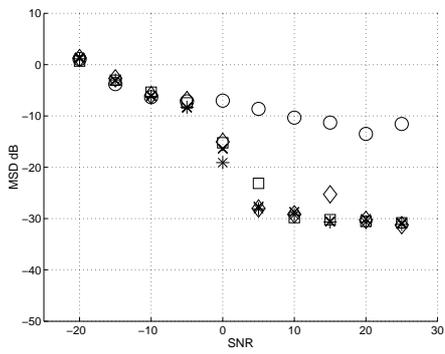


Fig. 15.87 Modelo 8. 10000 datos

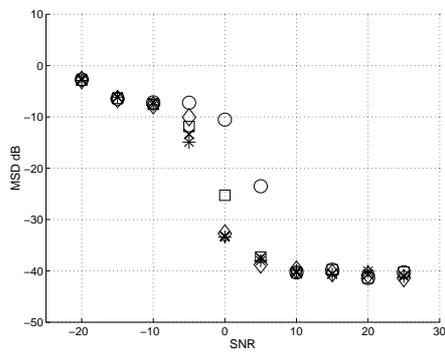


Fig. 15.88 Modelo 8. 100000 datos

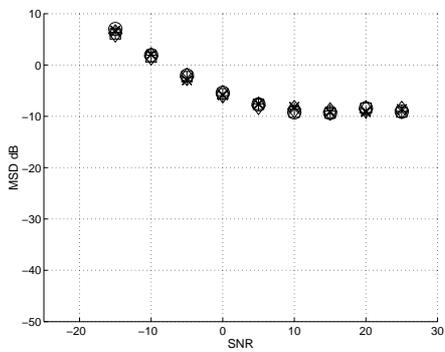


Fig. 15.89 Modelo 9. 100 datos

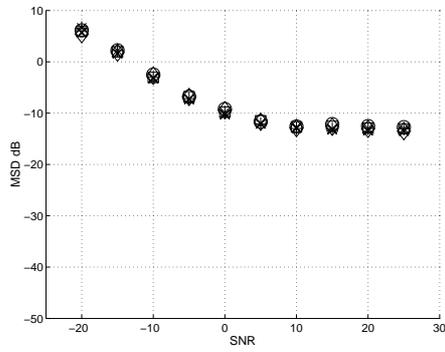


Fig. 15.90 Modelo 9. 1000 datos

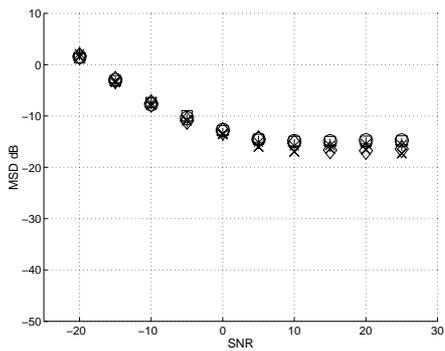


Fig. 15.91 Modelo 9. 10000 datos

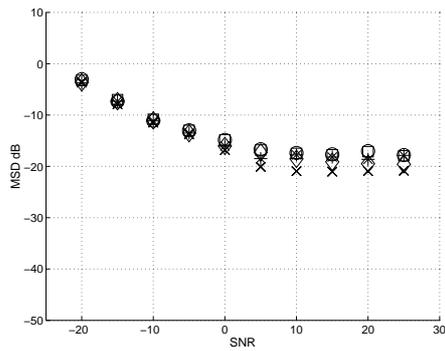


Fig. 15.92 Modelo 9. 100000 datos

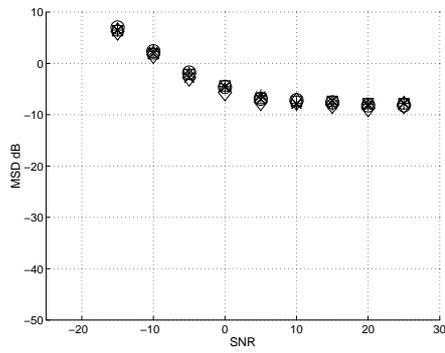


Fig. 15.93 Modelo 10. 100 datos

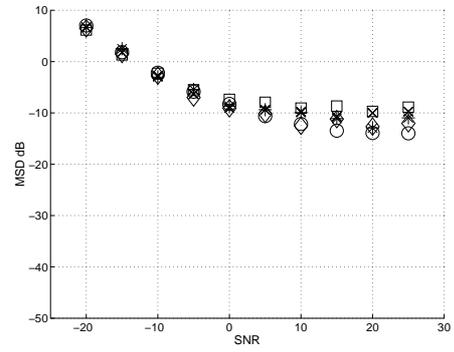


Fig. 15.94 Modelo 10. 1000 datos

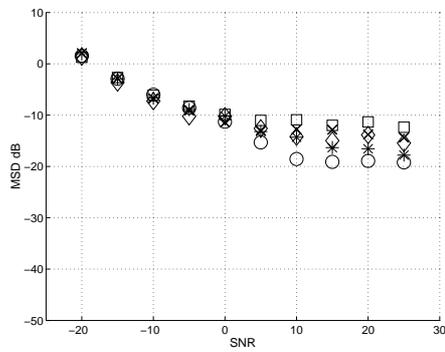


Fig. 15.95 Modelo 10. 10000 datos

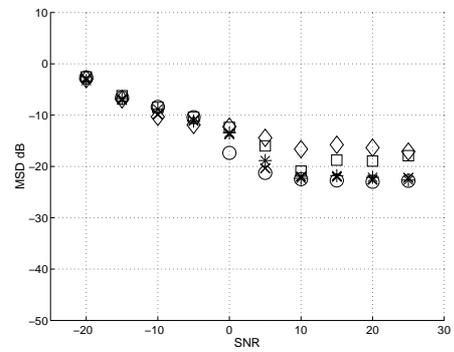


Fig. 15.96 Modelo 10. 100000 dat.

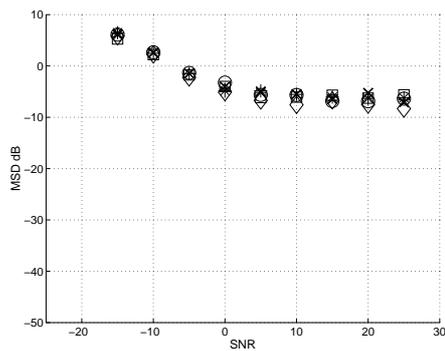


Fig. 15.97 Modelo 11. 100 datos

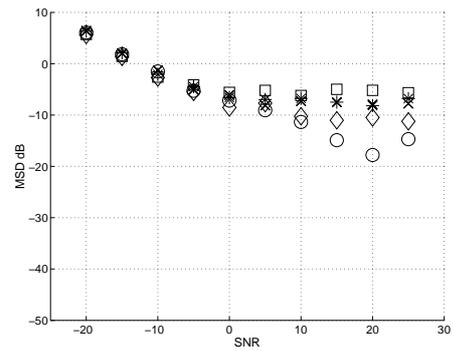


Fig. 15.98 Modelo 11. 1000 datos

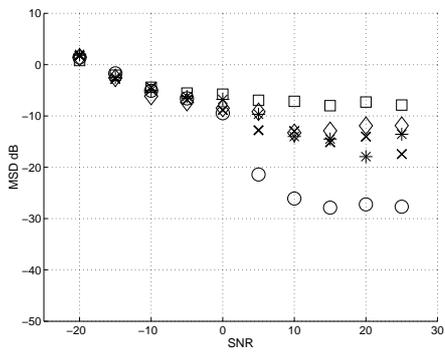


Fig. 15.99 Modelo 11. 10000 datos

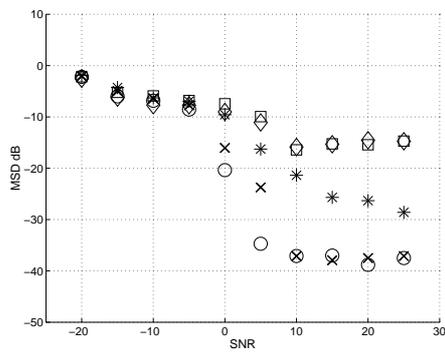


Fig. 15.100 Modelo 11.100000 dat.

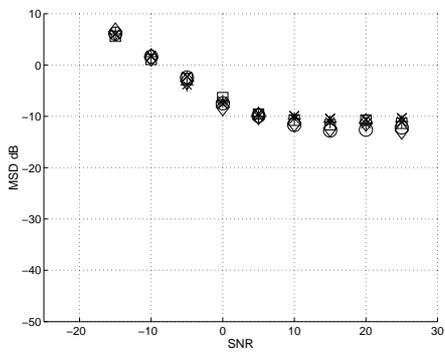


Fig. 15.101 Modelo 12. 100 datos

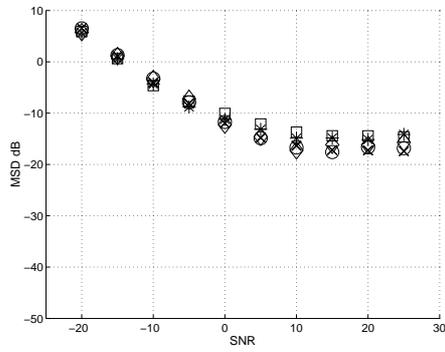


Fig. 15.102 Modelo 12. 1000 datos

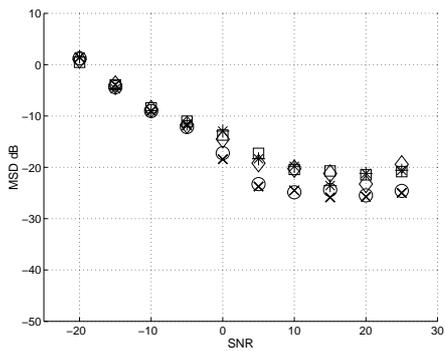


Fig. 15.103 Modelo 12. 10000 dat.

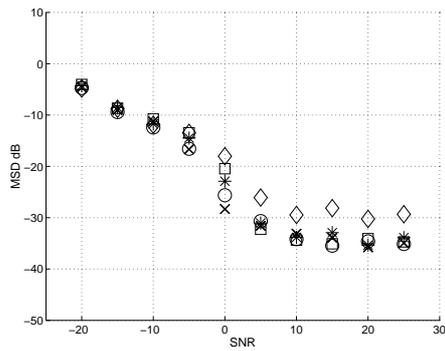


Fig. 15.104 Modelo 12.100000 dat.

Parte IV

Summary in English

Introduction

Many signal processing problems are defined in environments that can be time varying, or require a result immediately after every new data is obtained. In such situations, and in many others, the problem should be solved adaptatively. An adaptive algorithm is capable of improving (updating) the knowledge about the problem in an efficient way for every new sample that arrives.

The first adaptive algorithm to be developed was the Least Mean Square LMS algorithm [WH60] and it imposes an orthogonality condition between the estimation error and the data vector used in the estimation, as required by the orthogonality principle for linear estimation. In this set-up only the second order statistics is used to solve the problem, which is equivalent to solve the Wiener-Hopf equations (a well known linear system of equations).

The LMS lies in the stochastic gradient algorithms family. Generally speaking, the same problem can be solved using the Recursive Least Squares RLS algorithm, but because of its mathematical properties, it gives better results (estimates with lower variance using shorter data records) but with an increased computational cost. The RLS algorithm can be easily generalised to solve problems involving higher-order statistics (i.e. expected values of the product of more than two stochastic variables); this new algorithm is the Recursive Instrumental Variable RIV algorithm. In addition, sometimes the problem we are considering not only involves higher-order statistics but can also be overdetermined, think for instance in the blind identification of Moving Average models using the Giannakis-Mendel equations [GM89]. This situation can also be covered by an algo-

rithm of the Least Squares family: the Overdetermined Recursive Instrumental Variable ORIV algorithm.

Nevertheless, all these situations are not covered by stochastic gradient algorithms. The main goal of chapters 16 and 17 is to fill this gap. It is true that in [AAM96] the derived Generalised LMS algorithm can solve problems involving higher-order statistics, but it suffers from one important drawback: the matrix formed by the linear system of equations must be positive definite or negative definite, limiting the range of applicability. Also, no stochastic gradient algorithm has been designed to solve overdetermined systems of equations with higher-order statistics (the counterpart of ORIV in this family).

To be more precise, chapters 16 and 17 are devoted to seek a new algorithm that can solve problems with associated overdetermined systems of equations involving higher-order statistics, not suffering from the mentioned drawback of GLMS. The ultimate goal is to design an algorithm of the stochastic gradient family (and hence with reduced computational complexity) enjoying as much as the properties of the ORIV algorithm. To reach this, the first step is to theoretically analyse the ORIV algorithm in chapter 16. From this theoretical study, a new orthogonality condition is obtained that is used in chapter 17 to propose a new algorithm meeting all required properties.

Apart from this new orthogonality condition, the theoretical study covers the convergence in the mean and mean square for stationary and not stationary environments including estimation noise. Because good estimators require long memory (big forgetting factor) and tracking requires short memory (small forgetting factor), in average an optimal forgetting factor is found that weights both requirements. An analogous result for the RLS algorithm was found in [MB91].

To test the theoretical study the ORIV algorithm is applied to the blind identification of MA models. For this purpose a new general equation is derived in the time domain that contains as special cases the Giannakis-Mendel equation [GM89] and its generalisation to 3rd and 4th order cumulants [Rui95]. Also, the novel proposed algorithm is tested via simulation in the identification of MA models;

previously its convergence has been briefly theoretically studied.

The final part of this work, chapter 18, is concerned with the blind identification of linear systems (channels) on batch mode. This chapter aims to solve the problem using second order statistics of the output and as little higher-order statistics as possible. The use of higher-order statistics is a must, because only from second order statistics the phase response can not be determined.

A new method is derived that uses almost exclusively 2nd order statistics and uses higher-order statistics to compute just signs (a simple binary decision). This novel method is supposed to have all good properties of second order statistics (need of short data records, low estimate variances, etc.) and higher-order statistics (phase information) and to be only affected in an infinitesimal amount by the drawbacks of higher-order statistics (basically the need for longer data records).

This new method is compared with other techniques that rely on the Spectrally Equivalent Minimum Phase system complemented by a phase selection step (such as cumulant matching, constant modulus of equalised signal, etc. [Swa97], [ZMM93] and [BL95]). All these techniques are analysed to provide a better understanding of their performance.

Capítulo 16

Analysis of the Overdetermined Recursive Instrumental Variable ORIV algorithm. Applications to system identification

Introduction

In this chapter the convergence of the ORIV algorithm is theoretically analysed. The ORIV algorithm can be considered a generalization of the RLS to overdetermined problems involving higher-order statistics. For this reason the orthogonality condition is established between the estimation error and an instrumental variable vector (instead of the data vector). The ORIV algorithm is applied to blind identification of MA systems.

The convergence analysis for stationary systems is presented in section 16.2 where the convergence in the mean and mean square are considered, even in the presence of additive noise. In section 16.3 the environment is supposed to be non-stationary and again the convergence in the mean and mean square are studied. Section 16.4 summarises the previous analysis by deriving the

expression for the forgetting factor that makes the ORIV algorithm attains its best results. Finally, application to blind MA system identification is presented in section 16.5, where the difference between methods based on second and on higher order statistics are discussed.

16.1 The ORIV algorithm

The orthogonality condition between an instrumental variable vector $\tilde{\mathbf{x}}(n)$ and the output error $e(n)$ states that

$$\langle \tilde{\mathbf{x}}(n), e(n) \rangle = 0 \quad (16.1)$$

where \langle, \rangle stands for a scalar product defined for two numerical series like follows:

$$\langle \mathbf{x}(n), \mathbf{y}(n) \rangle = \sum_{i=1}^n \lambda^{n-i} \mathbf{x}(i) \mathbf{y}^t(i)$$

and $e(n) = d(n) - \mathbf{w}^t(n) \mathbf{x}(n)$. The scalar λ is the forgetting factor and takes values between 0 and 1. So (16.1) is equivalent to solve the system of equations:

$$\langle \tilde{\mathbf{x}}(n), \mathbf{x}(n) \rangle \mathbf{w}(n) = \langle \tilde{\mathbf{x}}(n), d(n) \rangle \quad (16.2)$$

where $\tilde{\mathbf{x}}(n) = [\tilde{x}(n), \tilde{x}(n-1), \dots, \tilde{x}(n-l+1)]^t$, $\mathbf{x}(n) = [x(n), x(n-1), \dots, x(n-q+1)]^t$ and the vector of unknowns $\mathbf{w}(n) = [w_1(n), \dots, w_q(n)]^t$. So knowing the series $\{x(n)\}$, $\{\tilde{x}(n)\}$ y $\{d(n)\}$ we have to compute the unknown vector $\{\mathbf{w}(n)\}$.

Equation (16.2) can be restated as:

$$\Phi(n) \mathbf{w}(n) = \mathbf{z}(n) \quad (16.3)$$

where the definition of each term is clear if it is compared with (16.2). This system is overdetermined because always $l > q$ is considered. The solution is obtained minimizing the squared error $\|\Phi(n) \mathbf{w}(n) - \mathbf{z}(n)\|^2$:

$$\mathbf{w}(n) = (\Phi^t(n) \Phi(n))^{-1} \Phi(n)^t \mathbf{z}(n) \quad (16.4)$$

The ORIV algorithm computes $\mathbf{w}(n)$, assuming that $\mathbf{w}(n-1)$, $x(n)$, $\tilde{x}(n)$ and $d(n)$ are known, in the way shown in Table 16.1. For a complete derivation of ORIV see [Fri84]. The update formula of $\mathbf{w}(n)$ has the same general form of the Recursive Least Squares RLS algorithm or the Recursive Instrumental Variable RIV algorithm, that is, a gain vector/matrix times the a priori error. The soft-constrained initialization is used in Table 16.1.

<p style="text-align: center;"><u>Initial Conditions</u></p> $\Phi^t(0) = \delta[\mathbf{I}_{q \times q} \mathbf{0}_{q \times l-q}]$ $\Gamma(0) = \frac{1}{\delta^2} \mathbf{I}_{q \times q}$ <p style="text-align: center;">$\delta =$ escalar arbitrario</p> $\mathbf{w}(0) = \mathbf{0}$ $\mathbf{z}(0) = \mathbf{0}$ <p style="text-align: center;"><u>Body of the algorithm</u></p> $\Phi(n) = \lambda \Phi(n-1) + \tilde{\mathbf{x}}(n) \mathbf{x}^t(n)$ $\mathbf{X}(n) = [\Phi^t(n-1) \tilde{\mathbf{x}}(n) \quad \mathbf{x}(n)]$ $\lambda^2 \mathbf{\Lambda}(n) = \begin{pmatrix} -\tilde{\mathbf{x}}^t(n) \tilde{\mathbf{x}}(n) & \lambda \\ \lambda & 0 \end{pmatrix}$ $\mathbf{K}(n) = \Gamma^{-1}(n-1) \mathbf{X}(n) [\lambda^2 \mathbf{\Lambda}(n) + \mathbf{X}^t(n) \Gamma^{-1}(n-1) \mathbf{X}(n)]^{-1}$ $\Gamma^{-1}(n) = \frac{1}{\lambda^2} [\Gamma^{-1}(n-1) - \mathbf{K}(n) \mathbf{X}^t(n) \Gamma^{-1}(n-1)]$ $\mathbf{v}(n) = \begin{pmatrix} \tilde{\mathbf{x}}^t(n) \mathbf{z}(n-1) \\ d(n) \end{pmatrix}$ $\boldsymbol{\alpha}(n) = \mathbf{v}(n) - \mathbf{X}^t(n) \mathbf{w}(n-1)$ $\mathbf{z}(n) = \lambda \mathbf{z}(n-1) + \tilde{\mathbf{x}}(n) d(n)$ $\mathbf{w}(n) = \mathbf{w}(n-1) + \mathbf{K}(n) \boldsymbol{\alpha}(n)$

Tabla 16.1 Iterative procedure for the ORIV algorithm

16.2 Convergence analysis for stationary systems

16.2.1 Convergence in the mean

The data series will be supposed to be ergodic, so temporal averages could be changed for ensemble averages. Under this condition, one can accept that for $n \rightarrow \infty$:

$$\Phi(n) \sim \frac{\mathbf{R}_{\tilde{\mathbf{x}}\mathbf{x}}}{1-\lambda} + \mathbf{R}_{\tilde{\mathbf{x}}\mathbf{x}}^e(n) \quad (16.5)$$

$$\mathbf{z}(n) \sim \frac{\mathbf{r}_{\tilde{\mathbf{x}}d}}{1-\lambda} + \mathbf{r}_{\tilde{\mathbf{x}}d}^e(n) \quad (16.6)$$

where $\mathbf{R}_{\tilde{x}x} = \mathbb{E}[\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)]$ can be considered the deterministic part of this temporal average and $\mathbf{R}_x^e(n)$ is the matrix of errors in the estimation, the estimation noise. Analogously, $\mathbf{r}_{d\tilde{x}} = \mathbb{E}[\tilde{\mathbf{x}}(n)d(n)]$ and $\mathbf{r}_{d\tilde{x}}^e(n)$ is the estimation noise. With this decomposition of the temporal averages, eq. (16.4) takes the form, for large enough n :

$$\begin{aligned} \mathbf{w}(n) \sim & [\mathbf{R}_{xx}^t \mathbf{R}_{\tilde{x}x} + (1 - \lambda) \mathbf{R}_{\tilde{x}x}^t \mathbf{R}_{\tilde{x}x}^e(n) + (1 - \lambda) \mathbf{R}_{xx}^{et}(n) \mathbf{R}_{\tilde{x}x} + \\ & + (1 - \lambda)^2 \mathbf{R}^{et}(n) \mathbf{R}^e(n)]^{-1} [\mathbf{R}_{xx}^t \mathbf{r}_{\tilde{x}d} + (1 - \lambda) \mathbf{R}_{\tilde{x}x}^t \mathbf{r}_{\tilde{x}d}^e(n) + \\ & + (1 - \lambda) \mathbf{R}_{\tilde{x}x}^{et}(n) \mathbf{r}_{\tilde{x}d} + (1 - \lambda)^2 \mathbf{R}^{et}(n) \mathbf{r}^e(n)] \quad (16.7) \end{aligned}$$

From now on subindexes will be dropped where they are no necessary. Like λ is near 1 the factor $(1 - \lambda)$ will be a perturbation parameter. In addition, the matrix $\Phi(n)$ will be considered as quasideterministic [EF86]. With these assumptions, the inverse that appears in the right hand side of (16.7) can be approximated, so (16.7) results in

$$\begin{aligned} \mathbf{w}(n) \sim & [\mathbf{I} - (1 - \lambda)(\mathbf{R}^t \mathbf{R})^{-1} \mathbf{R}^t \mathbf{R}^e(n) - (1 - \lambda)(\mathbf{R}^t \mathbf{R})^{-1} \mathbf{R}^{et}(n) \mathbf{R}](\mathbf{R}^t \mathbf{R})^{-1} \\ & [\mathbf{R}^t \mathbf{r} + (1 - \lambda) \mathbf{R}^t \mathbf{r}^e(n) + (1 - \lambda) \mathbf{R}^{et}(n) \mathbf{r}] \quad (16.8) \end{aligned}$$

Working in first order in $(1 - \lambda)$ and taking expected values:

$$\mathbb{E}[\mathbf{w}(n)] \sim (\mathbf{R}^t \mathbf{R})^{-1} \mathbf{R}^t \mathbf{r} = \mathbf{w}_0 \quad (16.9)$$

where \mathbf{w}_0 is the solution to $\mathbf{R} \mathbf{w}_0 = \mathbf{r}$, the vector we are looking for. There is one important remark to make: $\mathbf{w}(n)$ verifies that $\langle \tilde{\mathbf{x}}(n), e(n) \rangle$ is minimum (possibly zero) $\forall n$ in LS by construction, nevertheless $\langle \tilde{\mathbf{x}}(n), e_0(n) \rangle \neq 0$ because the expected value operator $\mathbb{E}[\cdot]$ is needed:

$$\mathbb{E}[\tilde{\mathbf{x}}(n)e_0(n)] = 0 \quad (16.10)$$

where $e_0(n) = d(n) - \mathbf{x}^t(n)\mathbf{w}_0$. Here we are keeping in mind that ORIV will be used with stochastic series although is designed for deterministic data.

Equation (16.9) shows that if we solve (16.2) (or (16.3)) involving stochastic series, let the algorithm evolve long enough in time and take expected value is equivalent to solve $\mathbf{R} \mathbf{w}_0 = \mathbf{r}$. Using ergodicity an equation made for stochastic processes has been solved using equations for deterministic processes. This is nothing new, but this argument will support our future analysis.

Effect of initial conditions

Combining the expressions in table 16.1, after some algebra the following equation can be derived:

$$\Delta \mathbf{w}(n) = \Delta \mathbf{w}(n - 1) + \mathbf{\Gamma}^{-1}(n) \mathbf{X}(n) \mathbf{\Lambda}^{-1}(n) \boldsymbol{\alpha}(n) \quad (16.11)$$

where $\Delta \mathbf{w}(n) = \mathbf{w}(n) - \mathbf{w}_0$. Multiplying by $\Gamma(n)$ and using its update expression, we get to

$$\begin{aligned} \Gamma(n)\Delta \mathbf{w}(n) &= \lambda^2 \Gamma(n-1)\Delta \mathbf{w}(n-1) + \\ &+ \lambda \mathbf{x}(n)\tilde{\mathbf{x}}^t(n)[\mathbf{z}(n-1) - \Phi(n-1)\mathbf{w}_0] + \Phi^t(n)\tilde{\mathbf{x}}(n)e_0(n) \end{aligned} \quad (16.12)$$

a difference equation which solution is:

$$\begin{aligned} \Gamma(n)\Delta \mathbf{w}(n) &= \lambda^{2n}\Gamma(0)\Delta \mathbf{w}(0) + \sum_{j=1}^n \lambda^{2(n-j)} \Phi^t(j)\tilde{\mathbf{x}}(j)e_0(j) + \\ &+ \sum_{i=1}^n \lambda \lambda^{2(n-i)} \mathbf{x}(i)\tilde{\mathbf{x}}^t(i)[\mathbf{z}(i-1) - \Phi(i-1)\mathbf{w}_0] \end{aligned} \quad (16.13)$$

To continue with the derivation a hypothesis, of the independence assumption type (HI), is needed :

HI1 $\Phi(n)$ and $\mathbf{z}(n)$ are independent of each other and of the rest of variables involved.

Taking expected value and taking into account the remark after (16.9), together with HI1, we get to:

$$\mathbb{E}[\Gamma(n)\Delta \mathbf{w}(n)] = \lambda^{2n}\Gamma(0)\Delta \mathbf{w}(0) \quad (16.14)$$

To study the convergence, n can be assumed large enough and use the decomposition given in (16.5) that allows us to state that $\Gamma(n)$ is independent of $\Delta \mathbf{w}(n)$ and gives an expected value of:

$$\mathbb{E}[\Gamma(n)] = (1 - \lambda)^{-2} \mathbf{R}^t \mathbf{R} + \mathbb{E}[\mathbf{R}^{et}(n)\mathbf{R}^e(n)]$$

in consequence

$$\begin{aligned} \mathbb{E}[\Gamma(n)]^{-1} &= \\ &= (1 - \lambda)^2 [\mathbf{I} + (1 - \lambda)^2 (\mathbf{R}^t \mathbf{R})^{-1} \mathbb{E}[\mathbf{R}^{et}(n)\mathbf{R}^e(n)]]^{-1} (\mathbf{R}^t \mathbf{R})^{-1} \approx \\ &\approx (1 - \lambda)^2 (\mathbf{R}^t \mathbf{R})^{-1} \end{aligned} \quad (16.15)$$

again λ is assumed to be near unity and also that $\Phi(n)$ is quasideterministic. With all of this we get finally to:

$$\mathbb{E}[\Delta \mathbf{w}(n)] = \lambda^{2n} (1 - \lambda)^2 (\mathbf{R}^t \mathbf{R})^{-1} \Gamma(0) \Delta \mathbf{w}(0) \quad (16.16)$$

Some comments could be made as a consequence of (16.16):

1. The influence of the initial conditions, $\Gamma(0)$ and $\mathbf{w}(0)$, falls exponentially like λ^{2n} .

2. $E[\mathbf{w}(n)] \rightarrow \mathbf{w}_0$ for large n , as it was known.
3. If $\lambda = 1$, $1 - \lambda$ has to be changed for $1/n$.
4. Expression (16.13) can be written in a more compact form as:

$$\mathbf{\Gamma}(n)\Delta\mathbf{w}(n) = \lambda^{2n}\mathbf{\Gamma}(0)\Delta\mathbf{w}(0) + \sum_{i=1}^n \lambda^{2(n-i)} \mathbf{X}(i)\mathbf{\Lambda}^{-1}(i)\boldsymbol{\alpha}_0(i)$$

where $\boldsymbol{\alpha}_0(n) = \mathbf{v}(n) - \mathbf{X}^t(n)\mathbf{w}_0$, that can be considered the generalization of the measurement error for overdetermined systems.

So (16.14) implies, using the stationarity of the processes involved, that

$$E[\mathbf{X}(n)\mathbf{\Lambda}^{-1}(n)\boldsymbol{\alpha}_0(n)] = 0 \quad (16.17)$$

that, in turn, can be understood as a generalization of (16.10). This expression will be exploited in the next chapter to derive a complete family of adaptive algorithms, so it is of great importance.

16.2.2 Mean Square Deviation

In this subsection the quantity $q(n) \equiv E[\Delta\mathbf{w}^t(n)\mathbf{Q}\Delta\mathbf{w}(n)]$ is computed, where \mathbf{Q} is any positive defined matrix (metric), that is to say, the mean square deviation of the $\mathbf{w}(n)$ estimates. We begin with (16.12) written in compact form:

$$\mathbf{\Gamma}(n)\Delta\mathbf{w}(n) = \lambda^2\mathbf{\Gamma}(n-1)\Delta\mathbf{w}(n-1) + \mathbf{X}(n)\mathbf{\Lambda}^{-1}(n)\boldsymbol{\alpha}_0(n)$$

Given that we want to compute the MSD in convergence we can suppose that n is large enough so that $\mathbf{\Gamma}^{-1}(n)\mathbf{\Gamma}(n-1) = \mathbf{I}$. In this approximation results:

$$\Delta\mathbf{w}(n) = \lambda^2\Delta\mathbf{w}(n-1) + \mathbf{\Gamma}^{-1}(n)\mathbf{X}(n)\mathbf{\Lambda}^{-1}(n)\boldsymbol{\alpha}_0(n)$$

Based on this difference equation, the following relation holds:

$$\begin{aligned} \Delta\mathbf{w}^t(n)\mathbf{Q}\Delta\mathbf{w}(n) &= \lambda^4\Delta\mathbf{w}^t(n-1)\mathbf{Q}\Delta\mathbf{w}(n-1) + \\ &+ 2\lambda^2\Delta\mathbf{w}^t(n-1)\mathbf{Q}\mathbf{\Gamma}^{-1}(n)\mathbf{X}(n)\mathbf{\Lambda}^{-1}(n)\boldsymbol{\alpha}_0(n) + \\ &+ [\mathbf{X}(n)\mathbf{\Lambda}^{-1}(n)\boldsymbol{\alpha}_0(n)]^t\hat{\boldsymbol{\beta}}_Q(n)[\mathbf{X}(n)\mathbf{\Lambda}^{-1}(n)\boldsymbol{\alpha}_0(n)] \end{aligned} \quad (16.18)$$

where $\hat{\boldsymbol{\beta}}_Q(n) \equiv \mathbf{\Gamma}^{-t}(n)\mathbf{Q}\mathbf{\Gamma}^{-1}(n)$. To continue, the expected value of each term is calculated. The first one is nothing else that $\lambda^4q(n-1)$. For the second one $\mathbf{\Gamma}(n)$ is replaced by its expected value and another independence assumption is introduced:

HI2 $\mathbf{w}(n-1)$ is independent of $\mathbf{X}(n)\mathbf{\Lambda}^{-1}(n)\boldsymbol{\alpha}_0(n)$.

So that we can conclude that this term is null using (16.17).

Let us study the third term. Because we are working in the lowest possible order in $(1 - \lambda)$ again $\mathbf{\Gamma}(n)$ is replaced by its expected value, resulting:

$$3^{\text{rd}} \text{ term} \xrightarrow{E[\cdot]} (1 - \lambda)^4 \text{Tr}\{\mathbf{V}_{X\alpha_0}\beta_Q\}$$

where

$$\mathbf{V}_{X\alpha_0} = E[\mathbf{X}(n)\mathbf{\Lambda}^{-1}(n)\boldsymbol{\alpha}_0(n)(\mathbf{X}(n)\mathbf{\Lambda}^{-1}(n)\boldsymbol{\alpha}_0(n))^t] \quad (16.19)$$

$$\beta_Q = (\mathbf{R}^t\mathbf{R})^{-t}\mathbf{Q}(\mathbf{R}^t\mathbf{R})^{-1} \quad (16.20)$$

Given that the factor $\mathbf{X}(n)\mathbf{\Lambda}^{-1}(n)\boldsymbol{\alpha}_0(n)$ is zero mean, $\mathbf{V}_{X\alpha_0}$ could be considered as its variance matrix. Finally, adding all the contributions:

$$q(n) = \lambda^4 q(n-1) + (1 - \lambda)^4 \text{Tr}\{\mathbf{V}_{X\alpha_0}\beta_Q\} \quad (16.21)$$

valid for n large enough. This difference equation gives a convergence value of:

$$q(\infty) = \frac{(1 - \lambda)^3}{(1 + \lambda^2)(1 + \lambda)} \text{Tr}\{\mathbf{V}_{X\alpha_0}\beta_Q\} \quad (16.22)$$

Some considerations can be made from (16.22):

1. For stationary systems, as is the case now, $\lambda \rightarrow 1$ and $(1 - \lambda)$ has to be changed for $1/n$ so that $q(\infty) \rightarrow 0$.
2. Both factors appearing in (16.22) depend on the instrumental variable \tilde{x} . The MSD in convergence could be reduced by an accurate choice of it.
3. The previous derivation as a function of $(1 - \lambda)$ is not really valid, because $\mathbf{V}_{X\alpha_0}$ has some terms on it like it will be seen in the next subsection.
4. Using (16.21) and (16.22) the following relation can be drawn:

$$q(n) = q(\infty) + \lambda^{4(n-n_0)}[q(n_0) - q(\infty)] \quad (16.23)$$

where $n > n_0 \gg 1$. This means that the effect of the initial conditions falls to 0 like λ^{4n} , for RIV is like λ^{2n} [Swa96]. In addition, the speed of convergence only depend on λ and not in the spread of eigenvalues of \mathbf{R} , as it is usual in other algorithms.

Despite point n.4, it does not eliminate the case in which, if the problem is ill-conditioned, numerical problems could arise that lead to a bad convergence.

16.2.3 MSD in noisy environments

When the series we are working with are contaminated with noise, the stochastic components $\mathbf{R}^e(n)$ and $\mathbf{r}^e(n)$ of (16.5) and (16.6) are more and more important. Here the influence of these components in (16.21) is considered:

- i) $\beta_Q = \mathbb{E}[\mathbf{\Gamma}^{-t}(n)\mathbf{Q}\mathbf{\Gamma}^{-1}(n)]$. Here the decomposition given in (16.5) is used. With it we see that the stochastic component in this term appears always in the form ‘term of \mathbf{R}^e/\mathbf{R} ’, which is negligible compared to the terms in which \mathbf{R}^e appears alone, just because $\mathbf{\Phi}(n)$ is quasi-deterministic.
- ii) $\mathbf{V}_{X\alpha_0}$. The terms in \mathbf{R}^e are not normalized by \mathbf{R} here. It is not possible, in consequence, to omit the stochastic component.

In short, the noise influence is bigger for $\mathbf{V}_{X\alpha_0}$ than for β_Q , so that case ii) has to be considered in detail. Unfolding the components according to the definition we get:

$$\begin{aligned} \mathbf{X}(n)\mathbf{\Lambda}^{-1}(n)\boldsymbol{\alpha}_0(n) &= \\ &= \lambda\mathbf{x}(n)\tilde{\mathbf{x}}^t(n)[z(n-1) - \mathbf{\Phi}(n-1)\mathbf{w}_0] + \mathbf{\Phi}^t(n)\tilde{\mathbf{x}}(n)e_0(n) \end{aligned} \quad (16.24)$$

assuming large n and using the remark after (16.9) results:

$$\begin{aligned} \mathbf{X}(n)\mathbf{\Lambda}^{-1}(n)\boldsymbol{\alpha}_0(n) &\underset{n \rightarrow \infty}{\sim} \\ &\sim \lambda\mathbf{x}(n)\tilde{\mathbf{x}}^t(n) \left[\frac{\mathbf{r}}{1-\lambda} + \mathbf{r}^e(n-1) - \frac{\mathbf{R}}{1-\lambda}\mathbf{w}_0 - \mathbf{R}^e(n-1)\mathbf{w}_0 \right] + \\ &\quad + \frac{\mathbf{R}^t}{1-\lambda}\tilde{\mathbf{x}}(n)e_0(n) + \mathbf{R}^{et}(n)\tilde{\mathbf{x}}(n)e_0(n) = \\ &= \lambda\mathbf{x}(n)\tilde{\mathbf{x}}^t(n)[\mathbf{r}^e(n-1) - \mathbf{R}^e(n-1)\mathbf{w}_0] + \frac{\mathbf{R}^t}{1-\lambda}\tilde{\mathbf{x}}(n)e_0(n) + \mathbf{R}^{et}(n)\tilde{\mathbf{x}}(n)e_0(n) \end{aligned} \quad (16.25)$$

The expected value of this expression is zero, as it was known for the general case. Here it has been proved for large n . Continuing with the derivation, the following product is constructed:

$$\begin{aligned} \mathbf{X}(n)\mathbf{\Lambda}^{-1}(n)\boldsymbol{\alpha}_0(n)[\mathbf{X}(n)\mathbf{\Lambda}^{-1}(n)\boldsymbol{\alpha}_0(n)]^t &= \\ &= \{\lambda\mathbf{x}(n)\tilde{\mathbf{x}}^t(n)[\mathbf{r}^e(n-1) - \mathbf{R}^e(n-1)\mathbf{w}_0] + \\ &\quad + (1-\lambda)^{-1}\mathbf{R}^t\tilde{\mathbf{x}}(n)e_0(n) + \mathbf{R}^{et}(n)\tilde{\mathbf{x}}(n)e_0(n)\} \\ &\quad \{\lambda\mathbf{x}(n)\tilde{\mathbf{x}}^t(n)[\mathbf{r}^e(n-1) - \mathbf{R}^e(n-1)\mathbf{w}_0] + \\ &\quad + (1-\lambda)^{-1}\mathbf{R}^t\tilde{\mathbf{x}}(n)e_0(n) + \mathbf{R}^{et}(n)\tilde{\mathbf{x}}(n)e_0(n)\}^t \end{aligned} \quad (16.26)$$

Operating and taking expected values, having in mind that $\mathbf{R}^e(n)$, $\mathbf{r}^e(n)$ and $e_0(n)$ are zero mean and independent of the rest of stochastic variables, results:

$$\mathbf{V}_{X\alpha 0} \xrightarrow{n \rightarrow \infty} \lambda^2 \mathbf{V}_X + \sigma_0^2 \left[\frac{\mathbf{A}}{(1-\lambda)^2} + \mathbf{V}_R \right] \quad (16.27)$$

where

$$\mathbf{V}_X = \mathbb{E}[\mathbf{x}(n)\tilde{\mathbf{x}}^t(n)\{\mathbf{V}\mathbf{r}^e + \mathbb{E}[\mathbf{R}^e(n)\mathbf{w}_0\mathbf{w}_0^t\mathbf{R}^{et}(n)]\}\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)] \quad (16.28)$$

$$\mathbf{V}\mathbf{r}^e = \mathbb{E}[\mathbf{r}^e(n)\mathbf{r}^{et}(n)] \quad (16.29)$$

$$\sigma_0^2 = \mathbb{E}[e_0^2(n)] \quad (16.30)$$

$$\mathbf{V}_R = \mathbb{E}[\mathbf{R}^{et}(n)\mathbf{R}_{\tilde{x}\tilde{x}}\mathbf{R}^e(n)] \quad (16.31)$$

$$\mathbf{A} = \mathbf{R}^t\mathbf{R}_{\tilde{x}\tilde{x}}\mathbf{R} \quad (16.32)$$

In this way, the expression (16.22) can be complemented with (16.27), assuming also that $\lambda \sim 1$:

$$\begin{aligned} q(\infty) &= (1-\lambda)\sigma_0^2 \frac{\text{Tr}\{\beta_Q \mathbf{A}\}}{4} + (1-\lambda)^3 \left[\sigma_0^2 \frac{\text{Tr}\{\beta_Q \mathbf{V}_R\}}{4} + \frac{\text{Tr}\{\beta_Q \mathbf{V}_X\}}{4} \right] \equiv \\ &\equiv q_0(\infty) + q_e(\infty) \end{aligned} \quad (16.33)$$

where $q_e(\infty)$ has all the dependency on the \mathbf{R} and \mathbf{r} estimators and will rise as the SNR decreases and $q_0(\infty)$ will not depend on the SNR. For stationary environments $q_0(\infty)$ falls like $1/n$ and $q_e(\infty)$ like $1/n^3$. That is, the measurement noise $e_0(n)$ is more important than the estimation noise. Besides, the term \mathbf{V}_X is affected by additive gaussian noise (the instrumental variable is chosen to assure that \mathbf{R} and \mathbf{r} are not affected by additive gaussian noise).

The results have been obtained as a perturbation series in $(1-\lambda)$. All the approximations in (16.18) should be made at the same time in all the terms, to sum, later on, all the terms that have the same order in the perturbation parameter. In the development carried out here, the $\mathbf{\Gamma}(n)$ matrix has been treated first and $\mathbf{V}_{X\alpha 0}$ afterwards, neglecting possible interactions between both contributions. Nevertheless, more rigorous calculations show that each term contribution is of the same order as given. That is, the measurement noise contributes like $(1-\lambda)$, the estimation noise like $(1-\lambda)^3$ and there is no term in $(1-\lambda)^2$.

16.2.4 MSD in noisy environments II

The definition of MSD will, from now on, be $D(n) = \mathbb{E}[\Delta \mathbf{w}^t(n)\Delta \mathbf{w}(n)]$, it is like $q(n)$ but taking $\mathbf{Q} = \mathbf{I}$ (the identity matrix) to simplify notation. In this subsection the effect of the additive noise in the MSD will be considered, but

using a different approach than the used in the last subsection. From (16.11) the following expression can be obtained:

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \mathbf{\Gamma}^{-1}(n)\mathbf{X}(n)\mathbf{\Lambda}^{-1}(n)\boldsymbol{\alpha}(n) \quad (16.34)$$

unfolding the components:

$$\begin{aligned} \mathbf{w}(n) = \mathbf{w}(n-1) + \mathbf{\Gamma}^{-1}(n)\{ & \mathbf{\Phi}^t(n)\tilde{\mathbf{x}}(n)e_0(n) + \lambda\mathbf{x}(n)\tilde{\mathbf{x}}^t(n)[\mathbf{z}(n-1) - \\ & - \mathbf{\Phi}(n-1)\mathbf{w}(n-1)] + \mathbf{\Phi}^t(n)\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)[\mathbf{w}_0 - \mathbf{w}(n-1)]\} \end{aligned} \quad (16.35)$$

That gives this expression for the error vector $\Delta\mathbf{w}(n) = \mathbf{w}(n) - \mathbf{w}_0$,

$$\begin{aligned} \Delta\mathbf{w}(n) = [\mathbf{I} - \lambda\mathbf{\Gamma}^{-1}(n)\mathbf{x}(n)\tilde{\mathbf{x}}^t(n)\mathbf{\Phi}(n-1) - \\ - \mathbf{\Gamma}^{-1}(n)\mathbf{\Phi}^t(n)\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)]\Delta\mathbf{w}(n-1) + \mathbf{\Gamma}^{-1}(n)\mathbf{\Phi}^t(n)\tilde{\mathbf{x}}(n)e_0(n) + \\ + \lambda\mathbf{\Gamma}^{-1}(n)\mathbf{x}(n)\tilde{\mathbf{x}}^t(n)\mathbf{z}(n-1) - \lambda\mathbf{\Gamma}^{-1}(n)\mathbf{x}(n)\tilde{\mathbf{x}}^t(n)\mathbf{\Phi}(n-1)\mathbf{w}_0 \end{aligned} \quad (16.36)$$

The approximations to be used are based on the conditions under which the results want to be obtained. The goal is to compute the MSD in convergence and with estimation noise, so n is supposed to be large and only the first terms in the expansion containing information of the estimation noise are considered:

$$\begin{aligned} \mathbf{\Gamma}^{-1}(n) \approx (1-\lambda)^2(\mathbf{R}^t\mathbf{R})^{-1} - \\ - (1-\lambda)^3(\mathbf{R}^t\mathbf{R})^{-1}[\mathbf{R}^t\mathbf{R}^e(n) + \mathbf{R}^{et}(n)\mathbf{R}](\mathbf{R}^t\mathbf{R})^{-1} \end{aligned} \quad (16.37)$$

$$\mathbf{\Phi}(n) \approx \frac{1}{1-\lambda}\mathbf{R} + \mathbf{R}^e(n) \quad (16.38)$$

$$\mathbf{z}(n) \approx \frac{1}{1-\lambda}\mathbf{r} + \mathbf{r}^e(n) \quad (16.39)$$

Taking into account these approximations and direct averaging [Kus84], (16.36) turns to:

$$\begin{aligned} \Delta\mathbf{w}(n) = \lambda^2\Delta\mathbf{w}(n-1) + (1-\lambda)(\mathbf{R}^t\mathbf{R})^{-1}\mathbf{R}^t\tilde{\mathbf{x}}(n)e_0(n) + \\ + (1-\lambda)^2(\mathbf{R}^t\mathbf{R})^{-1}\mathbf{R}^{et}(n)\tilde{\mathbf{x}}(n)e_0(n) - \\ - (1-\lambda)^2(\mathbf{R}^t\mathbf{R})^{-1}[\mathbf{R}^t\mathbf{R}^e(n) + \mathbf{R}^{et}(n)\mathbf{R}](\mathbf{R}^t\mathbf{R})^{-1}\mathbf{R}^t\tilde{\mathbf{x}}(n)e_0(n) - \\ - (1-\lambda)^3(\mathbf{R}^t\mathbf{R})^{-1}[\mathbf{R}^t\mathbf{R}^e(n) + \mathbf{R}^{et}(n)\mathbf{R}](\mathbf{R}^t\mathbf{R})^{-1}\mathbf{R}^{et}(n)\tilde{\mathbf{x}}(n)e_0(n) - \\ - (1-\lambda)^2\lambda(\mathbf{R}^t\mathbf{R})^{-1}\mathbf{x}(n)\tilde{\mathbf{x}}^t(n)[\mathbf{R}^e(n-1)\mathbf{w}_0 - \mathbf{r}^e(n-1)] + \\ + (1-\lambda)^3\lambda(\mathbf{R}^t\mathbf{R})^{-1}[\mathbf{R}^t\mathbf{R}^e(n) + \mathbf{R}^{et}(n)\mathbf{R}](\mathbf{R}^t\mathbf{R})^{-1} \\ \mathbf{x}(n)\tilde{\mathbf{x}}^t(n)[\mathbf{R}^e(n-1)\mathbf{w}_0 - \mathbf{r}^e(n-1)] \end{aligned} \quad (16.40)$$

Constructing the scalar product $\Delta\mathbf{w}^t(n)\Delta\mathbf{w}(n)$, taking expected value and neglecting for the moment terms in $(1-\lambda)^4$ emerges:

$$D(n) = \lambda^4 D(n-1) + (1-\lambda)^2 \sigma_0^2 \tilde{\gamma}$$

where only effects of the measurement noise are included. The terms in $(1 - \lambda)^4$ include information about the estimation noise and will be grouped in the terms V_e and ξ_e , with them finally we get to:

$$D(n) = \lambda^4 D(n-1) + (1 - \lambda)^2 \sigma_0^2 \tilde{\gamma} + (1 - \lambda)^4 \sigma_0^2 V_e + (1 - \lambda)^4 \lambda^2 \xi_e \quad (16.41)$$

where

$$\tilde{\gamma} = \text{Tr}\{\mathbf{R}(\mathbf{R}^t \mathbf{R})^{-2} \mathbf{R}^t \mathbf{R}_{\tilde{x}\tilde{x}}\} \quad (16.42)$$

$$\begin{aligned} V_e = & \text{E}[\tilde{\mathbf{x}}^t(n) \mathbf{R}^e(n) (\mathbf{R}^t \mathbf{R})^{-2} \mathbf{R}^{et}(n) \tilde{\mathbf{x}}(n)] - \\ & - 2\text{E}[\tilde{\mathbf{x}}^t(n) \mathbf{R}^e(n) (\mathbf{R}^t \mathbf{R})^{-1} [\mathbf{R}^t \mathbf{R}^e(n) + \mathbf{R}^{et}(n) \mathbf{R}] (\mathbf{R}^t \mathbf{R})^{-2} \mathbf{R}^t \tilde{\mathbf{x}}(n)] - \\ & - 2\text{E}[\tilde{\mathbf{x}}^t(n) \mathbf{R}^e(n) (\mathbf{R}^t \mathbf{R})^{-2} [\mathbf{R}^t \mathbf{R}^e(n) + \mathbf{R}^{et}(n) \mathbf{R}] (\mathbf{R}^t \mathbf{R})^{-1} \mathbf{R}^t \tilde{\mathbf{x}}(n)] + \\ & + \text{E}[\tilde{\mathbf{x}}^t(n) \mathbf{R} (\mathbf{R}^t \mathbf{R})^{-1} [\mathbf{R}^t \mathbf{R}^e(n) + \mathbf{R}^{et}(n) \mathbf{R}] (\mathbf{R}^t \mathbf{R})^{-2} \\ & \quad [\mathbf{R}^t \mathbf{R}^e(n) + \mathbf{R}^{et}(n) \mathbf{R}] (\mathbf{R}^t \mathbf{R})^{-1} \mathbf{R}^t \tilde{\mathbf{x}}(n)] \end{aligned} \quad (16.43)$$

$$\xi_e = \text{E}[(\mathbf{w}_0^t \mathbf{R}^{et}(n-1) - \mathbf{r}^{et}(n-1)) \tilde{\mathbf{x}}(n) \mathbf{x}^t(n) (\mathbf{R}^t \mathbf{R})^{-2} \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) (\mathbf{R}^e(n-1) \mathbf{w}_0 - \mathbf{r}^e(n-1))] \quad (16.44)$$

assuming $\lambda \sim 1$, in convergence results

$$D(\infty) = (1 - \lambda) \sigma_0^2 \frac{\tilde{\gamma}}{4} + (1 - \lambda)^3 \sigma_0^2 \frac{V_e}{4} + (1 - \lambda)^3 \frac{\xi_e}{4} \quad (16.45)$$

in complete similarity with (16.33) relating $\tilde{\gamma} \leftrightarrow \text{Tr}\{\boldsymbol{\beta}_Q \mathbf{A}\}$, $V_e \leftrightarrow \text{Tr}\{\boldsymbol{\beta}_Q \mathbf{V}_R\}$ and $\xi_e \leftrightarrow \text{Tr}\{\boldsymbol{\beta}_Q \mathbf{V}_X\}$, where the same case is studied from a different approximation, this means that these quantities are defined using the same components but with a different treatment.

16.3 Convergence analysis for non-stationary systems

16.3.1 Convergence in the mean

In order to study the tracking capabilities the starting point will be equation (16.34). In which the a priori error vector can be written as:

$$\begin{aligned} \boldsymbol{\alpha}(n) &= \mathbf{v}(n) - \mathbf{X}^t(n) \mathbf{w}(n-1) = \\ &= \begin{pmatrix} \tilde{\mathbf{x}}^t(n) [\mathbf{z}(n-1) - \boldsymbol{\Phi}(n-1) \mathbf{w}(n-1)] \\ d(n) - \mathbf{x}^t(n) \mathbf{w}(n-1) \end{pmatrix} = \\ &= \begin{pmatrix} 0 \\ \mathbf{e}_0(n) \end{pmatrix} + \begin{pmatrix} \tilde{\mathbf{x}}^t(n) [\mathbf{z}(n-1) - \boldsymbol{\Phi}(n-1) \mathbf{w}(n-1)] \\ \mathbf{x}^t(n) [\mathbf{w}_0(n) - \mathbf{w}(n-1)] \end{pmatrix} \end{aligned} \quad (16.46)$$

Using (16.46) and the definitions of the components involved, with a little computation:

$$\begin{aligned} \mathbf{X}(n)\mathbf{\Lambda}^{-1}(n)\boldsymbol{\alpha}(n) &= \\ &= \boldsymbol{\Phi}^t(n)\tilde{\mathbf{x}}(n)e_0(n) + \lambda\mathbf{x}(n)\tilde{\mathbf{x}}^t(n)[\mathbf{z}(n-1) - \boldsymbol{\Phi}(n-1)\mathbf{w}(n-1)] + \\ &\quad + \boldsymbol{\Phi}^t(n)\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)[\mathbf{w}_0(n) - \mathbf{w}(n-1)] \end{aligned} \quad (16.47)$$

In this particular study the only interest is in the behavior of ORIV when the unknown parameters change according to the series $\{\mathbf{w}_0(n)\}$ (non-stationary solution). For this reason the rest of contributions, such as the estimation noise, will be neglected. In addition, because only the convergence is of interest, n will be supposed large and we can assume $\mathbf{\Gamma}^{-1}(n) = (1-\lambda)^2(\mathbf{R}^t(n)\mathbf{R}(n))^{-1}$, $\mathbf{z}(n) = (1-\lambda)^{-1}\mathbf{r}(n)$ and $\boldsymbol{\Phi}(n) = (1-\lambda)^{-1}\mathbf{R}(n)$. Note that now, because the filter coefficients are time variant, the statistics of the processes involved can also be time variant. Taking expected value:

$$\begin{aligned} \mathbb{E}[\mathbf{X}(n)\mathbf{\Lambda}^{-1}(n)\boldsymbol{\alpha}(n)] &= \frac{1}{1-\lambda}\mathbf{R}^t(n)\mathbf{R}(n)[\mathbf{w}_0(n) - \mathbb{E}[\mathbf{w}(n-1)]] + \\ &\quad + \frac{\lambda}{1-\lambda}[\mathbf{R}^t(n)\mathbf{r}(n-1) - \mathbf{R}^t(n)\mathbf{R}(n-1)\mathbb{E}[\mathbf{w}(n-1)]] \end{aligned} \quad (16.48)$$

Going back to (16.34), replacing $\mathbf{r}(n-1)$ by $\mathbf{R}(n-1)\mathbf{w}_0(n-1)$ in (16.48) and taking again expected value we finally get to,

$$\begin{aligned} \mathbb{E}[\mathbf{w}(n)] &= \mathbb{E}[\mathbf{w}(n-1)] + \lambda(1-\lambda)[\mathbf{w}_0(n-1) - \mathbb{E}[\mathbf{w}(n-1)]] + \\ &\quad + (1-\lambda)(\mathbf{w}_0(n) - \mathbb{E}[\mathbf{w}(n-1)]) = \\ &= \lambda^2\mathbb{E}[\mathbf{w}(n-1)] + (1-\lambda)\mathbf{w}_0(n) + \lambda(1-\lambda)\mathbf{w}_0(n-1) \end{aligned} \quad (16.49)$$

where it was assumed that $[\mathbf{R}^t(n)\mathbf{R}(n)]^{-1}\mathbf{R}^t(n)\mathbf{R}(n-1) = \mathbf{I}$, or equivalently that the system evolves slowly enough with time to make this approximation. Expression (16.49) is a difference equation that has as a particular solution,

$$\mathbb{E}[\mathbf{w}(z)] = \mathbf{w}_0(z) - \lambda\frac{1-z^{-1}}{1-\lambda^2z^{-1}}\mathbf{w}_0(z)$$

or equivalently,

$$\mathbb{E}[\mathbf{w}(z)] = \mathbf{w}_0(z) - \lambda(1-z^{-1})\sum_{i=0}^{\infty}\lambda^{2i}z^{-i}\mathbf{w}_0(z)$$

This expression can be rewritten as a function of the time index n as follows,

$$\mathbb{E}[\mathbf{w}(n)] = (1-\lambda)\mathbf{w}_0(n) + \frac{1-\lambda^2}{\lambda}\sum_{i=1}^{\infty}\lambda^{2i}\mathbf{w}_0(n-i) \quad (16.50)$$

The general solution of the homogeneous part is $\mathbb{E}[\mathbf{w}(n)]_g = \lambda^{2n}\mathbb{E}[\mathbf{w}(0)]$ that tends to zero for large n . The convergence time τ of the homogenous part could be defined as follows:

$$\mathbb{E}[\mathbf{w}(\tau)] = \frac{\mathbb{E}[\mathbf{w}(0)]}{e} \rightarrow 2\tau \ln \lambda = -1$$

which means, assuming $\lambda \sim 1$, that

$$\tau = \frac{1}{2(1-\lambda)} \quad (16.51)$$

The bigger λ is, the longer takes to the solution to converge to the particular solution. On the contrary, the particular solution at time n has less contribution of previous times of $\mathbf{w}_0(n)$ the smaller λ is. This means that a small λ improves the tracking capability. As an example let us suppose that $\mathbf{w}_0(n) = n\mathbf{p}$, where \mathbf{p} is a constant vector. In this case (16.50) implies that $\mathbf{w}(n) = \mathbf{w}_0(n - \lambda/(1-\lambda^2))$ which means that the stationary solution is delayed with respect to the true solution an amount of time equal to $\lambda/(1-\lambda^2)$. Keep in mind that two different aspects have been discussed. One is the time that takes to the general solution of the homogenous part to get to zero and the other is the delay time that the stationary solution has respect to the true one.

As a verification of (16.50), if $\mathbf{w}_0(n) = \mathbf{w}_0$ we get that $E[\mathbf{w}(n)] = \mathbf{w}_0$ as was derived in the stationary case analysis.

16.3.2 Mean Square Deviation

In this subsection we will assume, for operational reasons, that the system parameters depend on time according to a lineal expression, as in the previous example:

$$\mathbf{w}_0(n) = n\mathbf{p}$$

Using the same equations and hypothesis that in the previous subsection it is easy to get to

$$\begin{aligned} \Delta\mathbf{w}(n) = & [\mathbf{I} - \lambda(1-\lambda)(\mathbf{R}^t(n)\mathbf{R}(n))^{-1}\mathbf{x}(n)\tilde{\mathbf{x}}^t(n)\mathbf{R}(n-1) - \\ & -(1-\lambda)(\mathbf{R}^t(n)\mathbf{R}(n))^{-1}\mathbf{R}^t(n)\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)]\Delta\mathbf{w}(n-1) + \\ & +(1-\lambda)(\mathbf{R}^t(n)\mathbf{R}(n))^{-1}\mathbf{R}^t(n)\tilde{\mathbf{x}}(n)e_0(n) + \\ & +(1-\lambda)(\mathbf{R}^t(n)\mathbf{R}(n))^{-1}\mathbf{R}^t(n)\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)\mathbf{p} - \mathbf{p} \end{aligned} \quad (16.52)$$

a non-homogeneous difference equation with a stochastic coefficient. Because $\lambda \sim 1$ the stochastic part of the coefficient is small and can be considered that behaves like its mean value. That is to say, the direct averaging technique is used, so that the equation turns to

$$\begin{aligned} \Delta\mathbf{w}(n) = & \lambda^2\Delta\mathbf{w}(n-1) + (1-\lambda)(\mathbf{R}^t(n)\mathbf{R}(n))^{-1}\mathbf{R}^t(n)\tilde{\mathbf{x}}(n)e_0(n) - \\ & -[\mathbf{I} - (1-\lambda)(\mathbf{R}^t(n)\mathbf{R}(n))^{-1}\mathbf{R}^t(n)\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)]\mathbf{p} \end{aligned} \quad (16.53)$$

constructing the product $\Delta \mathbf{w}^t(n)\Delta \mathbf{w}(n)$, taking expected value and taking into account that $e_0(n)$ is independent and zero mean, we get easily to

$$\begin{aligned} D(n) = & \lambda^4 D(n-1) - 2\lambda^2 \mathbb{E}[\Delta \mathbf{w}^t(n-1)]\lambda \mathbf{p} + \mathbf{p}^t \mathbf{p} - 2(1-\lambda)\mathbf{p}^t \mathbf{p} + \\ & + (1-\lambda)^2 \sigma_0^2 \text{Tr}[\mathbf{R}(n)(\mathbf{R}^t(n)\mathbf{R}(n))^{-2}\mathbf{R}^t(n)\mathbf{R}_{\tilde{x}\tilde{x}}(n)] + \\ & + (1-\lambda)^2 \mathbf{p}^t \mathbb{E}[\mathbf{x}(n)\tilde{\mathbf{x}}^t(n)\mathbf{R}(n)(\mathbf{R}^t(n)\mathbf{R}(n))^{-2}\mathbf{R}^t(n)\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)]\mathbf{p} \end{aligned} \quad (16.54)$$

Before going on further is desirable to make a remark: taking expected value in (16.53) results,

$$\mathbb{E}[\Delta \mathbf{w}(n)] = \lambda^2 \mathbb{E}[\Delta \mathbf{w}(n-1)] - \lambda \mathbf{p}$$

in convergence

$$\mathbb{E}[\Delta \mathbf{w}(\infty)] = -\frac{\lambda}{1-\lambda^2}\mathbf{p}$$

or taking as a proper solution for large n the expression:

$$\mathbb{E}[\mathbf{w}(n)] = \mathbf{w}_0(n) - \frac{\lambda}{1-\lambda^2}\mathbf{p} \quad n \gg 1 \quad (16.55)$$

so that this solution has a delay time equal to

$$\mathbb{E}[\mathbf{w}(n)] = \mathbf{w}_0(n - \tau') \quad \tau' = \frac{\lambda}{1-\lambda^2}$$

This delay time is the same as the one given after the comments on (16.51) for this very example, as expected.

If in equation (16.54) the expression in (16.55) is considered, we get to:

$$\begin{aligned} D(n) = & \lambda^4 D(n-1) + \frac{2\lambda^4 - 2\lambda^3 + \lambda^2 + 2\lambda - 1}{1-\lambda^2}\mathbf{p}^t \mathbf{p} + \\ & + (1-\lambda)^2 \sigma_0^2 \tilde{\gamma}(n) + (1-\lambda)^2 \mathbf{p}^t \mathbf{H}(n)\mathbf{p} \end{aligned} \quad (16.56)$$

where

$$\tilde{\gamma}(n) = \text{Tr}[\mathbf{R}(n)(\mathbf{R}(n)\mathbf{R}(n))^{-2}\mathbf{R}^t(n)\mathbf{R}_{\tilde{x}\tilde{x}}(n)] \quad (16.57)$$

$$\mathbf{H}(n) = \mathbb{E}[\mathbf{x}(n)\tilde{\mathbf{x}}^t(n)\mathbf{R}(n)(\mathbf{R}^t(n)\mathbf{R}(n))^{-2}\mathbf{R}^t(n)\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)] \quad (16.58)$$

The matrix \mathbf{H} and the scalar $\tilde{\gamma}$ depend on the time index n because the system is non-stationary. This expression is only valid for large n . Equation (16.56) converges to:

$$D(\infty) = \frac{1}{(1-\lambda)^2}\frac{\mathbf{p}^t \mathbf{p}}{4} + \frac{7}{2}\frac{\mathbf{p}^t \mathbf{p}}{4} + (1-\lambda)\sigma_0^2\frac{\tilde{\gamma}}{4} + (1-\lambda)\frac{\mathbf{p}^t \mathbf{H}\mathbf{p}}{4} \quad (16.59)$$

where λ is considered to be near 1, the following definitions are made

$$\mathbf{H} \equiv 4(1-\lambda) \lim_{n \rightarrow \infty} \sum_{i=0}^{n-1} \lambda^{4i} \mathbf{H}(n-i) \quad (16.60)$$

$$\tilde{\gamma} \equiv 4(1-\lambda) \lim_{n \rightarrow \infty} \sum_{i=0}^{n-1} \lambda^{4i} \tilde{\gamma}(n-i) \quad (16.61)$$

and the terms have been ordered in decreasing importance. Recall that the most influential term is the one due to the non-stationarity of the system, on the other hand the measurement noise is located in third place. The only contribution not considered in (16.59) is the estimation noise of \mathbf{R} and \mathbf{r} . This contribution has been studied previously.

16.4 Optimum tracking condition

Adding the contributions (16.59) and (16.45) an expression for MSD in convergence is obtained that includes all the factors previously discussed:

$$D(\infty) = \frac{1}{(1-\lambda)^2} \frac{\mathbf{p}^t \mathbf{p}}{4} + \frac{7}{2} \frac{\mathbf{p}^t \mathbf{p}}{4} + (1-\lambda) \frac{\sigma_0^2 \tilde{\gamma}}{4} + (1-\lambda) \frac{\mathbf{p}^t \mathbf{H} \mathbf{p}}{4} + (1-\lambda)^3 \frac{\sigma_0^2 V_e + \xi_e}{4} \quad (16.62)$$

In [MB91] a similar result is obtained for RLS: the measurement noise is affected by the $(1-\lambda)$ factor and the summand due to non-stationarity by $(1-\lambda)^{-2}$.

Which is the optimum λ_{opt} that gives a minimum $D(\infty)$? Minimizing (16.62) with respect to λ , supposing for simplicity that $V_e = 0$ and $\xi_e = 0$, results

$$\lambda_{opt} = 1 - \left[\frac{2\mathbf{p}^t \mathbf{p}}{\sigma_0^2 \tilde{\gamma} + \mathbf{p}^t \mathbf{H} \mathbf{p}} \right]^{1/3} \quad (16.63)$$

Is interesting to discuss two particular cases:

- a) When the lineal variation of the non-stationary model that we are tracking is very big, λ_{opt} decreases.
- b) When the measurement noise increases, λ_{opt} tends to unity, because a long memory for good estimators is needed.

Expression (16.62) represents the convergence value for the MSD considering the two error sources. The first one, expression (16.45), includes the MSD due to that the finite memory ($\lambda \neq 1$) of the algorithm increases the variance of the estimators of the quantities involved in the resolution of (16.4). The second one, expression (16.59), includes the MSD due to the estimation of non-stationary systems. It is known that (16.45) is the MSD in the stationary state caused by the fact that $\lambda \neq 1$. If λ where unity we would have that $1-\lambda$ turns to $1/n$ and this contribution will be zero for large n . Nevertheless, (16.59) appears in the need to track non-stationary systems (in particular in this case systems with linear temporal evolution) and logically, for $1-\lambda \rightarrow 1/n$ would diverge, because

if $\lambda = 1$ the tracking capability is destroyed. To sum up, a) and b) are two opposite situations, among them (16.63) weights both and gives the optimum value.

16.5 Applying ORIV to blind system identification

Let us consider the output of a causal, linear, time invariant, finite impulse response system given by

$$x(n) = \sum_{i=0}^q b_i w(n-i) \quad (16.64)$$

where this output is possibly contaminated by additive noise $v(n)$:

$$y(n) = x(n) + v(n)$$

In what follows it is assumed that:

1. The excitation sequence $\{w(n)\}$ is inaccessible and is a non-gaussian random process of zero mean, independent and identically distributed (i.i.d) with $0 < |\gamma_{kw}| < \infty$ and $0 < |\gamma_{k-1w}| < \infty$ $k > 3$, where γ_{kw} is the k^{th} -order cumulant of $w(n)$.
2. The additive noise $\{v(n)\}$ is a gaussian process, white or colored, independent of the input signal $\{w(n)\}$ and thus, of the output $\{x(n)\}$.
3. The system is possibly of non-minimum phase with $b_q \neq 0$.

16.5.1 Relations between cumulants and system coefficients

Under the previous conditions the Brillinger-Rosenblatt equation holds,

$$c_{kx}(\tau_1, \tau_2, \dots, \tau_{k-1}) = \gamma_{kw} \sum_{i=0}^q b_i b_{i+\tau_1} \dots b_{i+\tau_{k-1}} \quad (16.65)$$

and from it, the following expression can be derived

$$\sum_{i=0}^q b_i^k c_{kx}(0, \dots, 0, i-m) = \gamma_{kw} \sum_{i,l} b_i^k b_l^{k-1} b_{l+i-m}$$

summing over i and considering again (16.65) results

$$\sum_{i=0}^q b_i^k c_{kx}(0, \overset{(k-2)}{\dots}, 0, i-m) = \frac{\gamma_{kw}}{\gamma_{k+1w}} \sum_l^q b_l^{k-1} c_{k+1x}(0, \overset{(k-1)}{\dots}, 0, l-m)$$

Renaming l like i and taking into account the cumulant symmetries the desired expression is reached

$$\sum_{i=0}^q b_i^k c_{kx}(m-i, \overset{(k-1)}{\dots}, m-i) = \frac{\gamma_{kw}}{\gamma_{k+1w}} \sum_i^q b_i^{k-1} c_{k+1x}(m-i, \overset{(k)}{\dots}, m-i) \quad (16.66)$$

Depending on the value of k different particular equations can be obtained.

The Giannakis-Mendel equation

With $k = 2$ equation (16.66) becomes

$$\sum_{k=0}^q b_k^2 c_{2x}(m-k) = \frac{\sigma_w^2}{\gamma_3 w} \sum_{k=0}^q b_k c_{3x}(m-k, m-k) \quad (16.67)$$

that was first derived by Giannakis and Mendel and involves 2nd and 3rd order statistics.

Generalization of the GM equation

With $k = 3$ equation (16.66) turns to

$$\sum_{k=0}^q b_k^3 c_{3x}(m-k, m-k) = \frac{\gamma_{3w}}{\gamma_{4w}} \sum_{k=0}^q b_k^2 c_{4x}(m-k, m-k, m-k) \quad (16.68)$$

that is a generalization of the GM equation, previously derived by several authors [Rui95], [NKSK95] independently. Nevertheless, it is worth saying that this is the first time that a general expression, covering the GM equation and its generalization, is derived in the time domain (equation (16.68) was obtained in the time domain in [ARBC02a]).

16.5.2 Solving the equations using ORIV

Because expressions (16.67) and (16.68) relate the filter coefficients with the output cumulants they provide a way of identifying the system without knowledge of the input (blind identification). From both of them, making $-q \leq$

$m \leq 2q$, sets of linear equations are obtained. These equations involve the actual cumulants of the output which are not known because they need to be estimated, under these conditions the linear equations take the form:

$$\Phi(n)\mathbf{w}(n) = \mathbf{z}(n) \quad (16.69)$$

where the following definitions are made:

$$\Phi(n) = (\Phi_1(n) \mid \Phi_2(n))$$

$$\Phi_1(n) = \begin{pmatrix} \hat{c}_3(-q; n) & 0 & \dots & 0 \\ \hat{c}_3(-q+1; n) & \hat{c}_3(-q; n) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & & \vdots \\ \hat{c}_3(q; n) & \hat{c}_3(q-1; n) & \dots & \hat{c}_3(0; n) \\ 0 & \ddots & & \\ \vdots & & & \ddots \\ 0 & \dots & \dots & \hat{c}_3(q; n) \end{pmatrix} \quad 3q+1 \times q+1$$

$$\Phi_2(n) = \begin{pmatrix} 0 & \dots & \dots & 0 \\ \hat{c}_2(-q; n) & \dots & \dots & 0 \\ \hat{c}_2(-q+1; n) & \hat{c}_2(-q; n) & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ \hat{c}_2(q; n) & \hat{c}_2(q-1; n) & \dots & \hat{c}_2(1; n) \\ \vdots & & & \\ 0 & \dots & \dots & \hat{c}_2(q-1; n) \end{pmatrix} \quad 3q+1 \times q$$

(16.70)

$$\mathbf{w}(n) = (\epsilon(n), \epsilon b_1(n), \dots, \epsilon b_q(n), -b_1^2(n), \dots, -b_q^2(n))^t \quad \epsilon = \frac{\sigma_w^2}{\gamma_{3w}} \quad (16.71)$$

$$\mathbf{z}(n) = (\hat{c}_2(-q; n), \hat{c}_2(-q+1; n), \dots, \hat{c}_2(q; n), 0, \dots, 0)^t \quad (16.72)$$

where $\hat{c}_k(m; n)$ stands for the cumulant $c_{ky}(m, m, \dots, m)$ estimated using n samples, or equivalently, at time n . The vector $\mathbf{w}(n)$ is the vector of unknowns.

ORIV can solve this set of linear equations if the input data vector, instrumental variable vector and desired response are defined as:

$$\begin{aligned} \tilde{\mathbf{x}}(n) &= [y(n), \dots, y(n-3q)]^t \\ \mathbf{x}(n) &= [y^2(n-q), \dots, y^2(n-2q)|y(n-q-1), \dots, y(n-2q)]^t \\ d(n) &= y(n-q) \end{aligned} \quad (16.73)$$

Analogously, from the generalized GM equation, a set of linear equations can be formed:

$$\Phi(n)\mathbf{w}(n) = \mathbf{z}(n) \quad (16.74)$$

where:

$$\Phi(n) = (\Phi_1(n) \quad | \quad \Phi_2(n))$$

$$\Phi_1(n) = \begin{pmatrix} \hat{c}_4(-q; n) & 0 & \dots & 0 \\ \hat{c}_4(-q+1; n) & \hat{c}_4(-q; n) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \hat{c}_4(q; n) & \hat{c}_4(q-1; n) & \dots & \hat{c}_4(0; n) \\ 0 & \ddots & & \\ \vdots & & \ddots & \\ 0 & \dots & \dots & \hat{c}_4(q; n) \end{pmatrix} \quad 3q+1 \times q+1$$

$$\Phi_2(n) = \begin{pmatrix} 0 & \dots & \dots & 0 \\ \hat{c}_3(-q; n) & \dots & \dots & 0 \\ \hat{c}_3(-q+1; n) & \hat{c}_3(-q; n) & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ \hat{c}_3(q; n) & \hat{c}_3(q-1; n) & \dots & \hat{c}_3(1; n) \\ \vdots & & & \\ 0 & \dots & \dots & \hat{c}_3(q-1; n) \end{pmatrix} \quad 3q+1 \times q$$

(16.75)

$$\mathbf{w}(n) = (\epsilon(n), \epsilon b_1^2(n), \dots, \epsilon b_q^2(n), -b_1^3(n), \dots, -b_q^3(n))^t \quad \epsilon = \frac{\gamma_{3w}}{\gamma_{4w}} \quad (16.76)$$

$$\mathbf{z}(n) = (\hat{c}_3(-q; n), \hat{c}_3(-q+1; n), \dots, \hat{c}_3(q; n), 0, \dots, 0)^t \quad (16.77)$$

And again can be solved using ORIV if the following definitions are made:

$$\begin{aligned} \tilde{\mathbf{x}}(n) &= [y(n), \dots, y(n-3q)]^t \\ \mathbf{x}(n) &= [y^3(n-q), \dots, y^3(n-2q)|y^2(n-q-1), \dots, y^2(n-2q)]^t - \\ &\quad - 3\hat{\sigma}^2(n)[y(n-q), \dots, y(n-2q)|0, \dots, 0]^t \\ d(n) &= y^2(n-q) \end{aligned} \quad (16.78)$$

16.5.3 Simulation results

Equation (16.67) solved by ORIV constitutes an adaptive method based on second and third order cumulants, hence it will be called ORIV-C2C3 [FP89],

for blind identification of linear systems. On the other hand, equation (16.68) is based on third and fourth order cumulants and the method will be called ORIV-C3C4 [KJYS97].

The goal of these simulations is twofold: firstly to show that ORIV-C2C3 and ORIV-C3C4 actually work and secondly to decide which one gives better results under which conditions.

Simulations set-up

1. The input process $\{w(n)\}$ is i.i.d with one-sided exponential probability density function, of zero mean, skewness 2 and kurtosis 6.
2. The additive noise $\{v(n)\}$ is the result of filtering white gaussian noise through an autorregressive moving average ARMA filter with AR= $[1 \quad -2.2 \quad 1.77 \quad -0.52]$ and MA= $[1 \quad -1.25]$.
3. The filter coefficients $\{b_0, \dots, b_q\}$ are all real with $b_0 = 1$.

Identification of stationary systems. Model ER2

For these simulations the data series involved consists in 5000 samples of noisy output $y(n)$. The ORIV algorithm is applied to obtain an estimator of the filter coefficient for every instant of time, so a MSE as a function of n can be plotted (learning curve). 1000 runs are averaged.

As an example consider the MA model MA= $[1 \quad -1.25]$. The mean and variance of the estimators in the last iteration (5000) are shown on tables 16.2 and 16.3 for ORIV-C2C3 and ORIV-C3C4 respectively, for different SNR. Also, the learning curves are shown on figures 16.1(a) and 16.1(b) respectively.

SNR (dB)	Type 1		Type 2	
	b_1	var b_1	b_1	var b_1
20	-1.271	0.047	-1.250	0.036
10	-1.621	0.115	-1.181	0.036
5	-11.014	325.968	0.436	0.618
0	0.390	0.191	0.000	0.068
-5	0.419	0.220	0.000	0.190
-10	1.519	14.154	0.157	0.888

Tabla 16.2 Mean and variance of the estimators. ORIV-C2C3, model ER2

SNR (dB)	Type 1		Type 2	
	b_1	var b_1	b_1	var b_1
20	-1.250	0.035	-1.250	0.053
10	-1.250	0.048	-1.250	0.067
5	-1.270	0.074	-1.230	0.100
0	-1.450	1.230	-1.100	0.237
-5	-0.763	2.440	-0.480	0.648
-10	-0.017	2.160	0.034	0.815

Tabla 16.3 Mean and variance of the estimators. ORIV-C3C4, model ER2

Note that due to the structure of the vectors of unknowns in (16.71) and (16.76) two types of estimators can be obtained.

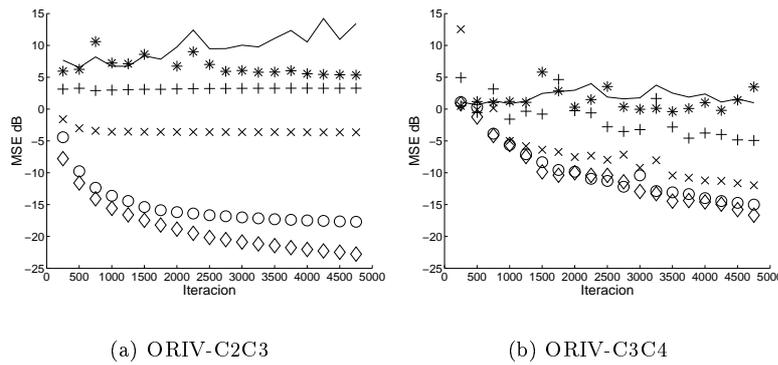


Fig. 16.1 Learning curves for model ER2. \diamond 20 dB; \circ 10 dB; \times 5 dB; $+$ 0 dB; $*$ -5 dB; $-$ -10 dB

Comments on the identification of stationary systems

The most important points to note are the following:

1. Both methods correctly identify non-minimum phase systems because they use higher-order statistics.

2. Other property of higher-order cumulants is that they vanish for gaussian processes. Thus, method ORIV-C3C4 is insensitive to additive gaussian noise, but not ORIV-C2C3 because it uses 2nd as well as third order statistics. This theoretical insensitivity to gaussian noise is translated to the simulations like follows: because cumulants have to be estimated, the estimates of the cumulant will have some variance (even if they are unbiased). Also, the second order cumulant (moment) of the noisy output $y(n)$ will differ from the second order cumulant of the system output $x(n)$, so this cumulant will be biased. In consequence, as the SNR decreases, method ORIV-C3C4 will give unbiased estimates with increasing variance, but method ORIV-C2C3 will give estimates with increasing bias and variance. Tables 16.2 and 16.3 confirm this comment.
3. ORIV-C2C3 converges quicker than ORIV-C3C4 because less samples are needed to estimate lower order statistics.
4. To sum up, ORIV-C3C4 gives better results for $\text{SNR} < 10$ dB.

Identification of non-stationary systems. Model NEL1

To test the tracking capabilities of these adaptive methods, for the next simulation it will be assumed that the filter coefficients evolve with time according to

$$\mathbf{b}(n) = \mathbf{b}_0 + n\mathbf{p}$$

where $\mathbf{b}_0 = [1 \ -1 \ 0.75]$ is the vector of initial values, $\mathbf{p} = [0 \ -1/8000 \ 1/8000]$ is the constant vector of the slope in the linear evolution, and n is the temporal index. The noisy output consists in a 2000 data series and the results are averaged for 1000 independent runs.

For non-stationary environments, the ORIV algorithm needs to set the forgetting factor λ to a value less than 1, in this case $\lambda = 0.9985$ chosen in a trail and error scheme.

The evolution of the mean of the type 2 estimators is shown in figures 16.2(a)-16.2(d) for methods ORIV-C2C3 and ORIV-C3C4. Analogously, the evolution of the MSE is displayed in figures 16.3(a) and 16.3(b) respectively.

Comments on the identification of non-stationary systems

The most important ideas that can be extracted from these simulations are:

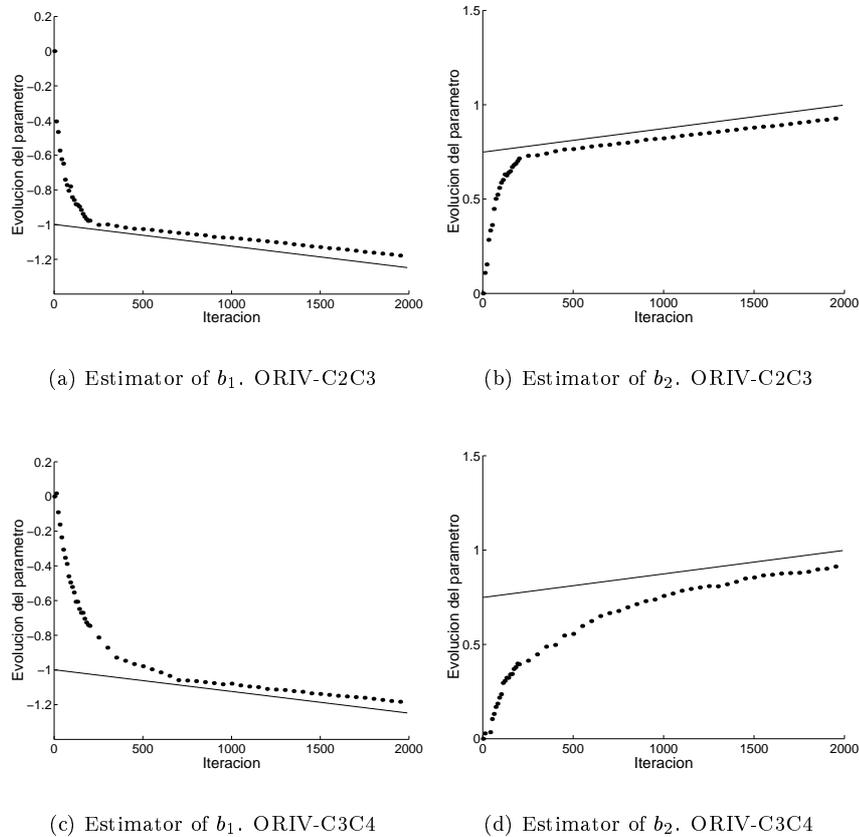


Fig. 16.2 Evolution of the mean of type 2 estimator for model *NEL1*

1. The existence of a delay time is confirmed. The methods never reach the exact value of the system coefficients but are tracking them at a certain distance (see figures 16.2(a) -16.2(d) for instance).
2. Due to the forgetting factor λ the estimation has to be done using a short set of samples (short memory). With this short memory a better job can be done with 2nd and 3rd order cumulants than with 3rd and 4th order cumulants, so method ORIV-C2C3 provides better estimates than ORIV-C3C4 because it has a smaller delay time.
3. Short memory is good for tracking because a more up-to-date information is contained in it, but in the contrary a long memory is required to compute good estimates. The optimum forgetting factor λ_{opt} appears because of these two opposite requirements. In fact these simulations are carried out

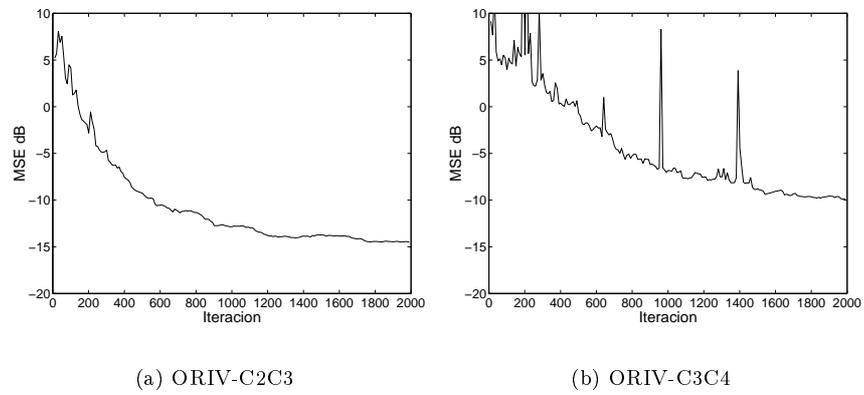


Fig. 16.3 *Learning curves for model NEL1*

with a λ_{opt} common to both methods, searched in a trail-error scheme, as stated before.

4. As it is known from the stationary case, ORIV-C2C3 converges faster to the steady state (of constant delay time).
5. To sum up, ORIV-C2C3 has better properties in non-stationary environments as expected, but, in the case of additive gaussian noise, ORIV-C3C4 should outperform it.

Capítulo 17

The Averaged and Overdetermined Generalised LMS algorithm AOGLMS. A new algorithm

Introduction

In this chapter the overdetermined and generalised orthogonality principle, derived in the previous chapter, is used to propose a number of stochastic gradient type algorithms. The goal of all this work is to reach to a new algorithm that meets all the required properties, i.e. one that works with overdetermined systems of equations and with higher-order statistics. This is tested theoretically and via simulations.

From the overdetermined and generalised orthogonality principle the Overdetermined and Generalised LMS algorithm is derived in section 17.1. This algorithm can be split into two different algorithms as explained in section 17.2, and from one of them the GLMS can be rederived. Except the GLMS algorithm, as shown in section 17.3, the rest of algorithms have the expected properties, but they suffer from a big computational burden. This burden is reduced in section 17.5 through the proposed Averaged, Overdetermined and Generalised LMS AOGLMS algorithm. In this very section the algorithm is studied theoretically and the results checked using simulations.

17.1 The Overdetermined and Generalised Orthogonality Principle

It is known that the Least Mean Square algorithm is built from an orthogonality condition between the input data vector and the estimation error. This reasoning can be generalised, and from any orthogonality condition an adaptive algorithm to impose that condition can be derived. Mathematically, it can be said that if the solution to a certain problem is attained when

$$E[\mathbf{f}(\mathbf{x}(n), e(n))] = 0$$

this recursion converges to the solution

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \mu \mathbf{f}(\mathbf{x}(n), e(n)) \quad (17.1)$$

In the next lines this procedure is used to obtain a new LMS type algorithm.

In the previous chapter it was shown that the ORIV algorithm imposes this condition (16.17):

$$E[\mathbf{X}(n)\mathbf{\Lambda}^{-1}(n)\boldsymbol{\alpha}_0(n)] = 0 \quad (17.2)$$

that can be regarded as an overdetermined orthogonality principle that generalises the one related to the LMS: $E[\mathbf{x}(n)e(n)] = 0$. The overdetermined character comes from the problem it was derived from, because this expression was obtained in the ORIV algorithm analysis that solves overdetermined systems. The matrix $\mathbf{X}(n)$ is the data matrix, the vector $\boldsymbol{\alpha}_0(n)$ is the error vector and the matrix $\mathbf{\Lambda}^{-1}(n)$ connects both quantities. The error vector takes a familiar form:

$$\boldsymbol{\alpha}_0(n) = \mathbf{v}(n) - \mathbf{X}^t(n)\mathbf{w}_0$$

where the different components are defined as:

$$\begin{aligned} \mathbf{X}(n) &= [\boldsymbol{\Phi}^t(n-1)\tilde{\mathbf{x}}(n) \quad \mathbf{x}(n)] \\ \mathbf{\Lambda}^{-1}(n) &= \begin{pmatrix} 0 & \lambda \\ \lambda & \tilde{\mathbf{x}}^t(n)\tilde{\mathbf{x}}(n) \end{pmatrix} \\ \mathbf{v}(n) &= \begin{pmatrix} \tilde{\mathbf{x}}^t(n)z(n-1) \\ d(n) \end{pmatrix} \end{aligned}$$

that depend on the data vector $\mathbf{x}(n)$, the instrumental variable vector $\tilde{\mathbf{x}}(n)$, the desired response $d(n)$, the autocorrelation matrix $\boldsymbol{\Phi}(n)$ between $\mathbf{x}(n)$ and $\tilde{\mathbf{x}}(n)$ and the correlation vector between $\tilde{\mathbf{x}}(n)$ and $d(n)$, $z(n)$. The meaning of all these magnitudes was explained in the previous chapter.

From the preceding comments a new recursive equation for the update of the parameters could be proposed as follows

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \mu \mathbf{X}(n)\mathbf{\Lambda}^{-1}(n)\boldsymbol{\alpha}(n) \quad (17.3)$$

where the error vector shown here is computed using $\mathbf{w}(n-1)$. This recursion would solve the overdetermined system of equations:

$$\mathbf{E}[\Phi(n)]\mathbf{w}_0(n) = \mathbf{E}[z(n)] \quad (17.4)$$

in an adaptive fashion. The statement of that problem, together with its resolution via ORIV was discussed in the previous chapter.

The resulting algorithm, having as a fundamental equation (17.3), is summarised in table 17.1. It will be called overdetermined and generalised LMS algorithm, OGLMS.

<p style="text-align: center;"><u>Initialitation</u></p> <p style="text-align: center;">$\mathbf{w}(0) = 0 \quad \Phi(0) = 0 \quad z(0) = 0$</p> <p style="text-align: center;">For every new $x(n)$, $\tilde{x}(n)$ and $d(n)$ do</p> <hr style="width: 80%; margin: auto;"/> <p style="text-align: center;">Build the vectors $\mathbf{x}(n)$ and $\tilde{\mathbf{x}}(n)$</p> <p style="text-align: center;">$\mathbf{X}(n) = [\Phi^t(n-1)\tilde{\mathbf{x}}(n) \quad \mathbf{x}(n)]$</p> <p style="text-align: center;">$\Lambda^{-1}(n) = \begin{pmatrix} 0 & \lambda \\ \lambda & \tilde{\mathbf{x}}^t(n)\tilde{\mathbf{x}}(n) \end{pmatrix}$</p> <p style="text-align: center;">$\mathbf{v}(n) = \begin{pmatrix} \tilde{\mathbf{x}}^t(n)z(n-1) \\ d(n) \end{pmatrix}$</p> <p style="text-align: center;">$\boldsymbol{\alpha}(n) = \mathbf{v}(n) - \mathbf{X}^t(n)\mathbf{w}(n-1)$</p> <p style="text-align: center;">$\mathbf{w}(n) = \mathbf{w}(n-1) + \mu\mathbf{X}(n)\Lambda^{-1}(n)\boldsymbol{\alpha}(n)$</p> <p style="text-align: center;">$\Phi(n) = \lambda\Phi(n-1) + \tilde{\mathbf{x}}(n)\mathbf{x}^t(n)$</p> <p style="text-align: center;">$z(n) = \lambda z(n-1) + \tilde{\mathbf{x}}(n)d(n)$</p>
--

Tabla 17.1 Iterative process for OGLMS

17.2 Some versions of the OGLMS algorithm

Instead of analysing the behaviour of OGLMS, in this section it will be simplified to yield 4 versions. One of these versions is the Generalised LMS algorithm developed in [AAM96].

To proceed with the simplification the gradient term in equation (17.3) is rewritten as a function of its components,

$$\begin{aligned} \mathbf{X}(n)\Lambda^{-1}(n)\boldsymbol{\alpha}(n) &= \Lambda\mathbf{x}(n)\tilde{\mathbf{x}}^t(n)[z(n-1) - \Phi(n-1)\mathbf{w}(n-1)] + \\ &+ \Phi^t(n)[\tilde{\mathbf{x}}(n)d(n) - \tilde{\mathbf{x}}(n)\mathbf{x}^t(n)\mathbf{w}(n-1)] \end{aligned} \quad (17.5)$$

the right hand side term is made up of two summands, each of which tends to zero when the expected value is taken with n large enough with $\mathbf{w}(n) = \mathbf{w}_0$. That is, only one could be used as a new gradient, simplifying the recursion.

Therefore, four different options can be proposed:

1st option: The first summand is taken as the gradient term,

$$\nabla(n) = \lambda \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) [z(n-1) - \Phi(n-1) \mathbf{w}(n-1)]$$

2nd option: Because the part containing information about the proximity to the solution is the difference, this gradient is proposed

$$\nabla(n) = z(n-1) - \Phi(n-1) \mathbf{w}(n-1)$$

The main difference with previous versions is that now it does not solve an overdetermined problem (just because dimensional homogeneity in (17.1)), so the quantities $\Phi(n)$ and $\mathbf{z}(n)$ have to be redimensioned, reducing the number of equations that the system (17.4) will contained.

This expression leads to a generalised formulation of the least squares problem when an instrumental variable is considered. That is, if the definitions $\mathbf{z}(n)$ and $\Phi(n)$ are taken into account, results:

$$\nabla = \sum_{i=0}^{n-1} \lambda^{n-i-1} \tilde{\mathbf{x}}(i) e(i, n-1)$$

that can be considered an orthogonality principle for deterministic series. This equation imposes orthogonality between the instrumental variable and the error at time i computed with the estimator of time $n-1$ given by $e(i, n-1) = d(i) - \mathbf{x}^t(i) \mathbf{w}(n-1)$.

3rd option: The second summand is taken as the gradient term,

$$\nabla(n) = \Phi^t(n) [\tilde{\mathbf{x}}(n) d(n) - \tilde{\mathbf{x}}(n) \mathbf{x}^t(n) \mathbf{w}(n-1)]$$

4th option: If only the difference part, containing enough information, is preserved,

$$\nabla(n) = \tilde{\mathbf{x}}(n) (d(n) - \mathbf{x}^t(n) \mathbf{w}(n-1))$$

Again, the problem is no more overdetermined and the GLMS algorithm is recovered. The main problem of this algorithm, as will be seen later on, as it is the case for the second option, is that not always converges, because the matrix \mathbf{R} must be positive definite or negative definite. Options 1 and 3 do not suffer from this disadvantage, neither the OGLMS algorithm does as can be tested in (8.6).

This technique, that provides a way of constructing stochastic gradient algorithms from orthogonality principles, has been shown to be very fruitful. Not only the GLMS algorithm has been rederived, provided with a complete formalism, but also 3 new algorithms are proposed. Only one of the laterals will be used to obtain a new algorithm having much more interesting properties.

17.3 The GLMS algorithm

The main goal of this section is to show that GLMS is not able to converge if the matrix \mathbf{R} is not positive definite or negative definite.

The key equation of this algorithm was displayed in option 4, that together with the steps described in table 17.2 form all the necessary iterative process.

<p><u>Initialitation</u></p> <p>$\mathbf{w}(0) = 0$</p> <p>For every new $x(n)$, $\tilde{x}(n)$ and $d(n)$ do</p> <hr style="width: 80%; margin: auto;"/> <p>Build the vectors $\mathbf{x}(n)$ and $\tilde{\mathbf{x}}(n)$</p> <p>$e(n) = d(n) - \mathbf{w}^H(n-1)\mathbf{x}(n)$</p> <p>$\mathbf{w}(n) = \mathbf{w}(n-1) + \mu\tilde{\mathbf{x}}(n)e^*(n)$</p>
--

Table 17.2 *Iterative process of the GLMS algorithm*

As can be seen, the computational burden is very low and the only difference with the LMS algorithm is the use of an instrumental variable $\tilde{x}(n)$ which allows it to work with higher-order statistics. The name GLMS stands because it generalizes LMS. Also, the possibility of complex variables will be considered. Because all the simplifications made on the generalised and overdetermined orthogonality condition (17.2), the orthogonality condition for GLMS becomes:

$$E[\tilde{\mathbf{x}}(n)e^*(n)] = 0 \quad (17.6)$$

17.3.1 Convergence in the mean

Here it is shown that the iterative process displayed on table 17.2 converges in fact in the mean to \mathbf{w}_0 . From the filter weights update equation in that table,

it easily follows that:

$$\begin{aligned} \mathbf{w}(n) - \mathbf{w}_0 = & \mathbf{w}(n-1) - \mathbf{w}_0 + \mu \tilde{\mathbf{x}}(n) d^*(n) - \mu \tilde{\mathbf{x}}(n) \mathbf{x}^H(n) \mathbf{w}(n-1) + \\ & + \mu \tilde{\mathbf{x}}(n) \mathbf{x}^H(n) \mathbf{w}_0 - \mu \tilde{\mathbf{x}}(n) \mathbf{x}^H(n) \mathbf{w}_0 \end{aligned} \quad (17.7)$$

or equivalently

$$\Delta \mathbf{w}(n) = [\mathbf{I} - \mu \tilde{\mathbf{x}}(n) \mathbf{x}^H(n)] \Delta \mathbf{w}(n-1) + \mu \tilde{\mathbf{x}}(n) e_0^*(n) \quad (17.8)$$

where $\Delta \mathbf{w}(n) = \mathbf{w}(n) - \mathbf{w}_0$. This is a non-homogeneous difference equation with stochastic coefficients. The interest lies in the expected value of this equation, for this reason the independence hypothesis are used¹:

Independence hypothesis

1. The input vectors $\mathbf{x}(1)$, $\mathbf{x}(2)$, ..., $\mathbf{x}(n)$ form an statistically independent series of vectors.
2. The instrumental variable vectors $\tilde{\mathbf{x}}(1)$, $\tilde{\mathbf{x}}(2)$, ..., $\tilde{\mathbf{x}}(n)$ form an statistically independent series of vectors.
3. The desired response scalars $d(1)$, $d(2)$, ..., $d(n)$ are a statistically independent series of scalars.
4. The vector series $\{\mathbf{x}(n)\}$ and $\{\tilde{\mathbf{x}}(n)\}$ are jointly stationary.
5. The estimation error $e_0(n)$ is independent of the rest of variables at previous times and is zero mean².
6. The variables $x(n)$, $\tilde{x}(n)$ and $d(n)$ are jointly gaussian.

From equation (17.8) is can be said that $\mathbf{w}(n)$ only depend on samples of $\{d(n)\}$, $\{\mathbf{x}(n)\}$ and $\{\tilde{\mathbf{x}}(n)\}$ up to time n and therefore is independent of $\{d(n+1)\}$, $\{\mathbf{x}(n+1)\}$ and $\{\tilde{\mathbf{x}}(n+1)\}$ by the use of the first three independence hypothesis.

With this in mind, taking expected value to (17.8) and considering (17.6):

$$\mathbf{E}[\Delta \mathbf{w}(n)] = [\mathbf{I} - \mu \mathbf{R}] \mathbf{E}[\Delta \mathbf{w}(n-1)]$$

this homogeneous and deterministic difference equation has one solution that can be quickly computed if it is substituted n times in itself:

$$\mathbf{E}[\Delta \mathbf{w}(n)] = [\mathbf{I} - \mu \mathbf{R}]^n \mathbf{E}[\Delta \mathbf{w}(0)]$$

¹This independence hypothesis are adapted from that used in the analysis of the LMS algorithm and are analogous to those used in the previous chapter.

²The final interpretation for $e_0(n)$ depends on the given problem to solve. Thus, depends on the definitions of $\mathbf{x}(n)$ and de $d(n)$.

working in the base where \mathbf{R} is a diagonal matrix, the components of the weights error vector $\Delta \mathbf{w}(n)$ will uncouple in the previous matrix expression:

$$E[\Delta w_i(n)] = [1 - \mu \lambda_i]^n E[\Delta w_i(0)]$$

turning into a scalar expression. It would be desirable that the solution to this equation tend to 0 when n is large, which can be got if $|1 - \mu \lambda_i| < 1 \quad \forall \lambda_i$ eigenvalue of \mathbf{R} . This condition on the step-size μ is equivalent to:

- i) If $\Re\{\lambda_i\} > 0$ then $0 < \mu < 2\Re\{1/\lambda_i\}$ should hold.
- ii) If $\Re\{\lambda_i\} < 0$ then $0 > \mu > 2\Re\{1/\lambda_i\}$ should hold.

In order to make the whole weights error vector converge to zero, all its components should converge to zero. This can only be accomplished if the matrix $\Re\{\mathbf{R}\}$ have a defined sign. Thus, if it is positive definite, $\mu > 0$ should hold and if negative definite $\mu < 0$ should hold.

These conditions are truly important and impose a strong restriction on the systems that can be solved using GLMS.

17.4 Analysis of the stochastic gradient algorithm based on the 1st option, OGLMS1

As it was commented on the preceding sections, the following recursion solves the system given in (17.4):

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \frac{\mu}{n-1} \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) [z(n-1) - \Phi(n-1) \mathbf{w}(n-1)]$$

This expression forms the fundamentals of an stochastic gradient algorithm that can be built from it, described in table 17.3, based on the overdetermined and generalised orthogonality principle, as it is known.

In the next section it will be shown that this algorithm converges in any circumstance regardless the matrix \mathbf{R} is definite or not, thus outperforming the GLMS algorithm.

<p><u>Initial Conditions</u></p> $\mathbf{w}(0) = 0 \quad \Phi(0) = 0 \quad z(0) = 0$
<p>With every new $x(n)$, $\tilde{x}(n)$ and $d(n)$ do</p> <hr style="width: 80%; margin: auto;"/> <p>Build the vectors $\mathbf{x}(n)$ and $\tilde{\mathbf{x}}(n)$</p>

$$\left[\begin{array}{l} \mathbf{w}(n) = \mathbf{w}(n-1) + \frac{\mu}{n-1} \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) [z(n-1) - \Phi(n-1) \mathbf{w}(n-1)] \\ \Phi(n) = \lambda \Phi(n-1) + \tilde{\mathbf{x}}(n) \mathbf{x}^t(n) \\ z(n) = \lambda z(n-1) + \tilde{\mathbf{x}}(n) d(n) \end{array} \right]$$

Tabla 17.3 Iterative process of the OGLMS1 algorithm

17.4.1 Convergence in the mean for stationary systems

Following similar steps as those given in the analysis of the GLMS algorithm, the previous equation can be rewritten as:

$$\begin{aligned} \Delta \mathbf{w}(n) = & \left[\mathbf{I} - \frac{\mu}{n-1} \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \Phi(n-1) \right] \Delta \mathbf{w}(n-1) + \\ & + \frac{\mu}{n-1} \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) [z(n-1) - \Phi(n-1) \mathbf{w}_0] \end{aligned} \quad (17.9)$$

In this equation n can be supposed to be large because only the analysis in convergence or near convergence is of interest. So, the hypothesis stating that $z(n)$ and $\Phi(n)$ are quasideterministic holds. That is, recalling expressions (16.5) and (16.6):

$$\begin{aligned} \Phi(n) & \sim n [\mathbf{R}_{\bar{x}x} + \mathbf{R}_{\bar{x}x}^e(n)] \\ z(n) & \sim n [\mathbf{r}_{d\bar{x}} + \mathbf{r}_{d\bar{x}}^e(n)] \end{aligned}$$

where the components of $\mathbf{R}^e(n)$ and $\mathbf{r}^e(n)$ are white noise of zero mean and are independent of the rest of stochastic variables involved in the problem. So, considering that $\mathbf{R}_{\bar{x}x} \mathbf{w}_0 = \mathbf{r}_{d\bar{x}}$ this expression can be verified:

$$\begin{aligned} \Delta \mathbf{w}(n) = & \left[\mathbf{I} - \frac{\mu}{n-1} \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \Phi(n-1) \right] \Delta \mathbf{w}(n-1) + \\ & + \mu \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \mathbf{s}^e(n-1) \end{aligned} \quad (17.10)$$

the vector $\mathbf{s}^e(n) = \mathbf{r}^e(n) - \mathbf{R}^e(n) \mathbf{w}_0$ has been defined, containing information related to the estimation noise of the magnitudes \mathbf{r} and \mathbf{R} ; it is of zero mean and independent of the rest of stochastic variables because its components are.

Taking expected value and recalling the independence hypothesis:

$$E[\Delta \mathbf{w}(n)] = [\mathbf{I} - \mu \mathbf{R}^t \mathbf{R}] E[\Delta \mathbf{w}(n-1)]$$

Working in the base where $\mathbf{R}^t \mathbf{R}$ is diagonal, the components of the weights error vector can decouple:

$$E[\Delta \mathbf{w}_i(n)] = [1 - \mu \lambda_i] E[\Delta \mathbf{w}_i(n-1)] \quad (17.11)$$

this difference equation tends to zero if $|1 - \mu\lambda_i| < 1$, that is, if

$$0 < \mu < \frac{\lambda_i}{2} \quad (17.12)$$

where it was taking into account that the matrix $\mathbf{R}^t \mathbf{R}$ is positive definite³ and its eigenvalues are positive.

To check the influence of the eigenvalues on the algorithm convergence, it is convenient to study the solution to (17.11). If this equation is substituted n times in itself, it follows that:

$$\mathbf{E}[\Delta \mathbf{w}_i(n)] = [1 - \mu\lambda_i]^n \mathbf{E}[\Delta \mathbf{w}_i(0)]$$

the closer the difference $1 - \mu\lambda_i$ be to zero, the fewer iterations are needed in order that the i th component of the weights error vector become zero. The speed of convergence for a particular component of the vector can be improved with an adequate choosing of the step-size parameter μ . The problem lies in the fact that a good choosing for a component can be a bad choosing for another (other eigenvalue), or even can diverge if (17.12) does not hold. If the values of the eigenvalues would be in a short interval, a step-size parameter could be chosen to secure a quick convergence of all the components. If the eigenvalue spread is large, the disparity of convergence speeds between components would be large and the overall speed would be determined by the slowest component.

The ORIV algorithm does not suffer from this disadvantage. For this algorithm the convergence in the mean was governed by the next expression (equation (16.16)):

$$\mathbf{E}[\Delta \mathbf{w}(n)] = \lambda^{2n} (1 - \lambda)^2 (\mathbf{R}^t \mathbf{R})^{-1} \mathbf{\Gamma}(0) \Delta \mathbf{w}(0)$$

in stationary environments, where the forgetting factor λ is not required, the change $(1 - \lambda) \rightarrow 1/n$ is made:

$$\mathbf{E}[\Delta \mathbf{w}(n)] = \frac{1}{n^2} (\mathbf{R}^t \mathbf{R})^{-1} \mathbf{\Gamma}(0) \Delta \mathbf{w}(0) \quad (17.13)$$

This expression clearly shows that the speed of convergence does not depend on any quantity (i.e. eigenvalue spread of $\mathbf{R}^t \mathbf{R}$) nor on any algorithm free parameter (because ORIV does not have any in stationary environments). In summary, it is not affected by the eigenvalue spread of the matrix $\mathbf{R}^t \mathbf{R}$. The weights error vector tends to zero like $1/n^2$.

How the computational burden of the OGLMS type algorithm has been achieved was eluding the computation of the matrix $\mathbf{\Gamma}(n)$. This matrix eliminates any dependency on the eigenvalues in the ORIV algorithm as the previous equation shows.

³Strictly speaking, the only secure fact is that it is positive semidefinite, to be positive definite \mathbf{R} should be full rank and that depends on the actual problem to be solved.

17.5 The Averaged, Overdetermined and Generalised LMS algorithm. A new algorithm

The main objective of this section is to reduce the computational burden of the OGLMS1 algorithm, that was proven to have very interesting properties like the ability to solve overdetermined system regardless the matrix \mathbf{R} is definite or not.

The problem we are trying to solve is the same that in previous sections and also the notation will be kept. Thus, recalling the update equation for the filter coefficients in the OGLMS1 algorithm, rewritten here for convenience,

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \frac{\mu}{n-1} \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) [\mathbf{z}(n-1) - \Phi(n-1) \mathbf{w}(n-1)] \quad (17.14)$$

it can be checked that most of the computational burden is concentrated in the computation of the matrix Φ and in the product of this matrix with the vector of unknowns, both of order lq . To reduce the number of operations per iteration the only way is to substitute this operations by any equivalent ones. One possible option is to find a simplified update equation for the averaged error $\bar{e}(n) = \mathbf{z}(n) - \Phi(n) \mathbf{w}(n)$, as shown in the following lines.

Firstly, the update equations of the involved quantities are used to rewrite it as

$$\begin{aligned} \bar{e}(n) = & \lambda \mathbf{z}(n-1) + \tilde{\mathbf{x}}(n) d(n) - \\ & - [\lambda \Phi(n-1) + \tilde{\mathbf{x}}(n) \mathbf{x}^t(n)] [\mathbf{w}(n-1) + \mu \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \bar{e}(n-1)] \end{aligned} \quad (17.15)$$

regrouping terms and recalling the definition of the a priori error, we get to

$$\bar{e}(n) = [\lambda - \Phi(n) \mu \mathbf{x}(n) \tilde{\mathbf{x}}^t(n)] \bar{e}(n-1) + \tilde{\mathbf{x}}(n) e(n)$$

This equations, which is the update equation for $\bar{e}(n)$, means that $\bar{e}(n)$ is a filtered version, averaged, of $\tilde{\mathbf{x}}(n) e(n)$. The filter is an autorregresive AR model of first order with a stochastic matrix coefficient.

To simplify this update process, the matrix should turn into a scalar free parameter of the algorithm:

$$\bar{e}(n) = f \bar{e}(n-1) + (1-f) e(n) \tilde{\mathbf{x}}(n)$$

so it means that the averaged error is defined as

$$\bar{e}(n) = \frac{1-f^n}{1-f} \sum_{i=1}^n f^{n-i} \tilde{\mathbf{x}}(i) e(i) \quad (17.16)$$

or equivalently

$$\bar{e}(n) = \hat{\mathbf{E}}[\tilde{\mathbf{x}}(n) e(n)]$$

therefore, the following equation is proposed as the key equation of the new algorithm

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \mu \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \bar{e}(n) \quad (17.17)$$

and the new algorithm will be called Averaged, Overdetermined and Generalized Least Mean Square AONGLMS algorithm. Note that now no division by n is needed because $\bar{e}(n)$ has been properly normalised. On addition, in some places $\bar{e}(n)$ will be replaced by its equivalent $\hat{E}[\tilde{\mathbf{x}}(n)e(n)]$ for a better interpretation.

The AONGLMS algorithm is displayed on table 17.4.

<p>Initialitation</p> $\mathbf{w}(0) = \mathbf{0}$ $\bar{e}(0) = \mathbf{0}$ <p>On the arrival of a new sample:</p> <p style="padding-left: 2em;">build $\tilde{\mathbf{x}}(n)$, $\mathbf{x}(n)$ y $d(n)$</p> $e(n) = d(n) - \mathbf{w}^t(n-1)\mathbf{x}(n)$ $\bar{e}(n) = f\bar{e}(n-1) + (1-f)\tilde{\mathbf{x}}(n)e(n)$ $\mathbf{w}(n) = \mathbf{w}(n-1) + \mu \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \bar{e}(n)$
--

Table 17.4 Iterative process for the AONGLMS algorithm

17.5.1 Convergence in the mean

The starting equation for this section is equation (17.17), the update equation for the filter coefficients. From it, the update equation of the weights error vector can be obtained, using definition (17.16), to yield:

$$\begin{aligned} \Delta \mathbf{w}(n) = & \Delta \mathbf{w}(n-1) - \mu \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \hat{E}[\tilde{\mathbf{x}}(n) \mathbf{x}^t(n) \Delta \mathbf{w}(n-1)] + \\ & + \mu \mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \hat{E}[\tilde{\mathbf{x}}(n) e_0(n)] \end{aligned} \quad (17.18)$$

Substituting the estimate of the expected value operator by its actual value (an averaged sum of stochastic variables), taking expected value and recalling the orthogonality principle $E[\tilde{\mathbf{x}}(n)e(n)] = 0$, the previous equation turns to

$$\begin{aligned} E[\Delta \mathbf{w}(n)] = & [\mathbf{I} - \mu(1-f)\mathbf{\Gamma}]E[\Delta \mathbf{w}(n-1)] + \\ & + \mu(1-f)\mathbf{R}^t \mathbf{R} \sum_{i=1}^{n-1} f^{n-i} E[\Delta \mathbf{w}(i-1)] + \mu(1-f)E[\mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \tilde{\mathbf{x}}(n) e_0(n)] \end{aligned} \quad (17.19)$$

where $\mathbf{\Gamma} = E[\mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \tilde{\mathbf{x}}(n) \mathbf{x}^t(n)]$, and contains two important factors:

- a) This equation involves all the weights error vectors that have some influence on the weight error vector at time n : from time 1 up to time $n - 1$.
- b) This equation has an independent term, which are the ‘remains’ of an estimate of the expected value operator that should vanish $\hat{E}[\tilde{\mathbf{x}}(n)e_0(n)]$. Due to this independent term the estimated filter coefficient will have a small bias.

The bias of the estimators is $E[\Delta \mathbf{w}(\infty)]$. This value can be obtained from (17.19), as usual, assuming that convergence has been reached so $E[\Delta \mathbf{w}(n)] = E[\Delta \mathbf{w}(n - 1)] = E[\Delta \mathbf{w}(\infty)]$ and solving for $E[\Delta \mathbf{w}(\infty)]$:

$$E[\Delta \mathbf{w}(\infty)] = (1 - f) \frac{\mathbf{I}}{(1 - f)\mathbf{\Gamma} - \mathbf{R}^t \mathbf{R}} E[\mathbf{x}(n) \tilde{\mathbf{x}}^t(n) \tilde{\mathbf{x}}(n) e_0(n)]$$

The bias is an infinitesimal value of order $(1 - f)$ and is independent of μ .

To get to that stationary solution, the root of the characteristic polynomial should be less than 1, otherwise the solution would diverge. The problem here is that the components of the weight error vector do not decouple in the base where the matrix $\mathbf{R}^t \mathbf{R}$ is diagonal. To have an idea of the range of values μ , λ_Γ and λ_R where convergence occurs, the case with only one unknown will be considered. For this special situation, the associate homogeneous equation takes the form of:

$$E[\Delta \mathbf{w}(n)] = [1 - \mu(1 - f)\lambda_\Gamma] E[\Delta \mathbf{w}(n - 1)] - \mu(1 - f) \sum_{i=1}^{n-1} \lambda_R f^{n-i} E[\Delta \mathbf{w}(n - i)]$$

The characteristic polynomial is built assuming that the solution is of the form $E[\Delta \mathbf{w}(n)] = x^n$:

$$x^n = [1 - \mu(1 - f)\lambda_\Gamma] x^{n-1} - \mu(1 - f)\lambda_R f \frac{x^{n-1} - f^{n-1}}{x - f}$$

and given that $f < 1$, for n large enough, it can be rewritten as:

$$x^{n-1} \left[x - 1 + \mu(1 - f)\lambda_\Gamma + \mu \frac{(1 - f)f\lambda_R}{x - f} \right] = 0$$

that apart from the trivial solution $x = 0$, other solutions are the roots of this polynomial:

$$x^2 + [\mu(1 - f)\lambda_\Gamma - (1 + f)]x + f[1 - \mu(1 - f)(\lambda_\Gamma - \lambda_R)] = 0$$

The solutions of this quadratic equation are:

$$x = \frac{1 + f - \mu(1 - f)\lambda_\Gamma}{2} \pm \frac{\sqrt{(1 + f - \mu(1 - f)\lambda_\Gamma)^2 - 4f[1 - \mu(1 - f)(\lambda_\Gamma - \lambda_R)]}}{2} \quad (17.20)$$

and x is a function of $x = x(\mu\lambda_R, \mu\lambda_\Gamma, f)$, from which some comments could be made:

- a) When $f \rightarrow 1$ (infinite memory) it follows that $x \rightarrow 1$ and there is no convergence.
- b) When the forgetting factor f is close to zero, the roots of the polynomial are $x = 1 - \mu\lambda_\Gamma$ (apart from the trivial solution). Under this condition a new algorithm can be derived that converges even in the case of non definite \mathbf{R} , having almost the same computational burden that the GLMS. This new algorithm is called CGLMS and is described in the main body of the Thesis (page 286).

17.5.2 Convergence in the mean square

To study how the weights error vector converges in the mean square, the starting equation is equation (17.18) rewritten here for clarity reasons

$$\begin{aligned} \Delta \mathbf{w}(n) = & \Delta \mathbf{w}(n-1) - \mu \mathbf{x}(n) \tilde{\mathbf{x}}(n) \hat{\mathbf{E}}[\tilde{\mathbf{x}}(n) \mathbf{x}^t(n) \Delta \mathbf{w}(n-1)] + \\ & + \mu \mathbf{x}(n) \tilde{\mathbf{x}}(n) \hat{\mathbf{E}}[\tilde{\mathbf{x}}(n) e_0(n)] \end{aligned} \quad (17.21)$$

It was shown in the previous section that the great difference between this analysis and the analysis of algorithms like the LMS, is how to treat the estimate of the expected value operator. It should yield simple enough mathematical expressions in order to allow some computations, but also it should bear enough information in order to get meaningful results. It can be checked that if the estimate is treated as a real expected value (i.e. independent of the rest of stochastic variables) no divergence is obtained from the equations.

In the next few lines the objective is to describe a first approximation to solve this problem. The estimate will be treated according to its definition, an averaged sum of stochastic variables. Also the independence hypothesis will be used. Therefore, the first of the terms of interest can be written as:

$$\begin{aligned} \hat{\mathbf{E}}[\tilde{\mathbf{x}}(n) \mathbf{x}^t(n) \Delta \mathbf{w}(n-1)] &= \frac{1-f}{1-f^n} \sum_{i=1}^n f^{n-i} \tilde{\mathbf{x}}(i) \mathbf{x}^t(i) \Delta \mathbf{w}(i-1) = \\ &= \frac{1-f}{1-f^n} \left(\tilde{\mathbf{x}}(n) \mathbf{x}^t(n) \Delta \mathbf{w}(n-1) + \sum_{i=1}^{n-1} f^{n-i} \tilde{\mathbf{x}}(i) \mathbf{x}^t(i) \Delta \mathbf{w}(i-1) \right) = \\ &= \frac{1-f}{1-f^n} \tilde{\mathbf{x}}(n) \mathbf{x}^t(n) \Delta \mathbf{w}(n-1) + \frac{1-f^{n-1}}{1-f^n} f \hat{\mathbf{E}}[\tilde{\mathbf{x}}(n-1) \mathbf{x}^t(n-1) \Delta \mathbf{w}(n-2)] \end{aligned} \quad (17.22)$$

where the part of the estimator with a bigger dependency on the rest of variables has been isolated. Once this is done, some simplifying hypothesis will be applied; in particular the estimator will be treated as a real expected value and, because n is considered to be large enough and the algorithm converges in the mean, it will be assumed that $E[\Delta \mathbf{w}(n-1)] = 0$. That is,

$$\begin{aligned} \hat{E}[\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)\Delta \mathbf{w}(n-1)] &\underset{n \rightarrow \infty}{\sim} (1-f)\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)\Delta \mathbf{w}(n-1) + \\ &+ f\mathbf{R} \sum_{i=1}^{n-1} f^{n-1-i} E[\Delta \mathbf{w}(i-1)] = \quad (17.23) \\ &= (1-f)\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)\Delta \mathbf{w}(n-1) \end{aligned}$$

Although it seems a very severe approximation, the final decision on its utility will be made in light of the obtained results. The meaning of this interpretation is quite transparent: ideally the estimate of the expected value is equal to the expected value of the given stochastic variable; that expected value, in this case, due to the independence hypothesis and to the convergence in the mean is zero,

$$\hat{E}[\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)\Delta \mathbf{w}(n-1)] = E[\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)\Delta \mathbf{w}(n-1)] = \mathbf{R}E[\Delta \mathbf{w}(n-1)] \underset{n \rightarrow \infty}{\rightarrow} 0$$

but the estimator, as such, presents an stochastic component which represents the term given by equation (17.23). A similar reasoning could be made to reach an analogous result for the second term of interest,

$$\hat{E}[\tilde{\mathbf{x}}(n)e_0(n)] \underset{n \rightarrow \infty}{\sim} (1-f)\tilde{\mathbf{x}}(n)e_0(n) \quad (17.24)$$

with the same meaning of the previous approximation: its mean value (zero again for this case) plus its corresponding stochastic component because it is an estimator.

Therefore, with all the comments up to this point, the difference equation that controls the evolution of the weights error vector can be written as,

$$\begin{aligned} \Delta \mathbf{w}(n) = &[\mathbf{I} - \mu(1-f)\mathbf{x}(n)\tilde{\mathbf{x}}^t(n)\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)]\Delta \mathbf{w}(n-1) + \\ &+ \mu(1-f)\mathbf{x}(n)\tilde{\mathbf{x}}^t(n)\tilde{\mathbf{x}}(n)e_0(n) \end{aligned} \quad (17.25)$$

Using the usual independence hypothesis, from this equation it follows that,

$$\begin{aligned} \mathbf{R}_\Delta(n) = &\mathbf{R}_\Delta(n-1) - \mu(1-f)\mathbf{\Gamma}\mathbf{R}_\Delta(n-1) - \mu(1-f)\mathbf{R}_\Delta(n-1)\mathbf{\Gamma} + \\ &+ \mu^2(1-f)^2 E[\boldsymbol{\psi}_1(n)\mathbf{R}_\Delta(n-1)\boldsymbol{\psi}_1(n)] + \mu^2(1-f)^2 E[e_o^2(n)\boldsymbol{\psi}_2(n)] \end{aligned} \quad (17.26)$$

where $\boldsymbol{\psi}_1(n) = \mathbf{x}(n)\tilde{\mathbf{x}}^t(n)\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)$; $\boldsymbol{\psi}_2(n) = \mathbf{x}(n)\tilde{\mathbf{x}}^t(n)\tilde{\mathbf{x}}(n)\tilde{\mathbf{x}}^t(n)\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)$ and $\mathbf{\Gamma} = E[\boldsymbol{\psi}_1(n)]$. Considering this new expression one more simplification could be made; given that the step-size is usually small and that the forgetting factor is close to one, the product $\mu(1-f)$ can be regarded as a perturbation

parameter and in this way the fourth summand of the second term can be neglected. The last summand can not be neglected because it plays a very important role in the difference equation, because it prevent the trivial solution to be a solution to this equation.

Working also in the base where the matrix $\mathbf{\Gamma}$ is diagonal⁴, the different diagonal components $D'_i(n)$ of $\mathbf{R}'_{\Delta}(n)$ decouple, resulting for everyone of them,

$$D'_i(n) = [1 - 2\mu(1 - f)\lambda_{\Gamma i}]D'_i(n - 1) + \mu^2(1 - f)^2\Psi'_{eii}$$

where $\lambda_{\Gamma i}$ is the i th eigenvalue of the matrix $\mathbf{\Gamma}$ and Ψ'_{eii} is the (i, i) element of the matrix $\Psi'_e = E[e_0^2(n)\psi_2(n)]$. Arriving finally to a scalar difference equation which can be easily solved, and from which aspects such as the convergence conditions, convergence time, etc. can be studied.

- A) The general solution of the homogeneous part will converge if $|1 - 2\mu(1 - f)\lambda_{\Gamma i}| \leq 1$ holds. Or equivalently, if

$$0 < \mu(1 - f) < \lambda_{\Gamma i}$$

- B) The convergence time, in this approximation of small step-size, is

$$\tau_i = \frac{1}{2\mu(1 - f)\lambda_{\Gamma i}}$$

- C) The i th diagonal component, that is (i, i) , converges to,

$$D'_i(\infty) = \frac{\mu(1 - f)}{2} \frac{\Psi'_{eii}}{\lambda_{\Gamma i}}$$

The goal of this section is to compute the Mean Square Deviation, MSD, $D(n) = \text{Tr}\{\mathbf{R}_{\Delta}(n)\}$ and in particular its convergence value and the time it takes to converge. The convergence value can be directly computed from C):

$$\begin{aligned} D(\infty) &= \text{Tr}\{\mathbf{R}_{\Delta}(\infty)\} = \text{Tr}\{\mathbf{R}'_{\Delta}(\infty)\} = \sum_{i=1}^q D'_i(\infty) = \frac{\mu(1 - f)}{2} \sum_{i=1}^q \frac{\Psi'_{eii}}{\lambda_{\Gamma i}} = \\ &= \frac{\mu(1 - f)}{2} \sum_{i=1}^q (\mathbf{\Gamma}'^{-1} \Psi'_e)_{ii} = \frac{\mu(1 - f)}{2} \text{Tr}\{\mathbf{\Gamma}'^{-1} \Psi'_e\} = \\ &= \frac{\mu(1 - f)}{2} \text{Tr}\{\mathbf{\Gamma}^{-1} \Psi_e\} \end{aligned} \tag{17.27}$$

⁴the ' denotes quantities in this new base.

And the overall convergence time can be computed as the averaged convergence time of all its components:

$$\begin{aligned}\tau &= \frac{1}{2\mu(1-f)} \frac{1}{q} \sum_{i=1}^q \lambda_{\Gamma^i}^{-1} = \frac{1}{2\mu(1-f)} \frac{1}{q} \text{Tr}\{\mathbf{\Gamma}'^{-1}\} = \\ &= \frac{1}{2\mu(1-f)} \frac{1}{q} \text{Tr}\{\mathbf{\Gamma}^{-1}\}\end{aligned}\quad (17.28)$$

As can be seen from the results of this section, the mean square behaviour of the AOGLMS algorithm depends, as far as the free parameters is concerned, on the product $\mu(1-f)$. This parameter has the same role that μ alone has in the LMS type algorithms: when it takes big values the convergence time decreases, but the convergence value increases. There are opposite interests between a small convergence time and a small convergence value, only one can be achieved at a time.

17.5.3 Analysis of the simulations

Once that the AOGLMS algorithm has been derived and that convergence in the mean and mean square have been proven theoretically, it is time to test the algorithm via simulations. It should be noted that only a brief description of the simulations is presented, the main work in this chapter has been the development of this new algorithm.

Here the problem of the blind identification of MA systems is considered using the Giannakis-Mendel equation. This very problem was considered in the analysis of the ORIV algorithm in the previous chapter. In the table 17.5 the coefficients of the used models are displayed (they are the same that the used in the ORIV analysis), together with the eigenvalues of the matrix $\mathbf{R}^t \mathbf{R}$, keeping in mind the following points:

1. The AOGLMS algorithm will be implemented in an overdetermined form, so it can be used to solve the $3q+1$ equations contained in the GM system of equations to identify the $2q+1$ unknowns of the problem.
2. To solve the problem it suffices to define the variables $\tilde{\mathbf{x}}(n)$, $\mathbf{x}(n)$ and $d(n)$ that appear in the table 17.4 as they were define for the ORIV algorithm.
3. So, the behaviour of the algorithm is determined by the eigenvalue spread of the matrix $\mathbf{R}^t \mathbf{R}$ where $\mathbf{R} = \text{E}[\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)]$.

Model	Coefficients	Eigenvalue spread
A	[1 - 0.8]	3.2211

B	[1 - 1.25]	13.3117
---	------------	---------

Tabla 17.5 *MA models identified using AOGLMS*

The MA models are driven by an i.i.d process, consisting in stochastic variables with an one-sided exponential probability density function, with zero mean, variance 1, skewness 2 and kurtosis 6. Series of 40000 and 100000 samples are available and the results are averaged over 200 independent runs. The additive noise was generated filtering white gaussian noise through an ARMA filter with coefficients MA= [1 - 1.25] and AR= [1 - 2.2 1.77 - 0.52] and the signal to noise ratio levels were 20, 15, 10, 5 y 0 dB.

The displayed figures are the best of a searching process: for each model and SNR the step-size parameter was varied according to the function $\mu = 10^{ex}$ where $ex = -6, -5.5, -5, \dots, -3.5$ and the forgetting factor f took the values $f = 0.9, 0.99$ and 0.999 . In figure 17.1(a) the evolution of the MSD for model A is shown for the indicated SNR, when $\mu = 10^{-4.5}$ and $f = 0.999$. The reading of the figures should be made attending to these relationships:

◇...20 dB
 ·...15 dB
 ○...10 dB
 × ...5 dB

Analogously, in figure 17.1(b) the results for model A but with $\mu = 10^{-4}$ and $f = 0.999$ are displayed. In both figures common aspects to the stochastic gradient type algorithms and those working with higher-order statistics can be spotted, like the fact that the smaller the step-size is $D(\infty)$ attains a smaller value; in turn are less sensitive to the SNR than other algorithms of this type but based exclusively on second order statistics. Anyway, the main goal of these figures is to show that AOGLMS converges in fact for these models, a very important point indeed because GLMS can not identify them as it will be proven in a few lines.

The evolution of the MSD for model B and for the same pairs of parameter values μ and f than for model A are presented in figures 17.1(c) and 17.1(d). It has a worse behaviour because the eigenvalue spread is bigger.

It would be very interesting to compare the behaviour of AOGLMS with that of GLMS under the same conditions, in the blind identification of the MA models using the GM equation. The first thing to note is that GLMS can not

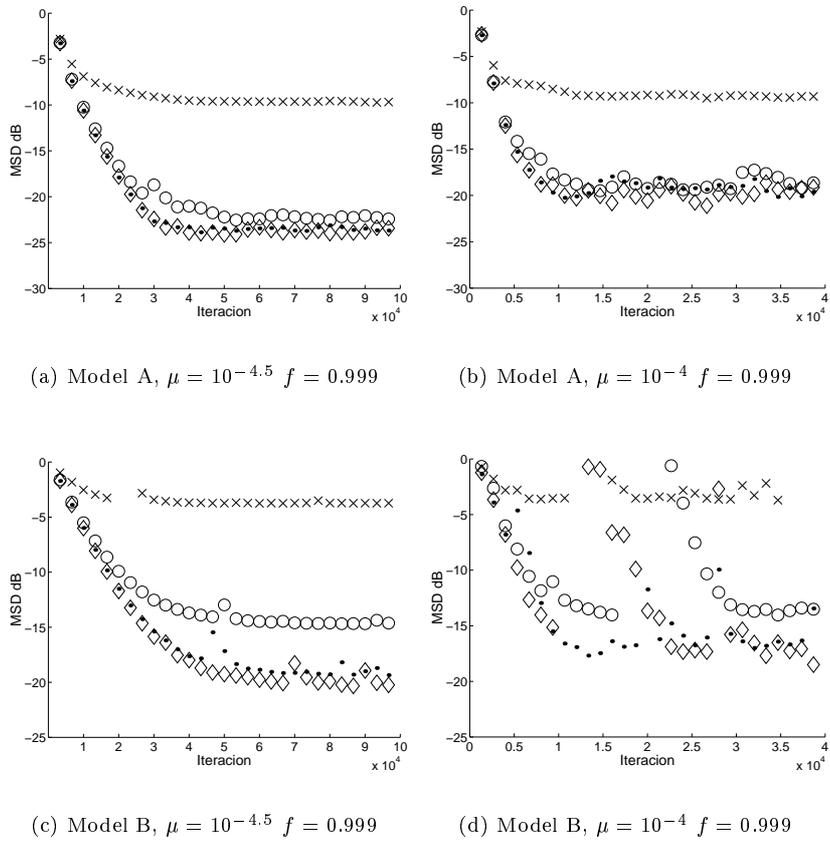


Fig. 17.1 Evolution of the MSD for models A and B

solve overdetermined systems of equations, so the length of the instrumental variable vector has to be shortened to match that of the data vector $\mathbf{x}(n)$, in that way the same number of unknowns and equations is attained. Furthermore, the convergence of the algorithm depends on the eigenvalues of the squared matrix $\mathbf{R} = E[\tilde{\mathbf{x}}(n)\mathbf{x}^t(n)]$. The spread of the real part of the eigenvalues for each model is detailed here:

modelo A	...	-0.8611
modelo B	...	-1.0213

For those models with a negative spread is not possible that GLMS converge, because that indicates that some eigenvalues has a negative real part and others a positive real part.

With these simulations two important ideas have been tested:

1. The AOGLMS algorithm can solve overdetermined systems. Unlike the GLMS algorithm.
2. The AOGLMS can solve systems with the matrix \mathbf{R} , associated to the problem, being not positive definite or not negative definite. Unlike the GLMS algorithm.

Capítulo 18

Blind identification of FIR filters using second-order statistics. Phase selection methods

Introduction

The problem of blind system identification has been addressed before in an adaptive fashion. In this chapter this problem will be studied but in batch mode, in a non-adaptive way. The interest lies in solving the problem benefiting from all the good things that second and higher order statistics can offer. The later provides phase information and the former the magnitude response of the system. To minimize the influence of the big variance associated with higher-order statistics based techniques only a sign (just a binary decision) is computed from this kind of statistics. This new method is based on explicit expressions relating the filter coefficients with the autocorrelation output and its compared with other methods relying in the Spectrally Equivalent Minimum Phase SEMP system. The performance of all these methods is explained.

The problem statement is given in section 18.1. The SE set for a first order channel is computed in section 18.2, the influence on these expressions of the errors in the autocorrelation sequence estimation is discussed in section 18.3. Analogously, the SE set for a 2nd order channel is computed in section 18.4, the sensitivity of these solutions is considered in section 18.5 and the meaning

of these sensitivity conditions is described in section 18.6. On all this basis, the new method for a first order channel is presented in section 18.7 and in section 18.8 for a second order channel. Finally, simulations are carried out for this new method and for a whole variety of methods based on the SEMP system in section 18.9.

18.1 Problem statement

Consider a linear filter given by its finite impulse response $\{h(n)\}$ excited by an independent and identically distributed input random process $\{w(n)\}$ of known statistics. The output of this filter is

$$x(n) = \sum_{i=0}^q h(i)w(n-i) \quad (18.1)$$

where q is the model order, where in general can be contaminated by additive white gaussian noise $\{v(n)\}$ of known variance. So that the total system output is

$$y(n) = x(n) + v(n) \quad (18.2)$$

The problem is to determine the system coefficients by measurements of the system output alone under the previous assumptions. Only the second order statistics information will be used in a first instance. In the following the autocorrelation sequence of the output is assumed known.

18.2 First order model

Let the system be defined by $\mathbf{h} = [h_0 \ h_1]$, so that the relation between the autocorrelation sequence of the output and these coefficients is

$$r_0 = \sigma_w^2(h_0^2 + h_1^2) \quad (18.3)$$

$$r_1 = \sigma_w^2 h_0 h_1 \quad (18.4)$$

where σ_w^2 is the variance of the input process. In the following a normalised autocorrelation sequence will be assumed, or equivalently a unit variance. It is not difficult to show that the filter coefficients satisfy the following equations

$$\hat{h}_0^2 = \frac{r_0 \pm \sqrt{r_0^2 - 4r_1^2}}{2} \quad (18.5)$$

$$\hat{h}_1 = \frac{r_1}{\hat{h}_0} \quad (18.6)$$

So that there are 4 different solutions to the problem: 2 from the \pm option in the expression for h_0^2 and 2 more because of the freedom in the choice of the sign once the square root is taken to h_0^2 . Basically there are two solutions, the other 2 are their opposites changing the sign of the estimated coefficients. If the input signal is symmetric, there is no way to distinguish between a filter or its opposite, because their outputs are statistically equivalent, so that only the positive values for h_0 will be considered.

Denoting by h_0^+ (h_0^-) the result of expression (18.5) taking the $+$ ($-$) sign, and h_1^+ (h_1^-) the results of (18.6) using h_0^+ (h_0^-), it is easy to show that $h_1^+ h_1^- = r_1$, so that

$$h_1^- = \frac{r_1}{h_1^+} = r_1 \frac{h_0^+}{r_1} = h_0^+$$

and analogously

$$h_1^+ = h_0^-$$

With this property the system of equations can be solved, with the solutions shown in table 18.1.

	1 st coefficient	2 nd coefficient
1st option	h_0	h_1
2nd option	h_1	h_0

Tabla 18.1 *Spectrally equivalent channels for a first order system*

This table shows that only one of the possible solutions of (18.5)-(18.6) need to be computed, the others are obtain from it. Similar scheme will be used for higher order filters to simplify the problem. This result could be also obtained by reflecting the zero of the first option with respect to the unit circle, keeping the same output power. As expected the solutions to equations (18.3)-(18.4) are all equivalent systems that have the same second order statistics i.e. power spectrum, but can have different phase response because its higher order statistics are not the same.

18.2.1 Characteristics of the solutions

Using the definitions of r_0 and r_1 given in (18.3) and (18.4) in the expression for \hat{h}_0 given in (18.5), results

$$\hat{h}_0^2 = \frac{h_1^2 + h_0^2 \pm \sqrt{(h_0^2 + h_1^2)^2 - 4h_0^2 h_1^2}}{2} = \frac{h_1^2 + h_0^2 \pm \sqrt{(h_0^2 - h_1^2)^2}}{2} \quad (18.7)$$

where the hat in h_0 denotes that it can be any estimate of the valid options for h_0 . According to this expression the radicand will be close to zero if $h_0^2 \sim h_1^2$, an estimate of this quantity could possibly be negative, so the channel coefficient could be complex even if its real value is real.

Choosing the + sign in (18.7) it follows that $\hat{h}_0^{+2} = h_0^2$ and that $\hat{h}_0^{-2} = h_1^2$ for the - sign, so these equations confirm that these solutions are blind to a change in sign in the filter coefficients. Using these results in (18.6), the possible values for the second filter coefficient are $\hat{h}_1^+ = h_1$ and $\hat{h}_1^- = h_0$. Confirming the 2 options displayed in table 18.1.

18.3 Sensitivity of the solutions to small variations in the autocorrelation sequence estimate

In actual applications the autocorrelation sequence has to be estimated from the channel output. The longer the data record available, the better the estimates, but in any case, there will be some difference between the estimated and theoretical ones. In this section the influence on the solution of this (small) difference will be studied. It will be done using partial derivatives of equations (18.5) and (18.6).

18.3.1 Variations for the h_0 coefficient

With respect to r_0

The partial derivative of \hat{h}_0 with respect to r_0 is

$$\frac{\partial \hat{h}_0}{\partial r_0} = \frac{\sqrt{2}}{4} \left[1 \pm \frac{r_0}{\sqrt{r_0^2 - 4r_1^2}} \right] \frac{1}{\left(r_0 \pm \sqrt{r_0^2 - 4r_1^2} \right)^{1/2}} \quad (18.8)$$

that can have a large value if any of the following happens:

case a) $r_0^2 \approx 4r_1^2$ that makes the first factor of (18.8) diverge. What does this condition impose on the actual filter coefficients? Using again the relations given in (18.3) and (18.4), it follows that $(h_0^2 + h_1^2)^2 \approx 4h_0^2h_1^2$ or equivalently $(h_0^2 - h_1^2)^2 \approx 0$ which leads to $|h_0| \approx |h_1|$. Under this

condition equation (18.8) takes the form of

$$\frac{\partial \hat{h}_0}{\partial r_0} \sim \frac{1}{4} \left[1 \pm \frac{r_0}{\epsilon} \right] \frac{\sqrt{2}}{(r_0 \pm \epsilon)^{1/2}} \approx \frac{1}{4} \left[1 \pm \frac{r_0}{\epsilon} \right] \frac{\sqrt{2}}{\sqrt{r_0}}$$

where $\epsilon \sim 0$. The only way of avoiding the first factor to diverge is setting r_0 close to zero, but this arrangement also causes the second factor to diverge, so divergence is unavoidable. In addition if r_0 is close to zero h_0 or h_1 have to be close to zero, which under the above condition means that both of them have to be close to zero, so that we are left with no system.

Case b) The second situation worth studying is that in which $r_0 \sim \pm \sqrt{r_0^2 - 4r_1^2}$, that is, when \hat{h}_0 is close to zero. Taking a look to table 18.1 this can be achieved if either h_0 or h_1 are close to zero, which means that r_1 is close to zero. Under this condition equation (18.8) takes the form of

$$\frac{\partial \hat{h}_0}{\partial r_0} \underset{r_1 \sim 0}{\sim} \frac{1}{4} \left[1 \pm \left(1 + 2 \frac{r_1^2}{r_0^2} \right) \right] \frac{\sqrt{2}}{\sqrt{r_0} \sqrt{1 \pm (1 - 2(r_1/r_0)^2)}} \quad (18.9)$$

where the Taylor expansion of $\sqrt{1+x}$ and $(1+x)^{-1}$ has been employed.

case b.1) Choosing the + sign, the partial derivative is

$$\frac{\partial \hat{h}_0}{\partial r_0} \sim \frac{1}{2} \frac{1}{\sqrt{r_0}}$$

if r_0 is finite also is the derivative. This expression can diverge if r_0 goes to zero, but it implies that both h_0 and h_1 are close to zero, i.e. there is no system. So, there is no problem for case b.1).

case b.2) Choosing the - sign in (18.9), the partial derivative can be written as

$$\frac{\partial \hat{h}_0}{\partial r_0} = \frac{-1}{2} \frac{r_1}{r_0^{3/2}}$$

which can have some problems when r_0 is close to zero but again this leads to a system with coefficients close to zero. Again, there is no divergence problem for case b.2).

Similar analysis can be carried out for the partial derivative of \hat{h}_0 with respect to r_1 and for the partial derivatives of \hat{h}_1 with respect to r_0 and r_1 , resulting in a range of situations of special sensitivity. These situations are displayed on table 18.2.

∂/∂	h_0	h_1
r_0	$h_0^2 \sim h_1^2$	$h_0^2 \sim h_1^2$ $\hat{h}_0 \sim 0$

r_1	$h_0^2 \sim h_1^2$	$\begin{matrix} h_0^2 \sim h_1^2 \\ \hat{h}_0 \sim 0 \end{matrix}$
-------	--------------------	--

Tabla 18.2 *High sensitivity conditions for the estimation*

18.4 Second order systems

Let us suppose that the channel of interest now is a second order linear time invariant system defined by its impulse response $\mathbf{h} = [h_0 \ h_1 \ h_2]$. The equation that relates these coefficients with the second order statistics of the output is

$$r_m = \sigma_w^2 \sum_{i=0}^2 h_i h_{i+m}$$

or in an explicit form for each possible value of m

$$r_0 = \sigma_w^2 (h_0^2 + h_1^2 + h_2^2) \quad (18.10)$$

$$r_1 = \sigma_w^2 (h_0 h_1 + h_1 h_2) \quad (18.11)$$

$$r_2 = \sigma_w^2 (h_0 h_2) \quad (18.12)$$

Again, as was done with the first order model a unit variance input process will be assumed (or the autocorrelation sequence renormalised divided by the input variance). The problem is to compute the filter coefficients once the autocorrelation sequence is known.

The coefficients can be computed as follows. From (18.12) it's direct to see that

$$\hat{h}_2 = \frac{r_2}{\hat{h}_0} \quad (18.13)$$

using this equation together with (18.11), it follows that

$$\hat{h}_1 = \frac{r_1}{\hat{h}_0 + r_2/\hat{h}_0} \quad (18.14)$$

In order to compute h_1 and h_2 , h_0 is needed. This can be done using (18.13) and (18.14) in (18.10) to give

$$r_0 = \hat{h}_0^2 + \frac{r_1^2 \hat{h}_0^2}{(\hat{h}_0^2 + r_2)^2} + \frac{r_2^2}{\hat{h}_0^2} \quad (18.15)$$

Solving (18.15) for h_0 and substituting in (18.13) and (18.14) the whole system can be identified. The problem with equations (18.13)-(18.15) is that a discussion is not straight forward, so that a different expression for them could be more satisfactory. In order to do that consider the square of (18.11)

$$r_1^2 = h_0^2 h_1^2 + h_1^2 h_2^2 + 2h_1^2 h_0 h_2 = h_1^2 (r_0 - h_1^2 + 2r_2) \quad (18.16)$$

and solving for h_1 gives

$$\hat{h}_1^2 = r_2 + r_0/2 \pm \sqrt{(r_2 + r_0/2)^2 - r_1^2} \quad (18.17)$$

once h_1 is known, eq. (18.14) can be used to compute h_0 , giving

$$\hat{h}_0 = \frac{r_1 \pm \sqrt{r_1^2 - 4\hat{h}_1^2 r_2}}{2\hat{h}_1} \quad (18.18)$$

and finally, (18.13) can be used to obtain h_2 from h_0 .

18.4.1 Discussion of the solution

In order to discuss the solution the second method will be used. From (18.17) 4 different values of h_1 are obtained: h_{11} , h_{12} , $-h_{11}$ and $-h_{12}$, corresponding to the choice of + or - in (18.17) and the election of the + or - once the square root is taken to compute h_1 from h_1^2 .

From them, 8 different values for h_0 can be calculated using (18.18), two for every previous value of h_1 . But due to the structure of (18.18) 4 are the opposite to the other 4, as shown in table 18.3

Values of h_1	Obtained values of h_0
h_{11}	h_{01} and h_{02}
h_{12}	h_{03} and h_{04}
$-h_{11}$	$-h_{01}$ and $-h_{02}$
$-h_{12}$	$-h_{03}$ and $-h_{04}$

Tabla 18.3 Values for h_0 derived from those for h_1

And similarly, for every value of h_0 a value of h_2 is obtained, where again half of them are the opposites of the other half using (18.13). From this discussion turns out that the solutions to equations (18.10)-(18.12) are 8 different systems, where half of them equal the other half with the sign changed. For symmetric

input processes there is no way to distinguish between a given system and its opposite, so only the first 4 solutions will be considered hereafter.

18.4.2 Characteristics of the solution

In what follows all the quantities involved are assumed to be real.

Possible values for \hat{h}_1 Using the relations between the filter coefficients and the autocorrelation sequence in (18.10)-(18.12), eq. (18.17) can be written in the following way

$$\hat{h}_1^2 = \frac{1}{2} \left[(h_0 + h_1)^2 + h_1^2 \pm \sqrt{((h_0 + h_2)^2 - h_1^2)^2} \right] \quad (18.19)$$

the radicand of this expression can be close to zero if $(h_0 + h_2)^2 \sim h_1^2$. But the estimate of a quantity close to zero can easily be negative so that a complex value can arise for \hat{h}_1 when a real value is expected. Situations like this are to be skipped. Continuing with the operations indicated in (18.19), two different solutions can be obtained:

$$\hat{h}_1^2 = h_1^2 \quad (18.20)$$

$$\hat{h}_1^2 = (h_0 + h_2)^2 \quad (18.21)$$

Possible values for \hat{h}_0 Using the definitions of the autocorrelation sequence once again in (18.18), the estimate of the coefficient h_0 is

$$\hat{h}_0 = \frac{h_0 h_1 + h_1 h_2 \pm \sqrt{(h_0 h_1 - h_1 h_2)^2 - 4\hat{h}_1^2 h_0 h_2}}{2\hat{h}_1} \quad (18.22)$$

under (18.20) this expression turns to

$$\hat{h}_0 = \frac{h_0 h_1 + h_1 h_2 \pm \sqrt{(h_0 - h_2)^2 h_1^2}}{2h_1}$$

where again there is more chances to get an undesired complex value if the radicand is close to zero, in this case it can happen if $h_0 \sim h_2$. From this, two different values for the filter coefficient can be obtained

$$\hat{h}_{01} = h_0 \quad (18.23)$$

$$\hat{h}_{02} = h_2 \quad (18.24)$$

On the other hand, under (18.21) eq. (18.22) turns to

$$\hat{h}_0 = \frac{h_1 \pm \sqrt{h_1^2 - h_0 h_2}}{2}$$

from which two different values can be obtained

$$\hat{h}_{03} = \frac{h_1 + \sqrt{h_1^2 - h_0 h_2}}{2} \quad (18.25)$$

$$\hat{h}_{04} = \frac{h_1 - \sqrt{h_1^2 - h_0 h_2}}{2} \quad (18.26)$$

Possible values for \hat{h}_2 Finally, using equation (18.13) the four possible values for \hat{h}_2 can be obtained from the four values of \hat{h}_0 , these are:

$$\hat{h}_{21} = h_2 \quad (18.27)$$

$$\hat{h}_{22} = h_0 \quad (18.28)$$

$$\hat{h}_{23} = \frac{h_1 - \sqrt{h_1^2 - h_0 h_2}}{2} \quad (18.29)$$

$$\hat{h}_{24} = \frac{h_1 + \sqrt{h_1^2 - h_0 h_2}}{2} \quad (18.30)$$

To summarise, the 4 second order systems that have the same autocorrelation sequence (spectrally equivalent) are shown in table 18.4

	\hat{h}_0	\hat{h}_1	\hat{h}_2
1st Option	h_0	h_1	h_2
2nd Option	h_2	h_1	h_0
3rd Option	$\frac{h_1 - \sqrt{h_1^2 - h_0 h_2}}{2}$	$h_0 + h_2$	$\frac{h_1 + \sqrt{h_1^2 - h_0 h_2}}{2}$
4th Option	$\frac{h_1 + \sqrt{h_1^2 - h_0 h_2}}{2}$	$h_0 + h_2$	$\frac{h_1 - \sqrt{h_1^2 - h_0 h_2}}{2}$

Tabla 18.4 Spectrally equivalent systems of a second order system

18.5 Sensitivity of the solution to the estimation of the autocorrelation sequence

Analogously as was done for the identification of the first order system, the effect of a small variation in the autocorrelation sequence estimators on the filter coefficients can be computed using (18.17), (18.18) and (18.13). The procedure is quite similar and will not be presented here, only the results are displayed in table 18.5.

∂/∂	h_0	h_1	h_2
r_0	$h_1 \sim 0$ $h_1 \sim 2\sqrt{h_0 h_2}$ $h_2 \sim \pm h_0$ $(h_0 + h_2)^2 \sim h_1^2$	$(h_0 + h_2)^2 \sim h_1^2$	$h_0 \sim 0$ $h_2 \sim 0$ $h_1 \sim 0$ $h_1 = 2\sqrt{h_0 h_2}$ $h_2 \sim \pm h_0$ $(h_0 + h_2)^2 \sim h_1^2$
r_1	$h_1 \sim 0$ $h_1 \sim 2\sqrt{h_0 h_2}$ $h_2 \sim \pm h_0$ $(h_0 + h_2)^2 \sim h_1^2$	$(h_0 + h_2)^2 \sim h_1^2$	$h_0 \sim 0$ $h_2 \sim 0$ $h_1 \sim 0$ $h_1 = 2\sqrt{h_0 h_2}$ $h_2 \sim \pm h_0$ $(h_0 + h_2)^2 \sim h_1^2$
r_2	$h_1 \sim 0$ $h_1 \sim 2\sqrt{h_0 h_2}$ $h_2 \sim \pm h_0$ $(h_0 + h_2)^2 \sim h_1^2$	$(h_0 + h_2)^2 \sim h_1^2$	$h_0 \sim 0$ $h_2 \sim 0$ $h_1 \sim 0$ $h_1 = 2\sqrt{h_0 h_2}$ $h_2 \sim \pm h_0$ $(h_0 + h_2)^2 \sim h_1^2$

Tabla 18.5 Summary of the high sensitivity situations

18.6 Meaning of the sensitivity conditions

A system is completely characterised by the location of its zeros up to a scale factor. It can be very instructive to translate the previous conditions on the coefficients to conditions on the location of the zeros. To do that recall that if a and b are the zeros of the system, its coefficients are related to them in the following way:

$$h_0 = h_0 \quad (18.31)$$

$$h_1 = -h_0(a + b) \quad (18.32)$$

$$h_2 = h_0 ab \quad (18.33)$$

where h_0 makes the role of the normalisation factor.

18.6.1 Meaning for \hat{h}_1

Considering the previous relationships, the condition under which bad estimation of \hat{h}_1 is possible turns to:

$$(h_0 + h_2)^2 \sim h_1^2 \Rightarrow (1 + ab)^2 \sim (a + b)^2 \quad (18.34)$$

Taking square root two different conditions on the zero result

1.

$$1 + ab \sim a + b \quad (18.35)$$

which implies that $a \sim 1$ and b takes any value.

2. or considering that the right hand side of 18.34 has different sign than the left hand side (after taking square root):

$$1 + ab \sim -(a + b) \quad (18.36)$$

that takes to $b \sim -1$ with an arbitrary a .

So that this condition translated to the zero location implies that no one of the zeros should be close to ± 1 . It means that if the zeros are real they cannot be near the unit circle, but if they are complex this condition does not prevent them to be near the unit circle. In fact, solving (18.35) and (18.36) for a and b a complex conjugate pair gives no solution except the ones already mentioned for real values.

18.6.2 Meaning for \hat{h}_0

Recall that this estimate is computed using \hat{h}_1 and so the causes of divergence depend on whether the value of \hat{h}_1 is h_1 or its equivalent $h_0 + h_2$.

case 1) Estimated as h_1 : there are two possible situation in which divergence may occur

1. $h_1 \sim 0$. According to eq. (18.32) this means that the zeros are the opposite of each other $a \sim -b$.
2. $h_0 \sim h_2$. Again using eq. (18.32) this means that the zeros are the inverse of each other $a \sim 1/b$.

case 2) Estimated as $h_0 + h_2$: analogously there are two possible conditions under which divergence may appear:

1. $h_0 + h_2 \sim 0$, or equivalently $h_0 \sim -h_2$. Which imposes on the zeros that they are inverse and opposite of each other $a \sim -1/b$.
2. $h_1^2 \sim 4h_0h_2$ that translating it to the zeros gives $(a - b)^2 \sim 0$ or finally $a \sim b$.

The two subcases of case 1 together with the two subcases of case 2 give in total 4 cases which correspond with the four different values that the estimate h_0 can attain. For this very reason, not always that any of this condition holds bad estimates are unavoidable; under a certain condition a bad estimate is obtained if this condition is derive for the especial case that this estimate is attained. For example, if the real value for \hat{h}_1 is h_1 , none of the subcases of case 2 can affect the results. Furthermore, for this particular value of \hat{h}_1 , h_1 , two different values of h_0 can be obtained, each of them will be affected from only one of the subcases of case 1. In a sense, every path followed to build the exact solution will be affected by a certain set of conditions. This set will not affect the solution if a different path is to be followed. For path is meant a given combination of $[\hat{h}_0 \hat{h}_1 \hat{h}_2]$.

Apart from this possible causes of divergence, an additional cause has to be added due to the fact that \hat{h}_0 is estimated using the estimate of h_1 . If \hat{h}_1 is badly estimated, \hat{h}_0 will also be, so the condition for bad estimation of h_1 has to be taken into account here.

18.6.3 Meaning for \hat{h}_2

The intrinsic cause of divergence for \hat{h}_2 is that $\hat{h}_0 \sim 0$, which can be achieved if $h_0 \sim 0$ or $h_2 \sim 0$. But this situation leads to a first order system and has to be studied using the analysis for this kind of system. Again, \hat{h}_2 could be badly estimated if \hat{h}_0 is badly estimated, so conditions under which \hat{h}_0 is badly estimated must be considered here.

18.7 A new method. Identifying a first order channel

In previous sections explicit expressions relating the filter coefficients with the autocorrelation sequence of the output of that filter were derived, although no unique solution exists but a whole range covering all spectrally equivalent SE

systems of the given one. Usually in that set of SE systems a search is carried out to choose the one with the correct phase.

It would be useful to avoid this searching procedure in order to save operations and time. In this section a new method [AMC02] is proposed that achieves this goal. It will be illustrated using the 1st order channel. Recall that expressions for the first order channel coefficients as a function of the autocorrelation sequence are

$$\hat{h}_0^2 = \left[r_0 \pm \sqrt{r_0^2 - 4r_1^2} \right] / 2 \quad (18.37)$$

$$\hat{h}_1 = \frac{r_1}{\hat{h}_0} \quad (18.38)$$

If in (18.37) the autocorrelation sequence is substituted by its value as a function of the filter coefficients results

$$\hat{h}_0^2 = \left[h_0^2 + h_1^2 \pm \sqrt{(h_0^2 - h_1^2)^2} \right] / 2 \quad (18.39)$$

From it, it is easy to see that the two possible values for the first coefficient estimate are h_0 and h_1 , depending on the election of the sign made just before the square root. So a correct election of that sign can lead us directly to the desired solution h_0 . The election can be done in the following way:

a) If $h_0^2 > h_1^2$, expression (18.39) turns, after taking the square root, to

$$\hat{h}_0^2 = [h_0^2 + h_1^2 \pm (h_0^2 - h_1^2)] / 2$$

and two different situations arise

a.1) Choosing the + sign: $\hat{h}_0 = h_0$.

a.2) Choosing the - sign: $\hat{h}_0 = h_1$.

b) If $h_0^2 < h_1^2$, expression (18.39) now turns to

$$\hat{h}_0^2 = [h_0^2 + h_1^2 \pm (h_1^2 - h_0^2)] / 2$$

and the election of the sign affects in the following way

b.1) Choosing the + sign: $\hat{h}_0 = h_1$.

b.2) Choosing the - sign: $\hat{h}_0 = h_0$.

Considering all the situations described here, the desired solution can always be obtained if equation (18.37) is substituted by

$$\hat{h}_0^2 = \left[r_0 + s \sqrt{r_0^2 - 4r_1^2} \right] / 2$$

where $s = \text{sign}[h_0^2 - h_1^2]$. The problem is now the estimation of this sign. It is clear that knowing this sign is equivalent to knowing the phase of the channel: if $|h_0| > |h_1|$ the channel is minimum phase and is maximum phase if $|h_1| > |h_0|$. So it can not be estimated using second order statistics of the output because is phase blind. One possible option is the use of cumulants because they provide an easy way to do this task. First note that s can also be written as $s = \text{sign}[1 - \bar{h}_1^2]$ where $\bar{h}_1 \equiv h_1/h_0$ and then with the help of the Brillinger-Rossenblatt [BR67] expression it is straight forward to verify that

$$\bar{h}_1 = \frac{c_{4x}(0, 1, 1)}{c_{4x}(0, 0, 1)} \quad (18.40)$$

Some comments should be made about (18.40)

1. The fourth order cumulant has been chosen in order to make it valid even in the case of symmetric input signals (giving place to symmetric output signals).
2. This expression can be regarded as a particular case of the q -slice formulas for normalised coefficients.
3. Can be generalised for any order of cumulants and any order of filters.

Once \hat{h}_0 has been correctly estimated among the possible options, expression (18.38) gives the correct value for \hat{h}_1 .

Note that the information needed from HOS is just a sign, so the method is not expected to suffer the HOS based methods drawbacks: need for long data records, high variance, etc. Furthermore, the only condition that this method needs to be applied is that the input process be white, no additional information about its statistics is necessary, unlike traditional SEMP based methods where some high order moment (cumulant) of the input is needed.

18.8 A new method. Identifying a second order channel

The technique just described here can be generalised to any filter order as far as explicit expressions for the filter coefficients are available as a function of

the autocorrelation sequence. To illustrate this point let us apply this method to a second order system.

Recall that the solutions for the filter coefficients are

$$\hat{h}_1^2 = r_2 + r_0/2 \pm \sqrt{(r_2 + r_0/2)^2 - r_1^2} \quad (18.41)$$

$$\hat{h}_0 = \frac{r_1 \pm \sqrt{r_1^2 - 4\hat{h}_1^2 r_2}}{2\hat{h}_1} \quad (18.42)$$

$$\hat{h}_2 = \frac{r_2}{\hat{h}_0} \quad (18.43)$$

Again the election of the sign in (18.41) and (18.42) is crucial to get the true filter. To see which sign is going to be chosen let us write (18.41) as a function of the true filter coefficients:

$$\hat{h}_1^2 = \left[(h_0 + h_2)^2 + h_1^2 \pm \sqrt{(h_1^2 - (h_0 + h_2)^2)^2} \right] / 2 \quad (18.44)$$

Two situations can arise

a) If $h_1^2 > (h_0 + h_2)^2$ expression (18.44) can be written as

$$\hat{h}_1^2 = \frac{1}{2} [(h_0 + h_2)^2 + h_1^2 \pm (h_1^2 - (h_0 + h_2)^2)]$$

and depending on the sign elected two possibilities can happen

- a.1) If the + sign is chosen, the result is $\hat{h}_1 = h_1$.
- a.2) On the other hand, if - is chosen $\hat{h}_1 = h_0 + h_2$.

b) If $h_1^2 < (h_0 + h_2)^2$ expression (18.44) turns to

$$\hat{h}_1^2 = \frac{1}{2} [(h_0 + h_2)^2 + h_1^2 \pm ((h_0 + h_2)^2 - h_1^2)]$$

and the possibilities are

- b.1) If + is chosen $\hat{h}_1 = h_0 + h_2$.
- b.2) If - is chosen $\hat{h}_1 = h_1$.

Considering all these situations, the true value for the second coefficient of the filter can be obtained if the expression (18.41) is modified to

$$\hat{h}_1^2 = r_2 + r_0/2 + s_1 \sqrt{(r_2 + r_0/2)^2 - r_1^2}$$

where $s_1 = \text{sign}[h_1^2 - (h_0 + h_2)^2]$. From \hat{h}_1 , the next coefficient \hat{h}_0 can be computed using (18.42). For this purpose (18.42) is written as a function of the true filter coefficients, assuming that \hat{h}_1 has been correctly estimated so it is equal to h_1 ,

$$\hat{h}_0 = \frac{h_1(h_0 + h_2) \pm \sqrt{h_1^2(h_0 - h_2)^2}}{2h_1} \quad (18.45)$$

Following the known technique,

a) If $h_1(h_0 - h_2) > 0$ expression (18.45) can be written as

$$\hat{h}_0 = \frac{h_1(h_0 + h_2) \pm h_1(h_0 - h_2)}{2h_1}$$

a.1) Using the + sign, the estimate takes the value $\hat{h}_0 = h_0$.

a.2) With the - sign, $\hat{h}_0 = h_2$.

b) If $h_1(h_0 - h_2) < 0$ expression (18.45) becomes

$$\hat{h}_0 = \frac{h_1(h_0 + h_2) \pm h_1(h_2 - h_0)}{2h_1}$$

b.1) Choosing the + sign results $\hat{h}_0 = h_2$.

b.2) Choosing the - sign results $\hat{h}_0 = h_0$.

Considering all these possible situations, expression (18.42) can be modified to yield always the value of the true h_0 coefficient,

$$\hat{h}_0 = \frac{r_1 + s_2 \sqrt{r_1^2 - 4\hat{h}_1^2 r_2}}{2\hat{h}_1}$$

where $s_2 = \text{sign}((h_1(h_0 - h_2)))$. Finally, from \hat{h}_0 the last coefficient h_2 can be estimated using (18.43).

Still remains the problem of estimating s_1 and s_2 . One simple approach consists in rewriting both signs as functions of the normalised coefficients,

$$s_1 = \text{sign}[\bar{h}_1^2 - (1 + \bar{h}_0)^2] \quad (18.46)$$

$$s_2 = \text{sign}[\bar{h}_1(1 - \bar{h}_2)] \quad (18.47)$$

where the normalised coefficients are defined as $\bar{h}_i \equiv h_i/h_0$ and computing them using the following expressions:

$$\bar{h}_i = \frac{c_{4x}(0, i, 2)}{c_{4x}(0, 0, 2)}$$

That can be easily verified using the Brillinger-Rossenblatt equations.

18.9 Analysis of the phase selection methods

All the methods based on the spectrally equivalent set share the same three fundamental steps:

1. The autocorrelation sequence of the system output has to be computed.
2. From it, the coefficients of the systems of the spectrally equivalent set are obtained.
3. Among all the SE systems, the one with the correct phase is chosen.

In a previous section the influence of a small deviation from the true value of the autocorrelation on the filter estimates has been considered. To completely analyse the methods, points 1 and 3 also need considering (this analysis is contained in [AMGR02]).

18.9.1 On the estimation of the autocorrelation sequence

Let $\{x(n)\}$ be the output of a causal and stationary moving average MA system, excited by an independent and identically distributed random process $\{w(n)\}$,

$$x(n) = \sum_{i=0}^q h_i w(n-i) \quad (18.48)$$

Usually the output is corrupted by additive noise $\{v(n)\}$, modelled here as gaussian and independent of the input,

$$y(n) = x(n) + v(n) \quad (18.49)$$

The problem treated in this section is to study the factors that affect the estimation of the autocorrelation sequence of $\{x(n)\}$, assuming known the noise power and access to measurements of $\{y(n)\}$ alone. The autocorrelation sequence will

be estimated, as usual, as a time average: given N data samples, $y(1), \dots, y(N)$, the estimate of $r_{yy}(m) = E[y(n)y(n+m)]$ is

$$\hat{r}_{yy}(m) = \frac{1}{N-m} \sum_{i=1}^{N-m} y(i)y(i+m) \quad (18.50)$$

Using equations (18.48) and (18.49) the previous expression can be expressed as

$$\begin{aligned} \hat{r}_{yy}(m) = \frac{1}{N-m} \sum_{i=1}^{N-m} \left[\sum_{k,l=1}^q b_k b_l w(i-k)w(i+m-l) + \right. \\ \left. + \sum_{k=1}^q b_k w(i-k)v(i+m) + \right. \\ \left. + \sum_{l=1}^q b_l w(i+m-l)v(i) + v(i)v(i+m) \right] \end{aligned} \quad (18.51)$$

That can be rewritten as a function of autocorrelation estimates

$$\begin{aligned} \hat{r}_{yy}(m) = \sum_{k,l=1}^q b_k b_l \hat{r}_{ww}(m+k-l) + \\ + \sum_{k=1}^q b_k [\hat{r}_{wv}(m+k) + \hat{r}_{wv}(m-k)] + \hat{r}_{vv}(m) \end{aligned} \quad (18.52)$$

The autocorrelation estimates of the right hand side have the following expected value

$$E[\hat{r}_{ww}(m)] = \sigma_w^2 \delta(m) \quad (18.53a)$$

$$E[\hat{r}_{wv}(m)] = 0 \quad (18.53b)$$

$$E[\hat{r}_{vv}(m)] = \sigma_v^2 \delta(m) \quad (18.53c)$$

and because they are unbiased, these are the values we are looking for. Where σ_w^2 and σ_v^2 are the variances of $\{w(n)\}$ and $\{v(n)\}$ respectively. Using (18.53a)-(18.53c), the expected value of (18.52) turns to

$$r_{yy}(m) = \sigma_w^2 \sum_k^q b_k b_{k+m} + \sigma_v^2 \delta(m)$$

or equivalently

$$r_{yy}(m) = r_{xx}(m) + \sigma_v^2 \delta(m)$$

So, knowing the noise power, the autocorrelation sequence of $\{x(n)\}$ can be obtained from that of $\{y(n)\}$

$$r_{xx}(m) = r_{yy}(m) - \sigma_v^2 \delta(m) \quad (18.54)$$

If instead of the true value for the autocorrelation of $\{y(n)\}$ in equation (18.54), the estimate appearing in (18.51) is to be used, different causes can make it be far from its true value. These can be obtained inspecting (18.52):

1. If \hat{r}_{ww} is far from its expected value in (18.53a).
2. If \hat{r}_{wv} is much deviated from zero, its expected value in (18.53b).
3. If \hat{r}_{vv} is far from its expected value in (18.53c).

Assuming that the only free parameters are the filter coefficients and the SNR (i.e. the noise power), the previous deviations affect more the value of (18.52) if,

1. The bigger are the filter coefficients, the bigger the influence of the second summand in the right hand side of (18.52) is.
2. The smaller the SNR, the bigger is the deviation of \hat{r}_{wv} from zero.

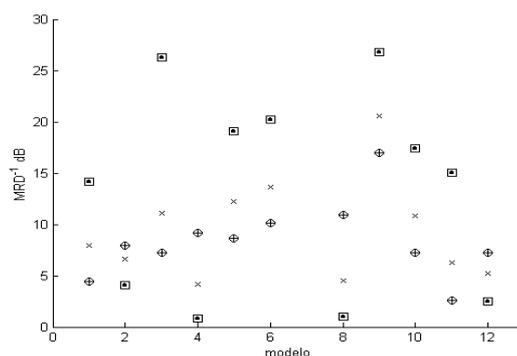


Fig. 18.1 *MDR for the studied models.* o SCuM_2; □ SCuM_1; ◇ SCuM_0; ● SCuM_1; + SCuM_2; x CuM

18.9.2 The minimum relative difference. A new magnitude

Once the SE set is computed, the system with the correct phase can be chosen based on a cumulant matching criteria. A given cumulant, say $c_{4y}(m, m, m) \equiv c_{4y}(m)$, is computed for every candidate and compared with the cumulant estimated from the system output. If $c_{4y}(m)$ is quite close for all the candidates

the selection process will be more difficult. This simple idea is contained in next lines.

Let $\Upsilon(m) = \{c_{4x}^{(1)}(m), \dots, c_{4x}^{(P)}(m)\}$ be the set of the P cumulants corresponding to the P candidate systems, in ascending order. Then the first element is subtracted from the second one and the difference divided by the former, for normalisation. In this way the first relative difference is computed. To compute the second relative difference the process is repeated for the second element of $\Upsilon(m)$. Once the $P - 1$ relative differences have been computed the minimum of this new set is called the Minimum Relative Difference MRD for this cumulant $c_{4y}(m)$ and is an indication of the difficulty of choosing the true candidate system based on a cumulant matching criteria. If all the diagonal cumulants are computed in the selection process, the MRD is computed as the averaged of $\text{MRD}(c_{4y}(m)) \forall m$.

The following proposition is not difficult to show,

Proposición 18.9.1 *The minimum relative difference (MRD) of a set of spectrally equivalent second order systems presents the following properties:*

1. $\text{MDR}(c_{4x}(m)) = \text{MDR}(c_{4x}(-m))$
2. $\text{MDR}(c_{4x}(0)) = 0$

18.9.3 Description of the studied methods

Basically two different kinds of methods can be applied to chose the system with the correct phase: equaliser based and not using the equaliser (or direct methods).

Methods based on the equaliser

All these methods compute an equaliser for every candidate and the output signal is then equalised. The correct system is chosen then according to different criteria:

Constant modulus CM If the input signal $\{w(n)\}$ has constant modulus (i.e. a binary signal), the equaliser giving an equalised signal with modulus closest to that constant value is chosen as the correct equaliser, and consequently, the candidate from which that equaliser was built is the correct system.

Closest kurtosis CK Now the correct system is selected when the equalised signal has a kurtosis closest to the kurtosis of the input process. Remember that the input process statistics is known.

Modified closest kurtosis MCK The previous criteria is reinforced here by imposing that the equalised signal must have the fourth order cumulant not lying in the main diagonal ($m \neq 0$) close to zero.

Direct methods

Cumulant matching CuM This method has been explained previously. It select the correct systems when the cumulants of a particular candidate match (in a least squares sense) the estimated output cumulants. In this chapter, all diagonal cumulants will be used ($m = -q, \dots, q$).

Simplified cumulant matching SCuM It is like the previous method but only one main diagonal cumulant is used. If diagonal cumulant $c_{4y}(m)$ is employed, the method will be called SCuM_ m .

18.9.4 Description of the models

To test the theoretical analysis developed in the previous sections, a set of simulations has been carried out in which 10 different second order models are identified. The input process $w(n)$ is a binary sequence of $\{+1, -1\}$ where each value has equal probability, fourth order stationary and white, with zero mean, $\sigma_w^2 = 1$, skewness $\gamma_{3w} = c_{3w}(0, 0) = 0$ and kurtosis $\gamma_{4w} = c_{4w}(0, 0, 0) = -2$.

From the 10 models studied in page 361 only 5 will be considered because they cover most of the most interesting properties to highlight here. The coefficients for these models, together with the locations of the zeros are displayed on table 18.6.

The properties of these models are:

Model 2 Is a minimum phase model, sensitive because it has opposite zeros.

Model 3 Is a minimum phase model, sensitive because one zero is near the unit circle.

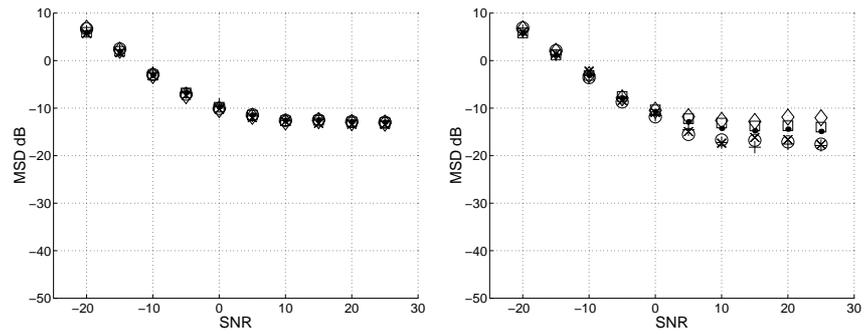
Model 4 Is a minimum phase model, sensitive because the zeros are the opposite of each other and are near the unit circle.

Model 8 Is a mixed phase model, sensitive because the zeros are the opposite and the inverse of each other.

Model 9 Is a mixed phase model, sensitive because both zeros are near the unit circle.

	h_0	h_1	h_2	z_1	z_2
model 2	1	0.05	-0.455	-0.7	0.65
model 3	1	1.5	0.54	-0.9	-0.6
model 4	1	-0.05	-0.765	0.9	-0.85
model 8	1	-0.5	-1.04	-0.8	1.3
model 9	1	2	0.99	-0.9	-1.1

Table 18.6 Coefficients of the studied models and location of their zeros



(a) Model 9. 1000 samples

(b) Model 4. 1000 samples

Fig. 18.2 Mean Square Deviation for models 9 and 4.

○ SCuM-2; □ SCuM-1; ◇ SCuM-0; ● SCuM-1; + SCuM-2; × CuM

18.9.5 Simulation results

For clarity reasons is convenient to divide the analysis of the methods in two groups. The first group would cover the cumulant matching and simplified cumulant matching criteria because is obvious that they are part of a same idea. The components of the second group will be unveiled later on.

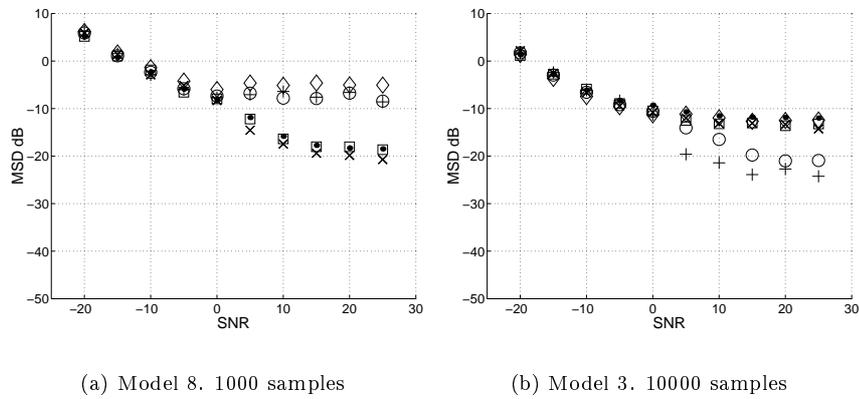


Fig. 18.3 Mean Square Deviation for model 8 and 3.
 ○ SCuM₋₂; □ SCuM₋₁; ◇ SCuM₀; ● SCuM₁; + SCuM₂; × CuM

First group. Methods CuM and SCuM_{*m*}

For this group the models 3, 4, 8 and 9 will be considered. The first thing to discuss is whether the autocorrelation sequence of the systems output can be easily estimated or not. Recall that, among other factors, the autocorrelation sequence ACS is less affected by estimation variance when the filter coefficients are small; models 4 and 8 have small coefficients so good ACS estimates are available, model 3 has coefficients just in the border between small and big and model 9 has big coefficients indeed. The adjective ‘small’ is applied on an empirical basis, a coefficient is said to be ‘big’ when it is bigger than 1.5.

From the estimated ACS the SE set is computed. For a sensitive model the SE set can be badly estimated if small variations in the estimated ACS with respect to the true ACS occur. All the models studied here are sensitive, so together with the comments made in the previous paragraph results that models 4 and 8 can have a correctly estimated SE set and model 9 has a badly estimated SE set. So, no method can identify correctly model 9 as can be checked in figure 18.2(a). The results for the rest of models are explained by the MRD. The inverse of the MRD for all the 10 models is displayed in figure 18.1.

When the inverse of the MRD($c_{4y}(m)$) is small (less than 10 dB), the method SCuM_{*m*} works properly and can identify the system. For example, for model 4 all the methods based on the cumulant matching idea have MRD⁻¹ below the 10 dB mark, so all these methods work properly, (with MSD around -20 dB) see figure 18.2(b). When some methods have the MRD⁻¹ below the 10 dB mark

and other above it, the formers will work properly and the later not; see model 8 for an example in figure 18.3(a). It is also worth noting that not always the method using more information gives the best results, in some situations using just 1 diagonal cumulant a better identification can be achieved than when using all slices (method CuM); see model 3 in figure 18.3(b) where methods SCuM₋₂ and SCuM₋₂ give better results than CuM.

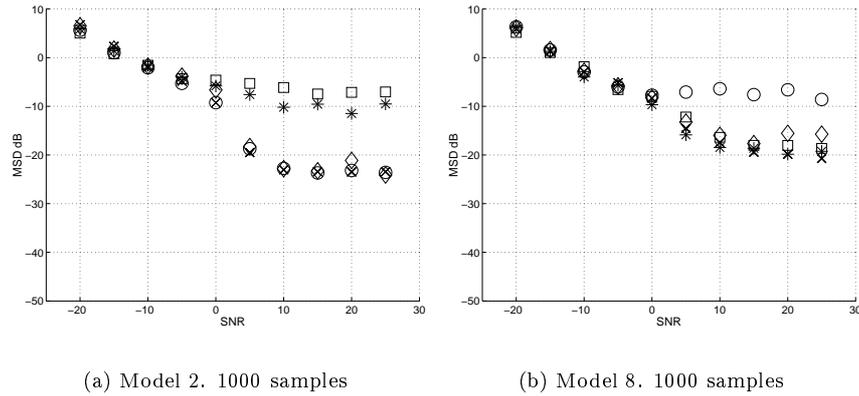


Fig. 18.4 Mean Square Deviation of the studied models.
 ○ RSCuM₋₂; □ RSCuM₋₁; ◇ EQU; * BISMH; x CuM

Second group. General comparison

It has been proven, theoretically and via simulations, that the methods SCuM_{-m} and SCuM_{4-m} give identical results; so in this general comparison only one of them will be displayed in the figures, and will be called RSCuM_{-m}. On the other hand, all the methods using an equaliser yield quite similar results, so the method CM will be chosen as the representative of all of them, a will be called, hereafter EQU. The CuM technique has been extensively used in many papers and will be considered here also. Finally, all these methods will be compared against the proposed new method for Blind Identification based on Second order statistics and Minimal Higher order statistics BISMH. Five methods in total.

The behaviour of the methods RSCuM₋₁, RSCuM₋₂ and CuM have been studied in the previous section. Now more attention will be paid to EQU and BISMH. For this second group of simulations only models 2, 8 and 9 will be discussed.

The technique BISMH basically shares the first two steps of the rest of methods: it has to estimate the ACS and then compute the true system directly using versions of the same equations used by the rest of methods to obtain the SE set. So aspects like the sensitivity of the models also affects BISMH. Apart from this, already commented, this new technique depend severely on the estimation of the signs s_1 and s_2 . For example, if $h_1 \sim 0$, $s_2 = \text{sign}[h_1(h_0 - h_2)]$ is close to zero and is more difficult to decide wheter it is positive or negative. Model 2 suffers from this inconvenient as it is shown on figure 18.4(a). But apart from this pathological example, the new method BISMH work as well as the best of the rest of methods; see model 8 in figure 18.4(b) for example.

The methods based on the equalizer, represented here by EQU, need the computation of the inverse filter of every candidate. This new step increments the choices of error accumulation, so sensitive models with big coefficients (thus, with badly estimated candidates) are more likely to give erroneus estimates of the true model. This is the case of model 9 in figure 18.5. In general, these techniques will depend on how accurate the inverse filter is. The inverse filters where computed as proposed in [ZMM93].

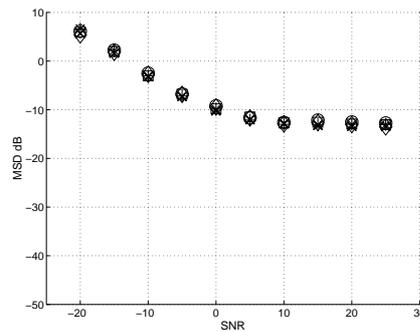


Fig. 18.5 Model 9 illustrates a case where EQU provides incorrect estimates.
 ○ RSCuM_2; □ RSCuM_1; ◇ EQU; * BISMH; x CuM

Conclusions

The main conclusions and results of this work will be presented grouped according to the chapter where they were presented.

Conclusions of chapter 16

- A complete convergence analysis of the ORIV algorithm has been presented in this chapter, covering stationary and non-stationary systems including estimation noise. Convergence in the mean and mean square have been proven for this algorithm.
- From this convergence analysis, a value for the optimum forgetting factor has been obtained as it was done for the RLS algorithm in [MB91].
- A second important product of this analysis is the derivation of the overdetermined and generalised orthogonality principle, which is the basis for chapter 17.
- A new general equation has been derived relating the coefficients of a linear system with the k th and $k + 1$ th order cumulants of the system output. This new equation has the Giannakis-Mendel equation and its generalisation to 3rd and 4th order cumulants as particular cases.
- These systems of equations have been solved by ORIV, forming methods ORIV-C2C3 and ORIV-C3C4. Both methods have been tested via simulations. ORIV-C3C4 gives better results if the $\text{SNR} < 10$ dB.

Conclusions of chapter 17

- The Overdetermined and Generalised Orthogonality Principle has been used to propose a new algorithm called the Averaged, Overdetermined and Generalised Least Mean Square AOGLMS algorithm and also to rederive the GLMS algorithm. This was achieved by the help of some intermediate algorithms like the OGLMS and OGLMS1.
- The novel AOGGLMS algorithm has been theoretically proved to have all the required properties: it is capable of solving overdetermined systems of equations involving higher-order statistics no matter whether the associated matrix is definite or not.
- The superior performance of AOGGLMS with respect to the existing GLMS algorithm is shown via simulations.

Conclusions of chapter 18

- A new method has been derived to blindly identify a linear system. It is based on the explicit expressions relating the second-order statistics of the system output and the system coefficients. It uses minimal higher-order statistics. So this method is called Blind Identification using Second-order statistics and Minimal Higher-order statistics BISMH. It benefits from all the advantages of 2nd and higher-order statistics and suffers from very few of its drawbacks.
- The techniques based on the computation of the Spectrally Equivalent Minimum Phase SEMP system to completely identify a linear system are analysed. In particular the Minimum Relative Difference MRD is defined to explain the observed behaviour of the cumulant matching techniques. Also other concepts like 'risk' or 'sensitive' are employed.
- Simulations show the excellent behaviour of BISMH when compared with the SEMP techniques. This behaviour is perfectly explained.

- It has been shown that not always the technique using more information yields better results, since the quality of this information is very important. This idea is hidden behind the MRD.

Conclusiones

Las principales aportaciones contenidas en esta Memoria y los principales resultados que en ella se han derivado se van a resumir en las próximas páginas agrupados atendiendo a la parte en la que fueron presentados. Asimismo se incluirán, allí donde sea procedente, las posibles líneas futuras de investigación que surgen de esta Memoria.

Conclusiones de la parte I

En esta parte de la Tesis Doctoral se ha llevado a cabo un análisis completo de los algoritmos adaptativos estándares de la familia del RLS: RLS, RIV y ORIV; así como una aplicación detallada a la resolución del problema de la identificación ciega de modelos MA, haciendo especial hincapié en las ventajas del empleo exclusivo de la estadística de alto orden. Algunas de las aportaciones más importantes que se han realizado son las siguientes:

- Se ha completado el análisis que sobre el algoritmo RIV se llevó a cabo en [Swa96]. En particular, como contribución nueva, se estudia la convergencia del MSD en situaciones estacionarias y no estacionarias bajo ruido de estimación. Esto permite obtener la expresión para el factor de olvido que minimiza el error cuadrático medio en convergencia, λ_{opt} , en función del ruido de estimación y del parámetro que controla la evolución temporal del sistema a identificar.
- Se ha realizado un análisis completo del algoritmo ORIV. Éste

incluye un análisis de la convergencia en media y en media cuadrática de los estimadores, para ambientes estacionarios y no estacionarios, tanto en presencia de ruido de estimación como sin él. Asimismo, y en analogía al estudio del algoritmo RIV, se ha derivado una expresión para λ_{opt} en función de las mismas variables. Los resultados son compatibles con estudios similares sobre el algoritmo RLS.

- Hay que destacar que dentro del análisis del algoritmo ORIV se ha obtenido una condición de ortogonalidad para situaciones sobredeterminadas que involucren estadística de alto orden. Este Principio de Ortogonalidad Sobredeterminado y Generalizado será la base de toda la segunda parte de la Memoria.
- Se ha discutido la idoneidad de las aproximaciones empleadas en el análisis, en particular se ha comprobado la equivalencia existente entre la técnica del promediado directo y la suposición de que el número de iteraciones ha alcanzado un valor elevado.
- Se ha obtenido en el dominio del tiempo una ecuación que relaciona los cumulantes de orden k y $k+1$ de la salida de un sistema MA con los coeficientes del mismo. Casos particulares importantes son la ecuación de Giannakis-Mendel ($k = 2$) y su generalización a estadística de alto orden exclusivamente ($k = 3$).
- Estas ecuaciones se han resuelto mediante el algoritmo ORIV para constituir los métodos ORIV-C2C3 y ORIV-C3C4 (ya existentes en la literatura). El comportamiento de ambos métodos se ha estudiado en una gran variedad de situaciones: ambientes contaminados por ruido y sin contaminar, tanto estacionarios como no estacionarios, donde la no estacionariedad venía dada por una evolución temporal lineal o aleatoria. Gracias al análisis teórico realizado se puede explicar satisfactoriamente el comportamiento observado, en particular a través de conceptos como longitud efectiva, tiempo de retardo, tiempo de convergencia, seguimiento límite, etc.
- En lo que respecta a la comparación ORIV-C2C3 con ORIV-C3C4 se ha concluido que el segundo proporciona un $\text{MSD}(\infty)$ menor que el primero en ambientes con $\text{SNR} < 10$ dB. En ausencia de ruido es conveniente emplear el método ORIV-C2C3. Además, ambos métodos identifican correctamente la fase del sistema.

- Se ha demostrado analíticamente la superioridad del comportamiento del algoritmo ORIV sobre el del algoritmo RIV. Dicha superioridad se ha corroborado mediante simulaciones. La comparación teórica se resume en la deducción del RIV a partir del ORIV, donde se aprecian claramente las diferencias entre uno y otro.

De todo el trabajo realizado aquí expuesto se concluye que se puede continuar fundamentalmente por una vía, siendo por tanto la principal línea futura de investigación:

- Reproducir numéricamente el comportamiento observado en las simulaciones para ORIV, quizás mediante el empleo de un modelo que permita un tratamiento matemático más sencillo.

Conclusiones de la parte II

Esta parte se centra en el estudio de los algoritmos adaptativos de la familia del LMS. El principal objetivo es encontrar los equivalentes, dentro de esta familia, a la familia del RLS. En particular los problemas que resuelve el RLS los puede resolver el LMS, y los que resuelva el RIV los puede resolver el GLMS generalmente. Sin embargo el ORIV no tiene equivalente en la familia de algoritmos de mínimo error cuadrático medio.

Las conclusiones a destacar de esta segunda parte son:

- Se ha empleado el Principio de Ortogonalidad Sobredeterminado y Generalizado en conjunción con la idea del gradiente estocástico para construir el algoritmo sobredeterminado y generalizado LMS, OGLMS.
- Tras comprobar su convergencia, el algoritmo OGLMS se toma como punto de partida para proponer los algoritmos OGLMS1 y OGLMS3. Además, del algoritmo OGLMS3 se reobtiene el algoritmo GLMS, ya presente en la literatura.

- De estos tres últimos algoritmos se analiza en primer lugar el GLMS. En su análisis se desprende que este algoritmo no puede converger cuando la matriz \mathbf{R} asociada al problema no sea definida. Además, dada su estructura este algoritmo no puede resolver problemas sobredeterminados. El análisis se ha llevado a cabo empleando 4 aproximaciones distintas: a orden μ , mediante promediado directo, factorización gaussiana y aproximación exacta.
- Se ha empleado el algoritmo GLMS para la identificación ciega de modelos AR complejos, excitados mediante constelaciones 4-QAM, empleando estadística de cuarto orden. Se han reproducido numéricamente los resultados de las simulaciones mediante las expresiones teóricas obtenidas anteriormente.
- Del análisis del comportamiento de OGLMS1 y OGLMS3 se desprende que ambos son capaces de trabajar con matrices \mathbf{R} no definidas y posiblemente sobredeterminadas, superando por tanto las limitaciones del GLMS. Pero dado que OGLMS1 presenta mejores propiedades que OGLMS3, se ha estudiado con más profundidad tanto su convergencia en media y en media cuadrática en ambientes estacionarios y no estacionarios.
- El mejor comportamiento de OGLMS1 con respecto a GLMS se debe a su mayor carga computacional por iteración. Esta desventaja se ha superado con la obtención del algoritmo promediado (averaged) y sobredeterminado (overdetermined) GLMS, AOGMS, como versión simplificada del OGLMS1.
- El análisis teórico de AOGMS se lleva a cabo en ambientes estacionarios. La principal novedad de este análisis es que tiene que tratar con una ecuación de actualización de los pesos que incluye un promedio temporal de los mismos; esto hace que el análisis se complique y sea necesario desarrollar nuevas técnicas, distintas de las clásicas empleadas en el análisis del LMS.
- El estudio mediante simulaciones de AOGMS corrobora la convergencia de este algoritmo con matrices \mathbf{R} no definidas y posiblemente sobredeterminadas y la imposibilidad de convergencia del algoritmo GLMS en estas circunstancias. Además se ha aplicado AOGMS a la identificación y ecualización ciega de canales AR alimentados por constelaciones de símbolos con distribuciones 4-QAM.

- La última contribución en esta parte ha sido la presentación de tres algoritmos nuevos dentro de la familia del LMS: CGLMS, especialmente ideado para trabajar con matrices \mathbf{R} no definidas, aunque presenta sesgo en presencia de ruido aditivo; el A1GLMS intenta solventar el problema del CGLMS basándose en la hipótesis de independencia; y el PNLMS, versión normalizada de GLMS. Tras un breve análisis teórico de ellos, las simulaciones corroboran el buen comportamiento de los mismos salvo para el A1GLMS, concluyendo por tanto que las hipótesis de independencia no son válidas para proponer un algoritmo basado en ellas (aunque sí para su análisis).

De las muchas líneas de investigación que quedan abiertas, los futuros esfuerzos investigadores deberían ir encaminados a,

- Dado que el algoritmo A1GLMS es el final del proceso de búsqueda y que reúne las características requeridas, sería conveniente, en primer lugar, llevar a cabo un análisis teórico completo para ambientes no estacionarios y contaminados por ruido. Un segundo paso requeriría intentar estabilizarlo en lo posible y acelerar su convergencia para que pueda competir con los algoritmos adaptativos basados exclusivamente en la estadística de segundo orden.

Conclusiones de la parte III

En esta última parte de la Memoria se han estudiado métodos de identificación ciega de sistemas cuya principal característica radica en que buena parte de su base se sustenta en la estadística de segundo orden. Estos métodos calculan el conjunto de sistemas espectralmente equivalentes del sistema problema, ya que la estadística de segundo orden no tiene acceso a la fase del mismo. En este conjunto se lleva a cabo la búsqueda del sistema con la fase correcta mediante el empleo de técnicas tales como el ajuste por cumulantes (cumulant matching) o imponiendo que la salida tenga módulo constante.

Con el fin de evitar este proceso de búsqueda se ha propuesto un nuevo método de Identificación Ciega con estadística de Segundo

orden y Mínima de Alto ICSMA.

Los puntos a destacar como aportaciones más importantes de esta parte de la Memoria son:

- Se han obtenido expresiones explícitas para los coeficientes de un sistema lineal en función de la secuencia de autocorrelación de su salida, para sistemas de orden 1 y 2. Se ha analizado la estructura de las soluciones y cuánto cambiarían al introducir un pequeño error en la autocorrelación. De esta manera se introduce el concepto de sistemas *sensibles*.
- A partir de las expresiones anteriores se deriva el método ICSMA. La mayor parte de su solución recae en estadística de segundo orden y la estadística de alto orden, portadora de la información de la fase, sólo se requiere para hacer una decisión binaria. Por tanto sus ventajas principales son: necesita menos datos y da estimadores menos sesgados que las técnicas que se basan en estadística de alto orden, lleva información sobre la fase al incluir sólo el uso justo de estadística de alto orden, elimina el proceso de búsqueda puesto que calcula directamente el sistema con la fase correcta, etc.
- Se han analizado mediante simulaciones 10 métodos para la identificación ciega de sistemas que comparten la filosofía del sistema SEMP (Spectrally Equivalent Minimum Phase), el ICSMA mas 9 métodos ya conocidos en la literatura pero cuyo comportamiento no se había justificado. El comportamiento observado de todos ellos se explica perfectamente mediante conceptos tales como la *sensibilidad* y el *riesgo* de los sistemas, que intervienen en la *calidad* de los estimadores de los coeficientes; la Mínima Diferencia Relativa MDR o la Diferencia Relativa Promedio DRP.
- Se ha comprobado que el nuevo método ICSMA funciona, al menos, tan bien como el mejor del resto de métodos estudiado en cualquier situación y para cualquier sistema.
- Se ha constatado que no siempre el método que emplea mayor información es el que proporciona los mejores estimadores, ya que la calidad de dicha información es crucial. Para los métodos basados en ajuste por cumulantes, esta calidad la determina la MDR: a mayor MDR mejores estimadores se obtendrán.

Las futuras líneas de trabajo dentro de esta última parte se pueden concentrar en dos puntos:

- ▶ Extender el método ICSMA a sistemas de orden mayor que dos. Esto pasa por encontrar expresiones explícitas para los coeficientes de los sistemas en función de la secuencia de autocorrelación de la salida o idear algún tipo de método recursivo tal y como se ha insinuado en esta Memoria.
- ▶ Llevar a cabo un análisis riguroso probabilístico de las cantidades MDR y DRP que aquí se han introducido de manera empírica.

Apéndices

Apéndice A

Lema de inversión matricial

Sean \mathbf{A} y \mathbf{B} dos matrices $m \times m$ definidas positivas relacionadas mediante la siguiente expresión:

$$\mathbf{A} = \mathbf{B}^{-1} + \mathbf{C}\mathbf{D}^{-1}\mathbf{E}^H \quad (\text{A.1})$$

donde \mathbf{D} es otra matriz definida positiva de dimensiones $n \times n$, \mathbf{C} es una matriz $m \times n$ y \mathbf{E} es una matriz $m \times n$. El lema de inversión matricial establece que la inversa de la matriz \mathbf{A} se puede expresar como

$$\mathbf{A}^{-1} = \mathbf{B} - \mathbf{B}\mathbf{C}(\mathbf{D} + \mathbf{E}^H\mathbf{B}\mathbf{C})^{-1}\mathbf{E}^H\mathbf{B} \quad (\text{A.2})$$

Esta afirmación se comprueba de manera sencilla sin más que llevar a cabo las operaciones pertinentes. Este lema también recibe el nombre de identidad de Woodbury.

Este lema es pieza fundamental en los capítulos 2, 3 y 4 para poder derivar los algoritmos RLS, RIV y ORIV respectivamente.

Apéndice B

Aproximaciones de la matriz $\Gamma(n)$

Si en la definición de $\Gamma(n)$ dada en (4.7) se emplea la aproximación para n grandes de (4.25) resulta:

$$(1 - \lambda)^2 \Gamma(n) = \mathbf{R}^t \mathbf{R} + (1 - \lambda) \mathbf{R}^t \mathbf{R}^e(n) + (1 - \lambda) \mathbf{R}^{et}(n) \mathbf{R} + (1 - \lambda)^2 \mathbf{R}^{et}(n) \mathbf{R}^e(n) \quad (\text{B.1})$$

considerando a la matriz $\Phi(n)$ como cuasideterminista (véase apéndice C) y a λ cercana a la unidad, la inversa se puede aproximar por

$$\Gamma^{-1}(n) \cong (1 - \lambda)^2 [\mathbf{I} - (1 - \lambda)(\mathbf{R}^t \mathbf{R})^{-1} \mathbf{R}^t \mathbf{R}^e(n) - (1 - \lambda)(\mathbf{R}^t \mathbf{R})^{-1} \mathbf{R}^{et}(n) \mathbf{R}] (\mathbf{R}^t \mathbf{R})^{-1} \quad (\text{B.2})$$

Esta aproximación contiene toda la información de primer orden (el orden más bajo en $(1 - \lambda)$ e información acerca de $\mathbf{R}^e(n)$) y será suficientemente buena para todos los cálculos que se lleven a cabo en el capítulo 4. Es importante recordar que la componente estocástica $\mathbf{R}^e(n)$ está bien modelada como de media zero e independiente del resto de variables aleatorias involucradas.

La aproximación más drástica para esta matriz se puede conseguir omitiendo los términos en el ruido de estimación:

$$\Gamma^{-1}(n) = (1 - \lambda)^2 (\mathbf{R}^t \mathbf{R})^{-1} \quad (\text{B.3})$$

cuya validez aumenta al aumentar n .

Apéndice C

Las hipótesis de independencia

En el análisis de ORIV llevado a cabo en el capítulo 4 se han utilizado principalmente dos recursos para hacer la matemática más tratable.

El primero consiste en suponer que las cantidades definidas como producto escalar $f(n) = \langle \mathbf{x}(n), \mathbf{y}(n) \rangle = \sum_{i=1}^n \lambda^{n-i} \mathbf{x}(i) \mathbf{y}(i)$, cuando n es suficientemente grande, son cuasideterminista: su componente aleatoria es mucho menor que su componente determinista. Es decir,

$$f(n) \underset{n \rightarrow \infty}{\sim} \mathbf{E}[f(n)] + f^e(n) \quad (\text{C.1})$$

con $\|f^e(n)\| \ll \|\mathbf{E}[f(n)]\|$, siendo $f^e(n)$ un proceso aleatorio de media nula e independiente del resto de procesos aleatorios [EF86]. Esta hipótesis (HC) se puede aceptar intuitivamente supuestas las series ergódicas. Además se ha demostrado para el caso particular de un seno chirped con ruido (ver [MB91]).

El segundo recurso está compuesto por el conjunto de hipótesis de independencia. Alguna de las cuales tiene una justificación más difícil, o simplemente no la tiene. Dichas hipótesis son las siguientes:

HI1 $\Phi(n)$ y $z(n)$ son independientes entre sí y del resto de variables.

Esto está relacionado con la suposición de que son cuasideterministas. Es más cierto conforme mayor es n .

HI2 $\mathbf{w}(n-1)$ es independiente de $\mathbf{x}(n)$ y de $\tilde{\mathbf{x}}(n)$. O lo que es lo mismo, teniendo en cuenta HI1 y las definiciones de (4.9), (4.10) y (4.39), $\mathbf{w}(n-1)$ es independiente de $\mathbf{X}(n) \mathbf{\Lambda}^{-1}(n) \boldsymbol{\alpha}_0(n)$.

Esta hipótesis es más cierta conforme aumenta n , porque en ese caso se puede utilizar (4.28) como aproximación de $\mathbf{w}(n-1)$ y se demuestra HI2 a la vista de la hipótesis de cuasideterminación.

Por tanto, las HI se pueden considerar, según el uso dado aquí¹, como corolario de la HC. Sin embargo, para n pequeños, las HI no tienen justificación porque al fin y al cabo, todas las cantidades definidas en ORIV dependen de $\{x(p), \dots, x(n), \tilde{x}(p), \dots, \tilde{x}(n)\}$ y éstas sí son dependientes entre sí normalmente.

De cualquier manera, los distintos análisis de ORIV hechos en este capítulo, que necesiten de HI, se pueden modificar para que sólo involucren $n > n_0$ con n_0 suficientemente grande (como en (4.50)), pudiendo ser admitidos desde el punto de vista lógico sin el mayor reparo.

Todavía sería posible hacer un comentario final en este apéndice, ya que se podría considerar como una hipótesis de independencia adicional. Recuerdese que se ha definido $e_0(n) = d(n) - \mathbf{w}_0^t \mathbf{x}(n)$, (ver comentario después de (4.29)), al que se ha llamado error de medición. Por definición de \mathbf{w}_0 , se tiene que $E[e_0(n) \tilde{\mathbf{x}}(n)] = 0$, es decir, $e_0(n)$ es independiente de $\tilde{\mathbf{x}}(n)$. Además es de media nula, ya que lo son $d(n)$ y $x(n)$. Por tanto, lo único que no tiene justificación plausible es que sea independiente de $x(n)$, como se supone a lo largo del capítulo 4. Y si es independiente de $x(n)$ y de $\tilde{x}(n)$ es independiente de todo lo demás, porque se definen en función de ellos.

Este problema, tal y como se ha planteado aquí, se corresponde con un problema clásico² de estimación de la serie $d(n)$ (respuesta deseada) en función de combinaciones lineales de $x(n)$ por medio de $\mathbf{w}(n)$. Sin embargo, si $d(n) = \mathbf{w}_0^t \mathbf{x}(n) + e_0(n)$, quiere decir que se podrá estimar salvo por $e_0(n)$, es como si se tuviera un error de medición, lo que justifica el nombre que se le ha dado.

En la situación más general posible, como es la que se tiene en la obtención de resultados, se basa en la analogía con el problema anterior para calcular \mathbf{w}_0 , pero no sería adecuado llamar a $d(n)$ respuesta deseada ni a $e_0(n)$ error de medición. Véanse por ejemplo las definiciones (5.31) o (5.37).

¹El uso genérico de HI está muy extendido en la literatura para el análisis de algoritmos adaptativos en general, aunque se sepa que es una aproximación errónea y poco realista en su acepción más radical: ' $\{x(p), \dots, x(n), \tilde{x}(p), \dots, \tilde{x}(n)\}$ son independientes' ¿Cómo van a ser independientes si ellas son las que llevan información sobre el sistema? [Wid76], [Maz79].

²Algo de esto se discute en el capítulo 2 de [WS85].

Apéndice D

Resolución de ecuaciones en diferencias de primer orden

En muchos análisis llevados a cabo en esta memoria se requiere la resolución de ecuaciones en diferencias de primer orden. Siempre se ha intentado mediante diversas técnicas que el coeficiente de la homogénea sea constante para facilitar los cálculos subsiguientes, sin embargo el coeficiente de la no homogénea puede ser constante en el tiempo o no, dependiendo principalmente de si el proceso subyacente tras la ecuación en diferencias es estacionario o no. Recuérdese por ejemplo el análisis de los algoritmos RIV y ORIV cuando trataban con sistemas con coeficientes constantes o cuando estos coeficientes evolucionan con el tiempo.

Es necesario por tanto dividir este apéndice en dos partes. La primera trata sobre las ecuaciones en diferencias con el coeficiente de la no homogénea constante y la segunda lo supone como función del índice temporal.

D.1 Coeficientes constantes

En este primer caso la ecuación tiene la forma:

$$m(n) = cm(n-1) + a \tag{D.1}$$

junto con la condición inicial de que $m(n) = 0 \ \forall n < 0$ y valor inicial para $m(0)$ conocido. La solución es la suma de dos contribuciones: la solución general de la homogénea y la particular de la no homogénea.

La ecuación homogénea asociada es:

$$m(n) = cm(n-1) \quad (\text{D.2})$$

cuya solución en función del valor inicial $m(0)$ se obtiene fácilmente si se sustituye esta expresión n veces en sí misma, llegando a $m(n) = c^n m(0)$. Notar que para que esta solución no diverja para tiempos grandes es necesario que $c < 1$, además esto contribuye a que el efecto de las condiciones iniciales vaya desapareciendo con el tiempo.

De igual manera la solución general se puede encontrar si se sustituye la expresión (D.1) n veces en sí misma:

$$m(n) = c^n m(0) + a \sum_{i=0}^{n-1} c^i \quad (\text{D.3})$$

a partir de la cual es sencillo encontrar el valor para $n \rightarrow \infty$ (valor en convergencia), teniendo en cuenta que $c < 1$:

$$m(\infty) = \frac{1}{1-c} a \quad (\text{D.4})$$

Despejando aquí a , sustituyendo su valor en (D.3) y realizando la sumatoria indicada, tras operar brevemente, se llega a una expresión para la solución general en función de la condición inicial, el valor en convergencia y el coeficiente c :

$$m(n) = m(\infty) + c^n [m(0) - m(\infty)] \quad (\text{D.5})$$

Es decir, estas tres magnitudes son suficientes para caracterizar exactamente la solución a una ecuación en diferencias de primer orden con coeficientes constantes.

D.1.1 Tiempos de convergencia

Es interesante, además de contar con expresiones explícitas de la solución en función del tiempo y del valor que alcanzan en convergencia, caracterizar la evolución de dicha solución mediante alguna clase de constante de tiempo.

La solución más general posible se construye a partir de la solución general de la homogénea más una particular de la no homogénea. Dependiendo del tipo de información que se quiera obtener se pueden definir dos tiempos de convergencia distintos:

Tiempo de convergencia asociado a la solución de la homogénea

Como la solución general de la ecuación en diferencias homogénea asociada tiende a cero, bajo condiciones concretas, a veces es necesario conocer el tiempo

que ha de transcurrir para aceptar que ha llegado a un valor tan bajo que se puede despreciar y por tanto la fase estacionaria está empezando. Este tiempo de convergencia se definirá como el número de iteraciones necesarias para que la solución general de la homogénea tienda a $1/e$ del valor inicial. Dado que esta solución tiene la forma

$$m(n) = c^n m(0)$$

resulta que el tiempo de convergencia asociado es

$$\tau_h = \frac{-1}{\ln c}$$

pero dado que $c \sim 1$ en muchas de las ocasiones en las que aparece una ecuación en diferencias a lo largo de esta memoria, la expresión anterior se puede simplificar dando

$$\tau_h = \frac{1}{1 - c}$$

Tiempo de convergencia asociado a la solución general

En otras circunstancias interesa más analizar el número de iteraciones necesarias para que la solución converja pero desde un punto de vista más general: se tienen en cuenta todos los sumandos que componen la solución, tanto la solución de la homogénea como la de la particular, es decir, se supone que

$$m(n) = m(\infty) + c^n [m(0) - m(\infty)]$$

y por tanto el tiempo de convergencia asociado vale:

$$\tau_g = \frac{\ln \frac{m(0)/e - m(\infty)}{m(0) - m(\infty)}}{\ln c} \quad (\text{D.6})$$

Hay ciertas ocasiones en las cuales $m(\infty) \sim 0$ y $c \sim 1$, por ejemplo cuando el tamaño del paso es pequeño en el algoritmo GLMS. En este caso es fácil comprobar que

$$\tau_h = \tau_g = \frac{1}{1 - c}$$

D.2 Coeficientes no constantes. Caso 1

En el caso de que el coeficiente de la parte no homogénea no sea constante la ecuación a resolver tiene la forma:

$$m(n) = cm(n-1) + a(n) \quad (\text{D.7})$$

La solución más general posible se puede obtener directamente si se sustituye esta expresión n veces en sí misma y se deduce una ley general de la forma:

$$m(n) = c^n m(0) + \sum_{i=0}^{n-1} c^i a(n-i) \quad (\text{D.8})$$

donde de nuevo se reconocen la solución general de la homogénea asociada, que tiende a cero siempre y cuando $c < 1$, y una solución particular de la no homogénea dependiente en este caso de la forma funcional de $a(n)$. Para esta situación el valor en convergencia es:

$$m(\infty) = \sum_{i=0}^{\infty} c^i a(n-i) \quad (\text{D.9})$$

Como se puede prever a partir de la ecuación (D.7), la dinámica de $m(n)$ es muy dependiente de la evolución de $a(n)$ con lo que una expresión de la sencillez de (D.5) no parece suficiente, tal y como se comprueba si en ella se sustituye (D.9) y se intenta recuperar (D.8).

D.3 Coeficientes no constantes. Caso 2

Para trabajar en el caso más general posible se va a suponer que tanto el coeficiente como el término independiente dependen del tiempo. Se quiere resolver, por tanto, una ecuación de la forma:

$$a(n) = [1 - c(n)]a(n-1) + c(n)a_0(n-1)$$

donde se ha supuesto cierta relación entre el coeficiente y el término independiente. Suponiendo que el término responsable de la excitación evoluciona según la expresión siguiente

$$a_0(n) = a_0(n-1) + \Delta a_0$$

la ecuación anterior se puede poner de forma equivalente:

$$\Delta a(n) = [1 - c(n)]\Delta a(n-1) - \Delta a_0$$

donde $\Delta a(n) = a(n) - a_0(n)$. A continuación se va a obtener una expresión explícita para $\Delta a(n)$, para ello se muestra su valor para los primeros valores de n :

$$\begin{aligned} \Delta a(1) &= [1 - c(1)]\Delta a(0) - \Delta a_0 \\ \Delta a(2) &= [1 - c(2)][1 - c(1)]\Delta a(0) - [1 - c(2)]\Delta a_0 - \Delta a_0 \\ \Delta a(3) &= [1 - c(3)][1 - c(2)][1 - c(1)]\Delta a(0) - [1 - c(3)][1 - c(2)]\Delta a_0 - \\ &\quad - [1 - c(3)]\Delta a_0 - \Delta a_0 \end{aligned} \quad (\text{D.10})$$

a partir de ellos se puede inducir su ley general:

$$\Delta a(n) = \left(\prod_{i=1}^n [1 - c(i)] \right) \Delta a(0) - \left(1 + \sum_{i=1}^{n-1} \prod_{k=i+1}^n [1 - c(k)] \right) \Delta a_0$$

que está compuesta por dos sumandos. El primero depende de las condiciones iniciales y el segundo del término que controla la excitación. En todos los análisis realizados es deseable que el efecto de las condiciones iniciales vaya desapareciendo por el tiempo, así se podrá converger a la solución verdadera independientemente de la información con la que se cuente al inicializar el algoritmo. Para que esto se cumpla es suficiente que

$$|1 - c(l)| < 1 \quad \forall l \in \mathcal{N}$$

Con lo cual, para n suficientemente grandes el primer término se hace muy pequeño (va tendiendo a cero) con lo que se puede escribir,

$$a(n) = a_0(n) - \left(1 + \sum_{i=1}^{n-1} \prod_{k=i+1}^n [1 - c(k)] \right) \Delta a_0$$

Esto significa que el segundo sumando representa el retraso con que la serie $a(n)$ rastrea o sigue la evolución de $a_0(n)$, es decir:

$$a(n) = a_0(n - \tau)$$

$$\tau = \lim_{n \rightarrow \infty} \left\{ 1 + \sum_{i=1}^{n-1} \prod_{k=i+1}^n [1 - c(k)] \right\}$$

Sería conveniente encontrar una expresión explícita para el retraso τ . Esto no es posible porque a lo largo del texto no se conoce la dependencia de los coeficientes c con el tiempo.

Apéndice E

Obtención del algoritmo RIV a partir del ORIV

En el capítulo 6 se comprobó que las expresiones que describen el comportamiento de RIV se pueden obtener a partir de aquellas que describen el comportamiento de ORIV. En este apéndice se quiere dar un paso más y mostrar que directamente del algoritmo ORIV se puede obtener el RIV realizando las simplificaciones necesarias.

Para comenzar se recordará la expresión de actualización de las incógnitas para ORIV dada en la ecuación (4.61):

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \mathbf{\Gamma}^{-1}(n)\mathbf{X}(n)\mathbf{\Lambda}^{-1}(n)\boldsymbol{\alpha}(n) \quad (\text{E.1})$$

y se estudiará cada matriz y vector por separado, teniendo en cuenta que RIV resuelve exactamente la ecuación $\mathbf{\Phi}(n)\mathbf{w}(n) = \mathbf{z}(n)$ y que ORIV lo hace por mínimos cuadrados. Por tanto, para el primero la diferencia $\mathbf{\Phi}(n)\mathbf{w}(n) - \mathbf{z}(n)$ es idénticamente cero y para ORIV no es más que otra fuente de error.

E.1 Análisis del producto $\mathbf{\Gamma}^{-1}(n)\mathbf{X}(n)\mathbf{\Lambda}^{-1}(n)$

Teniendo en cuenta las definiciones de estos vectores y matrices dados en (4.7), (4.9) y (4.10) y que para reducir ORIV a RIV hay que suponer que $\mathbf{\Phi}(n)$

es cuadrada, se puede escribir que:

$$\begin{aligned}
\mathbf{\Gamma}^{-1}(n)\mathbf{X}(n)\mathbf{\Lambda}^{-1}(n) &= \\
&= \mathbf{\Phi}^{-1}(n)\mathbf{\Phi}^{-t}(n)[\mathbf{\Phi}^t(n-1)\tilde{\mathbf{x}}(n) \quad \mathbf{x}(n)] \begin{pmatrix} 0 & \lambda \\ \lambda & \tilde{\mathbf{x}}^t(n)\tilde{\mathbf{x}}(n) \end{pmatrix} = \\
&= \mathbf{\Phi}^{-1}(n)\mathbf{\Phi}^{-t}(n)[\lambda\mathbf{x}(n) \quad \mathbf{\Phi}^t(n)\tilde{\mathbf{x}}(n)] = \\
&= [\lambda\mathbf{\Phi}^{-1}(n)\mathbf{\Phi}^{-t}(n)\mathbf{x}(n) \quad \mathbf{\Phi}^{-1}(n)\tilde{\mathbf{x}}(n)]
\end{aligned} \tag{E.2}$$

E.2 Análisis de $\boldsymbol{\alpha}(n)$

El error a priori de ORIV $\boldsymbol{\alpha}(n)$ está definido de la siguiente manera:

$$\begin{aligned}
\boldsymbol{\alpha}(n) &= \mathbf{v}(n) - \mathbf{X}^t(n)\mathbf{w}(n-1) = \\
&= \begin{pmatrix} \tilde{\mathbf{x}}^t(n)[z(n-1) - \mathbf{\Phi}(n-1)\mathbf{w}(n-1)] \\ d(n) - \mathbf{x}^t(n)\mathbf{w}(n-1) \end{pmatrix}
\end{aligned} \tag{E.3}$$

Recordando el comentario hecho con anterioridad se deduce inmediatamente que la primera componente es nula por definición, con lo que el vector de error a priori queda:

$$\boldsymbol{\alpha}(n) = \begin{pmatrix} 0 \\ d(n) - \mathbf{x}^t(n)\mathbf{w}(n-1) \end{pmatrix} \tag{E.4}$$

que como se ve elimina una fuente de error y por tanto de información.

E.3 Recuperación del algoritmo RIV

Si en la expresión (E.1) se tienen en cuenta las expresiones (E.2) y (E.4) se llega a:

$$\mathbf{w}(n) = \mathbf{w}(n-1) + \mathbf{\Phi}^{-1}(n)\tilde{\mathbf{x}}(n)e(n) \tag{E.5}$$

que coincide con la expresión de actualización de las incógnitas mediante RIV, ecuación (3.29), tal y como se quería obtener.

Básicamente se han hecho dos modificaciones:

1. Considerar la matriz $\mathbf{\Phi}(n)$ cuadrada.
2. Eliminar el error debido a la resolución mediante mínimos cuadrados, es decir, suponer que $\mathbf{\Phi}(n)\mathbf{w}(n) - z(n) = 0$.

Bibliografía

- [AAM96] D. Aboutajdine, A. Adib, and A. Meziane. Fast adaptive algorithms for AR parameters estimation using higher-order statistics. *IEEE transactions on signal processing*, 44(8), agosto 1996.
- [AB01] W. Ang and F. Boroujeny. A new class of gradient adaptive step-size LMS algorithm. *IEEE transactions on signal processing*, 49(4), abril 2001.
- [AJS99] M. Turki-Hadaj Alouane and M. Jaidane-Saidane. A non stationary RLS algorithm for adaptive tracking of markov time varying channel. In *Proc. ICASSP 1999*, pages 1269–1272, 1999.
- [AMC02] Enrique Alameda-Hernandez, Desmond C. McLernon, and Maria C. Carrion. A new method for blind identification of FIR channels based almost exclusively on second order statistics. In *36th Asilomar Conference on Signals, Systems and Computers*, Asilomar Conference Grounds. CA. USA, 2002.
- [AMGR02] Enrique Alameda-Hernandez, D. C. McLernon, Mounir Ghogho, and Diego P. Ruiz. On the applications of second-order statistics to blind FIR system identification. In *2nd IEEE International Symposium on Signal Processing and Information Technology ISSPIT*, Marrakech. Morocco, 2002.
- [ARBC02a] Enrique Alameda, Diego P. Ruiz, David Blanco, and M. Carmen Carrion. Blind identification of non-stationary MA systems. *Electronics Letters*, 38(19):1145–1147, 2002.
- [ARBC02b] Enrique Alameda-Hernandez, Diego P. Ruiz, David Blanco, and Maria C. Carrion. A new method for blind and adaptive identification of non-gaussian FIR channels. In *2nd IEEE International Symposium on Signal Processing and Information Technology ISSPIT*, Marrakech. Morocco, 2002.
- [ARBC02c] Enrique Alameda Hernandez, Diego P. Ruiz Padillo, David Blanco Navarro, and Maria C. Carrion Perez. Análisis del algoritmo

- de mínimo error cuadrático medio generalizado (GLMS) para la identificación ciega y adaptativa de canales. In *XVII Simposium Nacional de la Unión Científica Internacional de Radio URSI*, Alcalá de Henares. Madrid. España., 2002.
- [BBCC88] S. Bittanti, P. Bolzern, M. Campi, and E. Colleti. Deterministic convergence analysis of RLS estimators with different forgetting factors. In *Proceedings 27th Conference on Decision and Control*, pages 1530–1531, Austin. TX. USA, 1988.
- [Ber86] Neil J. Bershad. Analysis of the normalized LMS algorithm with gaussian inputs. *IEEE transactions on acoustics, speech and signal processing*, ASSP-34(4):793–806, 1986.
- [BHT63] B.P. Bogert, M.J.R. Healy, and J.W. Tukey. The frequency analysis of time series echoes: Cepstrum, pseudo-auto covariances, cross-cepstrum, and saphe cracking. In M. Rosenblatt, editor, *Proc. Symposium Time Series Analysis*, pages 209–243, New York, 1963. J.Wiley.
- [Bit83] Robert R. Bitmead. Convergence in distribution of LMS-type adaptive parameter estimates. *IEEE transactions on automatic control*, AC-28(1):54–60, 1983.
- [BKS93] K. M. Bucklew, T. Kurtz, and W. A. Sethares. Weak convergence and local stability properties of fixed stepsize recursive algorithms. *IEEE Trans. information theory*, 39:966–978, 1993.
- [BL95] Mouad Boumahdi and Jean-Louis Lacoume. Blind identification using the kurtoisis: Results of field data processing. In *International Conference on Acoustics, Speech and Signal Processing*, pages 1980–1983, 1995.
- [BM89] Jean-Luc Botto and George V. Moustakides. Stabilizing the fast kalman algorithms. *IEEE transactions o acoustics, speech and signal processing*, 37(9):1342–1348, 1989.
- [BM91] Neil J. Bershad and Odile M. Macchi. Adaptive recovery of a chirp sinusoid in noise, part 2: Performance of the LMS algorithm. *IEEE transactions on signal processing*, 39:595–602, 1991.
- [BR67] D.R. Brillinger and M. Rosenblatt. Computation and interpretation of the k-order spectra. *Spectral analysis of time series*, pages 153–188, 1967. Wiley.
- [BRAC00] David Blanco, Diego P. Ruiz, Enrique Alameda, and Maria C. Carrion. Parametric estimation of autorregresive linear models contaminated with gaussian noise. In *XV Symposium Nacional de la Unión Científica Internacional de Radio URSI*, Zaragoza. España., 2000.

- [BRAC01] David Blanco, Diego P. Ruiz, Enrique Alameda, and Maria C. Carrion. Radar target identification scheme using extinction pulses based on cross correlations. In *XVI Symposium Nacional de la Unión Científica Internacional de Radio URSI*, Villaviciosa de Odón. Madrid, 2001.
- [BRAC02a] David Blanco, Diego P. Ruiz, Enrique Alameda, and Maria C. Carrion. Radar target discrimination using discriminating waves in arbitrarily noisy environments. In *Mediterranean Microwave Symposium*, pages 217–220, Cáceres. España., 2002.
- [BRAC02b] David Blanco Navarro, Diego P. Ruiz Padillo, Enrique Alameda Hernandez, and Maria C. Carrion Perez. Desarrollo de pulsos de extinción mediante β -splines. nuevas condiciones e-pulso. In *XVII Symposium Internacional de la Unión Científica Internacional de Radio URSI*, pages 247–248, Alcalá de Henares. Madrid. España, 2002.
- [Bri77] D.R. Brillinger. The identification of a particular nonlinear time series system. *Biometrika*, 64:509–515, 1977.
- [Bru01] E.N. Bruce. *Biomedical Signal Processing and Signal Modeling*. Wiley, 2001.
- [BSN00] V. Buzenac-Settineri and M.Ñajim. OLRIV: A new fast adaptive algorithm for rectangular-block toeplitz systems. *IEEE transactions on signal processing*, 48(9):2519–2534, septiembre 2000.
- [CC99] Sergio Cruces and Luis Castedo. Stability analysis of adaptive algorithms for blind source separation of convolutive mixtures. *Signal processing*, 78:265–275, 1999.
- [Chi78] D.G. Childers. *Modern Spectrum Analysis*. IEEE press, New York, 1978.
- [CL93] Jian de Z. Chen and Zhiyue Lin. Adaptive cancellation of the respiratory artifact in surface recording of small intestinal electrical activity. *Compt. Biol. Med.*, 23(6):497–509, 1993.
- [CL99] S. Chen and B.L. Luk. Adaptive simulated annealing for optimization in signal processing applications. *Signal processing*, 79:117–128, 1999.
- [CN90] Hsing-Hsing Chiang and Chrysostomos L. Nikias. Adaptive deconvolution and identification of nonminimum phase FIR systems based on cumulants. *IEEE transactions on automatic control*, 35(1):36–47, 1990.

- [CRGM95] M.C. Carrion, D.P. Ruiz, A. Gallego, and J.A. Morente. FIR system identification using third- and fourth-order cumulants. *Electronics letters*, 31(8):612–614, 1995.
- [Cru99] Sergio A. Cruces Alvarez. *Una Visión Unificada de Los Algoritmos de Separación Ciega de Fuentes*. PhD thesis, Universidad de La Coruña, 1999.
- [CWM97] S. Chen, Y. Wu, and S. McLaughlin. Genetic algorithm optimization for blind channel identification with higher order cumulant fitting. *IEEE transactions on evolutionary computation*, 1(4):249–265, 1997.
- [DL01] Zhi Ding and Jing Liang. A cumulant matrix subspace algorithm for blind single fir channel identification. *IEEE transactions on signal processing*, 49(2):325–332, 2001.
- [DM94] Scott C. Douglas and Teresa H.Y. Meng. Normalized data nonlinearities for LMS adaptation. *IEEE transactions on signal processing*, 42(6):1352–1365, 1994.
- [Dou95] Scott Douglas, C. Exact expectation analysis of the LMS adaptive filter. *IEEE transactions on signal processing*, 43(12):2863–2871, 1995.
- [Dur60] J. Durbin. The fitting of time series models. *Rev. int. stat. inst.*, 28:233–244, 1960.
- [EF86] E. Eleftherious and D.D. Falconer. Tracking properties and steady state performance of RLS adaptive filtering algorithm. *IEEE transactions on acoustics, speech and signal processing*, ASSP-34:1097–1110, 1986.
- [FP89] B. Friedlander and B. Porat. Adaptive IRR algorithms based on higher-order statistics. *IEEE trans. Acou. Speech and Sign. Proc.*, 37(4):485–495, 1989.
- [Fri84] B. Friedlander. The overdetermined recursive instrumental variable method. *IEEE transactions on automatic control*, Ac-29(4):353–356, 1984.
- [FW85] Arie Feuer and Ehud Weinstein. Convergence analysis of LMS filters with uncorrelated gaussian data. *IEEE transactions on acoustics, speech and signal processing*, ASSP-33(1):222–230, 1985.
- [Gar88] W.B. Gardner. *Statistical Spectral Analysis*. Prentice-Hall, Englewood Cliffs, N.J., 1988.
- [Gau09] C.F. Gauss. *Theoria Motus Corporum Coelestium in Sectionibus Conicus Solem Ambientum*. Hamburgo, 1809. Traducción: Dover, Nueva York, 1963.

- [GBT99] George-Othon Glentis, Kostas Berberidis, and Sergios Theodoridis. Efficient least squares adaptive algorithms for FIR transversal filtering. *IEEE signal processing magazine*, pages 13–41, julio 1999.
- [Gia87] G.B. Giannakis. Cumulants: A powerful tool in signal processing. *Proc. IEEE*, 75:1333–1334, 1987.
- [GM89] G.B. Giannakis and J.M. Mendel. Identification of non-minimum phase systems using higher order statistics. *IEEE transactions on acoustic, speech and signal processing*, 37(3):360–377, 1989.
- [God74] D.N. Godard. Channel equalization using a kalman filter for fast data transmission. *IBM J. Res. Dev.*, 18:267–273, 1974.
- [GS90] G. B. Giannakis and A. Swami. On estimating non-causal non-minimum phase ARMA models of non-gaussian processes. *IEEE Trans. acoust. speech and signal processing*, 38:478–495, 1990.
- [GWK99] Saul B. Gelfand, Yongbein Wei, and James V. Krogmeier. The stability of variable step-size LMS algorithms. *IEEE transactions on signal processing*, 47:3277–3288, 1999.
- [Hay83] S. Haykin. *Nonlinear Methods of Spectral Analysis*. Springer-Verlag, Berlin, Alemania, segunda edición, 1983.
- [Hay88] Simon Haykin. *Digital Communications*. John Wiley and sons, New York, 1988.
- [Hay94] S. Haykin. *Blind Deconvolution*. Prentice-Hall PTR, New Jersey, 1994.
- [Hay96] S. Haykin. *Adaptive Filter Theory*. Prentice-Hall, Upper Saddle River, N.J., tercera edición, 1996.
- [Hay00] S. Haykin, editor. *Unsupervised Adaptive Filtering*, volume 1 y 2. John Wiley and sons., New York, 2000.
- [Hin90] M.J. Hinich. Detecting a transient signal by bispectral analysis. *IEEE Trans. Acoustics, Speech and Signal Processing*, 38(7):1277–1283, julio 1990.
- [HKGD71] P.J. Huber, B. Kleiner, T. Gasser, and G. Dumermuth. Statistical methods for investigating phase relations in stationary stochastic processes. *IEEE transactions audio electroacoustics*, AU-19:78–86, 1971.
- [HOS00] K. Hidaka, H. Ohmori, and A. Sano. Adaptive filter for a linear time-varying system with a high-order estimator. *IEE Proc. Control theory appl.*, 147(4):395–402, 2000.

- [HSK96] B. Hassibi, A.H. Sayed, and T. Kailath. H^∞ optimality of the LMS algorithm. *IEEE trans. signal process.*, 44:267–280, 1996.
- [HTL00] Pau-Lo Hsu, Tsung-Yu Tasi, and Fu-Ching Lee. Applications of a variable step size algorithm to QCEE adaptive IIR filters. *IEEE transactions on signal processing*, 48(1):250–254, 2000.
- [Huz81] M. Huzii. Estimation of coefficient of an autoregressive process by using higher order moments. *J. Time series analysis*, 2:87–93, 1981.
- [JK92] Björn Jelonnek and Karl-Dirk Kammeyer. Improved methods for the blind system identification using higher order statistics. *IEEE transactions on signal processing*, 40(12):2947–2960, 1992.
- [Joh95] C. Richard Johnson, JR. On the interaction of adaptive filtering, identification and control. *IEEE signal processing magazine*, pages 22–37, marzo 1995.
- [Kai99] Thomas Kaiser. Adaptive MA-parameter estimation with modulated cumulants. pages 387–391. IEEE signal processing workshop on HOS, 1999.
- [Kay88] M. S. Kay. *Modern Spectral Estimation: Theory and Application*. Prentice-hall, Englewood Cliffs, N.J., 1988.
- [KCM84] N. Kalouptsidis, G. Carayannis, and D. Manolakis. A fast covariance type algorithm for sequential least-squares filtering and prediction. *IEEE transactions on automatic control*, AC-29(8):752–755, 1984.
- [Kes86] S.B. Kesler. *Modern Spectrum Analysis, II*. IEEE press, New York, 1986.
- [KJYS97] Hyoungill Kim, Bumki Jeon, Taewon Yang, and Koeng-Mo Sung. Recursive estimation algorithm for FIR systems using the 3rd and 4th order cumulants. In *ICASSP 1997*, pages 248–251, 1997.
- [Kol39] A.N. Kolmogorov. Sur l’interpolation et extrapolation des suites stationnaires. *C.R. Acad. Sci.*, 208:2043–2045, 1939.
- [KP88] K.I Kim and E.J. Powers. A digital method for modeling quadratically nonlinear systems with a general random input. *IEEE Trans. on Acoustics, Speech and Signal Processing*, 36(11):1758–1769, noviembre 1988.
- [Kre45] M.G. Krein. On a problem of extrapolation of a.n. kolmogorov. *C.R. Akad. Nauk SSSR*, 46:306–309, 1945.

- [Kus84] H. J. Kushner. *Approximation and Weak Convergence Methods for Random Processes with Applications to Stochastic System Theory*. MIT Press, Cambridge, MA, 1984.
- [Leg10] A.M. Legendre. Méthode des moindres carrés, pour trouver le milieu le plus probable entre les résultats de différentes observations. *Mem. Inst. France*, pages 149–154, 1810.
- [Lev47] N. Levinson. The wiener RMS (root-mean-square) error criterion in filter design and prediction. *J. Math. phys.*, 25:261–278, 1947.
- [Lii82] K. S. Lii. Non-gaussian ARMA model identification and estimation. *Proc. Bus. and Econ. Statistics (ASA)*, pages 135–141, 1982.
- [LR82] K. S. Lii and M. Rosenblatt. Deconvolution and estimation of transfer function phase and coefficients for nongaussian linear processes. *Annals of statistics*, 10(4):1195–1208, 1982.
- [Mac95] Odile Macchi. *Adaptive Processing: The LMS Approach with Applications in Transmission*. Wiley, New York, 1995.
- [Mar87] S.L. Marple, Jr. *Digital Spectral Analysis with Applications*. Prentice-Hall, Englewood Cliffs, 1987.
- [Maz79] S.L. Mazo. On the independence theory of equalizer convergence. *Bell syst. tech j.*, 58:963–993, 1979.
- [MB91] O. Macchi and N.J. Bershad. Adaptive recovery of a chirped sinusoid in noise, part i: Performance of the RLS algorithm. *IEEE Transactions on acoustic, speech and signal processing*, 39:583–594, 1991.
- [MD95] Mohsen Montazeri and Pierre Duhamel. A set of algorithms linking NLMS and block RLS algorithms. *IEEE transactions on signal processing*, 43(2):444–453, 1995.
- [Men91] J.M. Mendel. Tutorial on higher-order statistics (spectra) in signal processing and system theory: Theoretical results and some applications. *Proceedings of the IEEE*, 79(3):278–305, 1991.
- [MT91] George V. Moustakides and Sergios Theodoridis. Fast newton transversal filters- a new class of adaptive estimation algorithms. *IEEE transactions on signal processing*, 39(10):2184–2193, 1991.
- [MU84] T. Matsuoka and T.J. Ulrych. Phase estimation using the bispectrum. *Proceedings of IEEE*, 72:1403–1411, octobre 1984.
- [Nan94] A.K. Nandi. Blind identification of FIR systems using third order cumulants. *Signal processing*, 39:131–147, 1994. Elsevier Science B.V.

- [NKSK95] Yoon Jeong Na, Kwang Soon Kim, Ickho Song, and Taehyun Kim. Identification of nonminimum phase FIR systems using the third- and fourth-order cumulants. *IEEE transactions on signal processing*, 43(8):2018–2022, 1995.
- [NP93] Chrysostomos L. Nikias and Athina P. Petropulu. *Higher-Order Spectra Analysis. A Nonlinear Signal Processing Framework*. PTR Prentice Hall, Inc., Englewood Cliffs, New Jersey, 1993.
- [NR87] C.L. Nikias and M.R. Raghuveer. Bispectrum estimation: A digital signal processing framework. *Proceedings IEEE*, 75(7):869–891, 1987.
- [OM00] A.G. Orozco Lugo and D.C. McLernon. Blind channel equalization using chirp modulating signals. In *International Conference on Acoustics, Speech and Signal Processing, ICASP-2000*, pages 2721–2724, 2000.
- [Opp69] A.V Oppenheim. *Digital Processing of Signals*, chapter 8. McGraw-Hill, New York, 1969. ed. P. Gold y C.M. Rader.
- [Pap84] A. Papoulis. *Probability, Random Variables and Stochastic Processes*. McGraw-Hill, 1984.
- [Pet90] A. Petropulu. *Signal/Image Reconstruction from the Phase of the Bispectrum*. PhD thesis, Northeastern University, Boston, 1990.
- [PF89] B. Porat and B. Friedlander. The square-root overdetermined recursive instrumental variable algorithm. *IEEE transactions on automatic control*, 34(6), junio 1989.
- [PF91] Boaz Porat and Benjamin Friedlander. Blind equalization of digital communication channels using high-order moments. *IEEE transactions on signal processing*, 39(2):522–526, 1991.
- [Pla50] R.L. Plackett. Some theorems in least squares. *biometrika*, 47:149, 1950.
- [PN91] A.P. Petropulu and C.L. Nikias. Blind deconvolution using reconstruction from partial information, based on higher-order spectra. In *Proc. ICASSP'91*, pages 1757–1760, Toronto, Canada, mayo 1991.
- [RABC00] Diego Ruiz Padillo, Enrique Alameda Hernández, David Blanco Navarro, and María Carrión Pérez. Identificación ciega y adaptativa de sistemas MA mediante estadística de alto orden. In *XV Symposium Nacional de la Unión Científica Internacional de Radio*, Zaragoza, 2000.

- [RBAC99a] Diego P. Ruiz, David Blanco, Enrique Alameda, and Maria C. Carrion. Cubic phase coupling detection using AR modeling of the trispectrum. In *XIV Simposium Nacional de la Unión Científica Internacional de Radio URSI*, Santiago de Compostela. España., 1999.
- [RBAC99b] Diego Pablo Ruiz, David Blanco, Enrique Alameda, and Maria C. Carrion. Non-cooperative radar target identification using higher-order correlations. In *XIV Simposium Nacional de la Unión Científica Internacional de Radio URSI*, Santiago de Compostela. España., 1999.
- [RBAC00] Diego P. Ruiz, David Blanco, Enrique Alameda, and Maria C. Carrion. Automated radar target discrimination scheme based on a preprocessing of the scattered signal. In *Mediterranean Microwaves Symposium*, Tetuan. Morocco, 2000.
- [RBC99] D.P. Ruiz Padillo, D. Blanco Navarro, and M.C. Carrión Pérez. Noncooperative radar target identification using fourth-order correlations. In *Proc. Of SPW-HOS'99*, pages 353–356, Caesarea, Israel, junio 1999.
- [Ree62] I. S. Reed. On a moment theorem for complex gaussian processes. *IRE Trans. information theory*, IT-8:194–195, 1962.
- [RN85] M. R. Raghuveer and C. L. Nikias. Bispectrum estimation: A parametric approach. *IEEE Trans. on acoust. speech and signal processing*, ASSP-33(4), 1985.
- [Rui95] D.P. Ruiz Padillo. *Desarrollo de Técnicas de Procesado de Señal Mediante Estadística de órdenes Superiores. Aplicación a la Discriminación de Blancos de Radar*. PhD thesis, Universidad de Granada, 1995.
- [RZ99] Michael Reuter and James R. Zeidler. Nonlinear effects in LMS adaptive equalizers. *IEEE transactions on signal processing*, 47(6):1570–1579, 1999.
- [Sch98] A. Schuster. On the investigation of hidden periodicities with applications to a supposed 26-day period of meteorological phenomena. *Terr. Magn. Atmos. Electr.*, 3:13–41, 1898.
- [Sch80] M. Schetzen. *The Volterra and Wiener Theories of Nonlinear Systems*. Wiley, New York, 1980.
- [SK94] A.H Sayed and T. Kailath. A state-space approach to adaptive RLS filtering. *IEEE signal Process. Mag.*, 11(3):18–60, 1994.

- [SM90] A. Swami and J. M. Mendel. ARMA parameter estimation using only output cumulants. *IEEE Trans. acoust., speech and signal processing*, 38:1257–1265, 1990.
- [SM92] A. Swami and J.M. Mendel. Identifiability of the parameters of an ARMA process using cumulants. *IEEE Trans. on Automat. Contr.*, 37:268–273, febrero 1992.
- [Sol97] Victor Solo. The stability of LMS. *IEEE transactions on signal processing*, 45(12):3017–3026, 1997.
- [Ste73] G.W. Stewart. *Introduction to Matrix Computation*. Academic Press, New York, 1973.
- [Swa96] A. Swami. Fast transversal version of the RIV algorithm. *Int. Journ. Adaptive Control and Signal Processing*, 10:267–281, 1996.
- [Swa97] Ananthram Swami. On some parameter estimation problems in alpha-stable processes. In *International Conference on Acoustics, Speech and Signal Processing*, pages 3541–3544, 1997.
- [TJ99] John R. Treichler and C. Richard Johnson. Practical blind demodulation: Theory and design. *ICASSP-99 tutorial*, 1999.
- [Tug87] J. Tugnait. Fitting non-causal AR signal plus noise models to noisy non-gaussian linear processes. *IEEE trans. Automat. Contr.*, 32:547–552, 1987.
- [Tug95] Jitendra K. Tugnait. Blind equalization and estimation of digital communication FIR channels using cumulant matching. *IEEE transactions on communications*, 43(2/3/4):1240–1244, 1995.
- [WH60] B. Widrow and M.E. Hoff. Adaptive switching circuits. *IRE WESCON Conv. Rec.*, 4:96–104, 1960.
- [Wid70] B. Widrow. *Aspects of Network and System Theory*, chapter Adaptive filters. Holt, Rinehart and Winston, New York, 1970. ed. R.E. Kalman and N. Declaris.
- [Wid76] B. Widrow. Stationary and nonstationary learning characteristics of the LMS adaptive filter. *Proceeding IEEE*, 64:1151–1162, 1976.
- [Wie49] N. Wiener. *Extrapolation, Interpolation, and Smoothing of Stationary Time Series, with Engineering Applications*. MIT press, Cambridge, Mass., 1949.
- [WS85] B. Widrow and S.D. Stearns. *Adaptive Signal Processing*. Prentice Hall, New Jersey, 1985.

- [Yul27] G.U. Yule. On a method of investigating periodicities in disturbed series, with special reference to Wölfer's sunspot numbers. *Philos. Trans. Royal Soc. London*, A226:267–298, 1927.
- [ZMM93] Fu-Chun Zheng, Stephen McLaughlin, and Bernard Mulgrew. Blind equalization of nonminimum phase channels: Higher order cumulant based algorithm. *IEEE transactions on signal processing*, 41(2):681–691, 1993.
- [ZR91] W. Zhang and M.R. Raghuveer. Nonparametric bispectrum-based time-delay estimators for multiple sensor data. *IEEE Trans. on Signal Processing*, 39(13):770–774, marzo 1991.
- [ZSL96] Xian-Da Zhang, Yu Song, and Yan-Da Li. Adaptive identification on nonminimum phase ARMA models using higher order cumulants alone. *IEEE transactions on signal processing*, 44(5):1285–1288, 1996.
- [ZV98] Tong G. Zhou and Anthony Vassiliou. A hybrid second- and higher-order statistical method for blind deconvolution of seismic data. *Proceedings of ICSP98*, pages 164–167, 1998.
- [ZZ94] Xian-Da Zhang and Yuan-Sheng Zhang. FIR system identification using higher-order statistics alone. *IEEE transactions on signal processing*, 42(10):2854–2858, 1994.