



UNIVERSIDAD DE A CORUÑA
Departamento de Electrónica y Sistemas

Tesis doctoral

**NUEVOS MÉTODOS DE MONTE CARLO
PARA LA DEMODULACIÓN
EN SISTEMAS CON DIVERSIDAD**

Autor: MANUEL A. VÁZQUEZ LÓPEZ
Director: JOAQUÍN MÍGUEZ ARENAS
DICIEMBRE, 2008

Dr. Joaquín Míguez Arenas
Profesor titular de Universidad
Departamento de Teoría de la Señal y Comunicaciones
Universidad Carlos III de Madrid

CERTIFICA:

Que la memoria titulada “Nuevos métodos de Monte Carlo para la demodulación en sistemas con diversidad”, ha sido realizada por D. Manuel A. Vázquez López bajo mi dirección en el Departamento de Electrónica y Sistemas de la Universidad de A Coruña y concluye la Tesis que presenta para optar al grado de Doctor.

A Coruña, 1 de diciembre de 2008

Fdo: Dr. Joaquín Míguez Arenas
Director de la Tesis Doctoral

Tesis Doctoral: NUEVOS MÉTODOS DE MONTECARLO PARA LA
DEMODULACIÓN EN SISTEMAS CON DIVERSIDAD

Autor: D. Manuel A. Vázquez López

Director: Dr. Joaquín Míguez Arenas

Tutor: Dr. Luis Castedo Ribas

Fecha: 1 de diciembre de 2008

Tribunal

Presidente: Pedro Crespo Bofill

Vocal 1: Diego Pablo Ruiz Padillo

Vocal 2: Carles Anton Haro

Vocal 3: Mónica Fernández Bugallo

Secretario: Luis Castedo Ribas

A mis padres y a Raquel

Agradecimientos

Unas palabras para la que, probablemente, será la sección más leída de la tesis...

En primer lugar, como no podía ser de otra manera, quiero expresar mi más sincero agradecimiento al Profesor Joaquín Míguez por guiarme con sus consejos y algún que otro “latigazo” (de oficio) en la realización de este trabajo...y en los años que lo precedieron, por estar siempre cerca a pesar de la distancia.

También me gustaría recordar aquí a todos aquellos que han compartido conmigo (en algún momento) el laboratorio 1.2 de investigación de FIC (ellos saben quienes son)...y en especial a Fran. Las conversaciones mantenidas durante los años de doctorado (algunas de ellas de carácter científico, incluso) hicieron el trabajo mucho más ameno.

Por último, aunque no por ello menos importante, también quería agradecer a Alfred que, sin dejar en ningún momento de salpicar y chapotear feliz en el agua (helada a veces), me prestara su apoyo incondicional.

Resumen

El uso de varias antenas en el transmisor y en el receptor de un sistema de comunicaciones inalámbricas da lugar a un canal con diversidad que habitualmente se denomina MIMO (*Multiple Input Multiple Output*). La principal característica de este tipo de canales es que su capacidad¹ aumenta, para una relación señal a ruido (SNR) fija, al incrementar el número de antenas transmisoras y receptoras. Una mayor capacidad del canal implica una mayor eficiencia espectral o, equivalentemente, una mayor velocidad de transmisión utilizando el mismo ancho de banda, un recurso cada vez más limitado y costoso. Es por ello que los sistemas MIMO son una tecnología clave en la cuarta generación (4G) de telefonía móvil.

La detección óptima (*maximum a posteriori* o MAP) en un sistema MIMO tiene una complejidad computacional que crece exponencialmente con el número de antenas transmisoras, lo que dificulta su implementación práctica. Recientemente se ha propuesto el uso de métodos secuenciales de Monte Carlo (SMC) para llevar a cabo la igualación cuasi-óptima de canales MIMO. Los métodos SMC, también conocidos como filtrado de partículas, son técnicas basadas en simulación cuyo objetivo es la aproximación recursiva de distribuciones de probabilidad de interés mediante muestras de otra distribución conocida como tentativa. Se utilizan principalmente para estimar la señal no observada en sistemas dinámicos en formato de espacio de estados. El proceso de transmisión a través de un canal MIMO se puede modelar como uno de tales sistemas, lo que permite aplicar la metodología SMC para detectar la secuencia de símbolos transmitidos.

La mayor parte de los métodos de filtrado de partículas que se pueden encontrar en la literatura para la igualación de canales MIMO tienen también una complejidad exponencial en el número de antenas transmisoras. En este trabajo se presentan algoritmos SMC que superan esta limitación mediante un adecuado diseño de las correspondientes funciones tentativas que hace posible muestrearlas con una carga computacional menor y que, en última instancia, da lugar a una complejidad cúbica en el número de antenas transmisoras. Los métodos propuestos están especialmente diseñados para canales MIMO (desconocidos) selectivos en tiempo y en frecuencia. Las simulaciones llevadas a cabo demuestran que alcanzan un rendimiento muy próximo al del igualador SMC óptimo con un coste computacional mucho menor.

La segunda parte de este trabajo aborda el problema de la igualación de un canal MIMO selectivo en frecuencia cuando la longitud de su respuesta al impulso, generalmente conocida como el orden del canal, es desconocida. Ésta es una situación muy común en la práctica, y la aproximación más habitual para afrontarla consiste en considerar un orden de canal elevado (posiblemente *sobreestimando* el auténtico) para de esa manera evitar la enorme pérdida de rendimiento que sufre un receptor que asume un orden inferior al verdadero. Pero la complejidad computacional de los detectores MAP y de máxima verosimilitud (MV) en un

¹La capacidad de un canal de comunicaciones es la máxima velocidad de transmisión a la que es posible alcanzar una probabilidad de error arbitrariamente pequeña (cuando el emisor y el receptor no están sujetos a ninguna restricción).

canal MIMO crece exponencialmente (además de con el número de antenas transmisoras) con la longitud de la respuesta al impulso del canal. Así pues, sobreestimar el orden del canal incrementa la complejidad computacional del receptor. Además, tal y como se verá en el apartado de simulaciones, también da lugar a una progresiva pérdida de rendimiento a medida que el orden de canal considerado se aleja (por encima) del verdadero. En esta tesis se presentan dos métodos de igualación de canales MIMO capaces de estimar el orden del canal. Uno de ellos es una extensión a los métodos SMC de complejidad reducida anteriormente propuestos que tiene en cuenta la incertidumbre asociada al orden del canal. El algoritmo resultante todavía tiene una complejidad cúbica en el número de antenas transmisoras, y únicamente experimenta una pérdida de rendimiento significativa con respecto a los métodos propuestos que sí conocen el orden del canal cuando la relación señal a ruido del sistema es alta. El otro método presentado está basado en la idea de procesado por supervivientes (PSP) y, a costa de una complejidad exponencial en el número antenas transmisoras (aunque lineal en el de valores que puede tomar el orden del canal), alcanza un rendimiento muy cercano al óptimo.

Summary

The use of several antennas in both the transmitter and the receiver of a wireless communications system leads to a channel with diversity, usually known as MIMO (*Multiple Input Multiple Output*). The main advantage of this kind of channels is that their capacity² increases, for a fixed signal to noise ratio (SNR), as the number of transmitting and receiving antennas grows. A higher channel capacity entails a higher spectral efficiency i.e., a higher transmission velocity employing the same bandwidth, the latter being a limited and expensive resource. This is the essential reason why MIMO systems are regarded as a key technology in the fourth generation (4G) of mobile telephony.

Optimal detection (*maximum a posteriori* or MAP) in a MIMO system has a complexity that grows exponentially with the number of transmitting antennas, which makes practical implementations hard. Recently, the application of sequential Monte Carlo (SMC) methods to quasi-optimal MIMO channel equalization has been proposed. SMC methods, also known as particle filtering, are simulation-based techniques that aim at the recursive approximation of probability distributions of interest by drawing samples from another distribution known as proposal or importance function. They are mainly used to estimate the hidden signal of a dynamic system in state-space form. The transmission through a MIMO channel can be modelled as such system, which allows for the SMC framework to be applied to the detection of the transmitted symbols sequence.

Most of the particle filtering methods that can be found in the literature for the equalization of MIMO channels also have an exponential complexity with respect to the number of transmitting antennas. In this work, new SMC algorithms that overcome this limitation are introduced. They rely on a convenient design of the proposal functions that allows for them to be sampled with a lower computational burden, which results in a cubic complexity on the number of transmitting antennas. The proposal methods have been specifically designed for (unknown) frequency and time selective MIMO channels, and the simulations carried out show that they achieve a performance very close to that of the SMC optimal equalizer but involve a much lower computational load.

The second part of the present work tackles the problem of the equalization of a MIMO channel when the length of its channel impulse response (CIR), generally referred to as the channel order, is unknown. This is a common situation in practice, and the most usual approach consists in considering a high channel order (that possibly *overestimates* the real one) in order to avoid the huge performance degradation that a receiver assuming a channel order below the true one suffers. However, the computational complexity of MAP and maximum likelihood (ML) detectors for MIMO channel not only grows exponentially with the number of transmitting antennas, but also with the length of the CIR. Therefore, overestimating the channel order increases the computational complexity of the receptor. Moreover, as it will be seen in the

²The capacity of a communications channel is the highest transmission velocity which allows for arbitrarily small error probability at the receiver (when neither transmitter nor the receiver are subject to any constraints).

simulations chapter, it also leads to a performance loss that aggravates as the assumed channel order gets far above the true one. In this thesis, two methods that estimate the channel order are introduced. One of them is an extension of the complexity-constrained methods proposed earlier, able to handle the uncertainty associated with the channel order. The resulting algorithms still have a cubic complexity on the number of input streams, and its performance only degrades, compared to that of the proposed methods that have perfect knowledge of the channel order, when the system SNR is high. The other presented method is based on the concept of per-survivor processing (PSP) and, at the expense of an exponential complexity on the number of transmitting antennas (although linear on the number of possible channel orders), achieves a nearly optimal performance.

Contenido

Lista de figuras	XIX
Lista de tablas	XXI
Lista de pseudocódigos	XXIII
1. Introducción	1
1.1. Comunicaciones inalámbricas	1
1.2. Panorama actual de las comunicaciones móviles	3
1.3. Múltiples antenas en transmisión y recepción	4
1.4. Métodos Secuenciales de Monte Carlo	5
1.5. Igualación de canales MIMO mediante métodos SMC	6
1.6. Canales MIMO de orden desconocido	7
1.7. Organización de la tesis	8
2. Métodos de Monte Carlo Secuenciales	11
2.1. Notación	11
2.2. Métodos de Monte Carlo	12
2.3. Muestreo enfatizado	14
2.4. Métodos Secuenciales de Monte Carlo	17
2.4.1. Sistemas dinámicos en formato de espacio de estados	18
2.4.2. Muestreo enfatizado secuencial	20
2.4.3. El problema de la degeneración de los pesos y el remuestreo	21
2.4.4. Convergencia	29
2.4.5. Resumen	30
2.5. Conclusiones	31
3. Modelo de señal	33
3.1. Modelado de sistemas de comunicaciones MIMO	33
3.1.1. Modelo apilado	37
3.2. Capacidad del canal MIMO	38
3.3. Planteamiento del problema	39

3.4. Conclusiones	39
4. Revisión de aproximaciones anteriores	41
4.1. Métodos convencionales para la igualación de canales MIMO	41
4.1.1. Receptores Lineales	41
4.1.2. Receptor realimentado	44
4.1.3. El algoritmo de Viterbi	45
4.1.4. Procesado por supervivientes	47
4.1.5. Algoritmo BCJR	48
4.1.6. Algoritmo EM	50
4.2. Igualación de canales MIMO basada en métodos SMC	52
4.2.1. Función tentativa óptima sin suavizado	53
4.2.2. Función tentativa óptima con suavizado	55
4.2.3. Detector M estocástico	57
4.2.4. Filtrado de partículas con estructura de árbol	60
4.3. Conclusiones	62
5. Nuevos métodos SMC de complejidad reducida para la igualación de canales MIMO	63
5.1. Variables auxiliares	64
5.1.1. Muestreo del canal	67
5.1.2. Muestreo de los datos	67
5.2. Distribución conjunta del canal y los datos	76
5.2.1. Muestreo del canal	78
5.2.2. Muestreo de los datos	80
5.3. Conclusiones	83
6. Igualación de canales MIMO de orden desconocido	87
6.1. Modelo de señal	89
6.2. Aproximaciones anteriores	89
6.2.1. Basadas en teoría de la información	90
6.2.2. CME	91
6.2.3. Combinación de las funciones de coste de identificación e igualación de canal	92
6.2.4. Métodos basados en subespacios	93
6.3. SMC utilizando la función tentativa óptima	93
6.4. Nuevos métodos para igualar canales de orden desconocido	95
6.4.1. Métodos SMC de complejidad reducida	95
6.4.2. Igualación mediante procesado por supervivientes	107
6.5. Conclusiones	112

7. Simulaciones	115
7.1. Nuevos métodos SMC para la igualación de canales MIMO	116
7.1.1. Métodos SMC basados en variables auxiliares	117
7.1.2. Aproximación de la distribución conjunta del canal y los datos	120
7.1.3. Influencia del número de partículas	125
7.2. Igualación de canales MIMO de orden desconocido	125
7.2.1. Métodos SMC de complejidad reducida para canales MIMO de orden desconocido	126
7.2.2. Igualación mediante procesado por supervivientes	128
7.3. Conclusiones	131
8. Conclusiones y líneas futuras	133
8.1. Conclusiones	133
8.1.1. Métodos SMC de complejidad reducida	134
8.1.2. Canales MIMO de orden desconocido	136
8.2. Líneas futuras de investigación	137
8.2.1. Evaluación del rendimiento de los métodos propuestos utilizando un demostrador MIMO	137
8.2.2. Paralelización	138
8.2.3. Aplicación en sistemas de localización	138
8.2.4. Orden de canal desconocido y variante con el tiempo	139
8.2.5. Codificación turbo	139
A. Derivación del filtro MMSE	141
B. Detección mediante el algoritmo de Viterbi	145
C. Estimación del canal mediante RLS	153
D. Cálculo de $p(\mathbf{y}_t \mathbf{s}_{0:t}, \mathbf{y}_{0:t-1})$	157
E. Aproximación de la esperanza de la matriz de correlación de las observaciones	159
F. Derivada de una traza respecto a una matriz	161

Lista de figuras

2.1.	Aproximación de una distribución $\mathcal{N}(0,1)$ por medio de (izquierda) 50 muestras y (derecha) 1000 muestras de la misma.	13
2.2.	Aproximación de una distribución laplaciana de parámetros $\mu = -1$ y $b = 0,3$ mediante muestras obtenidas de una gaussiana de media $\mu = 0,5$ y varianza $\sigma^2 = 0,4$	16
2.3.	Aproximación de una distribución laplaciana de parámetros $\mu = -1$ y $b = 0,3$ mediante muestras obtenidas de una gaussiana de media $\mu = -1$ y varianza $\sigma^2 = 0,2$	17
2.4.	Diagrama de un sistema dinámico en formato de espacio de estados. Los círculos representan funciones y los cuadrados vectores aleatorios.	19
2.5.	Descripción esquemática de una operación de remuestreo.	24
2.6.	Obtención de los índices de las muestras a replicar en el remuestreo multinomial.	26
2.7.	Obtención de los índices de las muestras a replicar en el remuestreo estratificado.	27
2.8.	Obtención de los índices de las muestras a replicar en el remuestreo sistemático.	28
2.9.	Operación de un filtro de partículas.	30
3.1.	Esquema de un sistema de comunicaciones MIMO.	34
3.2.	Matriz de canal conjunta \mathbf{H}_t	35
4.1.	Esquema de un receptor lineal.	42
4.2.	Esquema de un receptor DFE.	45
4.3.	Diagrama de <i>trellis</i> de un PSP.	49
6.1.	Dispersión de la señal transmitida en un canal selectivo en frecuencia.	87
6.2.	Cálculo de las estimaciones blandas considerando un orden de canal $m < \text{máx}\{\mathcal{M}\}$	105
6.3.	Esquema del algoritmo para la igualación de canales MIMO de orden desconocido basado en procesamiento por supervivientes.	109
7.1.	BER para distintos valores de la SNR (dB). Los algoritmos SMC utilizan $M = 192$ partículas. Los resultados están promediados sobre 2519 tramas de datos. Modelo de canal AR.	118

7.2.	MSE para distintos valores de la SNR (dB). Los algoritmos SMC utilizan $M = 192$ partículas. Los resultados están promediados sobre 2519 tramas de datos. Modelo de canal AR.	119
7.3.	BER para distintos valores de la SNR (dB). Los algoritmos SMC utilizan $M = 192$ partículas. Los resultados están promediados sobre 2037 tramas de datos. Modelo de autocorrelación de Clarke.	121
7.4.	MSE para distintos valores de la SNR (dB). Los algoritmos SMC utilizan $M = 192$ partículas. Los resultados están promediados sobre 2037 tramas de datos. Modelo de autocorrelación de Clarke.	122
7.5.	BER para distintos valores de la SNR (dB). Los algoritmos SMC utilizan $M = 192$ partículas. Los resultados están promediados sobre 2037 tramas de datos. Modelo de autocorrelación de Clarke.	123
7.6.	MSE para distintos valores de la SNR (dB). Los algoritmos SMC utilizan $M = 192$ partículas. Los resultados están promediados sobre 2037 tramas de datos. Modelo de autocorrelación de Clarke.	124
7.7.	Evolución de la probabilidad <i>a posteriori</i> de los diferentes órdenes de canal en el MLSD-m . Los resultados están promediados sobre 12092 tramas de datos. Modelo de autocorrelación de Clarke.	125
7.8.	BER para distintos valores de la SNR (dB). Los algoritmos SMC utilizan $M = 192$ partículas. Los resultados están promediados sobre 1185 tramas de datos. Modelo de autocorrelación de Clarke.	127
7.9.	Evolución de la probabilidad <i>a posteriori</i> de los diferentes órdenes de canal en el algoritmo USIS. Los resultados están promediados sobre 1185 tramas de datos. Modelo de autocorrelación de Clarke.	128
7.10.	BER para distintos valores de la SNR (dB). Los algoritmos SMC utilizan $M = 192$ partículas. Los resultados están promediados sobre 1185 tramas de datos. Modelo de autocorrelación de Clarke.	129
7.11.	Evolución de la probabilidad <i>a posteriori</i> de los diferentes órdenes de canal en el MLSD-m. Los resultados están promediados sobre 1185 tramas de datos. Modelo de autocorrelación de Clarke.	130
8.1.	Posible implementación paralela de los igualadores SMC propuestos. Cada partícula es propagada por un elemento de procesamiento diferente salvo cuando se lleva a cabo el remuestreo.	139
B.1.	Diagrama <i>trellis</i> típico.	146
B.2.	Primera etapa del proceso de detección mediante el algoritmo de Viterbi.	148
B.3.	Segunda etapa del proceso de detección mediante el algoritmo de Viterbi.	149
B.4.	Tercera etapa del proceso de detección mediante el algoritmo de Viterbi.	149
B.5.	Cuarta etapa del proceso de detección mediante el algoritmo de Viterbi.	150
B.6.	Quinta etapa del proceso de detección mediante el algoritmo de Viterbi.	151

Lista de tablas

7.1. Frecuencia de error relativa en la detección del orden del canal.	131
--	-----

Lista de pseudocódigos

1.	SIS	22
2.	Algoritmo de remuestreo multinomial	25
3.	Algoritmo de remuestreo residual	27
4.	Algoritmo SIS óptimo sin suavizado	55
5.	Algoritmo SIS óptimo suavizado	58
6.	SIS con variables auxiliares basado en la triangularización de la matriz de autocorrelación del canal	73
7.	SIS con variables auxiliares basado en estimadores lineales	75
8.	SIS basado en la aproximación conjunta del canal y los datos	84
9.	Algoritmo SIS óptimo suavizado para orden desconocido	96
10.	Algoritmo SIS de complejidad reducida para orden desconocido	108
11.	Algoritmo para la igualación de canales MIMO de orden desconocido basado en procesado por supervivientes	111
12.	Estimación del canal mediante RLS	155

Capítulo 1

Introducción

Vamos a abordar el problema de la detección de la secuencia de datos transmitida en sistemas de comunicación inalámbrica con múltiples antenas transmisoras y múltiples antenas receptoras (MIMO, del inglés *Multiple Input Multiple Output*). Se sabe que la complejidad del receptor óptimo es en este caso exponencial en el número de antenas emisoras y en longitud de la respuesta al impulso del canal [1]. La primera parte de este trabajo se dedica a explorar nuevas aproximaciones al problema basadas en métodos secuenciales de Monte Carlo que reducen sustancialmente dicha carga computacional sin que por ello se degrade el rendimiento de manera significativa. En la segunda parte de la tesis se estudian las dificultades adicionales que surgen cuando la longitud de la respuesta al impulso del canal es desconocida, y se proponen algoritmos capaces de operar en tales condiciones.

Este capítulo comienza con una breve panorámica de los actuales sistemas de comunicaciones inalámbricas y sus limitaciones, que sirve para poner en contexto el problema de la demodulación en canales MIMO. Se discuten a continuación los métodos secuenciales de Monte Carlo [2] como herramienta de tratamiento de señales y su aplicación a problemas de comunicaciones como el planteado en este trabajo. El capítulo concluye con una descripción de la organización del resto de la memoria.

1.1. Comunicaciones inalámbricas

Un sistema de comunicación inalámbrica permite transferir información entre dos entidades (el emisor y el receptor) sin que exista un medio físico guiado (un cable) entre ambos. La construcción del primer sistema de estas características se atribuye a Guglielmo Marconi en el año 1895. El radiotelégrafo ideado por éste probó que las ondas electromagnéticas, que Maxwell había caracterizado matemáticamente en el año 1865 y cuya existencia había demostrado Hertz en el 1887 [3], podían en verdad emplearse para transmitir información. Dicha información es ya a comienzos del siglo XX música y voz, lo cual supone el nacimiento de la radio moderna. Las aplicaciones que desde entonces han encontrado los sistemas de comunicación sin hilos son incontables. Algunas de ellas, como la radio y la televisión, han tenido una enorme repercusión

y ya forman parte del legado cultural del pasado siglo. Otras, como por ejemplo los sistemas globales de navegación por satélite (GNSS, del inglés *Global Navigation Satellite System*) tales como el GPS (*Global Positioning System*) americano o su equivalente europeo, GALILEO (todavía en desarrollo), son de uso cada vez más común gracias a los avances en la tecnología digital, que han permitido abaratar el coste del hardware necesario para sacarles partido.

Cualquier medio que permita la propagación de ondas electromagnéticas puede hacer el papel de canal en un sistema de comunicación inalámbrica. Así pues, aún cuando el emisor y el receptor estén muy alejados entre sí, todavía es posible disponer de un medio de transmisión (más o menos hostil, dependiendo de la distancia y el entorno que los separe) entre ellos. Esta es una de las principales ventajas de los sistemas de comunicaciones inalámbricas: permiten transmitir información entre localizaciones muy remotas (piénsese en los satélites que proveen la señal GPS, orbitando alrededor de la Tierra a distancias comprendidas entre los 5 y los 20 kilómetros [4]), entre las cuales sería difícil o imposible extender un cable o medio guiado.

Otra de las ventajas inherentes a un canal inalámbrico es la movilidad. Algunos autores [5] diferencian entre movilidad y portabilidad entendiendo que el primer término alude a la capacidad que tiene un usuario de beneficiarse de un mismo servicio en diferentes lugares (sin que ello suponga que el emisor o el receptor del sistema o sistemas de comunicación implicados se desplacen) y el segundo a la posibilidad de que el emisor o el receptor varíen su posición sin que se interrumpa el funcionamiento del sistema. Cuando aquí hablemos de movilidad nos estaremos refiriendo a esta segunda idea, en la que se sustentan aplicaciones tales como la telefonía móvil o las redes de ordenadores Wi-Fi.

Si bien las ventajas de un sistema comunicación inalámbrica son evidentes aún sin saber nada acerca de la implementación del mismo, no ocurre lo mismo con sus inconvenientes, cuya naturaleza es más técnica y dependiente de cada sistema particular. Aquí únicamente mencionaremos las más comunes sin centrarnos en ningún tipo de sistema concreto.

El canal radio, el medio de transmisión en un sistema de comunicación inalámbrica, es un medio hostil. Fenómenos como las pérdidas por propagación (la potencia de la señal que llega al receptor disminuye con la distancia que lo separa del emisor), el multitrayecto (que puede ocasionar interferencia intersimbólica) o la naturaleza cambiante del canal, hacen que la recuperación de la señal transmitida sea una tarea mucho más complicada que en un sistema con hilos [6]. Esto está estrechamente relacionado con el rendimiento del sistema, que tendrá una tasa de fallos mayor, lo cual a su vez repercute en la velocidad de transmisión.

Otros inconvenientes asociados a las comunicaciones inalámbricas tienen que ver con el hecho de utilizar un medio de transmisión compartido. En primer lugar, en un sistema de comunicaciones con hilos se puede controlar (en cierta medida) qué usuarios o dispositivos tienen acceso al medio. En un canal inalámbrico, un intruso situado en un entorno del emisor o del receptor podría “escuchar” las señales intercambiadas por éstos. Esto plantea problemas de seguridad y privacidad, que pueden combatirse mediante el uso de técnicas de cifrado [7], pero sólo hasta cierto punto. Por otro lado, el espectro electromagnético es un recurso limitado que debe ser compartido por los diferentes servicios que utilizan el canal radio, lo cual restringe el

ancho de banda de que dispone cada uno de ellos.

A todo lo dicho cabe añadir que, con frecuencia, los dispositivos emisor y/o receptor implicados en un sistema de comunicaciones inalámbricas son móviles y funcionan mediante baterías, por lo que su consumo de energía (y por tanto, su potencia de cálculo) está limitado

1.2. Panorama actual de las comunicaciones móviles

La telefonía móvil ha jugado desde sus orígenes, allá por los años ochenta, un papel fundamental en el desarrollo de los sistemas de comunicación inalámbrica. Las técnicas de transmisión empleadas por éstos sistemas se han ido sofisticando cada vez más para hacer posibles las velocidades de transmisión requeridas por los nuevos servicios demandados por los usuarios. La segunda generación de telefonía móvil (2G) sólo permite el envío señales de voz o tramas de datos cortas y, así, el estándar GSM (*Global System for Mobile Communications*) dominante en Europa tan solo es capaz de transmitir datos a 9,6 kbps [8]. Su combinación con GPRS (*General Packet Radio Service*), en lo que se conoce como 2.5G, ofrece la posibilidad de transferir contenidos multimedia, pero la tasa de transferencia es de tan solo 114 kbps [8]. La tercera generación (3G) vio la luz a comienzos de la presente década y es una implementación de la recomendación técnica IMT-2000 (*International Mobile Telecommunications by the year 2000*) [9]. Fue desarrollada teniendo en mente servicios tales como el envío de imágenes y vídeo, o la conexión a internet de banda ancha, y ya permite velocidades de hasta 2 Mbps. Esto abre la puerta a toda una serie de aplicaciones, impensables hace unos pocos años, de las que se puede beneficiar cualquier poseedor de un teléfono móvil de tercera generación (en España el número de terminales 3G vendidos supera actualmente los 10 millones).

El desarrollo de las comunicaciones inalámbricas no sólo ha mejorado las capacidades de los teléfonos celulares, sino que ha expandido el uso de estas tecnologías a entornos en los que, hasta hace muy poco, la transmisión con hilos era la única alternativa viable. Las redes de ordenadores inalámbricas, abanderadas por el logo Wi-Fi [10], han experimentado un enorme crecimiento en los últimos años y su presencia ya no se limita únicamente a lugares públicos, en los que constituyen una forma fácil y económica de ofrecer acceso a internet a un conjunto variable de usuarios. Hoy en día, son muchos los hogares que utilizan redes Wi-Fi, ya sea para hacer llegar la conexión a internet a cualquier parte de la casa o simplemente por comodidad (para prescindir de cableado). La tecnología Wi-Fi permite construir redes de área local inalámbricas (WLAN, del inglés *Wireless Local Area Network*) con velocidades de transmisión de hasta 30 Mbps. El estándar IEEE802.16, comúnmente denominado WiMAX (*Worldwide Interoperability for Microwave Access*), promete tasas de transferencia de 75 Mbps y un área de servicio de 50 kilómetros [11], con lo que ya podríamos hablar de redes inalámbricas de área metropolitana o WMAN (*Wireless Metropolitan Area Network*).

Además de WLAN y WMAN, existe un tercer tipo de redes inalámbricas pensado para aquellas situaciones en las que emisor y receptor se hallan muy próximos: son las redes inalámbricas de área personal o WPAN (*Wireless Personal Area Network*). Bluetooth permite

la comunicación inalámbrica entre dispositivos portátiles separados por distancias de hasta 200 metros con velocidades que alcanzan los 3 Mbps [9]. Otras tecnologías aún en desarrollo, como por ejemplo UWB (*Ultra Wide Band*) [12] (basada en la transmisión de pulsos de duración de muy corta, del orden de nanosegundos), permitirán velocidades de cientos de megabits por segundo.

1.3. Múltiples antenas en transmisión y recepción

Como ya se comentó anteriormente en este capítulo, el espectro radioeléctrico es un recurso limitado que debe ser compartido por los distintos usuarios y servicios del canal radio, por lo que constantemente se están buscando nuevas formas de conseguir un mejor aprovechamiento del mismo. Esto implica incrementar la eficiencia espectral, que no es sino la velocidad de transmisión por unidad de ancho de banda. En la siguiente generación de telefonía móvil (4G) se esperan velocidades de transmisión por encima de los 20 Mbps [10] que permitan el envío de voz, música y vídeo de alta fidelidad. Esto plantea un nuevo reto que precisa de técnicas de transmisión mucho más sofisticadas que las empleadas en 3G (cuya tasa de transferencia es 10 veces inferior).

Desde hace tiempo se sabe que la capacidad del canal radio se incrementa linealmente con el mínimo entre el número de antenas emisoras y el de receptoras [13]. Así pues, utilizando sistemas con varias antenas emisoras y varias antenas receptoras, denominados genéricamente MIMO, se pueden conseguir importantes mejoras en la velocidad de transmisión. Este tipo de sistemas han acaparado la atención de buena parte de la comunidad investigadora en los últimos años [14, 15, 16, 17, 18] y son la base del modelo de señal utilizado a lo largo de esta tesis.

Básicamente, un sistema MIMO añade una dimensión espacial al canal de comunicaciones. Si las antenas empleadas a ambos extremos del canal están lo suficientemente separadas entre sí (la distancia entre dos antenas cualesquiera en el emisor o en el receptor es al menos la mitad de la longitud de onda de la señal transmitida), podemos considerar que cada par antena transmisora-antena receptora constituye un canal de comunicaciones independiente de todos los demás. En tal caso, la información radiada por cada antena transmisora “viaja” a través de tantos subcanales diferentes como pares transmisor-receptor tenga el sistema. Esta diversidad es la que da lugar al mencionado incremento en la capacidad (y prestaciones, por tanto) de los sistemas MIMO.

Así pues, la ventaja de este tipo de sistemas frente a los sistemas de comunicación inalámbrica tradicionales constituidos por una única antena emisora y una única antena receptora, también conocidos como SISO (del inglés *Single Input Single Output*), es clara. Los inconvenientes son una complejidad mucho mayor en la detección de la señal transmitida, tanto computacional (los receptores requieren más memoria y potencia de cálculo) como conceptual (la derivación de los algoritmos es más complicada), y un mayor coste del sistema de comunicaciones. En relación a éste último, téngase en cuenta que aún así el coste asociado al despliegue de nuevas antenas es mucho menor que el asociado a adquirir más ancho de banda.

1.4. Métodos Secuenciales de Monte Carlo

Los métodos de Monte Carlo son técnicas que permiten hacer inferencia sobre una magnitud de interés utilizando muestras generadas mediante simulación. Por lo general se utilizan en aquellos problemas en los que, o bien no es posible obtener una solución analítica, o bien ésta es muy compleja.

La expresión “método de Monte Carlo” fue acuñada por el matemático polaco Stanislaw Ulam en el 1946 [19] y hace referencia a uno de los casinos más famosos de Mónaco (las actividades que tienen lugar en un casino son de naturaleza aleatoria y repetitiva, y esas son las cualidades inherentes a todo método de Monte Carlo). Sin embargo, 15 años antes Enrico Fermi ya había utilizado métodos de Monte Carlo en sus estudios relacionados con las propiedades del neutrón y lo había hecho basándose en técnicas de muestreo estadístico [20]. En el año 1954, los autores de [21], uno de los trabajos pioneros en el uso de métodos de Monte Carlo, afirman que éstos no son una innovación, sino un nombre nuevo para algo ya existente.

Los métodos secuenciales de Monte Carlo (SMC), también conocidos como filtros de partículas (FP), son una clase particular de métodos de Monte Carlo cuyo objetivo fundamental es la aproximación recursiva (secuencial) de distribuciones de probabilidad de interés mediante muestras. Éstas proceden, por lo general, de otra distribución distinta de la que se pretende aproximar y se ponderan mediante el principio de muestreo enfatizado (IS, del inglés *Importance Sampling*), de tal manera que cuando se habla de métodos secuenciales Monte Carlo (o de filtrado de partículas) el principio de IS también suele estar implícito.

La historia del primer proyecto ambicioso en el que se hizo uso de métodos de Monte Carlo es relatada en [20] por uno de los investigadores que participaron en el mismo. Se trataba de estudiar el comportamiento de los neutrones en un proceso de fisión. Para llevar a cabo dicha tarea disponían del primer ordenador de propósito general: el ENIAC (*Electronic Numerical Integrator And Computer*). Algunas pruebas y comparaciones con otras aproximaciones al problema probaron la validez de los métodos de Monte Carlo y éstos despertaron el interés de la comunidad científica. Si bien comenzaron a emplearse en el campo de la física [22], no tardaron en ser adoptados como una nueva herramienta por otras disciplinas. Algunos autores [2] refieren su uso en [23] para la simulación de largas cadenas de polímeros como la primera aplicación de los métodos de Monte Carlo secuenciales.

A pesar de todo lo expuesto, los métodos de Monte Carlo no comenzaron a hacerse populares hasta los años 90, cuando la potencia de los nuevos microprocesadores permitió mejorar su eficacia¹ y, sobre todo, “democratizar” su uso, ya que antes de la llegada de los ordenadores personales eran una herramienta al alcance de sólo unos pocos.

En el campo del procesado de señal, el interés despertado por los métodos secuenciales de Monte Carlo surgió a raíz del *bootstrap filter* propuesto en [24]. Dicho trabajo introdujo por primera vez la idea del remuestreo, fundamental en todo método de Monte Carlo secuencial, y

¹Como veremos en el Capítulo 2, el rendimiento de cualquier método de Monte Carlo depende en gran medida del número de muestras empleadas.

sentó las bases de la aplicación de esta metodología al problema de la estimación de la señal no observada en un sistema dinámico en formato de espacio de estados.

1.5. Igualación de canales MIMO mediante métodos secuenciales de Monte Carlo

El diseño de receptores para canales MIMO es conceptualmente complicado y su implementación computacionalmente costosa. En particular, la igualación óptima de canales MIMO tiene una complejidad exponencial en el número de antenas transmisoras [1], lo que la convierte en una opción inviable en la mayor parte de situaciones prácticas. Inevitablemente, este hecho ha atraído la atención de numerosos investigadores en el ámbito del tratamiento estadístico de señales, que han planteado la posibilidad de aplicar métodos secuenciales de Monte Carlo a éste y otros problemas que surgen en el campo de las comunicaciones digitales (véase el tutorial [2]). Así, son numerosos los trabajos ([17, 25, 26, 27, 16] entre otros) que han abordado el problema de la igualación cuasi-óptima (sin una degradación notable en el rendimiento) de canales MIMO utilizando técnicas SMC. En [17] se muestra que es posible conseguir tasas de error de bit (BER, del inglés *Bit Error Rate*) casi óptimas utilizando filtrado de partículas, pero la complejidad del esquema de muestreo propuesto sigue siendo exponencial en el número de flujos de entrada. Los autores de [26] reducen la complejidad de los algoritmos SMC agrupando partículas idénticas que son representadas como caminos en un árbol. Una aproximación similar es el detector M estocástico propuesto en [27] para el sistema de codificación BLAST (*Bell Labs layered space-time*) y aplicada en [28] al problema de la detección multiusuario en sistemas de acceso múltiple por división de código (CDMA, del inglés *Code Division Multiple Access*). Las técnicas propuestas en [26] y [27] se explican con más detalle en las secciones 4.2.4 y 4.2.3, respectivamente.

Otra forma de reducir la complejidad de la detección en sistemas MIMO es mediante la triangularización de la matriz de autocorrelación de canal. Ejemplos de aplicación de este método en el contexto de los sistemas CDMA se pueden encontrar en [1] y, en combinación con FP, en [29]. Esquemas similares (basados en la descomposición QR) para sistemas de multiplexación por división ortogonal de frecuencia (OFDM, *Orthogonal Frequency-Division multiplexing*) se describen en [30, 31]. La triangularización de la matriz de autocorrelación de canal también ha sido utilizada conjuntamente con técnicas de filtrado de partículas para la detección en canales MIMO genéricos. En [25, 16] los autores proponen un método iterativo (turbo) para la estimación del canal y la decodificación de los datos transmitidos. En cada iteración, dada la última estimación de canal disponible, la decodificación se lleva a cabo por medio de métodos secuenciales de Monte Carlo de manera similar a la descrita en [29]. Los datos detectados se utilizan entonces para actualizar la estimación del canal mediante un estimador convencional cuadrático.

Nótese que ninguno de los trabajos arriba mencionados aborda el problema de la estimación conjunta del canal MIMO y los datos utilizando filtros de partículas. Este hecho es llamativo

porque algunos autores han señalado que aunque este tipo de técnicas son adecuadas para problemas complejos que involucran la estimación de variables discretas (los símbolos, en este caso) y continuas (los coeficientes del canal), el problema que consiste únicamente en la detectar la secuencia transmitida se puede resolver más eficientemente con métodos determinísticos o semi-determinísticos [32].

En el Capítulo 5 se presentan nuevos métodos SMC para la estimación conjunta del canal y los datos en sistemas MIMO genéricos que evitan la complejidad exponencial en el número de antenas transmisoras. Han sido diseñados específicamente para canales selectivos en tiempo y en frecuencia, por lo que que son especialmente adecuados para sistemas de comunicaciones móviles o, en general, entornos cambiantes en los que no existe una buena línea de visión directa entre emisor y receptor. Si bien no alcanzan el rendimiento del igualador SMC óptimo explicado en la Sección 4.2.2, ofrecen un atractivo compromiso entre la complejidad y el rendimiento.

1.6. Canales MIMO de orden desconocido

En la igualación de canales MIMO selectivos en frecuencia normalmente se asume que la longitud de la respuesta al impulso del canal, también conocida como el orden del canal, es conocida. Sin embargo, esto no es cierto en la mayor parte de casos prácticos y, para evitar la importante degradación de rendimiento que ocurre cuando se subestima el orden del canal, suele asumirse un orden elevado, posiblemente mayor que el verdadero (en cuyo caso se dice que el orden del canal está siendo sobreestimado). Esta aproximación tiene dos inconvenientes: por un lado da lugar a una pérdida progresiva de rendimiento a medida que el orden considerado se aleja del verdadero y, por otro, incrementa la complejidad del receptor.

Son pocos los trabajos que han abordado el problema de la estimación del orden del canal y menos aún los que lo han hecho para el caso de un sistema MIMO. Métodos basados en teoría de la información, como el principio de la descripción de menor longitud (MDL, del inglés *Minimum Description Length*) [33] o el criterio de información de Akaike (AIC, del inglés *An Information Criterion*) [34] han sido propuestos a tal efecto, pero dichas técnicas tienden a sobreestimar el orden del canal cuando la SNR es alta [35, 36]. En [37] se propone una nueva aproximación para resolver el problema de la selección del modelo que mejor explica un conjunto de observaciones. El método, acuñado con el término CME (del inglés, *Conditional Model Estimator*), tiene su origen en la teoría de estadísticos suficientes y su superioridad frente al MDL es reivindicada por el autor del mismo. En [38] se utiliza para estimar el orden de un canal con múltiples entradas y una única salida (MISO, del inglés *Multiple Input Single Output*), y una extensión que permite su aplicación a canales MIMO es derivada (aunque no se muestran resultados numéricos). Sin embargo, CME está basado en la suposición de que el canal es fijo a lo largo de toda la trama de observaciones, y el procesado de éstas ha de llevarse a cabo por bloques, lo que impide su utilización de una manera adaptativa.

Una aproximación al problema completamente diferente, basada en la minimización de la combinación de dos funciones de coste, es la expuesta en [39]. Los inconvenientes de esta

técnica son que está pensada únicamente para canales con una única entrada y múltiples salidas (SIMO, *Single Input Multiple Output*) y que no considera canales variantes con el tiempo.

El uso de métodos de filtrado de partículas también ha sido propuesto para la igualación de canales selectivos en frecuencia de orden desconocido [40]. Aunque tales algoritmos son efectivos, su complejidad es prohibitiva incluso para el sencillo sistema SISO utilizado en [40]. Una técnica similar aplicada a sistemas OFDM se puede encontrar en [41].

En el Capítulo 6 se presentan dos nuevos algoritmos que resuelven el problema de la estimación de los datos y el canal (incluyendo su orden) en canales MIMO variantes con el tiempo. El primero de ellos extiende los métodos SMC del Capítulo 5 para manejar la incertidumbre asociada al orden del canal y lo hace manteniendo la misma complejidad (polinómica) en el número de antenas transmisoras. El segundo es un algoritmo determinista basado en el concepto de procesado por supervivientes y, a costa de una complejidad exponencial en el número de entradas, alcanza unos resultados comparables a los del algoritmo que detecta la secuencia más verosímil con pleno conocimiento de la información del estado del canal. Ambos métodos calculan, en cada instante de tiempo, estimaciones de la probabilidad *a posteriori* de cada uno de los posibles órdenes de canal considerados.

1.7. Organización de la tesis

El resto de la tesis está organizado de la siguiente manera. En el Capítulo 2 se explican los fundamentos de los métodos secuenciales de Monte Carlo y su aplicación a la estimación de la señal no observada en un sistema dinámico en formato de espacio de estados. El principio de muestreo enfatizado (tanto en su forma original como en su versión secuencial) y el remuestreo, necesario debido al problema de la degeneración de los pesos, son algunas de las ideas revisadas en esta parte de la tesis. En la sección dedicada al remuestreo se incluye además una breve descripción de algunos de los métodos más importantes. El capítulo se cierra con un resumen de los resultados de convergencia más relevantes en el contexto de los métodos SMC.

El modelo matemático que describe la transmisión en un sistema de comunicaciones MIMO selectivo en frecuencia y en tiempo se introduce en el Capítulo 3. Junto con algunas consideraciones acerca de la capacidad del canal MIMO, las ecuaciones que constituyen la base sobre la cual se diseñan todos los algoritmos presentados a lo largo de la tesis se encuentran en dicho capítulo.

En el Capítulo 4 se revisan brevemente algunos de los procedimientos ya existentes para la igualación de canales MIMO, distinguiendo claramente entre los que hacen uso de la metodología SMC y las técnicas convencionales de naturaleza determinista.

Los métodos de filtrado de partículas de complejidad reducida para la igualación de canales MIMO que constituyen el tema central de este trabajo se explican en el Capítulo 5, mientras que el Capítulo 6 está dedicado a la igualación de canales MIMO de orden desconocido. En éste último, el modelo de señal del Capítulo 3 es extendido para tener en cuenta la incertidumbre ahora asociada al orden del canal. Algunas aproximaciones anteriores

al problema son discutidas entonces, antes de introducir los métodos propuestos en este trabajo. Los resultados de las simulaciones llevadas a cabo para probar la eficacia de todos los algoritmos propuestos se encuentran en el Capítulo 7 junto con un análisis detallado de los mismos.

Por último, el Capítulo 8 está dedicado a las conclusiones y posibles líneas de investigación futura.

Capítulo 2

Métodos de Monte Carlo Secuenciales

En este capítulo se explican los fundamentos y la motivación de los métodos secuenciales de Monte Carlo (SMC). También conocidos como filtros de partículas (FP), son un conjunto de técnicas basadas en simulación cuyo objetivo es la aproximación recursiva de distribuciones de probabilidad. Se basan en el principio de muestreo enfatizado (IS, del inglés *Importance Sampling*) para obtener un conjunto de muestras que, junto con sus pesos asociados, constituyen una aproximación discreta de la función de densidad de probabilidad de interés.

Los métodos secuenciales de Monte Carlo fueron desarrollados en los años cincuenta para la simulación de largas cadenas de polímeros [21, 23]. Sin embargo, no fue hasta cuatro décadas después, coincidiendo con un sustancial incremento en la potencia de cálculo de los ordenadores, cuando suscitaron el interés de la comunidad científica y su ámbito de aplicación se extendió a multitud de campos como la física o la ingeniería [42].

Los FP suelen utilizarse para resolver el problema de la estimación de la señal no observada (el estado) en sistemas dinámicos en formato de espacio de estados. El filtro de Kalman (KF, del inglés *Kalman Filter*) proporciona la solución óptima a este problema, pero está basado en hipótesis de linealidad y gaussianidad del sistema y su rendimiento se degrada de forma significativa cuando éstas no se cumplen. Los métodos SMC son una familia de técnicas mucho más generales que el KF y no están sujetos a dichas restricciones, lo que los convierte en una herramienta más versátil (y por tanto con más posibilidades de aplicación práctica) que el KF.

2.1. Notación

Con objeto de simplificar la notación, vamos a denotar de la misma manera a una variable aleatoria (v.a.) y a su realización. Es decir, dada una v.a. X , escribiremos simplemente $p(x)$ (en lugar de $p(X = x)$) para indicar la probabilidad de que X tome el valor x . Además, utilizaremos $p(\cdot)$ indistintamente para referirnos a la función de densidad de probabilidad (f.d.p.) o a la función de masa de probabilidad (f.m.p.) del argumento. Cuando es una u otra (es decir, cuando la v.a. del argumento es continua y cuando es discreta) quedará claro a partir del contexto. Nótese que, en cualquier caso, una f.m.p. se puede escribir como una f.d.p. por medio de

funciones generalizadas delta de Dirac: si una v.a. discreta toma los valores x_1, x_2, \dots, x_N con probabilidades $p(x_1), p(x_2), \dots, p(x_N)$, respectivamente, entonces su f.m.p. se puede expresar como la f.d.p.

$$p(x) = \sum_{i=1}^N p(x_i) \delta(x - x_i) \quad (2.1)$$

donde $\delta(\cdot)$ es la función delta de Dirac.

La densidad condicional de una v.a., x , dada otra v.a., y , la denotaremos por $p(x|y)$. Desde un punto de vista bayesiano, si x es una v.a. de interés que estamos tratando de estimar e y una v.a. que aporta algún tipo de información acerca de x , entonces $p(x|y)$ es la densidad *a posteriori* de x y $p(x)$ su densidad *a priori*. A lo largo de la tesis será recurrente el uso de esta terminología.

Para distinguir entre variables escalares, vectoriales y matriciales seguiremos la aproximación habitual consistente en utilizar una en letra minúscula para escalares, una letra negrita en minúscula para los vectores, y una en letra negrita mayúscula para las matrices. Así, a representa un escalar, \mathbf{a} un vector y \mathbf{A} una matriz. Además, por $a(i)$ denotaremos el i -ésimo elemento del vector \mathbf{a} , y por $a(i, j)$ el elemento que ocupa la i -ésima fila y j -ésima columna de la matriz \mathbf{A} .

Por último, utilizaremos la notación $x_{0:t}$ para referirnos al conjunto $\{x_0, \dots, x_t\}$.

2.2. Métodos de Monte Carlo

Un método de Monte Carlo es una técnica que proporciona una solución aproximada a un problema cuantitativo haciendo inferencia a partir de muestras obtenidas por medio de simulación estadística.

Por lo general, se recurre a métodos de Monte Carlo para resolver problemas cuya solución analítica, o bien no existe, o bien es muy compleja. Su uso está extendido en áreas tales como las matemáticas aplicadas, la física, la química, la economía, la medicina... (véase [43]). Nosotros nos vamos a centrar en su utilidad a la hora de construir una representación de una f.d.p. mediante muestras. Cuando éstas proceden de la propia f.d.p. que se quiere aproximar, se habla de muestreo exacto. Es el caso que vamos a estudiar a continuación.

Supongamos que no conocemos la f.d.p. de x , pero sí tenemos una manera de obtener muestras de la misma. Entonces podemos utilizar M muestras aleatorias, independientes e idénticamente distribuidas (i.i.d.), $x^{(1)}, x^{(2)}, \dots, x^{(M)} \sim p(x)$, para aproximar su f.d.p. como

$$p_M(x) = \frac{1}{M} \sum_{i=1}^M \delta(x - x^{(i)}). \quad (2.2)$$

Obsérvese que, en última instancia, la probabilidad que $p_M(x)$ asigna a un determinado valor de la variable aleatoria, x' , está determinada por el número de muestras de dicho valor obtenidas, es decir, por la cardinalidad del conjunto $\{x^{(j)} : x^{(j)} = x', j = 1 \dots, M\}$.

A partir de p_M es posible aproximar el valor esperado de cualquier función integrable de x , $f(x)$, que nos interese. Sea $\mathbb{E}_{p(\cdot)}[\cdot]$ la esperanza del argumento respecto de la función de probabilidad en el subíndice. Si x es una v.a. continua con f.d.p. $p(x)$, el valor esperado de $f(x)$ viene dado por

$$\mathbb{E}_{p(x)}[f(x)] = \int f(x)p(x)dx \quad (2.3)$$

y se puede aproximar como

$$\mathbb{E}_{p(x)}[f(x)] \approx \mathbb{E}_{p_M(x)}[f(x)] = \int f(x)p_M(x)dx = \frac{1}{M} \sum_{i=1}^M f(x^{(i)}). \quad (2.4)$$

Nótese que la ecuación (2.4) puede emplearse para calcular momentos de x tales como la media ($f(x) = x$) o la varianza ($f(x) = (x - \mathbb{E}_{p(x)}[x])^2$).

La bondad de la aproximación dada por la ecuación (2.2) depende del número de muestras empleadas, de tal manera que [42]

$$\lim_{M \rightarrow \infty} \mathbb{E}_{p_M(x)}[f(x)] = \mathbb{E}_{p(x)}[f(x)]. \quad (2.5)$$

La Figura 2.1 ilustra la aproximación (2.2) por medio de un ejemplo. En ella, una distribución normal de media cero y varianza uno se aproxima por medio de métodos Monte Carlo con 50 (izquierda) y 1000 (derecha) muestras. Se puede ver claramente que al aumentar el número de muestras generadas, la aproximación construida (representada por medio de un histograma normalizado) se asemeja más a la f.d.p. gaussiana objetivo (dibujada superpuesta).

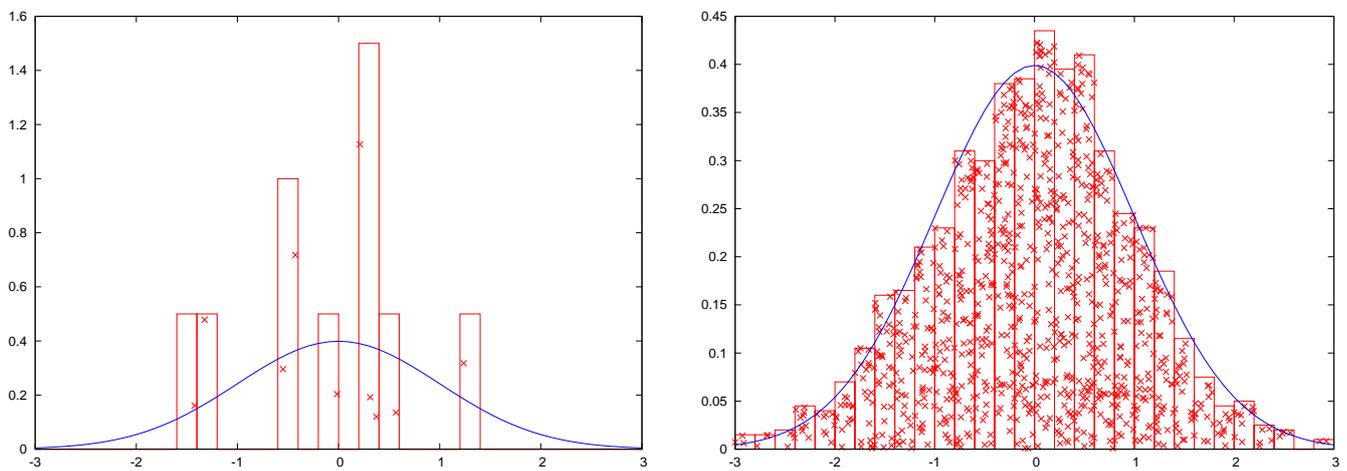


Figura 2.1: Aproximación de una distribución $\mathcal{N}(0, 1)$ por medio de (izquierda) 50 muestras y (derecha) 1000 muestras de la misma.

Así pues, la cantidad de muestras que se necesitan para resolver un determinado problema utilizando métodos de Monte Carlo depende por un lado de la complejidad del mismo y, por otro, del grado de precisión que queramos obtener en la solución.

2.3. Muestreo enfatizado

La sección 2.2 demostró la manera en que se pueden utilizar los métodos de Monte Carlo para representar una distribución de probabilidad, $p(x)$, por medio de un conjunto de muestras. Entonces, las muestras se obtenían directamente de la f.d.p. objetivo, $p(x)$, y hablábamos de muestreo exacto. Cuando esto no es posible (o deseable), el principio de muestreo enfatizado (IS, del inglés *Importance Sampling*) proporciona una manera alternativa de representar $p(x)$ mediante muestras tomadas de otra distribución alternativa conocida como “tentativa”.

Sea f una función arbitraria de x , cuya esperanza respecto de la densidad $p(x)$, conocida sólo hasta una constante de proporcionalidad c , puede escribirse como

$$\mathbb{E}_{p(x)} [f(x)] = \frac{\int cf(x)p(x)dx}{\int cp(x)dx} \quad (2.6)$$

donde el término $\int cp(x)dx = c$ es una constante de normalización. Si $q(x)$ es una densidad con el mismo soporte (dominio) que $p(x)$, se puede reescribir la ecuación (2.6) de la siguiente forma:

$$\mathbb{E}_{p(x)} [f(X)] = \frac{\int cf(x)p(x)\frac{q(x)}{q(x)}dx}{\int cp(x)\frac{q(x)}{q(x)}dx} = \frac{\int f(x)q(x)\tilde{w}(x)dx}{\int q(x)\tilde{w}(x)dx} \quad (2.7)$$

donde

$$\tilde{w}(x) = c\frac{p(x)}{q(x)} \quad (2.8)$$

es la función de peso.

Las integrales de la ecuación (2.7) se pueden aproximar generando M muestras aleatorias i.i.d., $x^{(1)}, x^{(2)}, \dots, x^{(M)}$, de la distribución $q(x)$. En tal caso, una estimación de $\mathbb{E}_{p(x)} [f(x)]$ sería:

$$\hat{\mathbb{E}}_{p(x),M} [f(x)] \approx \frac{1/M \sum_{i=1}^M f(x^{(i)})\tilde{w}(x^{(i)})}{1/M \sum_{i=1}^M \tilde{w}(x^{(i)})} = \sum_{i=1}^M f(x^{(i)})w^{(i)} \quad (2.9)$$

donde

$$w^{(i)} \triangleq w(x^{(i)}) = \frac{\tilde{w}(x^{(i)})}{\sum_{j=1}^M \tilde{w}(x^{(j)})} \propto \frac{p(x^{(i)})}{q(x^{(i)})} \quad (2.10)$$

es el peso normalizado de la i -ésima muestra.

La ecuación (2.9) implica que podemos aproximar la esperanza de una función arbitraria, $f(x)$, respecto de una densidad $p(x)$, tomando muestras de otra f.d.p. diferente, $q(x)$, conocida como función tentativa o propuesta. Para ello, es necesario asociar a cada muestra $x^{(i)}$ de $q(x)$ un peso, $w^{(i)}$, obtenido mediante la ecuación (2.10), que indica la calidad de la muestra. Una muestra junto con su peso correspondiente constituyen una partícula.

Intuitivamente, y a la vista de la ecuación (2.9), podemos interpretar el peso de una muestra, $w^{(i)}$, como la probabilidad acumulada por la región del espacio de estados en torno a dicha

muestra. Entonces, la densidad $p(x)$ se puede aproximar como:

$$p_M(x) = \sum_{i=1}^M w^{(i)} \delta(x - x^{(i)}) \quad (2.11)$$

verificándose [42] que

$$\lim_{M \rightarrow \infty} \mathbb{E}_{p_M(x)} [f(x)] = \mathbb{E}_{p(x)} [f(x)]. \quad (2.12)$$

Hasta este momento, la única condición que hemos impuesto sobre la función tentativa es que ha de tener el mismo soporte que la función objetivo. Formalmente, esto quiere decir que no existe ningún subconjunto no nulo del dominio de $p(x)$ con medida positiva donde la integral de la densidad $q(x)$ es nula, o equivalentemente:

$$p(x) > 0 \Rightarrow q(x) > 0 \forall x \in A : A \text{ tiene medida positiva.} \quad (2.13)$$

Garantizado esto, la ecuación (2.12) afirma que podemos aproximar el valor esperado de $f(x)$ respecto de $p(x)$ eligiendo $q(x)$ arbitrariamente. Sin embargo, el número de muestras necesarias para obtener una buena estimación de $\mathbb{E}_{p(x)} [f(x)]$ podría ser muy elevado (nótese que el resultado (2.12) es asintótico), por lo que generalmente es necesario diseñar cuidadosamente la función tentativa.

Saber que una v.a. continua (o discreta con un espacio muestral muy grande) toma un determinado valor con una probabilidad muy próxima a cero aporta muy poca información acerca de la misma. Por tanto, si tenemos un conjunto de muestras de una v.a., todas ellas con una probabilidad muy pequeña, sabemos en realidad muy poco acerca del experimento aleatorio asociado a dicha variable. Se necesitan muestras con una cierta probabilidad que den una idea acerca de “lo que suele ocurrir”, más que de “lo que casi nunca ocurre”. Bajo este punto de vista, si la forma de la función tentativa no es similar a la de la función objetivo, la mayor parte de las muestras obtenidas a partir de la primera tendrán una probabilidad despreciable en la segunda, y será necesario generar una gran cantidad de muestras de $q(x)$ para conseguir unas pocas que aporten alguna información acerca de $p(x)$.

A modo de ejemplo, supongamos que queremos aproximar una distribución laplaciana con parámetro de posición $\mu = -1$ y parámetro de escala $b = 0,3$, que haría el papel de $p(x)$ o función objetivo. Si la función tentativa utilizada, $q(x)$, es una normal con parámetro de posición (media) $\mu = 0,5$ y parámetro de escala (varianza) $\sigma^2 = 0,4$, podría ocurrir algo como lo que se muestra en la Figura 2.2: la mayor parte de las muestras obtenidas de $q(x)$ difícilmente habrían sido generadas por la distribución $p(x)$ y, por tanto, son poco representativas de la misma. El problema está en que la función tentativa elegida rara vez obtendrá muestras en la región del espacio muestral en la que la función objetivo condensa la mayor parte de su área (probabilidad), que es precisamente aquella en la que estamos interesados. En consecuencia, es necesario generar un gran número de muestras de $q(x)$ para encontrar alguna que sea significativa en $p(x)$.

Si las funciones tentativa y objetivo se “parecen”, las muestras obtenidas de la primera son representativas de la distribución de la segunda, es decir, podrían haber resultado de muestrear

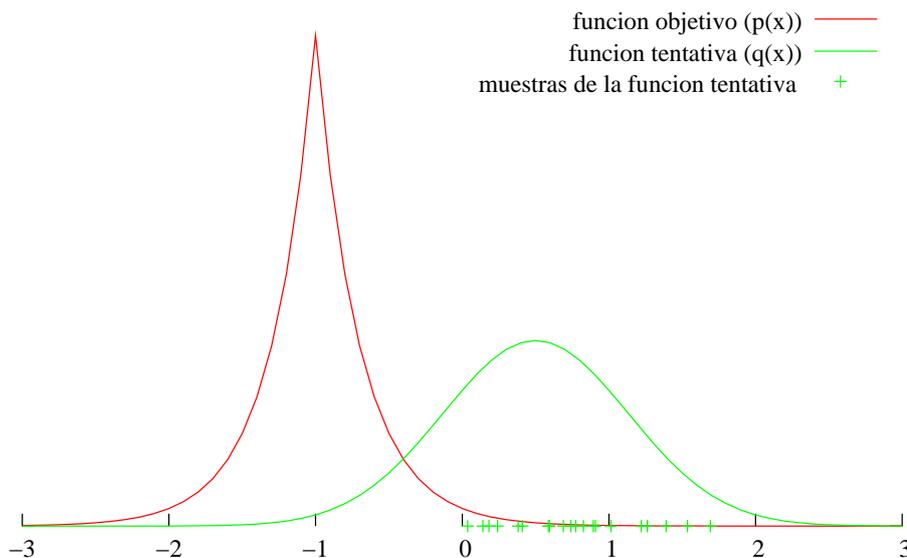


Figura 2.2: Aproximación de una distribución laplaciana de parámetros $\mu = -1$ y $b = 0,3$ mediante muestras obtenidas de una gaussiana de media $\mu = 0,5$ y varianza $\sigma^2 = 0,4$.

la segunda, y por medio del principio de IS, somos capaces de asignarles un peso acorde con la densidad de probabilidad que tiene la función objetivo $p(x)$ en esa región del espacio de estados. En la Figura 2.3 hemos utilizado como función tentativa una normal centrada en el mismo punto que la laplaciana y más apuntada (con menor varianza). Unas pocas muestras generadas a partir de esta función tentativa dan una idea bastante aproximada de cual es la forma de la función objetivo.

Es importante resaltar que los pesos de las muestras que se utilizan para aproximar $p(x)$ están normalizados, por lo que a la vista de ellos no es posible saber si dichas muestras procedentes de $q(x)$ son o no significativas en la distribución de $p(x)$. Así, M muestras de $q(x)$ cuyo peso sin normalizar es muy próximo a 0 y que, por tanto, apenas tienen relevancia en la distribución objetivo, tendrán un peso igual a $1/M$ tras la normalización.

A la vista de lo expuesto, está claro que la bondad de la aproximación que el principio de IS permite hacer de una distribución depende en gran parte de la función tentativa elegida. Si una función tentativa es tanto mejor cuanto más se parezca a la función objetivo, entonces la función tentativa óptima es la propia función a aproximar (y estaríamos hablando de muestreo exacto). Evidentemente, el principio de IS está pensado para aquellos casos en los que, o bien no es posible muestrear directamente de $p(x)$ (no es trivial muestrear de una distribución cualquiera), o bien no es deseable (por ser costoso). Es decir, su principal motivación es evitar tener que tomar muestras de la distribución a aproximar. Sin embargo, para calcular los pesos, ésta debe ser fácilmente evaluable, al menos hasta una constante de proporcionalidad: dada una muestra $x^{(i)}$ de $q(x)$, obtener su peso asociado implica calcular $cp(x^{(i)})$ para algún c (incluso desconocido). Así pues, la función objetivo es difícil (o imposible) de muestrear pero fácil de evaluar, y la función tentativa ha de ser fácil de evaluar y de muestrear.

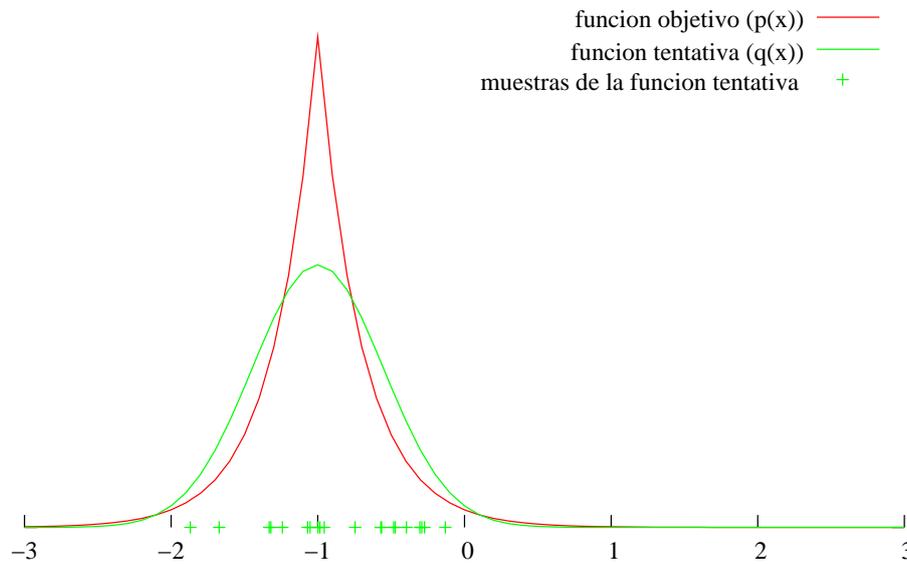


Figura 2.3: Aproximación de una distribución laplaciana de parámetros $\mu = -1$ y $b = 0,3$ mediante muestras obtenidas de una gaussiana de media $\mu = -1$ y varianza $\sigma^2 = 0,2$.

2.4. Métodos Secuenciales de Monte Carlo

Se entiende por métodos secuenciales de Monte Carlo un conjunto de técnicas basadas en simulación cuyo objetivo es la aproximación de una secuencia de distribuciones de probabilidad mediante muestras y pesos asociados. Las dos características fundamentales inherentes a esta metodología son:

- El cálculo de la distribución de probabilidad de interés ha de ser secuencial y recursivo, esto es, se van a obtener las partículas que aproximan la distribución en el instante de tiempo t a partir de las partículas que sirvieron para aproximarla en $t - 1$ y de la nueva observación.
- A las muestras (obtenidas, por lo general, muestreando de una función tentativa) se les asignan pesos por medio del principio de IS.

Es habitual clasificar como método de Monte Carlo cualquier técnica que resuelva un determinado problema mediante la simulación de muestras aleatorias. Así, los métodos de Monte Carlo se pueden aplicar a un conjunto muy heterogéneo de problemas. Los métodos secuenciales de Monte Carlo, por su parte, están pensados para ser aplicados principalmente a problemas de estimación que surgen en el contexto de los sistemas dinámicos en formato de espacio de estados.

2.4.1. Sistemas dinámicos en formato de espacio de estados

Multitud de problemas en el campo del procesamiento de señal se pueden describir mediante un sistema dinámico de la forma

$$\mathbf{x}_0 \sim p(\mathbf{x}_0) \quad (2.14)$$

$$\mathbf{x}_t = f_t(\mathbf{x}_{t-1}, \mathbf{u}_t) \quad (2.15)$$

$$\mathbf{y}_t = g_t(\mathbf{x}_t, \mathbf{v}_t) \quad (2.16)$$

donde $\mathbf{x}_t \in \mathbb{R}^{n_x}$ es el vector de estado, $\mathbf{y}_t \in \mathbb{R}^{n_y}$ es el vector de observaciones, f_t es la función de transición de estado, g_t es la función de observación, y \mathbf{u}_t y \mathbf{v}_t son vectores de ruido ($t \in \mathbb{N}$). Las funciones f_t , g_t y las distribuciones de probabilidad de los procesos de ruido, \mathbf{u}_t y \mathbf{v}_t , se suponen conocidas.

Las ecuaciones (2.15) y (2.16) determinan las dos secuencias relacionadas que caracterizan a todo sistema dinámico en formato de espacio de estados. La primera de ellas, la secuencia de estados $\{\mathbf{x}_t; t \in \mathbb{N}\}$, es una señal no observada definida por un proceso markoviano de primer orden ($p(\mathbf{x}_t | \mathbf{x}_{0:t-1}) = p(\mathbf{x}_t | \mathbf{x}_{t-1})$) cuya f.d.p. *a priori* en el instante inicial, $p(\mathbf{x}_0)$, es conocida. La otra, $\{\mathbf{y}_t; t \in \mathbb{N}\}$, es una secuencia de observaciones independientes entre sí dada la secuencia de estados. Las denotaremos por $\mathbf{x}_{0:t} \triangleq \{\mathbf{x}_0 \cdots \mathbf{x}_t\}$ e $\mathbf{y}_{1:t} \triangleq \{\mathbf{y}_1 \cdots \mathbf{y}_t\}$, respectivamente. Nótese que a cada una de las ecuaciones que describen el sistema se le puede asociar una densidad de probabilidad condicional: $p(\mathbf{x}_t | \mathbf{x}_{t-1})$ a la ecuación de estado y $p(\mathbf{y}_t | \mathbf{x}_t)$ a la ecuación de observación.

La Figura 2.4 muestra el diagrama de un sistema dinámico en formato de espacio de estados. El sistema se encuentra inicialmente en el estado \mathbf{x}_0 , que no tiene ninguna observación asociada. A partir de dicho estado el sistema evoluciona de acuerdo a la función de transición de estado, que depende del tiempo. Así, el estado del sistema en t ($t > 0$) se obtiene aplicando la función de transición de estado en el t -ésimo instante de tiempo, f_t , sobre el estado anterior, \mathbf{x}_{t-1} , y el vector de ruido \mathbf{u}_t . La ecuación de estado define formalmente esta transición. A modo de ejemplo, en la figura se puede ver como el estado \mathbf{x}_1 es el resultado de aplicar la función f_1 sobre los vectores \mathbf{x}_0 y \mathbf{u}_1 . En base al t -ésimo estado, \mathbf{x}_t , y al vector de ruido \mathbf{v}_t se genera la t -ésima observación, \mathbf{y}_t , mediante la función g_t . La ecuación (2.16) expresa esto matemáticamente. En la parte inferior de la Figura 2.4 podemos observar como \mathbf{x}_1 y \mathbf{v}_1 constituyen los argumentos de entrada de la función g_1 , que da lugar a la observación \mathbf{y}_1 .

El objetivo en todo sistema dinámico en formato de espacio de estados es estimar la secuencia de estados a partir de la secuencia de observaciones. Toda la información estadística relevante para llevar a cabo dicha tarea está contenida en la densidad *a posteriori*, $p(\mathbf{x}_{0:t} | \mathbf{y}_{0:t})$. A partir de ella, podemos evaluar cualquier estimador de la forma

$$\mathbb{E}_{p(\mathbf{x}_{0:t} | \mathbf{y}_{0:t})} [f(\mathbf{x}_{0:t})] = \int f(\mathbf{x}_{0:t}) p(\mathbf{x}_{0:t} | \mathbf{y}_{0:t}) d\mathbf{x}_{0:t} \quad (2.17)$$

siendo $f(\cdot)$ una función integrable.

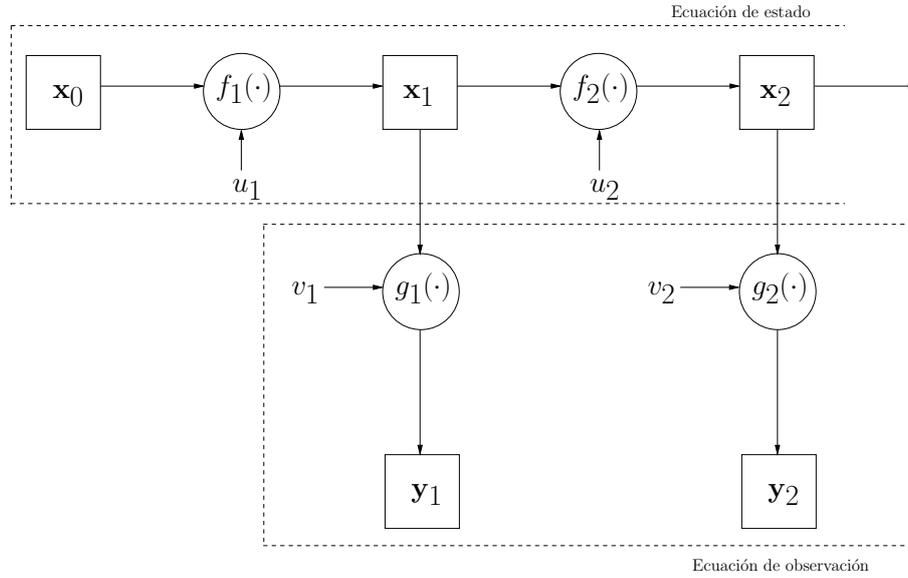


Figura 2.4: Diagrama de un sistema dinámico en formato de espacio de estados. Los círculos representan funciones y los cuadrados vectores aleatorios.

Aunque por conveniencia generalmente vamos a aproximar la f.d.p. *a posteriori* de la secuencia de estados completa, en la práctica son las densidades marginales las que son de interés. A partir de $p(\mathbf{x}_{0:t}|\mathbf{y}_{0:t})$ es fácil obtener la marginal $p(\mathbf{x}_t|\mathbf{y}_{0:t'})$, que se denomina densidad de filtrado si $t = t'$, densidad de suavizado si $t < t'$ y densidad de predicción si $t > t'$.

Como su propio nombre indica, una de las características inherentes a los métodos SMC es su habilidad para el procesamiento secuencial: recibida en el instante de tiempo t una nueva observación \mathbf{y}_t , el objetivo es estimar a partir de ella y de todas las observaciones anteriores el estado actual del sistema, \mathbf{x}_t , o la secuencia completa de estados $\mathbf{x}_{0:t}$.

La f.d.p. *a posteriori* de la secuencia de estados se puede descomponer aplicando el teorema de Bayes y la definición de probabilidad condicionada como

$$\begin{aligned}
 p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t}) &= \frac{p(\mathbf{y}_t|\mathbf{x}_{0:t}, \mathbf{y}_{1:t-1})p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t-1})}{p(\mathbf{y}_t|\mathbf{y}_{1:t-1})} \\
 &= \frac{p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{x}_{0:t-1}, \mathbf{y}_{1:t-1})}{p(\mathbf{y}_t|\mathbf{y}_{1:t-1})}p(\mathbf{x}_{0:t-1}|\mathbf{y}_{1:t-1}) \\
 &= \frac{p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{x}_{t-1})}{p(\mathbf{y}_t|\mathbf{y}_{1:t-1})}p(\mathbf{x}_{0:t-1}|\mathbf{y}_{1:t-1}), \tag{2.18}
 \end{aligned}$$

donde se ha tenido en cuenta que la distribución de la t -ésima observación está completamente determinada por el estado \mathbf{x}_t conocidas la función f_t y la distribución del ruido, por lo que $p(\mathbf{y}_t|\mathbf{x}_{0:t}, \mathbf{y}_{1:t-1}) = p(\mathbf{y}_t|\mathbf{x}_t)$. Asimismo, al ser la secuencia de estados un proceso markoviano se cumple que \mathbf{x}_t depende únicamente de \mathbf{x}_{t-1} y, por tanto, $p(\mathbf{x}_t|\mathbf{x}_{0:t-1}, \mathbf{y}_{1:t-1}) = p(\mathbf{x}_t|\mathbf{x}_{t-1})$.

Puesto que el término $p(\mathbf{y}_t|\mathbf{y}_{1:t-1})$ que aparece en (2.18) no depende de $\mathbf{x}_{0:t}$, finalmente

podemos escribir

$$p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t}) \propto p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{0:t-1}|\mathbf{y}_{1:t-1}). \quad (2.19)$$

La densidad de filtrado, $p(\mathbf{x}_t|\mathbf{y}_{0:t})$, también se puede calcular de forma recursiva aplicando sucesivamente las ecuaciones

$$\text{(predicción)} \quad p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) = \int p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})d\mathbf{x}_{t-1} \quad (2.20)$$

$$\text{(actualización)} \quad p(\mathbf{x}_t|\mathbf{y}_{1:t}) = \frac{p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{1:t-1})}{p(\mathbf{y}_t|\mathbf{y}_{1:t-1})}. \quad (2.21)$$

La técnica de predicción-actualización consiste en calcular la densidad predictiva $p(\mathbf{x}_t|\mathbf{y}_{1:t-1})$ utilizando la ecuación (2.20), para a continuación modificarla por medio de la ecuación (2.21) en base a la información que proporciona la observación t -ésima, \mathbf{y}_t , y obtener de esa manera la densidad de filtrado.

La constante de normalización que aparece en las ecuaciones (2.18) y (2.21) se puede reescribir por medio del teorema de las probabilidades totales como

$$p(\mathbf{y}_t|\mathbf{y}_{1:t-1}) = \int p(\mathbf{y}_t|\mathbf{x}_t, \mathbf{y}_{1:t-1})p(\mathbf{x}_t|\mathbf{y}_{1:t-1})d\mathbf{x}_t = \int p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{1:t-1})d\mathbf{x}_t. \quad (2.22)$$

El problema del método de predicción-actualización dado por las ecuaciones (2.20) y (2.21) es que es meramente formal, ya que en la práctica estas ecuaciones no suelen tener solución analítica. Únicamente la tienen si las ecuaciones de estado y observación son lineales y gaussianas (las funciones f_t y g_t son lineales y los vectores de ruido \mathbf{u}_t y \mathbf{v}_t gaussianos), o el conjunto de posibles estados es finito. En el primer caso las ecuaciones del filtro de Kalman [44] permiten hallar la media y covarianza de las densidades de filtrado y predicción, $p(\mathbf{x}_t|\mathbf{y}_{1:t-1})$ y $p(\mathbf{x}_t|\mathbf{y}_{1:t})$, respectivamente. Si el número de estado es finito, por otro lado, las integrales se pueden calcular como sumatorios sobre el conjunto de todos los posibles estados.

En la siguiente sección vamos a presentar una combinación del método de muestreo enfatizado con la descomposición (2.19) para obtener un algoritmo recursivo que aproxima la f.d.p. *a posteriori* $p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})$.

2.4.2. Muestreo enfatizado secuencial

Si aplicásemos directamente el método de muestreo enfatizado al problema de estimar la f.d.p. *a posteriori* $p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})$, la complejidad del algoritmo resultante aumentaría con el tiempo. Es decir, en el instante $t - 1$ obtendríamos una muestra $\mathbf{x}_{0:t-1}^{(i)} \sim q(\mathbf{x}_{0:t-1}|\mathbf{y}_{1:t-1})$ de dimensión $n_x \times (t - 1)$, que en el instante t se descartaría para generar una muestra $\mathbf{x}_{0:t}^{(i)} \sim q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})$ de tamaño $n_x \times t$. Así pues, la dimensión del espacio muestral sería cada vez mayor.

El objetivo de la estrategia de muestreo enfatizado secuencial (SIS, del inglés *Sequential Importance Sampling*) es evitar que la complejidad del método de IS aplicado al sistema dinámico descrito por las ecuaciones (2.15) y (2.16) se incremente con el tiempo. La

idea fundamental es reutilizar cada muestra, $\mathbf{x}_{0:t-1}^{(i)}$, simulada de acuerdo a la distribución $q(\mathbf{x}_{0:t-1}|\mathbf{y}_{1:t-1})$, para obtener una muestra $\mathbf{x}_{0:t}^{(i)}$ de $q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})$. Para ello se elige una función tentativa $q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})$ que se pueda descomponer de la forma:

$$q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t}) = q(\mathbf{x}_t|\mathbf{x}_{0:t-1}, \mathbf{y}_{1:t})q(\mathbf{x}_{0:t-1}|\mathbf{y}_{1:t-1}), \quad (2.23)$$

es decir, tal que se contenga a sí misma en el instante de tiempo anterior como distribución marginal. Se define recursivamente, por tanto, y aplicando repetidas veces la ecuación (2.23) sobre $q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})$ resulta

$$q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t}) = q(\mathbf{x}_0) \prod_{k=1}^t q(\mathbf{x}_k|\mathbf{x}_{0:k-1}, \mathbf{y}_{1:k}) \quad (2.24)$$

pudiéndose emplear como función tentativa inicial $q(\mathbf{x}_0) = p(\mathbf{x}_0)$, la función de densidad *a priori* del estado, que se ha supuesto conocida en la definición del sistema.

Para el caso de un sistema dinámico en formato de espacio de estados en el que se pretende aproximar la f.d.p. *a posteriori* del estado, $p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})$, el principio de IS asocia a cada muestra $\mathbf{x}_{0:t}^{(i)}$ un peso

$$\tilde{w}_t^{(i)} = \frac{p(\mathbf{x}_{0:t}^{(i)}|\mathbf{y}_{1:t})}{q(\mathbf{x}_{0:t}^{(i)}|\mathbf{y}_{1:t})}, \quad (2.25)$$

que por medio de las ecuaciones (2.19) y (2.23) verifica la proporcionalidad

$$\tilde{w}_t^{(i)} \propto \frac{p(\mathbf{y}_t|\mathbf{x}_t^{(i)})p(\mathbf{x}_t^{(i)}|\mathbf{x}_{t-1}^{(i)})p(\mathbf{x}_{0:t-1}^{(i)}|\mathbf{y}_{1:t-1})}{q(\mathbf{x}_t^{(i)}|\mathbf{x}_{0:t-1}^{(i)}, \mathbf{y}_{1:t})q(\mathbf{x}_{0:t-1}^{(i)}|\mathbf{y}_{1:t-1})} = \frac{p(\mathbf{y}_t|\mathbf{x}_t^{(i)})p(\mathbf{x}_t^{(i)}|\mathbf{x}_{t-1}^{(i)})}{q(\mathbf{x}_t^{(i)}|\mathbf{x}_{0:t-1}^{(i)}, \mathbf{y}_{1:t})} w_{t-1}^{(i)}. \quad (2.26)$$

Así pues, es posible calcular el peso de la muestra $\mathbf{x}_{0:t}^{(i)}$ a partir del peso normalizado, $w_{t-1}^{(i)}$, asociado a la muestra $\mathbf{x}_{0:t-1}^{(i)}$ en el instante de tiempo anterior.

En resumen, en el instante de tiempo t el algoritmo SIS actualiza la secuencia de estados que constituye la i -ésima muestra, $\mathbf{x}_{0:t-1}^{(i)}$, tomando una muestra, $\mathbf{x}_t^{(i)}$, de la distribución marginal $q(\mathbf{x}_t|\mathbf{x}_{0:t-1}, \mathbf{y}_{1:t})$, y calcula el peso de la secuencia $\mathbf{x}_{0:t}^{(i)}$ resultante utilizando la ecuación (2.26). El Pseudocódigo 1 muestra el esquema básico del algoritmo SIS.

Es importante tener presente que SIS no es más que un caso especial de IS en el que imponemos una condición a la función tentativa (ecuación (2.23)), para de ese modo permitir la actualización recursiva de las muestras y los pesos.

2.4.3. El problema de la degeneración de los pesos y el remuestreo

El mayor problema de la metodología SIS es que al cabo de unos pocos instantes de tiempo (o iteraciones del algoritmo) toda la probabilidad se concentra en unas pocas muestras y las restantes tienen un peso muy próximo a cero. Se conoce como el problema de la degeneración de los pesos. Es debido a que la varianza de los pesos de las partículas aumenta (estocásticamente)

Pseudocódigo 1 SIS

muestrear $\mathbf{x}_0^{(i)} \sim p(\mathbf{x}_0)$, $i = 1, \dots, M$
 asignar $w_0^{(i)} \leftarrow 1/M$, $i = 1, \dots, M$
para cada instante de tiempo $t > 0$
 muestrear $\mathbf{x}_t^{(i)} \sim q(\mathbf{x}_t | \mathbf{x}_{0:t-1}, \mathbf{y}_{1:t})$, $i = 1, \dots, M$
 calcular el peso, $\tilde{w}_t^{(i)}$, $i = 1, \dots, M$, asociado a la muestra $\mathbf{x}_{0:t}^{(i)}$ mediante la ecuación (2.26)
 normalizar el peso anterior para obtener $w_t^{(i)} \leftarrow \frac{\tilde{w}_t^{(i)}}{\sum_{i=1}^M \tilde{w}_t^{(i)}}$

con el tiempo [45]. Para el caso de muestreo exacto, en el que la función tentativa es la propia f.d.p. a aproximar, $q(\mathbf{x}_{0:t} | \mathbf{y}_{1:t}) = p(\mathbf{x}_{0:t} | \mathbf{y}_{1:t})$, se tiene que la media de los pesos es [46]

$$\mathbb{E}_{q(\mathbf{x}_{0:t} | \mathbf{y}_{1:t})} [w(\mathbf{x}_{0:t})] = \mathbb{E}_{q(\mathbf{x}_{0:t} | \mathbf{y}_{1:t})} \left[\frac{p(\mathbf{x}_{0:t} | \mathbf{y}_{1:t})}{q(\mathbf{x}_{0:t} | \mathbf{y}_{1:t})} \right] = 1 \quad (2.27)$$

y su varianza

$$\text{Var}_{q(\mathbf{x}_{0:t} | \mathbf{y}_{1:t})} [w(\mathbf{x}_{0:t})] = \mathbb{E}_{q(\mathbf{x}_{0:t} | \mathbf{y}_{1:t})} \left[\left(\frac{p(\mathbf{x}_{0:t} | \mathbf{y}_{1:t})}{q(\mathbf{x}_{0:t} | \mathbf{y}_{1:t})} - \mathbb{E}_{q(\mathbf{x}_{0:t} | \mathbf{y}_{1:t})} \left[\frac{p(\mathbf{x}_{0:t} | \mathbf{y}_{1:t})}{q(\mathbf{x}_{0:t} | \mathbf{y}_{1:t})} \right] \right)^2 \right] = 0. \quad (2.28)$$

Esta es la situación ideal y a la que queremos aproximarnos.

Como se explicó en la Sección 2.3, una muestra con un peso muy pequeño apenas contribuye a la aproximación de la f.d.p. objetivo. Entonces, cuando ocurre el problema de la degeneración de los pesos se está malgastando esfuerzo computacional en propagar¹ y actualizar trayectorias cuya relevancia a la hora de construir la representación de la distribución de probabilidad es mínima. A efectos prácticos, esto quiere decir que el número de partículas que se tienen para aproximar $p(\mathbf{x}_{0:t} | \mathbf{y}_{1:t})$, lo que se conoce como el número efectivo de partículas, es mucho menor que el tamaño muestral M . Una medida típica de la degeneración de los pesos en el algoritmo SIS es el tamaño efectivo de la muestra, M_{ef} , que se define como [45]

$$M_{ef} = \frac{M}{1 + \text{Var}(\tilde{w}_t^{(i)})} \quad (2.29)$$

donde $\tilde{w}_t^{(i)}$ son los pesos sin normalizar. Puesto que la expresión anterior no se puede calcular exactamente, una buena estimación de M_{ef} es [45]

$$\hat{M}_{ef} = \frac{1}{\sum_{i=1}^M (w_t^{(i)})^2} \quad (2.30)$$

donde $w_t^{(i)}$ son los pesos normalizados. Obsérvese que en el mejor de los casos, cuando todas las partículas tienen peso $1/M$, se verifica que $\hat{M}_{ef} = M$.

¹Por propagar una trayectoria se entiende añadir un nuevo elemento al final de la secuencia que la define.

El problema de la degeneración de los pesos está siempre presente en cualquier algoritmo SIS. Podrían pensarse como posibles soluciones utilizar un número muy grande de partículas, o elegir una función tentativa que muestree en las zonas del espacio de estados con mayor densidad de probabilidad *a posteriori*. Esto último equivale a elegir una “buena” función tentativa, lo cual no suele ser tarea fácil. En cualquier caso, ambas soluciones no hacen sino retrasar el problema.

Una tercera opción para resolver el problema es llevar a cabo pasos de remuestreo (en inglés, *resampling*). Debido a su demostrada eficacia, esta técnica se ha convertido en un paso imprescindible en cualquier algoritmo SIS [45, 47].

El remuestreo es una técnica aleatoria que elimina las partículas con pesos pequeños (que apenas van a contribuir a la aproximación por Monte Carlo de la distribución de probabilidad objetivo) al tiempo que replica aquellas que tienen pesos grandes. Nótese que de esta manera la aproximación se centra en las regiones del espacio de estados más prometedoras, es decir, con una mayor probabilidad *a posteriori*.

Claramente existen muchas formas de llevar a cabo lo arriba expuesto, constituyendo cada una de ellas un método de remuestreo diferente. La mayor parte de ellas se basan en tomar M muestras i.i.d. de la f.d.p. $\hat{p}_M(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})$, que resulta de aproximar $p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})$ mediante IS. Esto implica que tras el remuestreo todas las muestras tienen el mismo peso y, por tanto, se consideran igualmente “buenas”. Podemos pensar en el remuestreo como en una rutina de caja negra que recibe como entrada un vector con los pesos de las partículas, $\mathbf{w} = (w^{(1)}, \dots, w^{(M)})$, y devuelve como salida un vector $\mathbf{M} = (M_1, \dots, M_M)$, en el que M_i representa el número de replicas de la muestra i que conservaremos después del remuestreo. Si $M_i = 0$, entonces la i -ésima partícula es descartada. Es habitual imponer la condición $M = \sum_{i=1}^M M_i$, es decir, el número de muestras antes y después del remuestreo es el mismo. La Figura 2.5 ilustra lo que típicamente ocurre en un paso de remuestreo: las partículas con un peso significativo dan lugar a una o varias réplicas de sí mismas, mientras que aquellas con un peso próximo a cero son descartadas. Así, en el ejemplo de la figura, la primera muestra, $\mathbf{x}_{0:t}^{(1)}$, es descartada, mientras que la tercera se ve replicada tres veces, es decir, en este caso $M_1 = 0$ y $M_3 = 3$. Obsérvese que las copias de una muestra se sitúan justo encima de ella para indicar que representan el mismo punto del espacio muestral.

Debido al problema de la degeneración de los pesos, el remuestreo es un paso algorítmico imprescindible en cualquier algoritmo de filtrado de partículas. Sin embargo, introduce algunos problemas, tanto prácticos como teóricos. Entre éstos últimos se encuentra el hecho de que replicar las partículas con pesos altos al tiempo que se descartan aquellas que tienen uno pequeño da lugar a una pérdida de diversidad [24]. Además, el remuestreo incrementa la varianza en la estimación de la esperanza de una función arbitraria de $\mathbf{x}_{0:t}$ respecto de la f.d.p. que se está aproximando por SIS (aunque normalmente reduce la de estimaciones futuras [48, 49]).

Desde un punto de vista práctico, el remuestreo es una operación computacionalmente costosa y plantea un reto a la hora de paralelizar cualquier algoritmo de filtrado de

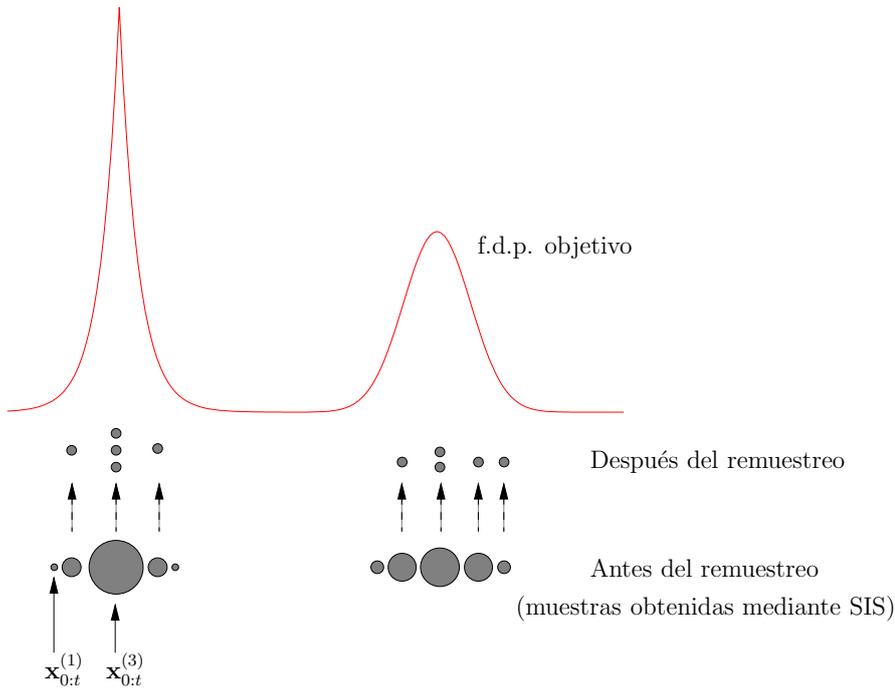


Figura 2.5: Descripción esquemática de una operación de remuestreo.

partículas que lo incluya. Cada trayectoria se propaga por medio de SIS y actualiza su peso independientemente de las demás, lo cual permite que todas lleven a cabo dichas tareas al mismo tiempo. Pero el remuestreo implica comparar unas partículas con otras para saber cuáles son las que interesa replicar y, por tanto, no es una operación que pueda paralelizarse de una forma sencilla. Trabajos recientes [50] han conseguido paliar en gran medida este problema que afecta a todos los métodos SMC.

Algunos algoritmos de filtrado de partículas, como el *bootstrap filter* [24], remuestrean en cada instante de tiempo. Otros lo hacen a intervalos regulares. Sin embargo, a causa de los problemas arriba mencionados, el remuestreo es una operación a evitar y la mayor parte de los métodos secuenciales de Monte Carlo utilizan algún tipo criterio para decidir en cada instante de tiempo si el problema de la degeneración de los pesos es lo bastante grave como para que se haga necesario un remuestreo. Es frecuente basar este criterio en el tamaño efectivo de la muestra, de tal manera que sólo se remuestrea cuando el ratio de remuestreo

$$\gamma = \frac{\hat{M}_{ef}}{M} \quad (2.31)$$

está por debajo de un cierto umbral. El rendimiento del algoritmo depende en parte de este parámetro y, si bien su elección no es tarea fácil, un umbral próximo a 1 asegura que la mayor parte de las partículas tienen un peso significativo.

Algoritmos de remuestreo clásicos

En la literatura se pueden encontrar varios de esquemas de remuestreo. El objetivo último de todo esquema de remuestreo es minimizar la varianza de los pesos, de tal manera que todas las partículas contribuyan en cierta medida a la aproximación por Monte Carlo de la distribución de interés.

Los métodos de remuestreo que se presentan a continuación se dice que son insesgados en el sentido de que $E[M_i] = Mw^{(i)}$, y se diferencian en el valor de $\text{Var}[M_i]$ y en la complejidad que llevan asociada.

El método de remuestreo más conocido es el multinomial [51]. Se basa en tomar una muestra de $\mathbf{M} = (M_1, \dots, M_M)$ asumiendo que es una v.a. multinomial de parámetros M y $\mathbf{w} = (w^{(1)}, \dots, w^{(M)})$. Eso nos da el número de replicas de cada muestra ($M_i, i = 1, \dots, M$), todas ellas con peso $1/M$, que debemos generar.

Una manera alternativa de ver este método de remuestreo es la siguiente. Si

$$\hat{p}_M(\mathbf{x}_{0:t}|\mathbf{y}_{1:t}) = \sum_{i=1}^M w^{(i)} \delta(\mathbf{x}_{0:t} - \mathbf{x}_{0:t}^{(i)}) \quad (2.32)$$

es una estimación por SIS de la f.d.p. *a posteriori* del estado, el remuestreo multinomial obtiene un nuevo conjunto de muestras, $\{\mathbf{x}_{0:t}^{(j(1))}, \dots, \mathbf{x}_{0:t}^{(j(M))}\}$ con $j(i) \in \{1, \dots, M\}$, mediante muestreo exacto de dicha estimación. La nueva estimación de la f.d.p. sería entonces:

$$\hat{p}'_M(\mathbf{x}_{0:t}|\mathbf{y}_{1:t}) = \frac{1}{M} \sum_{i=1}^M \delta(\mathbf{x}_{0:t} - \mathbf{x}_{0:t}^{(j(i))}). \quad (2.33)$$

El Pseudocódigo 2 describe la implementación.

Pseudocódigo 2 Algoritmo de remuestreo multinomial

dado un conjunto $\{\mathbf{x}_{0:t}^{(i)}, w_t^{(i)}\}_{i=1}^M$

- 1: muestrear M veces el índice $j \in \{1, \dots, M\}$ de acuerdo a la distribución discreta $\text{Prob}(j = l) = w_t^{(l)}, l = 1, \dots, M$ para generar $\{j^{(i)}\}_{i=1}^M$
 - 2: asignar $\mathbf{x}_{0:t}^{(i)} \leftarrow \mathbf{x}_{0:t}^{(j^{(i)})}, w_t^{(i)} \leftarrow 1/M, i = 1, \dots, M$
-

El primer paso del Pseudocódigo 2 es muy sencillo a partir de la inversa de la función de distribución asociada al conjunto de pesos normalizados, $w^{(i)}, i = 1, \dots, M$. Dicha función se define como $F_w^{-1}(u) = i$ para todo $u \in \left(\sum_{j=1}^{i-1} w^{(j)}, \sum_{j=1}^i w^{(j)}\right]$. Entonces basta con tomar M muestras de una distribución uniforme en el intervalo $[0, 1)$ y aplicar sobre cada una de ellas la función $F_w^{-1}(\cdot)$ para obtener los índices de las partículas a replicar. La Figura 2.6 ilustra este proceso por medio de un ejemplo en el que $M = 5$. De las 5 muestras de la distribución uniforme generadas, vemos que la primera se encuentra dentro del escalón asociado al primer

peso, lo cual dará lugar a una réplica de la primera partícula, $\mathbf{x}_{0:t}^{(1)}$. Las dos muestras siguientes están dentro del escalón debido al segundo peso, por lo que la segunda partícula, $\mathbf{x}_{0:t}^{(2)}$, se va a replicar dos veces. El proceso continúa de esta forma hasta agotar todas las muestras de la distribución uniforme. Obsérvese que ninguna muestra se sitúa en el escalón a que da lugar la quinta partícula, de tal manera que ésta será descartada (a pesar de que tiene un peso bastante significativo).

Muestras de una
U(0,1)

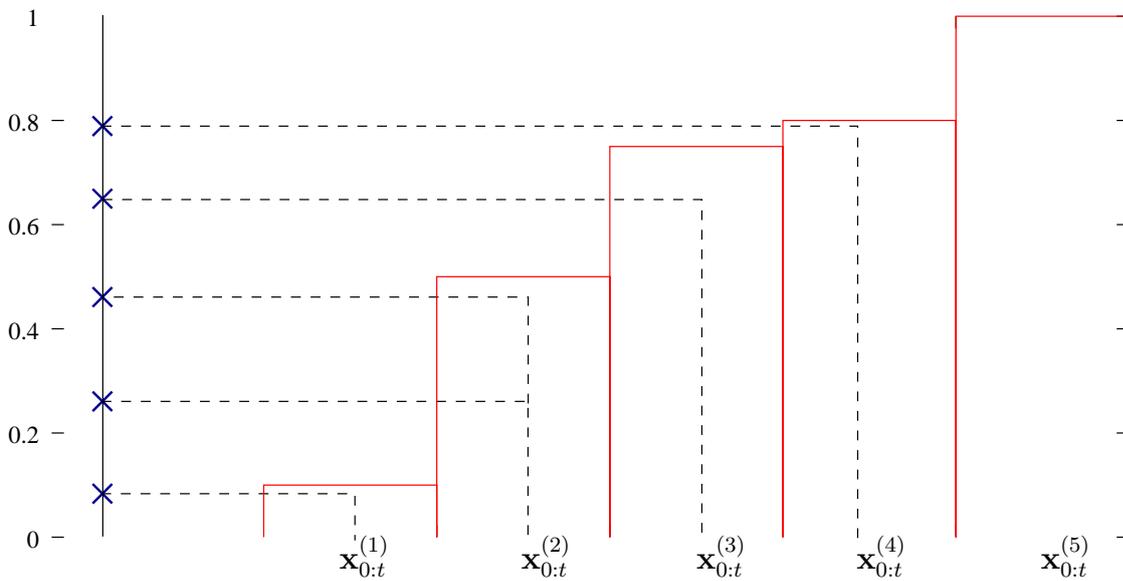


Figura 2.6: Obtención de los índices de las muestras a replicar en el remuestreo multinomial.

En el algoritmo de remuestreo multinomial la varianza del número de réplicas de la i -ésima partícula después del remuestreo es $\text{Var}[M_i] = Mw^{(i)}(1 - w^{(i)})$. El algoritmo de remuestreo residual [51] reduce esta varianza. Se caracteriza por introducir una componente determinista a la hora de obtener los distintos M_i , que ahora vamos a descomponer como $M_i = M_i^d + M_i^a$. Calcularemos en primer lugar la componente determinista de cada uno de los M_i como $M_i^d = \lfloor Mw^{(i)} \rfloor$, donde $\lfloor \alpha \rfloor$ es la parte entera de α . Esto supone $\sum_{i=1}^M M_i^d$ nuevas partículas obtenidas como réplicas de partículas viejas, y faltan $\bar{M} = M - \sum_{i=1}^M M_i^d$ partículas por elegir. Para ello vamos a aplicar el remuestreo multinomial sobre un nuevo conjunto de pesos calculados como $w'^{(i)} = \frac{1}{\bar{M}} (Mw^{(i)} - M_i^d)$, $i = 1, \dots, M$. El Pseudocódigo 3 describe la implementación del algoritmo. La varianza que se tiene para este esquema de remuestreo es $\text{Var}[M_i] = \bar{M}w'^{(i)}(1 - w'^{(i)})$, que es más pequeña que la que teníamos para el caso del remuestreo multinomial [52].

La menor varianza que se puede conseguir con un algoritmo de remuestreo insesgado la

Pseudocódigo 3 Algoritmo de remuestreo residual

- dado** un conjunto $\{\mathbf{x}_{0:t}^{(i)}, w_t^{(i)}\}_{i=1}^M$
- 1: $M_i^d \leftarrow \lfloor Mw_t^{(i)} \rfloor, i = 1, \dots, M$
 - 2: replicar M_i^d veces la partícula $\mathbf{x}_{0:t}^{(i)}, i = 1, \dots, M$, asignándole a cada réplica un peso $1/M$
 - 3: calcular el número de partículas que faltan por replicar: $\bar{M} \leftarrow M - \sum_{i=1}^M M_i^d$
 - 4: obtener los pesos modificados $w_t'^{(i)} \leftarrow \frac{1}{M} (Mw_t^{(i)} - M_i^d), i = 1, \dots, M$
 - 5: elegir las \bar{M} muestras que faltan por replicar aplicando el remuestreo multinomial sobre el conjunto $\{\mathbf{x}_{0:t}^{(i)}, w_t'^{(i)}\}_{i=1}^M$
-

alcanza el algoritmo de remuestreo estratificado [53] y es

$$\text{Var}[M_i] = \{Mw^{(i)}\} (1 - \{Mw^{(i)}\}) \tag{2.34}$$

donde $\{\alpha\} = \alpha - \lfloor \alpha \rfloor$.

El remuestreo estratificado consiste en dividir el intervalo $(0, 1]$ en M subintervalos disjuntos $(0, 1] = (0, \frac{1}{M}] \cup \dots \cup (\frac{M-1}{M}, 1]$, dentro de cada uno de los cuales se va a generar una muestra uniforme independiente de todas las demás. Es decir, la i -ésima muestra se obtiene de acuerdo a la distribución $U(\frac{i-1}{M}, \frac{i}{M})$. Una vez obtenidas las muestras, aplicaríamos el mismo procedimiento de inversión de la función de distribución de los pesos que veíamos para el caso del remuestreo multinomial. La Figura 2.7 utiliza los mismos pesos que el ejemplo de la Figura 2.6 para mostrar las diferencias entre ambos algoritmos de remuestreo.

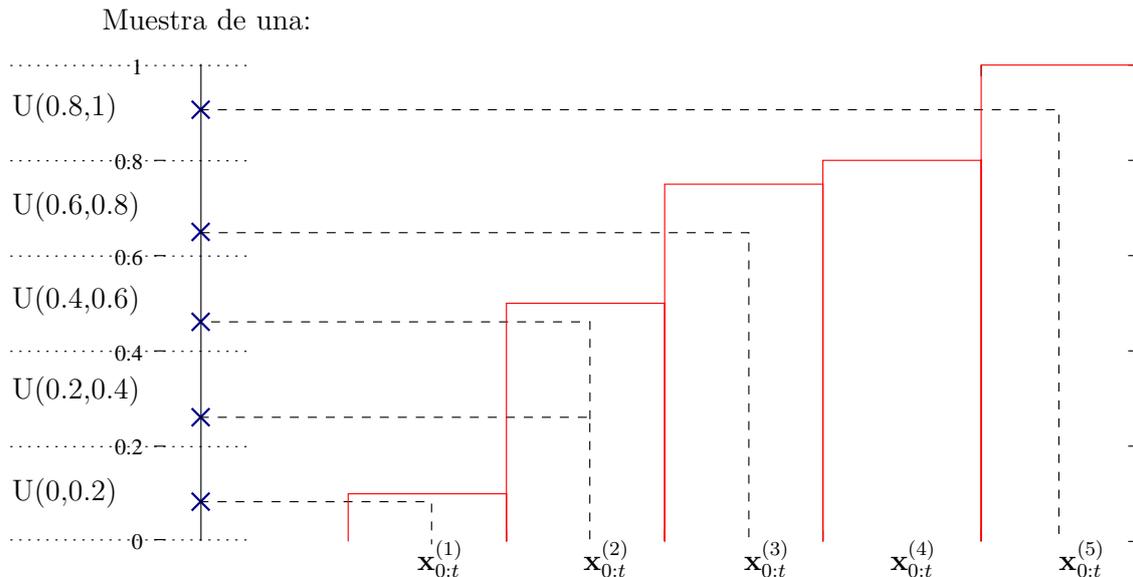


Figura 2.7: Obtención de los índices de las muestras a replicar en el remuestreo estratificado.

Se observa como, en este caso, exactamente una muestra es generada dentro de cada intervalo, de tal manera que una de ellas se encuentra ahora dentro del intervalo $(0, 8, 1]$, por lo que va a ser la partícula quinta la que se replique una vez y la cuarta la que se descarte.

Un algoritmo de remuestreo relacionado con el estratificado es el sistemático. Básicamente consiste en generar una muestra aleatoria uniforme en el intervalo $(0, \frac{1}{M}]$ y obtener las de los los restantes intervalos de manera determinista. Generaríamos entonces la primera muestra como $u_1 \sim U(0, \frac{1}{M})$ y, a partir de ella, las muestras u_2, \dots, u_M como $u_i = u_1 + \frac{i-1}{M}, i = 2, \dots, M$. La Figura 2.8 ilustra el modo de operar del algoritmo de remuestreo sistemático sobre el ejemplo anterior ($M = 5$). Únicamente la primera muestra es generada aleatoriamente (en el intervalo $(0, 0,2)$), y las demás se obtienen simplemente sumando $1/M$ a la anterior. Una vez tenemos muestras u_1, u_2, \dots, u_M de una v.a. uniforme², los índices de las partículas a replicar se obtendrían, al igual que en los algoritmos anteriores, invirtiendo la función de distribución de los pesos normalizados.

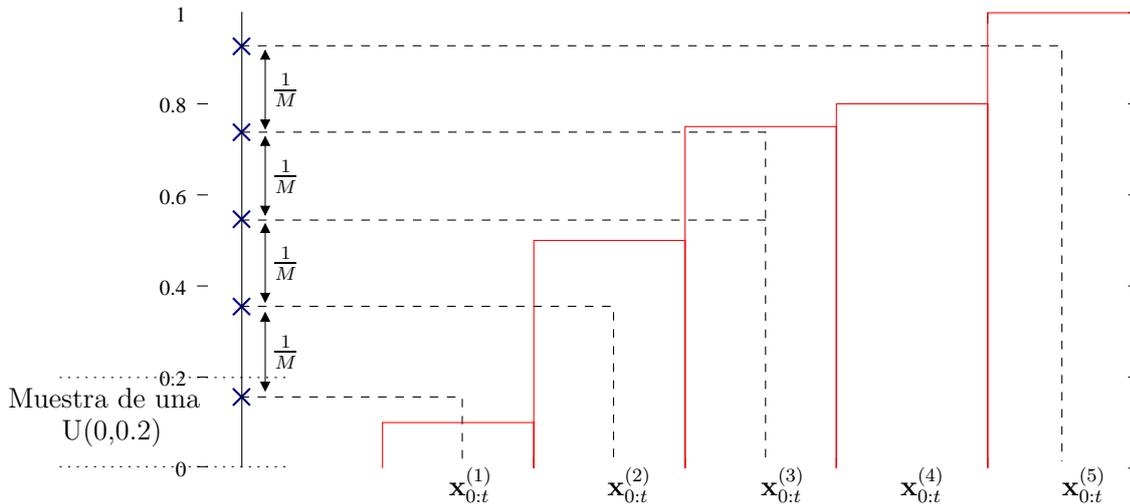


Figura 2.8: Obtención de los índices de las muestras a replicar en el remuestreo sistemático.

El remuestreo sistemático es muy simple y por ello es el método de remuestreo elegido en muchos algoritmos de filtrado de partículas. Sin embargo, al contrario que los algoritmos de remuestreo residual y estratificado, su rendimiento no siempre es superior al del remuestreo multinomial [51].

A pesar de que todos los algoritmos de remuestreo descritos en esta sección son insesgados en el sentido de que M_i tiende a $Mw^{(i)}$, estudios teóricos recientes [54, 55] demuestran que ésta no es una condición necesaria para que un algoritmo de filtrado de partículas funcione correctamente.

²Obsérvese que las muestras u_2, \dots, u_M son de naturaleza claramente pseudoaleatoria.

2.4.4. Convergencia

Debido a la normalización de los pesos y a los pasos de remuestreo, las partículas (una muestra y su peso asociado) que aproximan la distribución de probabilidad de interés en un método secuencial de Monte Carlo interactúan entre sí, por lo que son estadísticamente dependientes. Así pues, los resultados clásicos de convergencia para métodos de Monte Carlo basados en la hipótesis de independencia estadística no son aplicables a los filtros de partículas y la convergencia de éstos ha de ser estudiada de forma específica. Existen algunos trabajos que han abordado el problema de la convergencia de los métodos SMC. En esta sección únicamente enumeraremos los resultados más relevantes que se han derivado de ellos, así como algunos de los problemas que todavía están por resolver. Una aproximación más formal al problema se puede encontrar en [56].

Sea $p(x)$ una f.d.p. de interés, y sea $p_M(x)$ la estimación de dicha densidad obtenida utilizando M partículas. Diremos que la aproximación $p_M(x)$ converge a $p(x)$ o

$$\lim_{M \rightarrow \infty} p_M(x) = p(x) \quad (2.35)$$

si

$$\lim_{M \rightarrow \infty} \mathbb{E}_{p_M(x)} [f(x)] = \mathbb{E}_{p(x)} [f(x)] \quad (2.36)$$

para una función de x arbitraria, f .

A día de hoy todavía existen algunas dificultades teóricas para garantizar la convergencia de los métodos secuenciales de Monte Carlo en cualquier circunstancia. Si, por ejemplo, hay algún parámetro fijo aleatorio en el modelo dinámico podría ocurrir que éste no fuera ergódico, en cuyo caso la convergencia de un filtro de partículas no está asegurada. Sin embargo, imponiendo algunas restricciones es posible establecer la convergencia casi segura (con probabilidad 1) de los filtros de partículas. En términos de error cuadrático medio (MSE, del inglés *Mean Square Error*), la tasa de convergencia de un filtro de partículas es directamente proporcional al número de partículas empleadas por el mismo, lo cual es perfectamente coherente con lo que veíamos en la Sección 2.2 para métodos de Monte Carlo. Además, la tasa de convergencia es, en principio, independiente de la dimensión del espacio de estados. Sin embargo, para asegurar una determinada precisión en el MSE, el número de partículas, M , depende indirectamente de la dimensión del espacio de estados.

Aunque es mucho lo que se ha avanzado en los últimos años en este campo, todavía se están buscando resultados que garanticen la convergencia de los filtros de partículas relajando algunas de las condiciones más severas que suelen imponerse. Por ejemplo, algunos resultados parten de la hipótesis de que la función $f(x)$ tiene que estar acotada, lo cual es una limitación importante. Otros resultados imponen ciertas restricciones en el modelo dinámico que no siempre es posible garantizar en problemas reales.

2.4.5. Resumen

El objetivo de los métodos secuenciales de Monte Carlo es la estimación recursiva de distribuciones de probabilidad de interés. Para lograr dicho objetivo hacen uso de las dos herramientas que acabamos de explicar: SIS para obtener recursivamente muestras de la f.d.p. a aproximar, y pasos de remuestreo para evitar el problema de la degeneración de los pesos. Es habitual referirse a la combinación de SIS y remuestreo como *Sequential Importance Sampling with Resampling* o SIR. La Figura 2.9 ilustra el funcionamiento de un filtro de partículas de tamaño $M = 10$ que trata de estimar la densidad $p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})$.

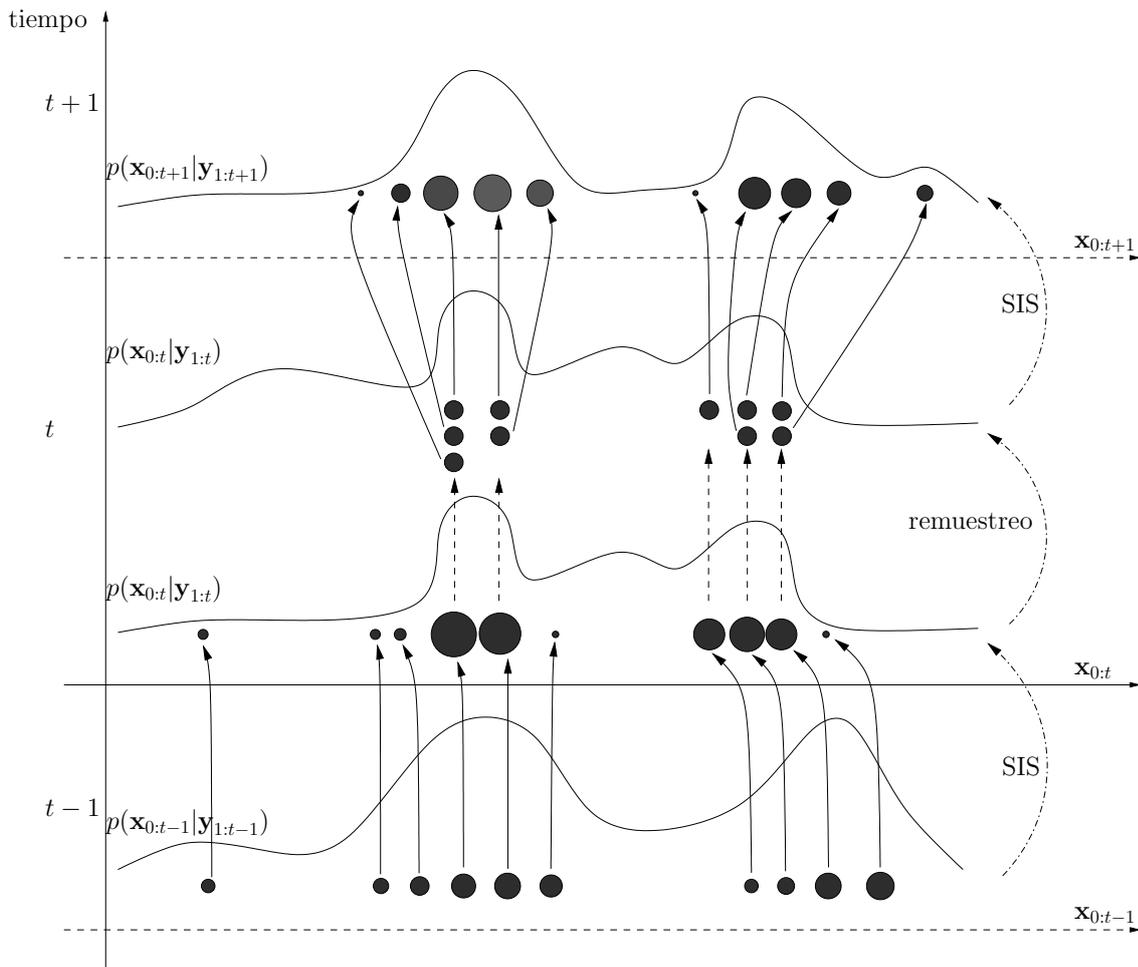


Figura 2.9: Operación de un filtro de partículas.

Las muestras están representadas por círculos cuya área indica la magnitud del peso que tienen asociado, de tal manera que muestras que se encuentran en regiones del espacio de estados más probables aparecen como círculos más grandes. Partimos en el instante de tiempo $t - 1$ de un conjunto de muestras que constituyen una representación de la densidad $p(\mathbf{x}_{0:t-1}|\mathbf{y}_{1:t-1})$. En el instante t se utiliza SIS para obtener una muestra de $p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})$ a partir

de una muestra de $p(\mathbf{x}_{0:t-1} | \mathbf{y}_{1:t-1})$. Para ello es necesario tomar una muestra, $\mathbf{x}_t^{(i)}$, de la función tentativa $q(\mathbf{x}_t | \mathbf{x}_{0:t-1}, \mathbf{y}_{1:t})$, y calcular el peso de la nueva muestra, $\mathbf{x}_{0:t}^{(i)}$, a partir del peso que tenía la muestra del instante anterior que la originó, $\mathbf{x}_{0:t-1}^{(i)}$. En la figura, las flechas continuas representan la actualización de las diferentes trayectorias o partículas (el paso SIS).

Nótese que la f.d.p. objetivo en el instante $t - 1$ es diferente a la f.d.p. objetivo en el instante t y, de hecho, tienen espacios de estados diferentes (de ahí que en la figura se haya dibujado un eje de abscisas distinto para cada instante de tiempo). Al comienzo del instante t se observa el problema de la degeneración de los pesos: sólo la mitad de las muestras que se utilizan para representar la f.d.p. objetivo tienen un peso significativo. Se hace necesario un paso de remuestreo que replique las muestras que acumulan una cierta probabilidad y descarte aquellas cuyo peso es despreciable y que, por tanto, apenas contribuyen a la representación de $p(\mathbf{x}_{0:t} | \mathbf{y}_{1:t})$. Después del remuestreo hay 10 muestras con idéntico peso situadas en las regiones más probables del espacio de estados de la densidad objetivo. La siguiente iteración del algoritmo las propagaría (por medio de SIS nuevamente) al instante de tiempo $t + 1$, y el proceso continuaría de esta manera hasta haber procesado toda la secuencia de observaciones.

2.5. Conclusiones

En este capítulo se ha hecho una breve introducción a los métodos secuenciales de Monte Carlo, cuyo objetivo fundamental es la aproximación recursiva de una secuencia de distribuciones de probabilidad mediante muestras y pesos asociados. Comenzamos explicando como representar una densidad de probabilidad de interés mediante muestras obtenidas directamente de la misma (muestreo exacto). Cuando esto no es posible, o resulta demasiado costoso, podemos generar muestras a partir de otra f.d.p., la función tentativa, y valernos del principio de IS para asignarles un peso acorde con la densidad de probabilidad que tendría nuestra densidad objetivo en un entorno de dichas muestras. Los métodos SMC, también conocidos como filtros de partículas, están pensados para la estimación secuencial de la f.d.p. de la señal no observada en un sistema dinámico en formato de espacio de estados. Sin embargo, el principio de IS, en su formulación más sencilla, no es adecuado para llevar a cabo una estimación recursiva de una secuencia de distribuciones de probabilidad. SIS es una modificación del principio de muestreo enfatizado que resuelve este problema mediante una adecuada elección de la función tentativa.

La estimación recursiva de distribuciones de probabilidad de interés mediante SIS es la esencia de casi todos los algoritmos de filtrado de partículas, pero la metodología no estaría completa sin el paso algorítmico conocido como remuestreo. Sin el remuestreo la mayor parte de las partículas tendrían un peso muy próximo a cero al cabo de unos pocos instantes de tiempo (lo que se conoce como el problema de la degeneración de los pesos), con lo que su contribución a la aproximación de la densidad objetivo sería prácticamente nula.

Capítulo 3

Modelo de señal

Un sistema de comunicaciones con diversidad, tanto en el transmisor como en el receptor, se denomina MIMO (del inglés *Multiple Input, Multiple Output*). En este capítulo se presenta un modelo matemático que explica tanto el proceso de transmisión a través de un canal MIMO como la evolución de éste con el tiempo. En lo sucesivo supondremos siempre que la diversidad se consigue mediante la utilización de varias antenas en los dos extremos del canal, aunque existen otros esquemas de transmisión que dan lugar a modelos de señal muy similares al presentado aquí (e.g., detección multiusuario en sistemas CDMA).

3.1. Modelado de sistemas de comunicaciones MIMO

La Figura 3.1 representa el diagrama de bloques de un sistema MIMO con N_t antenas transmisoras (entradas) y N_r antenas receptoras (salidas). Si las antenas están lo suficientemente separadas, se puede considerar que el canal de comunicaciones que existe entre cada par antena emisora-antena receptora es diferente. Por tanto, podemos hablar de $N_r \times N_t$ canales independientes, siendo $\mathbf{h}_{i,j,t}$ la respuesta al impulso¹ que caracteriza al canal existente entre las antenas i y j en el instante de tiempo t . Así, $\mathbf{h}_{i,j,t}$ determina de qué manera los símbolos transmitidos por la i -ésima antena emisora en el instante t y anteriores (considerando un canal selectivo en frecuencia) contribuyen al valor observado en la j -ésima antena receptora en dicho instante, $y_t(j)$.

En última instancia, el modelo equivalente en banda base y tiempo discreto de las observaciones de un sistema MIMO como el mostrado en la Figura 3.1 es [57]

$$\mathbf{y}_t = \sum_{l=0}^{m-1} \mathbf{H}_t(l) \mathbf{s}_{t-l} + \mathbf{g}_t, \quad t = 0, 1, \dots \quad (3.1)$$

donde

¹Nótese que estamos considerando t como un índice temporal discreto.

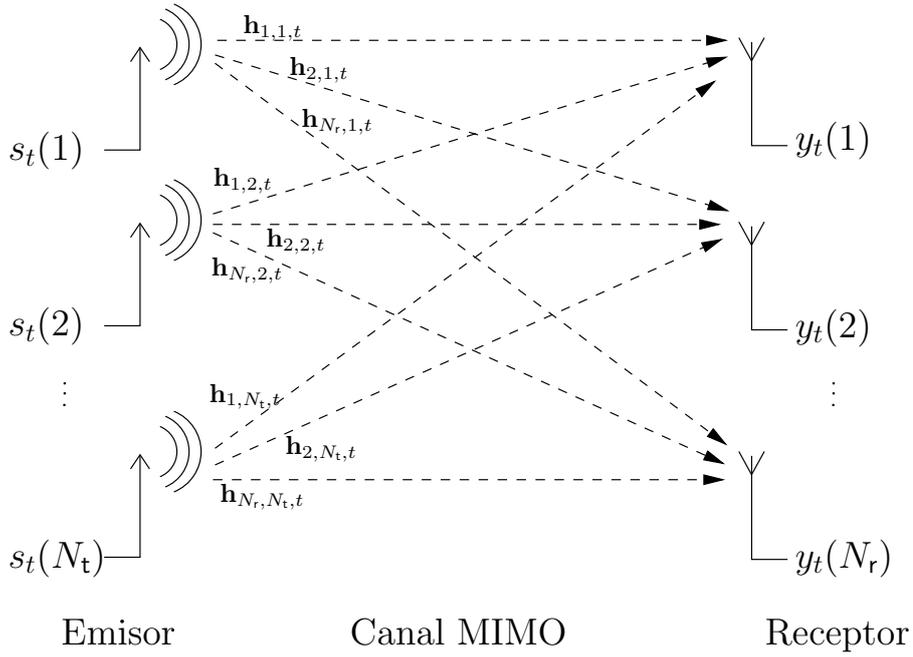


Figura 3.1: Esquema de un sistema de comunicaciones MIMO.

- $\mathbf{y}_t = [y_t(1), y_t(1), \dots, y_t(N_r)]^\top$ es un vector $N_r \times 1$ con las observaciones recogidas en el instante t .
- $\mathbf{H}_t(l), l = 0, \dots, m - 1$ es la respuesta al impulso del canal MIMO de dimensiones $N_r \times N_t$. La matriz $\mathbf{H}_t(l)$ contiene los coeficientes asociados al l -ésimo retardo del canal, y el número total de retardos, m , es lo que se denomina memoria u orden del canal.
- $\mathbf{s}_t = [s_t(1), \dots, s_t(N_t)]^\top$ es un vector columna de tamaño N_t que contiene los símbolos transmitidos en el instante t , y que se modelan como variables aleatorias discretas uniformes en el alfabeto \mathcal{S} . Vamos a suponer que tienen media cero, $\mathbb{E}[\mathbf{s}_t] = \mathbf{0}$, y que su matriz de covarianza es $\mathbb{E}[\mathbf{s}_t \mathbf{s}_t^\top] = \mathbf{I}_{N_t} \sigma_s^2$ (\mathbf{I}_{N_t} es la matriz identidad de orden N_t).
- \mathbf{g}_t es un proceso estocástico de ruido aditivo blanco gaussiano (AWGN (ruido aditivo blanco gaussiano)) de media cero y matriz de covarianza $\sigma_g^2 \mathbf{I}_{N_r}$.

Nótese que para un canal plano (sin memoria), en el cual sólo los símbolos transmitidos en el instante de tiempo t contribuyen a la t -ésima observación, se tiene $m = 1$. Es decir, aunque con frecuencia nos referiremos a m como la memoria del canal, dicho parámetro no representa el número de vectores de símbolos pasados que intervienen en la observación actual, sino el número total de vectores de símbolos que afectan a ésta.

Es importante resaltar, por otro lado, que en esta tesis se consideran exclusivamente constelaciones de símbolos reales y, asimismo, los procesos estocásticos del canal y del ruido son también reales. La extensión a alfabetos complejos, más generales, es conceptualmente

trivial, pero complica la notación utilizada en los algoritmos SMC, que constituyen el principal objeto de este trabajo.

La ecuación (3.1) se puede reescribir de una manera más compacta como

$$\mathbf{y}_t = \mathbf{H}_t \mathbf{s}_{t-m+1} + \mathbf{g}_t, \quad (3.2)$$

donde $\mathbf{s}_{t-m+1} = [\mathbf{s}_{t-m+1}^\top \cdots \mathbf{s}_t^\top]^\top$ es un vector $N_t m \times 1$ que contiene todos los símbolos implicados en la t -ésima observación y $\mathbf{H}_t = [\mathbf{H}_t(m-1) \cdots \mathbf{H}_t(0)]$ es la matriz de canal conjunta de tamaño $N_r \times N_t m$.

La notación \mathbf{b}_{t_1, t_2} , donde $\{\mathbf{b}_t\}$ es un proceso estocástico y t_1 y t_2 son dos instantes de tiempo discreto tales que $t_1 < t_2$, será habitual a lo largo de este trabajo para referirnos al vector (columna) que resulta de apilar verticalmente los vectores $\mathbf{b}_{t_1}, \mathbf{b}_{t_1+1}, \dots, \mathbf{b}_{t_2}$, es decir

$$\mathbf{b}_{t_1, t_2} = \begin{bmatrix} \mathbf{b}_{t_1} \\ \mathbf{b}_{t_1+1} \\ \vdots \\ \mathbf{b}_{t_2} \end{bmatrix}. \quad (3.3)$$

La matriz $\mathbf{H}_t(l)$ representa el l -ésimo retardo de la respuesta al impulso del canal, es decir, la parte de ésta que determina la contribución a la t -ésima observación de los símbolos transmitidos en el instante $t-l$. Agrupando las m matrices que representan los diferentes retardos del canal resulta la matriz \mathbf{H}_t , que contiene toda la información del estado del canal (CSI, del inglés *Channel State Information*) en el instante t . La Figura 3.2 relaciona las respuestas al impulso que aparecen indicadas en la Figura 3.1 con la estructura de esta matriz.

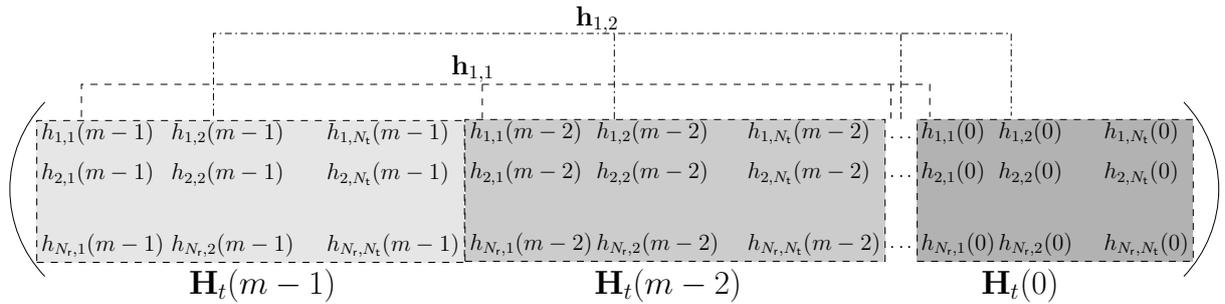


Figura 3.2: Matriz de canal conjunta \mathbf{H}_t .

Cada uno de los coeficientes de la respuesta al impulso del canal entre la i -ésima antena emisora y la j -ésima antena receptora, $\mathbf{h}_{i,j,t}$, representa un retardo diferente y, por tanto, se encuentra en una matriz $\mathbf{H}_t(\cdot)$ diferente. Así, si $h_t(k,l)$ es el coeficiente que se encuentra en la k -ésima fila y l -ésima columna de \mathbf{H}_t , la respuesta al impulso entre la i -ésima antena emisora y la j -ésima antena receptora es $\mathbf{h}_{i,j,t} = [h_t(i,j), h_t(i,j+N_t), h_t(i,j+2N_t), \dots, h_t(i,j+N_t(m-1))]$. Por simplicidad, en la figura

hemos optado por suprimir el índice temporal de los vectores que representan las respuestas al impulso entre cada par de antenas y, así, el l -ésimo elemento del vector $\mathbf{h}_{i,j,t}$ con $0 \leq l < m$, se denota por $h_{i,j}(l)$.

En cuanto a la variación del canal, éste se va a modelar como un proceso autorregresivo (AR) [58, 59],

$$\mathbf{H}_t(l) = \sum_{r=1}^R \alpha_r \mathbf{H}_{t-r}(l) + \mathbf{V}_t(l) \quad 0 \leq l \leq m-1, \quad (3.4)$$

donde R es el orden del proceso AR, α_r con $r = 1, \dots, R$ son sus coeficientes y $\mathbf{V}_t(l)$ es una matriz $N_r \times N_t$ de variables aleatorias gaussianas i.i.d. con media cero y varianza σ_v^2 .

Para facilitar la derivación de los algoritmos vamos a considerar un proceso AR de orden 1, es decir, $R = 1$, de tal forma que

$$\mathbf{H}_t(l) = \alpha \mathbf{H}_{t-1}(l) + \mathbf{V}_t(l) \quad 0 \leq l \leq m-1 \quad (3.5)$$

siendo $\alpha = \alpha_1$ el único coeficiente del proceso AR². Podemos reescribir la ecuación (3.5) utilizando la matriz de canal conjunta, \mathbf{H}_t , como

$$\mathbf{H}_t = \alpha \mathbf{H}_{t-1} + \mathbf{V}_t \quad (3.6)$$

donde $\mathbf{V}_t = [\mathbf{V}_t(m-1), \mathbf{V}_t(m-2), \dots, \mathbf{V}_t(0)]$ es una matriz $N_r \times N_t m$ de variables aleatorias gaussianas i.i.d. con media cero y varianza σ_v^2 .

A pesar de que el modelo (3.5) no tiene en cuenta principios físicos de propagación de ondas, al ser la simplicidad su objetivo primordial, mediante un adecuado ajuste de los parámetros α y σ_v^2 es posible obtener una gran variedad de funciones de autocorrelación para el canal. En concreto, es una práctica muy común en la literatura [58, 59, 60, 61] aproximar el modelo de correlación de Clarke por medio de un proceso AR.

A partir de las ecuaciones (3.1) y (3.5), es posible modelar el proceso de transmisión como un sistema dinámico en formato de espacio de estados:

$$\left. \begin{array}{l} \mathbf{s}_t \sim \mathcal{U}(\mathcal{S}^{N_t}) \\ \mathbf{H}_t(l) = \alpha \mathbf{H}_{t-1}(l) + \mathbf{V}_t(l) \quad 0 \leq l \leq m-1 \end{array} \right\} \text{ecuaciones de estado} \quad (3.7)$$

$$\left. \begin{array}{l} \mathbf{y}_t = \sum_{l=0}^{m-1} \mathbf{H}_t(l) \mathbf{s}_{t-l} + \mathbf{g}_t, \quad t = 0, 1, \dots \end{array} \right\} \text{ecuación de observación.} \quad (3.8)$$

El estado en el instante de tiempo t se compone de la información del estado del canal, $\{\mathbf{H}_t(l)\}_{l=0}^{m-1}$, que suponemos desconocido, y de los vectores de símbolos $\mathbf{s}_{t-m+1:t} = \{\mathbf{s}_{t-m+1}, \dots, \mathbf{s}_t\}$. Su evolución está gobernada por las ecuaciones en (3.7): el canal varía de acuerdo a un proceso AR de orden 1 con parámetros σ_v^2 y α conocidos, y los símbolos se modelan como variables aleatorias uniformes discretas independientes en el alfabeto \mathcal{S} , esto es, $\mathbf{s}_t \sim \mathcal{U}(\mathcal{S}^{N_t})$.

²No existe ninguna dificultad conceptual para extender los métodos descritos en esta tesis a modelos AR de orden superior.

3.1.1. Modelo apilado

Cuando el canal es selectivo en frecuencia ($m > 1$) se hace necesario llevar a cabo algún tipo de suavizado para conseguir una detección fiable de los datos transmitidos. El suavizado consiste en detectar el t -ésimo vector de símbolos transmitidos considerando las observaciones $\mathbf{y}_{t:t+a} = \{\mathbf{y}_t, \dots, \mathbf{y}_{t+a}\}$, siendo $a > 0$ el factor de suavizado. Para tener en cuenta todas las observaciones a lo largo de las cuales el canal dispersa la energía de \mathbf{s}_t es útil considerar el modelo apilado

$$\mathbf{y}_{t+a}^t = \mathbf{H}_{t,a} \mathbf{s}_{t+a}^{t-m+1} + \mathbf{g}_{t+a}^t, \quad (3.9)$$

donde $\mathbf{y}_{t+a}^t = [\mathbf{y}_t^\top \cdots \mathbf{y}_{t+a}^\top]^\top$ es un vector de tamaño $N_r(a+1) \times 1$ que contiene las observaciones apiladas, $\mathbf{s}_{t-m+1}^{t+a} = [\mathbf{s}_{t-m+1}^\top \cdots \mathbf{s}_{t+a}^\top]^\top$ tiene dimensiones $N_t(m+a) \times 1$, $\mathbf{g}_{t+a}^t = [\mathbf{g}_t^\top \cdots \mathbf{g}_{t+a}^\top]^\top$ y

$$\mathbf{H}_{t,a} = \begin{bmatrix} \mathbf{H}_t(m-1)^\top & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{H}_t(m-2)^\top & \mathbf{H}_{t+1}(m-1)^\top & \cdots & \mathbf{0} \\ \vdots & \mathbf{H}_{t+1}(m-2)^\top & \ddots & \vdots \\ \mathbf{H}_t(0)^\top & \vdots & \ddots & \mathbf{H}_{t+a}(m-1)^\top \\ \vdots & \mathbf{H}_{t+1}(0)^\top & \ddots & \mathbf{H}_{t+a}(m-2)^\top \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{H}_{t+a}(0)^\top \end{bmatrix}^\top \quad (3.10)$$

es la matriz de canal apilada, de dimensiones $N_r(a+1) \times N_t(m+a)$.

Modelo apilado eliminando la interferencia causal

Al vector de observaciones apiladas, \mathbf{y}_{t+a}^t , de la ecuación (3.9) contribuyen no sólo los vectores de símbolos desconocidos, $\mathbf{s}_{t:t+a}$, sino también los vectores $\mathbf{s}_{t-m+1:t-1}$ que ya han sido detectados en instantes de tiempo anteriores. Conocidas las matrices de canal $\mathbf{H}_{t:t+a}$, podemos suprimir la interferencia intersimbólica causal (debida a símbolos pasados) que éstos introducen en las observaciones para obtener

$$\bar{\mathbf{y}}_{t+a}^t = \mathbf{y}_{t+a}^t - \check{\mathbf{H}}_{t,a} \mathbf{s}_{t-1}^{t-m+1}, \quad (3.11)$$

donde

$$\check{\mathbf{H}}_{t,a} = \begin{bmatrix} \mathbf{H}_t(m-1) & \mathbf{H}_t(m-2) & \cdots & \mathbf{H}_t(1) \\ \mathbf{0} & \mathbf{H}_{t+1}(m-1) & \cdots & \mathbf{H}_{t+1}(2) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{H}_{t+a-1}(m-1) \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix} \quad (3.12)$$

es una matriz $N_r(a+1) \times N_t(m-1)$, cuyos elementos no nulos son los coeficientes de $\mathbf{H}_{t,a}$ que multiplicaban a los vectores de símbolos $\mathbf{s}_{t-m+1:t-1}$ en (3.9).

Entonces, la ecuación que describe este nuevo vector de observaciones transformadas es

$$\bar{\mathbf{y}}_{t+a}^t = \bar{\mathbf{H}}_{t,a} \mathbf{s}_{t+a}^t + \mathbf{g}_{t+a}^t, \quad (3.13)$$

donde

$$\bar{\mathbf{H}}_{t,a} = \begin{bmatrix} \mathbf{H}_t(0) & \mathbf{0} & \cdots & \cdots & \mathbf{0} \\ \mathbf{H}_{t+1}(1) & \mathbf{H}_{t+1}(0) & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{H}_{t+a}(m-1) & \mathbf{H}_{t+a}(m-2) & \cdots & \cdots & \mathbf{H}_{t+a}(0) \end{bmatrix} \quad (3.14)$$

es una matriz $N_r(a+1) \times N_t(a+1)$ y \mathbf{s}_{t+a}^t es un vector columna $N_t(a+1) \times 1$.

Obsérvese que para obtener el vector de observaciones transformadas $\bar{\mathbf{y}}_{t+a}^t$ mediante la ecuación (3.11) es necesario conocer el canal (las matrices $\mathbf{H}_{t:t+a-1}$) y los símbolos transmitidos desde el instante $t-m+1$ hasta el $t-1$ (para construir el vector $\mathbf{s}_{t-m+1:t-1}$). Normalmente sólo dispondremos de una estimación del canal y los símbolos detectados en los instantes de tiempo anteriores, $\mathbf{s}_{t-m+1:t-1}$, no siempre serán los verdaderos, por lo que (3.13) no será en general una igualdad sino una aproximación.

3.2. Capacidad del canal MIMO

La mayor ventaja de los sistemas MIMO respecto a los sistemas tradicionales que emplean una única antena en ambos lados del enlace radio tiene que ver con el concepto de capacidad. La capacidad de un canal es la máxima velocidad de transmisión a la que es posible una comunicación fiable entre un emisor y un receptor no sujetos a ninguna restricción. A finales de los años cuarenta, Shannon demostró que para cualquier tasa de transferencia inferior a la capacidad es posible encontrar un código que alcance una probabilidad de error arbitrariamente pequeña (distinta de cero) [62]. Sin embargo, dicho código podría tener un tamaño de bloque muy largo y la complejidad del emisor y el receptor podrían ser muy elevadas.

La capacidad de un canal MIMO que no es selectivo ni en tiempo ni en frecuencia e introduce ruido gaussiano es

$$C = \log \det \left(\mathbf{I} + \frac{\text{SNR}}{N_t} \mathbf{H} \mathbf{H}^T \right) \quad (3.15)$$

donde SNR (del inglés, *Signal-to-Noise-Ratio*) es la relación entre la potencia de la señal y la potencia del ruido en unidades naturales y \mathbf{H} es la matriz de canal. Si la base del logaritmo es 2 (lo más habitual), entonces la capacidad viene dada en *bits/s/Hz*.

Cuando el canal es selectivo en frecuencia y/o variante con el tiempo, como en nuestro caso, el cálculo de la capacidad es mucho más complejo. Sin embargo, en ambos casos se cumple que la pendiente de la curva de capacidad crece linealmente con el mínimo entre el número de

antenas emisoras y el de receptoras [63]. Este hecho constituye el principal atractivo teórico de los sistemas MIMO.

3.3. Planteamiento del problema

En la Sección 3.1 se presentaron las ecuaciones que modelan la transmisión en un sistema de comunicaciones MIMO. El emisor transmite una secuencia de símbolos que el receptor debe recuperar a partir de unas observaciones que contienen a dicha secuencia perturbada por el canal más un ruido. Por tanto, y basándonos en el modelo de transmisión descrito en el presente capítulo, el objetivo último de todos los algoritmos presentados en este trabajo será la detección de los símbolos transmitidos, $\mathbf{s}_{0:t}$, dadas las observaciones $\mathbf{y}_{0:t}$. Por lo general, el canal no es conocido y debe ser estimado como paso previo a una detección coherente de los datos. Nuestros algoritmos van a partir siempre de un canal desconocido y producirán estimaciones del mismo como subproducto del proceso de detección.

Aún cuando el canal es desconocido, para estimarlo es necesario conocer al menos su orden (el parámetro m del modelo). Puesto que subestimar el orden del canal (asumir un orden menor del que realmente tiene) prácticamente imposibilita la detección, es habitual considerar un orden “lo bastante grande” (posiblemente mayor del que realmente tiene, en cuyo caso estaríamos sobreestimando el orden del canal) como para evitar dicha situación. Sin embargo, sobreestimar el orden del canal normalmente incrementa la complejidad del receptor al tiempo que reduce su rendimiento en términos de probabilidad de error. Así, en el Capítulo 6 de la tesis nos planteamos el problema de la detección en sistemas MIMO en los que tanto los coeficientes del canal como el orden del mismo son desconocidos. Se trata, por tanto, de detectar la secuencia transmitida, $\mathbf{s}_{0:t}$, a partir de las observaciones estimando la matriz de canal, \mathbf{H}_t , que a su vez depende del parámetro m (la memoria del canal), que también es necesario estimar.

3.4. Conclusiones

En este capítulo se han presentado las ecuaciones matemáticas que definen el comportamiento del sistema MIMO que servirá de base para todos los algoritmos desarrollados a lo largo de la tesis. Hemos representado el proceso de transmisión en términos de un sistema dinámico en formato de espacio de estados al ser esta la manera más conveniente para el desarrollo de los algoritmos presentados en los capítulos siguientes.

El interés despertado en la actualidad por los sistemas MIMO se debe a que en ellos la capacidad del canal crece linealmente con el mínimo entre el número de antenas emisoras y el número de antenas receptoras. Esto se traduce en tasas de transferencia de información más elevadas empleando el mismo espectro radioeléctrico, es decir, en un mayor aprovechamiento del medio de transmisión.

Capítulo 4

Revisión de aproximaciones anteriores

Son muchas las técnicas propuestas en la literatura para llevar a cabo la igualación de canales MIMO. Este capítulo reúne algunas de las que hemos considerado más relevantes explicando brevemente en que consiste cada una de ellas. Las hemos clasificado en dos grupos según estén basadas o no en la metodología SMC. A las que no lo están las hemos denominado convencionales.

Igualación es el proceso que se lleva cabo en el receptor para anular el efecto que el canal tiene sobre la señal transmitida. En último término, el objetivo que persiguen todos los métodos descritos en este capítulo es detectar los símbolos transmitidos en el instante de tiempo t en base a las observaciones recogidas en dicho instante y, posiblemente, en instantes posteriores. Así pues, y a la vista de la ecuación (3.2), queremos detectar el vector \mathbf{s}_t (incluido en \mathbf{s}_{t-m+1}) a partir de \mathbf{y}_t . Cuando el canal tiene memoria ($m > 1$) también es útil considerar el modelo de observaciones apiladas dado por la ecuación (3.9).

4.1. Métodos convencionales para la igualación de canales MIMO

En esta sección estudiaremos algunos de los métodos clásicos para la igualación de canales MIMO. En su mayoría son técnicas que ya se aplicaron originariamente con éxito a sistemas con una única entrada y una única salida (SISO, del inglés *Single Input Single Output*), que siguen siendo válidos cuando tenemos varias antenas tanto en emisión como en recepción.

4.1.1. Receptores Lineales

Constituyen una de las técnicas más sencillas para igualar un canal MIMO y por ello son ampliamente utilizados. Un filtro lineal para extraer el i -ésimo elemento de \mathbf{s}_{t-m+1} en un modelo como el de la ecuación (3.2) no es más que un vector, $\mathbf{f}_{t,i} = [f_{t,i}(1), f_{t,i}(2), \dots, f_{t,i}(N_r)]^\top$, de dimensiones $N_r \times 1$, cuyo producto escalar por el vector de

observaciones da lugar a una estimación, $\tilde{s}_{t-m+1}(i)$, de dicho elemento, es decir,

$$\tilde{s}_{t-m+1}(i) = \mathbf{f}_{t,i}^\top \mathbf{y}_t. \quad (4.1)$$

Debido al ruido que afecta a la transmisión, la estimación dada por el filtro $\mathbf{f}_{t,i}$ no corresponde con ninguno de los símbolos del alfabeto \mathcal{S} . Se dice que $\tilde{s}_{t-m+1}(i)$ es una estimación blanda de $s_{t-m+1}(i)$ y, por medio de un detector de umbral, será mapeada a un símbolo de \mathcal{S} , que constituye una estimación dura de $s_{t-m+1}(i)$. La Figura 4.1 ilustra la ecuación (4.1) y muestra el esquema de un receptor lineal que obtiene una estimación blanda del i -ésimo símbolo en \mathbf{s}_{t-m+1} .

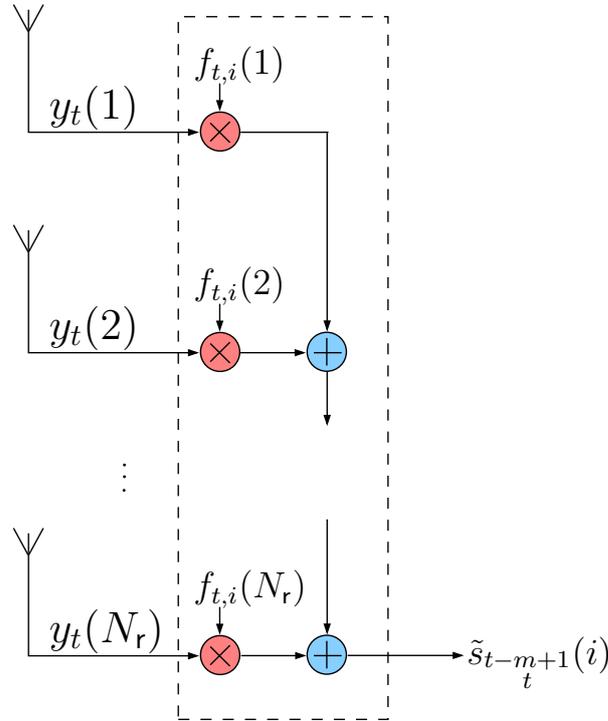


Figura 4.1: Esquema de un receptor lineal.

Para detectar el vector \mathbf{s}_{t-m+1} al completo se pueden agrupar los filtros necesarios para estimar cada uno de sus elementos en una matriz \mathbf{F}_t , de tamaño $N_r \times N_t$, cuya i -ésima columna se corresponde con el filtro $\mathbf{f}_{t,i}$ que utilizábamos en la ecuación (4.1) para obtener una estimación blanda de $s_{t-m+1}(i)$. Entonces,

$$\tilde{\mathbf{s}}_{t-m+1} = \mathbf{F}_t^\top \mathbf{y}_t \quad (4.2)$$

constituye una estimación blanda de \mathbf{s}_{t-m+1} .

Si, como es el caso que nos ocupa, el canal introduce dispersión temporal, una detección fiable de los símbolos transmitidos en el instante de tiempo t requiere tener en cuenta

observaciones posteriores a dicho instante (procedimiento al que nos referiremos como “suavizado”). Podemos obtener estimaciones blandas de los símbolos en \mathbf{s}_t teniendo en cuenta las observaciones $\mathbf{y}_{t:t+a}$ (donde $a > 0$ es el factor de suavizado) si construimos filtros que detecten en alguno de los modelos apilados de la Sección 3.1.1.

A la vista de las ecuaciones (3.9) y (3.13) podríamos estimar mediante filtrado lineal los vectores de símbolos $\mathbf{s}_{t:t+a}$ (los vectores de símbolos $\mathbf{s}_{t-m+1:t-1}$ contenidos en el vector $\mathbf{s}_{t-m+1:t+a}$ de la ecuación (3.9) ya han sido detectados entre los instantes $t - m + 1$ y $t - 1$), pero sólo el vector \mathbf{s}_t estaría siendo detectado con un factor de suavizado a . La detección del vector \mathbf{s}_{t+1} se efectuaría basándose en las observaciones $\mathbf{y}_{t+1:a}$ (supuesto conocido el canal, las observaciones anteriores a $t + 1$ no aportan nada) y el vector \mathbf{s}_{t+a} se estaría detectando sin llevar a cabo ningún tipo de suavizado.

De aquí en adelante supondremos que el canal tiene memoria y, ya sea en base a la ecuación (3.9) o a la (3.13), construiremos un filtro matricial, $\mathbf{F}_{t,a}$, que estime los símbolos transmitidos en t con factor de suavizado $a > 0$. A la hora de construir dicho filtro se pueden seguir diversos criterios. A continuación revisaremos dos de los más comunes.

Decorrelación

Una primera aproximación consiste en calcular $\mathbf{F}_{t,a}$ como la pseudoinversa [64] por la izquierda de la matriz de canal (o de una estimación de la misma en caso de que sea desconocida). Así, el filtro de decorrelación (también conocido como forzador de ceros) para estimar $\mathbf{s}_{t:t+a}$ en (3.13) es [1]

$$\mathbf{F}_{t,a}^\top = \left(\bar{\mathbf{H}}_{t,a}^\top \bar{\mathbf{H}}_{t,a} \right)^{-1} \bar{\mathbf{H}}_{t,a}^\top. \quad (4.3)$$

Nótese que la pseudoinversa por la izquierda de una matriz solo existe cuando el número de filas es mayor o igual que el de columnas (en éste último caso, tendríamos que la pseudoinversa sería equivalente a la inversa). Por tanto, solo podemos utilizar un receptor de decorrelación para llevar a cabo la detección en la ecuación (3.9) si $N_r(a+1) \geq N_t(m+a)$. Cuando $N_r = N_t$ (el número de antenas receptoras es igual al de transmisoras) esta condición se reduce a $m = 1$, es decir, a que el canal no tenga memoria. Para superar esta limitación recurrimos al modelo de observaciones transformadas que viene dado por la ecuación (3.13): siempre que tengamos al menos tantas antenas emisoras como receptoras es posible estimar $\mathbf{s}_{t:t+a}$ en esta ecuación mediante un filtro de decorrelación, ya que $N_r(a+1) \geq N_t(a+1)$ tan sólo implica $N_r \geq N_t$.

El receptor de decorrelación es óptimo en ausencia de ruido, pero cuando éste está presente, podría verse amplificado a la salida del filtro, que viene dada por

$$\begin{aligned} \tilde{\mathbf{s}}_{t:t+a} &= \mathbf{F}_{t,a}^\top \bar{\mathbf{y}}_{t:t+a} = \mathbf{F}_{t,a}^\top \bar{\mathbf{H}}_{t,a} \mathbf{s}_{t:t+a} + \mathbf{F}_{t,a}^\top \mathbf{g}_{t:t+a} \\ &= \left(\bar{\mathbf{H}}_{t,a}^\top \bar{\mathbf{H}}_{t,a} \right)^{-1} \bar{\mathbf{H}}_{t,a}^\top \bar{\mathbf{H}}_{t,a} \mathbf{s}_{t:t+a} + \mathbf{F}_{t,a}^\top \mathbf{g}_{t:t+a} = \mathbf{s}_{t:t+a} + \mathbf{F}_{t,a}^\top \mathbf{g}_{t:t+a}. \end{aligned} \quad (4.4)$$

Así pues, este receptor no tiene en cuenta en modo alguno la potencia del ruido gaussiano que afecta a las observaciones que va a procesar.

Error cuadrático medio mínimo

El criterio MMSE (*Minimum Mean Square Error*) [1] elige los coeficientes del filtro de manera que se minimice el error cuadrático medio cometido en la estimación de los símbolos transmitidos, es decir, $\mathbf{F}_{t,a}$ se calcula como

$$\mathbf{F}_{t,a} = \arg \min_{\mathbf{F}_{t,a}} \mathbb{E} \left[\left\| \mathbf{F}_{t,a}^\top \mathbf{y}_{t+a} - \mathbf{s}_{t-m+1} \right\|^2 \right], \quad (4.5)$$

donde $\|\cdot\|^2$ denota la norma cuadrática del vector correspondiente.

En el Apéndice A se resuelve el problema de optimización (4.5) para un sistema MIMO genérico. Utilizando la ecuación (A.13) obtenida al final de dicho apéndice tenemos que el filtro MMSE para detectar el vector de símbolos transmitidos, \mathbf{s}_{t-m+1} , en la ecuación (3.9) viene dado por

$$\mathbf{F}_{t,a} = (\mathbf{H}_{t,a} \sigma_s^2 \mathbf{H}_{t,a}^\top + \sigma_g^2 \mathbf{I})^{-1} \sigma_s^2 \mathbf{H}_{t,a}. \quad (4.6)$$

El receptor MMSE no amplifica el ruido a la salida del filtro, lo cual constituye su principal ventaja frente al receptor de decorrelación. Es más: cuando la varianza del ruido tiende a cero, el detector MMSE se puede reducir al de decorrelación. Así, para SNRs altas el rendimiento de ambos es prácticamente el mismo, mientras que para SNRs bajas el del receptor MMSE es sensiblemente mejor.

4.1.2. Receptor realimentado

La Figura 4.2 muestra el diagrama de bloques de un igualador con realimentación de decisiones (DFE, *Decision Feedback Equalizer*) típico, que no hace suavizado.

Este tipo de receptores consta de dos filtros lineales y un detector de umbral intercalado entre ambos. El filtro de realimentación o filtro regresivo se encarga de suprimir la interferencia intersimbólica (ISI) de las observaciones utilizando las estimaciones duras proporcionadas por el detector de umbral en instantes de tiempo anteriores (almacenadas por medio de una línea de elementos de retardo), mientras que el filtro hacia delante o progresivo compensa la distorsión introducida por el canal. La estimación (blanda) del vector de símbolos transmitidos, \mathbf{s}_t , a la salida del filtro viene dada por

$$\tilde{\mathbf{s}}_t = \mathbf{F}^\top \mathbf{y}_t - \mathbf{B}^\top \mathbf{s}_{t-1}. \quad (4.7)$$

El filtro hacia atrás se descompone en las submatrices que en la figura aparecen multiplicando a cada uno de los vectores de símbolos detectados en los instantes anteriores: $\mathbf{B}^\top = [\mathbf{B}(m-1), \dots, \mathbf{B}(-1)]$.

El DFE funciona mejor que un receptor lineal convencional cuando la probabilidad de error en las decisiones es baja (a efectos prácticos, cuando la SNR es alta) [57]. En tal caso, la no linealidad que introduce el detector de umbral permite eliminar el ruido asociado a estimaciones previas. Cuando la probabilidad de tomar una decisión incorrecta es alta, este tipo de receptores sufren el problema de la propagación de errores: una decisión incorrecta en un determinado

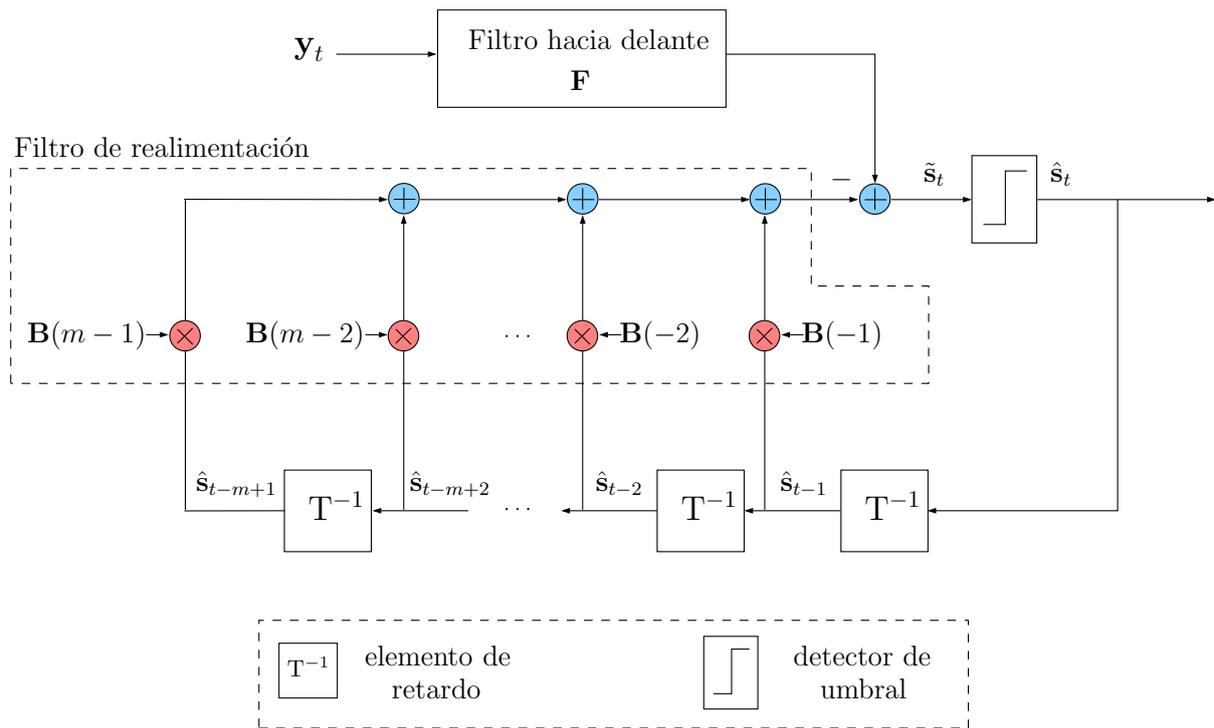


Figura 4.2: Esquema de un receptor DFE.

instante de tiempo hace que no se elimine correctamente la ISI en instantes futuros, lo cual dará lugar a malas estimaciones de los símbolos, que a su vez originarán más decisiones incorrectas.

El esquema DFE puede extenderse para obtener estimaciones suavizadas. La principal diferencia, en este caso, es que el filtro F debe mitigar también la ISI anticausal (debida a los símbolos transmitidos en instantes posteriores al de los símbolos que se están tratando de detectar).

4.1.3. El algoritmo de Viterbi

El algoritmo de Viterbi (VA, del inglés *Viterbi Algorithm*) es un algoritmo de programación dinámica¹ diseñado para encontrar la secuencia de estados que más verosímelmente da lugar a una determinada secuencia de observaciones.

Para poder aplicar el VA es necesario que el coste de cualquier secuencia de estados pueda descomponerse en la suma de los costes de cada una de las transiciones que la originan y que éstos sean no negativos. El algoritmo obtiene entonces la secuencia de menor coste eligiendo aquellas transiciones que minimizan dicha suma. A continuación veremos como el problema de encontrar la secuencia de vectores de símbolos transmitidos más verosímil en el modelo de

¹Un algoritmo de programación dinámica obtiene la solución a un problema dividiéndolo en subproblemas solapados que se resuelven secuencialmente.

señal del Capítulo 3 se reduce a minimizar un sumatorio de distancias entre vectores (que harán el papel de costes).

En el sistema dinámico descrito por las ecuaciones (3.1) y (3.5), la secuencia de vectores de símbolos más probablemente transmitida, $\mathbf{s}_{0:t}$, dadas las observaciones, $\mathbf{y}_{0:t}$, es aquella que maximiza la probabilidad *a posteriori*, es decir

$$\mathbf{s}_{0:t} = \arg \max_{\mathbf{s}_{0:t}} p(\mathbf{s}_{0:t} | \mathbf{y}_{0:t}). \quad (4.8)$$

Descomponiendo dicha probabilidad por medio del teorema de Bayes como

$$p(\mathbf{s}_{0:t} | \mathbf{y}_{0:t}) = \frac{p(\mathbf{y}_{0:t} | \mathbf{s}_{0:t}) p(\mathbf{s}_{0:t})}{p(\mathbf{y}_{0:t})}, \quad (4.9)$$

considerando todas las posibles secuencias de vectores transmitidos igualmente probables ($p(\mathbf{s}_{0:t})$ es constante), y teniendo en cuenta que $p(\mathbf{y}_{0:t})$ es una constante de normalización que no depende de $\mathbf{s}_{0:t}$, tenemos que el problema de optimización de la ecuación (4.8) se reduce a

$$\mathbf{s}_{0:t} = \arg \max_{\mathbf{s}_{0:t}} p(\mathbf{y}_{0:t} | \mathbf{s}_{0:t}). \quad (4.10)$$

y hemos pasado de un problema MAP (*Maximum A Posteriori*) de maximización de la probabilidad *a posteriori* a uno MV de maximización de la verosimilitud.

La verosimilitud anterior se puede descomponer a su vez como

$$p(\mathbf{y}_{0:t} | \mathbf{s}_{0:t}) = \prod_{i=0}^t p(\mathbf{y}_i | \mathbf{s}_{i-m+1:i}), \quad (4.11)$$

donde, atendiendo a la ecuación (3.2), el i -ésimo factor del producto es una normal de media $\boldsymbol{\mu}_i = \sum_{l=0}^{m-1} \mathbf{H}_i(l) \mathbf{s}_{i-l}$ y matriz de covarianza $\sigma_g^2 \mathbf{I}_{N_r}$. Sustituyendo las probabilidades por la expresión de su f.d.p. resulta

$$p(\mathbf{y}_{0:t} | \mathbf{s}_{0:t}) = \prod_{i=0}^t (2\pi\sigma_g^2)^{-N_r/2} \exp\left(-\frac{1}{2\sigma_g^2} (\mathbf{y}_i - \boldsymbol{\mu}_i)^\top (\mathbf{y}_i - \boldsymbol{\mu}_i)\right) \quad (4.12)$$

que debemos maximizar con respecto a $\mathbf{s}_{0:t}$ para obtener la secuencia más probablemente transmitida. En lugar de maximizar $p(\mathbf{y}_{0:t} | \mathbf{s}_{0:t})$, podemos maximizar su logaritmo natural

$$\begin{aligned} \log(p(\mathbf{y}_{0:t} | \mathbf{s}_{0:t})) &= \log\left(\prod_{i=0}^t (2\pi\sigma_g^2)^{-N_r/2}\right) + \log\left(\prod_{i=0}^t \exp\left(-\frac{1}{2\sigma_g^2} (\mathbf{y}_i - \boldsymbol{\mu}_i)^\top (\mathbf{y}_i - \boldsymbol{\mu}_i)\right)\right) \\ &= k_1 + \sum_{i=0}^t \log\left(\exp\left(-\frac{1}{2\sigma_g^2} (\mathbf{y}_i - \boldsymbol{\mu}_i)^\top (\mathbf{y}_i - \boldsymbol{\mu}_i)\right)\right) \\ &= k_1 + \sum_{i=0}^t \left(-\frac{1}{2\sigma_g^2} (\mathbf{y}_i - \boldsymbol{\mu}_i)^\top (\mathbf{y}_i - \boldsymbol{\mu}_i)\right) \\ &= k_1 - k_2 \sum_{i=0}^t (\mathbf{y}_i - \boldsymbol{\mu}_i)^\top (\mathbf{y}_i - \boldsymbol{\mu}_i) = k_1 - k_2 \sum_{i=0}^t \|\mathbf{y}_i - \boldsymbol{\mu}_i\|^2 \end{aligned} \quad (4.13)$$

donde $k_1 = \log \left(\prod_{i=0}^t (2\pi\sigma_g^2)^{-N_r/2} \right)$ y $k_2 = 1/2\sigma_g^2$ no dependen de $\mathbf{s}_{0:t}$.

A la vista de la ecuación (4.13), maximizar $\log(p(\mathbf{y}_{0:t}|\mathbf{s}_{0:t}))$ es equivalente a minimizar $\sum_{i=0}^t \|\mathbf{y}_i - \boldsymbol{\mu}_i\|^2$ y el problema de optimización que finalmente resulta,

$$\mathbf{s}_{0:t} = \arg \min_{\mathbf{s}_{0:t}} \sum_{i=0}^t \|\mathbf{y}_i - \boldsymbol{\mu}_i\|^2 = \arg \min_{\mathbf{s}_{0:t}} \sum_{i=0}^t \left\| \mathbf{y}_i - \sum_{l=0}^{m-1} \mathbf{H}_i(l) \mathbf{s}_{i-l} \right\|^2, \quad (4.14)$$

se puede resolver por medio del VA conociendo (o estimando) las matrices de canal $\mathbf{H}_i(l)$ para todo $i = 0, \dots, t$ y todo $l = 0, \dots, m-1$.

El Apéndice B explica más en detalle el algoritmo de Viterbi y muestra su funcionamiento sobre un sencillo ejemplo basado en el modelo de señal utilizado en esta tesis.

Uno de los mayores inconvenientes del algoritmo de Viterbi es que no permite comenzar la detección de los símbolos transmitidos hasta que se recibe la última observación de la trama. Además, es necesario tener almacenados tantos caminos como estados tenga el *trellis*, y su longitud aumenta con el tiempo. Una posibilidad para paliar en cierta medida este problema consiste en no almacenar en el instante de tiempo t la secuencia completa de estados (o de entradas) que determina cada uno de los caminos supervivientes, sino únicamente la parte de la misma comprendida entre los instantes $t-d$ y t , siendo d un parámetro de diseño. En tal caso, en el instante t se detectan los símbolos transmitidos en $t-d$ y, por tanto, sólo es necesario recibir las d primeras observaciones para comenzar la detección de los símbolos transmitidos. En [65] se prueba que utilizando un d igual a 4 ó 5 veces la memoria del canal se consigue un resultado muy cercano al que se tendría esperando hasta el final de la trama para comenzar la detección.

4.1.4. Procesado por supervivientes

Como se vio en la sección anterior, el algoritmo de Viterbi halla la secuencia de vectores de símbolos transmitidos que maximiza la probabilidad condicional de las observaciones, es decir, la secuencia más verosímil. Pero lo hace asumiendo que la información del estado del canal es perfectamente conocida en todo momento. Esto no es cierto en la práctica: nunca se conoce con precisión el canal y una buena estimación es lo máximo a que se puede aspirar. Por tanto, el uso de cualquier algoritmo que parta de que el canal es conocido debe ir precedido de un proceso de estimación del mismo.

Al explicar el algoritmo de Viterbi se vio una modificación del mismo que hacía posible comenzar la detección de los símbolos transmitidos antes de recibir la última observación de la trama. Dicha mejora permitía al VA detectar los datos transmitidos con un cierto retardo. Si se considera un canal variante en el tiempo, como ocurre en nuestro modelo de señal, podríamos pensar en utilizar un estimador de canal externo alimentado por los símbolos que el VA vaya detectando y que proporcione al mismo las estimaciones de canal que necesita. Sin embargo, el retardo mencionado haría que las estimaciones de canal no “llegaran a tiempo” al Viterbi, lo cual degradaría el rendimiento del algoritmo tanto más cuanto más rápidamente varíe el canal.

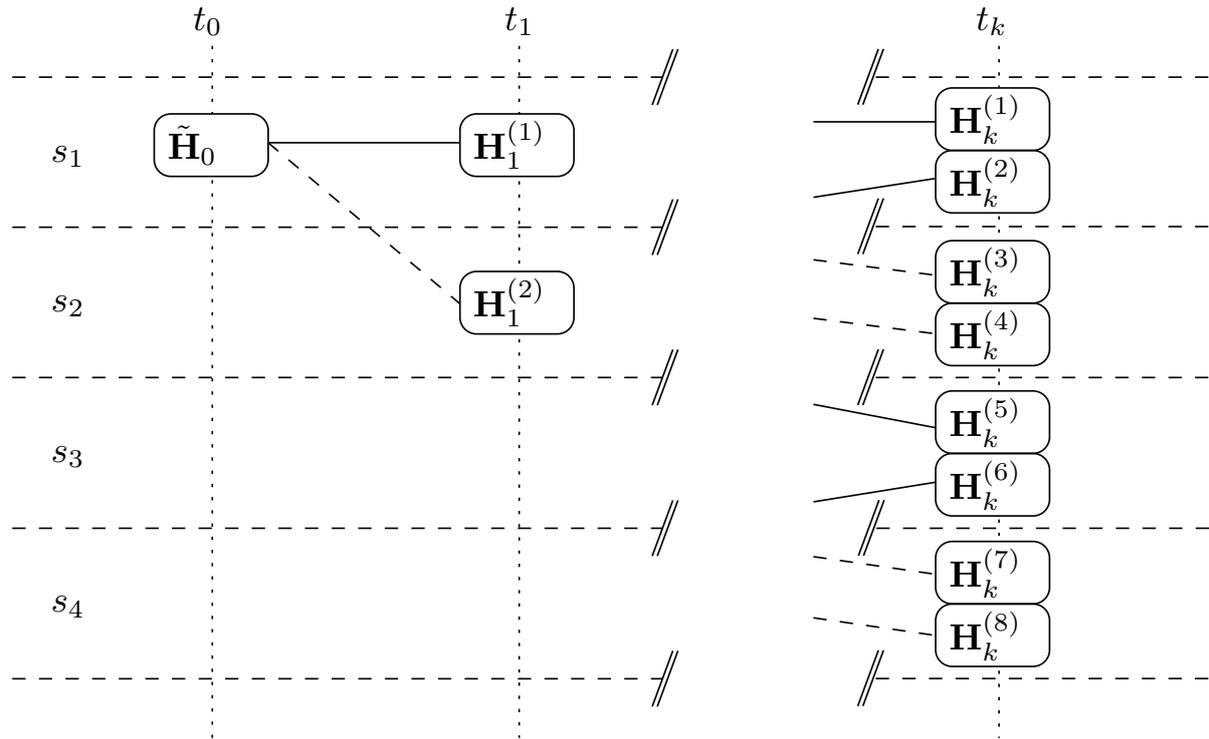
La situación anterior se podría evitar si la estimación de canal estuviera de alguna manera integrada en el propio algoritmo de Viterbi. Ésta es la idea fundamental en torno a la cual gira el procesado por superviviente (PSP) [66]. Al igual que el VA, es un algoritmo de búsqueda en *trellis*, pero ahora cada camino dispone de sus propios parámetros (a diferencia de lo que ocurría en el VA, donde todos los caminos compartían un mismo conjunto de parámetros), y éstos se irán actualizando a medida que aquel se va extendiendo (esto es, los nuevos vectores de símbolos que se incorporan a la secuencia detectada se utilizan para reestimar los parámetros).

Otra diferencia entre el VA y el PSP está en el número de caminos que mantienen por estado. Veíamos en la sección anterior como, en un determinado instante de tiempo, el VA almacenaba únicamente el mejor camino (el de menor coste) que llegaba a cada estado. Esto es debido a que, cuando el canal es perfectamente conocido, los costes están “bien” calculados, es decir, puesto que se conocen con exactitud los parámetros de los que depende no hay ninguna incertidumbre en los valores obtenidos. Cuando sólo disponemos de una estimación de los parámetros necesarios para calcular las métricas, éstas no son exactas. Para aliviar en cierta medida este problema, el PSP retiene en cada instante de tiempo un número P (por lo general, $P > 1$) de caminos por estado, cada uno de los cuales se denomina superviviente.

Para utilizar el PSP en nuestro modelo de señal, vamos a asociar a cada camino del *trellis* una estimación de la matriz de canal, \mathbf{H}_t , necesaria para calcular el coste de las transiciones entre estados. El PSP opera entonces de forma análoga al VA. En primer lugar, dado un camino (superviviente) hasta la t -ésima etapa del *trellis*, se calcula en base a la estimación de canal que tiene asociada el coste de cada posible transición que parta de su estado actual. Si el *trellis* tiene N_e estados y estamos manejando P supervivientes por estado, esto da lugar a $N_e P |\mathcal{S}^{N_t}|$ ramas o caminos que llegan hasta la etapa $t + 1$ del *trellis*. De entre ellos se eligen los P mejores (de menor coste) que llegan a cada estado y se actualizan sus correspondientes estimaciones de canal empleando el último vector de símbolos añadido al camino (el que etiqueta la última transición). La Figura 4.3 muestra el *trellis* asociado a un sistema como el que sirve para explicar el funcionamiento del VA en el apéndice B ($N_t = 1$ antena transmisora, $N_r = 1$ antena receptora y memoria $m = 3$), en el que tenemos $P = 2$ supervivientes por estado. El i -ésimo superviviente en el instante t determina una secuencia de vectores de símbolos transmitidos, $\mathbf{s}_{0:t}^{(i)}$, y tiene asociada una estimación de canal, $\mathbf{H}_t^{(i)}$. Nótese, asimismo, que las estimaciones de canal y secuencias de símbolos correspondientes a distintos supervivientes de un mismo estado son diferentes (aunque las secuencias de símbolos necesariamente han de tener en común los últimos $m - 1$ vectores, que son los que constituyen el estado).

4.1.5. Algoritmo BCJR

El algoritmo de Viterbi encuentra la secuencia de símbolos más probablemente transmitida de entre todas las posibles cuando éstas son equiprobables *a priori* (en ausencia de observaciones). Si no se da esta condición, el VA obtiene la secuencia más verosímil, que no es necesariamente la más probablemente transmitida dadas las observaciones disponibles (i.e., la que maximiza la probabilidad *a posteriori*).


 Figura 4.3: Diagrama de *trellis* de un PSP.

En 1974 Bahl, Cocke, Jelinek y Raviv propusieron un algoritmo (conocido por el acrónimo BCJR) [67] para calcular las probabilidades *a posteriori* de las transiciones y estados que atraviesa un proceso de Markov cuando las observaciones que genera están perturbadas por un canal discreto sin memoria. Aplicado a un modelo de señal como el del Capítulo 3, el algoritmo BCJR halla en cada instante de tiempo el vector de símbolos transmitidos que maximiza la probabilidad *a posteriori*. Es decir, encuentra el vector \mathbf{s}_t tal que

$$\mathbf{s}_t = \arg \max_{\mathbf{s}_t} p(\mathbf{s}_t | \mathbf{y}_{0:T}), \quad T \geq t. \quad (4.15)$$

Nótese que éste es un problema de optimización distinto al planteado por la ecuación (4.10), que es el que en última instancia resuelve el VA, y al planteado por la ecuación (4.8) que resulta de (4.10) cuando todas las secuencias de vectores de símbolos tienen la misma probabilidad de ocurrencia.

El BCJR opera procesando la secuencia de observaciones recibida dos veces: una hacia delante, en la que para cada instante de tiempo se calcula la probabilidad de cada posible estado del sistema dadas las observaciones recibidas hasta ese instante, y otra hacia atrás en la que se calcula la verosimilitud de cada estado en base únicamente a las observaciones que siguen al instante que se está procesando. Combinando ambas probabilidades se puede obtener la probabilidad *a posteriori* de cada estado en cada instante de tiempo, así como la probabilidad asociada a cada transición que haya podido tener lugar en el sistema.

El algoritmo BCJR juega un papel fundamental dentro del campo de la codificación de canal al ser utilizado (entre otros) por los códigos turbo [68].

4.1.6. Algoritmo EM

El algoritmo EM (del inglés, *Expectation Maximization*) [69] se utiliza para realizar estimación paramétrica basada en el criterio de máxima verosimilitud cuando, o bien no es posible maximizar directamente la función de verosimilitud, o bien el resultado de la misma da lugar a un procedimiento computacionalmente muy costoso. Algoritmos similares al EM (y que en último término se reducen a éste) se habían empleado con anterioridad, pero fueron Dempster, Laird y Rubin quienes, en el año 1977, enunciaron formalmente el algoritmo y acuñaron el término EM [70].

Supongamos que queremos estimar por máxima verosimilitud un parámetro θ (posiblemente vectorial), que toma valores en el conjunto Θ , dado un conjunto de observaciones relacionadas, \mathbf{y} . Es decir, queremos hallar

$$\hat{\theta} = \arg \max_{\theta \in \Theta} p(\mathbf{y}; \theta). \quad (4.16)$$

El algoritmo EM parte de la base de que existen un conjunto de datos no observados, que denotaremos por \mathbf{x} , cuyo conocimiento simplificaría notablemente el problema de estimación. Algunos autores asumen que los datos observados son una transformación no invertible (y que da lugar a una pérdida de información) de los no observados [71], por lo que éstos últimos contienen toda la información disponible para la estimación del parámetro θ . Otros autores consideran que los datos no observados completan de alguna manera a los observados y ambos son relevantes de cara a la estimación de θ . Nosotros seguiremos esta segunda aproximación y vamos a distinguir entre

- \mathbf{y} , el conjunto de datos observados,
- \mathbf{x} , el conjunto de datos no observados, y
- $\mathbf{z} = \{\mathbf{y}, \mathbf{x}\}$, el conjunto de datos completos.

Nótese que las f.d.p. de \mathbf{y} , \mathbf{x} y \mathbf{z} dependen del parámetro θ , por lo que vamos a denotarlas $p(\mathbf{y}; \theta)$, $p(\mathbf{x}; \theta)$ y $p(\mathbf{z}; \theta)$, respectivamente.

Simplemente utilizando la definición de probabilidad condicional, podemos escribir la verosimilitud para los datos completos como

$$p(\mathbf{z}; \theta) = p(\mathbf{x}, \mathbf{y}; \theta) = p(\mathbf{x}|\mathbf{y}; \theta)p(\mathbf{y}; \theta). \quad (4.17)$$

Tomando logaritmos a ambos lados de la igualdad y despejando el término que involucra la verosimilitud de las observaciones resulta

$$\log p(\mathbf{y}; \theta) = \log p(\mathbf{z}; \theta) - \log p(\mathbf{x}|\mathbf{y}; \theta). \quad (4.18)$$

Todos los términos de (4.18) son en última instancia funciones de θ (y de los vectores aleatorios \mathbf{y} y \mathbf{x}). Evaluando entonces dicha ecuación para un valor arbitrario θ' y tomando la esperanza respecto de $p(\mathbf{x}|\mathbf{y}; \hat{\theta})$, que denotaremos por $\mathbb{E}_{\mathbf{x}|\mathbf{y}; \hat{\theta}}[\cdot]$, tenemos

$$\log p(\mathbf{y}; \theta) = U(\theta', \hat{\theta}) - V(\theta', \hat{\theta}) \quad (4.19)$$

donde

$$U(\theta', \hat{\theta}) = \mathbb{E}_{\mathbf{x}|\mathbf{y}; \hat{\theta}} [\log p(\mathbf{z}; \theta)], \quad (4.20)$$

$$V(\theta', \hat{\theta}) = \mathbb{E}_{\mathbf{x}|\mathbf{y}; \hat{\theta}} [\log p(\mathbf{x}|\mathbf{y}; \theta)] \quad (4.21)$$

y se ha tenido en cuenta que, trivialmente, $\mathbb{E}_{\mathbf{x}|\mathbf{y}; \hat{\theta}} [\log p(\mathbf{y}; \theta)] = \log p(\mathbf{y}; \theta)$.

Intuitivamente, estamos calculando cuál sería en término medio la verosimilitud logarítmica del parámetro θ' a la vista de las observaciones, suponiendo que el verdadero parámetro del que dependen los datos (tantos los observados como los no observados) es $\hat{\theta}$.

Se puede demostrar [71] que la función V satisface

$$V(\theta', \hat{\theta}) \leq V(\hat{\theta}, \hat{\theta}), \forall \theta' \quad (4.22)$$

por lo que eligiendo θ' tal que

$$U(\theta', \hat{\theta}) \geq U(\hat{\theta}, \hat{\theta}), \quad (4.23)$$

tenemos

$$U(\theta', \hat{\theta}) - V(\theta', \hat{\theta}) \geq U(\hat{\theta}, \hat{\theta}) - V(\hat{\theta}, \hat{\theta}) \quad (4.24)$$

o, equivalentemente, según la ecuación (4.19),

$$\log p(\mathbf{y}; \theta') \geq \log p(\mathbf{y}; \hat{\theta}), \quad (4.25)$$

es decir, la verosimilitud (logarítmica) de θ' es mayor que la de $\hat{\theta}$ y de esta forma es posible construir una secuencia de estimaciones del parámetro θ no decrecientes en verosimilitud que tienden, por tanto, a un máximo de la función de verosimilitud $p(\mathbf{y}; \theta)$.

Así pues, partiendo de una estimación inicial, $\theta^{(0)}$, del parámetro θ , el algoritmo EM incrementa la verosimilitud para los datos observados iterando los siguientes pasos:

1. **Etapa E** (esperanza): calcular

$$U(\theta', \hat{\theta}^{(i)}) = \mathbb{E}_{\mathbf{x}|\mathbf{y}; \hat{\theta}^{(i)}} [\log p(\mathbf{z}; \theta')]. \quad (4.26)$$

Esto supone utilizar los datos incompletos (observados) y la estimación actual de los parámetros para obtener un conjunto de estadísticos suficientes de los datos completos.

2. **Etapa M** (maximización): hallar $\hat{\theta}^{(i+1)}$ tal que

$$\hat{\theta}^{(i+1)} = \arg \max_{\theta'} U(\theta', \hat{\theta}^{(i)}), \quad (4.27)$$

esto es, actualizar la estimación del parámetro maximizando la verosimilitud de los datos completos.

Las ecuaciones (4.26) y (4.27) son la base del algoritmo EM. Como en todo algoritmo iterativo determinista que trata de maximizar una función con máximos locales, la inicialización es importante y la capacidad del EM para encontrar el máximo global depende en buena medida del $\hat{\theta}^{(0)}$ elegido.

Nótese que EM es más una metodología que un algoritmo y para un mismo problema de estimación podemos derivar diferentes algoritmos EM de acuerdo a la información que identifiquemos como los datos no observados. Siempre se elige \mathbf{x} de tal manera que los cálculos en (4.26) y (4.27) sean más simples que el problema de optimización (4.16). El precio de éste decremento en la complejidad es una convergencia posiblemente lenta del algoritmo EM. Se debe buscar un compromiso entre complejidad computacional y velocidad de convergencia que, en última instancia, va a depender de cuanta información acerca del parámetro a estimar, θ , está contenida en los datos completos [70, 72, 73].

En nuestro modelo de señal, el parámetro a estimar en el instante t es el vector con los símbolos transmitidos, \mathbf{s}_t , y el conjunto de datos incompletos viene dado por el vector de observaciones \mathbf{y}_t . En cuanto a los datos no observados, la elección que hagamos de los mismos va a determinar el algoritmo resultante. En [74] se deriva un receptor multiusuario para un sistema CDMA que detecta el símbolo transmitido por el k -ésimo usuario considerando los transmitidos por los restantes el conjunto de los datos no observados. Así pues, dicho conjunto vendría dado en este caso por $\mathbf{x}_t^k = \{s_t(1), \dots, s_t(k-1), s_t(k+1), \dots, s_t(N)\}$ y, siendo $s_t^{(i)}(k)$ la estimación del k -ésimo elemento de \mathbf{s}_t en la i -ésima iteración del algoritmo, tendríamos el estadístico

$$U(s_t^{(i+1)}(k), s_t^{(i)}(k)) = \mathbb{E}_{\mathbf{x}_t^k | \mathbf{y}_t, s_t^{(i)}(k)} \left[\log p(\mathbf{y}_t, \mathbf{x}_t^k; s_t^{(i+1)}(k)) \right], \quad (4.28)$$

cuya maximización da lugar a una nueva estimación, $s_t^{(i+1)}(k)$, de $s_t(k)$, que tiene una verosimilitud mayor que la de $s_t^{(i)}(k)$.

Cuando el parámetro a estimar es multidimensional, como en este caso, una generalización del algoritmo EM conocida como SAGE (*Space Alternating Generalized EM* en inglés) [75] permite incrementar la velocidad de convergencia. La idea es aplicar de forma alternada el algoritmo EM sobre conjuntos disjuntos de parámetros. En el problema que nos ocupa, podríamos considerar cada conjunto formado por uno de los símbolos de \mathbf{s}_t . Un receptor CDMA que utiliza el algoritmo EM de esta manera se puede encontrar en [74].

4.2. Igualación de canales MIMO basada en métodos SMC

Para aplicar los métodos secuenciales de Monte Carlo al problema de la igualación de canales MIMO nos vamos a plantear la aproximación de la f.d.p. de los símbolos dadas las observaciones, es decir, $p(\mathbf{s}_{0:t} | \mathbf{y}_{0:t})$. Dicha densidad contiene toda la información estadística relevante para la estimación óptima (bayesiana) de $\mathbf{s}_{0:t}$.

Un algoritmo SMC cuya f.d.p. objetivo es $p(\mathbf{s}_{0:t} | \mathbf{y}_{0:t})$ proporciona en el instante de tiempo t un conjunto de muestras con pesos asociados (partículas), $\left\{ \mathbf{s}_{0:t}^{(i)}, w_t^{(i)} \right\}_{i=1}^M$, que constituyen una

representación discreta aleatoria de la la mencionada f.d.p. Esta puede aproximarse entonces como

$$\hat{p}_M(\mathbf{s}_{0:t}|\mathbf{y}_{0:t}) = \sum_{i=1}^M \delta(\mathbf{s}_{0:t} - \mathbf{s}_{0:t}^{(i)})w_t^{(i)}. \quad (4.29)$$

A partir de $\hat{p}_M(\mathbf{s}_{0:t}|\mathbf{y}_{0:t})$ es fácil obtener un estimador de cualquier función de los símbolos transmitidos que nos interese. Así, por ejemplo, podemos construir un detector MAP cuya salida sea el vector de símbolos

$$\hat{\mathbf{s}}_t^{MAP} = \arg \max_{\mathbf{s}_t} \left\{ \sum_{i=1}^M \delta(\mathbf{s}_t - \mathbf{s}_t^{(i)})w_t^{(i)} \right\}. \quad (4.30)$$

Esto se reduce a seleccionar la partícula que acumula el mayor peso (ha de tenerse en cuenta que algunas partículas pueden estar replicadas).

4.2.1. Función tentativa óptima sin suavizado

Supongamos que queremos llevar a cabo la detección del t -ésimo vector de símbolos transmitidos y para ello solo disponemos de las observaciones hasta el instante t , es decir, supongamos que no podemos (o no queremos) llevar a cabo ningún tipo de suavizado. Entonces, la f.d.p. a aproximar mediante SIS es $p(\mathbf{s}_{0:t}|\mathbf{y}_{0:t})$.

Cuando se explicó el muestreo enfatizado en la Sección 2.3, se dijo que la función tentativa óptima es la propia función a aproximar. Así pues, en este caso la función tentativa óptima es $q(\mathbf{s}_{0:t}|\mathbf{y}_{0:t}) = p(\mathbf{s}_{0:t}|\mathbf{y}_{0:t})$, que se puede descomponer aplicando el teorema de Bayes y la definición de probabilidad condicionada como

$$p(\mathbf{s}_{0:t}|\mathbf{y}_{0:t}) = \frac{p(\mathbf{y}_t|\mathbf{s}_{0:t}, \mathbf{y}_{0:t-1})p(\mathbf{s}_{0:t}|\mathbf{y}_{0:t-1})}{p(\mathbf{y}_t|\mathbf{y}_{0:t-1})} = \frac{p(\mathbf{y}_t|\mathbf{s}_{0:t}, \mathbf{y}_{0:t-1})p(\mathbf{s}_t|\mathbf{s}_{0:t-1}, \mathbf{y}_{0:t-1})p(\mathbf{s}_{0:t-1}|\mathbf{y}_{0:t-1})}{p(\mathbf{y}_t|\mathbf{y}_{0:t-1})}. \quad (4.31)$$

El vector \mathbf{s}_t es independiente de todos los vectores de símbolos y de todas las observaciones anteriores², $\mathbf{s}_{0:t-1}$ e $\mathbf{y}_{0:t-1}$, respectivamente, por lo que $p(\mathbf{s}_t|\mathbf{s}_{0:t-1}, \mathbf{y}_{0:t-1}) = p(\mathbf{s}_t)$, que suponemos constante. Además $p(\mathbf{y}_t|\mathbf{y}_{0:t-1})$ no depende de $\mathbf{s}_{0:t}$, de tal manera que

$$p(\mathbf{s}_{0:t}|\mathbf{y}_{0:t}) \propto p(\mathbf{y}_t|\mathbf{s}_{0:t}, \mathbf{y}_{0:t-1})p(\mathbf{s}_{0:t-1}|\mathbf{y}_{0:t-1}). \quad (4.32)$$

Obsérvese que, dados los vectores de símbolos $\mathbf{s}_{0:t}$, el sistema dinámico que definen las ecuaciones (3.2) y (3.6) es lineal en \mathbf{H}_t y gaussiano. El KF se puede integrar en el algoritmo SIS cuando el sistema dinámico es condicionalmente lineal, y el método resultante recibe el nombre de MKF (*Mixture Kalman Filter*) [76]. En el Apéndice D se explica como calcular la verosimilitud $p(\mathbf{y}_t|\mathbf{s}_{0:t}, \mathbf{y}_{0:t-1})$, que aparece en la ecuación (4.32), mediante el KF.

²Nótese que, aún cuando el canal es dispersivo, la primera observación que involucra al t -ésimo vector de símbolos es \mathbf{y}_t .

En la Sección 2.4.2 veíamos que el muestreo enfatizado secuencial requiere que la función tentativa en el instante t se incluya a sí misma en el instante anterior. En el modelo de señal que nos ocupa, esto se traduce en una función tentativa de la forma

$$q(\mathbf{s}_{0:t}|\mathbf{y}_{0:t}) = q(\mathbf{s}_t|\mathbf{s}_{0:t-1}, \mathbf{y}_{0:t})q(\mathbf{s}_{0:t-1}|\mathbf{y}_{0:t-1}). \quad (4.33)$$

Así pues, si la función tentativa es $q(\mathbf{s}_{0:t}|\mathbf{y}_{0:t}) = p(\mathbf{s}_{0:t}|\mathbf{y}_{0:t})$, tenemos que, a la vista de la ecuación (4.32), $q(\mathbf{s}_t|\mathbf{y}_{0:t-1}, \mathbf{y}_{0:t}) \propto p(\mathbf{y}_t|\mathbf{s}_{0:t}, \mathbf{y}_{0:t-1})$. Para poder tomar una muestra de \mathbf{s}_t , dicha función tentativa marginal debe ser normalizada, por lo que en última instancia vamos a muestrear de:

$$q(\mathbf{s}_t|\mathbf{s}_{0:t-1}, \mathbf{y}_{0:t}) = \frac{p(\mathbf{y}_t|\mathbf{s}_{0:t}, \mathbf{y}_{0:t-1})}{\sum_{\tilde{\mathbf{s}}_t \in \mathcal{S}^{N_t}} p(\mathbf{y}_t|\mathbf{s}_{0:t-1}, \tilde{\mathbf{s}}_t, \mathbf{y}_{0:t-1})}. \quad (4.34)$$

Si en el instante t disponemos de un conjunto de partículas, $\{\mathbf{s}_{0:t-1}^{(i)}, w_{t-1}^{(i)}\}_{i=1}^M$, que aproximan la f.d.p. $p(\mathbf{s}_{0:t-1}|\mathbf{y}_{0:t-1})$, la metodología SIS consiste en tomar una muestra $\mathbf{s}_t^{(i)} \sim q(\mathbf{s}_t|\mathbf{s}_{0:t-1}^{(i)}, \mathbf{y}_{0:t})$ y añadirla a $\mathbf{s}_{0:t-1}^{(i)}$ para obtener una muestra $\mathbf{s}_{0:t}^{(i)} \sim q(\mathbf{s}_{0:t}|\mathbf{y}_{0:t})$. El peso de las muestras en el instante t se puede calcular entonces, gracias a la descomposición recursiva de las f.d.p. objetivo y tentativa, a partir del peso que tenían en $t - 1$ como

$$\tilde{w}_t^{(i)} = \frac{p(\mathbf{s}_{0:t}^{(i)}|\mathbf{y}_{0:t})}{q(\mathbf{s}_{0:t}^{(i)}|\mathbf{y}_{0:t})} \propto \frac{p(\mathbf{y}_t|\mathbf{s}_{0:t}^{(i)}, \mathbf{y}_{0:t-1})p(\mathbf{s}_{0:t-1}^{(i)}|\mathbf{y}_{0:t-1})}{\frac{p(\mathbf{y}_t|\mathbf{s}_{0:t}^{(i)}, \mathbf{y}_{0:t-1})q(\mathbf{s}_{0:t-1}^{(i)}|\mathbf{y}_{0:t-1})}{\sum_{\tilde{\mathbf{s}}_t \in \mathcal{S}^{N_t}} p(\mathbf{y}_t|\mathbf{s}_{0:t-1}, \tilde{\mathbf{s}}_t, \mathbf{y}_{0:t-1})}} = \sum_{\tilde{\mathbf{s}}_t \in \mathcal{S}^{N_t}} p(\mathbf{y}_t|\mathbf{s}_{0:t-1}^{(i)}, \tilde{\mathbf{s}}_t, \mathbf{y}_{0:t-1})\tilde{w}_{t-1}^{(i)}. \quad (4.35)$$

y el algoritmo SIS con función tentativa óptima sin suavizado se resume en

$$\mathbf{s}_t^{(i)} \sim q(\mathbf{s}_t|\mathbf{s}_{0:t-1}^{(i)}, \mathbf{y}_{0:t}) = \frac{p(\mathbf{y}_t|\mathbf{s}_{0:t}^{(i)}, \mathbf{y}_{0:t-1})}{\sum_{\tilde{\mathbf{s}}_t \in \mathcal{S}^{N_t}} p(\mathbf{y}_t|\mathbf{s}_{0:t-1}^{(i)}, \tilde{\mathbf{s}}_t, \mathbf{y}_{0:t-1})} \quad (4.36)$$

$$\tilde{w}_t^{(i)} = \tilde{w}_{t-1}^{(i)} \sum_{\tilde{\mathbf{s}}_t \in \mathcal{S}^{N_t}} p(\mathbf{y}_t|\mathbf{s}_{0:t-1}^{(i)}, \tilde{\mathbf{s}}_t, \mathbf{y}_{0:t-1}) \quad (4.37)$$

más pasos de remuestreo cuando sea necesario. Téngase en cuenta, además, que cada nuevo vector de símbolos, $\mathbf{s}_t^{(i)}$, que se añade a la i -ésima partícula en el instante t es utilizado para actualizar su correspondiente KF de acuerdo a las ecuaciones que figuran en el Apéndice D.

En la ecuación (4.37), el factor por el que se multiplica el peso de la i -ésima partícula en el instante $t - 1$ para obtener el peso de la nueva muestra en el instante t , $\sum_{\tilde{\mathbf{s}}_t \in \mathcal{S}^{N_t}} p(\mathbf{y}_t|\mathbf{s}_{0:t-1}^{(i)}, \tilde{\mathbf{s}}_t, \mathbf{y}_{0:t-1})$, no depende del vector $\mathbf{s}_t^{(i)}$ muestreado. Es debido a que la función tentativa es la propia f.d.p. a aproximar, por lo que todas las muestras, $\mathbf{s}_t^{(i)}, i = 1, \dots, M$, obtenidas en el instante t son i.i.d. Las muestras $\mathbf{s}_{0:t}^{(i)}, i = 1, \dots, M$ tendrán en última instancia pesos diferentes debido a que cada vector de símbolos muestreado (nótese que $\mathbf{s}_{0:t}^{(i)}$ procede de muestrear individualmente los vectores $\mathbf{s}_0, \mathbf{s}_1, \dots, \mathbf{s}_t$) se utiliza para actualizar el KF que permite calcular la verosimilitud $p(\mathbf{y}_t|\mathbf{s}_{0:t-1}^{(i)}, \tilde{\mathbf{s}}_t, \mathbf{y}_{0:t-1})$, de tal manera que

$$\mathbf{s}_{0:t}^{(i)} \neq \mathbf{s}_{0:t}^{(j)} \Rightarrow p(\mathbf{y}_t|\mathbf{s}_{0:t-1}^{(i)}, \tilde{\mathbf{s}}_t, \mathbf{y}_{0:t-1}) \neq p(\mathbf{y}_t|\mathbf{s}_{0:t-1}^{(j)}, \tilde{\mathbf{s}}_t, \mathbf{y}_{0:t-1}).$$

Pseudocódigo 4 Algoritmo SIS óptimo sin suavizado

para cada instante de tiempo $t = 1, 2, \dots$
 para cada partícula $i = 1, \dots, M$
 sea $acum_{verosimil} \leftarrow 0$
 para todo $\mathbf{s}_t \in \mathcal{S}^{N_t}$
 $verosimil(\mathbf{s}_t) \leftarrow p(\mathbf{y}_t | \mathbf{s}_t, \mathbf{s}_{0:t-1}^{(i)}, \mathbf{y}_{0:t-1})$
 $acum_{verosimil} \leftarrow acum_{verosimil} + verosimil(\mathbf{s}_t)$
 muestrear $\mathbf{s}_t^{(i)} \sim q_t(\mathbf{s}_t) = verosimil(\mathbf{s}_t) / acum_{verosimil}$
 obtener el peso sin normalizar $\tilde{w}_t^{(i)} \leftarrow w_{t-1}^{(i)} acum_{verosimil}$
 actualizar el i -ésimo KF utilizando la muestra $\mathbf{s}_t^{(i)}$
 calcular $W \leftarrow \sum_{i=1}^M \tilde{w}_t^{(i)}$
 para cada partícula $i = 1, \dots, M$
 normalizar su peso $w_t^{(i)} \leftarrow \tilde{w}_t^{(i)} / W$
 estimar el tamaño efectivo de la muestra, $\hat{M}_{ef} \leftarrow \left(\sum_{i=1}^M w_t^{(i)2} \right)^{-1}$
 si $\hat{M}_{ef} < \gamma_r M$ **entonces**
 remuestrear

El Pseudocódigo 4 describe la implementación del algoritmo.

Este algoritmo presenta dos inconvenientes: por un lado su complejidad es exponencial en N_t (debemos calcular $|\mathcal{S}|^{N_t}$ veces la verosimilitud $p(\mathbf{y}_t | \mathbf{s}_{0:t-1}^{(i)}, \tilde{\mathbf{s}}_t, \mathbf{y}_{0:t-1})$, una por cada $\tilde{\mathbf{s}}_t \in \mathcal{S}^{N_t}$) y por otro, su rendimiento es muy pobre cuando $m > 1$. Ésto último es debido a que únicamente tiene en cuenta la observación t -ésima para detectar el t -ésimo vector de símbolos transmitidos (y la energía de éste se encuentra dispersa a lo largo de las observaciones $\mathbf{y}_{t:t+m-1}$). Por tanto, como se explicó anteriormente, para igualar canales con memoria es necesario algún tipo de suavizado para conseguir una detección fiable de los datos.

4.2.2. Función tentativa óptima con suavizado

En este caso, la función tentativa es también la propia función a aproximar, pero ahora nos planteamos la aproximación de $p(\mathbf{s}_{0:t} | \mathbf{y}_{0:t+a})$, donde a es el factor de suavizado. En el contexto de los métodos secuenciales de Monte Carlo el suavizado recibe el nombre de muestreo retardado [45, 58], ya que si el factor de suavizado es a , no se puede obtener la t -ésima muestra, $\mathbf{s}_t^{(i)}$, hasta haber recibido la observación $t + a$.

Nuevamente, nos interesa que la función tentativa en el instante t nos permita aprovechar las muestras que estaban disponibles en el instante $t - 1$. Por conveniencia vamos a elegir una función tentativa de la forma

$$q(\mathbf{s}_{0:t} | \mathbf{y}_{0:t+a}) = q(\mathbf{s}_t | \mathbf{s}_{0:t-1}, \mathbf{y}_{0:t+a}) q(\mathbf{s}_{0:t-1} | \mathbf{y}_{0:t-1}) \quad (4.38)$$

y definir la q marginal de la que se muestrea el vector \mathbf{s}_t como

$$\begin{aligned} q(\mathbf{s}_t | \mathbf{s}_{0:t-1}, \mathbf{y}_{0:t+a}) &= p(\mathbf{s}_t | \mathbf{s}_{0:t-1}, \mathbf{y}_{0:t+a}) = \frac{p(\mathbf{y}_{t:t+a} | \mathbf{s}_{0:t}, \mathbf{y}_{0:t-1}) p(\mathbf{s}_t | \mathbf{s}_{0:t-1}, \mathbf{y}_{0:t-1})}{p(\mathbf{y}_{t:t+a} | \mathbf{s}_{0:t-1}, \mathbf{y}_{0:t-1})} \\ &\propto p(\mathbf{y}_{t:t+a} | \mathbf{s}_{0:t}, \mathbf{y}_{0:t-1}) \end{aligned} \quad (4.39)$$

donde para llegar a la proporcionalidad final hemos utilizado que $p(\mathbf{y}_{t:t+a} | \mathbf{s}_{0:t-1}, \mathbf{y}_{0:t-1})$ es una constante de normalización que no depende de \mathbf{s}_t , y que $p(\mathbf{s}_t | \mathbf{s}_{0:t-1}, \mathbf{y}_{0:t+a})$ también es una constante al ser la probabilidad *a priori* de \mathbf{s}_t .

Para calcular la verosimilitud $p(\mathbf{y}_{t:t+a} | \mathbf{s}_{0:t}, \mathbf{y}_{0:t-1})$ son necesarios los vectores de símbolos $\mathbf{s}_{t+1:t+a}$. Aplicando el teorema de la probabilidad total tenemos que

$$\begin{aligned} p(\mathbf{y}_{t:t+a} | \mathbf{s}_{0:t}, \mathbf{y}_{0:t-1}) &= \sum_{\tilde{\mathbf{s}}_{t+1:t+a} \in \mathcal{S}^{N_t a}} p(\mathbf{y}_{t:t+a}, \tilde{\mathbf{s}}_{t+1:t+a} | \mathbf{s}_{0:t}, \mathbf{y}_{0:t-1}) \\ &= \sum_{\tilde{\mathbf{s}}_{t+1:t+a} \in \mathcal{S}^{N_t a}} p(\mathbf{y}_{t:t+a} | \mathbf{s}_{0:t}, \tilde{\mathbf{s}}_{t+1:t+a}, \mathbf{y}_{0:t-1}) p(\tilde{\mathbf{s}}_{t+1:t+a} | \mathbf{s}_{0:t}, \mathbf{y}_{0:t-1}) \\ &\propto \sum_{\tilde{\mathbf{s}}_{t+1:t+a} \in \mathcal{S}^{N_t a}} p(\mathbf{y}_{t:t+a} | \mathbf{s}_{0:t}, \tilde{\mathbf{s}}_{t+1:t+a}, \mathbf{y}_{0:t-1}) \end{aligned} \quad (4.40)$$

y la verosimilitud de las observaciones desde t hasta $t+a$ se puede descomponer entonces como

$$\begin{aligned} p(\mathbf{y}_{t:t+a} | \mathbf{s}_{0:t}, \tilde{\mathbf{s}}_{t+1:t+a}, \mathbf{y}_{0:t-1}) &= p(\mathbf{y}_{t+a} | \mathbf{s}_{0:t}, \tilde{\mathbf{s}}_{t+1:t+a}, \mathbf{y}_{0:t+a-1}) p(\mathbf{y}_{t:t+a-1} | \mathbf{s}_{0:t}, \tilde{\mathbf{s}}_{t+1:t+a-1}, \mathbf{y}_{0:t-1}) \\ &= \dots \\ &= \prod_{k=0}^a p(\mathbf{y}_{t+k} | \mathbf{s}_{0:t}, \tilde{\mathbf{s}}_{t+1:t+k}, \mathbf{y}_{0:t+k-1}) \end{aligned} \quad (4.41)$$

de tal manera que, finalmente,

$$p(\mathbf{y}_{t:t+a} | \mathbf{s}_{0:t}, \mathbf{y}_{0:t-1}) \propto \sum_{\tilde{\mathbf{s}}_{t+1:t+a} \in \mathcal{S}^{N_t a}} \prod_{k=0}^a p(\mathbf{y}_{t+k} | \mathbf{s}_{0:t}, \tilde{\mathbf{s}}_{t+1:t+k}, \mathbf{y}_{0:t+k-1}). \quad (4.42)$$

Cada una de las verosimilitudes que aparecen en el producto de (4.42) se puede calcular por medio del KF utilizando las ecuaciones del Apéndice D. Nótese que para calcular $p(\mathbf{y}_{t+k} | \mathbf{s}_{0:t}, \tilde{\mathbf{s}}_{t+1:t+k}, \mathbf{y}_{0:t+k-1})$ por medio de la ecuación (D.10) hacen falta la media y la covarianza predictivas de Kalman en el instante $t+k$. Así pues, para $k > 0$ es necesario avanzar k pasos en el KF, esto es, iterar sobre las ecuaciones (D.5)-(D.8) utilizando los vectores $\mathbf{s}_t, \tilde{\mathbf{s}}_{t+1}, \dots, \tilde{\mathbf{s}}_{t+k-1}$ para obtener $\bar{\mathbf{h}}_{t+k|t+k-1}$ y $\mathbf{C}_{t+k|t+k-1}$ (respectivamente, la media y covarianza predictivas calculadas por el KF para el canal).

La función tentativa óptima para muestrear \mathbf{s}_t se obtiene normalizando (4.39) sobre todos los posibles vectores de símbolos,

$$q(\mathbf{s}_t | \mathbf{s}_{0:t-1}, \mathbf{y}_{0:t+a}) = \frac{p(\mathbf{y}_{t:t+a} | \mathbf{s}_{0:t}, \mathbf{y}_{0:t-1})}{\sum_{\tilde{\mathbf{s}}_t \in \mathcal{S}^{N_t}} p(\mathbf{y}_{t:t+a} | \mathbf{s}_{0:t-1}, \tilde{\mathbf{s}}_t, \mathbf{y}_{0:t-1})}. \quad (4.43)$$

La función objetivo, por otro lado, se puede descomponer por medio del teorema de Bayes como:

$$\begin{aligned}
p(\mathbf{s}_{0:t}|\mathbf{y}_{0:t+a}) &= \frac{p(\mathbf{y}_{t:t+a}|\mathbf{s}_{0:t}, \mathbf{y}_{0:t-1})p(\mathbf{s}_{0:t}|\mathbf{y}_{0:t-1})}{p(\mathbf{y}_{t:t+a}|\mathbf{y}_{0:t-1})} \\
&= \frac{p(\mathbf{y}_{t:t+a}|\mathbf{s}_{0:t}, \mathbf{y}_{0:t-1})p(\mathbf{s}_t|\mathbf{s}_{0:t-1}, \mathbf{y}_{0:t-1})p(\mathbf{s}_{0:t-1}|\mathbf{y}_{0:t-1})}{p(\mathbf{y}_{t:t+a}|\mathbf{y}_{0:t-1})} \\
&\propto p(\mathbf{y}_{t:t+a}|\mathbf{s}_{0:t}, \mathbf{y}_{0:t-1})p(\mathbf{s}_{0:t-1}|\mathbf{y}_{0:t-1})
\end{aligned} \tag{4.44}$$

donde la proporcionalidad se debe a que $p(\mathbf{y}_{t:t+a}|\mathbf{y}_{0:t-1})$ no depende de $\mathbf{s}_{0:t}$, y la probabilidad *a priori* de los símbolos, $p(\mathbf{s}_t|\mathbf{s}_{0:t-1}, \mathbf{y}_{0:t-1})$, es una constante.

Entonces, dada una muestra, $\mathbf{s}_{0:t-1}^{(i)}$, de la secuencia de símbolos transmitidos hasta el instante $t - 1$, el algoritmo genera una muestra

$$\mathbf{s}_t^{(i)} \sim q(\mathbf{s}_t|\mathbf{s}_{0:t-1}^{(i)}, \mathbf{y}_{0:t+a}) = \frac{p(\mathbf{y}_{t:t+a}|\mathbf{s}_{0:t}^{(i)}, \mathbf{y}_{0:t-1})}{\sum_{\tilde{\mathbf{s}}_t \in \mathcal{S}^{N_t}} p(\mathbf{y}_{t:t+a}|\mathbf{s}_{0:t-1}^{(i)}, \tilde{\mathbf{s}}_t, \mathbf{y}_{0:t-1})} \tag{4.45}$$

que se añade a $\mathbf{s}_{0:t-1}^{(i)}$ para obtener la muestra $\mathbf{s}_{0:t}^{(i)}$, a la que se asignará un peso

$$\begin{aligned}
\tilde{w}_{t+a}^{(i)} &= \frac{p(\mathbf{s}_{0:t}^{(i)}|\mathbf{y}_{0:t+a})}{q(\mathbf{s}_{0:t}^{(i)}|\mathbf{y}_{0:t+a})} \\
&\propto \frac{p(\mathbf{y}_{t:t+a}|\mathbf{s}_{0:t}^{(i)}, \mathbf{y}_{0:t-1})}{\sum_{\tilde{\mathbf{s}}_t \in \mathcal{S}^{N_t}} p(\mathbf{y}_{t:t+a}|\mathbf{s}_{0:t-1}^{(i)}, \tilde{\mathbf{s}}_t, \mathbf{y}_{0:t-1})} \frac{p(\mathbf{s}_{0:t-1}^{(i)}|\mathbf{y}_{0:t-1})}{q(\mathbf{s}_{0:t-1}^{(i)}|\mathbf{y}_{0:t-1})} \\
&= w_{t-1}^{(i)} \sum_{\tilde{\mathbf{s}}_t \in \mathcal{S}^{N_t}} p(\mathbf{y}_{t:t+a}|\mathbf{s}_{0:t-1}^{(i)}, \tilde{\mathbf{s}}_t, \mathbf{y}_{0:t-1}).
\end{aligned} \tag{4.46}$$

El Pseudocódigo 5 muestra un resumen del algoritmo.

Obsérvese que, una vez más, al utilizar como función tentativa la propia f.d.p. a aproximar, las muestras son i.i.d. y el peso se actualiza independientemente del vector $\mathbf{s}_t^{(i)}$ muestreado.

El mayor problema de la función tentativa óptima con suavizado es la complejidad. Observando la función tentativa dada por la ecuación (4.43) podemos deducir que la carga computacional crece exponencialmente con en el número de antenas transmisoras, N_t , por un lado, y con el factor de suavizado, a , por el otro (esto último es debido a que el cálculo de la probabilidad $p(\mathbf{y}_{t:t+a}|\mathbf{s}_{0:t}, \mathbf{y}_{0:t-1})$ implica, a la vista de la ecuación (4.42), un sumatorio sobre todas las posibles secuencias de vectores $\mathbf{s}_{t+1:t+a} \in \mathcal{S}^{N_t a}$). Esta complejidad, $\mathcal{O}(\mathcal{S}^{N_t(a+1)})$, limita enormemente el uso práctico de este algoritmo.

4.2.3. Detector M estocástico

En [77] se presenta un método de filtrado de partículas para canales MIMO planos basado en la función tentativa óptima, que opera de manera diferente al expuesto en la Sección 4.2.1.

Pseudocódigo 5 Algoritmo SIS óptimo suavizado

para cada instante de tiempo $t = 1, 2, \dots$
para cada partícula $i = 1, \dots, M$
 sea $acum_{verosimil} \leftarrow 0$
para todo $\mathbf{s}_t \in \mathcal{S}^{N_t}$
 $verosimil(\mathbf{s}_t) \leftarrow 0$
 $P_o \leftarrow p(\mathbf{y}_t | \mathbf{s}_{0:t}, \mathbf{y}_{0:t-1})$
 guardar el estado del i -ésimo KF (media y covarianza predictivas y filtradas)
para todo $\tilde{\mathbf{s}}_{t+1:t+a} \in \mathcal{S}^{N_t d}$
 sea $P \leftarrow P_o$
para $k = 1, \dots, a$
 actualizar el i -ésimo KF con el vector \mathbf{s}_{t+k-1}
 calcular $P \leftarrow P \times p(\mathbf{y}_{t+k} | \mathbf{s}_{0:t}, \tilde{\mathbf{s}}_{t+1:t+k}, \mathbf{y}_{0:t+k-1})$
 $verosimil(\mathbf{s}_t) \leftarrow verosimil(\mathbf{s}_t) + P$
 restaurar el estado del i -ésimo KF
 $acum_{verosimil} \leftarrow acum_{verosimil} + verosimil(\mathbf{s}_t)$
 muestrear $\mathbf{s}_t^{(i)} \sim q_t(\mathbf{s}_t) = verosimil(\mathbf{s}_t) / acum_{verosimil}$
 obtener el peso sin normalizar $\tilde{w}_t^{(i)} \leftarrow w_{t-1}^{(i)} acum_{verosimil}$
 actualizar el i -ésimo KF utilizando la muestra $\mathbf{s}_t^{(i)}$
 calcular $W \leftarrow \sum_{i=1}^M \tilde{w}_t^{(i)}$
para cada partícula $i = 1, \dots, M$
 normalizar su peso $w_t^{(i)} \leftarrow \tilde{w}_t^{(i)} / W$
 estimar el tamaño efectivo de la muestra, $\hat{M}_{ef} \leftarrow \left(\sum_{i=1}^M w_t^{(i)2} \right)^{-1}$
si $\hat{M}_{ef} < \gamma_r M$ **entonces**
 remuestrear

En dicho trabajo se asume que el canal es conocido y el filtro de partículas se utiliza únicamente para detectar los símbolos transmitidos. Así pues, las muestras se toman en un espacio discreto, y una muestra en el instante $t - 1$ sólo puede dar lugar mediante SIS a un número limitado de muestras en el instante t . En nuestro caso, dada una muestra, $\mathbf{s}_{0:t-1}^{(i)}$, de la secuencia de símbolos transmitidos desde 0 hasta $t - 1$, la muestra obtenida para el instante t siguiendo la metodología SIS va a estar forzosamente en el conjunto $\left\{ \mathbf{s}_{0:t-1}^{(i)}, \tilde{\mathbf{s}}_t \right\}$, $\tilde{\mathbf{s}}_t \in \mathcal{S}^{N_t}$, de cardinalidad $|\mathcal{S}^{N_t}|$. Si en el instante $t - 1$ tenemos varias partículas cuya muestras (y pesos, por tanto) son iguales (lo cual ocurrirá con frecuencia debido a los pasos de remuestreo), la función tentativa de la que todas ellas van a muestrear el vector \mathbf{s}_t es la misma, puesto que, a la vista de la ecuación (4.34), ésta únicamente depende de la secuencia de símbolos ya muestreados, $\mathbf{s}_{0:t-1}$, y de las observaciones $\mathbf{y}_{0:t}$. Si dicha función tentativa condensa su masa de probabilidad en unos pocos vectores de símbolos de entre todos los posibles, muchas de esas partículas obtendrán la misma muestra para el vector \mathbf{s}_t y seguirán siendo idénticas entre sí en el instante t . Esto supone un gasto innecesario de recursos computacionales ya que una misma partícula está siendo actualizada varias veces.

El esquema en [77], además de evitar el problema arriba mencionado, introduce una mejora adicional. Supongamos que en el instante $t - 1$ disponemos de M partículas diferentes $\left\{ \mathbf{s}_{0:t-1}^{(i)}, w_{t-1}^{(i)} \right\}_{i=1}^M$ que aproximan la f.d.p. $p(\mathbf{s}_{0:t-1} | \mathbf{y}_{0:t-1})$. Para propagar una trayectoria cualquiera, $\mathbf{s}_{0:t-1}^{(i)}$, es necesario calcular la probabilidad de cada posible $\mathbf{s}_t \in \mathcal{S}^{N_t}$ dadas las observaciones $\mathbf{y}_{0:t}$ cuando la secuencia de vectores de símbolos anteriormente transmitidos es $\mathbf{s}_{0:t-1}^{(i)}$. A partir de dicha masa de probabilidad, el detector óptimo de la Sección 4.2.1 toma una muestra de \mathbf{s}_t para obtener la nueva trayectoria $\mathbf{s}_{0:t}^{(i)}$. Es decir, cada muestra en el instante $t - 1$, independientemente del peso que tenga, da lugar a exactamente una muestra en el instante t (se observa claramente en la Figura 2.9, donde cada flecha continua representa la propagación de una partícula). Los autores de [77] sugieren permitir que una muestra en el instante $t - 1$ pueda dar lugar a varias en el instante t . La idea es muestrear, en el instante t , M trayectorias de entre las $M |\mathcal{S}^{N_t}|$ que pueden surgir de propagar las trayectorias que había en el instante $t - 1$ (cada trayectoria del instante $t - 1$ se puede extender con tantos vectores de símbolos como haya en \mathcal{S}^{N_t}). Para ello, primeramente se calculan mediante (4.35) los pesos sin normalizar de las trayectorias $\left\{ \mathbf{s}_{0:t-1}^{(i)}, \tilde{\mathbf{s}}_t \right\}_{i=1}^M$ con $\tilde{\mathbf{s}}_t \in \mathcal{S}^{N_t}$. Entonces, y una vez normalizados éstos, el algoritmo de remuestreo óptimo descrito en [78] permite obtener un conjunto de M trayectorias diferentes que, en el instante t , aproximan la f.d.p. $p(\mathbf{s}_{0:t} | \mathbf{y}_{0:t})$.

Aunque el algoritmo propuesto es estocástico puesto que en él tiene lugar un proceso de muestreo, no es un algoritmo de filtrado de partículas en un sentido tradicional, ya que las muestras que aproximan la función objetivo no se generan haciendo uso de una función tentativa, sino que se obtienen explorando exhaustivamente el espacio de estados.

4.2.4. Filtrado de partículas con estructura de árbol

También está basado en la función tentativa óptima el método de filtrado de partículas propuesto en [26] para realizar detección de máxima verosimilitud en canales MIMO planos. Al igual que en [77], el canal se supone conocido y sólo se van a tomar muestras de los símbolos transmitidos, es decir, el espacio muestral es finito. Tomando esta observación como punto de partida, los autores presentan un algoritmo de búsqueda en árbol donde los nodos del mismo son grupos de partículas idénticas (misma secuencia detectada y mismo peso) y donde la función tentativa óptima se utiliza para generar los nodos que van a descender de uno dado.

Para poder detectar los símbolos transmitidos mediante una búsqueda en árbol, primeramente es necesario llevar a cabo una transformación sobre las observaciones recibidas (que vienen determinadas por la ecuación (3.2) en el modelo de señal que nos ocupa) de tal manera que éstas se puedan escribir como el producto de una matriz triangular por los símbolos desconocidos más un ruido: es lo que se denomina triangularizar la matriz de canal. La descomposición QR que proponen los autores de [26] requiere que las columnas de la matriz que se va a descomponer sean linealmente independientes [64], lo que implica que el número de filas ha de ser mayor o igual al de columnas. Esto no lo verifica la matriz de canal, \mathbf{H}_t , de nuestra ecuación de observación. Podemos recurrir entonces al modelo apilado de la ecuación (3.13) y aplicar la descomposición QR para obtener $\bar{\mathbf{H}}_{t,a} = \mathbf{Q}\mathbf{U}$, donde \mathbf{Q} es una matriz $N_r(a+1) \times N_t(a+1)$ tal que

$$\mathbf{Q}^\top \mathbf{Q} = \mathbf{I}_{N_t(a+1)} \quad (4.47)$$

y \mathbf{U} es una matriz $N_t(a+1) \times N_t(a+1)$ triangular superior. Multiplicando (3.13) por \mathbf{Q}^\top resulta

$$\begin{aligned} \mathbf{z} &= \mathbf{Q}^\top \bar{\mathbf{H}}_{t,a} \mathbf{s}_{t+a} + \mathbf{Q}^\top \mathbf{g}_{t+a} = \mathbf{Q}^\top \mathbf{Q} \mathbf{U} \mathbf{s}_{t+a} + \mathbf{Q}^\top \mathbf{g}_{t+a} \\ &= \mathbf{U} \mathbf{s}_{t+a} + \mathbf{Q}^\top \mathbf{g}_{t+a} \\ &= \mathbf{U} \mathbf{s}_{t+a} + \mathbf{g}'_{t+a} \end{aligned} \quad (4.48)$$

donde $\mathbf{z} = \mathbf{Q}^\top \bar{\mathbf{y}}_{t+a}$ y $\mathbf{g}'_{t+a} = \mathbf{Q}^\top \mathbf{g}_{t+a}$ son dos vectores de dimensiones $N_t(a+1) \times 1$. El primero de ellos es fácilmente calculable pues tanto \mathbf{Q}^\top como $\bar{\mathbf{y}}_{t+a}$ son conocidos. El segundo es un vector de ruido cuya media y covarianza se obtienen a partir de la media y covarianza de \mathbf{g}_{t+a} .

Para simplificar la notación, en lo que resta de sección vamos a denotar por N_s al número de símbolos del vector \mathbf{s}_{t+a} (que es también el de elementos de los vectores \mathbf{z} y \mathbf{g}'_{t+a}). Téngase presente que $N_s = N_t(a+1)$.

Podemos interpretar (4.48) como una ecuación de observación en la que la matriz de canal es triangular superior. Vamos a hacer explícita su estructura reescribiendo la ecuación (4.48)

como

$$\begin{bmatrix} z(1) \\ z(2) \\ \vdots \\ z(N_s) \end{bmatrix} = \begin{bmatrix} u(1,1) & u(1,2) & \cdots & u(1,N_s) \\ 0 & u(2,2) & \cdots & u(2,N_s) \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & u(N_s,N_s) \end{bmatrix} \begin{bmatrix} s_{t+a}^t(1) \\ s_{t+a}^t(2) \\ \vdots \\ s_{t+a}^t(N_s) \end{bmatrix} + \begin{bmatrix} \acute{g}_{t+a}^t(1) \\ \acute{g}_{t+a}^t(2) \\ \vdots \\ \acute{g}_{t+a}^t(N_s) \end{bmatrix}. \quad (4.49)$$

A la vista de la ecuación (4.49), es trivial obtener una estimación de $s_{t+a}^t(N_s)$, pues el último elemento de \mathbf{z} sólo depende de él. Para detectar, a continuación, $s_{t+a}^t(N_s - 1)$, obsérvese que en $z(N_s - 1)$ solo intervienen el símbolo anteriormente detectado y el nuevo símbolo que queremos detectar. Entonces, utilizando la estimación de $s_{t+a}^t(N_s)$, podemos detectar fácilmente $s_{t+a}^t(N_s - 1)$. En general, el i -ésimo elemento del vector de observaciones de la ecuación (4.49) se puede escribir como

$$z(i) = \sum_{j=i}^{N_s} u(i,j) s_{t+a}^t(j) + \acute{g}_{t+a}^t(i) \quad (4.50)$$

y es fácil obtener una estimación del i -ésimo símbolo de \mathbf{s}_{t+a}^t una vez detectados todos los que le siguen. Nótese que en este caso estaríamos tratando de aproximar en cada instante de tiempo una trayectoria en el espacio: cada nuevo elemento que se añade a la misma es un símbolo transmitido por una antena diferente en el mismo instante de tiempo.

Los elementos de \mathbf{s}_{t+a}^t sólo pueden tomar valores en el alfabeto \mathcal{S} de tal manera que a la hora de detectar $s_{t+a}^t(N_s)$ tenemos $|\mathcal{S}|$ posibilidades, cada una de las cuales da lugar a su vez a $|\mathcal{S}|$ posibilidades para el símbolo $s_{t+a}^t(N_s - 1)$, y esta estructura se repite hasta llegar al primer elemento de \mathbf{s}_{t+a}^t . Así pues, se puede llevar a cabo el proceso de detección como si de una búsqueda en árbol se tratara. Es lo que hacen los autores de [26] mediante métodos SMC. Supongamos que el número de partículas empleado por el algoritmo es M . En un primer momento todas forman parte de un mismo grupo que ocupa la raíz del árbol y sus pesos se inicializan a $1/M$. Calculamos entonces las probabilidades $p(s_{t+a}^t(N_s) = s | z(N_s))$ para todo $s \in \mathcal{S}$ y las utilizamos para decidir como se distribuyen las partículas en sus nodos hijo: $p(s_{t+a}^t(N_s) = s | z(N_s))$ es el porcentaje de partículas del nodo padre (M en este primer paso) que actualizarán su trayectoria con el símbolo s y pasarán a formar parte del nodo hijo en el que $s_{t+a}^t(N_s) = s$. En general, las partículas de un grupo en el i -ésimo nivel del árbol tienen asociada una (única) estimación de los $i - 1$ últimos símbolos del vector \mathbf{s}_{t+a}^t . A partir de ella se calculan las probabilidades

$$p(s_{t+a}^t(N_s - i) = s | z(N_s - i), s_{t+a}^t(N_s - i + 1), \cdots, s_{t+a}^t(N_s)), s \in \mathcal{S}$$

de acuerdo a las cuales se distribuyen las partículas del grupo entre sus nodos hijo. Si alguno de los nuevos nodos resultantes no contiene ninguna partícula, su camino correspondiente a través del árbol de partículas se elimina.

En cuanto al remuestreo, éste se produce cuando el número de grupos con una partícula es mayor que el 35 % de M y el número total de grupos es mayor que el 50 % de M .

Obsérvese que, al igual que en el esquema de la Sección 4.2.3, las muestras que aproximan la f.d.p. objetivo no se generan utilizando una función tentativa, lo que hace que estemos una vez más ante un método de filtrado de partículas atípico.

4.3. Conclusiones

En este capítulo hemos revisado algunas de las técnicas existentes para la igualación de canales MIMO. Las hemos clasificado en dos grupos atendiendo a si hacen uso o no de la metodología SMC. En el primero se encuentran los receptores lineales, los receptores realimentados y algunos algoritmos clásicos que tratan de maximizar la verosimilitud para encontrar la secuencia más probablemente transmitida, como el algoritmo de Viterbi, el PSP o el EM. Dentro del segundo grupo hemos incluido los algoritmos de filtrado de partículas que utilizan la función tentativa óptima (con y sin suavizado), y dos aproximaciones también basadas en métodos SMC y en la función tentativa óptima, pero que presentan ciertas particularidades que los diferencian claramente de los anteriores.

Capítulo 5

Nuevos métodos SMC de complejidad reducida para la igualación de canales MIMO

En el capítulo anterior se explicó como aplicar el algoritmo de Viterbi y el algoritmo BCJR al problema de la detección de los símbolos transmitidos a través de un canal MIMO. El primero minimiza la probabilidad de error de secuencia y el segundo la de error de símbolo pero ambos tienen el inconveniente de que su complejidad crece exponencialmente con el número de antenas transmisoras, N_t . En el presente capítulo proponemos varios métodos de filtrado de partículas que resuelven el problema de igualación de un canal MIMO con una complejidad polinómica en N_t sin que se produzca una degradación significativa en el rendimiento.

Una característica inherente a cualquier método de filtrado de partículas es que la aproximación que hace de una f.d.p. es tanto mejor cuanto mayor es el número de partículas empleadas. Obviamente más partículas implican más cálculos y la complejidad de este tipo de algoritmos va a depender invariablemente del número de partículas generadas. Sin embargo, esta dependencia es lineal y dado que la trayectoria y el peso de cada partícula se actualizan de manera independiente, es posible paralelizar un filtro de partículas repartiendo éstas entre distintos elementos de procesado, que únicamente se comunican entre sí cuando se hace necesario un paso de remuestreo [50].

Por otro lado, tal como se explicó en el Capítulo 2, el principio de muestreo enfatizado secuencial evita que la complejidad asociada a obtener una muestra de $p(\mathbf{s}_{0:t}|\mathbf{y}_{0:t})$ se incremente con el tiempo. En última instancia, en el instante t queremos obtener una muestra, $\mathbf{s}_t^{(i)}$, de la distribución $q(\mathbf{s}_t|\mathbf{s}_{0:t-1}, \mathbf{y}_{0:t})$, que añadida a $\mathbf{s}_{0:t-1}^{(i)}$ de lugar a una muestra $\mathbf{s}_{0:t}^{(i)}$ de $p(\mathbf{s}_{0:t}|\mathbf{y}_{0:t})$. Así pues, la complejidad de los métodos SMC viene dada fundamentalmente por el coste computacional que supone tomar una muestra, $\mathbf{s}_t^{(i)}$, de la función tentativa marginal $q(\mathbf{s}_t|\mathbf{s}_{0:t-1}, \mathbf{y}_{0:t})$.

Todos los métodos de filtrado de partículas vistos en el capítulo anterior para la igualación de canales MIMO (salvo el de “filtrado de partículas con estructura de árbol” de la Sección

4.2.4¹) tienen una complejidad tal que el coste asociado a generar una muestra $\mathbf{s}_t^{(i)} \sim q(\mathbf{s}_t | \mathbf{s}_{0:t-1}, \mathbf{y}_{0:t})$ depende exponencialmente del tamaño del vector \mathbf{s}_t , es decir, del número de antenas transmisoras, N_t . En este capítulo se presentan nuevos esquemas de filtrado de partículas que evitan dicha complejidad exponencial mediante un adecuado diseño de la función tentativa. Todos ellos son métodos de suavizado y aproximan recursivamente la f.d.p. $p(\mathbf{s}_t | \mathbf{y}_{0:t+a})$ (donde a es el factor de suavizado) para obtener un detector MAP de los símbolos transmitidos. Además, vamos a asumir siempre que el canal es desconocido, por lo que debe ser estimado de alguna manera, ya sea implícita o explícita.

Se ha dividido el capítulo en dos secciones. La primera se centra en esquemas de igualación basados en métodos SMC que utilizan el canal como variable auxiliar para facilitar el muestreo de los símbolos transmitidos (nótese que el objetivo del receptor es la detección fiable de los datos y no la estimación del canal *per se*). Aunque la complejidad de los métodos incluidos en dicha sección es ya polinómica en el número de antenas transmisoras, su uso conlleva la utilización de filtros de Kalman para integrar el canal desconocido, y la complejidad de éstos puede ser prohibitiva en algunas aplicaciones donde los recursos computacionales son particularmente limitados. En la segunda sección se rediseñan algunas de las funciones tentativas propuestas con el objeto de evitar el uso de filtros de Kalman y, de esa manera, simplificar los receptores correspondientes.

5.1. Métodos SMC basados en variables auxiliares

Para desarrollar el método propuesto en esta sección comenzaremos estudiando la descomposición de las f.d.p. que permiten diseñar el filtro de partículas. Como resultado obtendremos una serie de relaciones de las que derivaremos el esquema de muestreo y la ecuación de actualización de los pesos.

Cuando el canal es conocido, generar muestras de la f.d.p. *a posteriori* de la secuencia de los símbolos, $p(\mathbf{s}_{0:t} | \mathbf{H}_t, \mathbf{y}_{0:t+a})$, es relativamente sencillo. Cuando el canal es desconocido, una posibilidad es generar una muestra del mismo, utilizarla en la detección de los símbolos y posteriormente descartarla. En tal caso los coeficientes del canal hacen el papel de variables auxiliares [79], lo que implica muestrear en un espacio de dimensión mayor, pero secuencialmente en el espacio de los datos [29]. Así pues, y teniendo en cuenta que estamos intentando derivar un método de suavizado, nuestra f.d.p. objetivo podría ser $p(\mathbf{s}_{0:t}, \mathbf{H}_t | \mathbf{y}_{0:t+a})$, que es la probabilidad conjunta de los símbolos desde 0 hasta t y la matriz de canal \mathbf{H}_t dada la secuencia de observaciones hasta $t + a$. Sin embargo, esta densidad de probabilidad es difícilmente evaluable² si no conocemos los símbolos $\mathbf{s}_{t+1:t+a}$: aplicando el teorema de Bayes

¹Nótese, sin embargo, que este método asume que el canal es plano y perfectamente conocido.

²Recuérdese del Capítulo 2 que para calcular el peso asociado a una muestra es necesario evaluar la f.d.p. objetivo.

tenemos que

$$\begin{aligned} p(\mathbf{s}_{0:t}, \mathbf{H}_t | \mathbf{y}_{0:t+a}) &= \frac{p(\mathbf{y}_{t+a} | \mathbf{s}_{0:t}, \mathbf{H}_t, \mathbf{y}_{0:t+a-1}) p(\mathbf{s}_{0:t}, \mathbf{H}_t | \mathbf{y}_{0:t+a-1})}{p(\mathbf{y}_{t+a} | \mathbf{y}_{0:t+a-1})} \\ &\propto p(\mathbf{y}_{t+a} | \mathbf{s}_{0:t}, \mathbf{H}_t, \mathbf{y}_{0:t+a-1}) p(\mathbf{s}_{0:t}, \mathbf{H}_t | \mathbf{y}_{0:t+a-1}) \end{aligned} \quad (5.1)$$

donde el factor $p(\mathbf{s}_{0:t}, \mathbf{H}_t | \mathbf{y}_{0:t+a-1})$ es la probabilidad que estábamos tratando de aproximar desde un principio, pero ahora utilizando únicamente las observaciones hasta $t+a-1$. Entonces, aplicando repetidas veces el teorema de Bayes se puede llegar a

$$p(\mathbf{s}_{0:t}, \mathbf{H}_t | \mathbf{y}_{0:t+a}) \propto p(\mathbf{s}_{0:t-1}, \mathbf{H}_t | \mathbf{y}_{0:t-1}) \prod_{k=0}^a p(\mathbf{y}_{t+k} | \mathbf{s}_{0:t}, \mathbf{H}_t, \mathbf{y}_{0:t+k-1}). \quad (5.2)$$

La observación \mathbf{y}_{t+j} sigue una distribución normal de media $\mathbf{H}_{t+j} \mathbf{s}_{t+j-m+1}$ y covarianza $\sigma_g^2 \mathbf{I}_{N_r}$. Así pues, para $k=0$ el término $p(\mathbf{y}_{t+k} | \mathbf{s}_{0:t}, \mathbf{H}_t, \mathbf{y}_{0:t+k-1})$ se puede calcular fácilmente pues tanto su media, $\mathbf{H}_t \mathbf{s}_{t-m+1}$, como su matriz de covarianza son conocidas. Para $k > 0$, sin embargo, no disponemos de la $(t+j)$ -ésima matriz de canal, \mathbf{H}_{t+j} , ni de los símbolos $\mathbf{s}_{t+1:t+j}$, necesarios para construir el vector $\mathbf{s}_{t+j-m+1}$. Para evitar este problema nos planteamos la aproximación de la f.d.p. $p(\mathbf{s}_{0:t+a}, \mathbf{H}_{t:t+a} | \mathbf{y}_{0:t+a})$, que se puede descomponer como

$$\begin{aligned} p(\mathbf{s}_{0:t+a}, \mathbf{H}_{t:t+a} | \mathbf{y}_{0:t+a}) &\propto \prod_{k=0}^a p(\mathbf{y}_{t+k} | \mathbf{H}_{t+k}, \mathbf{s}_{t+k-m+1:t+k}) \prod_{l=1}^a p(\mathbf{H}_{t+l} | \mathbf{H}_{t+l-1}) \times \\ &\times p(\mathbf{H}_t | \mathbf{s}_{0:t-1}, \mathbf{y}_{0:t-1}) p(\mathbf{s}_{0:t-1} | \mathbf{y}_{0:t-1}). \end{aligned} \quad (5.3)$$

En lo que sigue vamos a utilizar una representación vectorial, \mathbf{h}_t , de la matriz de canal, \mathbf{H}_t , que nos permite modelar ésta como un vector aleatorio con su media y su covarianza. Se construye apilando verticalmente las columnas de \mathbf{H}_t , es decir,

$$\mathbf{h}_t = \begin{bmatrix} h_t(1, 1) \\ \vdots \\ h_t(N_r, 1) \\ \vdots \\ h_t(1, N_t N_r m) \\ \vdots \\ h_t(N_r, N_t N_r m) \end{bmatrix} \quad (5.4)$$

donde $h_t(i, j)$ es el elemento que ocupa la i -ésima fila y j -ésima columna de \mathbf{H}_t . Así pues, tanto \mathbf{h}_t como \mathbf{H}_t representan la información del estado del canal en el instante t y emplearemos una notación u otra según nos convenga.

Utilizando esta nueva representación del canal tenemos que todos los factores que aparecen en (5.3) son gaussianos y se pueden calcular fácilmente:

$$p(\mathbf{y}_{t+k} | \mathbf{H}_{t+k}, \mathbf{s}_{t+k-m+1:t+k}) \sim \mathcal{N} \left(\mathbf{H}_{t+k} \mathbf{s}_{t+k-m+1:t+k}, \sigma_g^2 \mathbf{I}_{N_r} \right) \quad (5.5)$$

$$p(\mathbf{H}_{t+l} | \mathbf{H}_{t+l-1}) = p(\mathbf{h}_{t+l} | \mathbf{h}_{t+l-1}) \sim \mathcal{N} (\alpha \mathbf{h}_{t+l-1}, \sigma_v^2 \mathbf{I}_{N_r N_t m}) \quad (5.6)$$

$$p(\mathbf{H}_t | \mathbf{s}_{0:t-1}, \mathbf{y}_{0:t-1}) = p(\mathbf{h}_t | \mathbf{s}_{0:t-1}, \mathbf{y}_{0:t-1}) \sim \mathcal{N} (\bar{\mathbf{h}}_{t|t-1}, \mathbf{C}_{t|t-1}) \quad (5.7)$$

siendo $\bar{\mathbf{h}}_{t|t-1}$ y $\mathbf{C}_{t|t-1}$, respectivamente, la media y covarianza predictivas de Kalman, es decir, la media y covarianza del vector de coeficientes del canal estimadas por el filtro de Kalman en el instante t dadas las observaciones y la secuencia de símbolos hasta el instante anterior.

La función tentativa que vamos a utilizar para aproximar $p(\mathbf{s}_{0:t+a}, \mathbf{H}_{t:t+a} | \mathbf{y}_{0:t+a})$ tiene la forma

$$q(\mathbf{s}_{0:t+a}, \mathbf{H}_{t:t+a} | \mathbf{y}_{0:t+a}) = q(\mathbf{s}_{t:t+a}, \mathbf{H}_{t:t+a} | \mathbf{y}_{0:t+a}, \mathbf{s}_{0:t-1}) q(\mathbf{s}_{0:t-1} | \mathbf{y}_{0:t-1}) \quad (5.8)$$

y utilizando las ecuaciones (5.3) y (5.8) resulta el siguiente método de filtrado de partículas

$$\left(\mathbf{s}_{t:t+a}^{(i)}, \mathbf{H}_{t:t+a}^{(i)} \right) \sim q(\mathbf{s}_{t:t+a}, \mathbf{H}_{t:t+a} | \mathbf{s}_{0:t-1}, \mathbf{y}_{0:t+a}) \quad (5.9)$$

$$w_{t+a}^{(i)} = w_{t+a-1}^{(i)} \frac{\prod_{k=0}^a p(\mathbf{y}_{t+k} | \mathbf{H}_{t+k}^{(i)}, \mathbf{s}_{t+k-m+1:t+k}^{(i)}) p(\mathbf{H}_{t+k}^{(i)} | \mathbf{s}_{0:t-1}^{(i)}, \mathbf{y}_{0:t-1}) \prod_{l=1}^a p(\mathbf{H}_{t+l}^{(i)} | \mathbf{H}_{t+l-1}^{(i)})}{q(\mathbf{s}_{t:t+a}^{(i)}, \mathbf{H}_{t:t+a}^{(i)} | \mathbf{y}_{0:t+a}, \mathbf{s}_{0:t-1}^{(i)})} \quad (5.10)$$

con $i = 1, \dots, M$ e intercalando pasos de remuestreo cuando sea necesario.

Las ecuaciones (5.9) y (5.10) definen un esquema de filtrado de partículas con variables auxiliares [79], siendo éstas los vectores de símbolos $\mathbf{s}_{t+1:t+a}^{(i)}$ y las matrices de canal $\mathbf{H}_{t:t+a}^{(i)}$. Las muestras de las variables auxiliares sirven únicamente para facilitar el muestreo de las variables de interés (en este caso los símbolos $\mathbf{s}_{0:t}^{(i)}$) y se descartan³ una vez obtenidas éstas para tener una medida de probabilidad discreta que aproxime $p(\mathbf{s}_{0:t} | \mathbf{y}_{0:t+a})$.

Vamos a presentar dos métodos secuenciales de Monte Carlo que hacen uso de las ecuaciones (5.9) y (5.10) y únicamente difieren en la forma de definir la función tentativa $q(\mathbf{s}_{t:t+a}, \mathbf{H}_{t:t+a} | \mathbf{s}_{0:t-1}, \mathbf{y}_{0:t+a})$. En ambos casos comenzaremos factorizando ésta como

$$q(\mathbf{s}_{t:t+a}, \mathbf{h}_{t:t+a} | \mathbf{y}_{0:t+a}, \mathbf{s}_{0:t-1}) = p(\mathbf{h}_t | \mathbf{s}_{0:t-1}, \mathbf{y}_{0:t-1}) \prod_{k=1}^a p(\mathbf{h}_{t+k} | \mathbf{h}_{t+k-1}) \times \\ \times q_{t+a}(\mathbf{s}_{t:t+a} | \mathbf{h}_{t:t+a}, \mathbf{y}_{0:t+a}, \mathbf{s}_{0:t-1}). \quad (5.11)$$

La ecuación (5.11) implica tomar muestras del canal desde el instante t hasta el $t+a$ y luego utilizarlas para obtener muestras de los vectores de símbolos $\mathbf{s}_{t:t+a}$. Así pues, el proceso de muestreo consta de dos fases: el muestreo del canal y el muestreo de los datos. Se describen en detalle en las secciones siguientes.

³Nótese que dadas dos v.a., X e Y , y una muestra de la distribución conjunta $(x, y) \sim (X, Y)$, descartando y obtenemos una muestra $x \sim X$.

5.1.1. Muestreo del canal

Atendiendo a la ecuación (5.11), $\mathbf{h}_t^{(i)}$ es una muestra de la distribución (gaussiana) $p(\mathbf{h}_t^{(i)} | \mathbf{s}_{0:t-1}^{(i)}, \mathbf{y}_{0:t-1})$, cuya media y covarianza pueden ser calculadas por medio de un filtro de Kalman operando en la partícula i . Vamos a denotar por $\bar{\mathbf{h}}_{t|t-1}^{(i)}$ y $\mathbf{C}_{t|t-1}^{(i)}$ a la media y covarianza, respectivamente, estimadas por el filtro de Kalman cuando la secuencia de símbolos transmitida es $\mathbf{s}_{0:t-1}^{(i)}$ (la correspondiente a la i -ésima partícula). Cada partícula va a tener asociado, por tanto, un KF diferente que se irá actualizando a medida que nuevos vectores de símbolos muestreados se añadan a la secuencia de la misma.

A partir de una muestra $\mathbf{h}_t^{(i)}$ de \mathbf{h}_t , podemos obtener muestras de los vectores (o matrices) de canal $\mathbf{h}_{t+1:t+a}$ (o $\mathbf{H}_{t+1:t+a}$) simplemente utilizando la ecuación (3.5) del sistema dinámico que define nuestro modelo de señal. Así, una muestra del canal para la i -ésima partícula en el instante $t+k$ se obtiene a partir de la muestra del canal tomada para esa misma partícula en el instante anterior, es decir, $p(\mathbf{h}_{t+k} | \mathbf{h}_{t+k-1}^{(i)}, \sigma_v^2)$. En última instancia, esto supone predecir el estado del canal en el instante $t+k$ utilizando su f.d.p. *a priori*.

5.1.2. Muestreo de los datos

Una vez disponemos de muestras de las matrices de canal, podemos utilizarlas para suprimir la interferencia intersimbólica causal, tal como se explicó en la Sección 3.1.1. Para ello vamos a construir una matriz $\check{\mathbf{H}}_{t,a}^{(i)}$ de la forma indicada en (3.12), pero en este caso utilizando las muestras $\mathbf{H}_{t:t+a}^{(i)}$ en lugar de las auténticas matrices de canal, es decir,

$$\check{\mathbf{H}}_{t,a}^{(i)} = \begin{bmatrix} \mathbf{H}_t^{(i)}(m-1) & \mathbf{H}_t^{(i)}(m-2) & \dots & \mathbf{H}_t^{(i)}(1) \\ \mathbf{0} & \mathbf{H}_{t+1}^{(i)}(m-1) & \dots & \mathbf{H}_{t+1}^{(i)}(2) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{H}_{t+a-1}^{(i)}(m-1) \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \end{bmatrix}. \quad (5.12)$$

Un problema adicional que surge es que desconocemos los vectores de símbolos $\mathbf{s}_{t-m+1:t-1}$ y es necesario conocerlos para restar su contribución de las observaciones. En su lugar vamos a utilizar las muestras asociadas a la partícula correspondiente, a saber, $\mathbf{s}_{t-m+1:t-1}^{(i)}$ para el caso de la i -ésima partícula. Utilizando entonces las muestras de los símbolos desde $t-m+1$ hasta $t-1$ y la matriz $\check{\mathbf{H}}_{t,a}^{(i)}$ (construida a partir de las muestras $\mathbf{H}_{t:t+a}^{(i)}$) podemos obtener un vector de observaciones transformadas,

$$\bar{\mathbf{y}}_{t+a}^{(i)} = \mathbf{y}_{t+a} - \check{\mathbf{H}}_{t,a}^{(i)} \mathbf{s}_{t-m+1}^{(i)}. \quad (5.13)$$

Éste es un vector en el que se ha suprimido la interferencia causal suponiendo que

- la secuencia de símbolos muestreados por la partícula entre los instantes $t-m+1$ y $t-1$ se corresponde con la secuencia de símbolos efectivamente transmitida entre dichos instantes, es decir, $\mathbf{s}_{t-m+1:t-1}^{(i)} = \mathbf{s}_{t-m+1:t-1}$, y que

- las muestras $\mathbf{H}_{t:t+a}^{(i)}$ constituyen la información del estado del canal entre los instantes t y $t + a$ o, equivalentemente, $\mathbf{H}_{t:t+a}^{(i)} = \mathbf{H}_{t:t+a}$.

Para cada partícula calcularemos un vector de observaciones transformadas diferente puesto que cada una tiene su propia secuencia de vectores de símbolos anteriormente transmitidos ($\mathbf{s}_{t-m+1:t-1}^{(i)}$) y, sobre todo⁴, para cada una vamos a generar (como variables auxiliares) diferentes muestras del canal, $\mathbf{H}_{t:t+a}^{(i)}$.

El vector de observaciones transformadas únicamente involucra a los vectores de símbolos $\mathbf{s}_{t:t+a}$, por lo que para la i -ésima partícula podemos escribir

$$\bar{\mathbf{y}}_{t+a}^{(i)} = \bar{\mathbf{H}}_{t,a}^{(i)} \mathbf{s}_{t+a} + \mathbf{g}_{t+a} \quad (5.14)$$

donde la matriz $\bar{\mathbf{H}}_{t,a}^{(i)}$ es análoga a la matriz $\bar{\mathbf{H}}_{t,a}$ dada por la ecuación (3.14), pero construida no a partir de las matrices de canal $\mathbf{H}_{t:t+a}$ sino a partir de sus correspondientes muestras $\mathbf{H}_{t:t+a}^{(i)}$, es decir,

$$\bar{\mathbf{H}}_{t,a}^{(i)} = \begin{bmatrix} \mathbf{H}_t^{(i)}(0) & \mathbf{0} & \cdots & \cdots & \mathbf{0} \\ \mathbf{H}_{t+1}^{(i)}(1) & \mathbf{H}_{t+1}^{(i)}(0) & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{H}_{t+a}^{(i)}(m-1) & \mathbf{H}_{t+a}^{(i)}(m-2) & \cdots & \cdots & \mathbf{H}_{t+a}^{(i)}(0) \end{bmatrix} \quad (5.15)$$

Los métodos de complejidad reducida que se presentan a continuación construyen su función tentativa basándose en la ecuación (5.14).

Muestreo mediante la triangularización de la matriz de autocorrelación del canal

Cualquier matriz simétrica definida positiva, \mathbf{A} , se puede factorizar como $\mathbf{A} = \mathbf{U}\mathbf{U}^H$, donde \mathbf{U} es una matriz triangular superior con las mismas dimensiones que \mathbf{A} . Vamos a utilizar este resultado, conocido como a descomposición de Cholesky [64], para reescribir la ecuación (5.14) de tal manera que las observaciones resulten de multiplicar una matriz triangular por los símbolos transmitidos, lo que nos va a permitir muestrear éstos secuencialmente. De esta manera se consigue que la complejidad del esquema de muestreo no dependa exponencialmente de la dimensión de \mathbf{s}_t .

Puesto que la matriz $\bar{\mathbf{H}}_{t,a}^{(i)}$ no es necesariamente simétrica y definida positiva, no podemos aplicar directamente la descomposición de Cholesky sobre ella. Para superar esta dificultad vamos a utilizar la matriz de autocorrelación empírica del canal,

$$\mathbf{R}_{t,a}^{(i)} = \bar{\mathbf{H}}_{t,a}^{(i)\top} \bar{\mathbf{H}}_{t,a}^{(i)} \quad (5.16)$$

⁴Los símbolos son v.a. discretas y varias partículas podrían tener idénticas muestras para la secuencia de símbolos $\mathbf{s}_{t-m+1:t-1}$. Por otro lado, los coeficientes de canal son v.a. continuas, por lo que la probabilidad de que dos muestras $\mathbf{H}_{t:t+a}^{(i)}$ y $\mathbf{H}_{t:t+a}^{(j)}$ sean iguales es, en general, 0.

Ésta es una matriz cuadrada de tamaño $N_t(a+1)$, definida positiva siempre y cuando $\bar{\mathbf{H}}_{t,a}^{(i)}$ tenga rango completo. Por el modelo de canal que estamos manejando la probabilidad de que esto no se cumpla es cero y podemos aplicar la descomposición de Cholesky a $\mathbf{R}_{t,a}^{(i)}$ para obtener:

$$\mathbf{R}_{t,a}^{(i)} = \bar{\mathbf{H}}_{t,a}^{(i)\top} \bar{\mathbf{H}}_{t,a}^{(i)} = \mathbf{U}_{t,a}^{(i)} \mathbf{U}_{t,a}^{(i)\top}, \quad (5.17)$$

donde $\mathbf{U}_{t,a}^{(i)}$ es una matriz triangular superior de dimensiones $N_t(a+1) \times N_t(a+1)$. Multiplicando (por la izquierda) en la ecuación (5.14) por el factor $\mathbf{U}_{t,a}^{(i)-1} \bar{\mathbf{H}}_{t,a}^{(i)\top}$ resulta

$$\begin{aligned} \mathbf{z}_{t,a}^{(i)} &= \mathbf{U}_{t,a}^{(i)-1} \bar{\mathbf{H}}_{t,a}^{(i)\top} \bar{\mathbf{H}}_{t,a}^{(i)} \mathbf{s}_{t+a} + \mathbf{U}_{t,a}^{(i)-1} \bar{\mathbf{H}}_{t,a}^{(i)\top} \mathbf{g}_{t+a} \\ &= \mathbf{U}_{t,a}^{(i)-1} \mathbf{U}_{t,a}^{(i)} \mathbf{U}_{t,a}^{(i)H} \mathbf{s}_{t+a} + \hat{\mathbf{g}}_{t+a}^{(i)} \\ &= \mathbf{U}_{t,a}^{(i)\top} \mathbf{s}_{t+a} + \hat{\mathbf{g}}_{t+a}^{(i)} \end{aligned} \quad (5.18)$$

donde $\mathbf{z}_{t,a}^{(i)} = \mathbf{U}_{t,a}^{(i)-1} \bar{\mathbf{H}}_{t,a}^{(i)\top} \bar{\mathbf{y}}_{t+a}^{(i)}$ y $\hat{\mathbf{g}}_{t+a}^{(i)} = \mathbf{U}_{t,a}^{(i)-1} \bar{\mathbf{H}}_{t,a}^{(i)\top} \mathbf{g}_{t+a}$ son vectores de dimensiones $N_t(a+1) \times 1$. La media y covarianza de $\hat{\mathbf{g}}_{t+a}^{(i)}$ se pueden calcular fácilmente utilizando la media y covarianza de $\mathbf{g}_{t+a} = [\mathbf{g}_t^\top, \dots, \mathbf{g}_{t+a}^\top]^\top$, que vienen dadas por

$$\mathbb{E} \left[\mathbf{g}_{t+a} \right] = \mathbf{0} \quad (5.19)$$

$$\Sigma_{\mathbf{g}_{t+a}} = \mathbb{E} \left[\mathbf{g}_{t+a} \mathbf{g}_{t+a}^\top \right] = \sigma_g^2 \mathbf{I}_{N_t(a+1)}. \quad (5.20)$$

A partir de ellas, la media y covarianza de $\hat{\mathbf{g}}_{t+a}^{(i)}$ que resultan son, respectivamente,

$$\mathbb{E} \left[\hat{\mathbf{g}}_{t+a}^{(i)} \right] = \mathbb{E} \left[\mathbf{U}_{t,a}^{(i)-1} \bar{\mathbf{H}}_{t,a}^{(i)\top} \mathbf{g}_{t+a} \right] = \mathbf{U}_{t,a}^{(i)-1} \bar{\mathbf{H}}_{t,a}^{(i)\top} \mathbb{E} \left[\mathbf{g}_{t+a} \right] = \mathbf{0} \quad (5.21)$$

y

$$\begin{aligned} \Sigma_{\hat{\mathbf{g}}_{t+a}^{(i)}} &= E \left[\hat{\mathbf{g}}_{t+a}^{(i)} \hat{\mathbf{g}}_{t+a}^{(i)\top} \right] = \mathbb{E} \left[\mathbf{U}_{t,a}^{(i)-1} \bar{\mathbf{H}}_{t,a}^{(i)\top} \mathbf{g}_{t+a} \mathbf{g}_{t+a}^\top \bar{\mathbf{H}}_{t,a}^{(i)} \mathbf{U}_{t,a}^{(i)-\top} \right] \\ &= \mathbf{U}_{t,a}^{(i)-1} \bar{\mathbf{H}}_{t,a}^{(i)\top} \mathbb{E} \left[\mathbf{g}_{t+a} \mathbf{g}_{t+a}^\top \right] \bar{\mathbf{H}}_{t,a}^{(i)} \mathbf{U}_{t,a}^{(i)-\top} = \sigma_g^2 \mathbf{U}_{t,a}^{(i)-1} \bar{\mathbf{H}}_{t,a}^{(i)\top} \bar{\mathbf{H}}_{t,a}^{(i)} \mathbf{U}_{t,a}^{(i)-\top}. \end{aligned} \quad (5.22)$$

La ecuación (5.18) nos permite muestrear uno a uno los símbolos que componen el vector \mathbf{s}_{t+a} en lugar de tener que muestrear el vector completo, que es lo que en última instancia da lugar a una complejidad exponencial en el número de antenas transmisoras y la memoria del canal. Para entender mejor el procedimiento vamos a reescribir la ecuación (5.18) haciendo

explícita la estructura de la matriz que multiplica a \mathbf{s}_{t+a} :

$$\mathbf{z}_{t,a}^{(i)} = \begin{bmatrix} u_{t,a}^{(i)}(1,1) & 0 & \dots & 0 \\ u_{t,a}^{(i)}(1,2) & u_{t,a}^{(i)}(2,2) & \ddots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ u_{t,a}^{(i)}(1,N_s) & u_{t,a}^{(i)}(2,N_s) & \dots & u_{t,a}^{(i)}(N_s,N_s) \end{bmatrix} \begin{bmatrix} s_t(1) \\ s_t(2) \\ \vdots \\ s_{t+a}(N_t) \end{bmatrix} + \mathbf{g}_{t+a}^{(i)}, \quad (5.23)$$

donde $N_s = N_t(a+1)$ es el número total de símbolos que vamos a muestrear. Nótese que en realidad solamente estamos interesados en los N_t primeros símbolos de este vector (los que componen el vector \mathbf{s}_t) y los demás únicamente se emplearán en el filtro de partículas a modo de variables auxiliares $\mathbf{s}_{t+1:t+a}$.

Dada la ecuación (5.23), los vectores de símbolos en $\mathbf{s}_{t:t+a}$ se pueden muestrear secuencialmente de forma relativamente sencilla. Para $j = 1, \dots, N_s = N_t(a+1)$ tenemos que

$$z_{t,a}^{(i)}(j) = \sum_{l=1}^j u_{t,a}^{(i)}(l,j) s_{t+\lfloor(l-1)/N_t\rfloor} (1 + ((l-1) \div N_t)) + \mathbf{g}_{t+a}^{(i)}(j), \quad (5.24)$$

donde $(x \div y)$ es el resto de x/y y se ha tenido en cuenta la correspondencia entre los elementos de \mathbf{s}_{t+a} y los símbolos en $\mathbf{s}_{t:t+a}$, que viene dada por

$$s_{t+a}^{(i)}(j) = s_{t+\lfloor(j-1)/N_t\rfloor} (1 + ((j-1) \div N_t)). \quad (5.25)$$

A la vista de la ecuación (5.24), $z_{t,a}^{(i)}(j)$ es una v.a. gaussiana cuya media viene dada por el término $\sum_{l=1}^j u_{t,a}^{(i)}(l,j) s_{t+\lfloor(l-1)/N_t\rfloor} (1 + ((l-1) \div N_t))$ y cuya varianza es la de $\mathbf{g}_{t+a}^{(i)}(j)$. Ésta última es el j -ésimo elemento en la diagonal de la matriz de covarianza calculada en (5.22).

Por tanto, la media de $z_{t,a}^{(i)}(j)$ depende de la j -ésima fila de la matriz $\mathbf{U}_{t,a}^{(i)\top}$ (que conocemos) y de los símbolos $s_t(1), s_t(2), \dots, s_{t+\lfloor(j-1)/N_t\rfloor} (1 + ((j-1) \div N_t))$. Pero los símbolos van a ser muestreados secuencialmente empezando por $s_t(1)$, por lo que para muestrear $s_{t+\lfloor(j-1)/N_t\rfloor} (1 + ((j-1) \div N_t))$ ya disponemos de muestras de todos los símbolos anteriores, $s_t(1), s_t(2), \dots, s_{t+\lfloor(j-2)/N_t\rfloor} (1 + ((j-2) \div N_t))$, que utilizaremos en la ecuación (5.24) para obtener una v.a. gaussiana,

$$z = z_{t,a}^{(i)}(j) - \sum_{l=1}^{j-1} u_{t,a}^{(i)}(l,j) s_{t+\lfloor(l-1)/N_t\rfloor} (1 + ((l-1) \div N_t)) \\ \sim \mathcal{N} \left(z | u_{t,a}^{(i)}(j,j) s_{t+\lfloor(j-1)/N_t\rfloor} (1 + ((j-1) \div N_t)), \Sigma_{\mathbf{g}_{t+a}^{(i)}}(j,j) \right), \quad (5.26)$$

cuya media únicamente depende del símbolo a muestrear.

Así pues, mediante la ecuación (5.26) podemos muestrear secuencialmente los símbolos en \mathbf{s}_{t+a} . Supongamos, por simplicidad, que $\mathcal{S} = \{-1, +1\}$ (la extensión a un alfabeto mayor

es trivial). Entonces, si $j = kN_t + n$, una muestra del símbolo $s_{t+a}^{(j)} = s_{t+k}(n)$ dadas las muestras $s_t^{(i)}(1), s_t^{(i)}(2), \dots, s_{t+k}^{(i)}(n-1)$, se obtiene de acuerdo a las probabilidades

$$\begin{aligned} Prob_{t+a}\{s_{t+k}(n) = 1\} &= Prob\{\gamma_{t+k}^{(i)}(n) > 0\} \\ Prob_{t+a}\{s_{t+k}(n) = -1\} &= Prob\{\gamma_{t+k}^{(i)}(n) < 0\} \end{aligned}$$

donde

$$\gamma_{t+k}(n) = \frac{z_{t,a}^{(i)}(j) - \sum_{l=1}^{j-1} u_{t,a}^{(i)}(l, j) s_{t+\lfloor (l-1)/N_t \rfloor}^{(i)} (1 + ((l-1) \div N_t))}{u_{t,a}^{(i)}(j, j)} \quad (5.27)$$

es una v.a. gaussiana de media $s_{t+k}(n)$ y varianza $u_{t,a}^{(i)-2}(j, j) \Sigma_{\mathbf{g}_{t+a}^{(i)}}(j, j)$.

El procedimiento anterior nos permite muestrear los símbolos en $\mathbf{s}_{t:t+a}$ secuencialmente. Primeramente, comenzamos aplicando la descomposición de Cholesky sobre la matriz de autocorrelación del canal en la ecuación (5.17). El resultado de dicha factorización permite obtener la ecuación (5.18), en la que un vector de observaciones transformadas, $\mathbf{z}_{t,a}^{(i)}$, se relaciona con los símbolos transmitidos a través de una matriz triangular. A partir de dicho vector, la ecuación (5.27) permite obtener una v.a. gaussiana para cada uno de los símbolos transmitidos cuya media es el símbolo a muestrear y cuya varianza se calcula utilizando la matriz triangular anteriormente mencionada y la covarianza dada por la ecuación (5.22). El valor de la v.a. calculada para un determinado símbolo, $\gamma_{t+k}(n)$, depende de todos los símbolos anteriores (excepto en el caso del primer símbolo a muestrear), por lo que el muestreo ha de llevarse a cabo empezando por $s_t(1)$ y terminando en $s_{t+a}(N_t)$.

La función tentativa para los datos es, entonces,

$$q_{t+a}(\mathbf{s}_{t:t+a} | \mathbf{h}_{t:t+a}, \mathbf{y}_{0:t+a}, \mathbf{s}_{0:t-1}) = \prod_{k=0}^a \prod_{n=1}^{N_t} Prob_{t+a}\{s_{t+k}(n)\}, \quad (5.28)$$

por lo que, atendiendo a la ecuación (5.11), la función tentativa que finalmente resulta para el método de filtrado de partículas es ⁵

$$q(\mathbf{s}_{t:t+a}, \mathbf{h}_{t:t+a} | \mathbf{y}_{0:t+a}, \mathbf{s}_{0:t-1}) \propto p(\mathbf{h}_t | \mathbf{s}_{0:t-1}, \mathbf{y}_{0:t-1}) \prod_{k'=1}^a p(\mathbf{h}_{t+k'} | \mathbf{h}_{t+k'-1}) \prod_{k=0}^a \prod_{n=1}^{N_t} Prob_{t+a}\{s_{t+k}(n)\}. \quad (5.29)$$

La ecuación de actualización de los pesos se obtiene sustituyendo (5.29) en (5.10)

$$\begin{aligned} w_{t+a}^{(i)} &\propto w_{t+a-1}^{(i)} \frac{\prod_{k=0}^a p(\mathbf{y}_{t+k} | \mathbf{H}_{t+k}^{(i)}, \mathbf{s}_{t+k-m+1:t+k}^{(i)}) p(\mathbf{H}_t^{(i)} | \mathbf{s}_{0:t-1}, \mathbf{y}_{0:t-1}) \prod_{l=1}^a p(\mathbf{H}_{t+l}^{(i)} | \mathbf{H}_{t+l-1}^{(i)})}{p(\mathbf{h}_t^{(i)} | \mathbf{s}_{0:t-1}, \mathbf{y}_{0:t-1}) \prod_{k'=1}^a p(\mathbf{h}_{t+k'}^{(i)} | \mathbf{h}_{t+k'-1}^{(i)}) \prod_{k=0}^a \prod_{n=1}^{N_t} Prob_{t+a}\{s_{t+k}(n)\}} \\ &\propto w_{t+a-1}^{(i)} \frac{\prod_{k=0}^a p(\mathbf{y}_{t+k} | \mathbf{H}_{t+k}^{(i)}, \mathbf{s}_{t+k-m+1:t+k}^{(i)})}{\prod_{k=0}^a \prod_{n=1}^{N_t} Prob_{t+a}\{s_{t+k}(n)\}} \end{aligned} \quad (5.30)$$

⁵Se ha renombrado a k' la variable k asociada al sumatorio de la ecuación (5.11) para distinguirla de la variable del mismo nombre asociada al sumatorio de la ecuación (5.28).

donde, una vez más, hemos tenido en cuenta que \mathbf{h}_t y \mathbf{H}_t son representaciones equivalentes de la misma información del estado del canal.

La propagación de una única partícula (actualización de su trayectoria) en el algoritmo de suavizado propuesto supone evaluar $|\mathcal{S}|N_t(a+1)$ probabilidades en lugar de las $|\mathcal{S}|^{N_t(a+1)}$ verosimilitudes que requiere el algoritmo SMC que utiliza la función tentativa óptima (con suavizado) descrito por las ecuaciones (4.45) y (4.46). Téngase en cuenta además que la complejidad de la descomposición de Cholesky es $\mathcal{O}((N_t(a+1))^2)$, por lo que la complejidad global del algoritmo por partícula es polinómica en el número de entradas, N_t , y el factor de suavizado a .

El Pseudocódigo 6 muestra la implementación del algoritmo. En él, denotamos por $\left[\mathbf{U}_{t,a}^{(i)\top}\right]_k$ a la k -ésima fila de la matriz $\mathbf{U}_{t,a}^{(i)\top}$.

Muestreo a partir de estimadores lineales

En esta sección se presenta otro método de filtrado de partículas para la igualación de canales MIMO basado en el uso de variables auxiliares. La idea fundamental consiste en utilizar estimaciones blandas de los símbolos transmitidos para diseñar una función tentativa eficiente. Pero para que la complejidad global del algoritmo resultante sea reducida, dichas estimaciones blandas han de obtenerse con alguna técnica de bajo coste computacional. Por su sencillez, el filtrado lineal es la técnica elegida en este caso.

Este método de filtrado de partículas, al igual que el basado en la triangularización de la matriz de autocorrelación del canal, va a utilizar la función tentativa dada por la ecuación (5.11). Dicha función tentativa está compuesta por dos partes estrechamente relacionadas: el muestreo del canal y el muestreo de los datos a partir de las muestras de aquel. El método presentado en esta sección tiene en común con el de la sección anterior la parte relacionada con el muestreo del canal. Es decir, ahora también vamos a tomar muestras del canal, $\mathbf{H}_{t:t+a}^{(i)}$, para construir la matriz $\bar{\mathbf{H}}_{t,a}^{(i)}$ que aparece en el modelo apilado sin interferencia intersimbólica causal dado por la ecuación (5.14). Lo que diferencia a este método del anterior es el esquema de muestreo de los datos, que explicaremos en detalle a continuación.

Partiendo de que disponemos de muestras del canal, $\mathbf{H}_{t:t+a}^{(i)}$, podemos calcular tanto el vector $\bar{\mathbf{y}}_{t:t+a}^{(i)}$ como la matriz $\bar{\mathbf{H}}_{t,a}^{(i)}$ de la ecuación (5.14) y utilizarlos para construir un receptor lineal (e.g., un detector de error cuadrático medio mínimo o un detector de decorrelación [1]) que detecte el vector $\mathbf{s}_{t:t+a}$ (recuérdese que $\mathbf{s}_{t:t+a} = [\mathbf{s}_t^\top, \mathbf{s}_{t+1}^\top, \dots, \mathbf{s}_{t+a}^\top]^\top$). Al filtro asociado a la i -ésima partícula que en el instante t detecta dicho vector de símbolos lo denotaremos por la matriz $\mathbf{F}_t^{(i)}$, de tamaño $N_r(a+1) \times N_t(a+1)$. Entonces, una estimación blanda del símbolo $s_{t+a}(k)$ para la i -ésima partícula viene dada por

$$\tilde{s}_{t+a}^{(i)}(k) = \mathbf{f}_t^{(i)\top}(k) \bar{\mathbf{y}}_{t:t+a}^{(i)}, \quad k = 1, \dots, N_t(a+1), \quad (5.31)$$

donde $\mathbf{f}_t^{(i)}(k)$ se refiere a la k -ésima columna de $\mathbf{F}_t^{(i)}$. A partir de esta estimación blanda

Pseudocódigo 6 SIS con variables auxiliares basado en la triangularización de la matriz de autocorrelación del canal

para cada instante de tiempo $t = 1, 2, \dots$

para cada partícula $i = 1, \dots, M$

muestrear $\mathbf{h}_{t:t+a}^{(i)} \sim p(\mathbf{h}_{t:t+a} | \mathbf{s}_{0:t-1}^{(i)}, \mathbf{y}_{0:t-1})$

apilar los vectores $\mathbf{s}_{t-m+1:t-1}^{(i)}$ para obtener $\mathbf{s}_{t-m+1}^{(i)}$

construir $\check{\mathbf{H}}_{t,a}^{(i)}$ y $\bar{\mathbf{H}}_{t,a}^{(i)}$ utilizando las ecuaciones (5.12) y (5.15), respectivamente

calcular $\check{\mathbf{y}}_{t+a}^{(i)} = \mathbf{y}_{t+a} - \check{\mathbf{H}}_{t,a}^{(i)} \mathbf{s}_{t-m+1}^{(i)}$

calcular $\mathbf{R}_{t,a}^{(i)} = \bar{\mathbf{H}}_{t,a}^{(i)H} \bar{\mathbf{H}}_{t,a}^{(i)}$

aplicar la descomposición de Cholesky sobre $\mathbf{R}_{t,a}^{(i)}$: $\mathbf{R}_{t,a}^{(i)} = \mathbf{U}_{t,a}^{(i)} \mathbf{U}_{t,a}^{(i)H}$

calcular el vector de observaciones transformadas $\mathbf{z}_{t,a}^{(i)} = \mathbf{U}_{t,a}^{(i)-1} \bar{\mathbf{H}}_{t,a}^{(i)H} \check{\mathbf{y}}_{t+a}^{(i)}$

calcular la matriz de covarianza de $\mathbf{g}_{t+a}^{(i)}$ mediante la ecuación (5.22)

sea $\mathbf{s}_{t+a}^{(i)} = \mathbf{0}_{N_t(a+1) \times 1}$

sea $q_{t:t+a} = 0$

para todo estimación blanda $s_{t+a}^{(j)}$ en el vector $\mathbf{s}_{t+a}^{(j)}$, $j = 1, \dots, N_t(a+1)$

$z \leftarrow z_{t,a}^{(j)} - \left[\mathbf{U}_{t,a}^{(j)\top} \right]_j \mathbf{s}_{t+a}^{(j)}$

sea $S_p \leftarrow 0$

para todo $s \in \mathcal{S}$

$Prob(s) \leftarrow \mathcal{N}(z | u_{t,a}^{(j)}(j, s), \Sigma_{\mathbf{g}_{t,a}^{(j)}}(j, j))$

acumular $S_p \leftarrow S_p + Prob(s)$

tomar una muestra $s_{t+a}^{(j)} \sim q(s_{t+a}^{(j)}) = Prob(s_{t+a}^{(j)}) / S_p$

acumular $q_{t:t+a} \leftarrow q_{t:t+a} \times q(s_{t+a}^{(j)})$

actualizar $\mathbf{s}_{t+a}^{(i)}$ con $s_{t+a}^{(j)}$

extraer $\mathbf{s}_t^{(i)}$ del vector $\mathbf{s}_{t+a}^{(i)}$

obtener el peso sin normalizar $\tilde{w}_t^{(i)}$ por medio de las ecuación (5.30)

actualizar el i -ésimo KF utilizando la muestra $\mathbf{s}_t^{(i)}$

calcular $W \leftarrow \sum_{i=1}^M \tilde{w}_t^{(i)}$

para cada partícula $i = 1, \dots, M$

normalizar su peso $w_t^{(i)} \leftarrow \tilde{w}_t^{(i)} / W$

estimar el tamaño efectivo de la muestra, $\hat{M}_{ef} \leftarrow \left(\sum_{i=1}^M w_t^{(i)2} \right)^{-1}$

si $\hat{M}_{ef} < \gamma_r M$ **entonces**

remuestrear

obtendremos una muestra de su correspondiente símbolo utilizando la función tentativa

$$s_{t+a}^{(i)}(k) \sim q(s_{t+a} | s_t(k)) = \frac{\mathcal{N}\left(\tilde{s}_{t+a}^{(i)}(k) | s_{t+a}(k), \sigma_{q,k}^2\right)}{\sum_{s \in \mathcal{S}} \mathcal{N}\left(\tilde{s}_{t+a}^{(i)}(k) | s, \sigma_{q,k}^2\right)}, \quad i = 1, \dots, M, \quad (5.32)$$

donde la varianza $\sigma_{q,k}^2$ debe ajustarse a la de la v.a. $\left(s_{t+a}(k) - \tilde{s}_{t+a}^{(i)}(k)\right)$.

Básicamente, lo que estamos haciendo en la ecuación (5.32) es asumir que una determinada estimación blanda sigue una distribución normal de media su correspondiente símbolo transmitido. Nótese que esto es simplemente un modelo, es decir, una estimación blanda no tiene por que seguir necesariamente una distribución gaussiana (de hecho, no será así por lo general) y podría haberse empleado algún otro modelo. Hemos considerado en última instancia la hipótesis de gaussianidad porque es efectiva y facilita el proceso de muestreo.

Aunque en el instante de tiempo t sólo estamos interesados en el vector de símbolos \mathbf{s}_t (los N_t primeros elementos de \mathbf{s}_{t+a}), para poder evaluar $q_{t+a}(\mathbf{s}_{t:t+a} | \mathbf{h}_{t:t+a}, \mathbf{y}_{0:t+a}, \mathbf{s}_{0:t-1})$ y actualizar el peso de la correspondiente partícula necesitamos muestras de todos los símbolos en \mathbf{s}_{t+a} . Puesto que cada uno es muestreado independientemente de todos los demás, la función tentativa para los datos que finalmente resulta es

$$q_{t+a}(\mathbf{s}_{t:t+a} | \mathbf{h}_{t:t+a}, \mathbf{y}_{0:t+a}, \mathbf{s}_{0:t-1}) = \prod_{k=1}^{N_t(a+1)} q_t(\mathbf{s}_{t+a}(k)), \quad (5.33)$$

y sólo es necesario calcular $N_t(a+1)|\mathcal{S}|$ probabilidades por partícula. Sustituyendo (5.33) en (5.11) se obtiene la función tentativa para el método de Monte Carlo,

$$q(\mathbf{s}_{t:t+a}, \mathbf{h}_{t:t+a} | \mathbf{y}_{0:t+a}, \mathbf{s}_{0:t-1}) = p(\mathbf{h}_t | \mathbf{s}_{0:t-1}, \mathbf{y}_{0:t-1}) \prod_{k'=1}^a p(\mathbf{h}_{t+k'} | \mathbf{h}_{t+k'-1}) \prod_{k=1}^{N_t(a+1)} q_t(\mathbf{s}_{t+a}(k)), \quad (5.34)$$

y a partir de ésta última y (5.10), la ecuación de actualización de los pesos es

$$\begin{aligned} w_{t+a}^{(i)} &\propto w_{t+a-1}^{(i)} \frac{\prod_{k=0}^a p(\mathbf{y}_{t+k} | \mathbf{H}_{t+k}^{(i)}, \mathbf{s}_{t+k-m+1:t+k}^{(i)}) p(\mathbf{H}_t^{(i)} | \mathbf{s}_{0:t-1}, \mathbf{y}_{0:t-1}) \prod_{l=1}^a p(\mathbf{H}_{t+l}^{(i)} | \mathbf{H}_{t+l-1}^{(i)})}{p(\mathbf{h}_t^{(i)} | \mathbf{s}_{0:t-1}, \mathbf{y}_{0:t-1}) \prod_{k'=1}^a p(\mathbf{h}_{t+k'}^{(i)} | \mathbf{h}_{t+k'-1}^{(i)}) \prod_{k=1}^{N_t(a+1)} q_t(\mathbf{s}_{t+a}^{(i)}(k))} \\ &\propto w_{t+a-1}^{(i)} \frac{\prod_{k=0}^a p(\mathbf{y}_{t+k} | \mathbf{H}_{t+k}^{(i)}, \mathbf{s}_{t+k-m+1:t+k}^{(i)})}{\prod_{k=1}^{N_t(a+1)} q_t(\mathbf{s}_{t+a}^{(i)}(k))} \end{aligned} \quad (5.35)$$

con $i = 1, \dots, M$. El Pseudocódigo 7 muestra el algoritmo completo.

La eficacia de la función tentativa así definida depende en gran medida de la elección que hagamos del filtro \mathbf{F}_t . Vamos a considerar dos filtros diferentes, un filtro MMSE y un filtro

Pseudocódigo 7 SIS con variables auxiliares basado en estimadores lineales

para cada instante de tiempo $t = 1, 2, \dots$

para cada partícula $i = 1, \dots, M$

muestrear $\mathbf{h}_{t:t+a}^{(i)} \sim p(\mathbf{h}_{t:t+a} | \mathbf{s}_{0:t-1}^{(i)}, \mathbf{y}_{0:t-1})$

apilar los vectores $\mathbf{s}_{t-m+1:t-1}^{(i)}$ para obtener $\mathbf{s}_{t-m+1}^{(i)}$

construir $\check{\mathbf{H}}_{t,a}^{(i)}$ y $\bar{\mathbf{H}}_{t,a}^{(i)}$ utilizando las ecuaciones (5.12) y (5.15), respectivamente

calcular $\check{\mathbf{y}}_{t+a}^{(i)} = \mathbf{y}_{t+a} - \check{\mathbf{H}}_{t,a}^{(i)} \mathbf{s}_{t-m+1}^{(i)}$

obtener un filtro matricial $\mathbf{F}_t^{(i)}$ a partir de $\bar{\mathbf{H}}_{t,a}^{(i)}$

calcular el vector de estimaciones blandas $\tilde{\mathbf{s}}_{t+a}^{(i)} = \mathbf{F}_t^{(i)\top} \check{\mathbf{y}}_{t+a}^{(i)}$

sea $q_{t:t+a} \leftarrow 0$

para todo símbolo $\tilde{s}_{t+a}^{(i)}(k)$ con $k = 1, \dots, N_t(a+1)$

sea $S_p \leftarrow 0$

para todo $s \in \mathcal{S}$

$Prob(s) \leftarrow \mathcal{N}\left(\tilde{s}_{t+a}^{(i)}(k) | s, \sigma_{q,k}^2\right)$

acumular $S_p \leftarrow S_p + Prob(s)$

tomar una muestra $s_{t+a}^{(i)}(k) \sim q(s_{t+a}(k)) = Prob(s_{t+a}(k)) / S_p$

acumular $q_{t:t+a} \leftarrow q_{t:t+a} \times q(s_{t+a}^{(i)}(k))$

actualizar $\mathbf{s}_{t+a}^{(i)}$ con $s_{t+a}^{(i)}(k)$

extraer $\mathbf{s}_t^{(i)}$ del vector $\mathbf{s}_{t+a}^{(i)}$

obtener el peso sin normalizar $\tilde{w}_t^{(i)}$ por medio de la ecuación (5.35)

actualizar el i -ésimo KF utilizando la muestra $\mathbf{s}_t^{(i)}$

calcular $W \leftarrow \sum_{i=1}^M \tilde{w}_t^{(i)}$

para cada partícula $i = 1, \dots, M$

normalizar su peso $w_t^{(i)} \leftarrow \tilde{w}_t^{(i)} / W$

estimar el tamaño efectivo de la muestra, $\hat{M}_{ef} \leftarrow \left(\sum_{i=1}^M w_t^{(i)2}\right)^{-1}$

si $\hat{M}_{ef} < \gamma_r M$ **entonces**

remuestrear

de decorrelación. Ambos fueron explicados en detalle en la Sección 4.1.1 y aquí únicamente adaptaremos al modelo dado por la ecuación (5.14) las ecuaciones mostradas en dicha sección. Así, el filtro MMSE es en este caso

$$\mathbf{F}_t^{(i)} = \left(\bar{\mathbf{H}}_{t,a}^{(i)} \sigma_s^2 \bar{\mathbf{H}}_{t,a}^{(i)\top} + \Sigma_{\mathbf{g}} \Big|_{t+a} \right)^{-1} \sigma_s^2 \bar{\mathbf{H}}_{t,a}^{(i)} \quad (5.36)$$

y el de decorrelación

$$\mathbf{F}_t^{(i)\top} = \left(\bar{\mathbf{H}}_{t,a}^{(i)\top} \bar{\mathbf{H}}_{t,a}^{(i)} \right)^{-1} \bar{\mathbf{H}}_{t,a}^{(i)\top}. \quad (5.37)$$

Puesto que cada uno de ellos calcula diferentes estimaciones blandas de los símbolos transmitidos, cada uno de ellos dará lugar a una función tentativa diferente. Además, el filtro elegido también influye en la varianza de las estimaciones blandas,

$$\sigma_{q,k}^2 = \mathbb{E} \left[s_{t+a} \Big|_t(k) - \tilde{s}_{t+a}^{(i)} \Big|_t(k) \right] = \mathbb{E} \left[\mathbf{f}_t^{(i)\top} \Big|_t(k) \bar{\mathbf{y}}_{t+a}^{(i)} - \tilde{s}_{t+a}^{(i)} \Big|_t(k) \right] \quad (5.38)$$

que es $\sigma_{q,k}^2 = 1 - \bar{\mathbf{h}}_{t,a}^{(i)\top} \Big|_t(k) \mathbf{f}_t^{(i)} \Big|_t(k)$ (donde $\bar{\mathbf{h}}_{t,a}^{(i)} \Big|_t(k)$ es la k -ésima columna de $\bar{\mathbf{H}}_{t,a}^{(i)}$) para el caso del filtro MMSE y $\sigma_{q,k}^2 = \mathbf{f}_t^{(i)\top} \Big|_t(k) \mathbf{f}_t^{(i)} \Big|_t(k)$ en el caso del filtro de decorrelación.

La operación más costosa que involucra este método, tanto en el caso del algoritmo basado en el filtro MMSE como en el del que hace uso del filtro de decorrelación, es el cálculo de la inversa de una matriz (ecuaciones (5.36) y (5.37), respectivamente). En ambos casos el número de operaciones requeridas es $\mathcal{O}((N_t(a+1))^3)$, por lo que la complejidad global del método de filtrado de partículas es polinómica en N_t .

5.2. Aproximación de la distribución conjunta del canal y los datos

Los algoritmos SMC presentados en la sección anterior ya evitan la complejidad exponencial asociada al método de filtrado de partículas con función tentativa óptima explicado en la Sección 4.2.2, que era el objetivo fundamental que nos planteábamos al comienzo de este capítulo. Sin embargo, todos ellos involucran operaciones computacionalmente costosas, como son la inversión de una matriz o la descomposición de Cholesky. Además, tanto el método basado en esta última como el basado en el uso de filtros lineales, utilizan el filtro de Kalman para obtener muestras del canal. Puesto que cada partícula va a tener asociado su propio filtro de Kalman, esto supone mantener un banco de M filtros y la actualización de cada uno de ellos en cada instante de tiempo conlleva todos los cálculos dados por las ecuaciones (D.5)-(D.8) del Apéndice D. Todo ello podría dificultar la utilización de estos métodos en aquellas situaciones en las que los recursos computacionales son más escasos.

En esta sección introducimos otro método para la igualación de canales MIMO que rebaja la complejidad (polinómica) de los métodos presentados en la sección anterior, ofreciendo así más posibilidades de aplicación práctica.

La idea fundamental es ahora la aproximación de la densidad de suavizado conjunta del canal y los datos, $p(\mathbf{s}_{0:t}, \mathbf{H}_{0:t} | \mathbf{y}_{0:t+a})$. Al igual que ocurría en la sección anterior con $p(\mathbf{s}_{0:t}, \mathbf{H}_t | \mathbf{y}_{0:t+a})$, esta probabilidad es difícil de calcular⁶ sin conocer los vectores de símbolos $\mathbf{s}_{t+1:t+a}$ y las matrices de canal $\mathbf{H}_{t+1:t+a}$. Por ello, en lugar de $p(\mathbf{s}_{0:t}, \mathbf{H}_{0:t} | \mathbf{y}_{0:t+a})$, vamos a tratar de aproximar la f.d.p. $p(\mathbf{s}_{0:t+a}, \mathbf{H}_{0:t+a} | \mathbf{y}_{0:t+a})$, que admite la descomposición

$$\begin{aligned}
p(\mathbf{s}_{0:t+a}, \mathbf{H}_{0:t+a} | \mathbf{y}_{0:t+a}) &= \frac{p(\mathbf{y}_{t:t+a} | \mathbf{s}_{0:t+a}, \mathbf{H}_{0:t+a}, \mathbf{y}_{0:t-1}) p(\mathbf{s}_{0:t+a}, \mathbf{H}_{0:t+a} | \mathbf{y}_{0:t-1})}{p(\mathbf{y}_{t:t+a} | \mathbf{y}_{0:t-1})} \\
&\propto p(\mathbf{y}_{t:t+a} | \mathbf{s}_{0:t+a}, \mathbf{H}_{0:t+a}, \mathbf{y}_{0:t-1}) p(\mathbf{s}_{0:t+a}, \mathbf{H}_{0:t+a} | \mathbf{y}_{0:t-1}) \\
&\propto \prod_{k=0}^a p(\mathbf{y}_{t+k} | \mathbf{H}_{t+k}, \mathbf{s}_{t+k-m+1:t+k}) \prod_{k'=0}^a p(\mathbf{H}_{t+k'} | \mathbf{H}_{t+k'-1}) \times \\
&\times p(\mathbf{s}_{0:t-1}, \mathbf{H}_{0:t-1} | \mathbf{y}_{0:t-1})
\end{aligned} \tag{5.39}$$

donde se ha tenido en cuenta que, por ser el canal un proceso AR de orden 1, el estado del mismo en un instante dado, t , sólo depende del estado en el instante anterior, $t - 1$, de tal manera que

$$p(\mathbf{H}_{t:t+a} | \mathbf{s}_{0:t+a}, \mathbf{H}_{0:t-1}, \mathbf{y}_{0:t-1}) = p(\mathbf{H}_{t:t+a} | \mathbf{H}_{t-1}). \tag{5.40}$$

Además, los posibles vectores de símbolos transmitidos son *a priori* equiprobables, lo que implica que $p(\mathbf{s}_{t:t+a} | \mathbf{s}_{0:t-1}, \mathbf{H}_{0:t-1}, \mathbf{y}_{0:t-1})$ es una constante.

La f.d.p. objetivo definida en (5.39) es muy similar a la de los métodos de la sección anterior, que venía dada por la ecuación (5.3). Obsérvese que los factores no recursivos únicamente se diferencian en la densidad marginal asociada a la matriz de canal \mathbf{H}_t . En los métodos basados en el uso de variables auxiliares (Sección 5.1) ésta era $p(\mathbf{H}_t | \mathbf{s}_{0:t-1}, \mathbf{y}_{0:t-1})$ y necesitábamos el filtro de Kalman para calcularla. Ahora es sencillamente $p(\mathbf{H}_t | \mathbf{H}_{t-1})$, lo cual quiere decir que vamos a utilizar el estado del canal en el instante $t - 1$ (o una muestra del mismo) para calcular su probabilidad en t . De esta manera evitamos el uso de un banco de M filtros de Kalman con la consiguiente reducción en complejidad.

La función tentativa la definiremos ahora como

$$\begin{aligned}
q(\mathbf{s}_{0:t+a}, \mathbf{H}_{0:t+a} | \mathbf{y}_{0:t+a}) &= q(\mathbf{s}_{t:t+a} | \mathbf{s}_{0:t-1}, \mathbf{H}_{t:t+a}, \mathbf{y}_{t:t+a}) q(\mathbf{H}_{t:t+a} | \mathbf{H}_{t-1}) \times \\
&\times q(\mathbf{s}_{0:t-1}, \mathbf{H}_{0:t-1} | \mathbf{y}_{0:t-1}),
\end{aligned} \tag{5.41}$$

lo que implica que, una vez más, vamos a tomar muestras de las matrices de canal, $\mathbf{H}_{t:t+a}$, que luego servirán para obtener muestras de los vectores de símbolos, $\mathbf{s}_{t:t+a}$. Pero en este caso no se van a descartar todas las muestras del canal una vez tengamos muestras de los símbolos transmitidos: la muestra de \mathbf{H}_t se añadirá a la secuencia de matrices de canal de la partícula correspondiente del mismo modo que la muestra de \mathbf{s}_t se añade a su trayectoria de vectores de

⁶Habría que integrarla sobre todas las posibles matrices de canal $\mathbf{H}_{t+1:t+a}$ y sobre todos los posibles vectores de símbolos $\mathbf{s}_{t+1:t+a}$. Esto último daría lugar a una complejidad exponencial en el número de antenas transmisoras y la memoria del canal, que es lo que estamos tratando de evitar.

símbolos. Se utilizará en el instante $t + 1$ para obtener una muestra, $\mathbf{H}_{t+1}^{(i)}$, de \mathbf{H}_{t+1} y calcular $p(\mathbf{H}_{t+1}^{(i)}|\mathbf{H}_t^{(i)})$. En el instante t sí se descartan, sin embargo, las muestras, $\mathbf{H}_{t+1:t+a}^{(i)}$, del canal en los instantes $t + 1, \dots, t + a$ una vez finalizado el proceso de muestreo de los datos. Muestras de las matrices de canal \mathbf{H}_{t+k} , $k > 0$ pasarán a incorporarse a las trayectorias de las diferentes partículas en el instante $t + k$, pues será ése el instante en el que dispongamos de un mayor número de observaciones relacionadas. Nótese que este mismo razonamiento es el que nos lleva a descartar las muestras de los vectores de símbolos $\mathbf{s}_{t+1:t+a}$ una vez actualizado el peso de la partícula.

La ecuación de actualización de los pesos que resulta de las ecuaciones (5.39) y (5.41) es

$$w_{t+a}^{(i)} \propto w_{t+a-1}^{(i)} \frac{\prod_{k=0}^a p(\mathbf{y}_{t+k}|\mathbf{H}_{t+k}^{(i)}, \mathbf{s}_{t+k-m+1:t+k}^{(i)}) \prod_{k'=0}^a p(\mathbf{H}_{t+k'}^{(i)}|\mathbf{H}_{t+k'-1}^{(i)})}{q(\mathbf{s}_{t:t+a}^{(i)}|\mathbf{s}_{0:t-1}^{(i)}, \mathbf{H}_{t:t+a}^{(i)}, \mathbf{y}_{t:t+a}) q(\mathbf{H}_{t:t+a}^{(i)}|\mathbf{H}_{t-1}^{(i)})}, \quad (5.42)$$

y los términos $q(\mathbf{H}_{t:t+a}^{(i)}|\mathbf{H}_{t-1}^{(i)})$ y $q(\mathbf{s}_{t:t+a}^{(i)}|\mathbf{s}_{0:t-1}^{(i)}, \mathbf{H}_{t:t+a}^{(i)}, \mathbf{y}_{t:t+a})$, que se corresponden con las funciones tentativas marginales para el canal y los datos, respectivamente, serán definidos a continuación.

5.2.1. Muestreo del canal

La ecuación de actualización de los pesos (5.42) hace posible prescindir de los bancos de filtros de Kalman en [58, 80, 81]. Sin embargo, aún tenemos que diseñar una función tentativa que evite las inversiones de matrices, así como cualquier otra operación computacionalmente costosa. Para ello vamos a utilizar un banco de estimadores de canal adaptativos que cumplirán la función que antes desempeñaba el filtro de Kalman pero con una complejidad mucho menor. Consideraremos dos algoritmos diferentes para la estimación del canal: el LMS (*Least Mean Squares*) y el RLS (*Recursive Least Squares*). Serán explicados brevemente en la presente sección.

Independientemente del estimador de canal empleado, en el instante t disponemos de una estimación de canal en el instante anterior, $\tilde{\mathbf{h}}_{t-1}^{(i)}$, para cada partícula $i \in \{1, \dots, M\}$. Entonces, teniendo en cuenta el modelo de canal dado por la ecuación (3.6), vamos a tomar una muestra, $\mathbf{h}_t^{(i)}$, de \mathbf{h}_t a partir de una distribución gaussiana de media $\alpha \tilde{\mathbf{h}}_{t-1}^{(i)}$ y matriz de covarianza (diagonal) $\sigma_H^2 \mathbf{I}$, donde σ_H^2 es un parámetro de diseño⁷. Las restantes muestras del canal, $\mathbf{h}_{t+1:t+a}^{(i)}$, necesarias en el instante t se obtienen simplemente utilizando el modelo AR. Así pues,

$$\mathbf{H}_{t:t+a}^{(i)} = \mathbf{h}_{t:t+a}^{(i)} \sim q_t(\mathbf{h}_{t:t+a}|\tilde{\mathbf{h}}_{t-1}^{(i)}) = \mathcal{N}(\mathbf{h}_t|\alpha \tilde{\mathbf{h}}_{t-1}^{(i)}, \sigma_H^2 \mathbf{I}) \prod_{k=1}^a \mathcal{N}(\mathbf{h}_{t+k}|\alpha \mathbf{h}_{t+k-1}^{(i)}, \sigma_v^2 \mathbf{I}). \quad (5.43)$$

Nótese que la distribución de la que se muestrea \mathbf{H}_t no es la asociada a la ecuación (3.6) (que definía el modelo de canal) pues difieren en la varianza⁸. Las restantes muestras, $\mathbf{H}_{t+1:t+a}$,

⁷Recuérdese la equivalencia entre \mathbf{h}_t y \mathbf{H}_t que la ecuación (5.4) hace explícita.

⁸Eligiendo $\sigma_H^2 = \sigma_v^2$ se verifica esto y conseguimos simplificar la ecuación de actualización de los pesos.

sí que se obtienen de acuerdo a dicha distribución. Por tanto,

$$q_t(\mathbf{h}_{t:t+a}|\mathbf{h}_{t-1}) = \mathcal{N}(\mathbf{h}_t|\alpha\tilde{\mathbf{h}}_{t-1}, \sigma_H^2\mathbf{I}) \prod_{k=1}^a p(\mathbf{h}_{t+k}|\mathbf{h}_{t+k-1}). \quad (5.44)$$

Una vez tenemos las muestras del canal $\mathbf{h}_{t:t+a}^{(i)}$ (o $\mathbf{H}_{t:t+a}^{(i)}$, según convenga), podemos utilizarlas (véase la Sección 5.2.2) para obtener una muestra, $\mathbf{s}_t^{(i)}$, del vector de símbolos transmitidos en el instante t , que además servirá para actualizar el estimador de canal correspondiente.

Estimación de canal basada en LMS

El estimador de error cuadrático medio mínimo (MMSE) para el canal dados los vectores de símbolos $\mathbf{s}_{0:t}^{(i)}$ y las observaciones $\mathbf{y}_{0:t}$ es

$$\tilde{\mathbf{H}}_t^{(i)} = \arg \min_{\mathbf{H}_t} \left\{ \mathbb{E} \left[\left\| \mathbf{y}_t - \mathbf{H}_t \mathbf{s}_{t-m+1}^{(i)} \right\|^2 \right] \right\}. \quad (5.45)$$

El algoritmo LMS (*Least Mean Squares*) [82] es la manera más sencilla de resolver (5.45) adaptativamente. La estimación del canal que proporciona el LMS en el instante t viene dada por

$$\tilde{\mathbf{H}}_t^{(i)} = \tilde{\mathbf{H}}_{t-1}^{(i)} - \rho \left(\tilde{\mathbf{H}}_{t-1}^{(i)} \mathbf{s}_{t-m+1}^{(i)} - \mathbf{y}_t \right) \mathbf{s}_{t-m+1}^{(i)\top}, \quad (5.46)$$

donde $\rho \ll 1$ es un parámetro de incremento, que afecta a la convergencia del algoritmo.

Estimación recursiva del canal basada en mínimos cuadrados

El algoritmo LMS tiene a su favor una complejidad computacional muy baja. Sin embargo, es bien sabido que su convergencia es por lo general lenta y sus capacidades de seguimiento limitadas. Para superar esos inconvenientes vamos a utilizar el algoritmo de mínimos cuadrados recursivo (RLS, del inglés *Recursive Least Squares*) exponencialmente ponderado [82]. El estimador de canal correspondiente se define como

$$\tilde{\mathbf{H}}_t^{(i)} = \arg \min_{\mathbf{H}} \left\{ \sum_{k=0}^t \lambda_{\text{RLS}}^{t-k} \left\| \mathbf{y}_k - \mathbf{H} \mathbf{s}_{k-m+1}^{(i)} \right\|^2 \right\}, \quad (5.47)$$

donde $0 < \lambda_{\text{RLS}} < 1$ es un factor de olvido. El Apéndice C explica como utilizar el algoritmo RLS para resolver recursivamente la secuencia de problemas planteados por (5.47). Utilizando (C.11), (C.12) y (C.15) (e inicializando la matriz de correlación, $\mathbf{R}_t = \sum_{k=0}^t \lambda_{\text{RLS}}^{t-k} \mathbf{s}_{k-m+1}^{(i)} \mathbf{s}_{k-m+1}^{(i)\top}$, a la identidad) las ecuaciones que finalmente resultan en este caso

son

$$\mathbf{R}_0^{(i)-1} \propto \mathbf{I}_{N_t m} \quad (\text{inicialización}) \quad (5.48)$$

$$\mathbf{g}_t^{(i)} = \frac{\lambda_{\text{RLS}}^{-1} \mathbf{R}_{t-1}^{(i)-1} \mathbf{s}_{k-m+1}^{(i)}}{1 + \lambda_{\text{RLS}}^{-1} \mathbf{s}_{k-m+1}^{(i)\top} \mathbf{R}_{t-1}^{(i)-1} \mathbf{s}_{k-m+1}^{(i)}} \quad (5.49)$$

$$\tilde{\mathbf{H}}_t^{(i)\top} = \tilde{\mathbf{H}}_{t-1}^{(i)\top} + \mathbf{g}_t^{(i)} \left(\mathbf{y}_t^\top - \mathbf{s}_{k-m+1}^{(i)\top} \tilde{\mathbf{H}}_{t-1}^{(i)\top} \right) \quad (5.50)$$

$$\mathbf{R}_t^{(i)-1} = \lambda_{\text{RLS}}^{-1} \left(\mathbf{I}_{N_t m} - \mathbf{g}_t^{(i)} \mathbf{s}_{k-m+1}^{(i)\top} \right) \mathbf{R}_{t-1}^{(i)-1}. \quad (5.51)$$

Reescribiendo la ecuación (5.51) como

$$\mathbf{R}_t^{(i)-1} = \lambda_{\text{RLS}}^{-1} \left(\mathbf{R}_{t-1}^{(i)-1} - \mathbf{g}_t^{(i)} \mathbf{s}_{k-m+1}^{(i)\top} \mathbf{R}_{t-1}^{(i)-1} \right) \quad (5.52)$$

se puede ver claramente que la operación computacionalmente más costosa involucrada en las ecuaciones (5.48)-(5.51) es el producto de una matriz por un vector. Así pues, la complejidad de este método de estimación de canal es cuadrática en $N_t m$.

5.2.2. Muestreo de los datos

Vamos a hacer uso de las muestras de las matrices de canal en los instantes $t, \dots, t+a$, para obtener la función tentativa de los datos por medio del filtro MMSE. Empezaremos construyendo la matriz de canal apilada $\mathbf{H}_{t,a}^{(i)}$ a la manera de (3.10) pero utilizando las muestras $\mathbf{H}_{t:t+a}^{(i)}$ en lugar de las verdaderas matrices de canal, es decir,

$$\mathbf{H}_{t,a}^{(i)} = \begin{bmatrix} \mathbf{H}_t^{(i)}(m-1)^\top & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{H}_t^{(i)}(m-2)^\top & \mathbf{H}_{t+1}^{(i)}(m-1)^\top & \dots & \mathbf{0} \\ \vdots & \mathbf{H}_{t+1}^{(i)}(m-2)^\top & \ddots & \vdots \\ \mathbf{H}_t^{(i)}(0)^\top & \vdots & \ddots & \mathbf{H}_{t+a}^{(i)}(m-1)^\top \\ \vdots & \mathbf{H}_{t+1}^{(i)}(0)^\top & \ddots & \mathbf{H}_{t+a}^{(i)}(m-2)^\top \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{H}_{t+a}^{(i)}(0)^\top \end{bmatrix}^\top. \quad (5.53)$$

Sustituyendo en la ecuación (3.9) la matriz $\mathbf{H}_{t,a}$ por la aproximación (5.53) resulta la ecuación

$$\mathbf{y}_{t+a} = \mathbf{H}_{t,a}^{(i)} \mathbf{s}_{t+a} + \mathbf{g}_{t+a}, \quad (5.54)$$

que nos servirá para obtener una estimación de los símbolos transmitidos. A la vista de las ecuaciones (A.1) y (A.13) del Apéndice A, el filtro MMSE para estimar en este caso el vector \mathbf{s}_{t+a}^{t-m+1} es

$$\mathbf{F}_{t+a}^{(i)\top} = \sigma_s^2 \mathbf{H}_{t,a}^{(i)\top} \left(\mathbf{H}_{t,a}^{(i)} \sigma_s^2 \mathbf{H}_{t,a}^{(i)\top} + \sigma_g^2 \mathbf{I} \right)^{-1} \quad (5.55)$$

y se hace necesario invertir la matriz $\mathbf{H}_{t,a}^{(i)} \sigma_s^2 \mathbf{H}_{t,a}^{(i)\top} + \sigma_g^2 \mathbf{I}$. Sin embargo, como se puede apreciar en la ecuación (A.9), esta matriz no es sino la esperanza de la matriz de correlación de las observaciones, $\mathbb{E} \left[\mathbf{y}_{t+a} \mathbf{y}_{t+a}^\top \right]$, que denotaremos por $\mathbf{R}_{y,t}$. Para evitar invertir dicha matriz vamos a aproximar esta esperanza como

$$\mathbf{R}_{y,t} = \mathbb{E} \left[\mathbf{y}_{t+a} \mathbf{y}_{t+a}^\top \right] \approx \frac{1}{\Lambda_t} \sum_{k=0}^t \lambda_{\text{MMSE}}^{t-k} \mathbf{y}_{k+a} \mathbf{y}_{k+a}^\top \quad (5.56)$$

donde λ_{MMSE} es un factor de olvido y

$$\Lambda_t = \sum_{k=0}^t \lambda_{\text{MMSE}}^k = \frac{\lambda_{\text{MMSE}}^{t+1} - 1}{\lambda_{\text{MMSE}} - 1}. \quad (5.57)$$

El Apéndice E contiene los pasos necesarios para aproximar la esperanza de la matriz de correlación de las observaciones sin necesidad de llevar a cabo inversiones de matrices. Aplicando la ecuación (E.9) que finalmente resulta en dicho apéndice al caso que nos ocupa tenemos

$$\begin{aligned} \mathbf{R}_{y,t}^{-1} &= \mathbb{E} \left[\mathbf{y}_{t+a} \mathbf{y}_{t+a}^\top \right]^{-1} = \left(\mathbf{H}_{t,a}^{(i)} \sigma_s^2 \mathbf{H}_{t,a}^{(i)\top} + \sigma_g^2 \mathbf{I} \right)^{-1} \\ &\approx \frac{\Lambda_t}{\lambda_{\text{MMSE}} \Lambda_{t-1}} \left(\mathbf{I}_{N_r(a+1)} - \mathbf{g}_{t,a} \mathbf{y}_{t+a}^\top \frac{\Lambda_t}{\lambda_{\text{MMSE}} \Lambda_{t-1}} \right) \mathbf{R}_{y,t-1}^{-1}. \end{aligned} \quad (5.58)$$

donde

$$\mathbf{g}_{t,a} = \frac{\mathbf{R}_{y,t-1}^{-1} \mathbf{y}_{t+a}}{\Lambda_t + \mathbf{y}_{t+a}^\top \frac{\Lambda_t}{\lambda_{\text{MMSE}} \Lambda_{t-1}} \mathbf{R}_{y,t-1}^{-1} \mathbf{y}_{t+a}} \quad (5.59)$$

Un banco de filtros MMSE (uno por cada símbolo a detectar) para la i -ésima partícula, con $i = 1, \dots, M$, se calcula entonces, utilizando la ecuación (5.58), como

$$\mathbf{F}_{t+a}^{(i)\top} = \sigma_s^2 \mathbf{H}_{t,a}^{(i)\top} \left(\mathbf{H}_{t,a}^{(i)} \sigma_s^2 \mathbf{H}_{t,a}^{(i)\top} + \sigma_g^2 \mathbf{I} \right)^{-1} = \sigma_s^2 \mathbf{H}_{t,a}^{(i)\top} \mathbf{R}_{y,t}^{-1} \quad (5.60)$$

inicializando $\mathbf{R}_{y,t}^{-1}$ a la matriz identidad, es decir, $\mathbf{R}_{y,0}^{-1} = \mathbf{I}_{N_r(a+1)}$.

Nótese que en este caso no hemos utilizado el modelo apilado eliminando la interferencia intersimbólica dado por la ecuación (3.13). Si lo hiciéramos, el cálculo de $\mathbf{R}_{y,t}^{-1}$ sería dependiente de la partícula (del vector concreto de observaciones transformadas obtenido

mediante la ecuación (3.11)) con el consiguiente incremento en la complejidad. En el presente método, además, no es tan importante en términos de coste computacional el tamaño de la matriz de canal que aparece en la ecuación de observación empleada pues no es necesario invertirla o descomponerla mediante Cholesky.

El filtro matricial $\mathbf{F}_{t+a}^{(i)\top}$ tiene dimensiones $N_t(m+a) \times N_r(a+1)$ y al multiplicarlo por el vector de observaciones apiladas, \mathbf{y}_{t+a} , vamos a obtener estimaciones blandas de todos los símbolos en $\mathbf{s}_{t-m+1:t+a}$ (los vectores apilados en \mathbf{s}_{t-m+1}). Sólo estamos interesados en estimaciones blandas de los símbolos $\mathbf{s}_{t:t+a}$, por lo que para obtener una estimación blanda de $\mathbf{s}_{t:t+a}$ vamos a descartar las $N_t(m-1)$ primeras filas del filtro multiplicándolo por la izquierda por la matriz $\mathbf{E} = [\mathbf{0}_{N_t(a+1) \times N_t(m-1)} \mathbf{I}_{N_t(a+1)}]$. Entonces, el vector con las estimaciones blandas de los símbolos transmitidos entre los instantes t y $t+a$ obtenido para la i -ésima partícula es

$$\tilde{\mathbf{s}}_{t+a}^{(i)} = \mathbf{E} \mathbf{F}_{t+a}^{(i)\top} \mathbf{y}_{t+a} \quad (5.61)$$

y, si $j = N_t k + q$ para enteros $k, q \geq 0$, entonces $\tilde{s}_{t+a}^{(i)}(j)$ es una estimación blanda del símbolo $s_{t+k}(q)$.

A partir de una estimación blanda de un determinado símbolo, podemos una vez más asignar probabilidades a los distintos valores que puede tomar éste. En concreto, vamos a calcular la probabilidad de que $s_{t+k}(q)$ tome el valor $s \in \mathcal{S}$ como $\mathcal{N}\left(\tilde{s}_{t+a}^{(i)}(j) | s, \sigma_q^2\right)$ (donde σ_q^2 es un parámetro de diseño) y a tomar una muestra del símbolo $s_{t+k}(q)$ utilizando la función tentativa

$$q(s_{t+k}(q)) = \frac{\mathcal{N}\left(\tilde{s}_{t+a}^{(i)}(j) | s_{t+k}(q), \sigma_q^2\right)}{\sum_{s \in \mathcal{S}} \mathcal{N}\left(\tilde{s}_{t+a}^{(i)}(j) | s, \sigma_q^2\right)}. \quad (5.62)$$

La función tentativa para el muestreo de los datos que finalmente resulta es

$$q(\mathbf{s}_{t:t+a} | \mathbf{s}_{0:t-1}^{(i)}, \mathbf{H}_{t:t+a}^{(i)}, \mathbf{y}_{t:t+a}) = \prod_{k=0}^a \prod_{q=1}^{N_t} q(s_{t+k}(q)). \quad (5.63)$$

Sustituyendo las ecuaciones (5.44) y (5.63) en (5.41) se obtiene la función tentativa que define el método de filtrado de partículas y sustituyéndolas en (5.42), la ecuación de actualización de

los pesos

$$\begin{aligned}
w_{t+a}^{(i)} &\propto w_{t+a-1}^{(i)} \frac{\prod_{k=0}^a p(\mathbf{y}_{t+k} | \mathbf{H}_{t+k}^{(i)}, \mathbf{s}_{t+k-m+1:t+k}^{(i)}) \prod_{k'=0}^a p(\mathbf{H}_{t+k'}^{(i)} | \mathbf{H}_{t+k'-1}^{(i)})}{q(\mathbf{s}_{t:t+a}^{(i)} | \mathbf{s}_{0:t-1}^{(i)}, \mathbf{H}_{t:t+a}^{(i)}, \mathbf{y}_{t:t+a}) q(\mathbf{H}_{t:t+a}^{(i)} | \mathbf{H}_{t-1}^{(i)})} \\
&= w_{t+a-1}^{(i)} \frac{\prod_{k=0}^a p(\mathbf{y}_{t+k} | \mathbf{H}_{t+k}^{(i)}, \mathbf{s}_{t+k-m+1:t+k}^{(i)}) \prod_{k'=0}^a p(\mathbf{H}_{t+k'}^{(i)} | \mathbf{H}_{t+k'-1}^{(i)})}{\prod_{k=0}^a \prod_{q=1}^{N_t} q(s_{t+k}^{(i)}(q)) \mathcal{N}(\mathbf{h}_t^{(i)} | \alpha \tilde{\mathbf{h}}_{t-1}^{(i)}, \sigma_H^2 \mathbf{I}) \prod_{k'=1}^a p(\mathbf{h}_{t+k'}^{(i)} | \mathbf{h}_{t+k'-1}^{(i)})} \\
&= w_{t+a-1}^{(i)} \frac{\prod_{k=0}^a p(\mathbf{y}_{t+k} | \mathbf{H}_{t+k}^{(i)}, \mathbf{s}_{t+k-m+1:t+k}^{(i)}) p(\mathbf{H}_t^{(i)} | \mathbf{H}_{t-1}^{(i)})}{\prod_{k=0}^a \prod_{q=1}^{N_t} q(s_{t+k}^{(i)}(q)) \mathcal{N}(\mathbf{h}_t^{(i)} | \alpha \tilde{\mathbf{h}}_{t-1}^{(i)}, \sigma_H^2 \mathbf{I})}. \tag{5.64}
\end{aligned}$$

Si hacemos el parámetro de diseño σ_H^2 igual a la varianza del proceso del canal, σ_v^2 , entonces $\mathcal{N}(\mathbf{h}_t^{(i)} | \alpha \tilde{\mathbf{h}}_{t-1}^{(i)}, \sigma_H^2 \mathbf{I}) = p(\mathbf{H}_t^{(i)} | \mathbf{H}_{t-1}^{(i)})$ y la ecuación (5.64) se simplifica a

$$w_{t+a}^{(i)} \propto w_{t+a-1}^{(i)} \frac{\prod_{k=0}^a p(\mathbf{y}_{t+k} | \mathbf{H}_{t+k}^{(i)}, \mathbf{s}_{t+k-m+1:t+k}^{(i)})}{\prod_{k=0}^a \prod_{q=1}^{N_t} q(s_{t+k}^{(i)}(q))}. \tag{5.65}$$

El Pseudocódigo (8) muestra un resumen del algoritmo.

Debido al producto de matrices en (5.60), la complejidad global del método propuesto sigue siendo cúbica en el número de antenas transmisoras, N_t , y en el factor de suavizado a , al igual que en el caso de los métodos de la Sección 5.1. Sin embargo, la constante asociada a la complejidad del nuevo método es mucho menor puesto que la estimación de la matriz de canal tiene ahora complejidad cuadrática y, además, el cálculo de los filtros lineales para la detección de los símbolos se lleva a cabo de una manera adaptativa.

5.3. Conclusiones

Los métodos SMC óptimos para la igualación de canales MIMO tienen una complejidad que crece exponencialmente con el número de antenas transmisoras. En este capítulo nos hemos planteado la búsqueda de nuevos métodos de Monte Carlo secuenciales que resuelvan el problema con una complejidad polinómica. Las dos aproximaciones presentadas se basan en un adecuado diseño de la función tentativa que permite obtener muestras de los símbolos transmitidos en un determinado instante sin necesidad de explorar todo el espacio de posibles vectores construidos a partir de los mismos.

El primer esquema propuesto se caracteriza por utilizar el canal como una variable auxiliar de la función tentativa con la intención de facilitar el muestreo de los símbolos transmitidos. Para obtener muestras del canal, cada partícula mantiene un filtro de Kalman cuya media y covarianza vienen dadas por secuencia detectada por dicha partícula. Los algoritmos resultantes tienen una complejidad cúbica y, como se verá en capítulo de simulaciones, alcanzan un rendimiento muy cercano al del filtro de partículas con función tentativa óptima (y complejidad exponencial).

Pseudocódigo 8 SIS basado en la aproximación conjunta del canal y los datos

para cada instante de tiempo $t = 1, 2, \dots$

para cada partícula $i = 1, \dots, M$

muestrear $\mathbf{H}_{t:t+a}^{(i)} \sim q(\mathbf{H}_{t:t+a} | \mathbf{H}_{t-1}^{(i)})$

construir $\mathbf{H}_{t,a}^{(i)}$ utilizando (5.53)

obtener un filtro matricial $\mathbf{F}_{t+a}^{(i)\top}$ a partir de $\mathbf{H}_{t,a}^{(i)}$ según la ecuación (5.60)

calcular el vector de estimaciones blandas $\tilde{\mathbf{s}}_{t+a}^{(i)} = \mathbf{E}\mathbf{F}_{t+a}^{(i)\top} \mathbf{y}_{t+a}$

sea $q_{t:t+a} \leftarrow 0$

para todo estimación blanda $\tilde{s}_{t+a}^{(i)}(k)$ con $k = 1, \dots, N_t(a+1)$

sea $S_p \leftarrow 0$

para todo $s \in \mathcal{S}$

$Prob(s) \leftarrow \mathcal{N}\left(\tilde{s}_{t+a}^{(i)}(k) | s, \sigma_q^2\right)$

acumular $S_p \leftarrow S_p + Prob(s)$

tomar una muestra $s_{t+a}^{(i)}(k) \sim q(s_{t+a}(k)) = Prob(s_{t+a}(k)) / S_p$

acumular $q_{t:t+a} \leftarrow q_{t:t+a} \times q(s_{t+a}^{(i)}(k))$

actualizar $\mathbf{s}_{t+a}^{(i)}$ con $s_{t+a}^{(i)}(k)$

extraer $\mathbf{s}_t^{(i)}$ del vector $\mathbf{s}_{t+a}^{(i)}$

obtener el peso sin normalizar $\tilde{w}_t^{(i)}$ por medio de las ecuación (5.64)

actualizar el i -ésimo estimador de canal utilizando la muestra $\mathbf{s}_t^{(i)}$

calcular $W \leftarrow \sum_{i=1}^M \tilde{w}_t^{(i)}$

para cada partícula $i = 1, \dots, M$

normalizar su peso $w_t^{(i)} \leftarrow \tilde{w}_t^{(i)} / W$

estimar el tamaño efectivo de la muestra, $\hat{M}_{ef} \leftarrow \left(\sum_{i=1}^M w_t^{(i)2}\right)^{-1}$

si $\hat{M}_{ef} < \gamma_r M$ **entonces**

remuestrear

El segundo método introducido en el presente capítulo aproxima la f.d.p. conjunta del canal y los datos para de esa manera evitar el uso de filtros de Kalman. En este caso, cada partícula emplea un estimador de canal adaptativo simple (como puede ser el RLS) para obtener muestras del canal que son utilizadas por una versión simplificada de un receptor lineal MMSE para obtener estimaciones blandas de los símbolos transmitidos, las cuales a su vez se utilizarán en el proceso de muestreo de los datos. La complejidad de los algoritmos que se derivan de este método también es cúbica en el número de flujos de entrada. Sin embargo, a costa de una ligera degradación en el rendimiento, el número de cálculos que precisan realizar es mucho menor que en el caso anterior (la constante asociada a la complejidad es más pequeña), lo que hace que sean especialmente adecuados en aquellas situaciones en las que los recursos computacionales son muy limitados.

Capítulo 6

Igualación de canales MIMO de orden desconocido

En cualquier sistema de comunicaciones la señal recibida es una versión distorsionada de la señal enviada. Si además el canal es selectivo en frecuencia (o, equivalentemente, dispersivo en tiempo), un pulso de duración T en el emisor dará lugar a uno de duración $T' > T$ en el receptor. Ello implica que la energía de un símbolo transmitido se dispersa a lo largo de m períodos de símbolo en la señal recibida, siendo m la memoria u orden del canal¹. El orden del canal en el ejemplo de la figura es $m = 3$ porque, de acuerdo a nuestro modelo, la energía de un símbolo transmitido en el instante t se recibe en los instantes $t, t + 1, \dots, t + m - 1$.

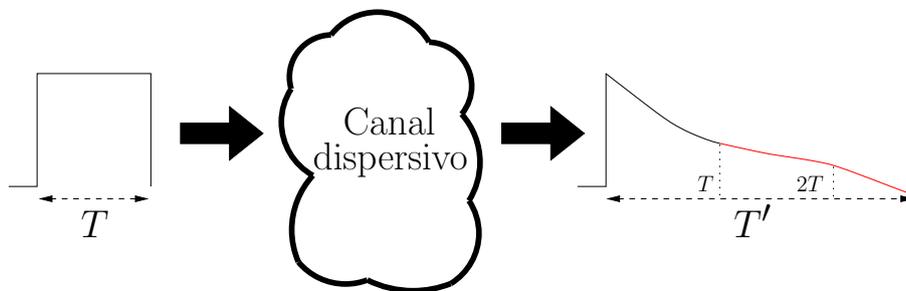


Figura 6.1: Dispersión de la señal transmitida en un canal selectivo en frecuencia.

Hasta ahora hemos supuesto que el orden del canal es conocido. Sin embargo, esto rara vez ocurre en la práctica y suele asumirse que tiene un determinado valor como paso previo al proceso de detección de los símbolos transmitidos. Podría ocurrir que dicho valor fuera mayor que el verdadero orden del canal. En tal caso se dice que el orden del canal está siendo sobreestimado. Por el contrario, cuando consideramos un orden menor del que realmente tiene estamos subestimando el orden del canal.

¹Nótese que estamos definiendo la memoria de un canal MIMO como el número *total* de vectores de símbolos que intervienen en una determinada observación, por lo que $m > 0$. Algunos autores consideran que la memoria es el número de vectores de símbolos *pasados* que intervienen en la observación. En ese caso se tiene $m \geq 0$.

Subestimar el orden del canal supone no tener en cuenta, en cada una de las observaciones del modelo de señal, uno o más vectores de símbolos que intervienen en la misma. Es por ello que subestimando el orden del canal no es posible realizar una detección fiable de los datos transmitidos. Para evitar dicha situación es una práctica común [83] asumir que el canal tiene un orden mayor del esperado, lo cual supone, con una alta probabilidad², sobreestimarlos. Esta aproximación evita el problema anteriormente mencionado, pero tiene dos inconvenientes importantes. Por un lado incrementa la complejidad del receptor en la medida en la que ésta dependa de las dimensiones de la matriz de canal y, por otro, da lugar a una progresiva pérdida de rendimiento a medida que el orden de canal considerado se aleja del verdadero orden. Esto último es debido a que es necesario estimar una matriz de canal más grande partiendo de las mismas observaciones.

Además de todo lo mencionado, hay que tener en cuenta que, al depender el orden del canal del período de símbolo, en la elección de éste ha de tenerse en cuenta la velocidad de transmisión, que es un parámetro variable en muchos sistemas.

Todo lo arriba expuesto sugiere la necesidad de buscar métodos para la igualación de canales MIMO que, por un lado, nunca subestimen el orden del canal y, por otro, traten de minimizar la penalización computacional y de rendimiento que supone sobreestimarlos. Se ha dividido el capítulo en cuatro secciones. La primera explica brevemente el modelo de señal considerado, que es una extensión al del Capítulo 3. En la segunda se revisan algunos de los trabajos anteriores que se pueden encontrar en la literatura abordando el problema. La mayor parte de ellos parten de que el canal es invariante en el tiempo y no son, por tanto, directamente aplicables a la igualación del canal selectivo en tiempo considerado en esta tesis. Además, la técnica basada en la intersección de dos funciones de coste [39] ha sido diseñada para canales con múltiples salidas pero una única entrada (SIMO, del inglés *Single Input Multiple Output*).

La tercera sección se dedica a la derivación del método de filtrado de partículas óptimo para la igualación de canales MIMO de orden desconocido. La complejidad del algoritmo resultante es prohibitiva incluso en los sistemas más sencillos, lo que (a día de hoy) impide su aplicación en la mayor parte de situaciones prácticas.

En la última sección del capítulo se explican los métodos propuestos en el presente trabajo para la estimación del orden del canal. El primero de ellos extiende el método de filtrado de partículas de complejidad reducida visto en la Sección 5.2 para tener en cuenta la incertidumbre asociada al orden del canal. El método resultante tiene complejidad polinómica en el número de antenas transmisoras y el número de órdenes de canal considerados, lo que permitiría a un hardware paralelo con elementos de cómputo muy simples resolver un problema extremadamente complejo como es la estimación del orden del canal en un sistema MIMO. El segundo método propuesto, basado en la metodología del procesado por supervivientes [66], ofrece un rendimiento muy superior al del primero a costa de un cierto incremento en la complejidad. Sin embargo, ésta se encuentra aún muy por debajo de la del método de filtrado

²Puesto que no se conoce el orden del canal, asignándole un determinado valor no es posible saber si lo estamos subestimando, sobreestimando o, incluso, si el orden considerado es el verdadero.

de partículas con función tentativa óptima que resuelve el mismo problema.

6.1. Modelo de señal

El modelo de señal considerado en este capítulo es esencialmente el descrito en el Capítulo 3 con la particularidad de que, al ser en este caso desconocido el orden del canal, m , vamos a considerarlo una variable aleatoria discreta³ en el conjunto $\mathcal{M} = \{1, 2, \dots, |\mathcal{M}|\}$ con distribución de probabilidad *a priori* uniforme, es decir,

$$p(m) = \frac{1}{|\mathcal{M}|}, \quad m \in \mathcal{M}. \quad (6.1)$$

La ecuación que describe el vector de observaciones recogidas en el instante t sigue siendo

$$\mathbf{y}_t = \mathbf{H}_t \mathbf{s}_{t-m+1} + \mathbf{g}_t, \quad (6.2)$$

donde $\mathbf{s}_{t-m+1} = [\mathbf{s}_{t-m+1}^\top \cdots \mathbf{s}_t^\top]^\top$ es un vector $N_t m \times 1$ que apila todos los vectores de símbolos implicados en la t -ésima observación y $\mathbf{H}_t = [\mathbf{H}_t(m-1) \cdots \mathbf{H}_t(0)]$ es la matriz de canal conjunta de tamaño $N_r \times N_t m$. Sin embargo, y puesto que m es ahora desconocido, no sabemos cuantos vectores de símbolos intervienen en una observación, lo cual a su vez determina el número de columnas de la matriz de canal. Por otro lado, aún sin conocer las dimensiones de ésta, podemos seguir modelando su evolución como un proceso AR de orden 1 de tal manera que

$$\mathbf{H}_t(l) = \alpha \mathbf{H}_{t-1}(l) + \mathbf{V}_t(l) \quad 0 \leq l \leq m-1 \quad (6.3)$$

donde α es el coeficiente del proceso AR y $\mathbf{V}_t(l)$ es una matriz $N_r \times N_t$ de variables aleatorias gaussianas i.i.d. con media cero y varianza σ_v^2 .

En cuanto a la notación introducida en este capítulo, los algoritmos de la Sección 6.4 obtienen estimaciones de la misma matriz de canal considerando diferentes órdenes de canal. Para diferenciarlas vamos a denotar por $\mathbf{H}_{m,t}$ a la matriz de canal en el instante t asumiendo que m es el verdadero orden del canal. Por otro lado, dado el alfabeto \mathcal{S} en el que pueden tomar valores los símbolos, nos referiremos al conjunto de todos los posibles vectores de símbolos de tamaño N_t como \mathcal{S}_t^N .

6.2. Aproximaciones anteriores

En esta sección revisaremos los fundamentos de los métodos más relevantes que se pueden encontrar en la literatura para afrontar el problema de la igualación de canales (no necesariamente MIMO) de orden desconocido. En cada caso, señalaremos las principales ventajas e inconvenientes que presentan a la hora de aplicarlos al modelo de señal considerado.

³Nótese que aunque estamos modelando el orden del canal como una v.a., éste es en realidad un parámetro determinista que no conocemos.

6.2.1. Basadas en teoría de la información

MDL

El principio de la descripción de menor longitud (MDL, del inglés *Minimum Description Length*) [84] es un método de inferencia inductiva [85, 86] que tiene por objeto elegir la hipótesis que mejor explica una secuencia de datos de entre un conjunto de hipótesis candidatas. Es decir, aborda el problema que en la literatura se conoce como “selección de modelo”.

MDL parte de la base de que cualquier regularidad que se halle en una secuencia de datos se puede utilizar para comprimir dicha secuencia, esto es, para describirla empleando menos símbolos de información de los que se necesitan para representarla literalmente (tal como se recibe). Entonces, dada una secuencia de datos y un conjunto de hipótesis (o modelos) que podrían explicarla, el principio MDL propone elegir aquella hipótesis que permita una mayor compresión de los datos. En cierta manera esto equivale a elegir la explicación más sencilla, lo que hace que este método guarde una estrecha relación con el razonamiento de Occam [87].

Para describir los datos (para comprimirlos) necesitamos un método de descripción (un código), es decir, un lenguaje formal que permita expresar las propiedades de los mismos. Podríamos elegir a tal efecto un lenguaje de programación de propósito general y surge entonces el concepto de complejidad de Kolmogorov de una secuencia [88], que se define como la longitud del programa más corto que imprime la secuencia y termina. Obviamente, la longitud del programa que lleva a cabo una determinada tarea depende (entre otros factores) del lenguaje de programación empleado. Sin embargo, el teorema de invarianza [33] nos dice que si la secuencia de datos es lo suficientemente larga, la diferencia entre la longitud de dos programas que la impriman, escritos en lenguajes de programación diferentes, es una constante que no depende de la longitud de la secuencia. Entonces, podemos identificar al programa de menor longitud que imprime una secuencia de datos como el mejor modelo que la ajusta. Esta versión idealizada del MDL tiene dos problemas: por un lado no es posible encontrar el programa de menor longitud que imprime una secuencia de datos dada y, por otro, en la práctica se trabaja con secuencias pequeñas de datos por lo que el teorema de invarianza no es aplicable y el lenguaje de programación empleado es relevante.

Una manera de llevar a la práctica las ideas arriba expuestas es lo que algunos autores denominan la versión en dos partes del principio MDL [33]. Sean $\mathcal{H}^{(1)}, \dots, \mathcal{H}^{(N_H)}$ la lista de hipótesis (o modelos) considerados para explicar una secuencia de datos D . La mejor hipótesis, $H \in \mathcal{H}^{(1)}, \dots, \mathcal{H}^{(N_H)}$, es aquella que minimiza la suma $L(H) + L(D|H)$ donde, $L(H)$ es la longitud en bits de la descripción de la hipótesis, y $L(D|H)$ es la longitud en bits de la descripción de los datos cuando se comprimen o codifican con la ayuda de la hipótesis.

El principio de MDL es un método muy general para seleccionar un modelo y, como tal, se podría aplicar al problema de la estimación del orden de un canal MIMO selectivo en frecuencia. Sin embargo, se sabe que en este caso el MDL tiende a sobreestimar el orden del canal [35, 36], por lo que no constituye una buena aproximación al problema.

Criterio de información de Akaike

El criterio de información de Akaike (AIC, del inglés *An Information Criterion*) [34] se utiliza para evaluar la bondad de ajuste de un modelo a un conjunto de datos. Tiene su origen en el concepto de divergencia de Kullback-Leibler, que es una medida del parecido entre dos distribuciones de probabilidad [89].

Dada una secuencia de datos y un conjunto modelos que podrían explicarla, el modelo al que mejor se ajustan los datos según el AIC es aquel que minimiza la métrica

$$AIC = 2k - 2 \ln L \quad (6.4)$$

donde k es el número de parámetros del modelo y L es el valor máximo de la verosimilitud que se puede conseguir con el mismo. Así pues, a la vista de (6.4), un buen modelo es aquel que explica un buen porcentaje de los datos observados (la verosimilitud es alta cuando a los parámetros se les asigna el valor correcto) y lo hace con el menor número de parámetros posibles (previniendo así la sobreestimación del modelo).

Al igual que el MDL, el criterio de información de Akaike es un método muy general para discriminar entre modelos que compiten entre sí a la hora de explicar una secuencia de observaciones y, como tal, también es aplicable al problema de la selección del orden del canal. Sin embargo, también tiene el problema de que tiende a sobreestimar el orden del canal [35, 36].

6.2.2. CME

Los anteriores métodos basados en teoría de la información se derivan utilizando argumentos asintóticos [37], por lo que desde un punto de vista teórico únicamente son válidos cuando el conjunto de datos es muy grande. Su aplicación a problemas como el que nos ocupa, en el que tenemos un conjunto limitado de datos para ayudarnos a decidir entre un modelo y otro no siempre está justificada.

Un criterio para resolver el problema de la selección de modelos que no tiene esta limitación es el descrito en [37]. El autor propone una aproximación al problema completamente diferente basada en la teoría de estadísticos suficientes. La idea fundamental es considerar que los datos se componen de dos partes. Una de ellas, bajo la hipótesis de que un determinado modelo es el correcto, constituye un estadístico suficiente para los parámetros desconocidos de dicho modelo. La otra no depende de éstos, sino de su dimensión, y es la que permite discriminar entre modelos.

El método expuesto en [37], acuñado por el autor con el término CME (del inglés, *Conditional Model Estimator*), funciona sin ningún tipo de conocimiento *a priori* acerca de los parámetros del modelo, lo cual es una ventaja pues estos son, por lo general, desconocidos. Otra ventaja fundamental del CME es que la regla de decisión a la que da lugar está estrechamente relacionada con la maximización de la probabilidad de una decisión correcta.

En [38] se aplica el criterio CME para estimar el orden de un canal MIMO cuasi-estático selectivo en frecuencia. El método resultante asigna un valor, $CME(m)$, a cada posible orden

del canal $m \in \mathcal{M}$, y el orden preferido es aquel que minimiza dicho estadístico. Adaptando la ecuación obtenida en [38] a nuestro modelo de señal resulta

$$\text{CME}(m) = \frac{1}{\sigma_g^2} \sum_{t=0}^K \left(\mathbf{y}_t - \mathbf{H}_m \mathbf{s}_{t-m+1} \right)^\top \left(\mathbf{y}_t - \mathbf{H}_m \mathbf{s}_{t-m+1} \right) + \frac{N_r}{2} \ln |\mathbf{S}_K^\top \mathbf{S}_K| \quad (6.5)$$

donde \mathbf{H}_m es la matriz de canal asumiendo orden m , K es la longitud de la trama y

$$\mathbf{S}_K = \begin{bmatrix} s_0(1) & s_{-1}(1) & s_{-2}(1) & \cdots & s_{-m+1}(1) & s_0(2) & \cdots & s_{-m+1}(N_t) \\ s_1(1) & s_0(1) & s_{-1}(1) & \cdots & s_{-m+2}(1) & s_1(2) & \cdots & s_{-m+1}(N_t) \\ \vdots & \vdots \\ s_{K-1}(1) & s_{K-2}(1) & s_{K-3}(1) & \cdots & s_{K-m}(1) & s_{K-1}(2) & \cdots & s_{K-m}(N_t) \end{bmatrix}. \quad (6.6)$$

En la práctica ni el canal ni los símbolos serán conocidos y deben ser estimados considerando cada uno de los posibles órdenes de canal $m \in \mathcal{M}$. Los autores de [38] utilizan el algoritmo EM a tal efecto.

El método descrito en [38] para estimar el orden de un canal MIMO solamente es válido cuando éste último es invariante en el tiempo⁴. Aunque a la vista de la ecuación (6.5) la extensión a un canal selectivo en tiempo puede parecer trivial, hay que tener en cuenta que el criterio CME ha sido derivando asumiendo que los parámetros que determinan el modelo (los coeficientes del canal en este caso), cuyo número queremos determinar, son fijos. Cuando esta condición no se cumple, como ocurre al asumir un canal variante con el tiempo, el rendimiento del algoritmo puede verse seriamente afectado.

Otra desventaja de este método es que realiza un procesamiento por bloques de los datos (observaciones), por lo que no puede ser utilizado en un sistema adaptativo.

6.2.3. Combinación de las funciones de coste de identificación e igualación de canal

La función de coste que se utiliza en trabajos como [90] para la identificación ciega de canales SIMO es monótona decreciente en relación al orden del canal. Por otro lado, la función de coste empleada con frecuencia en la igualación ciega de este tipo de sistemas (véase e.g. [88]) es monótona creciente respecto del mismo parámetro. Teniendo en cuenta esto, los autores de [39] construyen una nueva función de coste combinación de ambas que alcanza un mínimo en el verdadero orden del canal.

Las ventajas del método son, por un lado, que funciona tanto en SNRs altas como bajas y, por otro, que como subproducto del mismo se obtienen estimaciones del canal y filtros para llevar a cabo su igualación. Sus desventajas son que únicamente es para sistemas SIMO y que no considera canales variantes con el tiempo.

⁴Obsérvese que en la ecuación (6.5) se ha suprimido el subíndice temporal que acompaña a la matriz de canal a lo largo de toda la tesis.

6.2.4. Métodos basados en subespacios

Considerando el canal MIMO como un sistema lineal e invariante en el tiempo, en [91] se propone un esquema para la estimación del orden del canal basado en la relación isomórfica entre los subespacios de entrada y salida. Su mayor atractivo reside en que admite la posibilidad que cada antena transmisora tenga su propio orden de canal.

Una limitación de este esquema es que requiere que el número de antenas transmisoras sea estrictamente menor que el de receptoras. Además, ha sido derivado asumiendo que la transmisión no se ve afectada por ruido o, equivalentemente, que las observaciones están completamente determinadas por el canal y la secuencia transmitida. Por ello, en presencia de ruido es necesario disponer de una SNR muy alta (a partir de 20 dB) para obtener un buen rendimiento. Éste también depende en buena medida del número de observaciones, es decir, de la longitud de trama, de tal manera que tramas más largas incrementan la fiabilidad del método. Por otro lado, este esquema sólo es inherentemente válido para canales invariantes en el tiempo, lo que impide su aplicación al sistema de comunicaciones objeto de estudio.

6.3. SMC utilizando la función tentativa óptima

Sea o no conocido el orden del canal, en el instante t queremos detectar la secuencia transmitida, $\mathbf{s}_{0:t}$, dadas las observaciones, $\mathbf{y}_{0:t+a}$, donde a es el factor de suavizado. Así pues, nos planteamos una vez más⁵ la aproximación de la f.d.p. $p(\mathbf{s}_{0:t}|\mathbf{y}_{0:t+a})$.

Cuando el orden del canal es conocido, utilizando un factor de suavizado $a = m - 1$ tenemos en cuenta todas las observaciones que involucran a un determinado vector de símbolos a la hora de detectarlo. Si queremos garantizar esto mismo cuando el orden es desconocido debemos elegir $a = \max\{\mathcal{M}\} - 1$, lo cual supone suavizar utilizando tantos vectores de observaciones futuras como requiera el máximo valor que puede tomar el orden del canal.

Las funciones de densidad de probabilidad objetivo y tentativa que vamos a manejar en este método tienen exactamente la misma forma que las de la Sección 4.2.2. Así pues, por la ecuación (4.44) tenemos

$$p(\mathbf{s}_{0:t}|\mathbf{y}_{0:t+a}) \propto p(\mathbf{y}_{t:t+a}|\mathbf{s}_{0:t}, \mathbf{y}_{0:t-1})p(\mathbf{s}_{0:t-1}|\mathbf{y}_{0:t-1}), \quad (6.7)$$

y por (4.38)

$$q(\mathbf{s}_{0:t}|\mathbf{y}_{0:t+a}) = q(\mathbf{s}_t|\mathbf{s}_{0:t-1}, \mathbf{y}_{0:t+a})q(\mathbf{s}_{0:t-1}|\mathbf{y}_{0:t-1}) \quad (6.8)$$

donde $q(\mathbf{s}_t|\mathbf{s}_{0:t-1}, \mathbf{y}_{0:t+a})$ es la función tentativa marginal de la que se muestrea el vector de símbolos \mathbf{s}_t , y puede escribirse, atendiendo a la ecuación (4.43), como

$$q(\mathbf{s}_t|\mathbf{s}_{0:t-1}, \mathbf{y}_{0:t+a}) \propto \frac{p(\mathbf{y}_{t:t+a}|\mathbf{s}_{0:t}, \mathbf{y}_{0:t-1})}{\sum_{\tilde{\mathbf{s}}_t \in \mathcal{S}^{N_t}} p(\mathbf{y}_{t:t+a}|\mathbf{s}_{0:t-1}, \tilde{\mathbf{s}}_t, \mathbf{y}_{0:t-1})}. \quad (6.9)$$

⁵Nótese que ésta es la misma distribución que tratábamos de aproximar en la Sección 4.2.2.

La diferencia entre este método de filtrado de partículas y el de la Sección 4.2.2, que suponía conocido el orden del canal, está en el cálculo de la verosimilitud $p(\mathbf{y}_{t:t+a}|\mathbf{s}_{0:t}, \mathbf{y}_{0:t-1})$. Cuando el orden de canal es conocido dicha probabilidad puede calcularse mediante el filtro de Kalman. Cuando el orden del canal es desconocido podemos aplicar el teorema de la probabilidad total para calcular $p(\mathbf{y}_{t:t+a}|\mathbf{s}_{0:t}, \mathbf{y}_{0:t-1})$ como

$$\begin{aligned} p(\mathbf{y}_{t:t+a}|\mathbf{s}_{0:t}, \mathbf{y}_{0:t-1}) &= \sum_{m \in \mathcal{M}} p(\mathbf{y}_{t:t+a}, m|\mathbf{s}_{0:t}, \mathbf{y}_{0:t-1}) \\ &= \sum_{m \in \mathcal{M}} p(\mathbf{y}_{t:t+a}|m, \mathbf{s}_{0:t}, \mathbf{y}_{0:t-1})p(m, |\mathbf{s}_{0:t-1}, \mathbf{y}_{0:t-1}), \end{aligned} \quad (6.10)$$

donde se ha tenido en cuenta que el t -ésimo vector de símbolos, \mathbf{s}_t , no aporta ninguna información a la hora de calcular la probabilidad del orden del canal m dadas las observaciones $\mathbf{y}_{0:t-1}$, por lo que $p(m, |\mathbf{s}_{0:t}, \mathbf{y}_{0:t-1}) = p(m, |\mathbf{s}_{0:t-1}, \mathbf{y}_{0:t-1})$.

Cada una de las verosimilitudes $p(\mathbf{y}_{t:t+a}|m, \mathbf{s}_{0:t}, \mathbf{y}_{0:t-1})$ con $m \in \mathcal{M}$ asume un orden de canal diferente y puede ser evaluada mediante la ecuación (4.42) utilizando un filtro de Kalman adaptado al tamaño de la matriz de canal que corresponda en cada caso⁶. El término $p(m, |\mathbf{s}_{0:t-1}, \mathbf{y}_{0:t-1})$ es la probabilidad *a posteriori* del orden del canal en el instante $t - 1$ y se puede calcular hasta una constante de proporcionalidad mediante el teorema de Bayes como

$$\begin{aligned} p(m, |\mathbf{s}_{0:t-1}, \mathbf{y}_{0:t-1}) &= \frac{p(\mathbf{y}_{t-1}|m, \mathbf{s}_{0:t-1}, \mathbf{y}_{0:t-2})p(m, |\mathbf{s}_{0:t-1}, \mathbf{y}_{0:t-2})}{p(\mathbf{y}_{t-1}|\mathbf{s}_{0:t-1}, \mathbf{y}_{0:t-2})} \\ &\propto p(\mathbf{y}_{t-1}|m, \mathbf{s}_{0:t-1}, \mathbf{y}_{0:t-2})p(m, |\mathbf{s}_{0:t-2}, \mathbf{y}_{0:t-2}) \\ &\propto p(m) \prod_{k=0}^{t-1} p(\mathbf{y}_k|m, \mathbf{s}_{0:k}, \mathbf{y}_{0:k-1}) \end{aligned} \quad (6.11)$$

donde $p(m)$ es la f.m.p. *a priori* de m .

La función tentativa de la que vamos a muestrear los símbolos transmitidos en el instante t se obtiene sustituyendo (6.10) en (6.9), de tal manera que una muestra del vector de símbolos \mathbf{s}_t para la i -ésima partícula se genera de acuerdo a

$$\mathbf{s}_t^{(i)} \sim q(\mathbf{s}_t|\mathbf{s}_{0:t-1}^{(i)}, \mathbf{y}_{0:t+a}) \propto \frac{\sum_{m \in \mathcal{M}} p(\mathbf{y}_{t:t+a}|m, \mathbf{s}_{0:t-1}^{(i)}, \mathbf{s}_t, \mathbf{y}_{0:t-1})p(m, |\mathbf{s}_{0:t-1}^{(i)}, \mathbf{y}_{0:t-1})}{\sum_{\tilde{\mathbf{s}}_t \in \mathcal{S}^{N_t}} \sum_{m \in \mathcal{M}} p(\mathbf{y}_{t:t+a}|m, \mathbf{s}_{0:t-1}^{(i)}, \tilde{\mathbf{s}}_t, \mathbf{y}_{0:t-1})p(m, |\mathbf{s}_{0:t-1}^{(i)}, \mathbf{y}_{0:t-1})} \quad (6.12)$$

⁶En la Sección 4.2.2 no hacíamos explícito el condicionamiento de las observaciones respecto del orden del canal porque suponíamos que éste era un parámetro conocido del sistema. Sin embargo, este conocimiento estaba implícito en el filtro de Kalman construido para calcular la probabilidad $p(\mathbf{y}_{t:t+a}|\mathbf{s}_{0:t}, \mathbf{y}_{0:t-1})$, y podríamos haber escrito ésta última como $p(\mathbf{y}_{t:t+a}|m, \mathbf{s}_{0:t}, \mathbf{y}_{0:t-1})$

y su peso se actualiza mediante la ecuación

$$\begin{aligned}
w_{t+a}^{(i)} &= \frac{p(\mathbf{s}_{0:t}^{(i)}|\mathbf{y}_{0:t+a})}{q(\mathbf{s}_{0:t}^{(i)}|\mathbf{y}_{0:t+a})} \propto \frac{p(\mathbf{y}_{t:t+a}|\mathbf{s}_{0:t}^{(i)}, \mathbf{y}_{0:t-1})p(\mathbf{s}_{0:t-1}^{(i)}|\mathbf{y}_{0:t-1})}{q(\mathbf{s}_t^{(i)}|\mathbf{s}_{0:t-1}^{(i)}, \mathbf{y}_{0:t+a})q(\mathbf{s}_{0:t-1}^{(i)}|\mathbf{y}_{0:t-1})} = w_{t+a-1}^{(i)} \frac{p(\mathbf{y}_{t:t+a}|\mathbf{s}_{0:t}^{(i)}, \mathbf{y}_{0:t-1})}{q(\mathbf{s}_t^{(i)}|\mathbf{s}_{0:t-1}^{(i)}, \mathbf{y}_{0:t+a})} \\
&= w_{t+a-1}^{(i)} \frac{p(\mathbf{y}_{t:t+a}|\mathbf{s}_{0:t}^{(i)}, \mathbf{y}_{0:t-1})}{\frac{p(\mathbf{y}_{t:t+a}|\mathbf{s}_{0:t}^{(i)}, \mathbf{y}_{0:t-1})}{\sum_{\tilde{\mathbf{s}}_t \in \mathcal{S}^{N_t}} p(\mathbf{y}_{t:t+a}|\mathbf{s}_{0:t-1}^{(i)}, \tilde{\mathbf{s}}_t, \mathbf{y}_{0:t-1})}} = w_{t+a-1}^{(i)} \sum_{\tilde{\mathbf{s}}_t \in \mathcal{S}^{N_t}} p(\mathbf{y}_{t:t+a}|\mathbf{s}_{0:t-1}^{(i)}, \tilde{\mathbf{s}}_t, \mathbf{y}_{0:t-1}) \\
&= w_{t+a-1}^{(i)} \sum_{\tilde{\mathbf{s}}_t \in \mathcal{S}^{N_t}} \sum_{m \in \mathcal{M}} p(\mathbf{y}_{t:t+a}|m, \mathbf{s}_{0:t-1}, \tilde{\mathbf{s}}_t, \mathbf{y}_{0:t-1})p(m, |\mathbf{s}_{0:t-1}, \mathbf{y}_{0:t-1}). \tag{6.13}
\end{aligned}$$

Tanto (6.12) como (6.13) van a hacer uso de las ecuaciones (6.11) y (4.42). Ésta última la reescribiremos ahora por conveniencia como

$$p(\mathbf{y}_{t:t+a}|m, \mathbf{s}_{0:t}^{(i)}, \mathbf{y}_{0:t-1}) = \sum_{\tilde{\mathbf{s}}_{t+1:t+a} \in \mathcal{S}^{N_t a}} \prod_{k=0}^a p(\mathbf{y}_{t+k}|m, \mathbf{s}_{0:t}^{(i)}, \tilde{\mathbf{s}}_{t+1:t+k}, \mathbf{y}_{0:t+k-1}), \tag{6.14}$$

haciendo de esta manera explícito el orden del canal.

La ecuación (6.14) contiene un sumatorio sobre todas las posibles combinaciones de vectores de símbolos de suavizado, $\tilde{\mathbf{s}}_{t+1:t+a} \in \mathcal{S}^{N_t a}$, lo cual implica una vez más una complejidad exponencial en el número de antenas transmisoras, N_t , y el máximo factor de suavizado, $a = \max\{\mathcal{M}\} - 1$. Por otro lado, en la ecuación de actualización de los pesos, el término $p(\mathbf{y}_{t:t+a}|m, \mathbf{s}_{0:t}^{(i)}, \mathbf{y}_{0:t-1})$ desarrollado en (6.14) está dentro de un sumatorio sobre todos los órdenes de canal en \mathcal{M} , lo cual sugiere una complejidad lineal en el número de órdenes de canal considerados. Así pues, la complejidad asociada a propagar una trayectoria (muestrear un nuevo vector de símbolos) y actualizar su peso correspondiente mediante la ecuación (6.13) es $\mathcal{O}(|\mathcal{M}|\mathcal{S}^{N_t(a+1)})$. El Pseudocódigo 9 muestra un resumen del algoritmo resultante.

6.4. Nuevos métodos para la igualación de canales de orden desconocido

La enorme complejidad del algoritmo anterior motivó en esta tesis el desarrollo de nuevas técnicas que resolvieran el mismo problema con una carga computacional aceptable. En lo que resta de capítulo se presentan dos nuevos métodos para la igualación de canales MIMO cuando el orden del canal es desconocido. El primero de ellos es también un algoritmo de filtrado de partículas, mientras que el segundo es un algoritmo de programación dinámica basado en la idea de procesado por supervivientes [66].

6.4.1. Métodos SMC de complejidad reducida

El método explicado en esta sección extiende el esquema de complejidad reducida de la Sección 5.2 para incorporar y explotar la información de la f.d.p. *a posteriori*

Pseudocódigo 9 Algoritmo SIS óptimo suavizado para orden desconocido

para cada instante de tiempo $t = 1, 2, \dots$
para cada partícula $i = 1, \dots, M$
 sea $acum_{verosimil} \leftarrow 0$
para todo $\mathbf{s}_t \in \mathcal{S}^{N_t}$
 $verosimil(\mathbf{s}_t) \leftarrow 0$
para todo $m \in \mathcal{M}$
 $P_o(m) \leftarrow p(\mathbf{y}_t | m, \mathbf{s}_{0:t}, \mathbf{y}_{0:t-1})$
 guardar el estado del i -ésimo KF (media y covarianza predictivas y filtradas) para orden de canal m
para todo $\tilde{\mathbf{s}}_{t+1:t+a} \in \mathcal{S}^{N_{t+a}}$
 sea $P(m) \leftarrow P_o(m)$
para $k = 1, \dots, a$
 actualizar el i -ésimo KF para orden de canal m con el vector \mathbf{s}_{t+k-1}
 $P(m) \leftarrow P(m) \times p(\mathbf{y}_{t+k} | m, \mathbf{s}_{0:t}, \tilde{\mathbf{s}}_{t+1:t+k}, \mathbf{y}_{0:t+k-1})$
 $P \leftarrow P \times p(m, |\mathbf{s}_{0:t-1}^{(i)}, \mathbf{y}_{0:t-1})$
 $verosimil(\mathbf{s}_t) \leftarrow verosimil(\mathbf{s}_t) + P(m)$
 restaurar el estado del i -ésimo KF para orden de canal m
 calcular la probabilidad *a posteriori* del orden del canal necesaria en $t+1$ utilizando (6.11), $p(m, |\mathbf{s}_{0:t}^{(i)}, \mathbf{y}_{0:t}) \leftarrow P_o(m)p(m, |\mathbf{s}_{0:t-1}^{(i)}, \mathbf{y}_{0:t-1})$
 $acum_{verosimil} \leftarrow acum_{verosimil} + verosimil(\mathbf{s}_t)$
 muestrear $\mathbf{s}_t^{(i)} \sim q_t(\mathbf{s}_t) = verosimil(\mathbf{s}_t) / acum_{verosimil}$
 obtener el peso sin normalizar $\tilde{w}_t^{(i)} \leftarrow w_{t-1}^{(i)} acum_{verosimil}$
para todo $m \in \mathcal{M}$
 actualizar el i -ésimo KF para orden de canal m utilizando la muestra $\mathbf{s}_t^{(i)}$
 calcular $W \leftarrow \sum_{i=1}^M \tilde{w}_t^{(i)}$
para cada partícula $i = 1, \dots, M$
 normalizar su peso $w_t^{(i)} \leftarrow \tilde{w}_t^{(i)} / W$
 estimar el tamaño efectivo de la muestra, $\hat{M}_{ef} \leftarrow \left(\sum_{i=1}^M w_t^{(i)2} \right)^{-1}$
si $\hat{M}_{ef} < \gamma_r M$ **entonces**
 remuestrear

del orden del canal. En dicha sección nos planteábamos la aproximación de la densidad $p(\mathbf{s}_{0:t+a}, \mathbf{H}_{0:t+a} | \mathbf{y}_{0:t+a})$ considerando el orden del canal como un parámetro conocido del sistema. Entonces incluíamos la estimación del canal en el filtro de partículas y obteníamos muestras del mismo para facilitar la detección de los datos. Para utilizar la misma aproximación al problema cuando el orden del canal es desconocido vamos tomar muestras del canal considerando cada uno de los posibles órdenes, $m \in \mathcal{M}$. Así, nuestra f.d.p. objetivo es ahora $p(\mathbf{s}_{0:t+a}, \{\mathbf{H}_{m;0:t+a}\}_{m \in \mathcal{M}} | \mathbf{y}_{0:t+a})$, donde $\mathbf{H}_{m;0:t+a}$ es la secuencia de matrices de canal (de tamaño $N_r \times N_t m$) entre los instantes 0 y $t+a$ asumiendo que m es el orden del canal.

Al igual que en el método de la Sección 6.3, vamos a elegir como factor de suavizado $a = \max\{\mathcal{M}\} - 1$ de tal manera que a la hora de detectar un determinado vector de símbolos se tengan en cuenta todas las observaciones a lo largo de las cuales se distribuye su energía aún cuando el orden del canal sea máximo.

Utilizando la notación indicada, la densidad objetivo se puede desarrollar por medio del teorema de Bayes como

$$\begin{aligned} p(\mathbf{s}_{0:t+a}, \{\mathbf{H}_{m;0:t+a}\}_{m \in \mathcal{M}} | \mathbf{y}_{0:t+a}) &= p(\mathbf{y}_{t:t+a} | \mathbf{s}_{0:t+a}, \{\mathbf{H}_{m;0:t+a}\}_{m \in \mathcal{M}}, \mathbf{y}_{0:t-1}) \times \\ &\quad \times \frac{p(\mathbf{s}_{0:t+a}, \{\mathbf{H}_{m;0:t+a}\}_{m \in \mathcal{M}} | \mathbf{y}_{0:t-1})}{p(\mathbf{y}_{t+a} | \mathbf{y}_{0:t-1})} \\ &\propto p(\mathbf{y}_{t:t+a} | \mathbf{s}_{0:t+a}, \{\mathbf{H}_{m;0:t+a}\}_{m \in \mathcal{M}}, \mathbf{y}_{0:t-1}) \times \\ &\quad \times p(\mathbf{s}_{0:t+a}, \{\mathbf{H}_{m;0:t+a}\}_{m \in \mathcal{M}} | \mathbf{y}_{0:t-1}). \end{aligned} \quad (6.15)$$

Reescribiendo el último término que aparece en el lado derecho de (6.15) mediante la definición de probabilidad condicionada resulta

$$\begin{aligned} p(\mathbf{s}_{0:t+a}, \{\mathbf{H}_{m;0:t+a}\}_{m \in \mathcal{M}} | \mathbf{y}_{0:t+a}) &= p(\mathbf{y}_{t:t+a} | \mathbf{s}_{0:t+a}, \{\mathbf{H}_{m;0:t+a}\}_{m \in \mathcal{M}}, \mathbf{y}_{0:t-1}) \times \\ &\quad \times p(\{\mathbf{H}_{m;t:t+a}\}_{m \in \mathcal{M}} | \mathbf{s}_{0:t+a}, \{\mathbf{H}_{m;0:t-1}\}_{m \in \mathcal{M}}, \mathbf{y}_{0:t-1}) \times \\ &\quad \times p(\mathbf{s}_{0:t+a}, \{\mathbf{H}_{m;0:t-1}\}_{m \in \mathcal{M}} | \mathbf{y}_{0:t-1}) \end{aligned} \quad (6.16)$$

y aplicando nuevamente el teorema de Bayes y la definición de probabilidad condicionada sobre $p(\mathbf{s}_{0:t+a}, \{\mathbf{H}_{m;0:t-1}\}_{m \in \mathcal{M}} | \mathbf{y}_{0:t-1})$ tenemos

$$\begin{aligned} p(\mathbf{s}_{0:t+a}, \{\mathbf{H}_{m;0:t+a}\}_{m \in \mathcal{M}} | \mathbf{y}_{0:t+a}) &\propto p(\mathbf{y}_{t:t+a} | \mathbf{s}_{0:t+a}, \{\mathbf{H}_{m;0:t+a}\}_{m \in \mathcal{M}}, \mathbf{y}_{0:t-1}) \times \\ &\quad \times p(\{\mathbf{H}_{m;t:t+a}\}_{m \in \mathcal{M}} | \{\mathbf{H}_{m;0:t-1}\}_{m \in \mathcal{M}}) \times \\ &\quad \times p(\mathbf{s}_{t:t+a} | \mathbf{s}_{0:t-1}, \{\mathbf{H}_{m;0:t-1}\}_{m \in \mathcal{M}}, \mathbf{y}_{0:t-1}) \times \\ &\quad \times p(\mathbf{s}_{0:t-1}, \{\mathbf{H}_{m;0:t-1}\}_{m \in \mathcal{M}} | \mathbf{y}_{0:t-1}) \\ &\propto p(\mathbf{y}_{t:t+a} | \mathbf{s}_{0:t+a}, \{\mathbf{H}_{m;0:t+a}\}_{m \in \mathcal{M}}, \mathbf{y}_{0:t-1}) \times \\ &\quad \times p(\{\mathbf{H}_{m;t:t+a}\}_{m \in \mathcal{M}} | \{\mathbf{H}_{m;0:t-1}\}_{m \in \mathcal{M}}) \times \\ &\quad \times p(\mathbf{s}_{0:t-1}, \{\mathbf{H}_{m;0:t-1}\}_{m \in \mathcal{M}} | \mathbf{y}_{0:t-1}) \end{aligned} \quad (6.17)$$

donde la proporcionalidad se debe a que $p(\mathbf{s}_{t:t+a} | \mathbf{s}_{0:t-1}, \{\mathbf{H}_{m;0:t-1}\}_{m \in \mathcal{M}}, \mathbf{y}_{0:t-1})$ es una constante por ser la probabilidad *a priori* de los símbolos⁷. Por otro lado, al ser el canal un proceso de

⁷Nótese que las observaciones $\mathbf{y}_{0:t-1}$ no contienen ninguna información acerca de los vectores $\mathbf{s}_{t:t+a}$.

Markov, la matriz de canal en un instante de tiempo dado asumiendo un determinado orden de canal sólo depende de la matriz de canal (asumiendo el mismo orden) en el instante anterior o, equivalentemente,

$$p(\mathbf{H}_{m;t} | \mathbf{H}_{m;t-1}, \cdot) = p(\mathbf{H}_{m;t} | \mathbf{H}_{m;t-1}), \quad (6.18)$$

por lo que podemos escribir

$$p(\{\mathbf{H}_{m;t:t+a}\}_{m \in \mathcal{M}} | \mathbf{s}_{0:t+a}, \{\mathbf{H}_{m;0:t-1}\}_{m \in \mathcal{M}}, \mathbf{y}_{0:t-1}) = \prod_{m \in \mathcal{M}} \prod_{k=0}^a p(\mathbf{H}_{m;t+k} | \mathbf{H}_{m;t+k-1}). \quad (6.19)$$

En cuanto a la verosimilitud $p(\mathbf{y}_{t:t+a} | \mathbf{s}_{0:t+a}, \{\mathbf{H}_{m;0:t+a}\}_{m \in \mathcal{M}}, \mathbf{y}_{0:t-1})$, para calcularla sin conocer el orden del canal vamos a recurrir al teorema de la de la probabilidad total del mismo modo que hicimos en la sección anterior para calcular $p(\mathbf{y}_{t:t+a} | \mathbf{s}_{0:t}, \mathbf{y}_{0:t-1})$. Así pues, vamos a descomponer dicha probabilidad como

$$\begin{aligned} p(\mathbf{y}_{t:t+a} | \mathbf{s}_{0:t+a}, \{\mathbf{H}_{m;0:t+a}\}_{m \in \mathcal{M}}, \mathbf{y}_{0:t-1}) &= \sum_{m \in \mathcal{M}} p(\mathbf{y}_{t:t+a}, m | \mathbf{s}_{0:t+a}, \{\mathbf{H}_{m;0:t+a}\}_{m \in \mathcal{M}}, \mathbf{y}_{0:t-1}) \\ &= \sum_{m \in \mathcal{M}} p(\mathbf{y}_{t:t+a} | m, \mathbf{s}_{0:t+a}, \{\mathbf{H}_{m;0:t+a}\}_{m \in \mathcal{M}}, \mathbf{y}_{0:t-1}) \times \\ &\quad \times p(m | \mathbf{s}_{0:t+a}, \{\mathbf{H}_{m;0:t+a}\}_{m \in \mathcal{M}}, \mathbf{y}_{0:t-1}) \\ &= \sum_{m \in \mathcal{M}} p(\mathbf{y}_{t:t+a} | \mathbf{s}_{0:t+a}, \mathbf{H}_{m;0:t+a}, \mathbf{y}_{0:t-1}) \times \\ &\quad \times p(m | \mathbf{s}_{0:t+a}, \{\mathbf{H}_{m;0:t+a}\}_{m \in \mathcal{M}}, \mathbf{y}_{0:t-1}) \\ &= \sum_{m \in \mathcal{M}} p(\mathbf{y}_{t:t+a} | \mathbf{s}_{0:t+a}, \mathbf{H}_{m;0:t+a}, \mathbf{y}_{0:t-1}) \times \\ &\quad \times p(m | \mathbf{s}_{0:t-1}, \{\mathbf{H}_{m;0:t-1}\}_{m \in \mathcal{M}}, \mathbf{y}_{0:t-1}) \end{aligned} \quad (6.20)$$

donde hemos utilizado que, trivialmente,

$$p(\mathbf{y}_{t:t+a} | m, \mathbf{s}_{0:t+a}, \{\mathbf{H}_{m;0:t+a}\}_{m \in \mathcal{M}}, \mathbf{y}_{0:t-1}) = p(\mathbf{y}_{t:t+a} | \mathbf{s}_{0:t+a}, \mathbf{H}_{m;0:t+a}, \mathbf{y}_{0:t-1}), \quad (6.21)$$

es decir, si m es el verdadero orden del canal, podemos descartar las estimaciones de canal que suponen cualquiera de los restantes órdenes, $\{m' \in \mathcal{M}, m' \neq m\}$, puesto que ya no van a ser de utilidad en el cálculo de la probabilidad de las observaciones $\mathbf{y}_{t:t+a}$. Además,

$$p(m | \mathbf{s}_{0:t+a}, \{\mathbf{H}_{m;0:t+a}\}_{m \in \mathcal{M}}, \mathbf{y}_{0:t-1}) = p(m | \mathbf{s}_{0:t-1}, \{\mathbf{H}_{m;0:t-1}\}_{m \in \mathcal{M}}, \mathbf{y}_{0:t-1}), \quad (6.22)$$

porque los símbolos y la secuencia de matrices de canal desde t hasta $t+a$ (ésta última considerando todos los posibles órdenes de canal, $\{\mathbf{H}_{m;0:t+a}\}_{m \in \mathcal{M}}$) no aportan ninguna información acerca de m sin las correspondientes observaciones.

La probabilidad *a posteriori* del orden del canal se puede calcular por medio del teorema de Bayes como

$$\begin{aligned}
 p(m|\mathbf{s}_{0:t-1}, \{\mathbf{H}_{m';0:t-1}\}_{m' \in \mathcal{M}}, \mathbf{y}_{0:t-1}) &= p(\mathbf{y}_{0:t-1}|m, \mathbf{s}_{0:t-1}, \{\mathbf{H}_{m';0:t-1}\}_{m' \in \mathcal{M}}, \mathbf{y}_{0:t-2}) \times \\
 &\quad \times \frac{p(m|\mathbf{s}_{0:t-1}, \{\mathbf{H}_{m';0:t-1}\}_{m' \in \mathcal{M}}, \mathbf{y}_{0:t-2})}{p(\mathbf{y}_{0:t-1}|\mathbf{s}_{0:t-1}, \{\mathbf{H}_{m';0:t-1}\}_{m' \in \mathcal{M}}, \mathbf{y}_{0:t-2})} \\
 &\propto p(\mathbf{y}_{0:t-1}|m, \mathbf{s}_{0:t-1}, \{\mathbf{H}_{m';0:t-1}\}_{m' \in \mathcal{M}}, \mathbf{y}_{0:t-2}) \times \\
 &\quad \times p(m|\mathbf{s}_{0:t-1}, \{\mathbf{H}_{m';0:t-1}\}_{m' \in \mathcal{M}}, \mathbf{y}_{0:t-2}) \\
 &= p(\mathbf{y}_{0:t-1}|m, \mathbf{s}_{0:t-1}, \mathbf{H}_{m;0:t-1}, \mathbf{y}_{0:t-2}) \times \\
 &\quad \times p(m|\mathbf{s}_{0:t-2}, \{\mathbf{H}_{m';0:t-2}\}_{m' \in \mathcal{M}}, \mathbf{y}_{0:t-2}). \quad (6.23)
 \end{aligned}$$

Si expandimos cada una de las probabilidades *a posteriori* del orden del canal que surgen en el desarrollo de (6.23) podemos llegar a

$$p(m|\mathbf{s}_{0:t-1}, \{\mathbf{H}_{m;0:t-1}\}_{m \in \mathcal{M}}, \mathbf{y}_{0:t-1}) \propto p(m) \prod_{k=0}^{t-1} p(\mathbf{y}_{0:k}|m, \mathbf{s}_{0:k}, \mathbf{H}_{m;0:k}, \mathbf{y}_{0:k-1}), \quad (6.24)$$

lo que implica que $p(m|\mathbf{s}_{0:t-1}, \{\mathbf{H}_{m;0:t-1}\}_{m \in \mathcal{M}}, \mathbf{y}_{0:t-1})$ únicamente depende de las matrices de canal que asumen orden m , de tal manera que

$$p(m|\mathbf{s}_{0:t-1}, \{\mathbf{H}_{m;0:t-1}\}_{m \in \mathcal{M}}, \mathbf{y}_{0:t-1}) = p(m|\mathbf{s}_{0:t-1}, \mathbf{H}_{m;0:t-1}, \mathbf{y}_{0:t-1}). \quad (6.25)$$

Sustituyendo (6.19) y (6.20) en (6.17) resulta la f.d.p. objetivo que consideraremos de aquí en adelante:

$$\begin{aligned}
 p(\mathbf{s}_{0:t+a}, \{\mathbf{H}_{m;0:t+a}\}_{m \in \mathcal{M}} | \mathbf{y}_{0:t+a}) &= \sum_{m' \in \mathcal{M}} p(\mathbf{y}_{t:t+a} | \mathbf{s}_{0:t+a}, \mathbf{H}_{m';0:t+a}, \mathbf{y}_{0:t-1}) \times \\
 &\quad \times p(m' | \mathbf{s}_{0:t-1}, \mathbf{H}_{m';0:t-1}, \mathbf{y}_{0:t-1}) \prod_{m'' \in \mathcal{M}} \prod_{k=0}^a p(\mathbf{H}_{m'';t+k} | \mathbf{H}_{m'';t+k-1}) \times \\
 &\quad \times p(\mathbf{s}_{0:t-1}, \{\mathbf{H}_{m;0:t-1}\}_{m \in \mathcal{M}} | \mathbf{y}_{0:t-1}). \quad (6.26)
 \end{aligned}$$

Al igual que ocurría en el esquema de la Sección 5.2, las verosimilitudes que aparecen en las ecuaciones obtenidas se pueden calcular fácilmente al estar condicionadas en el canal. De esa manera evitamos el uso de bancos de filtros de Kalman.

Para aproximar la densidad $p(\mathbf{s}_{0:t+a}, \{\mathbf{H}_{m;0:t+a}\}_{m \in \mathcal{M}} | \mathbf{y}_{0:t+a})$ vamos a definir una función tentativa de la forma

$$\begin{aligned}
 q(\mathbf{s}_{0:t+a}, \{\mathbf{H}_{m;0:t+a}\}_{m \in \mathcal{M}} | \mathbf{y}_{0:t+a}) &= q(\mathbf{s}_{t:t+a} | \mathbf{s}_{0:t-1}, \{\mathbf{H}_{m;0:t+a}\}_{m \in \mathcal{M}}, \mathbf{y}_{0:t+a}) \times \\
 &\quad \times q(\{\mathbf{H}_{m;t:t+a}\}_{m \in \mathcal{M}} | \mathbf{s}_{0:t-1}, \{\mathbf{H}_{m;0:t-1}\}_{m \in \mathcal{M}}, \mathbf{y}_{0:t+a}) \times \\
 &\quad \times q(\mathbf{s}_{0:t-1}, \{\mathbf{H}_{m;0:t-1}\}_{m \in \mathcal{M}} | \mathbf{y}_{0:t+a}). \quad (6.27)
 \end{aligned}$$

Utilizando (6.26) y (6.27) para obtener la ecuación de actualización de los pesos, resulta el siguiente método de filtrado de partículas

$$\left\{ \mathbf{H}_{m;t:t+a}^{(i)} \right\}_{m \in \mathcal{M}} \sim q(\{ \mathbf{H}_{m;t:t+a} \}_{m \in \mathcal{M}} | \mathbf{s}_{0:t-1}, \{ \mathbf{H}_{m;0:t-1} \}_{m \in \mathcal{M}}, \mathbf{y}_{0:t+a}) \quad (6.28)$$

$$\mathbf{s}_{t:t+a}^{(i)} \sim q(\mathbf{s}_{t:t+a} | \mathbf{s}_{0:t-1}, \{ \mathbf{H}_{m;0:t+a} \}_{m \in \mathcal{M}}, \mathbf{y}_{0:t+a}) \quad (6.29)$$

$$\begin{aligned} w_{t+a}^{(i)} &= \frac{p(\mathbf{s}_{0:t+a}^{(i)}, \{ \mathbf{H}_{m;0:t+a}^{(i)} \}_{m \in \mathcal{M}} | \mathbf{y}_{0:t+a})}{q(\mathbf{s}_{0:t+a}^{(i)}, \{ \mathbf{H}_{m;0:t+a}^{(i)} \}_{m \in \mathcal{M}} | \mathbf{y}_{0:t+a})} \\ &= \frac{\sum_{m' \in \mathcal{M}} p(\mathbf{y}_{t:t+a} | \mathbf{s}_{0:t+a}, \mathbf{H}_{m';0:t+a}, \mathbf{y}_{0:t-1})}{q(\mathbf{s}_{t:t+a}^{(i)} | \mathbf{s}_{0:t-1}, \{ \mathbf{H}_{m;0:t+a}^{(i)} \}_{m \in \mathcal{M}}, \mathbf{y}_{0:t+a})} \times \\ &\quad \times \frac{p(m' | \mathbf{s}_{0:t-1}, \mathbf{H}_{m';0:t-1}, \mathbf{y}_{0:t-1}) \prod_{m'' \in \mathcal{M}} \prod_{k=0}^a p(\mathbf{H}_{m'';t+k} | \mathbf{H}_{m'';t+k-1})}{q(\{ \mathbf{H}_{m;t:t+a} \}_{m \in \mathcal{M}} | \mathbf{s}_{0:t-1}, \{ \mathbf{H}_{m;0:t-1} \}_{m \in \mathcal{M}}, \mathbf{y}_{0:t+a})} w_{t+a-1}^{(i)}. \end{aligned} \quad (6.30)$$

Así pues, una vez más en la función tentativa aparecen dos densidades marginales que servirán, una para el muestreo del canal, y otra para el muestreo de los datos. Vamos a examinar cada una de ellas en una sección diferente.

Muestreo del canal

La forma de la función $q(\{ \mathbf{H}_{m;t:t+a} \}_{m \in \mathcal{M}} | \mathbf{s}_{0:t-1}, \{ \mathbf{H}_{m;0:t-1} \}_{m \in \mathcal{M}}, \mathbf{y}_{0:t+a})$ implica que tenemos que tomar muestras de la secuencia de matrices de canal, $\mathbf{H}_{m;t:t+a}$, para cada uno de los posibles órdenes de canal $m \in \mathcal{M}$. Como ya se dijo anteriormente, este método de filtrado de partículas no es sino una extensión del de la Sección 5.2 para manejar la incertidumbre asociada al orden del canal. Para muestrear el canal en dicha sección utilizábamos el algoritmo RLS para obtener una estimación del mismo en cada instante de tiempo y para cada partícula, de manera que necesitábamos un banco de M algoritmos RLS, uno por partícula. Ahora tendremos un algoritmo RLS para cada posible orden del canal, $m \in \mathcal{M}$, en cada partícula, lo que da un total de $|\mathcal{M}|M$ estimadores RLS.

Vamos a obtener las muestras del canal considerando un determinado orden de canal independientemente de las de los restantes órdenes, lo que da lugar a una función tentativa de la forma

$$\begin{aligned} q(\{ \mathbf{H}_{m;t:t+a} \}_{m \in \mathcal{M}} | \mathbf{s}_{0:t-1}, \{ \mathbf{H}_{m;0:t-1} \}_{m \in \mathcal{M}}, \mathbf{y}_{0:t+a}) &= \\ \prod_{m \in \mathcal{M}} q(\mathbf{H}_{m;t:t+a} | \mathbf{s}_{0:t-1}, \{ \mathbf{H}_{m;0:t-1} \}_{m \in \mathcal{M}}, \mathbf{y}_{0:t+a}). \end{aligned} \quad (6.31)$$

Cada uno de los factores $q(\mathbf{H}_{m;t:t+a} | \mathbf{s}_{0:t-1}, \{ \mathbf{H}_{m;0:t-1} \}_{m \in \mathcal{M}}, \mathbf{y}_{0:t+a})$ en (6.31) se puede escribir atendiendo a la ecuación (6.18) como

$$q(\mathbf{H}_{m;t:t+a} | \mathbf{s}_{0:t-1}, \{ \mathbf{H}_{m;0:t-1} \}_{m \in \mathcal{M}}, \mathbf{y}_{0:t+a}) = q(\mathbf{H}_{m;t:t+a} | \mathbf{s}_{0:t-1}, \mathbf{H}_{m;0:t-1}, \mathbf{y}_{0:t+a}), \quad (6.32)$$

de tal manera que sustituyendo (6.32) en (6.31) resulta

$$q(\{\mathbf{H}_{m;t:t+a}\}_{m \in \mathcal{M}} | \mathbf{s}_{0:t-1}, \{\mathbf{H}_{m;0:t-1}\}_{m \in \mathcal{M}}, \mathbf{y}_{0:t+a}) = \prod_{m \in \mathcal{M}} q(\mathbf{H}_{m;t:t+a} | \mathbf{s}_{0:t-1}, \mathbf{H}_{m;0:t-1}, \mathbf{y}_{0:t+a}). \quad (6.33)$$

Las muestras del canal en un determinado instante se van a obtener por medio de las correspondientes estimaciones en el instante anterior. En particular, una muestra, $\mathbf{H}_{m;t}^{(i)}$ (o $\mathbf{h}_{m;t}^{(i)}$, recurriendo a su representación vectorial, más conveniente en este caso), de la matriz de canal en el instante t considerando m como el verdadero orden del canal, se va a generar de acuerdo a la distribución gaussiana,

$$\mathbf{h}_{m;t}^{(i)} \sim q(\mathbf{h}_{m;t} | \mathbf{s}_{0:t-1}^{(i)}, \mathbf{h}_{m;0:t-1}^{(i)}, \mathbf{y}_{0:t+a}) = \mathcal{N}(\mathbf{h}_{m;t} | \alpha \tilde{\mathbf{h}}_{m;t-1}^{(i)}, \sigma_H^2 \mathbf{I}), \quad (6.34)$$

donde $\tilde{\mathbf{h}}_{m;t-1}^{(i)}$ es la salida del estimador de canal de la partícula i para orden m en el instante $t-1$ (obtenida a partir de las secuencias de observaciones, $\mathbf{s}_{0:t-1}^{(i)}$, y símbolos, $\mathbf{y}_{0:t+a}$), α es el coeficiente del modelo de canal AR y σ_H^2 es un parámetro de diseño. Las restantes muestras del canal desde $t+1$ hasta $t+a$ se obtienen a partir de $\mathbf{h}_{m;t}^{(i)}$ utilizando directamente el modelo AR:

$$\begin{aligned} \mathbf{h}_{m;t+1:t+a}^{(i)} \sim q(\mathbf{h}_{m;t+1:t+a} | \mathbf{h}_{m;t}^{(i)}) &= p(\mathbf{h}_{m;t+1:t+a} | \mathbf{h}_{m;t}^{(i)}) = \prod_{k=1}^a \mathcal{N}(\mathbf{h}_{m;t+k} | \alpha \mathbf{h}_{m;t+k-1}^{(i)}, \sigma_v^2 \mathbf{I}) \\ &= \prod_{k=1}^a p(\mathbf{h}_{m;t+k} | \mathbf{h}_{m;t+k-1}^{(i)}). \end{aligned} \quad (6.35)$$

Utilizando las ecuaciones (6.34) y (6.35) podemos desarrollar (6.33) como

$$\begin{aligned} q(\{\mathbf{H}_{m;t:t+a}\}_{m \in \mathcal{M}} | \mathbf{s}_{0:t-1}, \{\mathbf{H}_{m;0:t-1}\}_{m \in \mathcal{M}}, \mathbf{y}_{0:t+a}) &= \prod_{m \in \mathcal{M}} q(\mathbf{H}_{m;t:t+a} | \mathbf{s}_{0:t-1}, \mathbf{H}_{m;0:t-1}, \mathbf{y}_{0:t+a}) \\ &= \prod_{m \in \mathcal{M}} q(\mathbf{H}_{m;t} | \mathbf{s}_{0:t-1}, \mathbf{H}_{m;0:t-1}, \mathbf{y}_{0:t+a}) \times \\ &\quad \times q(\mathbf{H}_{m;t+1:t+a} | \mathbf{H}_{m;t}) \\ &= \prod_{m \in \mathcal{M}} \mathcal{N}(\mathbf{h}_{m;t} | \alpha \tilde{\mathbf{h}}_{m;t-1}, \sigma_H^2 \mathbf{I}) \times \\ &\quad \times \prod_{k=1}^a p(\mathbf{h}_{m;t+k} | \mathbf{h}_{m;t+k-1}), \end{aligned} \quad (6.36)$$

que constituye la expresión final para la función tentativa del canal.

El estimador de canal que nos va a proporcionar estimaciones del canal en el instante anterior, $\tilde{\mathbf{h}}_{m;t-1}^{(i)}$, para cada partícula i y cada posible orden de canal, $m \in \mathcal{M}$, es el algoritmo RLS, que se encuentra formulado y derivado en el Apéndice C. Aplicando (C.11),

(C.12) y (C.15) al caso que nos ocupa e inicializando la matriz de correlación, $\mathbf{R}_{m;t} = \sum_{k=0}^t \lambda_{\text{RLS}}^{t-k} \mathbf{s}_{k-m+1}^{(i)} \mathbf{s}_{k-m+1}^{(i)\top}$, a la identidad, resultan las ecuaciones

$$\mathbf{R}_{m;0}^{(i)-1} \propto \mathbf{I}_{N_t m} \quad (\text{inicialización}) \quad (6.37)$$

$$\mathbf{g}_{m;t}^{(i)} = \frac{\lambda_{\text{RLS}}^{-1} \mathbf{R}_{m;t-1}^{(i)-1} \mathbf{s}_{k-m+1}^{(i)}}{1 + \lambda_{\text{RLS}}^{-1} \mathbf{s}_{k-m+1}^{(i)\top} \mathbf{R}_{m;t-1}^{(i)-1} \mathbf{s}_{k-m+1}^{(i)}} \quad (6.38)$$

$$\tilde{\mathbf{H}}_{m;t}^{(i)\top} = \tilde{\mathbf{H}}_{m;t-1}^{(i)\top} + \mathbf{g}_{m;t}^{(i)} \left(\mathbf{y}_t^\top - \mathbf{s}_{k-m+1}^{(i)\top} \tilde{\mathbf{H}}_{m;t-1}^{(i)\top} \right) \quad (6.39)$$

$$\mathbf{R}_{m;t}^{(i)-1} = \lambda_{\text{RLS}}^{-1} \left(\mathbf{I}_{N_t m} - \mathbf{g}_{m;t}^{(i)} \mathbf{s}_{k-m+1}^{(i)\top} \right) \mathbf{R}_{m;t-1}^{(i)-1}, \quad (6.40)$$

donde el subíndice m que acompaña a la matriz de correlación de los símbolos, $\mathbf{R}_{m;0}^{(i)-1}$, y al vector de ganancia, $\mathbf{g}_{m;t}^{(i)}$, se refiere (igual que en el caso de la matriz de canal $\mathbf{H}_{m;t-1}^{(i)}$) al orden de canal considerado.

Mediante las ecuaciones (6.37)-(6.40) podemos obtener estimaciones de la matriz de canal recursivamente en cada instante de tiempo y considerando cada uno de los posibles órdenes $m \in \mathcal{M}$.

Muestreo de los datos

Una vez tenemos muestras del canal entre los instantes t y $t+a$ considerando cada uno de los órdenes posibles, $\{\mathbf{H}_{m;t:t+a}\}_{m \in \mathcal{M}}$, vamos utilizarlas para facilitar la detección de los símbolos transmitidos. La forma de la función tentativa para los datos, que ya vimos en la ecuación (6.29), muestra esto explícitamente.

En la Sección 5.2, cuando el orden del canal era conocido, las muestras del canal servían para construir un estimador lineal que daba lugar a estimaciones blandas de los símbolos transmitidos, las cuales a su vez permitían obtener probabilidades para muestrear dichos símbolos. Ahora estamos considerando $|\mathcal{M}|$ órdenes de canal diferentes, cada uno de los cuales tiene asociada una estimación diferente para la secuencia de matrices de canal entre los instantes t y $t+a$, y queremos obtener probabilidades para los símbolos transmitidos que tengan en cuenta toda esta información. Para ello vamos a utilizar el filtro MMSE de complejidad reducida,

explicado en la Sección 5.2.2, que ahora viene dado por las ecuaciones

$$\mathbf{R}_{y,0}^{-1} = \mathbf{I}_{N_r m} \quad (6.41)$$

$$\mathbf{g}_{t,m-1} = \frac{\mathbf{R}_{y,t-1}^{-1} \mathbf{y}_{t+m-1}^t}{\Lambda_t + \mathbf{y}_{t+m-1}^T \frac{\Lambda_t}{\lambda_{\text{MMSE}} \Lambda_{t-1}} \mathbf{R}_{y,t-1}^{-1} \mathbf{y}_{t+m-1}^t} \quad (6.42)$$

$$\mathbf{R}_{y,t}^{-1} \approx \frac{\Lambda_t}{\lambda_{\text{MMSE}} \Lambda_{t-1}} \left(\mathbf{I}_{N_r m} - \mathbf{g}_{t,m-1} \mathbf{y}_{t+m-1}^T \frac{\Lambda_t}{\lambda_{\text{MMSE}} \Lambda_{t-1}} \right) \mathbf{R}_{y,t-1}^{-1} \quad (6.43)$$

$$\mathbf{F}_{m;t}^{(i)\top} = \sigma_s^2 \mathbf{H}_{m;t,m-1}^{(i)\top} \mathbf{R}_{y,t}^{-1} \quad (6.44)$$

donde la matriz $\mathbf{H}_{m;t,m-1}^{(i)}$ se construye de la manera indicada en (3.10) pero utilizando las muestras $\mathbf{H}_{m;t,t+m-1}^{(i)}$ (que asumen orden m) en lugar de las auténticas matrices de canal, es decir

$$\mathbf{H}_{m;t,m-1}^{(i)} = \begin{bmatrix} \mathbf{H}_{m;t}^{(i)}(m-1)^\top & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{H}_{m;t}^{(i)}(m-2)^\top & \mathbf{H}_{m;t+1}^{(i)}(m-1)^\top & \cdots & \mathbf{0} \\ \vdots & \mathbf{H}_{m;t+1}^{(i)}(m-2)^\top & \ddots & \vdots \\ \mathbf{H}_{m;t}^{(i)}(0)^\top & \vdots & \ddots & \mathbf{H}_{m;t+m-1}^{(i)}(m-1)^\top \\ \vdots & \mathbf{H}_{m;t+1}^{(i)}(0)^\top & \ddots & \mathbf{H}_{m;t+m-1}^{(i)}(m-2)^\top \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{H}_{m;t+m-1}^{(i)}(0)^\top \end{bmatrix}^\top \quad (6.45)$$

Utilizando filtros lineales así contruidos podemos obtener vectores de estimaciones blandas de los símbolos transmitidos del mismo modo que hacíamos en el Capítulo 5, pero ahora considerando cada uno de los órdenes de canal posibles,

$$\tilde{\mathbf{s}}_{m;t+m-1}^{(i)} = \begin{bmatrix} \tilde{\mathbf{s}}_{m;t}^{(i)} \\ \tilde{\mathbf{s}}_{m;t+1}^{(i)} \\ \vdots \\ \tilde{\mathbf{s}}_{m;t+m-1}^{(i)} \end{bmatrix} = \mathbf{E}_m \mathbf{F}_{m;t}^{(i)\top} \mathbf{y}_{t+m-1}^t \quad (6.46)$$

donde $\mathbf{E}_m = [\mathbf{0}_{N_t(a+1) \times N_t(m-1)} \mathbf{I}_{N_t m}]$ y $\tilde{\mathbf{s}}_{m;t}^{(i)}$ es un vector que contiene las estimaciones blandas de los símbolos en \mathbf{s}_t asumiendo que el orden de canal es m .

El filtro $\mathbf{F}_{m;t}^{(i)}$ tiene dimensiones $N_r m \times N_t(2m-1)$ y permite obtener estimaciones blandas de los vectores de símbolos desde $t-m+1$ hasta $t+m-1$. Teniendo en cuenta que la matriz \mathbf{E}_m suprime las estimaciones blandas correspondientes a los vectores de símbolos $\mathbf{s}_{t-m+1:t-1}$, ya detectados en instantes de tiempo anteriores, tenemos que el resultado de multiplicar $\mathbf{E}_m \mathbf{F}_{m;t}^{(i)\top}$ por las observaciones en (6.46) es un vector que apila las estimaciones blandas de los vectores $\mathbf{s}_{t:t+m-1}$. Así pues, cuando $m = \max\{\mathcal{M}\}$, el filtro a que dan lugar las ecuaciones (6.41)-(6.44)

nos permite obtener las estimaciones blandas de los símbolos entre t y $t + \text{máx}\{\mathcal{M}\} - 1 = t + a$ que queremos. Sin embargo, cuando $m < \text{máx}\{\mathcal{M}\}$ (el orden de canal considerado no es el máximo que puede tomar el canal), es necesario aplicar una secuencia de filtros

$$\mathbf{F}_{m;t}^{(i)}, \dots, \mathbf{F}_{m;t+|\mathcal{M}|-m}^{(i)}, \quad (6.47)$$

donde

$$\mathbf{F}_{m;t+j}^{(i)} = \sigma_s^2 \mathbf{H}_{m;t+j,m-1}^{(i)\top} \mathbf{R}_{y,t+j}^{-1}, j = 0, \dots, |\mathcal{M}| - m \quad (6.48)$$

para obtener estimaciones blandas de los vectores $\mathbf{s}_{t:t+a}$. La Figura 6.2 ilustra la construcción de los $\text{máx}\{\mathcal{M}\} - m + 1 = a - m + 2$ filtros. En ella se aprecia como las matrices de canal $\mathbf{H}_{m;t:t+a}$ son utilizadas en grupos m para obtener una secuencia de matrices de canal apiladas, $\mathbf{H}_{m;t+j,m-1}^{(i)}$ con $0 \leq j \leq a - m + 1$, que sirven para obtener, mediante la ecuación (6.48), una secuencia de filtros. Cada uno de éstos junto con el correspondiente vector de observaciones apiladas da lugar a un vector de estimaciones blandas, de tal manera que se tienen $a - m + 2$ vectores de estimaciones blandas. Todos los elementos del último de estos vectores y los N_t primeros de los restantes (resaltados en rojo en la figura) constituirán las estimaciones blandas para $\mathbf{s}_{t:t+a}$. Obsérvese que cuanto menor sea el orden de canal considerado, m , mayor es el número de filtros necesarios. Es debido a que no se utiliza un filtro lineal para obtener una estimación blanda de cada uno de los vectores $\mathbf{s}_{t:t+a}$, sino que las estimaciones blandas correspondientes a los m últimos vectores de dicha secuencia se calculan mediante un único filtro (el último).

Siguiendo el procedimiento arriba descrito somos capaces de obtener estimaciones blandas,

$$\tilde{s}_{m;t}(1), \dots, \tilde{s}_{m;t}(N_t), \tilde{s}_{m;t+1}(1), \dots, \tilde{s}_{m;t+a}(N_t),$$

de los símbolos

$$s_t(1), \dots, s_t(N_t), s_{t+1}(1), \dots, s_{t+a}(N_t)$$

considerando cada uno de los posibles órdenes de canal $m \in \mathcal{M}$, que nos van a servir para construir la función tentativa para los datos.

Teniendo en cuenta que $\tilde{s}_{m;t+k}(q)$ es una estimación blanda de $s_{t+k}(q)$ considerando m como el auténtico orden del canal, vamos a calcular la probabilidad de que $s_{t+k}(q)$ tome el valor $s \in \mathcal{S}$, cuando el orden del canal es m , como $\mathcal{N}(\tilde{s}_{m;t+k}(q)|s, \sigma_q^2)$ donde σ_q^2 es un parámetro de diseño. Normalizando dichas probabilidades obtenemos una función tentativa para cada uno de los símbolos en $\mathbf{s}_{t:t+a}$ y cada orden de canal $m \in \mathcal{M}$,

$$q_m(s_{t+k}(q)) = \frac{\mathcal{N}(\tilde{s}_{m;t+k}(q)|s_{t+k}(q), \sigma_q^2)}{\sum_{s \in \mathcal{S}} \mathcal{N}(\tilde{s}_{m;t+k}(q)|s, \sigma_q^2)}. \quad (6.49)$$

Podemos combinar las probabilidades asociadas a los diferentes órdenes utilizando la probabilidad *a posteriori* del orden del canal para obtener una función tentativa independiente del orden de canal considerado:

$$q(s_{t+k}(q)) = \sum_{m \in \mathcal{M}} p(m|\mathbf{s}_{0:t-1}, \mathbf{H}_{m;0:t-1}, \mathbf{y}_{0:t-1}) q_m(s_{t+k}(q)) \quad (6.50)$$

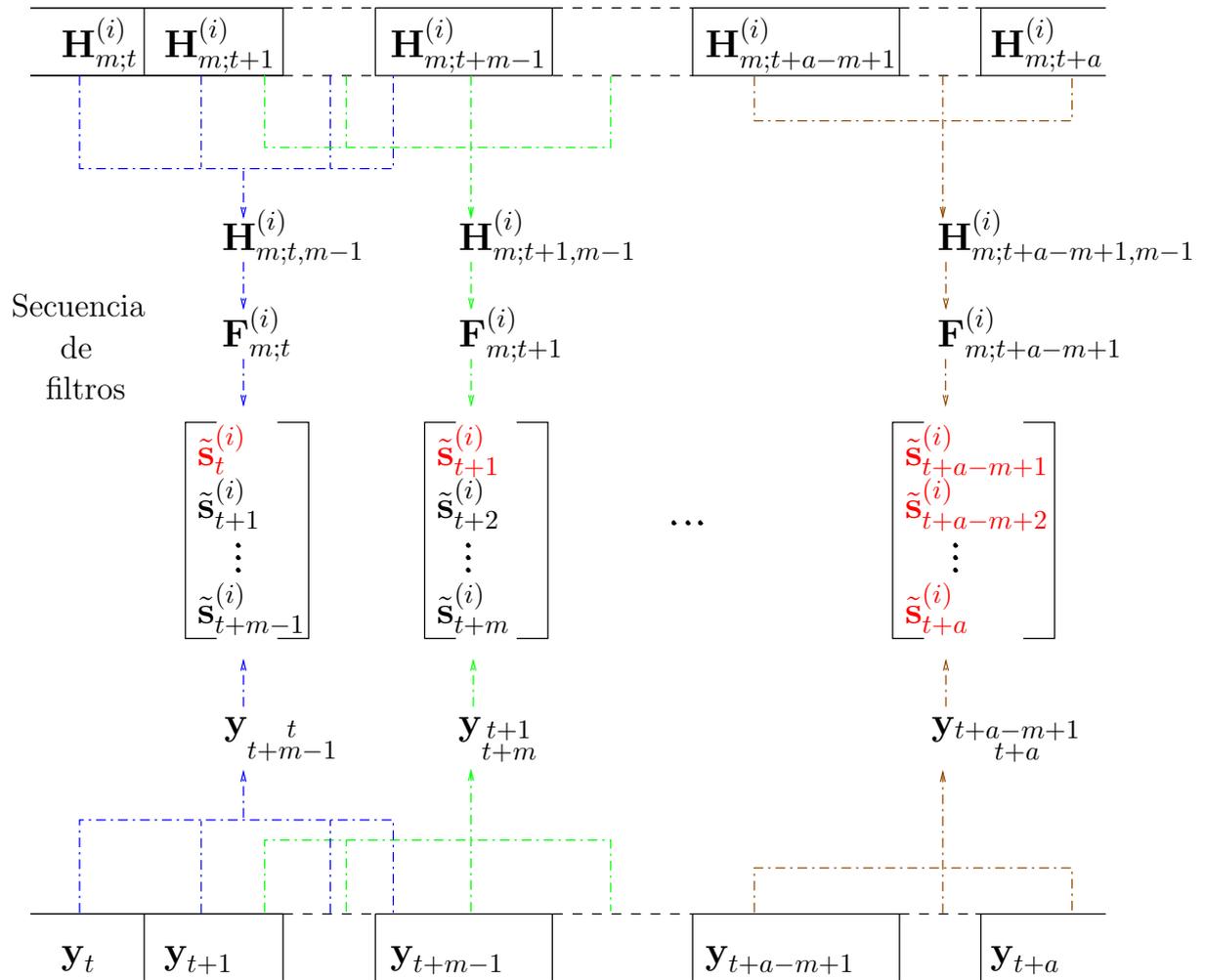


Figura 6.2: Cálculo de las estimaciones blandas considerando un orden de canal $m < \max\{\mathcal{M}\}$.

Puesto que cada símbolo se va a muestrear independientemente de todos los demás, la función tentativa que resulta para los datos es

$$q(\mathbf{s}_{t:t+a} | \mathbf{s}_{0:t-1}, \{\mathbf{H}_{m;0:t+a}\}_{m \in \mathcal{M}}, \mathbf{y}_{0:t+a}) = \prod_{k=0}^a \prod_{q=1}^{N_t} q(s_{t+k}(q)). \quad (6.51)$$

El método de filtrado de partículas que finalmente resulta viene dado por las ecuaciones (6.28) y (6.29) para muestrear el canal y los datos, respectivamente, y por la ecuación de actualización de los pesos que se obtiene al sustituir (6.36) y (6.51) en (6.30):

$$\begin{aligned} w_{t+a}^{(i)} &= w_{t+a-1}^{(i)} \frac{\sum_{m' \in \mathcal{M}} p(\mathbf{y}_{t:t+a} | \mathbf{s}_{0:t+a}^{(i)}, \mathbf{H}_{m';0:t+a}^{(i)}, \mathbf{y}_{0:t-1})}{\prod_{k'=0}^a \prod_{q=1}^{N_t} q(s_{t+k'}(q))} \times \\ &\times \frac{p(m' | \mathbf{s}_{0:t-1}^{(i)}, \mathbf{H}_{m';0:t-1}^{(i)}, \mathbf{y}_{0:t-1}) \prod_{m'' \in \mathcal{M}} \prod_{k=0}^a p(\mathbf{H}_{m'';t+k}^{(i)} | \mathbf{H}_{m'';t+k-1}^{(i)})}{\prod_{m \in \mathcal{M}} \mathcal{N}(\mathbf{h}_{m;t}^{(i)} | \alpha \tilde{\mathbf{h}}_{m;t-1}^{(i)}, \sigma_H^2 \mathbf{I}) \prod_{k=1}^a p(\mathbf{h}_{m;t+k}^{(i)} | \mathbf{h}_{m;t+k-1}^{(i)})}. \end{aligned} \quad (6.52)$$

Dado que $\mathbf{H}_{m;t}$ y $\mathbf{h}_{m;t}$ son representaciones equivalentes⁸ de una y la misma cosa (el canal en el instante t cuando su orden es m), tenemos que

$$\begin{aligned} \prod_{k=0}^a p(\mathbf{H}_{m'';t+k}^{(i)} | \mathbf{H}_{m'';t+k-1}^{(i)}) &= \prod_{k=0}^a p(\mathbf{h}_{m'';t+k}^{(i)} | \mathbf{h}_{m'';t+k-1}^{(i)}) \\ &= p(\mathbf{h}_{m'';t}^{(i)} | \mathbf{h}_{m'';t-1}^{(i)}) \prod_{k=1}^a p(\mathbf{h}_{m'';t+k}^{(i)} | \mathbf{h}_{m'';t+k-1}^{(i)}) \\ &= p(\mathbf{H}_{m'';t}^{(i)} | \mathbf{H}_{m'';t-1}^{(i)}) \prod_{k=1}^a p(\mathbf{h}_{m'';t+k}^{(i)} | \mathbf{h}_{m'';t+k-1}^{(i)}), \end{aligned} \quad (6.53)$$

lo que permite reescribir (6.52) como

$$\begin{aligned} w_{t+a}^{(i)} &= w_{t+a-1}^{(i)} \frac{\sum_{m' \in \mathcal{M}} p(\mathbf{y}_{t:t+a} | \mathbf{s}_{0:t+a}^{(i)}, \mathbf{H}_{m';0:t+a}^{(i)}, \mathbf{y}_{0:t-1})}{\prod_{k'=0}^a \prod_{q=1}^{N_t} q(s_{t+k'}(q))} \times \\ &\times \frac{p(m' | \mathbf{s}_{0:t-1}^{(i)}, \mathbf{H}_{m';0:t-1}^{(i)}, \mathbf{y}_{0:t-1}) \prod_{m'' \in \mathcal{M}} p(\mathbf{H}_{m'';t}^{(i)} | \mathbf{H}_{m'';t-1}^{(i)})}{\prod_{m \in \mathcal{M}} \mathcal{N}(\mathbf{h}_{m;t}^{(i)} | \alpha \tilde{\mathbf{h}}_{m;t-1}^{(i)}, \sigma_H^2 \mathbf{I})}. \end{aligned} \quad (6.54)$$

La ecuación (6.54) se puede simplificar si hacemos el parámetro de diseño σ_H^2 igual a la varianza del proceso AR, σ_v^2 . En tal caso, la ecuación de actualización de los pesos es

$$w_{t+a}^{(i)} = w_{t+a-1}^{(i)} \frac{\sum_{m' \in \mathcal{M}} p(\mathbf{y}_{t:t+a} | \mathbf{s}_{0:t+a}^{(i)}, \mathbf{H}_{m';0:t+a}^{(i)}, \mathbf{y}_{0:t-1}) p(m' | \mathbf{s}_{0:t-1}^{(i)}, \mathbf{H}_{m';0:t-1}^{(i)}, \mathbf{y}_{0:t-1})}{\prod_{k'=0}^a \prod_{q=1}^{N_t} q(s_{t+k'}(q))}. \quad (6.55)$$

⁸Véase la ecuación (5.4), que muestra explícitamente la relación entre ambas.

El Pseudocódigo 10 resume el algoritmo expuesto. La complejidad del mismo es lineal en el número de órdenes de canal posibles (la cardinalidad del conjunto \mathcal{M}) a la vista de la ecuación (6.55), y cúbica en el de antenas transmisoras y receptoras a la vista de las ecuaciones (6.37)-(6.40), que definen el algoritmo RLS para la estimación del canal, y de las ecuaciones (6.41)-(6.44), utilizadas en el receptor MMSE para calcular las estimaciones blandas de los símbolos transmitidos que sirven obtener muestras de los mismos.

6.4.2. Igualación mediante procesado por supervivientes

En la Sección 4.1.4 se explicó como aplicar el procesado por supervivientes para igualar canales MIMO de orden conocido. En esta sección se presenta un nuevo algoritmo basado en PSP capaz de manejar la incertidumbre asociada al orden del canal. En última instancia se trata un método determinista que lleva a cabo una búsqueda en un grafo tipo *trellis*, cuyo rendimiento (y complejidad, por tanto) se puede ajustar mediante el número de supervivientes.

El algoritmo PSP se utiliza por lo general para resolver problemas de detección de máxima verosimilitud cuando la presencia de ciertas variables desconocidas impide utilizar el algoritmo de Viterbi. En este caso queremos encontrar la secuencia de vectores de símbolos, $\mathbf{s}_{0:t}^{MLSD}$, que maximiza la probabilidad de las observaciones, $\mathbf{y}_{0:t}$, sin conocer el canal ni el orden del mismo. El problema de optimización planteado se puede formular como

$$\mathbf{s}_{0:t}^{MLSD} = \arg \max_{\mathbf{s}_{0:t}} p(\mathbf{y}_{0:t} | \mathbf{s}_{0:t}). \quad (6.56)$$

Una secuencia de vectores de símbolos (o trayectoria) candidata a ser la solución a dicho problema es lo que se denomina un superviviente. Vamos a denotar por $\mathbf{s}_{0:t}^{(i)}$ al i -ésimo superviviente en el instante t . Entonces, si P es el número total⁹ de supervivientes considerados por el algoritmo, en dicho instante de tiempo dispondremos de un conjunto de P trayectorias, $\Theta_t = \{\mathbf{s}_{0:t}^{(i)}\}_{i=1}^P$, cada una de ellas con una verosimilitud asociada, $p(\mathbf{y}_{0:t} | \mathbf{s}_{0:t}^{(i)})$. Haciendo uso de ellas podemos aproximar (6.56) como

$$\tilde{\mathbf{s}}_{0:t}^{MLSD} = \arg \max_{\mathbf{s}_{0:t} \in \Theta_t} p(\mathbf{y}_{0:t} | \mathbf{s}_{0:t}), \quad (6.57)$$

Explicaremos el funcionamiento del algoritmo describiendo los pasos necesarios para obtener un conjunto de supervivientes en el instante t , Θ_t , partiendo del conjunto de supervivientes en el instante anterior, Θ_{t-1} . Supongamos que para cada uno de los supervivientes en $t-1$, $\mathbf{s}_{0:t-1}^{(i)}$ con $i = 1, \dots, P$, conocemos su verosimilitud, $p(\mathbf{y}_{0:t-1} | \mathbf{s}_{0:t-1}^{(i)})$, y las probabilidades *a posteriori* para el orden del canal que tiene asociadas, $p(m | \mathbf{y}_{0:t-1}, \mathbf{s}_{0:t-1}^{(i)})$, $m \in \mathcal{M}$. El algoritmo propuesto consiste en la iteración de los cuatro pasos siguientes:

1. Extender cada trayectoria, $\mathbf{s}_{0:t-1}^{(i)}$, para generar $|\mathcal{S}^{N_t}|$ potenciales supervivientes, $\{\mathbf{s}_{0:t-1}^{(i)}, \mathbf{s}_t^{(j)}\}_{j=1}^{|\mathcal{S}^{N_t}|}$, en el instante t , uno por cada posible valor de $\mathbf{s}_t \in \mathcal{S}_t^N$.

⁹En el PSP clásico se considera un cierto número de supervivientes por estado del *trellis*, de tal manera que el número total de supervivientes está determinado por el número de estados del *trellis*.

Pseudocódigo 10 Algoritmo SIS de complejidad reducida para orden desconocido

para cada partícula $i = 1, \dots, M$

$$p(m|\mathbf{s}_{0:0}^{(i)}, \mathbf{H}_{m;0:0}^{(i)}, \mathbf{y}_{0:0}) = p(m) = \frac{1}{|\mathcal{M}|}$$

para cada instante de tiempo $t = 1, 2, \dots$

para cada partícula $i = 1, \dots, M$

para todo $m \in \mathcal{M}$

muestrear $\mathbf{H}_{m;t:t+a}^{(i)} \sim q(\mathbf{H}_{m;t:t+a}|\mathbf{H}_{m;t-1}^{(i)})$

para $j = 0$ hasta $j = \text{máx}\{\mathcal{M}\} - m$

construir $\mathbf{H}_{m;t+j,m-1}^{(i)}$ utilizando (6.45)

obtener un filtro matricial $\mathbf{F}_{m;t+j}^{(i)\top}$ a partir de $\mathbf{H}_{m;t+j,m-1}^{(i)}$ según la ecuación (6.48)

calcular el vector de estimaciones blandas $\tilde{\mathbf{s}}_{m;t+j,m-1}^{(i)} = \mathbf{E}_m \mathbf{F}_{m;t+j}^{(i)\top} \mathbf{y}_{t+j+m-1}$

para cada estimación blanda, $\tilde{\mathbf{s}}_{m;t+j,m-1}^{(i)}(k)$ con $k = 1, \dots, N_t m$

sea $S_p \leftarrow 0$

para todo $s \in \mathcal{S}$

$$Prob(s, m, k + jN_t) \leftarrow \mathcal{N}\left(\tilde{\mathbf{s}}_{m;t+j,m-1}^{(i)}(k) | s, \sigma_q^2\right)$$

acumular $S_p \leftarrow S_p + Prob(s, m, k + jN_t)$

para todo $s \in \mathcal{S}$

$$Prob(s, m, k + jN_t) \leftarrow Prob(s, m, k + jN_t) / S_p \text{ \{normalización\}}$$

para $k = 1, \dots, N_t(a+1)$

$$s_{t+a}^{(i)}(k) \sim q(s_{t+a} | s_t^{(i)}(k)) = \sum_{m \in \mathcal{M}} p(m|\mathbf{s}_{0:t-1}^{(i)}, \mathbf{H}_{m;0:t-1}^{(i)}, \mathbf{y}_{0:t-1}) Prob(s_{t+a} | s_t^{(i)}(k), m, k)$$

\{muestreo\}

acumular $q_{t:t+a} \leftarrow q_{t:t+a} \times q(s_{t+a}^{(i)}(k))$

actualizar $\mathbf{s}_{t+a}^{(i)}$ con $s_{t+a}^{(i)}(k)$

extraer $\mathbf{s}_t^{(i)}$ del vector $\mathbf{s}_{t+a}^{(i)}$

obtener el peso sin normalizar $\tilde{w}_{t+a}^{(i)}$ por medio de las ecuación (6.55)

para todo $m \in \mathcal{M}$

actualizar el i -ésimo estimador de canal para orden m utilizando la muestra $\mathbf{s}_t^{(i)}$

calcular $p(m|\mathbf{s}_{0:t}^{(i)}, \mathbf{H}_{m;0:t}^{(i)}, \mathbf{y}_{0:t})$ utilizando (6.23)

calcular $W \leftarrow \sum_{i=1}^M \tilde{w}_t^{(i)}$

para cada partícula $i = 1, \dots, M$

normalizar su peso $w_t^{(i)} \leftarrow \tilde{w}_t^{(i)} / W$

estimar el tamaño efectivo de la muestra, $\hat{M}_{ef} \leftarrow \left(\sum_{i=1}^M w_t^{(i)2}\right)^{-1}$

si $\hat{M}_{ef} < \gamma_r M$ **entonces**

remuestrear

2. Para cada potencial superviviente en el conjunto $\Theta_t^* = \left\{ \mathbf{s}_{0:t-1}^{(i)}, \mathbf{s}_t^{(j)} \right\}_{i=1, j=1}^{P, |\mathcal{S}^{N_t}|}$, calcular su verosimilitud asociada, $p(\mathbf{y}_{0:t} | \mathbf{s}_{0:t-1}^{(i)}, \mathbf{s}_t^{(j)})$.
3. Construir el nuevo conjunto de supervivientes $\Theta_t = \{ \mathbf{s}_{0:t}^{(i)} \}_{i=1}^P$ eligiendo en Θ_t^* las trayectorias asociados a las P verosimilitudes mayores.
4. Para cada superviviente en Θ_t , actualizar sus correspondientes probabilidades *a posteriori* para el orden del canal, $p(m | \mathbf{y}_{0:t}, \mathbf{s}_{0:t}^{(i)})$, $m \in \mathcal{M}$.

La Figura 6.3 ilustra de qué manera el algoritmo obtiene un conjunto de supervivientes en el instante t a partir del conjunto de supervivientes en el instante $t - 1$ utilizando los pasos enumerados.

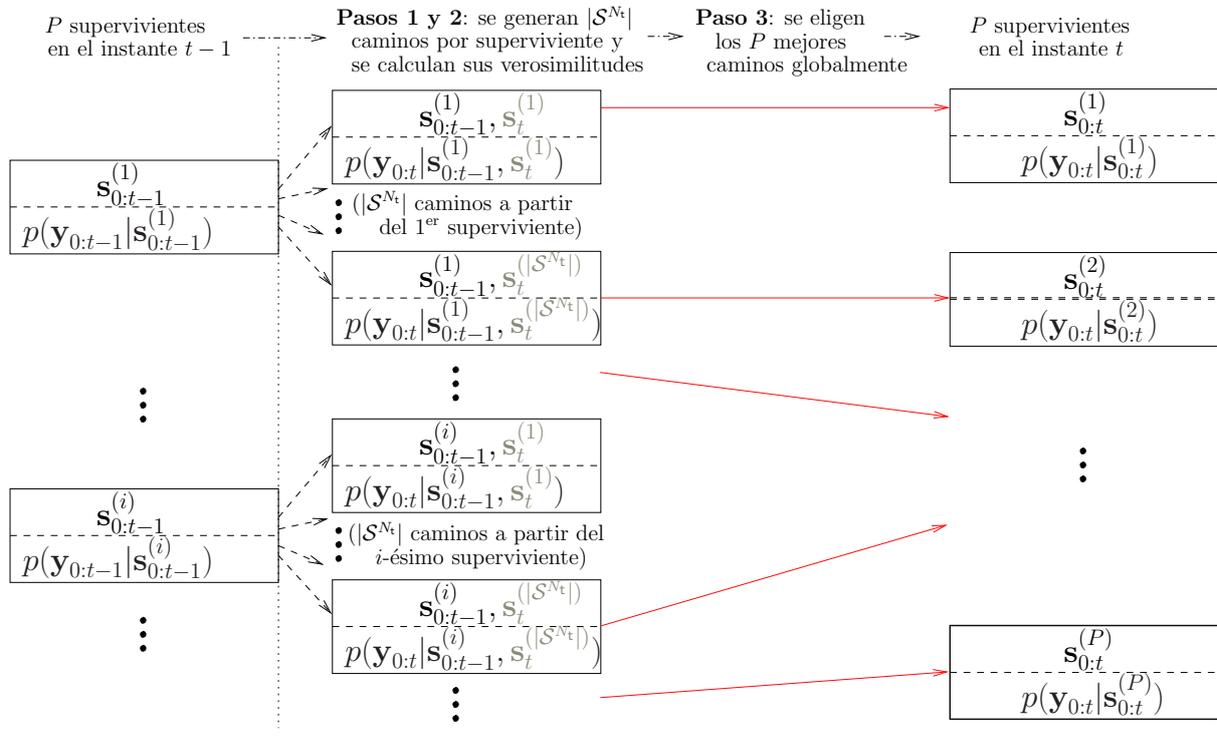


Figura 6.3: Esquema del algoritmo para la igualación de canales MIMO de orden desconocido basado en procesado por supervivientes.

Los pasos 1 y 3 son muy sencillos y no precisan de mayor explicación. En relación con los pasos 2 y 4, a continuación vamos a describir en detalle como calcular las verosimilitudes, $p(\mathbf{y}_{0:t} | \mathbf{s}_{0:t-1}^{(i)}, \mathbf{s}_t^{(j)})$, y las probabilidades *a posteriori* para el orden del canal, $p(m | \mathbf{y}_{0:t}, \mathbf{s}_{0:t}^{(i)})$, a partir de $p(\mathbf{y}_{0:t-1} | \mathbf{s}_{0:t-1}^{(i)})$ y $p(m | \mathbf{y}_{0:t-1}, \mathbf{s}_{0:t-1}^{(i)})$, respectivamente.

La verosimilitud en el instante t de la j -ésima extensión del i -ésimo superviviente se puede descomponer como

$$p(\mathbf{y}_{0:t} | \mathbf{s}_{0:t-1}^{(i)}, \mathbf{s}_t^{(j)}) = p(\mathbf{y}_t | \mathbf{y}_{0:t-1}, \mathbf{s}_{0:t-1}^{(i)}, \mathbf{s}_t^{(j)}) p(\mathbf{y}_{0:t-1} | \mathbf{s}_{0:t-1}^{(i)}) \quad (6.58)$$

simplemente utilizando la definición de probabilidad condicional. El término $p(\mathbf{y}_{0:t-1} | \mathbf{s}_{0:t-1}^{(i)})$ se corresponde con la verosimilitud de uno de los supervivientes en el instante $t - 1$, por lo que es conocido. En cuanto al otro factor que aparece en el lado derecho de (6.58), aplicando el teorema de la probabilidad total resulta

$$\begin{aligned} p(\mathbf{y}_t | \mathbf{y}_{0:t-1}, \mathbf{s}_{0:t-1}^{(i)}, \mathbf{s}_t^{(j)}) &= \sum_{m \in \mathcal{M}} p(\mathbf{y}_t, m | \mathbf{y}_{0:t-1}, \mathbf{s}_{0:t-1}^{(i)}, \mathbf{s}_t^{(j)}) \\ &= \sum_{m \in \mathcal{M}} p(\mathbf{y}_t | m, \mathbf{y}_{0:t-1}, \mathbf{s}_{0:t-1}^{(i)}, \mathbf{s}_t^{(j)}) p(m | \mathbf{y}_{0:t-1}, \mathbf{s}_{0:t-1}^{(i)}). \end{aligned} \quad (6.59)$$

Sustituyendo entonces (6.59) en (6.58) tenemos que la expresión final para calcular la verosimilitud de la secuencia $\{\mathbf{s}_{0:t-1}^{(i)}, \mathbf{s}_t^{(j)}\}$ es

$$p(\mathbf{y}_{0:t} | \mathbf{s}_{0:t-1}^{(i)}, \mathbf{s}_t^{(j)}) = p(\mathbf{y}_{0:t-1} | \mathbf{s}_{0:t-1}^{(i)}) \sum_{m \in \mathcal{M}} p(\mathbf{y}_t | m, \mathbf{y}_{0:t-1}, \mathbf{s}_{0:t-1}^{(i)}, \mathbf{s}_t^{(j)}) p(m | \mathbf{y}_{0:t-1}, \mathbf{s}_{0:t-1}^{(i)}). \quad (6.60)$$

Los términos $p(m | \mathbf{y}_{0:t-1}, \mathbf{s}_{0:t-1}^{(i)})$, $m \in \mathcal{M}$ son las probabilidades para los diferentes órdenes de canal condicionadas en la secuencia de símbolos asociada al i -ésimo superviviente en $t - 1$ y han sido calculadas en dicho instante de tiempo. Así pues, solo falta calcular la probabilidad $p(\mathbf{y}_t | m, \mathbf{y}_{0:t-1}, \mathbf{s}_{0:t-1}^{(i)}, \mathbf{s}_t^{(j)})$, que es gaussiana y se puede obtener, tal como se explica en el Apéndice D, mediante el filtro de Kalman¹⁰.

Para obtener $p(\mathbf{y}_t | \mathbf{y}_{0:t-1}, \mathbf{s}_{0:t-1}^{(i)}, \mathbf{s}_t^{(j)})$ mediante (6.59) vamos a necesitar $|\mathcal{M}|$ filtros de Kalman, cada uno de ellos considerando un orden de canal diferente, para calcular las probabilidades $p(\mathbf{y}_t | m, \mathbf{y}_{0:t-1}, \mathbf{s}_{0:t-1}^{(i)}, \mathbf{s}_t^{(j)})$, $m \in \mathcal{M}$. Sin embargo, en la expresión dada en el Apéndice D para calcular $p(\mathbf{y}_t | m, \mathbf{y}_{0:t-1}, \mathbf{s}_{0:t-1}^{(i)}, \mathbf{s}_t^{(j)})$ únicamente intervienen la media y covarianza predictivas de Kalman, y éstas no dependen del vector $\mathbf{s}_t^{(j)}$, por lo que el número total de pasos de filtrado requeridos por el algoritmo en cada instante de tiempo es $P|\mathcal{M}|$, i.e., no depende del tamaño de la constelación de posibles vectores de símbolos, \mathcal{S}_t^N .

Una vez disponemos de las verosimilitudes $p(\mathbf{y}_t | m, \mathbf{y}_{0:t-1}, \mathbf{s}_{0:t-1}^{(i)}, \mathbf{s}_t^{(j)})$ para todo $m \in \mathcal{M}$, la actualización recursiva de las probabilidades *a posteriori* del orden del canal requerida por el paso 4 es mucho más simple. Después de seleccionar los nuevos supervivientes en el paso 3, para cada camino en Θ_t y para cada orden de canal $m \in \mathcal{M}$ tenemos

$$p(m | \mathbf{y}_{0:t}, \mathbf{s}_{0:t}^{(i)}) = p(\mathbf{y}_t | m, \mathbf{y}_{0:t-1}, \mathbf{s}_{0:t}^{(i)}) p(m | \mathbf{y}_{0:t-1}, \mathbf{s}_{0:t-1}^{(i)}). \quad (6.61)$$

¹⁰En el Apéndice D se aborda el cálculo de $p(\mathbf{y}_t | \mathbf{s}_{0:t}, \mathbf{y}_{0:t-1})$ y no el de $p(\mathbf{y}_t | m, \mathbf{s}_{0:t}, \mathbf{y}_{0:t-1})$. Sin embargo, puesto que se hace asumiendo que el orden del canal, m , es un parámetro conocido del sistema, tenemos que $p(\mathbf{y}_t | \mathbf{s}_{0:t}, \mathbf{y}_{0:t-1}) = p(\mathbf{y}_t | m, \mathbf{s}_{0:t}, \mathbf{y}_{0:t-1})$.

El Pseudocódigo 11 muestra un resumen del algoritmo. Obsérvese que en la fase de inicialización del algoritmo se crea un único superviviente cuya trayectoria está vacía. El conjunto de supervivientes se irá llenando a medida que nuevos caminos o trayectorias se expandan a partir de él, pero pueden ser necesarios varios instantes de tiempo antes de disponer de un conjunto de P supervivientes distintos.

Pseudocódigo 11 Algoritmo para la igualación de canales MIMO de orden desconocido basado en procesado por supervivientes

construir un superviviente inicial con

- una trayectoria vacía
- verosimilitud asociada igual a 1
- $p(m) = \frac{1}{|\mathcal{M}|}$ para cada posible orden de canal $m \in \mathcal{M}$

para cada instante de tiempo $t = 1, 2, \dots$

para cada superviviente $i = 1, \dots, P$

para todo $\mathbf{s}_t^{(j)} \in \mathcal{S}^{N_t}$, $j = 1, \dots, |\mathcal{S}^{N_t}|$

para todo $m \in \mathcal{M}$

calcular $p(\mathbf{y}_t | m, \mathbf{y}_{0:t-1}, \mathbf{s}_{0:t-1}^{(i)}, \mathbf{s}_t^{(j)})$ por medio del filtro de Kalman

calcular $p(\mathbf{y}_{0:t} | \mathbf{s}_{0:t-1}^{(i)}, \mathbf{s}_t^{(j)})$ utilizando (6.60)

añadir $\{\mathbf{s}_{0:t-1}^{(i)}, \mathbf{s}_t^{(j)}\}$ al conjunto Θ_t^* de potenciales supervivientes

construir Θ_t eligiendo los P mejores elementos en Θ_t^*

para cada superviviente $i = 1, \dots, P$

para todo $m \in \mathcal{M}$

actualizar el i -ésimo KF para orden de canal m utilizando la muestra $\mathbf{s}_t^{(i)}$

calcular $p(m | \mathbf{y}_{0:t}, \mathbf{s}_{0:t}^{(i)})$ utilizando (6.61)

La aproximación propuesta difiere del PSP clásico, en el cual todos los posibles estados del diagrama *trellis*, cada uno de ellos con al menos un superviviente, se consideran en todo momento. En nuestro caso no estamos imponiendo tal restricción y el método propuesto obvia en cierto modo el concepto de estado para operar únicamente con supervivientes. Dicho de otra manera, estamos eligiendo los P mejores supervivientes de Θ_t^* independientemente del estado que tienen asociado en el diagrama de *trellis*. Puesto que manejamos $|\mathcal{M}|$ órdenes de canal, y cada uno de ellos requiere un filtro de Kalman cuya actualización tiene complejidad $\mathcal{O}(N_r^3)$, la complejidad total del algoritmo es $\mathcal{O}(N_r^3 P |\mathcal{M}| |\mathcal{S}^{N_t}|)$.

Para cada camino superviviente calculamos, por un lado, su verosimilitud, y por otro, las probabilidades *a posteriori* de los diferentes órdenes de canal cuando es la secuencia efectivamente transmitida. Éstas últimas se pueden utilizar para obtener estimaciones MAP del orden del canal,

$$\hat{m}_t^{(i)} = \arg \max_{m \in \mathcal{M}} p(m | \mathbf{y}_{0:t}, \mathbf{s}_{0:t}^{(i)}), \quad (6.62)$$

para cualquier superviviente y en cualquier instante de tiempo. Como se verá en las simulaciones, la probabilidad *a posteriori* del auténtico orden de canal crece rápidamente muy por encima de las probabilidades *a posteriori* de los restantes órdenes en el conjunto \mathcal{M} . Así pues, en una implementación práctica del algoritmo podríamos fijar el valor del orden del canal en todos los supervivientes después de un determinado número de iteraciones (es decir, $m_t^{(i)} = \hat{m}_{t_0}^{(i)}$ para todo $t > t_0$) o cuando la máxima probabilidad *a posteriori* sobrepasa un determinado umbral predefinido, $0 < \gamma < 1$ (es decir, si $p(\hat{m}_{t'}^{(i)} | \mathbf{y}_{0:t'}, \mathbf{s}_{0:t'}^{(i)}) > \gamma$ en t' , entonces $m_t^{(i)} = \hat{m}_{t'}^{(i)}$ para todo $t > t'$). Esto reduce la complejidad del algoritmo a $\mathcal{O}(N_r^3 P |\mathcal{S}^{N_t}|)$, pues en tal caso únicamente se necesita un filtro de Kalman por superviviente.

Otra posible manera de reducir la complejidad del algoritmo consiste en sustituir el banco de filtros de Kalman por un banco de estimadores de canal lineales, como pueden ser el RLS o el LMS. La probabilidad $p(\mathbf{y}_t | m, \mathbf{y}_{0:t-1}, \mathbf{s}_{0:t-1}^{(i)}, \mathbf{s}_t^{(j)})$ necesaria en (6.60) se aproximaría entonces como

$$\begin{aligned} p(\mathbf{y}_t | m, \mathbf{y}_{0:t-1}, \mathbf{s}_{0:t-1}^{(i)}, \mathbf{s}_t^{(j)}) &\approx p(\mathbf{y}_t | m, \tilde{\mathbf{H}}_{m;t}, \mathbf{s}_{0:t-1}^{(i)}, \mathbf{s}_t^{(j)}) \\ &= \mathcal{N} \left(\mathbf{y}_t | \tilde{\mathbf{H}}_{m;t} \mathbf{s}_{t-m+1}^{(i)}, \sigma_g^2 \mathbf{I}_{N_r} \right) \end{aligned} \quad (6.63)$$

donde $\tilde{\mathbf{H}}_{m;t}$ es una estimación de la matriz de canal en el instante t suponiendo orden de canal m . Si las estimaciones $\tilde{\mathbf{H}}_{m;t}$ se calculan utilizando las secuencias de símbolos y de observaciones hasta el instante $t-1$, i.e., $\mathbf{s}_{0:t-1}^{(i)}$ e $\mathbf{y}_{0:t-1}$, respectivamente, el número de pasos de estimación de canal necesarios en un determinado instante de tiempo es $P|\mathcal{M}|$.

6.5. Conclusiones

En este capítulo se ha considerado el problema de la igualación de canales MIMO cuyo orden de canal es desconocido. La aproximación clásica al problema consiste en asignarle a éste un valor supuestamente mayor del que realmente tiene (sobrestimarlo) para de esa manera evitar la enorme pérdida de rendimiento que ocurre en un receptor que considera un orden de canal menor que el verdadero, es decir, que lo subestima. Sin embargo, dicha aproximación tiene dos problemas fundamentales: un incremento en la complejidad computacional y una degradación progresiva del rendimiento del receptor a medida que el orden considerado se aleja (por encima) del verdadero.

Aunque el problema de la estimación del orden de un canal de comunicaciones es bien conocido, son pocos los trabajos que lo han abordado y menos aún los que lo han hecho en el contexto de un sistema MIMO. En la primera parte de este capítulo, tras introducir el modelo de señal, se han revisado brevemente algunos de los esfuerzos más notables encaminados a la estimación del orden del canal que se pueden encontrar en la literatura. En cada caso, se han señalado las principales ventajas e inconvenientes del método así como las dificultades que surgen a la hora de aplicarlo a nuestro modelo de señal.

En la última sección del capítulo se presentaron dos nuevas aproximaciones para la igualación de canales MIMO de orden desconocido. La primera de ellas extiende uno de los métodos de filtrado de partículas vistos en el Capítulo 5 para manejar la incertidumbre asociada al desconocimiento del orden del canal. La complejidad del algoritmo resultante es cúbica en el número de antenas transmisoras y lineal en el número de órdenes de canal posibles considerados, por lo que constituye una alternativa de bajo coste computacional al igualador SMC que utiliza la función tentativa óptima cuando el orden de canal es desconocido (también derivado en este capítulo).

El segundo esquema propuesto es determinista y está basado en el concepto de procesado por supervivientes. La complejidad del mismo es exponencial en el número de entradas y lineal en el de potenciales valores que puede tomar el orden del canal, pero a expensas de ella (tal como se verá en el capítulo de simulaciones) exhibe un rendimiento muy cercano al del algoritmo de Viterbi con canal conocido.

Capítulo 7

Simulaciones

Para probar el rendimiento de los algoritmos propuestos hemos simulado un sistema con $N_t = 2$ antenas transmisoras y $N_r = 3$ antenas receptoras. La transmisión se lleva a cabo en bloques de 300 vectores de símbolos, 10 de los cuales (al principio de la trama) son conocidos y constituyen una secuencia de entrenamiento que se utiliza para obtener una estimación inicial del canal. El orden de este último es $m = 3$, de tal manera que el número de coeficientes de la matriz de canal es $N_t \times N_r \times m = 18$. La modulación considerada es BPSK, por lo que el alfabeto es binario, $\mathcal{S} = \{-1, +1\}$.

Por simplicidad, todos los algoritmos se han derivado suponiendo que el canal es un proceso autorregresivo de primer orden que viene dado por la ecuación (3.6). Sin embargo, para probar la robustez de los algoritmos propuestos cuando no se cumple dicha hipótesis, también se han llevado a cabo simulaciones considerando canales generados mediante el modelo de Clarke [92]. En este caso, el canal es un proceso estocástico gaussiano (multidimensional) estacionario en sentido amplio y de dispersión incorrelada (WSSUS, del inglés *Wide-Sense Stationary Uncorrelated Scattering*). Cada uno de los coeficientes, $h_t(i, j)$, de la matriz de canal \mathbf{H}_t se modela entonces como una variable aleatoria gaussiana de media 0 y varianza σ_l^2 . La función de autocovarianza del proceso discreto $\{h_t(i, j)\}_{t \in \mathbb{N}}$ es

$$\phi(k) = \mathbb{E}[h_t(i, j)h_{t+k}(i, j)] = \sigma_l^2 \mathcal{J}_0(2\pi F_d k T) \quad (7.1)$$

donde $\mathcal{J}_0(\cdot)$ es la función Bessel de primera especie y orden cero, T es el período de símbolo y $F_d = (v_m/c)F_c$ es el espectro Doppler máximo, siendo $v_m = 50$ Km/h la velocidad del móvil (transmisor), c la velocidad de la luz y $F_c = 2$ GHz la frecuencia de portadora. Siempre vamos a considerar un perfil de potencia plano por lo que la varianza, σ_l^2 , es la misma para todos los coeficientes de la matriz de canal, independiente del retardo al que estén asociados.

Todos los algoritmos se van a evaluar en términos de la tasa de error de bit (BER, del inglés *Bit Error Rate*) y del error que cometen en la estimación del canal. En relación a esto último, la métrica utilizada es el error cuadrático medio (MSE, *Mean Square Error*) normalizado, que se define como

$$MSE_t = \frac{(\mathbf{h}_t - \tilde{\mathbf{h}}_t)^\top (\mathbf{h}_t - \tilde{\mathbf{h}}_t)}{\mathbf{h}_t^\top \mathbf{h}_t} \quad (7.2)$$

donde $\hat{\mathbf{h}}_t$ es la estimación del canal \mathbf{h}_t . A modo de referencia, en todas las gráficas de MSE se muestra el valor de dicha métrica obtenido por un filtro de Kalman cuando la secuencia de símbolos es conocida (la curva aparece etiquetada como “KF (símbolos conocidos)”). Este procedimiento es óptimo siempre y cuando el canal sea un proceso autorregresivo.

Para tener una idea del máximo rendimiento que puede alcanzar un algoritmo en términos de BER, se ha simulado el algoritmo de Viterbi (explicado en la Sección 4.1.3 y el Apéndice B) considerando que el canal es conocido.

Lo que resta del capítulo se ha dividido en dos secciones, cada una de ellas asociada a un capítulo diferente de la tesis. La primera está dedicada al análisis del rendimiento de los métodos SMC de complejidad reducida propuestos en el Capítulo 5, mientras que en la segunda se muestran los resultados obtenidos en la simulación de los algoritmos presentados en el Capítulo 6, que abordan el problema del orden de canal desconocido. Además, la estructura de cada sección está estrechamente relacionada con la del capítulo correspondiente.

Todas las gráficas que veremos a lo largo de este capítulo proceden de la simulación de muchas tramas de datos en las que los símbolos, el canal y el ruido son generados independientemente. En cada caso se indicará el número exacto de tramas utilizadas para calcular el promedio.

7.1. Nuevos métodos SMC de complejidad reducida para la igualación de canales MIMO

En esta sección compararemos el rendimiento de los algoritmos presentados a lo largo del Capítulo 5 con algunas de las aproximaciones anteriores descritas en el Capítulo 4. En particular, en todas las gráficas de BER se incluye el rendimiento de:

- Un receptor lineal MMSE análogo al dado por la ecuación (5.36) que utiliza las estimaciones de canal proporcionadas por un filtro de Kalman. Se trata, pues, de un esquema determinista dirigido por decisión (DD) similar al propuesto en [59]. Lo denotaremos por “KF+MMSE”.
- Un receptor MMSE idéntico al anterior pero, en este caso, utilizando un filtro de Kalman alimentado con los símbolos verdaderos. Las estimaciones de la matriz de canal son, por tanto, óptimas siempre y cuando el canal evolucione siguiendo el proceso AR que describe la ecuación (3.6). En las gráficas nos referiremos a él como “KF (símbolos conocidos)+MMSE”.
- El algoritmo SIS con función tentativa óptima sin suavizado descrito en la Sección 4.2.1 y el Pseudocódigo 4. Servirá para ilustrar el máximo rendimiento que puede alcanzar un algoritmo SMC que no hace ningún tipo de suavizado. Se ha etiquetado como “SIS opt (sin suavizado)”.

- El algoritmo SIS con función tentativa óptima (y suavizado) explicado en 4.2.2, cuya implementación viene dada por el Pseudocódigo 5. Se denota por “SIS opt.”.
- El PSP clásico explicado en la Sección 4.1.4 utilizando un número de supervivientes equivalente al número de partículas empleadas por los algoritmos SMC (así, si M es el número de partículas consideradas por los algoritmos SMC, el número de supervivientes por estado del PSP es $P = \frac{M}{|S|^{N_t(m-1)}}$) y estimando el canal por medio del filtro de Kalman.

Vamos a comparar los algoritmos mencionados, por un lado, con los igualadores SMC basados en variables auxiliares explicados en la Sección 5.1 y, por el otro, con los métodos de la Sección 5.2 que tratan de aproximar la distribución conjunta del canal y los datos.

7.1.1. Métodos SMC basados en variables auxiliares

En esta sección estudiaremos el rendimiento de los métodos SMC propuestos en 5.1:

- el basado en la triangularización de la matriz de autocorrelación del canal, que denotaremos por “MKF-Cholesky”,
- el algoritmo SMC basado en estimadores lineales utilizando el filtro MMSE, etiquetado como “MKF-MMSE“, y
- el algoritmo SMC basado en estimadores lineales utilizando el filtro de decorrelación, al que nos referiremos como “MKF-Dec”.

La Figura 7.1 muestra la tasa de error de bit de los distintos algoritmos para diferentes valores de la SNR cuando el número de partículas es $M = 192$. Los tres igualadores SMC propuestos obtienen un rendimiento muy cercano al del algoritmo de filtrado de partículas que utiliza la función tentativa óptima (SIS opt.) con una pérdida de apenas medio decibelio para tasas de error de bit entre 10^{-3} y 10^{-4} . La pérdida con respecto al algoritmo de Viterbi (suponiendo el canal conocido) y al PSP, ambos de complejidad exponencial en $N_t(m-1)$, es de aproximadamente 2 dB en 10^{-3} . Por otro lado, cualquiera de los filtros de partículas propuestos supera en rendimiento tanto al igualador SMC óptimo sin suavizado, como a los dos algoritmos que combinan el filtro de Kalman para la estimación de canal con un receptor MMSE para la detección. En relación a esto último cabe destacar que el rendimiento del MKF-Dec (el que mejor funciona de los igualadores SMC propuestos, en este caso) consigue una tasa de error de bit casi diez veces inferior a la del algoritmo etiquetado KF (símbolos conocidos)+MMSE, que detecta por medio de un filtro MMSE utilizando las estimaciones de canal proporcionadas por un KF cuya entrada es la secuencia de vectores de símbolos efectivamente transmitida.

Un fenómeno que llama la atención en la Figura 7.1 es la ausencia de mejoría en el rendimiento del SIS opt (sin suavizado) en las regiones altas de la SNR. Es debido a que al aumentar la SNR (o, equivalentemente, disminuir la varianza del ruido) las f.d.p. objetivo y

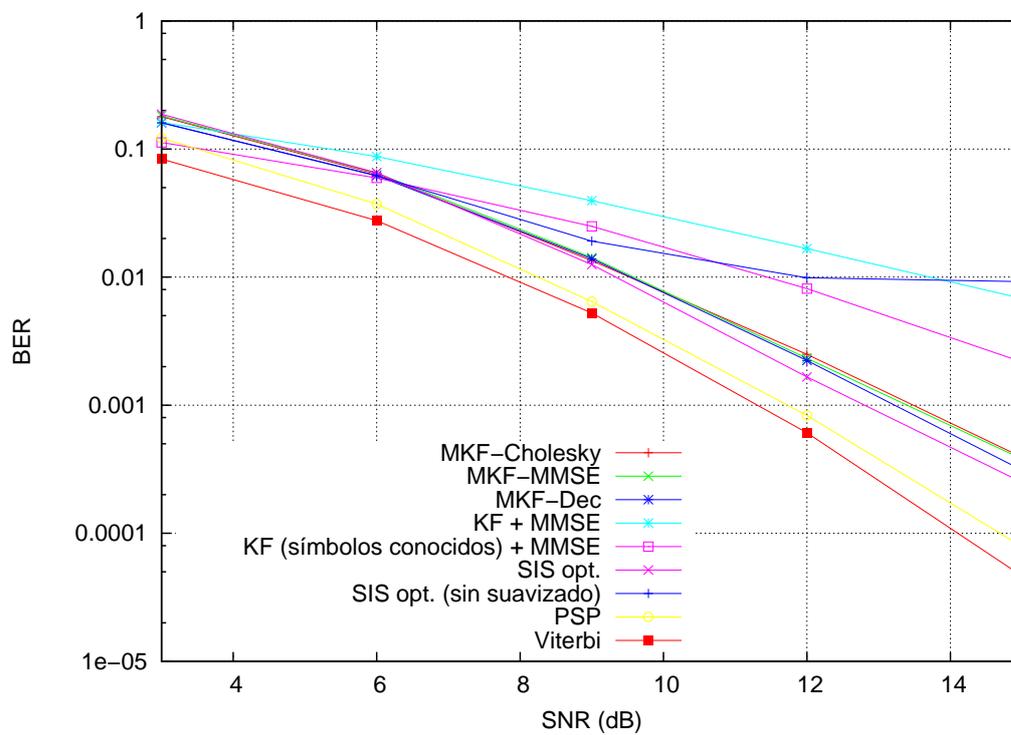


Figura 7.1: BER para distintos valores de la SNR (dB). Los algoritmos SMC utilizan $M = 192$ partículas. Los resultados están promediados sobre 2519 tramas de datos. Modelo de canal AR.

tentativa¹ se vuelven más apuntadas, de tal manera que se necesita una función tentativa muy próxima a la objetivo para que las muestras tomadas de la misma tengan una verosimilitud significativa. El SIS opt (sin suavizado), como su propio nombre indica, no lleva a cabo ningún tipo de suavizado y las diferencias entre las probabilidades que asigna a los distintos vectores de símbolos se diluyen a medida que aumenta la SNR, lo que dificulta que se muestree el vector de símbolos realmente transmitido.

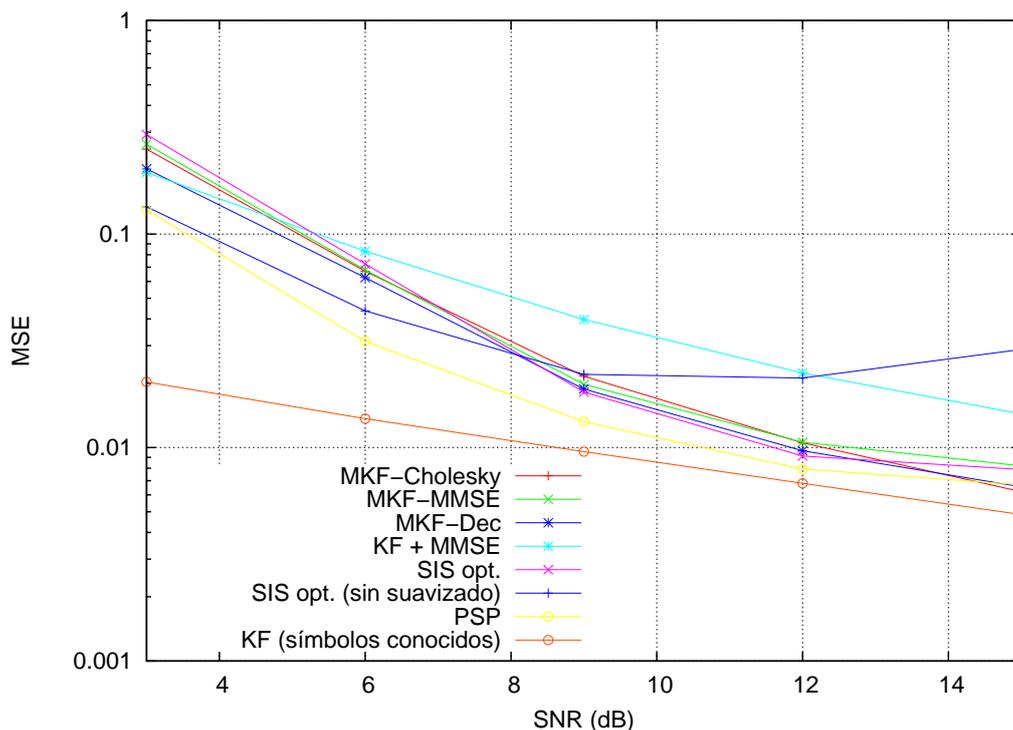


Figura 7.2: MSE para distintos valores de la SNR (dB). Los algoritmos SMC utilizan $M = 192$ partículas. Los resultados están promediados sobre 2519 tramas de datos. Modelo de canal AR.

El MSE alcanzado por los diferentes algoritmos en la estimación de la matriz de canal se muestra en la Figura 7.2. Al igual que ocurría con la BER, en términos de MSE el comportamiento de los métodos SMC propuestos también es muy similar y a medida que aumenta la SNR su rendimiento se aproxima al obtenido por un filtro de Kalman alimentado con los símbolos verdaderos (KF (símbolos conocidos)). Nótese, asimismo, que el error cometido en la estimación de canal por cualquiera de los algoritmos de filtrado de partículas, salvo el SIS opt (sin suavizado), es considerablemente menor que el cometido por un KF utilizando los símbolos detectados por un receptor MMSE (KF+MMSE).

¹Nótese que ambas son, efectivamente, funciones de densidad de probabilidad (continuas, pues) y asignan un valor a cada posible vector $\mathbf{s}_t \in \mathbb{R}^2$, aunque solo estemos interesados en aquellos que pertenecen al alfabeto, $\mathbf{s}_t \in \mathcal{S}^{N_t}$.

Obsérvese que, también en términos de MSE, el rendimiento del SIS opt (sin suavizado) es muy pobre, llegando incluso a empeorar para relaciones señal a ruido altas. La explicación hay que buscarla en el rendimiento del algoritmo en términos de BER, que a partir de SNR 12 prácticamente se mantiene constante. Cuando la varianza del ruido es alta, el KF actualiza su estimación del canal basándose más en la ecuación de estado (la estimación del canal en el instante anterior) que en las observaciones. Cuando la varianza del ruido es pequeña (la SNR es alta) se da la situación contraria y la bondad de la estimación del canal está condicionada por la exactitud de la ecuación de observación, que en este caso depende de los símbolos detectados (véase la ecuación (D.2) en el Apéndice D).

En las simulaciones consideradas hasta ahora, el canal ha sido generado siguiendo el modelo AR utilizado en la derivación de los algoritmos (véase la ecuación (3.6)). Sin embargo, para poner a prueba la robustez de los igualadores propuestos también se han llevado a cabo simulaciones considerando canales generados mediante el modelo clásico de Clarke [92], ya explicado anteriormente. La Figura 7.3 muestra la tasa de error de bit frente a la SNR de los algoritmos cuando la autocorrelación del canal viene dada por dicho modelo. Con respecto a las simulaciones hechas considerando un modelo de canal AR, los algoritmos MKF-Cholesky y MKF-Dec experimentan una ligera pérdida de rendimiento de aproximadamente 0.5 dB. Los algoritmos MKF-MMSE, Viterbi y PSP, por otro lado, son más robustos y no sufren una degradación del rendimiento perceptible. En cuanto al igualador SMC óptimo, una disminución de su BER en SNRs altas acorta la distancia que lo separa del PSP.

Estos resultados pueden justificarse a la vista de las curvas de MSE, que se muestran en la Figura 7.4. El error cometido en la estimación del canal por el filtro de Kalman alimentado con los símbolos verdaderos (que sirve como cota inferior del menor error que se puede cometer en la estimación del canal) es ligeramente inferior ahora que en el caso del canal AR. Es decir, los canales generados siguiendo el modelo de Clarke (suponiendo que el móvil se desplaza a 50 Km/h) son menos hostiles que los generados considerando el modelo AR especificado. Es por ello que el SIS opt. (y en menor medida el SIS opt (sin suavizado)) obtienen un mejor rendimiento (tanto en términos de MSE como de BER). Los igualadores SMC propuestos, por otro lado, sufren una pequeña degradación en el rendimiento debido a que el desajuste en el modelo tiene importancia a la hora de tomar muestras del canal², que siempre se hace asumiendo un modelo autorregresivo.

En cualquier caso, puesto que el rendimiento de los algoritmos es esencialmente el mismo independientemente del modelo de canal considerado, de aquí en adelante únicamente se mostrarán las gráficas correspondientes al modelo de Clarke, que consideramos más realista.

7.1.2. Aproximación de la distribución conjunta del canal y los datos

En esta sección se analizan los resultados numéricos obtenidos para el algoritmo de filtrado de partículas presentado en la Sección 5.2, que trataba de aproximar la f.d.p. conjunta del canal

²Recuérdese que ni el SIS opt. ni el SIS opt (sin suavizado) precisan tomar muestras del canal.

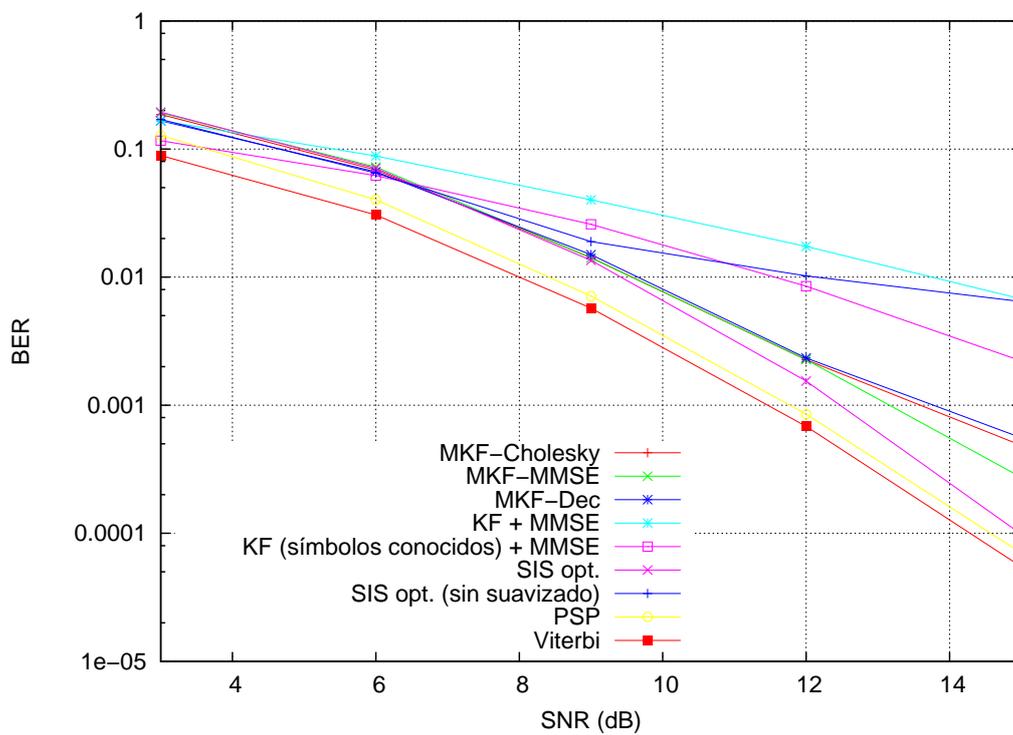


Figura 7.3: BER para distintos valores de la SNR (dB). Los algoritmos SMC utilizan $M = 192$ partículas. Los resultados están promediados sobre 2037 tramas de datos. Modelo de autocorrelación de Clarke.

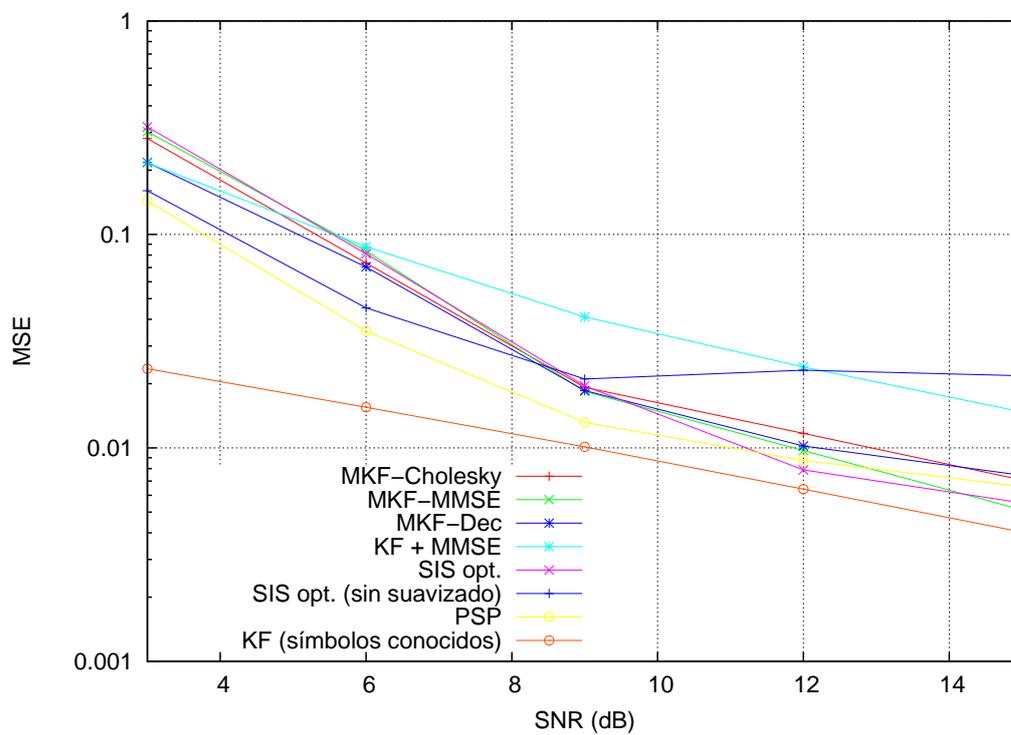


Figura 7.4: MSE para distintos valores de la SNR (dB). Los algoritmos SMC utilizan $M = 192$ partículas. Los resultados están promediados sobre 2037 tramas de datos. Modelo de autocorrelación de Clarke.

y los datos (el canal ya no es, por tanto, una variable auxiliar sino que está incluido en la f.d.p. objetivo). Se han simulado dos versiones del mismo que únicamente difieren en el estimador de canal utilizado,

- el “RLS-SIS”, que hace uso de un procedimiento RLS, y
- el “LMS-SIS”, que utiliza el algoritmo LMS

(ambos métodos de estimación fueron explicados en la Sección 5.2.1).

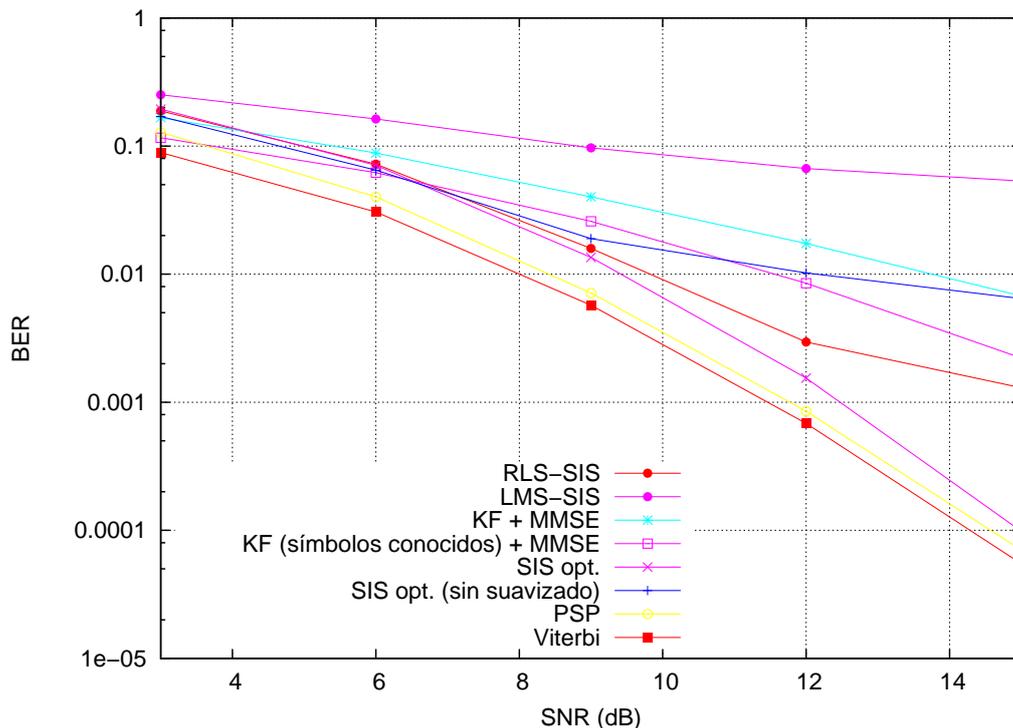


Figura 7.5: BER para distintos valores de la SNR (dB). Los algoritmos SMC utilizan $M = 192$ partículas. Los resultados están promediados sobre 2037 tramas de datos. Modelo de autocorrelación de Clarke.

La Figura 7.5 muestra la BER de los dos algoritmos basados en la aproximación de la distribución conjunta del canal y los datos para distintos valores de la SNR cuando el número de partículas es 192. Claramente la elección del estimador de canal es un factor decisivo en el rendimiento del método explicado en 5.2: para una SNR de 15 dB la tasa de error de bit del LMS-SIS se sitúa en torno a 10^{-1} mientras que el del RLS-SIS se acerca a 10^{-3} . Para esa misma BER el RLS-SIS se encuentra a 3 dB del igualador SMC óptimo y a 5 dB del algoritmo de Viterbi con canal conocido. Por tanto, el rendimiento de este método de filtrado de partículas (tanto utilizando el estimador de canal RLS como el LMS) es, para relaciones señal a ruido altas, considerablemente inferior al alcanzado por los métodos SMC basados en el uso de variables auxiliares de la Sección 5.1 (cerca de 2 dB separan sus curvas de rendimiento

si se comparan las figuras 7.3 y 7.5). Recuérdese sin embargo que, aunque la complejidad de ambos métodos es polinómica, la constante asociada al RLS-SIS es mucho más pequeña que la asociada a cualquiera de los métodos presentados en la Sección 5.1 y, por tanto, la carga computacional del algoritmo es considerablemente menor. Además, la curva de BER de éste último algoritmo está por debajo de la del algoritmo que detecta la secuencia transmitida utilizando las estimaciones de canal proporcionadas por un filtro de Kalman alimentado con los símbolos verdaderos (KF (símbolos conocidos)+MMSE), cuya complejidad es también cúbica en el número de flujos de entrada. El RLS-SIS también es claramente superior al SIS opt (sin suavizado), cuyo rendimiento se ve seriamente afectado por la dispersión temporal del canal.

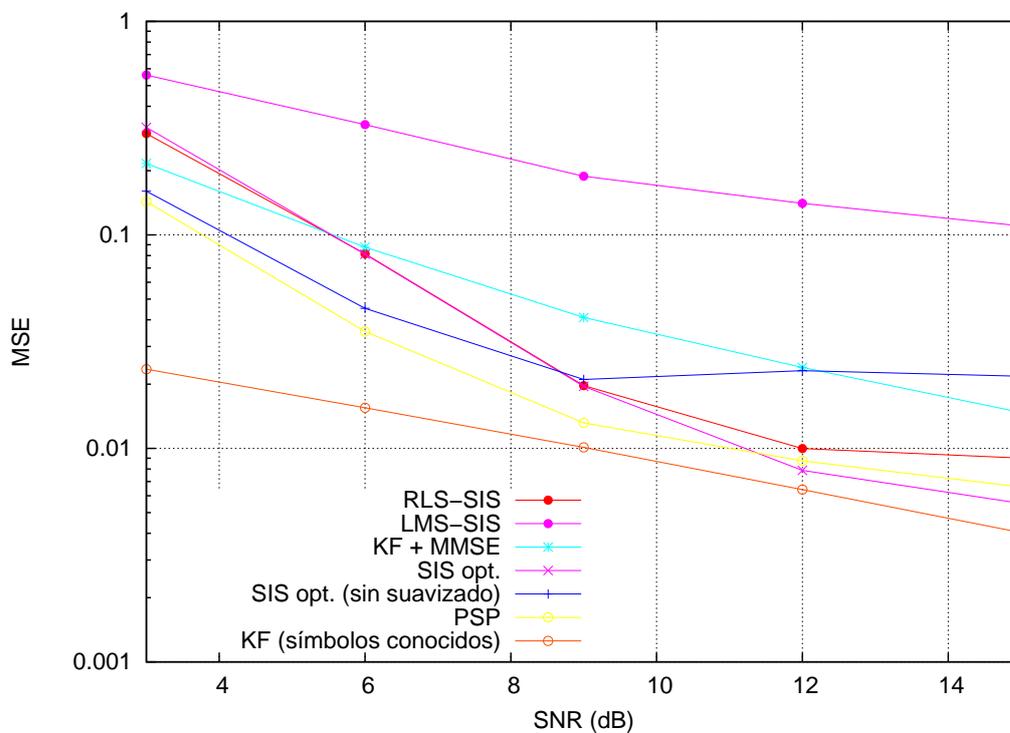


Figura 7.6: MSE para distintos valores de la SNR (dB). Los algoritmos SMC utilizan $M = 192$ partículas. Los resultados están promediados sobre 2037 tramas de datos. Modelo de autocorrelación de Clarke.

El error cometido por los distintos algoritmos en la estimación del canal se puede observar en la Figura 7.6. En este caso las diferencias entre el RLS-SIS y el SIS opt., aunque todavía apreciables, son menos evidentes y el RLS-SIS obtiene un rendimiento muy bueno, superior al del KF+MMSE. El LMS-SIS, por su parte, no es capaz de seguir la evolución del canal y para $\text{SNR} = 15$ dB el MSE que comete en la estimación del canal es un orden de magnitud superior al del RLS-SIS.

7.1.3. Influencia del número de partículas

Una de los mayores atractivos de los métodos de Monte Carlo secuenciales es que la aproximación que hacen de una distribución de probabilidad es tanto mejor cuanto mayor es el número de partículas empleadas. Así pues, es posible disminuir la tasa de error de bit de los algoritmos propuestos incrementando el número de partículas que utilizan. Se ilustra en la Figura 7.7, en la que se puede observar como la BER de los distintos igualadores decrece con el número de partículas para una SNR fija de 15 dB.

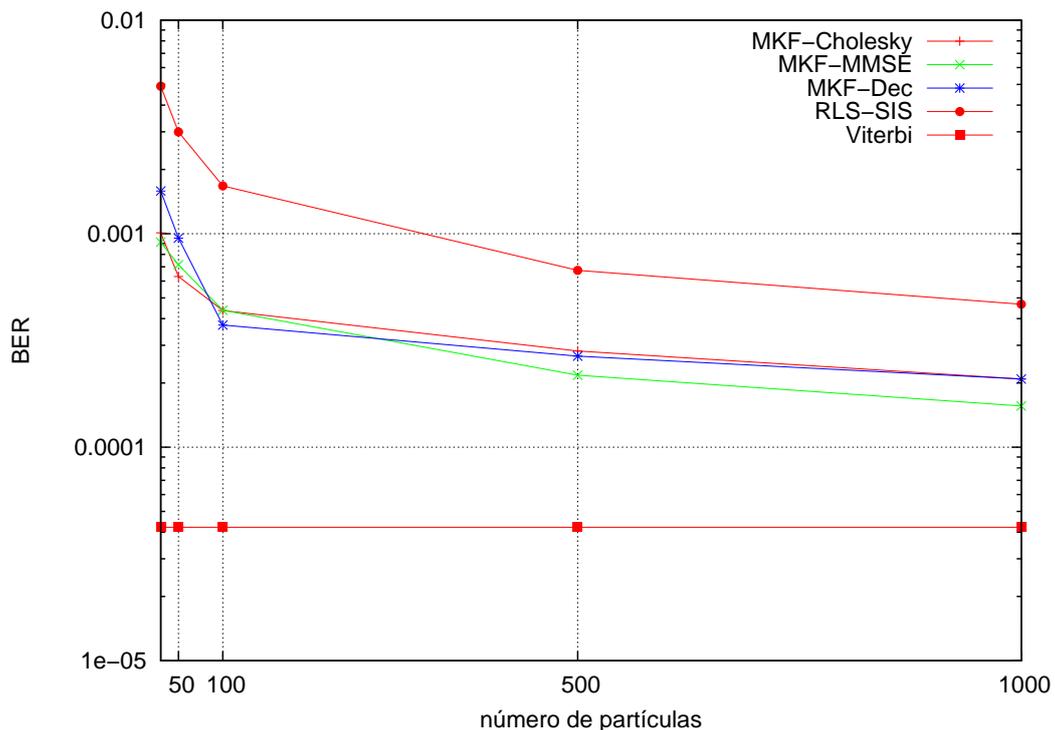


Figura 7.7: Evolución de la probabilidad *a posteriori* de los diferentes órdenes de canal en el MLSD-m . Los resultados están promediados sobre 12092 tramas de datos. Modelo de autocorrelación de Clarke.

7.2. Igualación de canales MIMO de orden desconocido

Los métodos de igualación para canales MIMO propuestos en el Capítulo 6 son capaces de manejar la incertidumbre asociada a un orden de canal desconocido y calcular la probabilidad *a posteriori* de cada uno de los posibles valores de éste.

Vamos a suponer que el orden del canal, m , puede tomar valores en el conjunto $\mathcal{M} = \{2, 3, 4, 5\}$ y vamos a comparar el rendimiento de los algoritmos propuestos en el Capítulo 6 con:

- un algoritmo de filtrado de partículas basado en la aproximación de la distribución conjunta del canal y los datos (Sección 5.2) haciendo uso del filtro RLS y con perfecto conocimiento del orden del canal, es decir, el RLS-SIS;
- el PSP convencional descrito en la Sección 4.1.4 cuando el orden de canal es conocido con tantos supervivientes como partículas utiliza el método SMC propuesto;
- el algoritmo de Viterbi cuando el canal es conocido.

Estudiaremos los dos métodos propuestos en el Capítulo 6 para la igualación de canales MIMO de orden desconocido en secciones separadas, manteniendo así la estructura empleada en su explicación.

7.2.1. Métodos SMC de complejidad reducida para canales MIMO de orden desconocido

En primer lugar se muestran los resultados obtenidos para el método SMC de complejidad reducida presentado en la Sección 6.4.1, etiquetado en las figuras como “USIS” (del inglés *Unknown channel order SIS*). Además de comparar su rendimiento con el del RLS-SIS explicado en la Sección 5.2, vamos a hacerlo también con tres versiones de dicho algoritmo considerando un orden de canal incorrecto:

- el “RLS-SIS ($m = 2$)”, que subestima el orden del canal,
- el “RLS-SIS ($m = 4$)”, que lo sobreestima por una unidad, y
- el “RLS-SIS ($m = 5$)”, que lo sobreestima por dos unidades.

La Figura 7.8 muestra la tasa de error de bit de los algoritmos para distintos valores de la SNR cuando el número de partículas es 192. Lo primero que se aprecia es que la curva correspondiente al algoritmo RLS-SIS subestimando el orden del canal es prácticamente plana y su valor es de 0,2 para $SNR = 15$ dB, por lo que la utilidad de dicho algoritmo es casi nula. Esto da una idea de la importancia que tiene una correcta estimación del orden del canal y explica por que suele asumirse un valor alto para él (posiblemente mucho mayor del esperado). Por otro lado, cuando el orden de canal considerado está una unidad por encima del auténtico, el RLS-SIS experimenta en este caso una pérdida de aproximadamente 3 dB para tasas de error de bit cercanas a 10^{-3} . Esta pérdida se acentúa notablemente cuando el orden de canal considerado está dos unidades por encima del verdadero y la diferencia entre la curva del RLS-SIS y la del RLS-SIS ($m = 5$) es ya de 5 dB para un BER de 10^{-2} . Así pues, si bien sobreestimar el orden del canal por una unidad es aceptable en términos de rendimiento³, éste se degrada paulatinamente a medida que la diferencia entre el orden asumido por el receptor y el verdadero crece, y tanto es así que el rendimiento del RLS-SIS considerando $m = 5$ es ya muy pobre.

³Nótese que, en cualquier caso, existe un incremento en la complejidad computacional.

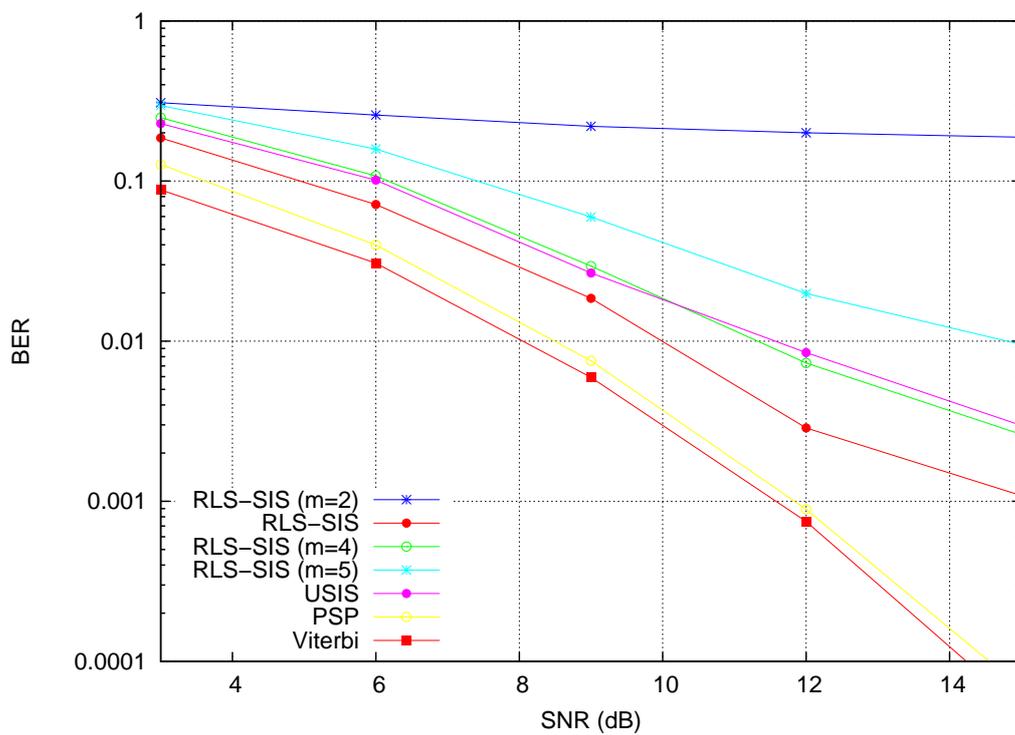


Figura 7.8: BER para distintos valores de la SNR (dB). Los algoritmos SMC utilizan $M = 192$ partículas. Los resultados están promediados sobre 1185 tramas de datos. Modelo de autocorrelación de Clarke.

El algoritmo propuesto, el USIS, tiene un rendimiento similar al RLS-SIS ($m = 4$), estando su curva ligeramente por debajo en regiones bajas de la SNR y ligeramente por encima en las altas. Por tanto, aunque debido al desconocimiento del orden de canal, m , el filtro de partículas propuesto sufre una pérdida de rendimiento, ésta es todavía tolerable, y más aún considerando que se trata de un algoritmo de complejidad cúbica.

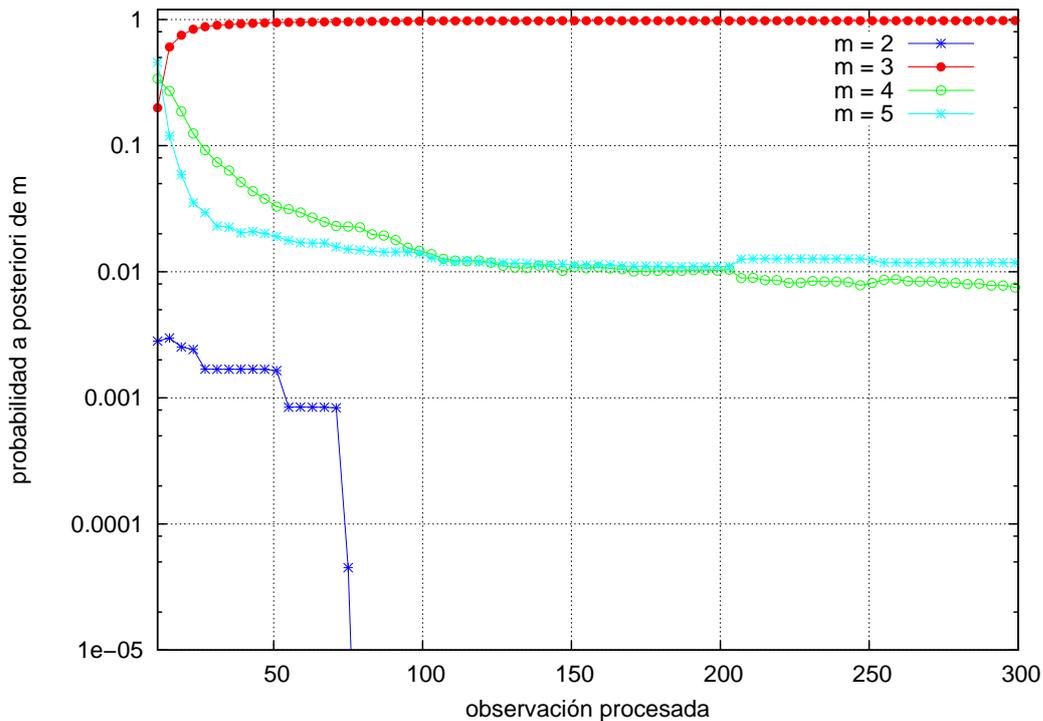


Figura 7.9: Evolución de la probabilidad *a posteriori* de los diferentes órdenes de canal en el algoritmo USIS. Los resultados están promediados sobre 1185 tramas de datos. Modelo de autocorrelación de Clarke.

En cuanto a la estimación del orden de canal, la Figura 7.9 muestra la evolución de las probabilidades *a posteriori* para los diferentes órdenes de canal calculadas por el USIS a medida que se va procesando la trama cuando la SNR es 15 dB. Se puede observar que la probabilidad del orden correcto alcanza muy rápidamente un valor muy cercano a 1 (aproximadamente 0,9 una vez procesadas las 50 primeras observaciones) mientras que la de los restantes órdenes de canal decrece muy rápidamente (desde el valor que tenían *a priori*, $\frac{1}{|\mathcal{M}|} = 0,25$), situándose hacia la mitad de la trama alrededor de 0,1, para el caso de $m = 4$ o $m = 5$, y a valores prácticamente nulos para $m = 2$.

7.2.2. Igualación mediante procesado por supervivientes

A continuación se muestran resultados numéricos obtenidos por el método basado en PSP para la igualación de canales MIMO de orden desconocido. Se han simulado dos versiones del

mismo:

- el “MLSD-m”, que es el método descrito en la Sección 6.4.2, utilizando filtros de Kalman para la estimación del canal, y
- el “MLSD-m (RLS)”, que estima el canal por medio del procedimiento RLS (una de las modificaciones propuestas para reducir la complejidad del método).

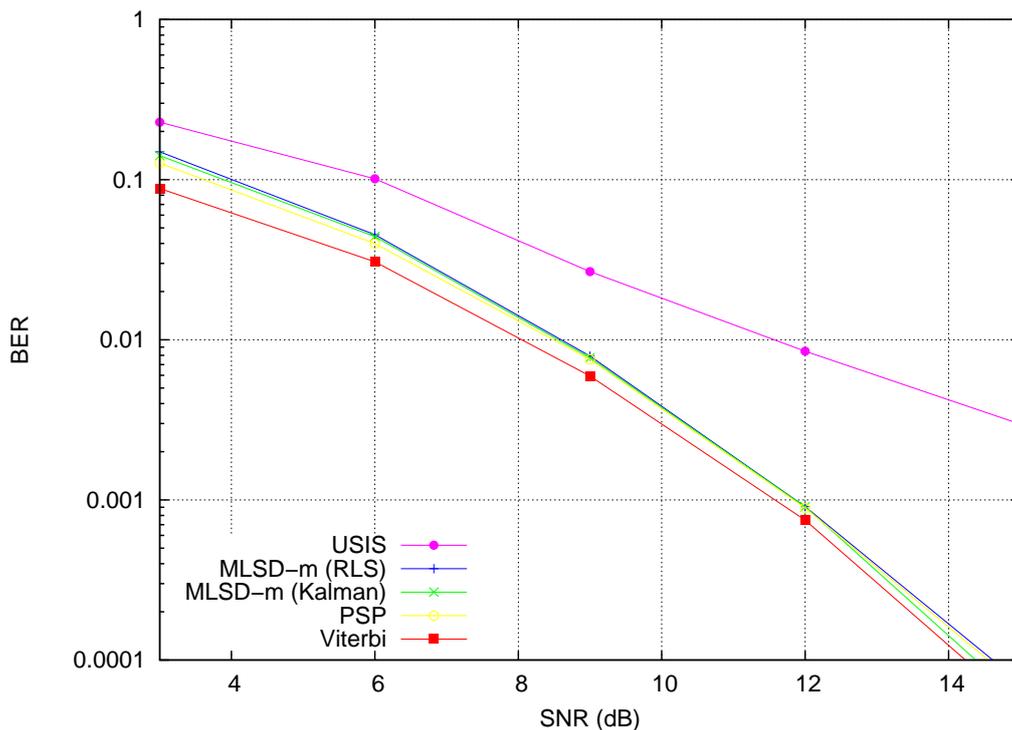


Figura 7.10: BER para distintos valores de la SNR (dB). Los algoritmos SMC utilizan $M = 192$ partículas. Los resultados están promediados sobre 1185 tramas de datos. Modelo de autocorrelación de Clarke.

El rendimiento de ambos algoritmos se puede ver en la Figura 7.10, junto con el obtenido por el USIS en las mismas circunstancias. Se observa que las curvas de BER de ambos algoritmos están muy próximas a la óptima, que viene dada por el Viterbi (con canal conocido), por lo que su rendimiento es muy superior al del igualador SMC para orden de canal desconocido (el USIS) descrito en la Sección 6.4.1. Sin embargo, nótese que dicho rendimiento se consigue a costa de una complejidad exponencial en el número de antenas transmisoras que éste último evita. En la figura también se aprecia como el MLSD-m consigue una tasa de error de bit ligeramente inferior a la del PSP cuando el orden del canal es conocido para SNRs altas. Es debido a que, tal como se vio en la Sección 6.4.2, el MLSD-m selecciona en todo momento los P mejores supervivientes globalmente (y no por estado).

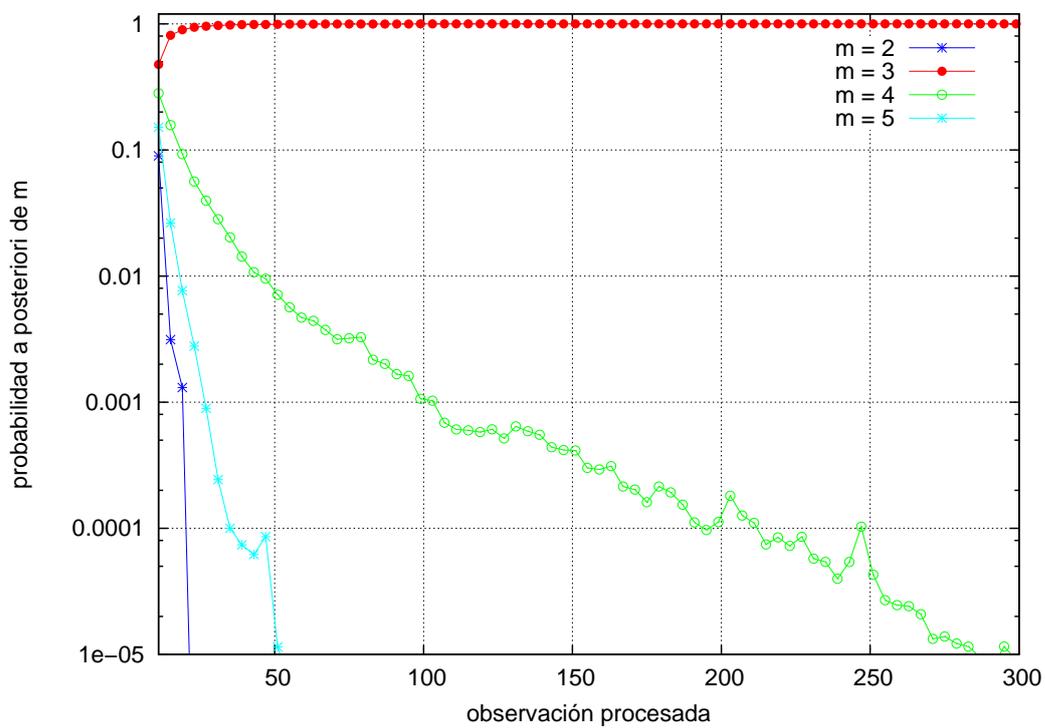


Figura 7.11: Evolución de la probabilidad *a posteriori* de los diferentes órdenes de canal en el MLSD-m. Los resultados están promediados sobre 1185 tramas de datos. Modelo de autocorrelación de Clarke.

	<i>SNR</i>				
	3	6	9	12	15
USIS	0.432068	0.246414	0.079325	0.032911	0.018565
MLSD-m	0.117300	0.011814	0.000844	0.000000	0.000000
MLSD-m (RLS)	0.168776	0.028692	0.001688	0.000000	0.000000
CME supervisado	0.001688	0.000000	0.000000	0.000000	0.000000

Tabla 7.1: Frecuencia de error relativa en la detección del orden del canal.

La Figura 7.11 muestra la evolución de las probabilidades *a posteriori* para los diferentes órdenes de canal calculadas por el MLSD-m cuando la SNR es 15 dB. La probabilidad asociada al verdadero orden de canal es prácticamente 1 apenas procesadas las 50 primeras observaciones de la trama, mientras que las de los restantes órdenes, especialmente las correspondientes a $m = 2$ y $m = 5$, decrecen rápidamente hasta alcanzar valores próximos a 0. Los resultados obtenidos para el MLSD-m (RLS) son análogos.

En la Tabla 7.1 se compara la frecuencia de error relativa en la detección del orden del canal de los algoritmos propuestos para la igualación de canales MIMO de orden desconocido. Se ha obtenido considerando que el orden detectado por un algoritmo para una determinada SNR es aquel que tiene una mayor probabilidad al final de la trama. También se incluyen los resultados que se obtendrían aplicando el criterio CME explicado en la Sección 6.2.2 cuando los símbolos son perfectamente conocidos (etiquetado “CME supervisado”). Los resultados están promediados sobre 1185 tramas.

Para SNRs altas el CME siempre detecta correctamente el orden del canal cuando se conoce perfectamente la secuencia de símbolos transmitida. Sin embargo, ésta hipótesis carece de sentido en una situación práctica y las simulaciones llevadas a cabo demuestran que cuando algunos de los símbolos empleados en el cálculo son erróneos el CME no funciona en absoluto (la frecuencia de error relativa en la detección del orden del canal es siempre 1).

7.3. Conclusiones

En este capítulo se ha estudiado por medio de simulaciones el rendimiento de todos los algoritmos propuestos a lo largo de la tesis. La primera sección se ha dedicado a los métodos SMC de complejidad reducida para la igualación de canales MIMO presentados en el Capítulo 5. Distinguíamos dos grupos: el de los que introducían el canal como una variable auxiliar de la función tentativa y el de los que trataban de aproximar la distribución conjunta del canal y los datos. En el primero se encuentran los algoritmos MKF-Cholesky, MKF-Dec y MKF-MMSE, que tienen una complejidad cúbica en el número de antenas transmisoras. Los tres obtienen un rendimiento muy cercano al del igualador SMC óptimo (tanto en términos de BER como de MSE cometido en la estimación del canal), que tiene una complejidad exponencial en el mismo parámetro. Las curvas de BER de los tres algoritmos son muy similares pero el MKF-MMSE

proporciona resultados ligeramente mejores cuando la SNR es alta.

Los otros algoritmos de complejidad reducida propuestos, cuya característica fundamental es la aproximación de la distribución conjunta del canal y los símbolos, son el RLS-SIS y el LMS-SIS. El rendimiento de ambos (y en especial el del LMS-SIS, cuyo estimador de canal no es capaz seguir la evolución de éste) es considerablemente peor que el de cualquiera de los algoritmos anteriormente mencionados. Sin embargo, y a pesar de que su complejidad es también cúbica en el número de antenas transmisoras, el número de operaciones que requieren es mucho menor (la constante asociada a la complejidad es más pequeña).

En todo el rango de relaciones señal a ruido consideradas, la tasa de error de bit obtenida por cualquiera de los igualadores SMC de complejidad reducida propuestos es menor que la de un receptor MMSE cuyas estimaciones de canal vienen dadas por un filtro de Kalman utilizando la secuencia de vectores de símbolos efectivamente transmitida. En todos los casos la complejidad es cúbica en el número de antenas transmisoras. Para el caso de los métodos SMC, es además lineal en el número de partículas, pero esta dificultad se puede superar mediante el uso de hardware paralelo.

Todas las gráficas de BER muestran el rendimiento tanto del igualador SMC óptimo como del PSP. Ambos tienen la misma complejidad computacional (exponencial en el número de antenas transmisoras y la memoria del canal) y la tasa de error de bit del segundo está siempre por debajo de la del primero. Por tanto, la relación rendimiento-coste computacional es mejor en el PSP que en el SIS opt.. Téngase en cuenta, sin embargo, que la complejidad de cualquiera de ellos es prohibitiva (el SIS opt. únicamente se ha simulado a modo de referencia) y al aumentar el número de antenas transmisoras por encima de un cierto umbral se hace necesario recurrir a algoritmos cuya complejidad no crezca tan rápidamente con este número.

La segunda sección de este capítulo analiza los resultados numéricos obtenidos de la simulación de los algoritmos propuestos para la igualación de canales MIMO de orden desconocido. El USIS extiende los métodos SMC que aproximaban la distribución conjunta del canal y los datos (en concreto el RLS-SIS) para manejar la incertidumbre que ahora afecta al orden del canal. El rendimiento del algoritmo es aproximadamente el mismo que el obtenido por el RLS-SIS sobreestimando el orden del canal por una unidad. Para una BER de 10^{-2} , esto supone una pérdida de aproximadamente unos 2 dB con respecto al RLS-SIS que conoce el orden del canal, pero una ganancia superior a los 3 dB con respecto al mismo algoritmo cuando lo sobreestima por dos unidades. Por otro lado, subestimar el orden del canal prácticamente imposibilita una detección fiable de los símbolos transmitidos.

La complejidad del USIS es lineal en el número de posibles órdenes de canal considerados, pero sigue siendo cúbica en el número de antenas transmisoras.

El otro algoritmo propuesto para la igualación de canales MIMO de orden desconocido es el MLSD-m. Está basado en procesado por supervivientes y, a costa de una complejidad exponencial en el número de antenas transmisoras, alcanza un rendimiento comparable al del PSP convencional que conoce el verdadero orden del canal (y al del algoritmo de Viterbi, por tanto).

Capítulo 8

Conclusiones y líneas futuras

8.1. Conclusiones

El objetivo primordial del presente trabajo ha sido la aplicación de los métodos secuenciales de Monte Carlo (SMC) al problema de la igualación de canales MIMO selectivos en frecuencia y en tiempo. Un sistema de comunicaciones MIMO emplea varias antenas tanto en emisión como en recepción, lo que da lugar a un incremento en la capacidad del canal. Esto se traduce en un mejor aprovechamiento del medio de transmisión, lo cual permite aumentar las prestaciones del sistema. Pero para ello son necesarias técnicas de procesado de señal más sofisticadas capaces de explotar la dimensión espacial del canal.

Los métodos SMC, también conocidos como filtros de partículas, son técnicas estocásticas cuyo objetivo es la aproximación recursiva de distribuciones de probabilidad de interés mediante muestras y pesos asociados. Normalmente las muestras se obtienen de una distribución que recibe el nombre de tentativa, y se ponderan mediante el principio de muestreo enfatizado secuencial (SIS). Los filtros de partículas se utilizan principalmente para estimar la señal no observada en sistemas dinámicos en formato de espacio de estados. El filtro de Kalman (KF) proporciona la solución óptima a este problema cuando las ecuaciones del sistema verifican las hipótesis de linealidad y gaussianidad. Sin embargo, su rendimiento se degrada de manera significativa cuando no se dan tales condiciones. Los métodos de Monte Carlo secuenciales no imponen tales restricciones, lo que los convierte en una herramienta mucho más versátil y general.

Las ecuaciones que utilizamos en el Capítulo 3 para modelar el proceso de transmisión en un sistema MIMO se pueden interpretar como las ecuaciones que definen un sistema dinámico en formato de espacio de estados. La secuencia de matrices de canal¹ y la de símbolos transmitidos constituyen, en este caso, la señal no observada (el estado) que queremos estimar a partir de la secuencia de observaciones. Desde este punto de vista es posible aplicar la metodología SMC a la igualación de canales MIMO, lo cual constituye el tema principal de la tesis.

En la literatura se pueden encontrar algoritmos de filtrado de partículas que resuelven el

¹Recuérdese que estamos tratando con canales variantes en el tiempo.

problema de la igualación cuasi-óptima de canales MIMO, pero su complejidad es por lo general exponencial en el número de antenas transmisoras. Recordemos que la motivación para utilizar un sistema MIMO es en última instancia conseguir un aumento en la capacidad del canal, y una condición necesaria para ello es incrementar el número antenas transmisoras. Un algoritmo cuyo coste computacional depende exponencialmente del número de antenas transmisoras limita (en la medida en que lo dicte la potencia de cálculo disponible) el número de ellas que se pueden emplear y, por tanto, la ganancia en la capacidad que se puede obtener mediante el uso de un sistema MIMO.

La principal aportación de este trabajo es el desarrollo de métodos SMC de complejidad reducida para la igualación de canales MIMO. Los algoritmos a que dan lugar las técnicas propuestas tienen una complejidad cúbica en el número de antenas transmisoras y exhiben un rendimiento muy cercano al del igualador SMC óptimo (derivado en el Capítulo 4). Además, están diseñados para canales (desconocidos) variantes en el tiempo y selectivos en frecuencia, por lo que son especialmente adecuados en el ámbito de los sistemas de comunicación móviles.

También se ha dedicado una parte importante de la tesis al problema que se plantea a la hora de igualar un canal MIMO cuando el orden del mismo es desconocido. La aproximación habitual consistente en sobreestimar dicho parámetro (asumir un determinado orden de canal que, supuestamente, estará siempre por encima del verdadero) degrada el rendimiento del receptor al tiempo que incrementa su complejidad computacional. En el Capítulo 6 hemos presentado dos métodos capaces de manejar la incertidumbre debida a un orden de canal desconocido: uno de ellos basado en filtrado de partículas y el otro en procesado por supervivientes.

8.1.1. Métodos SMC de complejidad reducida

En el Capítulo 5 se presentaron dos métodos de filtrado de partículas para la igualación de canales MIMO cuya complejidad computacional es cúbica en el número de antenas transmisoras. Ambos están basados en el uso de una función tentativa cuidadosamente diseñada para evitar calcular en cada instante de tiempo la probabilidad todos los posibles vectores de símbolos transmitidos.

En el primer método el canal se incluye en la f.d.p. objetivo a modo de variable auxiliar para facilitar la detección de los símbolos transmitidos. Las muestras del mismo se obtienen en cada partícula utilizando la media y covarianza estimadas por un filtro de Kalman a partir de la secuencia de símbolos asociada a la partícula. Haciendo uso de esta metodología se han derivado los algoritmos MKF-Cholesky, MKF-MMSE y MKF-Dec. Se diferencian en la manera de llevar a cabo el muestreo de los símbolos, pero todos ellos emplean el KF para la estimación del canal (de ahí el prefijo MKF, *Mixture Kalman Filter*).

El MKF-Cholesky se basa en triangularización de la matriz de autocorrelación del canal (mediante la descomposición de Cholesky) para dotar a las observaciones de una estructura que permite muestrear los elementos del vector de símbolos transmitidos secuencialmente. Los algoritmos MKF-MMSE y MKF-Dec, por otro lado, emplean receptores lineales (el

filtro MMSE y el de decorrelación, respectivamente), contruidos utilizando las muestras del canal, para obtener estimaciones blandas de los símbolos transmitidos que sirven, a su vez, para generar muestras de los mismos. Los tres algoritmos tienen una tasa de error de bits (BER) muy similar a la del igualador SMC óptimo (cuya complejidad es exponencial en el número de antenas transmisoras), la cual a su vez está próxima a la obtenida por un detector de máxima verosimilitud con conocimiento perfecto del canal. La curva de BER del MKF-MMSE, el algoritmo que demuestra un mejor rendimiento cuando la SNR es alta, se encuentra a aproximadamente 0,5 dB de la del SIS opt. y a algo más de 1 dB de la del PSP.

El otro método SMC para la igualación de canales MIMO propuesto en el Capítulo 5 aborda el problema desde un punto de vista diferente. La idea fundamental es considerar el canal no como una variable auxiliar que se muestrea por conveniencia y luego se descarta, sino como parte integrante de la f.d.p. objetivo. Se trata pues, en este caso, de aproximar la distribución conjunta del canal y los datos. Este nuevo planteamiento evita el uso de filtros de Kalman, con lo que se consigue una reducción en la complejidad del método. Las muestras del canal se obtienen entonces con la ayuda de simples estimadores de canal adaptativos (empleando algoritmos tales como el RLS o el LMS). A partir de ellas, y mediante una versión simplificada del receptor MMSE, se calculan estimaciones blandas de los símbolos transmitidos que servirán para tomar muestras de los mismos. Los algoritmos RLS-SIS y LMS-SIS implementan el método descrito haciendo uso de estimadores de canal RLS, el primero, y LMS el segundo. La complejidad de ambos es también cúbica en el número de antenas transmisoras, pero el número de operaciones que precisan es sensiblemente menor que el requerido por cualquiera de los algoritmos explicados en la Sección 5.1 (la constante asociada a la complejidad es menor) y así se ha constatado en los tiempos de ejecución de las simulaciones correspondientes. En cuanto al rendimiento, el RLS-SIS sólo experimenta una cierta pérdida con respecto a los algoritmos basados en el uso de variables auxiliares cuando la relación señal a ruido (SNR) es alta. En tal caso, alrededor de 3 dB separan su curva de BER de la del SIS opt. y 3,5 de la del PSP (cuyo tasa de error de bit, es prácticamente equiparable a la que proporciona el algoritmo de Viterbi con el canal conocido). El LMS-SIS, por su parte, muestra un rendimiento bastante pobre, debido fundamentalmente a que el estimador LMS no es capaz de seguir la evolución del canal. Así pues, estos algoritmos encuentran su utilidad en aquellas situaciones en las que la SNR del sistema no es muy alta o los recursos computacionales son especialmente limitados.

En todas las simulaciones llevadas a cabo, la curva de BER del PSP está por debajo de la de cualquiera de los igualadores SMC propuestos, incluido el SIS opt.. Teniendo en cuenta que tanto el PSP como el SIS opt. tienen la misma complejidad computacional, podemos concluir que el PSP proporciona una mejor relación rendimiento/coste computacional. Sin embargo, mientras que los métodos de filtrado de partículas son susceptibles de ser paralelizados (véase [50, 93]) (lo que permite reducir drásticamente el tiempo de ejecución), no ocurre lo mismo con PSP, cuya implementación no permite un alto grado de paralelismo debido al acoplamiento existente entre los distintos supervivientes. Así pues, utilizando hardware paralelo el SIS opt. puede alcanzar mayores velocidades de procesamiento con una pérdida muy pequeña de

rendimiento (en torno a 0,5 dB). Téngase en cuenta, en cualquier caso, que tanto la complejidad del PSP como la del SIS opt. son exponenciales en el número de antenas transmisoras y en la memoria del canal, lo que hace que ambos algoritmos sean inviables en la mayoría de situaciones prácticas.

Un resultado a destacar es que el rendimiento de cualquiera de los algoritmos SMC de complejidad reducida presentados en el Capítulo 5 supera en términos de BER al del método etiquetado “KF (símbolos conocidos)+MMSE”. Éste consiste en un receptor lineal MMSE suavizado cuyas estimaciones de canal vienen dadas por un filtro de Kalman alimentado con los símbolos verdaderos, y su complejidad es también cúbica en el número de antenas transmisoras.

El material de esta parte de la tesis ha dado lugar a las publicaciones [80, 81, 94, 95].

8.1.2. Canales MIMO de orden desconocido

La longitud (en períodos de símbolo) de la respuesta al impulso de un canal de comunicaciones es lo que habitualmente se conoce como el orden del mismo. La mayor parte de los trabajos que se pueden encontrar en la literatura abordando el problema de la igualación de un canal MIMO asumen que dicho parámetro es conocido. Esta situación difícilmente se da en la realidad y es una práctica habitual considerar un orden de canal elevado que, supuestamente, nunca estará por debajo del verdadero. La motivación detrás de esta última aproximación es evitar la enorme pérdida de rendimiento que sufre un receptor que subestima el orden del canal². Pero sobreestimar el orden del canal (suponer un orden mayor del que realmente tiene) tampoco es deseable pues da lugar a un incremento en la complejidad computacional del receptor, por un lado, y a una disminución de su rendimiento, por el otro, que además se acentúa a medida que el orden considerado se aleja (por encima) del auténtico orden de canal. En relación a esto último, nótese que el canal de un sistema de comunicaciones móvil está cambiando constantemente, por lo que evitar en todo momento subestimar el orden del mismo (considerando un orden de canal muy elevado) puede suponer sobreestimarlos por varias unidades, incrementando considerablemente la complejidad del receptor.

En la literatura se pueden encontrar diversas aproximaciones para detectar el orden de un canal, pero ninguna de ellas lo hace para el caso de un canal MIMO variante con el tiempo. En el Capítulo 6 se han propuesto dos algoritmos que estiman la probabilidad *a posteriori* del orden del canal en un sistema MIMO. Ambos consideran éste último como una variable aleatoria discreta con distribución *a priori* uniforme que toma valores en un conjunto conocido. El primero es un algoritmo de filtrado de partículas que extiende los métodos de complejidad reducida derivados en la Sección 5.2 asociando a cada partícula un banco de estimadores de canal, uno por cada posible valor que puede tomar el orden del canal. Haciendo uso de ellos y de receptores lineales se calculan probabilidades para muestrear los símbolos transmitidos que no dependen del verdadero orden del canal. A la vista de las simulaciones realizadas, el algoritmo resultante, denominado USIS (del inglés *Unknown channel order-SIS*), tiene aproximadamente

²Este hecho es corroborado por algunas de las simulaciones llevadas a cabo, cuyos resultados se pueden encontrar en el Capítulo 7.

la misma tasa de error de bits que un RLS-SIS sobreestimando el orden del canal por una unidad. Esto supone una considerable pérdida de rendimiento con respecto al igualador SMC óptimo en las regiones altas de la SNR. Sin embargo, a pesar de la incertidumbre asociada al orden del canal, el USIS todavía tiene una complejidad cúbica en el número de antenas transmisoras.

El otro algoritmo propuesto para la igualación de canales MIMO de orden desconocido es determinista y está basado en el concepto de procesado por supervivientes. Un superviviente es una secuencia de vectores de símbolos posiblemente transmitida a la que el receptor asocia un coste. Manteniendo un banco de filtros de Kalman en cada superviviente, uno por cada posible orden del canal, es posible calcular en cada instante de tiempo la verosimilitud de cada potencial superviviente independientemente del orden de canal desconocido. El algoritmo selecciona entonces globalmente los P mejores y descarta los restantes. Se ha denominado MLS-D-m porque es una aproximación de un detector de máxima verosimilitud (*Maximum Likelihood Sequence Detector* o MLS-D), que desconoce el valor del parámetro m (el orden del canal). A costa de una complejidad exponencial en el número de antenas transmisoras, este sencillo esquema exhibe un rendimiento cercano al del PSP convencional y, por extensión, al del algoritmo de Viterbi que conoce perfectamente el canal. Al final de la Sección 6.4.2 se propone una modificación del algoritmo (utilizando estimadores de canal RLS en lugar de filtros de Kalman) que consigue prácticamente el mismo rendimiento con una reducción importante en la complejidad.

Una característica de los métodos propuestos es que calculan en cada instante de tiempo la f.m.p. *a posteriori* del orden del canal. Esto permite en un momento dado, y cuando se cumple una determinada condición (como puede ser, que una de las probabilidades está por encima de un determinado umbral), tomar una decisión dura acerca del orden del canal y conmutar a un algoritmo que considera éste como un parámetro conocido, con la consiguiente reducción en la complejidad e incremento en el rendimiento del algoritmo.

Los resultados obtenidos en esta parte del trabajo han dado lugar a las publicaciones [96, 97, 98].

8.2. Líneas futuras de investigación

En esta sección se dan algunas ideas para extender los resultados de la tesis. Aspectos relacionados con un análisis más exhaustivo de los métodos propuestos, que tienen interés de cara a una implementación práctica de los mismos, o ideas que pueden servir de base para nuevos trabajos son algunos de los puntos tratados aquí.

8.2.1. Evaluación del rendimiento de los métodos propuestos utilizando un demostrador MIMO

Todos los métodos propuestos se han sido diseñado (por simplicidad) suponiendo que el canal sigue un proceso autorregresivo de primer orden. Éste no es, sin embargo, el modelo

empleado en el capítulo de simulaciones. Con la intención de probar la robustez de los distintos algoritmos, los canales considerados en dicho capítulo se han generado de acuerdo al modelo de Clarke [92], ampliamente utilizado y, posiblemente, más realista. Nótese sin embargo que, tanto en un caso como en otro, efectivamente sólo se trata de modelos y, como tales, únicamente son esquemas teóricos ideados para explicar el funcionamiento de un sistema complejo (el canal de comunicaciones, en este caso). Por tanto, los resultados obtenidos en un sistema MIMO real podrían diferir de los predichos por medio de simulaciones. Una manera de evaluar el rendimiento de un igualador MIMO en una situación real es haciendo uso de un demostrador. Así pues, una extensión interesante a los resultados presentados en este trabajo sería estudiar los problemas prácticos que surgen en la implementación real de los métodos propuestos.

8.2.2. Paralelización

El tiempo de ejecución de un método de filtrado de partículas se puede reducir considerablemente mediante el uso de hardware paralelo. Cada partícula puede ser propagada independientemente de todas las demás salvo cuando se hace necesario un paso de remuestreo. Es lo que representa la Figura 8.1, en la que elementos de procesado diferentes se encargan de propagar la trayectoria y actualizar el peso de partículas diferentes hasta que tiene lugar una operación de remuestreo en el instante $t+k$. Los métodos de remuestreo tradicionales requieren que todas las partículas sean examinadas a la vez para determinar cuales se van a replicar. Éste es el mayor cuello de botella que existe en la paralelización de un filtro de partículas. Sin embargo, ya se ha comentado en el Capítulo 2 que recientemente se han propuesto diversos algoritmos de remuestreo que palían en cierta medida este problema. Qué ganancia en la velocidad de procesamiento es posible obtener de esta forma y cuales son los algoritmos más adecuados para el problema que nos ocupa son cuestiones a las que no se ha dado respuesta en el presente trabajo.

8.2.3. Aplicación en sistemas de localización

Todos los métodos de igualación propuestos generan estimaciones del canal como subproducto del proceso de detección de los símbolos transmitidos. Podemos aprovechar este hecho para construir un sistema de localización basado en asociar a cada posición un valor del canal (MIMO) que se establecería entre un receptor ubicado en ella y una serie de balizas convenientemente distribuidas en una región del espacio de interés. La posición de éstas es conocida y también se dispone de información acerca del estado del canal entre ellas y varios receptores imaginarios ubicados en diferentes posiciones de la región de interés. Todo ello sirve para, mediante interpolación, calcular la posición del receptor dada la estimación que hace del canal.

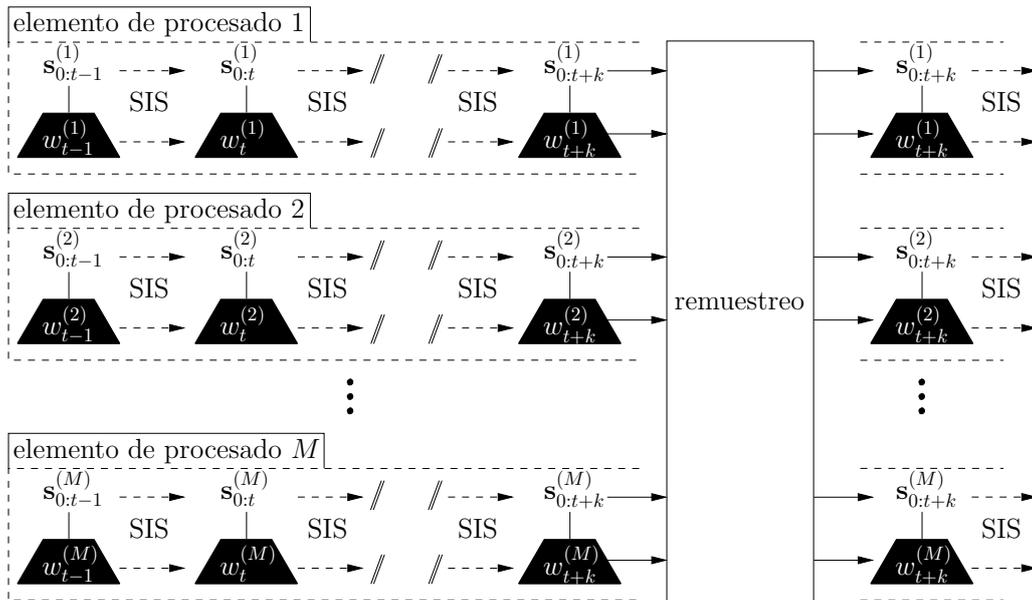


Figura 8.1: Posible implementación paralela de los igualadores SMC propuestos. Cada partícula es propagada por un elemento de procesamiento diferente salvo cuando se lleva a cabo el remuestreo.

8.2.4. Orden de canal desconocido y variante con el tiempo

El Capítulo 6 de la tesis se dedicó a la igualación de canales MIMO de orden desconocido. Sin embargo, se hizo asumiendo que dicho parámetro se mantiene constante a lo largo de toda la trama de datos. Si el tiempo que abarca ésta es muy grande, podría ocurrir que el orden del canal variase durante la misma, lo cual complicaría su estimación. Una dificultad adicional surge si además consideramos un orden de canal diferente en cada uno de los subcanales que constituyen un canal MIMO.

8.2.5. Codificación turbo

Cualquiera de los igualadores propuestos lleva a cabo una detección blanda de los símbolos transmitidos, es decir, calcula probabilidades para los mismos. Es por ello que los algoritmos propuestos son susceptibles de ser empleados en un lazo de codificación turbo. Aunque en la literatura ya se pueden encontrar trabajos que combinan filtros de partículas con turbo códigos (véase [25], por ejemplo), es interesante estudiar el rendimiento de un esquema de tales características utilizando los métodos propuestos (dadas las diferencias conceptuales y de rendimiento existentes entre éstos y los filtros de partículas que típicamente se encuentran en la bibliografía).

Apéndice A

Derivación del filtro MMSE

Sea

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{g} \quad (\text{A.1})$$

la ecuación que modela los valores observados en las antenas receptoras de un sistema MIMO, cuando \mathbf{s} es el vector de símbolos transmitidos, \mathbf{H} la matriz de canal y \mathbf{g} el ruido que afecta a la transmisión. Vamos a suponer que, además del vector de observaciones \mathbf{y} , está disponible la matriz de canal \mathbf{H} . Asimismo, $\mathbb{E}[\mathbf{s}] = \mathbf{0}$ (la media del vector de símbolos transmitidos es cero), $\mathbb{E}[\mathbf{ss}^\top] = \sigma_s^2 \mathbf{I}$ (los símbolos son independientes entre sí) y $\mathbb{E}[\mathbf{gg}^\top] = \sigma_g^2 \mathbf{I}$ (las componentes de ruido que afectan a las diferentes antenas receptoras también son independientes), siendo \mathbf{I} una matriz identidad del tamaño adecuado.

Nos planteamos encontrar la estimación, $\hat{\mathbf{s}}$, del vector \mathbf{s} que minimiza el error cuadrático medio por medio de un filtro (matricial) lineal \mathbf{F} . Es decir, queremos hallar \mathbf{F} tal que

$$\mathbf{F} = \arg \min_{\mathbf{F}} \mathbb{E} \left[|\mathbf{F}^\top \mathbf{y} - \mathbf{s}|^2 \right]. \quad (\text{A.2})$$

Para ello derivaremos la esperanza de la ecuación (A.2) con respecto a la matriz \mathbf{F}^\top ¹. Dicha esperanza se puede desarrollar como

$$\begin{aligned} \mathbb{E} \left[|\mathbf{F}^\top \mathbf{y} - \mathbf{s}|^2 \right] &= \mathbb{E} \left[(\mathbf{F}^\top \mathbf{y} - \mathbf{s})^\top (\mathbf{F}^\top \mathbf{y} - \mathbf{s}) \right] \\ &= \mathbb{E} \left[\text{Tr} \left\{ (\mathbf{F}^\top \mathbf{y} - \mathbf{s}) (\mathbf{F}^\top \mathbf{y} - \mathbf{s})^\top \right\} \right] \\ &= \mathbb{E} \left[\text{Tr} \left\{ (\mathbf{F}^\top \mathbf{y} - \mathbf{s}) (\mathbf{y}^\top \mathbf{F} - \mathbf{s}^\top) \right\} \right] \\ &= \mathbb{E} \left[\text{Tr} \left\{ (\mathbf{F}^\top \mathbf{y} \mathbf{y}^\top \mathbf{F} - \mathbf{F}^\top \mathbf{y} \mathbf{s}^\top - \mathbf{s} \mathbf{y}^\top \mathbf{F} + \mathbf{ss}^\top) \right\} \right] \\ &= \text{Tr} \left\{ \mathbf{F}^\top \mathbb{E} [\mathbf{yy}^\top] \mathbf{F} \right\} - \text{Tr} \left\{ \mathbf{F}^\top \mathbb{E} [\mathbf{ys}^\top] \right\} - \text{Tr} \left\{ \mathbb{E} [\mathbf{sy}^\top] \mathbf{F} \right\} + \\ &\quad + \text{Tr} \left\{ \mathbb{E} [\mathbf{ss}^\top] \right\}, \end{aligned} \quad (\text{A.3})$$

donde $\text{Tr} \{ \cdot \}$ denota la traza de la matriz en el argumento.

¹Se deriva con respecto a la matriz traspuesta para facilitar los cálculos

La derivada con respecto a \mathbf{F}^\top de la esperanza en (A.3) es la suma de las derivadas con respecto a \mathbf{F}^\top de cada uno de los sumandos y éstas se calculan utilizando las reglas de derivación recogidas en el Apéndice F. En concreto,

$$\frac{\partial}{\partial \mathbf{F}^\top} \text{Tr} \{ \mathbf{F}^\top \mathbb{E} [\mathbf{y}\mathbf{y}^\top] \mathbf{F} \} = \mathbf{F}^\top \mathbb{E} [\mathbf{y}\mathbf{y}^\top]^\top + \mathbf{F}^\top \mathbb{E} [\mathbf{y}\mathbf{y}^\top] \quad (\text{A.4})$$

$$\frac{\partial}{\partial \mathbf{F}^\top} \text{Tr} \{ \mathbf{F}^\top \mathbb{E} [\mathbf{y}\mathbf{s}^\top] \} = \mathbb{E} [\mathbf{y}\mathbf{s}^\top]^\top = \mathbb{E} [\mathbf{s}\mathbf{y}^\top] \quad (\text{A.5})$$

$$\frac{\partial}{\partial \mathbf{F}^\top} \text{Tr} \{ \mathbb{E} [\mathbf{s}\mathbf{y}^\top] \mathbf{F} \} = \frac{\partial}{\partial \mathbf{F}^\top} \text{Tr} \{ \mathbf{F}^\top \mathbb{E} [\mathbf{s}\mathbf{y}^\top]^\top \} = \mathbb{E} [\mathbf{s}\mathbf{y}^\top] \quad (\text{A.6})$$

$$\frac{\partial}{\partial \mathbf{F}^\top} \text{Tr} \{ \mathbb{E} [\mathbf{s}\mathbf{s}^\top] \} = 0, \quad (\text{A.7})$$

donde $\frac{\partial}{\partial A} f$ es la derivada de la función escalar f respecto de la matriz A . Utilizando las ecuaciones (A.4)-(A.7) resulta

$$\frac{\partial}{\partial \mathbf{F}^\top} \mathbb{E} [|\mathbf{F}^\top \mathbf{y} - \mathbf{s}|^2] = \mathbf{F}^\top \mathbb{E} [\mathbf{y}\mathbf{y}^\top]^\top + \mathbf{F}^\top \mathbb{E} [\mathbf{y}\mathbf{y}^\top] - 2\mathbb{E} [\mathbf{s}\mathbf{y}^\top]. \quad (\text{A.8})$$

Las esperanzas que aparecen en (A.8) se calculan como:

$$\begin{aligned} \mathbb{E} [\mathbf{y}\mathbf{y}^\top] &= \mathbb{E} [(\mathbf{H}\mathbf{s} + \mathbf{g})(\mathbf{H}\mathbf{s} + \mathbf{g})^\top] = \mathbb{E} [(\mathbf{H}\mathbf{s} + \mathbf{g})(\mathbf{s}^\top \mathbf{H}^\top + \mathbf{g}^\top)] \\ &= \mathbb{E} [\mathbf{H}\mathbf{s}\mathbf{s}^\top \mathbf{H}^\top + \mathbf{H}\mathbf{s}\mathbf{g}^\top + \mathbf{g}\mathbf{s}^\top \mathbf{H}^\top + \mathbf{g}\mathbf{g}^\top] \\ &= \mathbb{E} [\mathbf{H}\mathbf{s}\mathbf{s}^\top \mathbf{H}^\top] + \mathbb{E} [\mathbf{H}\mathbf{s}\mathbf{g}^\top] + \mathbb{E} [\mathbf{g}\mathbf{s}^\top \mathbf{H}^\top] + \mathbb{E} [\mathbf{g}\mathbf{g}^\top] \\ &= \mathbf{H}\mathbb{E} [\mathbf{s}\mathbf{s}^\top] \mathbf{H}^\top + \mathbf{H}\mathbb{E} [\mathbf{s}\mathbf{g}^\top] + \mathbb{E} [\mathbf{g}\mathbf{s}^\top] \mathbf{H}^\top + \mathbb{E} [\mathbf{g}\mathbf{g}^\top], \\ &= \mathbf{H}\sigma_s^2 \mathbf{H}^\top + \sigma_g^2 \mathbf{I} \end{aligned} \quad (\text{A.9})$$

$$\begin{aligned} \mathbb{E} [\mathbf{s}\mathbf{y}^\top] &= \mathbb{E} [\mathbf{s}(\mathbf{H}\mathbf{s} + \mathbf{g})^\top] = \mathbb{E} [\mathbf{s}(\mathbf{s}^\top \mathbf{H}^\top + \mathbf{g}^\top)] = \mathbb{E} [\mathbf{s}\mathbf{s}^\top \mathbf{H}^\top + \mathbf{s}\mathbf{g}^\top] \\ &= \mathbb{E} [\mathbf{s}\mathbf{s}^\top] \mathbf{H}^\top + \mathbb{E} [\mathbf{s}\mathbf{g}^\top] = \sigma_s^2 \mathbf{H}^\top, \end{aligned} \quad (\text{A.10})$$

donde se ha tenido en cuenta que, puesto que los símbolos transmitidos son independientes del ruido y $\mathbb{E} [\mathbf{s}] = \mathbf{0}$, $\mathbb{E} [\mathbf{s}\mathbf{g}^\top] = \mathbb{E} [\mathbf{s}] \mathbb{E} [\mathbf{g}^\top] = \mathbf{0}$, y lo mismo ocurre con $\mathbb{E} [\mathbf{g}\mathbf{s}^\top] = \mathbb{E} [\mathbf{s}\mathbf{g}^\top]^\top$.

Obsérvese que, a la vista de (A.9), $\mathbb{E} [\mathbf{y}\mathbf{y}^\top] = \mathbb{E} [\mathbf{y}\mathbf{y}^\top]^\top$, por lo que

$$\frac{\partial}{\partial \mathbf{F}^\top} \mathbb{E} [|\mathbf{F}^\top \mathbf{y} - \mathbf{s}|^2] = 2\mathbf{F}^\top \mathbb{E} [\mathbf{y}\mathbf{y}^\top] - 2\mathbb{E} [\mathbf{s}\mathbf{y}^\top] \quad (\text{A.11})$$

y sustituyendo (A.9) y (A.10) en (A.11) resulta

$$\frac{\partial}{\partial \mathbf{F}^\top} \mathbb{E} [|\mathbf{F}^\top \mathbf{y} - \mathbf{s}|^2] = 2\mathbf{F}^\top (\mathbf{H}\sigma_s^2 \mathbf{H}^\top + \sigma_g^2 \mathbf{I}) - 2\sigma_s^2 \mathbf{H}^\top. \quad (\text{A.12})$$

Finalmente, igualando a cero la matriz de derivadas parciales para hallar el mínimo de la función

tenemos

$$\begin{aligned}
 2\mathbf{F}^\top (\mathbf{H}\sigma_s^2\mathbf{H}^\top + \sigma_g^2\mathbf{I}) - 2\sigma_s^2\mathbf{H}^\top &= 0 \\
 \Downarrow \\
 \mathbf{F}^\top (\mathbf{H}\sigma_s^2\mathbf{H}^\top + \sigma_g^2\mathbf{I}) &= \sigma_s^2\mathbf{H}^\top \\
 \Downarrow \\
 \mathbf{F}^\top &= \sigma_s^2\mathbf{H}^\top (\mathbf{H}\sigma_s^2\mathbf{H}^\top + \sigma_g^2\mathbf{I})^{-1} \tag{A.13}
 \end{aligned}$$

Apéndice B

Detección mediante el algoritmo de Viterbi

En un sistema dinámico como el de la Figura 2.4, una secuencia de $T + 1$ estados (desde 0 hasta T) queda completamente determinada por T transiciones entre estados, cada una de las cuales produce una salida. El algoritmo de Viterbi asocia en cada instante de tiempo un coste a cada posible transición en función de lo que difiere su correspondiente salida de la observación. Así, una secuencia de estados tiene un coste que viene dado por la suma de los costes de las transiciones que la originan, y el algoritmo procesa las observaciones secuencialmente calculando en cada instante el coste de todos los posibles caminos que llegan hasta cada uno de los estados. Pero en un instante dado, $t < T$, se pueden descartar todos los caminos que llegan a un determinado estado salvo uno, el llamado camino superviviente, que es el camino de menor coste que llega hasta ese estado en dicho instante de tiempo. El resto de los caminos que llegan al estado en el instante t nunca podrán dar lugar a una secuencia de menor coste en el instante T , ya que que las secuencias de estados que pueden ocurrir a partir de $t + 1$ sólo dependen del estado en t y no de los estados anteriores (proceso markoviano), por lo que el coste de cualquier secuencia que vaya de $t + 1$ a T sumado al coste del camino superviviente será menor o igual que el coste de esa misma secuencia más el coste de cualquier otro camino distinto del superviviente (cuyo coste es mayor o igual que el de éste).

Para ilustrar el funcionamiento del algoritmo de Viterbi se utiliza lo que se conoce como diagrama de trellis. Una diagrama de *trellis* muestra las transiciones que pueden ocurrir en el sistema junto con sus entradas y salidas correspondientes. Es una representación similar al diagrama de árbol, pero más compacta, en la que cada estado del sistema tiene asociada una secuencia de nodos (puntos), uno por cada instante de tiempo considerado en el *trellis*. Las secuencias de nodos de los diferentes estados se disponen verticalmente para formar una cuadrícula que permite representar las transiciones entre estados mediante flechas que unen dos nodos situados en instantes de tiempo diferentes. Cada flecha (transición) se etiqueta mediante una expresión de la forma x/y , donde x es la entrada que origina la transición e y es la salida de la misma. La Figura B.1 muestra un diagrama de *trellis* típico asociado a un sistema con cuatro estados, s_1, s_2, s_3 y s_4 , dos entradas, x_1 y x_2 , y cuatro salidas, y_1, y_2, y_3 e y_4 .

El *trellis* de la figura representa un sistema invariante en el tiempo, es decir, la salida del

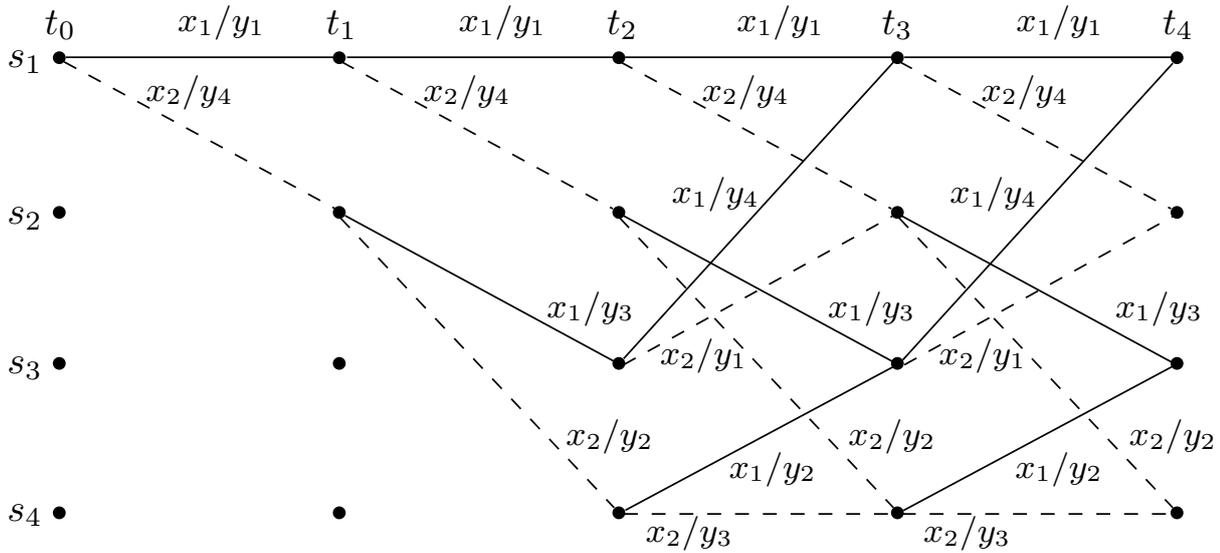


Figura B.1: Diagrama *trellis* típico.

sistema cuando se produce una determinada transición depende únicamente del estado origen y de la entrada (obsérvese que el *trellis* se repite a partir del instante t_3 y que la primera parte del mismo se dibuja únicamente para indicar que durante los primeros instantes de tiempo, y partiendo de un estado inicial dado, s_1 , no son posibles todas las transiciones). Sin embargo, la salida asociada a una determinada transición podría depender también del tiempo, es decir, de los instantes de tiempo entre los cuales tiene lugar. Es el caso en nuestro modelo de señal, en el que el canal cambia en cada instante.

En la sección 4.1.3 demostramos que el problema de optimización que plantea la detección MV en el modelo del Capítulo 3 tiene en último término la forma requerida por el VA. En efecto, el sistema dinámico en formato de espacio de estados que define dicho modelo de señal se puede representar por medio de un *trellis*. A la vista de la ecuación (3.1), la t -ésima observación, \mathbf{x}_t , depende de los símbolos transmitidos en el instante actual y en los $m - 1$ anteriores (debido a la memoria del canal). Entonces, podemos pensar en la secuencia $\mathbf{s}_{t-m+1:t-1}$ como en el estado del sistema en el instante $t - 1$ y en \mathbf{s}_t como en la entrada que hace que el sistema transite al estado $\mathbf{s}_{t-m+2:t}$ a la vez que genera la salida \mathbf{x}_t . Así, tenemos que los vectores de símbolos transmitidos definen la secuencia de estados por la que pasa el sistema y cada uno de ellos da lugar a una observación relacionada, de tal manera que se cumple una de las hipótesis de las que parte el algoritmo de Viterbi: cada estado tiene su observación correspondiente, o lo que es lo mismo, las secuencias están alineadas.

Explicaremos el funcionamiento del VA sobre un ejemplo basado en el modelo de señal del Capítulo 3. Vamos a considerar un sistema con una sola antena emisora ($N_t = 1$) y una sola antena receptora ($N_r = 1$) transmitiendo a través de un canal con memoria $m = 3$. En este caso, los vectores de observaciones, \mathbf{x}_t , y ruido, \mathbf{g}_t , que aparecen en la

ecuación (3.1) son en realidad escalares, y lo mismo ocurre con el t -ésimo vector de símbolos transmitidos, \mathbf{s}_t , al estar considerando una única antena emisora. Además, vamos a considerar una modulación BPSK (*Binary Phase-Shift Keying*), es decir, el alfabeto de los símbolos es $\mathcal{S} = \{-1, +1\}$. Sea entonces $s_{1:4} = [-1, +1, -1, -1]$ la secuencia de símbolos transmitidos y $x_{1:4} = [-1,66473, 0,13799, -0,87508, -1,17578]$ las observaciones que se obtienen cuando la varianza del ruido es $\sigma_g^2 = 0,01$. Supondremos, asimismo, que el canal variante con el tiempo es conocido, y que sus coeficientes durante los cuatro primeros instantes son:

$$\begin{aligned}\mathbf{H}_1 &= [0,2 \ 0,5 \ 1] \\ \mathbf{H}_2 &= [0,25 \ 0,45 \ 1,05] \\ \mathbf{H}_3 &= [0,22 \ 0,48 \ 1,1] \\ \mathbf{H}_4 &= [0,18 \ 0,40 \ 0,95]\end{aligned}$$

El *trellis* necesario para llevar a cabo el proceso de detección en este ejemplo es básicamente el que aparece en la Figura B.1, pero adecuando las entradas y las salidas de las transiciones a aquellas que determinan los parámetros concretos de nuestro sistema. Cada estado del *trellis* representa una de las posibles secuencias de $m - 1$ últimos vectores de símbolos transmitidos que, en este caso, son de longitud 1, por lo que cada estado almacena $m - 1 = 2$ símbolos. Entonces, el número de posibles estados es $|\mathcal{S}|^2 = 4$. Por otro lado, y puesto hay una única antena emisora, el número de posibles entradas es 2 (el cardinal del alfabeto \mathcal{S}). Así pues, tendremos un diagrama de *trellis* con 4 estados y 2 entradas, cuyas transiciones serán las mismas que las del *trellis* de la Figura B.1.

Cuando hay sólo dos posibles entradas es habitual indicar a cual de ellas se refiere una transición por el trazo de la rama (en lugar de hacerlo explícitamente en la etiqueta). En los *trellis* que ilustran el ejemplo dibujaremos con trazo continuo aquellas transiciones cuya entrada es -1 y con trazo discontinuo aquellas cuya entrada es $+1$. Además, por simplicidad, cada transición no se etiqueta con su salida correspondiente, sino con su coste (que es la diferencia entre la salida y el valor observado).

Vamos a comenzar la detección en el estado $[-1, -1]$, lo cual supone asumir que esos fueron los dos últimos símbolos transmitidos antes del comienzo de la trama, $s_{1:4}$, que estamos considerando. En el instante t_1 , podemos seguir en el estado t_0 o pasar al estado $[-1, +1]$, según el símbolo s_1 haya sido un -1 o un $+1$, respectivamente. En el primer caso, la salida sería $\mathbf{H}_1 [-1, -1, -1]^T = -1,70$ y el coste de la transición correspondiente $|x_1 - (-1,70)|^2 = 0,001244$ (en las figuras solo aparecerán indicadas dos cifras decimales). Si, por el contrario, el símbolo transmitido fuera $+1$, la salida sería $\mathbf{H}_1 [-1, -1, +1]^T = 0,30$ y el coste $|x_1 - 0,30|^2 = 3,86$. La Figura B.2 ilustra este primer paso del algoritmo, tras el cual tenemos un camino de coste 0,001244 que llega al estado $[-1, -1]$ y otro de coste 3,86 que llega a $[-1, +1]$. Puesto que son los únicos caminos que llegan a dichos estados, ambos serán caminos supervivientes.

Después de procesar la segunda observación (Figura B.3) tenemos un camino que llega a cada uno de los posibles estados. Cada uno de ellos constituye, pues, el mejor camino que llega

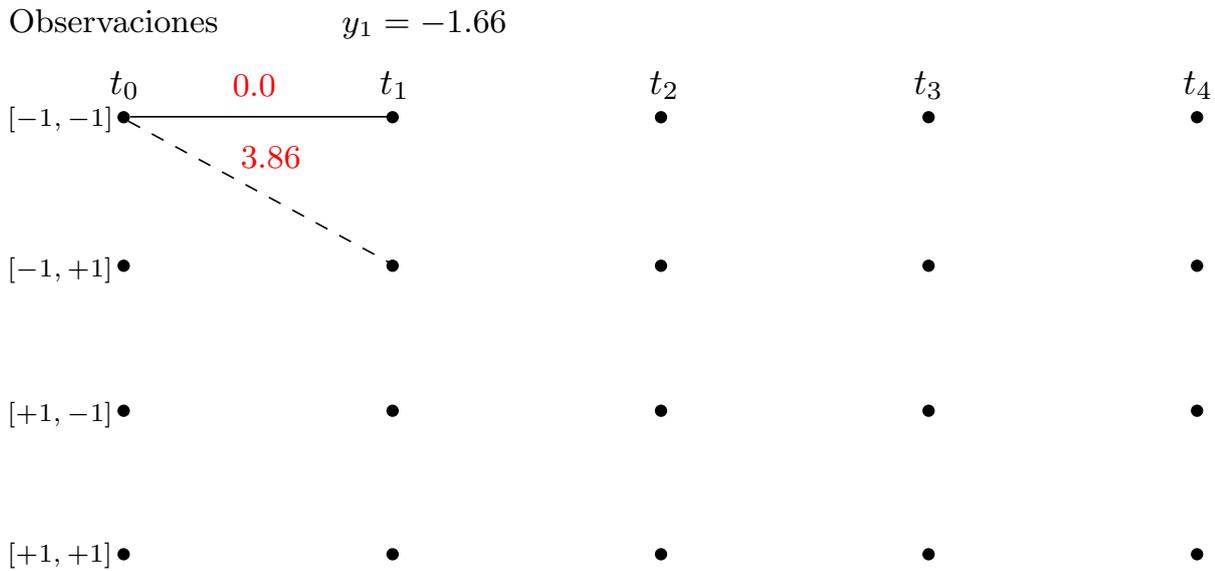


Figura B.2: Primera etapa del proceso de detección mediante el algoritmo de Viterbi.

a ese estado y todos son caminos supervivientes.

La Figura B.4 muestra el aspecto del *trellis* en el tercer instante de tiempo. A cada estado llegan ahora dos caminos:

- Al estado $[-1, -1]$ llegan los caminos
 - $[-1, -1] \xrightarrow{-1} [-1, -1] \xrightarrow{-1} [-1, -1] \xrightarrow{-1} [-1, -1]$. Coste: $0 + 3,56 + 0,86 = 4,42$
 - $[-1, -1] \xrightarrow{+1} [-1, +1] \xrightarrow{-1} [+1, -1] \xrightarrow{-1} [-1, -1]$. Coste: $3,86 + 0,98 + 0,24 = 5,08$
- Al estado $[-1, +1]$ llegan los caminos
 - $[-1, -1] \xrightarrow{-1} [-1, -1] \xrightarrow{-1} [-1, -1] \xrightarrow{+1} [-1, +1]$. Coste: $0 + 3,56 + 1,63 = 5,19$
 - $[-1, -1] \xrightarrow{+1} [-1, +1] \xrightarrow{-1} [+1, -1] \xrightarrow{+1} [-1, +1]$. Coste: $3,86 + 0,98 + 2,94 = 7,78$
- Al estado $[+1, -1]$ llegan los caminos
 - $[-1, -1] \xrightarrow{-1} [-1, -1] \xrightarrow{+1} [-1, +1] \xrightarrow{-1} [+1, -1]$. Coste: $0 + 0,04 + 0 = 0,04$
 - $[-1, -1] \xrightarrow{+1} [-1, +1] \xrightarrow{+1} [+1, +1] \xrightarrow{-1} [+1, -1]$. Coste: $3,86 + 1,24 + 0,23 = 5,33$
- Al estado $[+1, +1]$ llegan los caminos
 - $[-1, -1] \xrightarrow{-1} [-1, -1] \xrightarrow{+1} [-1, +1] \xrightarrow{+1} [+1, +1]$. Coste: $0 + 0,04 + 5 = 5,04$
 - $[-1, -1] \xrightarrow{+1} [-1, +1] \xrightarrow{+1} [+1, +1] \xrightarrow{+1} [+1, +1]$. Coste: $3,86 + 1,24 + 7,16 = 12,26$

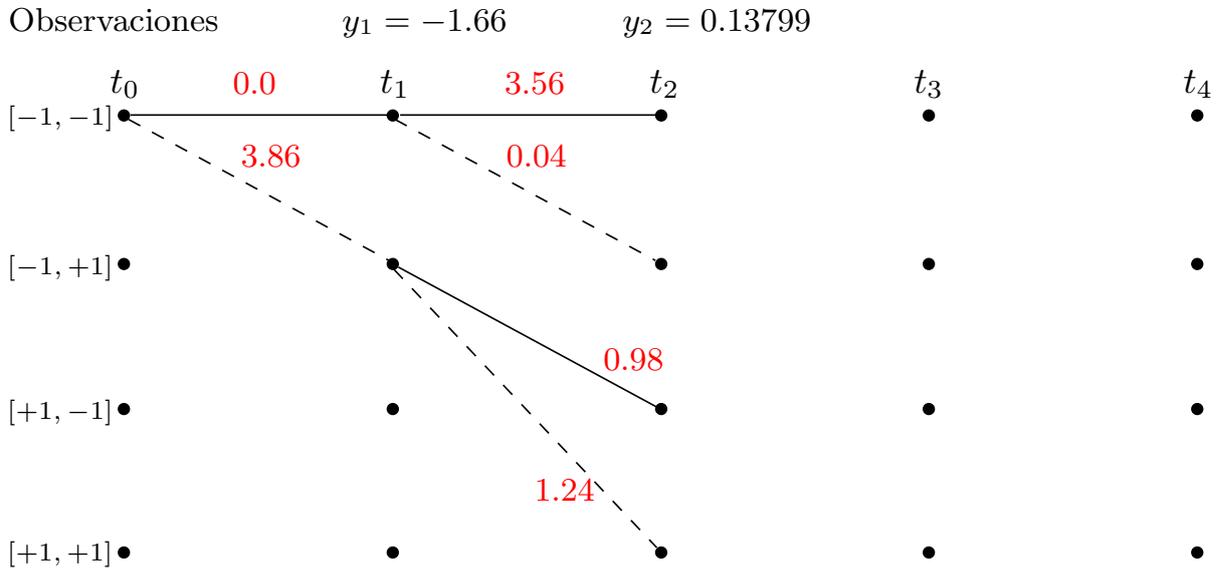


Figura B.3: Segunda etapa del proceso de detección mediante el algoritmo de Viterbi.

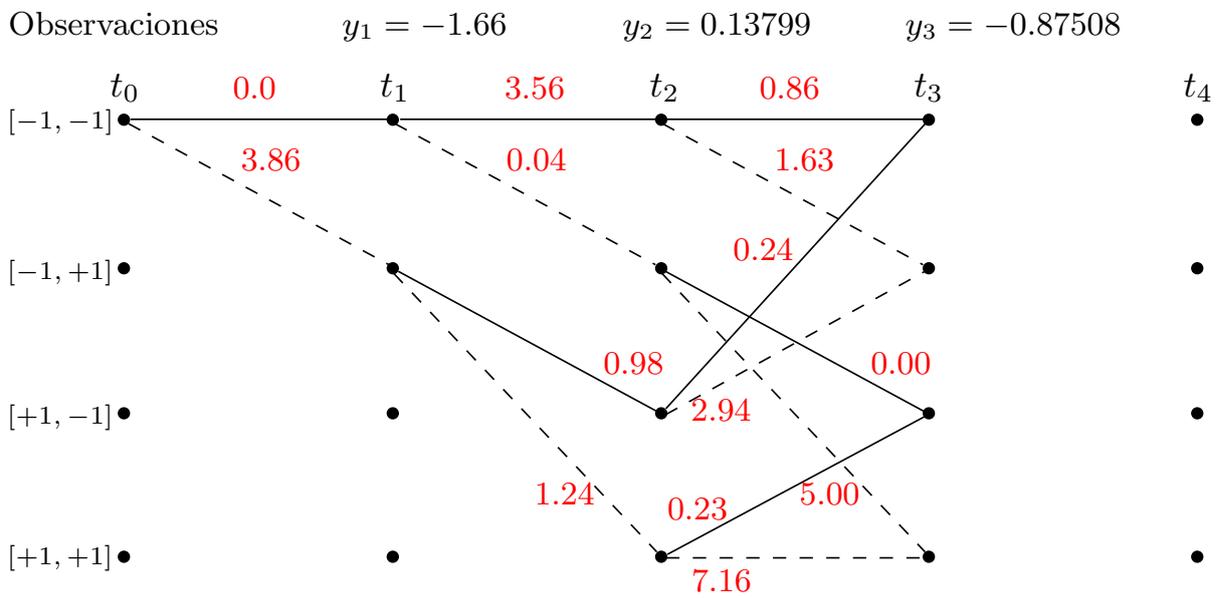


Figura B.4: Tercera etapa del proceso de detección mediante el algoritmo de Viterbi.

El camino superviviente (el de menor coste) es en todos los casos el primero, y el segundo camino que llega a cada uno de los estados puede ser descartado para dar lugar al *trellis* de la Figura B.5, en la que únicamente se muestran los caminos supervivientes hasta t_3 .

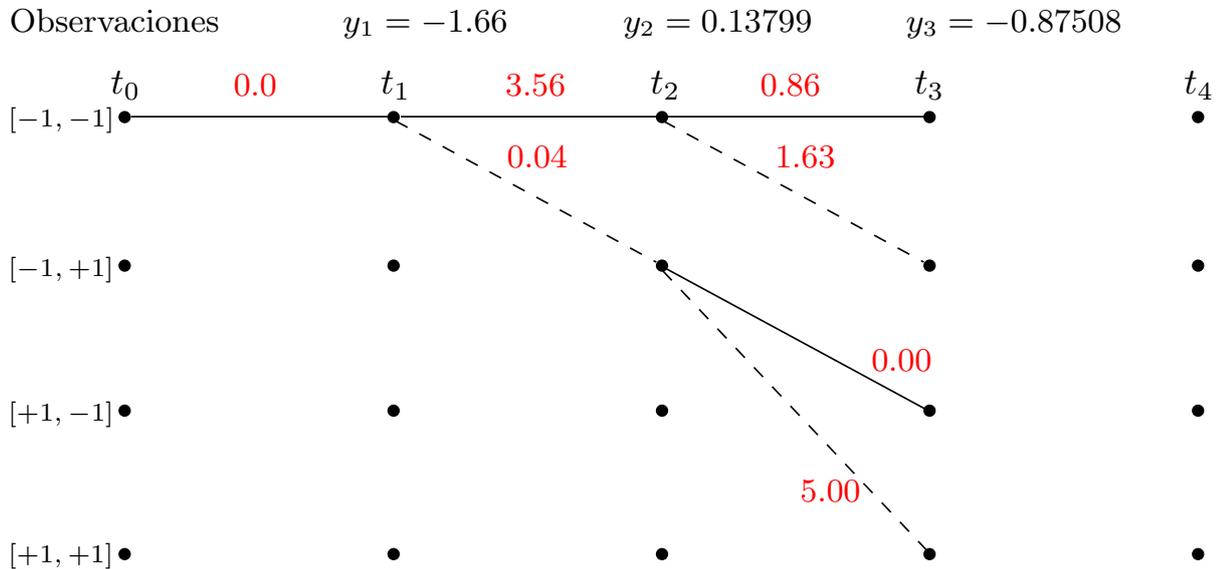


Figura B.5: Cuarta etapa del proceso de detección mediante el algoritmo de Viterbi.

En t_4 (Figura B.6) tenemos, una vez más, que a cada estado llegan dos caminos procedentes de los caminos supervivientes del instante de tiempo anterior, y a partir de los costes de éstos calcularemos los de aquellos. Si denotamos por \mathcal{C}_t^i al mejor camino que llega en el instante de tiempo t hasta el estado i ($\mathcal{C}_3^{[-1,-1]} = [-1, -1] \xrightarrow{-1} [-1, -1] \xrightarrow{-1} [-1, -1] \xrightarrow{-1} [-1, -1]$, por ejemplo), y por $c(\mathcal{C}_t^i)$ a su coste asociado, entonces:

- Al estado $[-1, -1]$ llegan los caminos
 - $\mathcal{C}_3^{[-1,-1]} \xrightarrow{-1} [-1, -1]$. Coste: $c(\mathcal{C}_3^{[-1,-1]}) + 0,13 = 4,42 + 0,13 = 4,55$
 - $\mathcal{C}_3^{[+1,-1]} \xrightarrow{-1} [-1, -1]$. Coste: $c(\mathcal{C}_3^{[+1,-1]}) + 0 = 0,04 + 0 = 0,04$
- Al estado $[-1, +1]$ llegan los caminos
 - $\mathcal{C}_3^{[-1,-1]} \xrightarrow{+1} [-1, +1]$. Coste: $c(\mathcal{C}_3^{[-1,-1]}) + 2,39 = 4,42 + 2,39 = 6,81$
 - $\mathcal{C}_3^{[+1,-1]} \xrightarrow{+1} [-1, +1]$. Coste: $c(\mathcal{C}_3^{[+1,-1]}) + 3,63 = 0,04 + 3,63 = 3,67$
- Al estado $[+1, -1]$ llegan los caminos
 - $\mathcal{C}_3^{[-1,+1]} \xrightarrow{-1} [+1, -1]$. Coste: $c(\mathcal{C}_3^{[-1,+1]}) + 0,20 = 5,19 + 0,20 = 5,39$
 - $\mathcal{C}_3^{[+1,+1]} \xrightarrow{-1} [+1, -1]$. Coste: $c(\mathcal{C}_3^{[+1,+1]}) + 0,65 = 5,04 + 0,65 = 5,69$

■ Al estado $[+1, +1]$ llegan los caminos

- $\mathcal{C}_3^{[-1,+1]} \xrightarrow{+1} [+1, +1]$. Coste: $c(\mathcal{C}_3^{[-1,+1]}) + 5,50 = 5,19 + 5,50 = 10,69$
- $\mathcal{C}_3^{[+1,+1]} \xrightarrow{+1} [+1, +1]$. Coste: $c(\mathcal{C}_3^{[+1,+1]}) + 7,32 = 5,04 + 7,32 = 12,36$

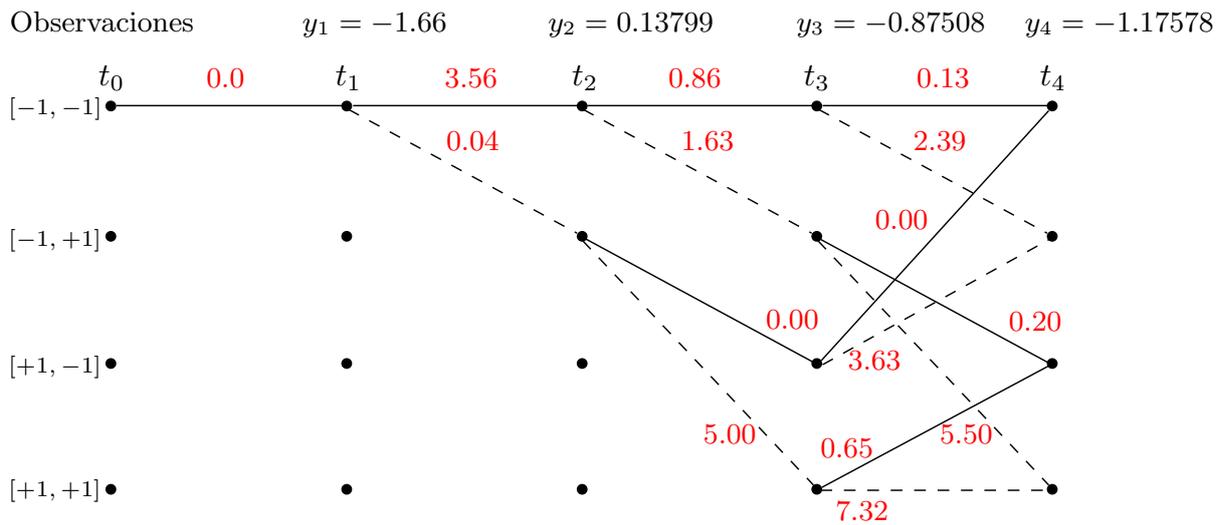


Figura B.6: Quinta etapa del proceso de detección mediante el algoritmo de Viterbi.

En el instante t_4 el mejor camino global (el camino superviviente de menor coste) es el segundo camino que llega al estado $[-1, -1]$. Las entradas que dan lugar a ese camino $(-1, +1, -1, -1)$ constituyen la secuencia más probablemente transmitida, que en este caso se corresponde con la secuencia efectivamente transmitida.

Apéndice C

Estimación del canal mediante RLS

Consideremos el modelo MIMO genérico dado por la ecuación

$$\mathbf{y}_t = \mathbf{H}_t \mathbf{s}_t + \mathbf{g}_t. \quad (\text{C.1})$$

Dadas las observaciones $\mathbf{y}_{0:t}$ y los símbolos $\mathbf{s}_{0:t}$, el estimador mínimos cuadrados (LS, del inglés *Least Squares*) para la matriz de canal en el instante t es

$$\mathbf{H}_t = \arg \min_{\mathbf{H}} \{\Upsilon(\mathbf{H})\} = \arg \min_{\mathbf{H}} \left\{ \sum_{k=0}^t \lambda^{t-k} \|\mathbf{y}_k - \mathbf{H} \mathbf{s}_k\|^2 \right\} \quad (\text{C.2})$$

donde λ es un factor de olvido. Comenzaremos desarrollando $\Upsilon(\mathbf{H})$ como

$$\begin{aligned} \Upsilon(\mathbf{H}) &= \sum_{k=0}^t \lambda^{t-k} \|\mathbf{y}_k - \mathbf{H} \mathbf{s}_k\|^2 = \sum_{k=0}^t \lambda^{t-k} \text{Tr} \{ (\mathbf{y}_k - \mathbf{H} \mathbf{s}_k)(\mathbf{y}_k - \mathbf{H} \mathbf{s}_k)^\top \} \\ &= \sum_{k=0}^t \lambda^{t-k} \text{Tr} \{ (\mathbf{y}_k - \mathbf{H} \mathbf{s}_k)(\mathbf{y}_k^\top - \mathbf{s}_k^\top \mathbf{H}^\top) \} = \sum_{k=0}^t \lambda^{t-k} \text{Tr} \{ \mathbf{y}_k \mathbf{y}_k^\top - \\ &\quad - \mathbf{y}_k \mathbf{s}_k^\top \mathbf{H}^\top - \mathbf{H} \mathbf{s}_k \mathbf{y}_k^\top + \mathbf{H} \mathbf{s}_k \mathbf{s}_k^\top \mathbf{H}^\top \} = \sum_{k=0}^t \lambda^{t-k} [\text{Tr} \{ \mathbf{y}_k \mathbf{y}_k^\top \} - \\ &\quad - \text{Tr} \{ \mathbf{y}_k \mathbf{s}_k^\top \mathbf{H}^\top \} - \text{Tr} \{ \mathbf{H} \mathbf{s}_k \mathbf{y}_k^\top \} + \text{Tr} \{ \mathbf{H} \mathbf{s}_k \mathbf{s}_k^\top \mathbf{H}^\top \}]. \end{aligned} \quad (\text{C.3})$$

Para hallar la matriz de canal, \mathbf{H} , que minimiza la función escalar de variable matricial $\Upsilon(\mathbf{H})$, vamos a calcular las derivadas parciales de ésta respecto de los elementos de aquella:

$$\begin{aligned} \frac{\partial}{\partial \mathbf{H}} \Upsilon(\mathbf{H}) &= \sum_{k=0}^t \lambda^{t-k} \left[\frac{\partial}{\partial \mathbf{H}} \text{Tr} \{ \mathbf{y}_k \mathbf{y}_k^\top \} - \frac{\partial}{\partial \mathbf{H}} \text{Tr} \{ \mathbf{y}_k \mathbf{s}_k^\top \mathbf{H}^\top \} - \frac{\partial}{\partial \mathbf{H}} \text{Tr} \{ \mathbf{H} \mathbf{s}_k \mathbf{y}_k^\top \} + \right. \\ &\quad \left. + \frac{\partial}{\partial \mathbf{H}} \text{Tr} \{ \mathbf{H} \mathbf{s}_k \mathbf{s}_k^\top \mathbf{H}^\top \} \right]. \end{aligned} \quad (\text{C.4})$$

Utilizando las reglas del Apéndice F podemos desarrollar la ecuación (C.4) como

$$\begin{aligned}\frac{\partial}{\partial \mathbf{H}} \Upsilon(\mathbf{H}) &= \sum_{k=0}^t \lambda^{t-k} [-\mathbf{s}_k \mathbf{y}_k^\top - \mathbf{s}_k \mathbf{y}_k^\top + \mathbf{s}_k \mathbf{s}_k^\top \mathbf{H}^\top + \mathbf{s}_k \mathbf{s}_k^\top \mathbf{H}^\top] \\ &= -2 \sum_{k=0}^t \lambda^{t-k} \mathbf{s}_k \mathbf{y}_k^\top + 2 \sum_{k=0}^t \lambda^{t-k} \mathbf{s}_k \mathbf{s}_k^\top \mathbf{H}^\top.\end{aligned}\quad (\text{C.5})$$

e igualando a cero resulta

$$\begin{aligned}0 &= -2 \sum_{k=0}^t \lambda^{t-k} \mathbf{s}_k \mathbf{y}_k^\top + 2 \sum_{k=0}^t \lambda^{t-k} \mathbf{s}_k \mathbf{s}_k^\top \mathbf{H}^\top. \\ &\Downarrow \\ \sum_{k=0}^t \lambda^{t-k} \mathbf{s}_k \mathbf{s}_k^\top \mathbf{H}^\top &= \sum_{k=0}^t \lambda^{t-k} \mathbf{s}_k \mathbf{y}_k^\top \\ &\Downarrow \\ \mathbf{H}^\top &= \left(\sum_{k=0}^t \lambda^{t-k} \mathbf{s}_k \mathbf{s}_k^\top \right)^{-1} \sum_{k=0}^t \lambda^{t-k} \mathbf{s}_k \mathbf{y}_k^\top. \\ &= \mathbf{R}_t^{-1} \mathbf{P}_t\end{aligned}\quad (\text{C.6})$$

Los factores $\mathbf{R}_t = \sum_{k=0}^t \lambda^{t-k} \mathbf{s}_k \mathbf{s}_k^\top$ y $\mathbf{P}_t = \sum_{k=0}^t \lambda^{t-k} \mathbf{s}_k \mathbf{y}_k^\top$ que resultan en la ecuación (C.6) son, respectivamente, la matriz de correlación de los símbolos y la matriz de correlación cruzada. Podemos escribirlos como

$$\mathbf{R}_t = \sum_{k=0}^t \lambda^{t-k} \mathbf{s}_k \mathbf{s}_k^\top = \sum_{k=0}^{t-1} \lambda^{t-k-1} \mathbf{s}_k \mathbf{s}_k^\top + \mathbf{s}_t \mathbf{s}_t^\top = \lambda \mathbf{R}_{t-1} + \mathbf{s}_t \mathbf{s}_t^\top \quad (\text{C.7})$$

$$\mathbf{P}_t = \sum_{k=0}^t \lambda^{t-k} \mathbf{s}_k \mathbf{y}_k^\top = \sum_{k=0}^{t-1} \lambda^{t-k-1} \mathbf{s}_k \mathbf{y}_k^\top + \mathbf{s}_t \mathbf{y}_t^\top = \lambda \mathbf{P}_{t-1} + \mathbf{s}_t \mathbf{y}_t^\top. \quad (\text{C.8})$$

Para evitar invertir \mathbf{R}_t podemos utilizar el lema de inversión, que afirma que si \mathbf{A} y \mathbf{B} son dos matrices definidas positivas tales que

$$\mathbf{A} = \mathbf{B}^{-1} + \mathbf{C} \mathbf{D}^{-1} \mathbf{C}^\top, \quad (\text{C.9})$$

donde \mathbf{C} y \mathbf{D} son matrices de dimensiones compatibles y definidas positivas, entonces

$$\mathbf{A}^{-1} = \mathbf{B} - \mathbf{B} \mathbf{C} (\mathbf{D} + \mathbf{C}^\top \mathbf{B} \mathbf{C})^{-1} \mathbf{C}^\top \mathbf{B}. \quad (\text{C.10})$$

Definiendo

$$\begin{aligned}\mathbf{A} &= \mathbf{R}_t \\ \mathbf{B}^{-1} &= \lambda \mathbf{R}_{t-1} \Rightarrow \mathbf{B} = \lambda^{-1} \mathbf{R}_{t-1}^{-1} \\ \mathbf{C} &= \mathbf{s}_t \\ \mathbf{D} &= 1\end{aligned}$$

tenemos que

$$\begin{aligned}\mathbf{R}_t^{-1} &= \lambda^{-1}\mathbf{R}_{t-1}^{-1} - \lambda^{-1}\mathbf{R}_{t-1}^{-1}\mathbf{s}_t \left(1 + \mathbf{s}_t^\top \lambda^{-1}\mathbf{R}_{t-1}^{-1}\mathbf{s}_t\right)^{-1} \mathbf{s}_t^\top \lambda^{-1}\mathbf{R}_{t-1}^{-1} \\ &= \lambda^{-1}\mathbf{R}_{t-1}^{-1} - \frac{\lambda^{-1}\mathbf{R}_{t-1}^{-1}\mathbf{s}_t\mathbf{s}_t^\top \lambda^{-1}\mathbf{R}_{t-1}^{-1}}{1 + \mathbf{s}_t^\top \lambda^{-1}\mathbf{R}_{t-1}^{-1}\mathbf{s}_t} = \lambda^{-1}\mathbf{R}_{t-1}^{-1} - \mathbf{g}_t\mathbf{s}_t^\top \lambda^{-1}\mathbf{R}_{t-1}^{-1}\end{aligned}\quad (\text{C.11})$$

donde

$$\mathbf{g}_t = \frac{\lambda^{-1}\mathbf{R}_{t-1}^{-1}\mathbf{s}_t}{1 + \mathbf{s}_t^\top \lambda^{-1}\mathbf{R}_{t-1}^{-1}\mathbf{s}_t} \quad (\text{C.12})$$

es un vector de ganancia.

La ecuación (C.12) se puede reescribir como

$$\begin{aligned}\mathbf{g}_t &= \lambda^{-1}\mathbf{R}_{t-1}^{-1}\mathbf{s}_t - \mathbf{g}_t\mathbf{s}_t^\top \lambda^{-1}\mathbf{R}_{t-1}^{-1}\mathbf{s}_t \\ &= \left(\lambda^{-1}\mathbf{R}_{t-1}^{-1} - \mathbf{g}_t\mathbf{s}_t^\top \lambda^{-1}\mathbf{R}_{t-1}^{-1}\right)\mathbf{s}_t\end{aligned}\quad (\text{C.13})$$

y sustituyendo (C.11) en (C.12) resulta

$$\mathbf{g}_t = \mathbf{R}_t^{-1}\mathbf{s}_t. \quad (\text{C.14})$$

Volviendo ahora a la ecuación (C.6)

$$\begin{aligned}\mathbf{H}_t^\top &= \mathbf{R}_t^{-1}\mathbf{P}_t \\ &= \mathbf{R}_t^{-1}(\lambda\mathbf{P}_{t-1} + \mathbf{s}_t\mathbf{y}_t^\top)\end{aligned}\quad (\text{C.15a})$$

$$\begin{aligned}&= \lambda\mathbf{R}_t^{-1}\mathbf{P}_{t-1} + \mathbf{R}_t^{-1}\mathbf{s}_t\mathbf{y}_t^\top \\ &= \lambda\left(\lambda^{-1}\mathbf{R}_{t-1}^{-1} - \mathbf{g}_t\mathbf{s}_t^\top \lambda^{-1}\mathbf{R}_{t-1}^{-1}\right)\mathbf{P}_{t-1} + \mathbf{R}_t^{-1}\mathbf{s}_t\mathbf{y}_t^\top\end{aligned}\quad (\text{C.15b})$$

$$\begin{aligned}&= \mathbf{R}_{t-1}^{-1}\mathbf{P}_{t-1} - \mathbf{g}_t\mathbf{s}_t^\top \mathbf{R}_{t-1}^{-1}\mathbf{P}_{t-1} + \mathbf{R}_t^{-1}\mathbf{s}_t\mathbf{y}_t^\top \\ &= \mathbf{H}_{t-1}^\top - \mathbf{g}_t\mathbf{s}_t^\top \mathbf{H}_{t-1}^\top + \mathbf{R}_t^{-1}\mathbf{s}_t\mathbf{y}_t^\top\end{aligned}\quad (\text{C.15c})$$

$$= \mathbf{H}_{t-1}^\top - \mathbf{g}_t\mathbf{s}_t^\top \mathbf{H}_{t-1}^\top + \mathbf{g}_t\mathbf{y}_t^\top \quad (\text{C.15d})$$

$$= \mathbf{H}_{t-1}^\top + \mathbf{g}_t(\mathbf{y}_t^\top - \mathbf{s}_t^\top \mathbf{H}_{t-1}^\top) \quad (\text{C.15e})$$

donde las ecuaciones (C.8), (C.11), (C.6) y (C.14) se utilizan en (C.15a), (C.15b), (C.15c) y (C.15d), respectivamente. El Pseudocódigo 12 resume el algoritmo que finalmente resulta.

Pseudocódigo 12 Estimación del canal mediante RLS

inicializar \mathbf{R}_0^{-1} a la matriz identidad

inicializar la estimación del canal \mathbf{H}_0^\top

para cada nuevo vector de símbolos \mathbf{s}_t

 calcular el vector de ganancia \mathbf{g}_t por medio de la ecuación (C.12)

 obtener la estimación del canal en el instante t utilizando (C.15e)

 actualizar la matriz de correlación de los símbolos usando (C.11)

Apéndice D

Cálculo de $p(\mathbf{y}_t | \mathbf{s}_{0:t}, \mathbf{y}_{0:t-1})$

Si \mathbf{h}_t es un vector $N_r N_t m \times 1$ construido apilando las $N_t m$ columnas de la matriz \mathbf{H}_t (si $\mathbf{H}_t = [\mathbf{h}_t(1), \mathbf{h}_t(2), \dots, \mathbf{h}_t(N_t m)]$ donde $\mathbf{h}_t(i)$ es la i -ésima columna de \mathbf{H}_t , entonces $\mathbf{h}_t = [\mathbf{h}_t^\top(1), \mathbf{h}_t^\top(2), \dots, \mathbf{h}_t^\top(N_t m)]^\top$) y agrupamos en una matriz

$$\mathbf{S}_t = [s_{t-m+1}(1)\mathbf{I}_{N_r}, \dots, s_{t-m+1}(N_t)\mathbf{I}_{N_r}, \dots, s_t(1)\mathbf{I}_{N_r}, \dots, s_t(N_t)\mathbf{I}_{N_r}], \quad (\text{D.1})$$

de tamaño $N_r \times N_r N_t m$ los símbolos implicados en la t -ésima observación, podemos reescribir las ecuaciones (3.2) y (3.6) que definen nuestro modelo de señal como

$$\mathbf{y}_t = \mathbf{S}_t \mathbf{h}_t + \mathbf{g}_t \quad (\text{D.2})$$

y

$$\mathbf{h}_t = \alpha \mathbf{h}_{t-1} + \mathbf{v}_t, \quad (\text{D.3})$$

respectivamente, donde \mathbf{v}_t es un vector aleatorio gaussiano de media cero y varianza σ_v^2 .

Supongamos ahora que conocemos la f.d.p. *a priori* del canal, $p(\mathbf{h}_{-1})$, y que ésta es gaussiana. Entonces, la f.d.p. *a posteriori* $p(\mathbf{h}_t | \mathbf{y}_{0:t}, \mathbf{s}_{0:t})$ también es gaussiana y se puede calcular recursivamente mediante el KF [82]. Si

$$p(\mathbf{h}_{t-1} | \mathbf{y}_{0:t-1}, \mathbf{s}_{0:t-1}) = \mathcal{N}(\mathbf{h}_{t-1} | \bar{\mathbf{h}}_{t-1|t-1}, \mathbf{C}_{t-1|t-1}), \quad (\text{D.4})$$

entonces $p(\mathbf{h}_t | \mathbf{y}_{0:t-1}, \mathbf{s}_{0:t-1}) = \mathcal{N}(\mathbf{h}_t | \bar{\mathbf{h}}_{t|t-1}, \mathbf{C}_{t|t-1})$, donde

$$\bar{\mathbf{h}}_{t|t-1} = \alpha \bar{\mathbf{h}}_{t-1|t-1} \quad (\text{D.5})$$

$$\mathbf{C}_{t|t-1} = \alpha^2 \mathbf{C}_{t-1|t-1} + \sigma_v^2 \mathbf{I}_{N_r N_t m}, \quad (\text{D.6})$$

y $p(\mathbf{h}_t | \mathbf{y}_{0:t}, \mathbf{s}_{0:t}) = \mathcal{N}(\mathbf{h}_t | \bar{\mathbf{h}}_{t|t}, \mathbf{C}_{t|t})$, donde

$$\bar{\mathbf{h}}_{t|t} = \bar{\mathbf{h}}_{t|t-1} + \mathbf{K}_t (\mathbf{y}_t - \mathbf{S}_t \bar{\mathbf{h}}_{t|t-1}) \quad (\text{D.7})$$

$$\mathbf{C}_{t|t} = (\mathbf{I}_{N_r N_t m} - \mathbf{K}_t \mathbf{S}_t) \mathbf{C}_{t|t-1}, \quad (\text{D.8})$$

siendo $\mathbf{K}_t = \mathbf{C}_{t|t-1} \mathbf{S}_t^{(i)\top} \left(\mathbf{S}_t \mathbf{C}_{t|t-1} \mathbf{S}_t^{(i)\top} + \sigma_{g,t}^2 \mathbf{I}_{N_r} \right)^{-1}$ la matriz de ganancia de Kalman.

El vector $\bar{\mathbf{h}}_{t|t-1}$ y la matriz $\mathbf{C}_{t|t-1}$ son, respectivamente, la media y la covarianza predictivas de Kalman en el instante t , es decir, la estimación que el KF hace de la media y de la covarianza de \mathbf{h}_t utilizando únicamente las observaciones hasta el instante $t - 1$. Las estimaciones que obtiene de dichos parámetros cuando emplea también la t -ésima observación son la media y la covarianza filtradas de Kalman, que se denotan por $\bar{\mathbf{h}}_{t|t}$ y $\mathbf{C}_{t|t}$, respectivamente.

Si aplicamos ahora el teorema de la probabilidad total sobre $p(\mathbf{y}_t | \mathbf{s}_{0:t}, \mathbf{y}_{0:t-1})$ tenemos

$$\begin{aligned} p(\mathbf{y}_t | \mathbf{s}_{0:t}, \mathbf{y}_{0:t-1}) &= \int p(\mathbf{y}_t, \mathbf{h}_t | \mathbf{s}_{0:t}, \mathbf{y}_{0:t-1}) d\mathbf{h}_t \\ &= \int p(\mathbf{y}_t | \mathbf{h}_t, \mathbf{s}_{0:t}, \mathbf{y}_{0:t-1}) p(\mathbf{h}_t | \mathbf{s}_{0:t}, \mathbf{y}_{0:t-1}) d\mathbf{h}_t \\ &= \int p(\mathbf{y}_t | \mathbf{h}_t, \mathbf{S}_t) p(\mathbf{h}_t | \mathbf{y}_{0:t-1}, \mathbf{s}_{0:t-1}) d\mathbf{h}_t \end{aligned} \quad (\text{D.9})$$

donde se ha utilizado que $p(\mathbf{h}_t | \mathbf{s}_{0:t}, \mathbf{y}_{0:t-1}) = p(\mathbf{h}_t | \mathbf{s}_{0:t-1}, \mathbf{y}_{0:t-1})$, ya que los símbolos transmitidos en el instante t no aportan ninguna información acerca del canal en dicho instante si no se conoce la observación \mathbf{y}_t .

Dado el canal, \mathbf{h}_t , y matriz con los símbolos que intervienen en la t -ésima observación, \mathbf{S}_t , el vector de observaciones \mathbf{y}_t es gaussiano de media $\mathbf{S}_t \mathbf{h}_t$ y covarianza $\sigma_{g,t}^2 \mathbf{I}_{N_r}$. Así pues, $p(\mathbf{y}_t | \mathbf{s}_{0:t}, \mathbf{y}_{0:t-1})$ es la integral del producto de dos densidades gaussianas (cuya media y covarianza conocemos) y, por tanto, también es gaussiana. Sustituyendo las probabilidades de la ecuación (D.9) por la expresión de su f.d.p. se puede llegar a:

$$\begin{aligned} p(\mathbf{y}_t | \mathbf{s}_{0:t}, \mathbf{y}_{0:t-1}) &= \frac{|\mathbf{B}^{-1}|^{1/2}}{(2\pi\sigma_{g,t}^2)^{N_r/2} |\mathbf{C}_{t|t-1}|^{1/2}} \exp \left\{ -\frac{1}{2} \left[\mathbf{y}_t^\top (\sigma_{g,t}^2 \mathbf{I}_{N_r})^{-1} \mathbf{y}_t + \right. \right. \\ &\quad \left. \left. + \bar{\mathbf{h}}_{t|t-1}^\top \mathbf{C}_{t|t-1}^{-1} \bar{\mathbf{h}}_{t|t-1} - \mathbf{a}^H \mathbf{B}^{-1} \mathbf{a} \right] \right\} \end{aligned} \quad (\text{D.10})$$

donde $\mathbf{B} = \mathbf{S}_t^\top (\sigma_{g,t}^2 \mathbf{I}_{N_r})^{-1} \mathbf{S}_t + \mathbf{C}_{t|t-1}^{-1}$ y $\mathbf{a} = \mathbf{S}_t^\top (\sigma_{g,t}^2 \mathbf{I}_{N_r})^{-1} \mathbf{y}_t + \mathbf{C}_{t|t-1}^{-1} \bar{\mathbf{h}}_{t|t-1}$. Esta fórmula constituye una expresión cerrada para $p(\mathbf{y}_t | \mathbf{s}_{0:t}, \mathbf{y}_{0:t-1})$ que depende de la media y covarianza predictivas calculadas por el KF. Manipulando (D.10) se puede demostrar que $p(\mathbf{y}_t | \mathbf{s}_{0:t}, \mathbf{y}_{0:t-1})$ es gaussiana [76].

Apéndice E

Aproximación de la esperanza de la matriz de correlación de las observaciones

Sea un sistema de comunicaciones MIMO genérico cuyas observaciones vienen dadas por la ecuación

$$\mathbf{y}_t = \mathbf{H}\mathbf{s}_t + \mathbf{g}_t \quad (\text{E.1})$$

donde \mathbf{y}_t es el vector de observaciones, \mathbf{s}_t el de símbolos transmitidos, \mathbf{H} la matriz de canal y \mathbf{g}_t el ruido aditivo blanco gaussiano. La esperanza de la matriz de correlación de las observaciones viene dada, tal como se explicó en el apéndice A, por

$$\mathbf{R} = \mathbb{E} [\mathbf{y}_t \mathbf{y}_t^\top] = \mathbf{H} \sigma_s^2 \mathbf{H}^\top + \sigma_g^2 \mathbf{I}. \quad (\text{E.2})$$

Si disponemos de un conjunto de observaciones, $\mathbf{y}_1, \mathbf{y}_1, \dots, \mathbf{y}_t$, podemos aproximarla utilizando la media muestral

$$\mathbf{R}_{y,t} = \mathbb{E} [\mathbf{y}_t \mathbf{y}_t^\top] = \frac{1}{t} \sum_{k=1}^t \mathbf{y}_k \mathbf{y}_k^\top. \quad (\text{E.3})$$

Sin embargo, a la vista de la ecuación (E.2), si la matriz de canal \mathbf{H} varía con el tiempo, entonces también lo hará \mathbf{R} . Así pues, si queremos calcular la esperanza de la matriz de correlación de las observaciones en el instante t , es deseable tener más en cuenta las observaciones t -ésima e inmediatamente anteriores (obtenidas con valores más recientes de la matriz de canal) que las primeras. Para ello, en lugar de aproximar \mathbf{R} por la media de la ecuación (E.3), vamos a hacerlo utilizando una media ponderada:

$$\mathbf{R}_{y,t} = \mathbb{E} [\mathbf{y}_t \mathbf{y}_t^\top] = \frac{1}{\Lambda_t} \sum_{k=1}^t \lambda^{t-k} \mathbf{y}_k \mathbf{y}_k^\top \quad (\text{E.4})$$

donde λ es un factor de olvido y

$$\Lambda_t = \sum_{k=1}^t \lambda^k = \frac{\lambda^{t+1} - 1}{\lambda - 1}. \quad (\text{E.5})$$

La segunda igualdad en (E.5) es debida a que Λ_t es la suma de los elementos de una progresión geométrica de razón λ .

Desarrollando entonces el sumatorio en (E.4) resulta

$$\begin{aligned}
\mathbf{R}_{y,t} &= \mathbb{E} [\mathbf{y}_t \mathbf{y}_t^\top] = \frac{1}{\Lambda_t} \sum_{k=1}^t \lambda^{t-k} \mathbf{y}_k \mathbf{y}_k^\top = \frac{\lambda}{\Lambda_t} \sum_{k=1}^t \lambda^{(t-1)-k} \mathbf{y}_k \mathbf{y}_k^\top \\
&= \frac{\lambda}{\Lambda_t} \left(\sum_{k=1}^{t-1} \lambda^{(t-1)-k} \mathbf{y}_k \mathbf{y}_k^\top + \lambda^{-1} \mathbf{y}_t \mathbf{y}_t^\top \right) = \frac{\lambda}{\Lambda_t} \sum_{k=1}^{t-1} \lambda^{(t-1)-k} \mathbf{y}_k \mathbf{y}_k^\top + \frac{1}{\Lambda_t} \mathbf{y}_t \mathbf{y}_t^\top \\
&= \frac{\lambda \Lambda_{t-1}}{\Lambda_t} \left(\frac{1}{\Lambda_{t-1}} \sum_{k=1}^{t-1} \lambda^{(t-1)-k} \mathbf{y}_k \mathbf{y}_k^\top \right) + \frac{1}{\Lambda_t} \mathbf{y}_t \mathbf{y}_t^\top = \frac{\lambda \Lambda_{t-1}}{\Lambda_t} \mathbf{R}_{y,t-1} + \frac{1}{\Lambda_t} \mathbf{y}_t \mathbf{y}_t^\top. \quad (\text{E.6})
\end{aligned}$$

y definiendo

$$\begin{aligned}
\mathbf{A} &= \mathbf{R}_{y,t} \\
\mathbf{B}^{-1} &= \frac{\lambda \Lambda_{t-1}}{\Lambda_t} \mathbf{R}_{y,t-1} \Rightarrow \mathbf{B} = \frac{\Lambda_t}{\lambda \Lambda_{t-1}} \mathbf{R}_{y,t-1}^{-1} \\
\mathbf{C} &= \mathbf{y}_t \\
\mathbf{D}^{-1} &= \frac{1}{\Lambda_t} \Rightarrow \mathbf{D} = \Lambda_t
\end{aligned}$$

podemos utilizar el lema de inversión dado por las ecuaciones (C.9) y (C.10) del Apéndice C para calcular la inversa de $\mathbf{R}_{y,t}$ como

$$\begin{aligned}
\mathbf{R}_{y,t}^{-1} &= \frac{\Lambda_t}{\lambda \Lambda_{t-1}} \mathbf{R}_{y,t-1}^{-1} - \frac{\Lambda_t}{\lambda \Lambda_{t-1}} \mathbf{R}_{y,t-1}^{-1} \mathbf{y}_t \left(\Lambda_t + \mathbf{y}_t^\top \frac{\Lambda_t}{\lambda \Lambda_{t-1}} \mathbf{R}_{y,t-1}^{-1} \mathbf{y}_t \right)^{-1} \mathbf{y}_t^\top \frac{\Lambda_t}{\lambda \Lambda_{t-1}} \mathbf{R}_{y,t-1}^{-1} \\
&= \frac{\Lambda_t}{\lambda \Lambda_{t-1}} \left(\mathbf{R}_{y,t-1}^{-1} - \frac{\mathbf{R}_{y,t-1}^{-1} \mathbf{y}_t \mathbf{y}_t^\top \frac{\Lambda_t}{\lambda \Lambda_{t-1}} \mathbf{R}_{y,t-1}^{-1}}{\Lambda_t + \mathbf{y}_t^\top \frac{\Lambda_t}{\lambda \Lambda_{t-1}} \mathbf{R}_{y,t-1}^{-1} \mathbf{y}_t} \right). \quad (\text{E.7})
\end{aligned}$$

Agrupando finalmente algunos de los términos de (E.7) en un vector de ganancia

$$\mathbf{g}_t = \frac{\mathbf{R}_{y,t-1}^{-1} \mathbf{y}_t}{\Lambda_t + \mathbf{y}_t^\top \frac{\Lambda_t}{\lambda \Lambda_{t-1}} \mathbf{R}_{y,t-1}^{-1} \mathbf{y}_t} \quad (\text{E.8})$$

resulta

$$\mathbf{R}_{y,t}^{-1} = \frac{\Lambda_t}{\lambda \Lambda_{t-1}} \left(\mathbf{R}_{y,t-1}^{-1} - \mathbf{g}_t \mathbf{y}_t^\top \frac{\Lambda_t}{\lambda \Lambda_{t-1}} \mathbf{R}_{y,t-1}^{-1} \right) = \frac{\Lambda_t}{\lambda \Lambda_{t-1}} \left(\mathbf{I} - \mathbf{g}_t \mathbf{y}_t^\top \frac{\Lambda_t}{\lambda \Lambda_{t-1}} \right) \mathbf{R}_{y,t-1}^{-1}. \quad (\text{E.9})$$

Apéndice F

Derivada de una traza respecto a una matriz

En este apéndice se enumeran algunas relaciones útiles para el cálculo de la matriz de derivadas parciales de una función escalar respecto a una matriz \mathbf{X} . Puesto que habitualmente nos encontramos con problemas de optimización cuadráticos en el parámetro de interés, los resultados mostrados se refieren a funciones escalares obtenidas como la traza de un producto de matrices lineal o cuadrático en \mathbf{X} . Todas las relaciones son fáciles de demostrar y verificar.

En cuanto a la notación utilizada, $\frac{\partial}{\partial \mathbf{X}}(f)$ es una matriz cuyos coeficientes son las derivadas parciales de la función escalar f respecto de los correspondientes elementos en \mathbf{X} . Es decir, si denotamos por $[\mathbf{X}]_{i,j}$ al elemento que ocupa la i -ésima fila y j -ésima columna de la matriz \mathbf{X} , entonces

$$\left[\frac{\partial}{\partial \mathbf{X}} f \right]_{i,j} = \frac{\partial f}{\partial [\mathbf{X}]_{i,j}}. \quad (\text{F.1})$$

Aquí únicamente se muestran las relaciones utilizadas a lo largo de la tesis. Véase [99] para una colección más completa de resultados.

$$\begin{aligned} \frac{\partial}{\partial \mathbf{X}} \text{Tr} \{ \mathbf{C} \} &= 0 \\ \frac{\partial}{\partial \mathbf{X}} \text{Tr} \{ \mathbf{X} \mathbf{B} \mathbf{X}^\top \} &= \mathbf{X} \mathbf{B}^\top + \mathbf{X} \mathbf{B} \\ \frac{\partial}{\partial \mathbf{X}} \text{Tr} \{ \mathbf{X}^\top \mathbf{B} \mathbf{X} \} &= \mathbf{B} \mathbf{X} + \mathbf{B}^\top \mathbf{X} \\ \frac{\partial}{\partial \mathbf{X}} \text{Tr} \{ \mathbf{A} \mathbf{X} \mathbf{B} \} &= \mathbf{A}^\top \mathbf{B}^\top \\ \frac{\partial}{\partial \mathbf{X}} \text{Tr} \{ \mathbf{X} \mathbf{B} \} &= \mathbf{B}^\top \\ \frac{\partial}{\partial \mathbf{X}} \text{Tr} \{ \mathbf{X}^\top \mathbf{B} \} &= \mathbf{B} \\ \frac{\partial}{\partial \mathbf{X}} \text{Tr} \{ \mathbf{A} \mathbf{X}^\top \} &= \mathbf{A} \end{aligned}$$

Abreviaturas

- 2G (sistemas de comunicaciones móviles de) segunda generación
- 3G (sistemas de comunicaciones móviles de) tercera generación
- 4G (sistemas de comunicaciones móviles de) cuarta generación
- AIC *An Information Criterion* (un criterio de información)
- AR autorregresivo (referido a un proceso estocástico)
- AWGN *Additive White Gaussian Noise* (ruido blanco aditivo gaussiano)
- BER *Bit Error Rate* (tasa de error de bit)
- BPSK *Binary Phase-Shift Keying*
- CDMA *Code Division Multiple Access* (acceso múltiple por división de código)
- CME *Conditional Model Estimator*
- CSI *Channel State Information* (información del estado del canal)
- DD dirigido por decisión
- DFE *Decision Feedback Equalizer* (igualador de decisiones realimentadas)
- EM (algoritmo) *Expectation-Maximization* (esperanza-maximización)
- f.d.p. función de densidad de probabilidad
- f.m.p. función de masa de probabilidad
- FP filtros de partículas
- GPS *Global Positioning System* (sistema de posicionamiento global)
- i.i.d. independientes e idénticamente distribuidas
- IS *Importance Sampling* (muestreo enfatizado)

- ISI *inter-symbolic interference* (interferencia intersimbólica)
- kbps kilo bits por segundo
- KF *Kalman Filter* (filtro de Kalman)
- LMS *Least Mean Squares* (mínimos cuadrados medios)
- LS *Least Squares* (mínimos cuadrados)
- MAP Maximum A Posteriori
- Mbps Mega bits por segundo
- MDL *Minimum Description Length* (descripción de menor longitud)
- MIMO *Multiple Input Multiple Output* (múltiples entradas múltiples salidas)
- MISO *Multiple Input Single Output* (múltiples entradas una única salida)
- MKF Mixture Kalman Filter
- MMSE *Minimum Mean Square Error* (error cuadrático medio mínimo)
- MSE *Mean Square Error* (error cuadrático medio)
- MV Máxima Verosimilitud
- OFDM *Orthogonal Frequency-Division Multiplexing* (multiplexación por división ortogonal de frecuencia)
- PSP *per-survivor processing* (procesado por supervivientes)
- RLS *Recursive Least Squares* (mínimos cuadrados recursivo)
- SAGE *Space Alternating Generalized EM* (EM generalizado con alternancia de espacio de estados)
- SIMO *Single Input Multiple Output* (una única entrada múltiples salidas)
- SIR *Sequential Importance Sampling with Resampling* (muestreo enfatizado secuencial con remuestreo)
- SIS *Sequential Importance Sampling* (muestreo enfatizado secuencial)
- SISO *Single Input Single Output* (única entrada única salida)
- SMC secuenciales de Monte Carlo (métodos)

SNR *Signal-to-Noise-Ratio* (relación señal a ruido)

UWB *Ultra Wide Band* (banda ultra ancha)

v.a. variable aleatoria

VA *Viterbi Algorithm* (algoritmo de Viterbi)

WiMAX *Worldwide Interoperability for Microwave Access*

WLAN *Wireless Local Area Network* (red inalámbrica de área local)

WMAN *Wireless Metropolitan Area Network* (red inalámbrica de área metropolitana)

WPAN *Wireless Personal Area Network* (red inalámbrica de área personal)

WSSUS *Wide-Sense stationary uncorrelated scattering* (estacionario en sentido amplio y de dispersión incorrelada)

Bibliografía

- [1] S. Verdú, *Multiuser Detection*. Cambridge (UK): Cambridge University Press, 1998.
- [2] P. M. Djurić, J. H. Kotecha, J. Zhang, Y. Huang, T. Ghirmai, M. F. Bugallo, and J. Míguez, “Particle filtering,” *IEEE Signal Processing Magazine*, vol. 20, no. 5, pp. 19–38, September 2003.
- [3] F. Bueche, *Ciencias físicas*. Reverté, 1979.
- [4] P. Misra and P. Enge, *Global positioning system*, 2nd ed. Ganga-Jamuna Press, 2006.
- [5] J. H. Schiller, *Mobile communications*, 2nd ed. Addison-Wesley, 2003.
- [6] T. S. Rappaport, *Wireless Communications*. Upper Saddle River, NJ (USA): Prentice-Hall, 1996.
- [7] R. K. Nichols and P. C. Lekkas, *Wireless Security: Models, Threats, and Solutions*. McGraw-Hill Professional, 2001.
- [8] N. Chandler and S. Jones, *The Basics of Cellular*. Intl. Engineering Consortiu, 2006.
- [9] A. Boukerche, *Handbook of Algorithms For Wireless Networking and Mobile Computing*. CRC Press, 2005.
- [10] J. Wang, *Broadband Wireless Communications: 3G, 4G and Wireless LAN*. Springer, 2001.
- [11] S. Ahson and M. Ilyas, *WiMAX Handbook*. CRC Press, 2008.
- [12] X. Shen, *Ultra-Wideband Wireless Communications and Networks*. John Wiley, 2007.
- [13] E. Telatar, “Capacity of multi-antenna gaussian channels,” *European transactions on telecommunications*, vol. 10, pp. 585–595, December 1999.
- [14] A. Gorokhov and P. Loubaton, “Blind identification of MIMO-FIR systems: A generalized linear prediction approach,” *Signal Processing*, vol. 73, no. 1-2, pp. 105–124, January 1999.

- [15] A. N. Mody and G. L. Stuber, "Synchronization for MIMO OFDM systems," in *Proceedings of IEEE ICC*, 2001.
- [16] B. Dong and X. Wang, "Sampling-based soft equalization for frequency-selective MIMO channels," *IEEE Transactions on Communications*, vol. 53, no. 2, pp. 278–288, February 2005.
- [17] D. Guo and X. Wang, "Blind detection in MIMO systems via sequential Monte Carlo," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 3, pp. 464–473, April 2003.
- [18] S. ten Brink and B. M. Hochwald, "Achieving near-capacity on a multiple-antenna channel," *IEEE Transactions on Wireless Communications*, vol. 51, no. 3, pp. 389–399, March 2003.
- [19] P. Hoffman, *The Man Who Loved Only Numbers : The Story of Paul Erdos and the Search for Mathematical Truth*. Hyperion, 1998.
- [20] N. Metropolis, "The beginning of the Monte Carlo method," *Los Alamos Science*, vol. Special Issue, pp. 125–130, 1987.
- [21] J. M. Hammersley and K. W. Morton, "Poor man's Monte Carlo," *Journal of the Royal Statistical Society*, pp. 23–38, 1954.
- [22] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equations of state calculations by fast computing machines," *Journal of Chemical Physics*, vol. 21 or 6, pp. 1087–1091, 1953.
- [23] M. Rosenbluth and A. Rosenbluth, "Monte Carlo calculation of the average extension of molecular chains," *Journal of Chemical Physics*, vol. 23, pp. 356–359, February 1955.
- [24] N. Gordon, D. Salmond, and A. F. M. Smith, "Novel approach to nonlinear and non-Gaussian Bayesian state estimation," *IEE Proceedings-F*, vol. 140, pp. 107–113, 1993.
- [25] B. Dong, X. Wang, and A. Doucet, "A new class of soft MIMO demodulation algorithms," *IEEE Transactions on Signal Processing*, vol. 51, no. 11, pp. 2752–2763, November 2003.
- [26] T. Bertozzi and D. L. Ruyet, "Iterative detection in MIMO channels using particle filtering," in *Proceedings of the IEEE ICC*, vol. 5, June 2004, pp. 2694–2698.
- [27] Y. Huang, J. Zhang, and P. M. Djurić, "Bayesian detection for BLAST," *IEEE Transactions on Signal Processing*, vol. 53, no. 3, pp. 1086–1096, March 2005.
- [28] Y. Huang, J. Zhang, I. T. Luna, P. M. Djurić, and D. P. R. Padillo, "Adaptive blind multiuser detection over flat fast fading channels using particle filtering," *EURASIP J. Wirel. Commun. Netw.*, vol. 2005, no. 2, pp. 130–140, 2005.

- [29] J. Zhang, Y. Huang, and P. M. Djurić, “Multiuser detection with particle filtering,” in *Proceedings of XI EUSIPCO*, September 2002.
- [30] K. J. Kim and R. A. Iltis, “A sequential Monte-Carlo Kalman filter based delay and channel estimation method in the MIMO-OFDM system,” in *Proceedings of VTC 2004*, vol. 1, September 2004, pp. 26–29.
- [31] K. J. Kim, J. Yue, R. A. Iltis, and J. D. Gibson, “A QRD-M/Kalman filter-based detection and channel estimation algorithm for MIMO-OFDM systems,” *IEEE Transactions on Wireless Communications*, vol. 4, no. 2, pp. 710–721, March 2005.
- [32] E. Punskeya, A. Doucet, and W. J. Fitzgerald, “On the use and misuse of particle filtering in digital communications,” in *Proceedings of EUSIPCO*, September 2002.
- [33] P. D. Grünwald, I. J. Myung, and M. A. Pitt, *Advances in Minimum Description Length. Theory and Applications*. The MIT Press, 2005.
- [34] H. Akaike, “A new look at the statistical model identification,” *IEEE Transactions on Automatic Control*, vol. 19, no. 6, p. 716–723, December 1974.
- [35] A. Liavas, P. Regalia, and J. Delmas, “Blind channel approximation: effective channel order determination,” *IEEE Transactions Signal Processing*, vol. 47, no. 12, pp. 3336–3344, December 1999.
- [36] A. Liavas and P. Regalia, “On the behavior of information theoretic criteria for model order selection,” *IEEE Transactions Signal Processing*, vol. 49, no. 8, pp. 1689–1695, August 2001.
- [37] S. Kay, “Conditional model order estimation,” *IEEE Transactions on Signal Processing*, vol. 49, no. 9, pp. 1900–1917, September 2001.
- [38] D. K. C. So and R. S. Cheng, “Iterative EM receiver for space-time coded systems in MIMO frequency-selective fading channels with channel gain and order estimation,” *IEEE Transactions on Wireless Communications*, vol. 3, no. 6, pp. 1928–1935, November 2004.
- [39] J. Vía, I. Santamaría, and J. Pérez, “Effective channel order estimation based on combined identification/equalization,” *IEEE Transactions on Signal Processing*, vol. 54, no. 9, pp. 3518–3526, September 2006.
- [40] J. Míguez and P. M. Djurić, “Blind equalization of frequency-selective channels by sequential importance sampling,” *IEEE Transactions on Signal Processing*, vol. 52, no. 10, pp. 2738–2748, October 2004.

- [41] D. Yee, J. P. Reilly, and T. Kirubarajan, "A blind sequential Monte Carlo detector for OFDM systems in the presence of phase noise, multipath fading, and channel order uncertainty," *IEEE Transactions on Signal Processing*, vol. 55, no. 9, pp. 4581–4598, September 2007.
- [42] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*. Springer, 2004.
- [43] A. Doucet, N. de Freitas, and N. Gordon, Eds., *Sequential Monte Carlo Methods in Practice*. New York (USA): Springer, 2001.
- [44] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Journal of Basic Engineering*, vol. 82, pp. 35–45, 1960.
- [45] A. Doucet, S. Godsill, and C. Andrieu, "On sequential Monte Carlo Sampling methods for Bayesian filtering," *Statistics and Computing*, vol. 10, no. 3, pp. 197–208, 2000.
- [46] E. Punskeya, "Sequential Monte Carlo methods for digital communications," Ph.D. dissertation, St. John's College, 2003.
- [47] D. Crisan and A. Doucet, "A survey of convergence results on particle filtering," *IEEE Transactions Signal Processing*, vol. 50, no. 3, pp. 736–746, March 2002.
- [48] J. S. Liu and R. Chen, "Blind deconvolution via sequential imputations," *Journal of the American Statistical Association*, vol. 90, no. 430, pp. 567–576, June 1995.
- [49] A. Doucet, N. de Freitas, and N. Gordon, "An introduction to sequential Monte Carlo methods," in *Sequential Monte Carlo Methods in Practice*, A. Doucet, N. de Freitas, and N. Gordon, Eds. Springer, 2001, ch. 1, pp. 4–14.
- [50] M. Bolic, P. Djuric, and H. Sangjin, "Resampling algorithms and architectures for distributed particle filters," *IEEE Transactions on Signal Processing*, vol. 53, no. 7, pp. 2442–2450, July 2005.
- [51] R. Douc, O. Cappe, and E. Moulines, "Comparison of resampling schemes for particle filtering," in *Proceedings of the ISPA 2005*, September 2005.
- [52] J. S. Liu and R. Chen, "Sequential Monte Carlo methods for dynamic systems," *Journal of the American Statistical Association*, vol. 93, no. 443, pp. 1032–1044, September 1998.
- [53] G. Kitagawa, "Monte Carlo filter and smoother for non-gaussian nonlinear state space models," *Journal of Computational and Graphical Statistics*, vol. 5, no. 1, pp. 1–25, March 1996.
- [54] D. Crisan, P. del Moral, and T. Lyons, "Non-linear filtering using branching and interacting particle systems," *Markov Processes Related Fields*, vol. 5, no. 3, pp. 293–319, 1999.

- [55] D. Crisan, "Particle filters - a theoretical perspective," in *Sequential Monte Carlo Methods in Practice*, A. Doucet, N. de Freitas, and N. Gordon, Eds. Springer, 2001, ch. 2, pp. 17–42.
- [56] P. D. Moral, *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*. Springer, 2004.
- [57] J. Míguez and L. Castedo, "Space-time channel estimation and soft detection in time-varying multiaccess channels," *Signal Processing*, vol. 83, no. 2, pp. 389–411, January 2003.
- [58] J. Zhang and P. M. Djurić, "Joint estimation and decoding of space-time trellis codes," *Journal of Applied Signal Processing*, vol. 3, pp. 305–315, 2002.
- [59] C. Komnikakis, C. Fragouli, A. H. Sayeed, and R. D. Wesel, "Multi-input multi-output fading channel tracking and equalization using kalman estimation," *IEEE Transactions Signal Processing*, vol. 50, no. 5, pp. 1065–1076, May 2002.
- [60] Y. Huang and P. M. Djurić, "A blind particle filtering detector of signals transmitted over flat fading channels," *IEEE Transactions on Signal Processing*, vol. 52, no. 7, pp. 1891–1900, July 2004.
- [61] F. Babich and G. Lombardi, "A markov model for the mobile propagation channel," *IEEE Transactions on Vehicular Technology*, vol. 49, no. 1, pp. 63–73, January 2000.
- [62] C. Shannon, "A mathematical theory of communication," *Bell system technical journal*, vol. 27, pp. 379–423, 1948.
- [63] G. G. Raleigh and J. M. Cioffi, "Spatio-temporal coding for wireless communication," *IEEE Transactions Communications*, vol. 46, no. 3, pp. 357–366, March 1998.
- [64] G. Strang, *Linear Algebra, Second Edition*. New York: Academic Press, Inc., 1980.
- [65] J. Heller and I. Jacobs, "Viterbi decoding for satellite and space communications," *IEEE Transactions on Communications Technology*, vol. COM19, no. 5, pp. 835–848, October 1971.
- [66] S. G. Glisic and P. A. Leppanen, *Wireless Communications: TDMA Versus CDMA*. Norwell, MA, USA: Kluwer Academic Publishers, 1997.
- [67] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Transactions on Communications Technology*, vol. 20, no. 2, pp. 284–287, March 1974.

- [68] C. Berrou, A. Glavieux, and P. Thitimajshima, “Near shannon limit error-correcting coding and decoding:turbo-codes,” in *IEEE International Conference on Communications (ICC) 1993*, vol. 2, May 1993, pp. 1064–1070.
- [69] G. J. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*. New York: Wiley Series in Probability and Statistics, John Wiley & Sons, 1997.
- [70] A. P. . Dempster, N. M. Laird, and D. B. Rubin, “Maximum likelihood from incomplete data via the EM algorithm,” *J. Roy. Stat. Soc., Ser. B*, vol. 39, no. 1, pp. 1–38, 1977.
- [71] C. W. Therrien, *Discrete Random Signals and Statistical Signal Processing*. Englewood Cliffs (NJ, USA): Prentice-Hall, 1992.
- [72] A. O. Hero and J. A. Fessler, “Convergence in norm for alternating expectation-maximization (EM) type algorithms,” *Statistica Sinica*, vol. 5, no. 1, January 1995.
- [73] X. L. Meng and D. Rubin, “On the global and componentwise rates of convergence of the EM algorithm,” *Linear algebra & its applicat.*, vol. 199, pp. 413–425, 1994.
- [74] L. B. Nelson and V. Poor, “Iterative multiuser receivers for CDMA channels: An EM-based approach,” *IEEE Transactions Communications*, vol. 44, no. 12, pp. 1700–1710, December 1996.
- [75] J. A. Fessler and A. O. Hero, “Space-alternating generalized expectation-maximization algorithm,” *IEEE Transactions Signal Processing*, vol. 42, no. 10, pp. 2664–2677, October 1994.
- [76] R. Chen and J. S. Liu, “Mixture Kalman filters,” *Journal of the Royal Statistics Society B*, vol. 62, pp. 493–508, 2000.
- [77] Y. Huang, J. M. Zhang, and P. M. Djurić, “Bayesian detection for BLAST,” *IEEE Transactions Signal Processing*, vol. 53, no. 3, pp. 1086– 1096, March 2005.
- [78] P. Fearnhead, “Sequential Monte Carlo method in filter theory,” Ph.D. dissertation, Oxford University, 1998.
- [79] M. K. Pitt and N. Shephard, “Auxiliary variable based particle filters,” in *Sequential Monte Carlo Methods in Practice*, A. Doucet, N. de Freitas, and N. Gordon, Eds. Springer, 2001, ch. 13, pp. 273–293.
- [80] M. A. Vázquez and J. Míguez, “A complexity-constrained particle filtering algorithm for MAP equalization of frequency-selective MIMO channels,” in *Proceedings of the IEEE ICASSP*, May 2005.
- [81] M. A. Vázquez, M. F. Bugallo, and J. Míguez, “Novel SMC techniques for blind equalization of flat-fading MIMO channels,” in *Proceedings of the IEEE VTC*, June 2005.

- [82] S. Haykin, *Adaptive Filter Theory, 4th Edition*. Prentice Hall, Information and System Sciences Series, 2001.
- [83] J. Míguez and L. Castedo, “Semiblind maximum likelihood demodulation for CDMA systems,” *IEEE Transactions Vehicular Technology*, vol. 51, no. 4, pp. 775–781, July 2002.
- [84] J. Rissanen, “Modeling by the shortest data description,” *Automatica*, pp. 465–471, 1978.
- [85] R. Solomonoff, “A formal theory of inductive inference, part I,” *Information and Control*, vol. 7, no. 1, pp. 1–22, March 1964.
- [86] ———, “A formal theory of inductive inference, part II,” *Information and Control*, vol. 7, no. 2, pp. 224–254, June 1964.
- [87] M. J. Kearns and U. V. Vazirani, *An Introduction to Computational Learning Theory*. Cambridge, Massachusetts: MIT press, 1994.
- [88] M. Li and P. Vitányi, *An Introduction to Kolmogorov Complexity and Its Applications*. New York: Springer-Verlag, 1997.
- [89] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York (USA): Wiley-Interscience, 1991.
- [90] G. Xu, H. Liu, L. Tong, and T. Kailath, “A least-squares approach to blind channel identification,” *IEEE Transactions Signal Processing*, vol. 43, no. 12, pp. 2982–2993, December 1995.
- [91] D. Kotoulas, P. Koukoulas, and N. Kalouptsidis, “Subspace projection based blind channel order estimation of MIMO systems,” *IEEE Transactions on Signal Processing*, vol. 54, no. 4, pp. 1351–1363, 2006.
- [92] W. C. Jakes, *Microwave Mobile Communications*. John Wiley & Sons, 1974.
- [93] J. Míguez, “Analysis of parallelizable resampling algorithms for particle filtering,” *Signal Processing*, vol. 87, no. 12, pp. 3155–3174, 2007.
- [94] M. A. Vázquez and J. Míguez, “Sequential map equalization of MIMO channels and its application to UWB communications,” in *Proceedings of the WSA 2006*, Reisingburg, Germany, March 2006.
- [95] M. A. Vázquez, M. F. Bugallo, and J. Míguez, “Sequential Monte Carlo methods for complexity-constrained MAP equalization of dispersive MIMO channels,” *Signal Processing*, vol. 88, pp. 1017–1034, 2008.

- [96] M. A. Vázquez and J. Míguez, “Sequential MAP equalization of MIMO channels when the channel order is unknown,” in *Proceedings of the WSA 2007*, Vienna, Austria, February 2007.
- [97] —, “A per-survivor processing algorithm for maximum likelihood equalization of MIMO channels with unknown order,” in *Proceedings of the WSA 2008*, Darmstadt, Germany, February 2008.
- [98] —, “Maximum likelihood sequence detection in time and frequency selective MIMO channels with unknown order,” *IEEE Transactions on Vehicular Technology*, vol. 57, no. 0, 2009.
- [99] K. T. Fang and Y. T. Zhang, *Generalized Multivariate Analysis*. Springer, 1990.

Índice alfabético

- trellis* (diagrama de), 145
- AIC, *véase* Akaike (criterio de información de)
- Akaike (criterio de información de), 91
- autorregresivo (proceso), 36
- BCJR, 48
- capacidad, 38
- Cholesky (descomposición de), 68
- CME, 91
- decorrelación (detector de), 43
- degeneración de los pesos (problema de la), 21
- DFE, *véase* receptor realimentado
- EM, 50
- EM generalizado, 52
- error cuadrático medio mínimo, 44
- esperanza-maximización, *véase* EM
- estimación blanda, 42
- estimación conjunta del canal y los datos, 76
- estimación dura, 42
- factor de olvido, 79
- filtrado de partículas, *véase* métodos secuenciales de Monte Carlo
- filtro lineal, *véase* receptor lineal
- función tentativa, 14
- interferencia intersimbólica, 44
- interferencia intersimbólica anticausal, 45
- IS, *véase* muestreo enfatizado
- ISI, *véase* interferencia intersimbólica
- ISI anticausal, *véase* interferencia intersimbólica anticausal
- lema de inversión, 154
- LMS, 78, 79
- LS, *véase* mínimos cuadrados (estimador)
- mínimos cuadrados (estimador), 153
- mínimos cuadrados recursivo, 79
- métodos de Monte Carlo, 5
- métodos secuenciales de Monte Carlo, 5, 17
- Markov (proceso de), 98
- markoviano (proceso), 18
- MDL, 90
- MKF, 53
- MMSE, *véase* error cuadrático medio mínimo
- muestreo enfatizado, 14
- muestreo enfatizado secuencial, 20
- muestreo exacto, 12
- muestreo retardado, 55
- partícula, 14
- procesado por supervivientes, 47
- programación dinámica, 45
- PSP, *véase* procesado por supervivientes
- ratio de remuestreo, 24
- receptor lineal, 41
- receptor realimentado, 44
- remuestreo, 21
 - estratificado, 27
 - multinomial, 25
 - residual, 26
 - sistemático, 28
- RLS, *véase* mínimos cuadrados recursivo
- SAGE, *véase* EM generalizado
- selección de modelo, 90

SIR, 30

SIS, *véase* muestreo enfatizado secuencial

sistemas dinámicos en formato de espacio de estados, 18

SMC, *véase* métodos secuenciales de Monte Carlo

suavizado, 37, 43

superviviente, 48

tamaño efectivo de la muestra, 22

triangularización de la matriz de canal, 68

variables auxiliares, 64

Viterbi, 45

Wi-Fi, 3

WiMAX, 3

WLAN, 3

WMAN, 3

WPAN, 3