

Estudiando el libro “Fundamentals of Statistical Signal
Processing: Estimation Theory” de Steven M. Kay

Ernesto López

21 de febrero de 2023

Índice general

1. Introducción	2
1.1. Problemas	2
1.1.1. Problema 1	2
1.1.2. Problema 2	3
1.1.3. Problema 3	4
1.1.4. Problema 4	5
1.1.5. Problema 5	7
2. Estimación insesgada de varianza mínima	8
2.1. Problemas	8
2.1.1. Problema 1	8
2.1.2. Problema 2	9
2.1.3. Problema 3	9
2.1.4. Problema 4	9
2.1.5. Problema 5	10
2.1.6. Problema 6	12
2.1.7. Problema 7	13
2.1.8. Problema 8	14
2.1.9. Problema 9	15
2.1.10. Problema 10	17
2.1.11. Problema 11	18
3. Cota Inferior de Cramer-Rao	19
3.1. Cota Inferior de Cramer-Rao: parámetro escalar	19
3.1.1. Teorema	19
3.1.2. Transformación de parámetros	24
3.2. Cota Inferior de Cramer-Rao: parámetro vectorial	24
3.2.1. Teorema	24
3.2.2. Ejemplo: nivel de DC en WGN	29
3.2.3. Ejemplo: ajuste de recta	30
3.3. CRLB general para señales en WGN	32
3.3.1. Ejemplo: fase de una senoide	33
3.4. CRLB para el caso gaussiano general	33
3.4.1. Teorema	33
3.4.2. Ejemplo: parámetros de una señal en WGN	39
3.4.3. Ejemplo: nivel de DC aleatorio en WGN	39
3.4.4. Ejemplo: estimación de los parámetros de una senoide	41
3.5. CRLB asintótica para procesos WSS gaussianos	42
3.5.1. Teorema	43
3.5.2. Ejemplo: estimación de los parámetros de un modelo autorregresivo	47
3.6. Problemas	50
3.6.1. Problema 1	50
3.6.2. Problema 2	51
3.6.3. Problema 3	54

3.6.4. Problema 4	56
3.6.5. Problema 5	56
3.6.6. Problema 6	57
3.6.7. Problema 7	58
3.6.8. Problema 8	59
3.6.9. Problema 9	61
3.6.10. Problema 10	63
3.6.11. Problema 11	65
3.6.12. Problema 12	66
3.6.13. Problema 13	67
3.6.14. Problema 14	68
3.6.15. Problema 15	69
3.6.16. Problema 16	71
3.6.17. Problema 17	72
3.6.18. Problema 20	73
4. Modelos lineales	76
4.1. Introducción	76
4.2. Definición y propiedades	76
4.3. Teorema: estimador MVU para el modelo lineal	77
4.4. Ejemplos	78
4.4.1. Ajuste de curva	78
4.4.2. Análisis de Fourier	79
4.4.3. Identificación de un sistema	81
4.5. Extensión del modelo lineal	88
4.5.1. Ejemplo: nivel de DC en ruido coloreado	89
4.6. Problemas	90
4.6.1. Problema 1	90
4.6.2. Problema 2	91
4.6.3. Problema 3	92
4.6.4. Problema 4	92
4.6.5. Problema 5	93
4.6.6. Problema 6	94
4.6.7. Problema 7	96
4.6.8. Problema 8	97
4.6.9. Problema 9	98
4.6.10. Problema 10	99
4.6.11. Problema 11	100
4.6.12. Problema 12	101
4.6.13. Problema 13	103
4.6.14. Problema 14	104
5. Estimación insesgada de varianza mínima general	105
5.1. Factorización de Neyman-Fisher	105
5.1.1. Teorema	105
5.1.2. Ejemplo: nivel de DC en WGN	108
5.1.3. Ejemplo: fase de una senoide	109
5.2. Uso del estadístico suficiente para encontrar el estimador MVU	110
5.2.1. Ejemplo: nivel de DC en WGN	110
5.2.2. Teorema de Rao-Blackwell-Lehmann-Scheffe	111
5.2.3. Ejemplo: media de ruido uniforme	113
5.3. Extensión a un parámetro vectorial	117
5.3.1. Teorema: Factorización de Neyman-Fisher (parámetro vectorial)	117
5.3.2. Ejemplo: estimación de los parámetros de una senoide	118
5.3.3. Ejemplo: Nivel de DC en WGN con potencia del ruido desconocida	118
5.3.4. Teorema: Rao-Blackwell-Lehmann-Scheffe (parámetro vectorial)	119

5.3.5.	Ejemplo: Nivel de DC en WGN con potencia del ruido desconocida (continuación)	119
5.4.	Problemas	121
5.4.1.	Problema 1	121
5.4.2.	Problema 2	122
5.4.3.	Problema 3	122
5.4.4.	Problema 4	123
5.4.5.	Problema 5	123
5.4.6.	Problema 6	124
5.4.7.	Problema 7	124
5.4.8.	Problema 8	125
5.4.9.	Problema 9	125
5.4.10.	Problema 10	126
5.4.11.	Problema 11	127
5.4.12.	Problema 12	129
5.4.13.	Problema 13	129
5.4.14.	Problema 14	131
5.4.15.	Problema 15	132
5.4.16.	Problema 16	135
5.4.17.	Problema 17	138
5.4.18.	Problema 18	141
5.4.19.	Problema 19	143
6.	Mejor estimador lineal insesgado	145
6.1.	Introducción	145
6.2.	Estimador BLUE	145
6.2.1.	Parámetro escalar	145
6.2.2.	Parámetro vectorial	145
6.3.	Ejemplo de procesamiento de señales	146
6.3.1.	Ejemplo: localización de una fuente	146
6.4.	Problemas	150
6.4.1.	Problema 1	150
6.4.2.	Problema 2	150
6.4.3.	Problema 3	151
6.4.4.	Problema 4	153
6.4.5.	Problema 5	153
6.4.6.	Problema 6	155
6.4.7.	Problema 7	155
6.4.8.	Problema 8	156
6.4.9.	Problema 9	158
6.4.10.	Problema 10	160
6.4.11.	Problema 11	161
6.4.12.	Problema 12	165
6.4.13.	Problema 13	166
6.4.14.	Problema 14	166
6.4.15.	Problema 15	168
6.4.16.	Problema 16	168
7.	Estimación de máxima verosimilitud	171
7.1.	Estimador de máxima verosimilitud	171
7.1.1.	Definición	171
7.1.2.	Propiedades asintóticas del MLE	171
7.1.3.	Propiedad de invarianza del MLE	174
7.1.4.	MLE para el modelo lineal	174
7.2.	Ejemplos básicos	176
7.2.1.	Ejemplo: MLE de los parámetros de ruido de Laplace	176
7.2.2.	Ejemplo: MLE de la fase de una senoide	177

7.2.3.	Ejemplo: nivel de DC en ruido no independiente no gaussiano	179
7.3.	Algoritmo esperanza-maximización	180
7.3.1.	Introducción	180
7.3.2.	El algoritmo EM	181
7.3.3.	Ejemplo: estimación de las frecuencias de mezcla de sinusoides	182
7.4.	MLE asintótico	184
7.4.1.	Ejemplo: proceso de media móvil gaussiano	185
7.5.	Ejemplos de procesamiento de señales	187
7.5.1.	Ejemplo: estimación de los parámetros de una senoide	187
7.5.2.	Ejemplo: estimación de los parámetros de un modelo autorregresivo	190
7.6.	Problemas	192
7.6.1.	Problema 1	192
7.6.2.	Problema 2	193
7.6.3.	Problema 3	194
7.6.4.	Problema 4	195
7.6.5.	Problema 5	195
7.6.6.	Problema 6	196
7.6.7.	Problema 7	197
7.6.8.	Problema 8	198
7.6.9.	Problema 9	199
7.6.10.	Problema 10	199
7.6.11.	Problema 11	200
7.6.12.	Problema 12	201
7.6.13.	Problema 13	201
7.6.14.	Problema 14	201
7.6.15.	Problema 15	204
7.6.16.	Problema 16	206
7.6.17.	Problema 17	207
7.6.18.	Problema 18	208
7.6.19.	Problema 19	209
7.6.20.	Problema 20	211
7.6.21.	Problema 21	213
7.6.22.	Problema 22	213
7.6.23.	Problema 23	214
7.6.24.	Problema 24	215
7.6.25.	Problema 25	216
7.6.26.	Problema 26	219
8.	Mínimos cuadrados	222
8.1.	Enfoque de mínimos cuadrados	222
8.2.	Mínimos cuadrados lineales	222
8.2.1.	Caso escalar	222
8.2.2.	Caso vectorial	223
8.2.3.	Interpretación geométrica	224
8.3.	Mínimos cuadrados de orden recursivo	225
8.3.1.	Algoritmo LS de orden recursivo	227
8.4.	Mínimos cuadrados secuencial	231
8.4.1.	Caso escalar	231
8.4.2.	Caso vectorial	234
8.5.	Mínimos cuadrados restringidos	238
8.6.	Mínimos cuadrados no lineales	240
8.7.	Ejemplos de procesamiento de señales	242
8.7.1.	Ejemplo: diseño de filtros digitales	242
8.7.2.	Ejemplo: estimación de los parámetros AR de un modelo ARMA	247
8.7.3.	Ejemplo: cancelador de ruido adaptativo	248
8.8.	Problemas	250

8.8.1. Problema 1	250
8.8.2. Problema 2	252
8.8.3. Problema 3	252
8.8.4. Problema 4	254
8.8.5. Problema 5	254
8.8.6. Problema 6	255
8.8.7. Problema 7	256
8.8.8. Problema 8	257
8.8.9. Problema 9	258
8.8.10. Problema 10	258
8.8.11. Problema 11	259
8.8.12. Problema 12	260
8.8.13. Problema 13	261
8.8.14. Problema 14	262
8.8.15. Problema 15	263
8.8.16. Problema 16	265
8.8.17. Problema 17	267
8.8.18. Problema 18	267
8.8.19. Problema 19	268
8.8.20. Problema 20	269
8.8.21. Problema 21	270
8.8.22. Problema 22	271
8.8.23. Problema 23	272
8.8.24. Problema 24	274
8.8.25. Problema 25	275
8.8.26. Problema 26	276
8.8.27. Problema 27	277
8.8.28. Problema 28	279
9. Método de los momentos	285
9.1. Introducción	285
9.2. Método de los momentos	286
9.2.1. Parámetro escalar	286
9.2.2. Parámetro vectorial	287
9.3. Evaluación estadística de los estimadores	287
9.3.1. Evaluación cuando el registro de datos es grande	288
9.3.2. Evaluación cuando la SNR es grande	290
9.4. Ejemplo de procesamiento de señales	291
9.5. Problemas	294
9.5.1. Problema 1	294
9.5.2. Problema 2	295
9.5.3. Problema 3	296
9.5.4. Problema 4	297
9.5.5. Problema 5	297
9.5.6. Problema 6	298
9.5.7. Problema 7	298
9.5.8. Problema 8	300
9.5.9. Problema 9	300
9.5.10. Problema 10	302
9.5.11. Problema 11	303
9.5.12. Problema 12	303
9.5.13. Problema 13	305

10.La filosofía Bayesiana	306
10.1. Conocimiento a priori y estimación	306
10.2. Elección de la PDF a priori	310
10.2.1. Ejemplo: nivel de DC en WGN con PDF a priori gaussiana	310
10.3. Propiedades de la PDF gaussiana	313
10.4. Modelo lineal bayesiano	315
10.5. Problemas	317
10.5.1. Problema 1	317
10.5.2. Problema 2	317
10.5.3. Problema 3	319
10.5.4. Problema 4	320
10.5.5. Problema 5	321
10.5.6. Problema 6	322
10.5.7. Problema 7	323
10.5.8. Problema 8	324
10.5.9. Problema 9	326
10.5.10. Problema 10	327
10.5.11. Problema 11	328
10.5.12. Problema 12	328
10.5.13. Problema 13	329
10.5.14. Problema 14	330
10.5.15. Problema 15	333
10.5.16. Problema 16	334
10.5.17. Problema 17	335
 11.Estimadores bayesianos generales	 336
11.1. Funciones de costo	336
11.2. Estimadores de error cuadrático medio mínimo	338
11.2.1. Análisis de Fourier bayesiano	339
11.3. Estimadores máximo a posteriori	342
11.3.1. Ejemplo: PDF exponencial	342
11.4. Descripción del desempeño	345
11.4.1. Caso escalar	345
11.4.2. Caso vectorial	346
11.5. Ejemplo de procesamiento de señales	349
11.6. Problemas	351
11.6.1. Problema 1	351
11.6.2. Problema 2	353
11.6.3. Problema 3	354
11.6.4. Problema 4	355
11.6.5. Problema 5	357
11.6.6. Problema 6	357
11.6.7. Problema 7	360
11.6.8. Problema 8	360
11.6.9. Problema 9	361
11.6.10. Problema 10	362
11.6.11. Problema 11	362
11.6.12. Problema 12	363
11.6.13. Problema 13	364
11.6.14. Problema 14	364
11.6.15. Problema 15	366
11.6.16. Problema 16	367
11.6.17. Problema 17	370
11.6.18. Problema 18	372

12.Estimadores bayesianos lineales	375
12.1. Estimación MMSE lineal	375
12.1.1. Ejemplo: nivel de DC en WGN con PDF a priori uniforme	376
12.2. Interpretación geométrica	378
12.2.1. Ejemplo: estimación con vectores ortogonales	381
12.3. El estimador LMMSE vectorial	381
12.4. Estimación LMMSE secuencial	383
12.4.1. Estimador LMMSE secuencial para el modelo lineal bayesiano	388
12.4.2. Análisis de Fourier bayesiano	392
12.5. Ejemplos de procesamiento de señales: filtrado de Wiener	394
12.6. Problemas	401
12.6.1. Problema 1	401
12.6.2. Problema 2	404
12.6.3. Problema 3	405
12.6.4. Problema 4	406
12.6.5. Problema 5	407
12.6.6. Problema 6	407
12.6.7. Problema 7	408
12.6.8. Problema 8	409
12.6.9. Problema 9	409
12.6.10. Problema 10	410
12.6.11. Problema 11	412
12.6.12. Problema 12	414
12.6.13. Problema 13	414
12.6.14. Problema 14	417
12.6.15. Problema 15	418
12.6.16. Problema 16	419
12.6.17. Problema 17	421
12.6.18. Problema 18	421
12.6.19. Problema 19	423
12.6.20. Problema 20	424
13.Filtros de Kalman	426
13.1. Introducción	426
13.2. Modelos de señales dinámicas	426
13.3. Filtro de Kalman escalar	432
13.4. Comparación entre los filtros de Wiener y de Kalman	440
13.5. Filtro de Kalman vectorial	442
13.5.1. Estado vectorial-observación escalar	442
13.5.2. Estado vectorial-observación vectorial	443
13.6. Filtro de Kalman extendido	447
13.7. Ejemplos de procesamiento de señales	452
13.7.1. Estimación de canal variante en el tiempo	452
13.7.2. Seguimiento de un vehículo	456
13.8. Problemas	462
13.8.1. Problema 1	462
13.8.2. Problema 2	462
13.8.3. Problema 3	463
13.8.4. Problema 4	464
13.8.5. Problema 5	466
13.8.6. Problema 6	466
13.8.7. Problema 7	467
13.8.8. Problema 8	467
13.8.9. Problema 9	469
13.8.10. Problema 10	470
13.8.11. Problema 11	472

13.8.12.Problema 12	473
13.8.13.Problema 13	474
13.8.14.Problema 14	475
13.8.15.Problema 15	475
13.8.16.Problema 16	476
13.8.17.Problema 17	477
13.8.18.Problema 18	477
13.8.19.Problema 19	478
13.8.20.Problema 20	479
13.8.21.Problema 21	479
13.8.22.Problema 22	480
13.8.23.Problema 23	481
A. Identidad matricial de Woodbury	483
B. Sumatoria finita de algunas secuencias sinusiodales	485
C. Momento cuarto de la media muestral	488
D. Derivadas de algunas funciones escalares multivariadas	490
E. Transformada discreta de Fourier real	492
F. PDF condicional de variables aleatorias gaussianas	495
F.1. Caso escalar	495
F.2. Caso multivariado	497
G. Propiedades de matrices en bloques	499
G.1. Determinante	499
G.2. Matriz inversa	499
G.3. Matriz diagonal en bloques	500
H. Conceptos básicos de álgebra lineal	501
H.1. Conceptos básicos	501
H.2. Propiedades y teoremas	502
I. Ecuación analítica de la elipse	505
I.1. Elipse estándar	505
I.2. Elipse rotada centrada en el origen	505
I.3. Ecuación general	507

Prefacio

El presente documento consiste en apuntes sobre el libro “Fundamentals of Statistical Signal Processing: Estimation Theory” [1]. Se hace hincapié en la solución de los problemas del final de cada capítulo del libro, pero también se pretende incluir el desarrollo de algunos conceptos teóricos y demostraciones de forma mas explícita.

Capítulo 1

Introducción

1.1. Problemas

1.1.1. Problema 1

En un sistema de radar, un estimador $\hat{\tau}_0$ del retardo τ_0 de ida y vuelta tiene función de densidad de probabilidad (*probability density function*, PDF) $\mathcal{N}(\tau_0, \sigma_{\tau_0}^2)$. Si se desea encontrar la distancia R , proponer un estimador \hat{R} y determinar su PDF. Además, determinar la desviación estándar $\sigma_{\hat{\tau}_0}$ de forma que el 99 % del tiempo la distancia estimada difiera menos de 100 m de la distancia verdadera. Considerar que la velocidad de propagación electromagnética es $c = 3 \times 10^8$ m/s.

Solución Como τ_0 es el tiempo de propagación de ida y vuelta y por lo tanto corresponde a una distancia recorrida de $2R$, se cumple que $R = \tau_0 c/2$. De esta forma, un estimador de R es

$$\hat{R} = \frac{\hat{\tau}_0 c}{2},$$

que tiene distribución normal debido a que $\hat{\tau}_0$ tiene distribución normal, con media y varianza respectivamente

$$E(\hat{R}) = \frac{E(\hat{\tau}_0)c}{2} = \frac{\tau_0 c}{2} = R, \quad \text{var}(\hat{R}) = \frac{\text{var}(\hat{\tau}_0)c^2}{4} = \frac{\sigma_{\tau_0}^2 c^2}{4}.$$

Se concluye que

$$\hat{R} \sim \mathcal{N}\left(R, \frac{\sigma_{\tau_0}^2 c^2}{4}\right).$$

Se quiere calcular $\sigma_{\hat{\tau}_0}$ tal que la distancia estimada difiera menos de 100 m de la distancia verdadera el 99 % del tiempo, es decir,

$$\Pr\left\{|\hat{R} - R| \leq 100\right\} \geq 0.99 \quad \Rightarrow \quad \Pr\left\{-100 \leq \hat{R} - R \leq 100\right\} \geq 0.99.$$

Dividiendo entre la desviación estándar de \hat{R} se tiene que

$$\Pr\left\{-\frac{100}{\left(\frac{\sigma_{\tau_0} c}{2}\right)} \leq \frac{\hat{R} - R}{\left(\frac{\sigma_{\tau_0} c}{2}\right)} \leq \frac{100}{\left(\frac{\sigma_{\tau_0} c}{2}\right)}\right\} \geq 0.99.$$

De esta forma, la variable aleatoria tiene distribución normal estandarizada $\mathcal{N}(0, 1)$ y es posible emplear las tablas, como por ejemplo, la tabla de la distribución normal acumulada desde la media¹. Teniendo en cuenta que la densidad de probabilidad estándar es una función simétrica, la expresión es equivalente a

$$\Pr\left\{0 \leq \frac{\hat{R} - R}{\left(\frac{\sigma_{\tau_0} c}{2}\right)} \leq \frac{100}{\left(\frac{\sigma_{\tau_0} c}{2}\right)}\right\} \geq \frac{0.99}{2} = 0.495,$$

¹[https://en.wikipedia.org/wiki/Standard_normal_table#Cumulative_from_mean_\(0_to_Z\)](https://en.wikipedia.org/wiki/Standard_normal_table#Cumulative_from_mean_(0_to_Z))

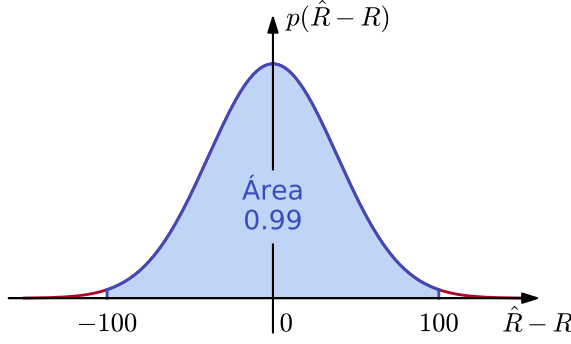


Figura 1.1: Densidad de probabilidad de la variable aleatoria $\hat{R} - R \sim \mathcal{N}(0, \sigma_{\hat{\tau}_0}^2 c^2 / 4)$. Se quiere determinar $\sigma_{\hat{\tau}_0}$ de forma que $\Pr\{|\hat{R} - R| \leq 100\} \geq 0.99$.

y recurriendo a la tabla, se obtiene que

$$\frac{100}{\left(\frac{\sigma_{\hat{\tau}_0} c}{2}\right)} \geq 2.58.$$

Finalmente, despejando se llega a que

$$\sigma_{\hat{\tau}_0} \leq \frac{200}{2.58c} = \frac{200}{2.58 \times 3 \times 10^8} \text{s} \approx 2.6 \times 10^{-7} \text{s} = 0.26 \mu\text{s}.$$

1.1.2. Problema 2

Un parámetro desconocido θ influencia en el resultado de un experimento que se modela mediante la variable aleatoria x . La PDF de x es

$$p(x; \theta) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(x - \theta)^2\right].$$

Se realizan una serie de experimentos, y se observa que x está siempre en el intervalo $[97, 103]$. En base a ello, el investigador concluye que θ debe ser 100. ¿Es correcta esta afirmación?

Solución Considérese el caso en que θ es efectivamente 100. De esta forma,

$$p(x; \theta) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(x - 100)^2\right].$$

La probabilidad de que x esté en el intervalo $[97, 103]$ es

$$\begin{aligned} \Pr\{97 \leq x \leq 103 \mid \theta = 100\} &\stackrel{(a)}{=} \Pr\left\{\frac{97 - \mu}{\sigma} \leq \frac{x - \mu}{\sigma} \leq \frac{103 - \mu}{\sigma} \mid \theta = 100\right\} \\ &\stackrel{(b)}{=} \Pr\left\{-3 \leq \frac{x - \mu}{\sigma} \leq 3\right\} \\ &\stackrel{(c)}{=} 2 \Pr\left\{0 \leq \frac{x - \mu}{\sigma} \leq 3\right\} \\ &\stackrel{(d)}{=} 2 \times 0.49865, \end{aligned}$$

donde en (a) se normalizó la variable aleatoria x como $(x - \mu)/\sigma \sim \mathcal{N}(0, 1)$, siendo μ y σ la media y la desviación estándar de x respectivamente, en (b) se consideró que $\mu = \theta = 100$ y $\sigma = 1$, en (c) que la densidad de probabilidad normal es simétrica y el valor en (d) se obtiene de la tabla de la distribución normal acumulada desde la media. Se concluye que

$$\Pr\{97 \leq x \leq 103 \mid \theta = 100\} = 0.9973$$

en un único experimento, por lo que el valor $\theta = 100$ es altamente probable.

Considérese el caso en que el valor verdadero de θ es por ejemplo 99. En este caso, se cumple que

$$\begin{aligned}\Pr\{97 \leq x \leq 103 \mid \theta = 99\} &= \Pr\left\{\frac{97 - \mu}{\sigma} \leq \frac{x - \mu}{\sigma} \leq \frac{103 - \mu}{\sigma} \mid \theta = 99\right\} \\ &\stackrel{(a)}{=} \Pr\left\{-2 \leq \frac{x - \mu}{\sigma} \leq 4\right\} \\ &= \Pr\left\{0 \leq \frac{x - \mu}{\sigma} \leq 2\right\} + \Pr\left\{0 \leq \frac{x - \mu}{\sigma} \leq 4\right\} \\ &= 0.47725 + 0.49997,\end{aligned}$$

donde en (a) se consideró que $\mu = \theta = 99$ y $\sigma = 1$, resultando en

$$\Pr\{97 \leq x \leq 103 \mid \theta = 99\} = 0.97722$$

en un único experimento. Este razonamiento muestra que la obtención de x en el intervalo $[97, 103]$ en un único experimento tiene probabilidad muy alta tanto si $\theta = 100$ o $\theta = 99$, y por lo tanto, el valor verdadero de θ podría ser cualquiera de ellos. Se concluye que la afirmación de que θ debe ser 100 no es correcta.

Supóngase ahora que el experimento se realiza 100 veces. Asumiendo que los experimentos son independientes, se tiene que

$$\begin{aligned}\Pr\{x \in [97, 103] \text{ en los 100 experimentos} \mid \theta = 100\} &= (\Pr\{97 \leq x \leq 103 \mid \theta = 100\})^{100} \\ &= 0.9973^{100} \\ &\approx 0.76,\end{aligned}$$

mientras que

$$\begin{aligned}\Pr\{x \in [97, 103] \text{ en los 100 experimentos} \mid \theta = 99\} &= (\Pr\{97 \leq x \leq 103 \mid \theta = 99\})^{100} \\ &= 0.97722^{100} \\ &\approx 0.1.\end{aligned}$$

Como la probabilidad de obtener x en el intervalo $[97, 103]$ en los 100 experimentos es considerablemente mas alta si $\theta = 100$ que si $\theta = 99$, podría afirmarse con mayor certeza que θ debe ser 100.

1.1.3. Problema 3

Sea $\mathbf{x} = \theta + \mathbf{w}$, donde \mathbf{w} es una variable aleatoria con PDF $p_w(w)$. Si θ es un parámetro determinístico, encontrar la PDF de \mathbf{x} en términos de p_w , la cual se denotará $p(x; \theta)$. Luego, asumiendo que θ es una variable aleatoria independiente de \mathbf{w} , encontrar la PDF condicional $p(x \mid \theta)$. Finalmente, determinar $p(x \mid \theta)$ sin asumir que θ y \mathbf{w} son independientes. ¿Qué se puede decir respecto a $p(x; \theta)$ y $p(x \mid \theta)$?²

Solución En el caso en que θ es un parámetro determinístico, la PDF de \mathbf{x} es la PDF de \mathbf{w} desplazada una cantidad θ hacia la derecha, es decir,

$$p(x; \theta) = p_w(x - \theta).$$

Para encontrar $p(x \mid \theta)$ en el caso en que θ y \mathbf{w} son variables aleatorias independientes, puede partirse

²En este problema, las letras en negrita representan variables aleatorias escalares, a diferencia del resto del documento, en donde las letras en negrita representan variables aleatorias vectoriales.

de la distribución de probabilidad condicional,

$$\begin{aligned}
F_{x|\theta}(x|\theta) &= \Pr\{\mathbf{x} \leq x | \boldsymbol{\theta} = \theta\} \\
&= \Pr\{\boldsymbol{\theta} + \mathbf{w} \leq x | \boldsymbol{\theta} = \theta\} \\
&\stackrel{(a)}{=} \Pr\{\mathbf{w} \leq x - \theta | \boldsymbol{\theta} = \theta\} \\
&\stackrel{(b)}{=} \frac{\Pr\{\mathbf{w} \leq x - \theta, \boldsymbol{\theta} = \theta\}}{\Pr\{\boldsymbol{\theta} = \theta\}} \\
&\stackrel{(c)}{=} \frac{\Pr\{\mathbf{w} \leq x - \theta\} \Pr\{\boldsymbol{\theta} = \theta\}}{\Pr\{\boldsymbol{\theta} = \theta\}} \\
&= \Pr\{\mathbf{w} \leq x - \theta\} \\
&= F_w(x - \theta),
\end{aligned}$$

donde en (a) se substituyó la variable aleatoria $\boldsymbol{\theta}$ por el valor dado por el condicional, en (b) se aplicó la definición de probabilidad condicional y en (c) se consideró que las variables aleatorias \mathbf{w} y $\boldsymbol{\theta}$ son independientes. Derivando respecto a x , se obtiene la PDF buscada,

$$p(x|\theta) = p_w(x - \theta).$$

Este razonamiento no es completamente formal debido a que el evento $\{\boldsymbol{\theta} = \theta\}$ tiene en general probabilidad nula. Para formalizarlo, habría que considerar el evento $\{\theta \leq \boldsymbol{\theta} \leq \theta + \Delta\theta\}$ y tomar el límite $\Delta\theta \rightarrow 0$. El mismo razonamiento puede hacerse directamente a partir de las PDF, pero se eligió incluirlo por hacerlo mas explícito. Usando las densidades, la deducción es como sigue,

$$p(x|\theta) = \frac{p_{x\theta}(x, \theta)}{p(\theta)} = \frac{p_{w\theta}(x - \theta, \theta)}{p(\theta)} = \frac{p_w(x - \theta)p(\theta)}{p(\theta)} = p_w(x - \theta).$$

Se concluye que en el caso en que θ y \mathbf{w} son independientes,

$$p(x; \theta) = p(x|\theta).$$

Se considera a continuación el caso en que θ y \mathbf{w} no son independientes. Nuevamente, partiendo de la distribución condicional, se cumple que

$$\begin{aligned}
F_{x|\theta}(x|\theta) &= \Pr\{\mathbf{x} \leq x | \boldsymbol{\theta} = \theta\} \\
&= \Pr\{\boldsymbol{\theta} + \mathbf{w} \leq x | \boldsymbol{\theta} = \theta\} \\
&= \Pr\{\mathbf{w} \leq x - \theta | \boldsymbol{\theta} = \theta\} \\
&= F_{w|\theta}(x - \theta | \theta),
\end{aligned}$$

y derivando respecto a x se llega a que

$$p(x|\theta) = p_{w|\theta}(x - \theta | \theta),$$

que en general no es igual a $p(x; \theta)$.

1.1.4. Problema 4

Se quiere estimar el valor A del nivel de corriente continua (*direct current*, DC) en ruido blanco gaussiano (*white gaussian noise*, WGN),

$$x[n] = A + w[n], \quad n = 0, 1, \dots, N - 1,$$

donde $w[n]$ es no correlacionado con media nula, y cada muestra tiene varianza $\sigma^2 = 1$. Se consideran los dos estimadores

$$\begin{aligned}
\hat{A} &= \frac{1}{N} \sum_{n=0}^{N-1} x[n] \\
\check{A} &= \frac{1}{N+2} \left(2x[0] + \sum_{n=1}^{N-2} x[n] + 2x[N-1] \right)
\end{aligned}$$

¿Cual de los dos es mejor? La respuesta anterior, ¿depende del valor de A ?

Solución Se comenzará analizando el sesgo de ambos estimadores. Para el estimador \hat{A} se cumple que

$$E(\hat{A}) = E\left(\frac{1}{N} \sum_{n=0}^{N-1} x[n]\right) = \frac{1}{N} \sum_{n=0}^{N-1} E(x[n]) = \frac{1}{N} \sum_{n=0}^{N-1} A = \frac{1}{N} NA = A, \quad (1.1)$$

y para el estimador \check{A} se tiene que

$$\begin{aligned} E(\check{A}) &= \frac{1}{N+2} \left[2E(x[0]) + \sum_{n=1}^{N-2} E(x[n]) + 2E(x[N-1]) \right] \\ &= \frac{1}{N+2} \left(2A + \sum_{n=1}^{N-2} A + 2A \right) \\ &= \frac{1}{N+2} (2A + (N-2)A + 2A) \\ &= \frac{1}{N+2} (N+2)A \\ &= A. \end{aligned}$$

Se concluye que ambos estimadores son insesgados. Teniendo esto en cuenta, el mejor estimador es el de menor varianza. La varianza del estimador \hat{A} es

$$\begin{aligned} \text{var}(\hat{A}) &= \text{var}\left(\frac{1}{N} \sum_{n=0}^{N-1} x[n]\right) \\ &\stackrel{(a)}{=} \frac{1}{N^2} \sum_{n=0}^{N-1} \text{var}(x[n]) \\ &\stackrel{(b)}{=} \frac{1}{N^2} \sum_{n=0}^{N-1} \text{var}(w[n]) \\ &= \frac{1}{N^2} \sum_{n=0}^{N-1} \sigma^2 \\ &= \frac{1}{N^2} N\sigma^2 \\ &= \frac{\sigma^2}{N}, \end{aligned} \quad (1.2)$$

donde en (a) se empleó que las variables aleatorias $x[n]$ son independientes debido a que las variables aleatorias $w[n]$ son independientes y que si a y b son constantes y X es una variable aleatoria se cumple que $\text{var}(aX + b) = a^2 \text{var}(X)$. Esta propiedad de la varianza también se empleó en (b). Teniendo en cuenta que en este caso $\sigma^2 = 1$, se obtiene que

$$\text{var}(\hat{A}) = \frac{1}{N}.$$

La varianza de \check{A} es

$$\begin{aligned} \text{var}(\check{A}) &= \frac{1}{(N+2)^2} \left[4\text{var}(x[0]) + \sum_{n=1}^{N-2} \text{var}(x[n]) + 4\text{var}(x[N-1]) \right] \\ &= \frac{1}{(N+2)^2} [4\text{var}(w[0]) + (N-2)\text{var}(w[n]) + 4\text{var}(w[N-1])] \\ &= \frac{1}{(N+2)^2} [4 + (N-2) + 4] \\ &= \frac{N+6}{(N+2)^2}. \end{aligned}$$

Notando que

$$\begin{aligned}
 \text{var}(\check{A}) - \text{var}(\hat{A}) &= \frac{N+6}{(N+2)^2} - \frac{1}{N} \\
 &= \frac{(N+6)N - (N+2)^2}{N(N+2)^2} \\
 &= \frac{(N^2 + 6N) - (N^2 + 4N + 4)}{N(N+2)^2} \\
 &= \frac{2N - 4}{N(N+2)^2} \\
 &> 0 \quad \text{si } N \geq 3,
 \end{aligned}$$

se concluye que el estimador \hat{A} es mejor por tener menor varianza. Como la varianza de los estimadores es independiente de A , el desempeño del estimador \hat{A} es mejor para todo valor de A .

1.1.5. Problema 5

Para los mismos datos del problema anterior se propone el siguiente estimador:

$$\hat{A} = \begin{cases} x[0], & \text{si } \frac{A}{\sigma^2} = A^2 > 1000 \\ \frac{1}{N} \sum_{n=0}^{N-1} x[n], & \text{si } \frac{A}{\sigma^2} = A^2 \leq 1000. \end{cases}$$

La fundamentación de este estimador es que para relaciones señal a ruido (*signal-to-noise ratio*, SNR) suficientemente altas, no se necesita reducir el efecto del ruido mediante el promediado y de esta forma, se reduce el costo computacional. Comentar sobre este enfoque.

Solución \hat{A} no es un estimador, ya que para implementarlo se requiere el conocimiento de A para calcular la SNR A/σ^2 .

Capítulo 2

Estimación insesgada de varianza mínima

2.1. Problemas

2.1.1. Problema 1

Se observan los datos $\{x[0], x[1], \dots, x[N-1]\}$, donde las muestras $x[n]$ son independientes e idénticamente distribuidas (IID) como $\mathcal{N}(0, \sigma^2)$. Se quiere estimar la varianza σ^2 como

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{n=0}^{N-1} x^2[n].$$

¿Es este estimador insesgado? Calcular la varianza de $\hat{\sigma}^2$ e indicar que ocurre cuando $N \rightarrow \infty$.

Solución Para verificar si el estimador es insesgado hay que confirmar que $E(\hat{\sigma}^2) = \sigma^2$. En este caso, se cumple que

$$E(\hat{\sigma}^2) = \frac{1}{N} \sum_{n=0}^{N-1} E(x^2[n]) \stackrel{(a)}{=} \frac{1}{N} \sum_{n=0}^{N-1} \sigma^2 = \frac{1}{N} N \sigma^2 = \sigma^2,$$

donde en (a) se tuvo en cuenta que $\sigma^2 \triangleq \text{var}(x[n]) = E(x^2[n]) - E^2(x[n]) = E(x^2[n])$ ya que los $x[n]$ tienen media nula. Se concluye que $\hat{\sigma}^2$ es insesgado.

La varianza de $\hat{\sigma}^2$ es

$$\begin{aligned} \text{var}(\hat{\sigma}^2) &= \frac{1}{N^2} \sum_{n=0}^{N-1} \text{var}(x^2[n]) \\ &\stackrel{(a)}{=} \frac{1}{N^2} \sum_{n=0}^{N-1} 2\sigma^4 \\ &= \frac{1}{N^2} 2N\sigma^4 \\ &= \frac{2\sigma^4}{N}, \end{aligned}$$

donde en (a) se tuvo en cuenta que

$$\text{var}(x^2[n]) = E(x^4[n]) - E^2(x^2[n]) = 3\sigma^4 - \sigma^4 = 2\sigma^4,$$

ya que los momentos centrales segundo y cuarto de una variable aleatoria normal son σ^2 y $3\sigma^4$ respectivamente (ver el capítulo 5 de [2], por ejemplo). Además, cuando $N \rightarrow \infty$, se cumple que

$$\lim_{N \rightarrow \infty} \text{var}(\hat{\sigma}^2) = \lim_{N \rightarrow \infty} \frac{2\sigma^4}{N} = 0.$$

Como el estimador es insesgado, esto implica que la PDF de los datos se concentra cada vez mas en torno al valor verdadero del parámetro a medida que se incrementa la cantidad de datos. Un estimador con esta característica se llama *consistente*, como se verá mas detalladamente en el problema de la sección 2.1.8.

2.1.2. Problema 2

Se consideran los datos $\{x[0], x[1], \dots, x[N-1]\}$, donde cada muestra tiene distribución $\mathcal{U}(0, \theta)$ y son IID. Encontrar un estimador de θ . El rango de θ es $0 < \theta < \infty$.

Solución Una posibilidad es buscar un estimador a partir de la media muestral \bar{x} . Puede verse que $E(x[n]) = \theta/2$,

$$E(x[n]) = \int_{-\infty}^{\infty} xp(x; \theta) dx = \frac{1}{\theta} \int_0^{\theta} x dx = \frac{1}{\theta} \left(\frac{x^2}{2} \Big|_0^{\theta} \right) = \frac{1}{\theta} \left(\frac{\theta^2}{2} - 0 \right) = \frac{\theta}{2}.$$

Por lo tanto, la esperanza de la media muestral también es $\theta/2$,

$$E(\bar{x}) = \frac{1}{N} \sum_{n=0}^{N-1} E(x[n]) = \frac{1}{N} \sum_{n=0}^{N-1} \frac{\theta}{2} = \frac{1}{N} N \frac{\theta}{2} = \frac{\theta}{2}.$$

Como $2E(\bar{x}) = \theta$, un estimador insesgado de θ es

$$\hat{\theta} = 2\bar{x}.$$

2.1.3. Problema 3

Considérese las observaciones

$$x[n] = A + w[n], \quad n = 0, 1, \dots, N-1,$$

donde A es el parámetro a estimar y $w[n]$ es WGN. El parámetro A puede tomar cualquier valor en el intervalo $-\infty < A < \infty$. Un estimador razonable de A es la media muestral,

$$\hat{A} = \frac{1}{N} \sum_{n=0}^{N-1} x[n].$$

Demostrar que la PDF de \hat{A} es $\mathcal{N}(A, \sigma^2/N)$.

Solución \hat{A} es una variable aleatoria gaussiana por ser una combinación lineal de variables aleatorias gaussianas (ver los capítulos 6 y 7 de [2]), por lo que para determinar su PDF basta con calcular la media y la varianza. La media y la varianza de \hat{A} ya se calcularon en la sección 1.1.4 (ecuaciones 1.1 y 1.2 respectivamente), y son

$$E(\hat{A}) = A \quad \text{var}(\hat{A}) = \frac{\sigma^2}{N}.$$

Se concluye que $\hat{A} \sim \mathcal{N}(A, \sigma^2/N)$.

2.1.4. Problema 4

La tasa del latido del corazón de un paciente se mide automáticamente con una computadora cada 100 ms. En 1 s las medidas $\{\hat{h}_1, \hat{h}_2, \dots, \hat{h}_{10}\}$ se promedian para obtener \hat{h} . Si $E(\hat{h}_i) = \alpha h$ para alguna constante α y $\text{var}(\hat{h}_i) = 1$ para cada i , determinar si el promediado mejora el estimador cuando $\alpha = 1$ y $\alpha = 1/2$. Asumir que las medidas son no correlacionadas.

Solución Se comenzará estudiando el sesgo de \hat{h} en función del parámetro α . Como

$$\hat{h} = \frac{1}{10} \sum_{i=1}^{10} \hat{h}_i,$$

la esperanza de \hat{h} es

$$E(\hat{h}) = \frac{1}{10} \sum_{i=1}^{10} E(\hat{h}_i) = \frac{1}{10} \sum_{i=1}^{10} \alpha h = \frac{1}{10} 10 \alpha h = \alpha h.$$

Se concluye que el estimador \hat{h} es insesgado solo si $\alpha = 1$. Además, la varianza de \hat{h} es

$$\text{var}(\hat{h}) = \text{var}\left(\frac{1}{10} \sum_{i=1}^{10} \hat{h}_i\right) \stackrel{(a)}{=} \frac{1}{10^2} \sum_{i=1}^{10} \text{var}(\hat{h}_i) = \frac{1}{100} \sum_{i=1}^{10} 1 = \frac{1}{100} 10 = \frac{1}{10},$$

donde en (a) se consideró que las medidas \hat{h}_i son no correlacionadas. Este resultado indica que el promediado de las 10 medidas reduce la varianza 1/10 respecto a una medida individual. En la figura 2.1

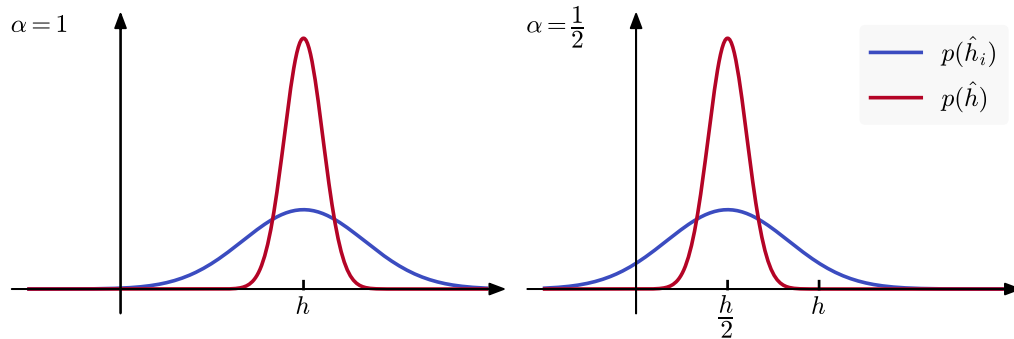


Figura 2.1: PDF de una medida individual \hat{h}_i y de las medidas promediadas \hat{h} para $\alpha = 1$ y $\alpha = 1/2$ en el caso gaussiano.

se muestran las PDF en el caso gaussiano. Con $\alpha = 1/2$, el promediado hace que la PDF esté más concentrada en torno al valor incorrecto de h , haciendo que la probabilidad de que \hat{h} esté cerca de h se reduzca. En el caso con $\alpha = 1$ en que el estimador es insesgado, el promediado es favorable, ya que incrementa la probabilidad de que \hat{h} esté cercano al valor verdadero h .

2.1.5. Problema 5

Dos muestras $\{x[0], x[1]\}$ son observadas independientemente de una distribución $\mathcal{N}(0, \sigma^2)$. El estimador

$$\hat{\sigma}^2 = \frac{1}{2}(x^2[0] + x^2[1])$$

es insesgado. Encontrar la PDF de $\hat{\sigma}^2$ para determinar si es simétrica respecto a σ^2 .

Solución En el problema de la sección 2.1.1 se demostró que el estimador $\hat{\sigma}^2$ es insesgado. Por otro lado, puede demostrarse que si x es una variable aleatoria $\mathcal{N}(0, 1)$, x^2 es una variable aleatoria chi-cuadrado con un grado de libertad (ver el capítulo 5 de [2]), es decir, $x^2 \sim \chi^2(1)$. Además, se cumple que si

$$z = x_1^2 + \dots + x_n^2$$

donde $x_i, i = 1, \dots, n$, son variables aleatorias $\mathcal{N}(0, 1)$ independientes, la variable aleatoria z es chi-cuadrado con n grados de libertad (ver el capítulo 7 de [2]). La PDF de una variable aleatoria $z \sim \chi^2(n)$

es

$$p(z) = \begin{cases} \frac{z^{n/2-1}}{2^{n/2}\Gamma(n/2)} e^{-z/2}, & z \geq 0 \\ 0, & \text{en otro caso} \end{cases}, \quad (2.1)$$

donde $\Gamma(\alpha)$ es la función gama, definida como

$$\Gamma(\alpha) = \int_0^\infty z^{\alpha-1} e^{-z} dz. \quad (2.2)$$

Puede demostrarse que si n es un número entero positivo, se cumple que $\Gamma(n) = (n-1)!$ (ver el capítulo 4 de [2]). Como

$$x[n] \sim \mathcal{N}(0, \sigma^2) \quad \Rightarrow \quad \frac{x[n]}{\sigma} \sim \mathcal{N}(0, 1) \quad \Rightarrow \quad \frac{x^2[n]}{\sigma^2} \sim \chi^2(1),$$

y por lo tanto,

$$y = \frac{1}{\sigma^2} (x^2[0] + x^2[1]) \sim \chi^2(2).$$

De esta forma, sustituyendo $n = 2$ en la ecuación 2.1, la PDF $p(y)$ de la variable aleatoria y es

$$p(y) = \begin{cases} \frac{e^{-y/2}}{2\Gamma(1)}, & y \geq 0 \\ 0, & \text{en otro caso} \end{cases},$$

y sustituyendo $\alpha = 1$ en la función gamma (ecuación 2.2) se tiene que

$$\Gamma(1) = \int_0^\infty e^{-y} dy = -e^{-y} \Big|_0^\infty = -(0 - 1) = 1.$$

Se concluye que

$$p(y) = \frac{e^{-y/2}}{2}, \quad y \geq 0.$$

Teniendo en cuenta que

$$\hat{\sigma}^2 = \frac{\sigma^2}{2} y,$$

la PDF de $\hat{\sigma}^2$ puede calcularse a partir de la PDF de y mediante el teorema fundamental de transformaciones de variables aleatorias (ver el capítulo 5 de [2]), que indica que si x es una variable aleatoria con PDF $p_x(x)$, la PDF $p_y(y)$ de la variable aleatoria $y = g(x)$ es

$$p_y(y) = \sum_i \frac{p_x(x_i)}{|g'(x_i)|}, \quad (2.3)$$

donde x_i son las raíces de $y = g(x)$ para un valor específico de y , es decir, $x_i = g^{-1}(y)$ para el valor de y que se quiere calcular la PDF. Como en este caso la transformación es inyectiva, con

$$y = \frac{2}{\sigma^2} \hat{\sigma}^2,$$

y la derivada de la transformación es $\hat{\sigma}^{2'}(y) = \sigma^2/2$, se obtiene que

$$p(\hat{\sigma}^2) = \frac{p(2\hat{\sigma}^2/\sigma^2)}{\sigma^2/2} = \frac{e^{-\hat{\sigma}^2/\sigma^2}}{\sigma^2}, \quad \hat{\sigma}^2 \geq 0,$$

es decir, $\hat{\sigma}^2$ tiene PDF exponencial con parámetro $\lambda = 1/\sigma^2$. Claramente, la PDF no es simétrica respecto a la media σ^2 , como se muestra en la figura 2.2.

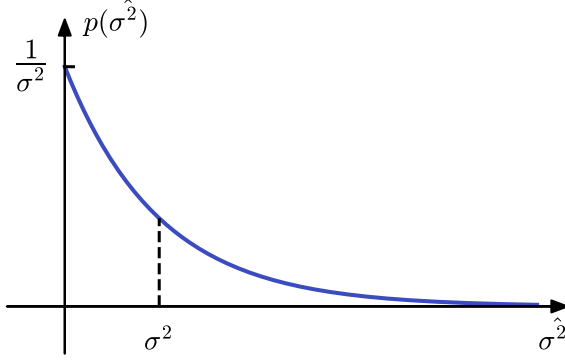


Figura 2.2: Densidad de probabilidad de la variable aleatoria $\hat{\sigma}^2$.

2.1.6. Problema 6

Para el problema de estimación del nivel de DC en WGN (ver sección 1.1.4) se propone el siguiente estimador mas genérico:

$$\hat{A} = \sum_{n=0}^{N-1} a_n x[n].$$

Encontrar los valores de los a_n de forma que el estimador sea insesgado y minimicen la varianza. Sugerencia: usar multiplicadores de Lagrange usando la condición de no sesgo como la ecuación de restricción.

Solución Para encontrar los puntos estacionarios de una función $f(x)$ sujeto a las condiciones de igualdad $g_i(x) = 0$, $i = 1, \dots, m$, hay que encontrar los puntos estacionarios de la función Lagrangiana¹, dada por

$$\mathcal{L}(x, \lambda) = f(x) - \sum_{i=1}^m \lambda_i g_i(x).$$

La función a minimizar es la varianza de \hat{A} ,

$$\text{var}(\hat{A}) = \sum_{n=0}^{N-1} a_n^2 \text{var}(x[n]) = \sigma^2 \sum_{n=0}^{N-1} a_n^2,$$

y la restricción es que \hat{A} sea insesgado es

$$E(\hat{A}) = \sum_{n=0}^{N-1} a_n E(x[n]) = A \sum_{n=0}^{N-1} a_n = A,$$

donde la última igualdad es la condición de insesgado, y resulta en

$$\sum_{n=0}^{N-1} a_n = 1.$$

Por lo tanto, la función Lagrangiana resulta en este caso en

$$\mathcal{L}(a_0, \dots, a_{N-1}, \lambda) = \sigma^2 \sum_{n=0}^{N-1} a_n^2 - \lambda \left(\sum_{n=0}^{N-1} a_n - 1 \right).$$

Los puntos estacionarios de $\mathcal{L}(a_0, \dots, a_{N-1}, \lambda)$ se obtienen derivando respecto a los a_i , $i = 0, \dots, N-1$, igualando a 0 y resolviendo el sistema de N ecuaciones con N incógnitas,

$$\frac{\partial \mathcal{L}(a_0, \dots, a_{N-1}, \lambda)}{\partial a_i} = 2\sigma^2 a_i - \lambda = 0 \quad \Rightarrow \quad a_i = \frac{\lambda}{2\sigma^2}, \quad i = 0, \dots, N-1. \quad (2.4)$$

¹Ver https://en.wikipedia.org/wiki/Lagrange_multiplier

Esto indica que los a_i tienen todos el mismo valor. Sustituyendo los resultados en la ecuación de la restricción, se obtiene el valor de λ ,

$$\sum_{n=0}^{N-1} a_n = \sum_{n=0}^{N-1} \frac{\lambda}{2\sigma^2} = \frac{N\lambda}{2\sigma^2} = 1 \quad \Rightarrow \quad \lambda = \frac{2\sigma^2}{N}$$

y sustituyendo el valor de λ en las ecuaciones 2.4 de los coeficientes se llega a que

$$a_i = \frac{1}{N}, \quad i = 0, \dots, N-1.$$

Se concluye que el estimador \hat{A} insesgado que minimiza la varianza es la media muestral.

2.1.7. Problema 7

Se proponen dos estimadores insesgados cuyas varianzas cumplen que $\text{var}(\hat{\theta}) < \text{var}(\check{\theta})$. Si ambos estimadores son gaussianos, demostrar que

$$\Pr \{|\hat{\theta} - \theta| > \epsilon\} < \Pr \{|\check{\theta} - \theta| > \epsilon\}$$

para todo $\epsilon > 0$. Este resultado indica que el estimador de menor varianza es mejor ya que su PDF está mas concentrada en torno al valor verdadero.

Solución Como los estimadores son insesgados, se cumple que $E(\hat{\theta}) = E(\check{\theta}) = \theta$, y por lo tanto, $\hat{\theta} - \theta \sim \mathcal{N}(0, \text{var}(\hat{\theta}))$ y $\check{\theta} - \theta \sim \mathcal{N}(0, \text{var}(\check{\theta}))$. De esta forma,

$$\begin{aligned} \Pr \{|\hat{\theta} - \theta| > \epsilon\} &= \Pr \{\hat{\theta} - \theta > \epsilon\} + \Pr \{\hat{\theta} - \theta < -\epsilon\} \\ &= 2 \Pr \{\hat{\theta} - \theta > \epsilon\} \\ &= 2 \Pr \left\{ \frac{\hat{\theta} - \theta}{\sqrt{\text{var}(\hat{\theta})}} > \frac{\epsilon}{\sqrt{\text{var}(\hat{\theta})}} \right\}, \quad \text{con } \frac{\hat{\theta} - \theta}{\sqrt{\text{var}(\hat{\theta})}} \sim \mathcal{N}(0, 1). \end{aligned}$$

Realizando el mismo razonamiento para $\check{\theta}$, se obtiene que

$$\Pr \{|\check{\theta} - \theta| > \epsilon\} = 2 \Pr \left\{ \frac{\check{\theta} - \theta}{\sqrt{\text{var}(\check{\theta})}} > \frac{\epsilon}{\sqrt{\text{var}(\check{\theta})}} \right\}, \quad \text{con } \frac{\check{\theta} - \theta}{\sqrt{\text{var}(\check{\theta})}} \sim \mathcal{N}(0, 1).$$

Considérese la distribución de probabilidad acumulada complementaria de una variable aleatoria normal estándar, definida como

$$f(z) = \Pr\{Z > z\} = \frac{1}{\sqrt{2\pi}} \int_z^\infty e^{-t^2/2} dt. \quad (2.5)$$

De esta forma,

$$\Pr \{|\hat{\theta} - \theta| > \epsilon\} = 2f\left(\frac{\epsilon}{\sqrt{\text{var}(\hat{\theta})}}\right) \quad \text{y} \quad \Pr \{|\check{\theta} - \theta| > \epsilon\} = 2f\left(\frac{\epsilon}{\sqrt{\text{var}(\check{\theta})}}\right).$$

$f(z)$ es una función decreciente con z , y por lo tanto, como

$$\text{var}(\hat{\theta}) < \text{var}(\check{\theta}) \quad \Rightarrow \quad \frac{\epsilon}{\sqrt{\text{var}(\hat{\theta})}} > \frac{\epsilon}{\sqrt{\text{var}(\check{\theta})}}$$

se cumple que

$$f\left(\frac{\epsilon}{\sqrt{\text{var}(\hat{\theta})}}\right) < f\left(\frac{\epsilon}{\sqrt{\text{var}(\check{\theta})}}\right),$$

concluyendo que

$$\Pr \{|\hat{\theta} - \theta| > \epsilon\} < \Pr \{|\check{\theta} - \theta| > \epsilon\}.$$

En la figura 2.3 se muestran todas las PDF involucradas en el problema. Las áreas sombreadas son las probabilidades.

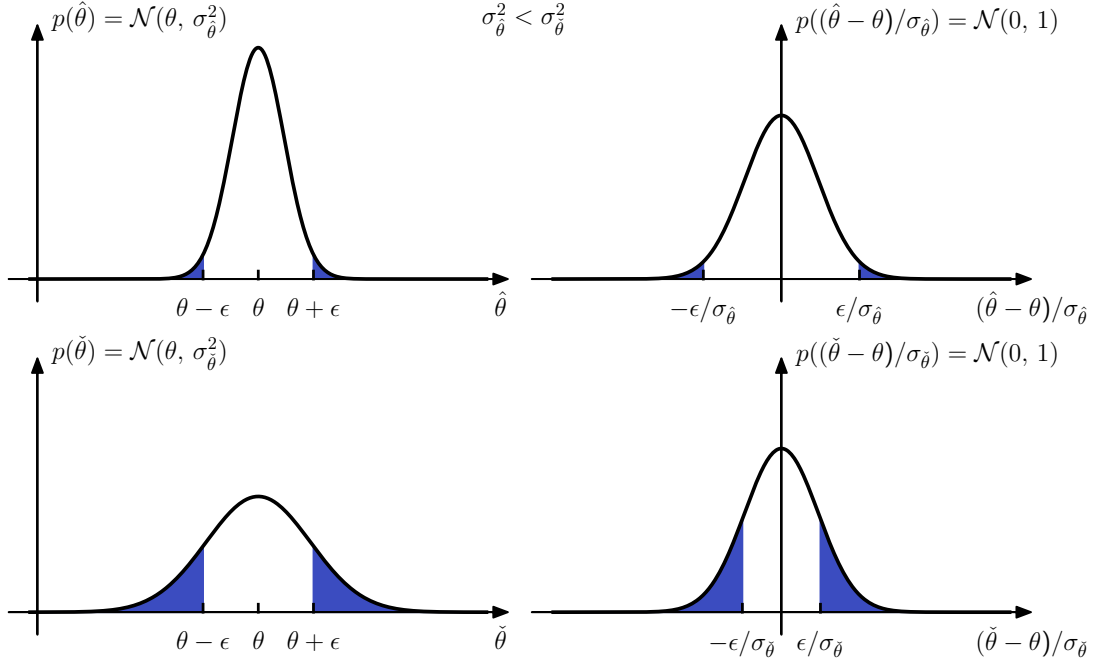


Figura 2.3: PDF y PDF normalizada de $\hat{\theta}$ (arriba) y $\check{\theta}$ (abajo). Las áreas sombreadas son $\Pr\{|\hat{\theta} - \theta| > \epsilon\}$ y $\Pr\{|\check{\theta} - \theta| > \epsilon\}$.

2.1.8. Problema 8

Para el problema de estimación del nivel de DC en WGN, usar el resultado del problema de la sección 2.1.3 para demostrar que $\hat{A} \rightarrow A$ cuando $N \rightarrow \infty$. Para hacerlo, demostrar que

$$\lim_{N \rightarrow \infty} \Pr\{|\hat{A} - A| > \epsilon\} = 0$$

para todo $\epsilon > 0$. En ese caso, se dice que el estimador \hat{A} es *consistente*. Investigar que sucede si se emplea el estimador alternativo

$$\check{A} = \frac{1}{2N} \sum_{n=0}^{N-1} x[n].$$

Solución En el problema de la sección 2.1.3, se demostró que $\hat{A} \sim \mathcal{N}(A, \sigma^2/N)$. Por lo tanto, siguiendo un razonamiento análogo al del problema anterior se tiene que

$$\Pr\{|\hat{A} - A| > \epsilon\} = 2 \Pr\{\hat{A} - A > \epsilon\} = 2 \Pr\left\{\frac{\hat{A} - A}{\sigma/\sqrt{N}} > \frac{\epsilon}{\sigma/\sqrt{N}}\right\} = 2f\left(\frac{\epsilon}{\sigma/\sqrt{N}}\right),$$

donde la función f es la distribución de probabilidad acumulada complementaria de una variable aleatoria normal estándar, definida en la ecuación 2.5. Como

$$\forall \epsilon > 0, \quad \lim_{N \rightarrow \infty} \frac{\epsilon}{\sigma/\sqrt{N}} = \infty \quad \text{y además} \quad \lim_{z \rightarrow \infty} f(z) = \lim_{z \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \int_z^{\infty} e^{-t^2/2} dt = 0,$$

se concluye que

$$\lim_{N \rightarrow \infty} \Pr\{|\hat{A} - A| > \epsilon\} = 0 \quad \forall \epsilon > 0. \quad (2.6)$$

Esta ecuación es la definición de *convergencia en probabilidad* de una variable aleatoria (ver el capítulo 7 de [2]), es decir, cuando se verifica la ecuación 2.6, se dice que la variable aleatoria \hat{A} converge en

probabilidad a la constante A . Cuando el estimador de un parámetro converge en probabilidad al valor verdadero del parámetro, se dice que el estimador es *consistente*.

En el caso del estimador alternativo \check{A} , se cumple que

$$E(\check{A}) = \frac{1}{2N} \sum_{n=0}^{N-1} E(x[n]) = \frac{1}{2N} NA = \frac{A}{2},$$

es decir, el estimador no es insesgado. Por otro lado, la varianza del estimador es

$$\text{var}(\check{A}) = \frac{1}{4N^2} \sum_{n=0}^{N-1} \text{var}(x[n]) = \frac{1}{4N^2} N\sigma^2 = \frac{\sigma^2}{4N}.$$

Se concluye que $\check{A} \sim \mathcal{N}(A/2, \sigma^2/4N)$. Por lo tanto,

$$\lim_{N \rightarrow \infty} \text{var}(\check{A}) = \lim_{N \rightarrow \infty} \frac{\sigma^2}{4N} = 0,$$

lo que implica que la PDF de \check{A} se concentra cada vez mas en torno a su media $A/2$. De esta forma,

$$\lim_{N \rightarrow \infty} \Pr \{ |\check{A} - A| > \epsilon \} = 1$$

para todo $\epsilon \in (0, A/2)$, como se muestra en la figura 2.4. Por lo tanto, \check{A} no converge en probabilidad a A , así que no es consistente.

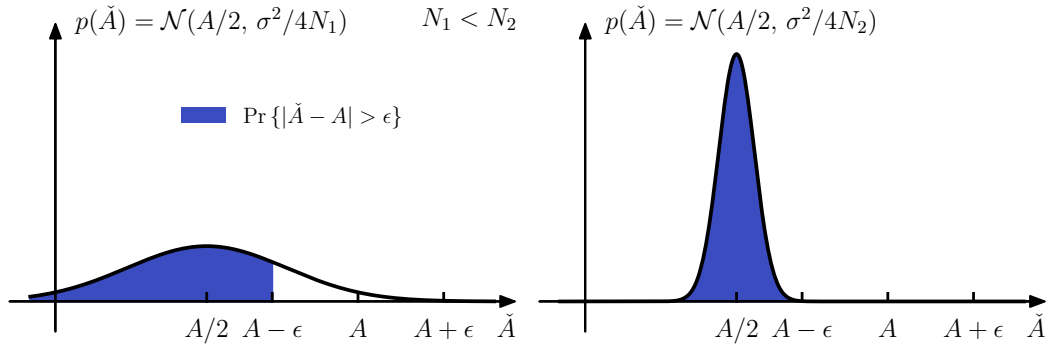


Figura 2.4: PDF de $\check{A} \sim \mathcal{N}(A/2, \sigma^2/4N)$ para dos valores N_1 y N_2 de N con $N_1 < N_2$. El área sombreada es $\Pr \{ |\check{A} - A| > \epsilon \}$, que tiende a 1 cuando $N \rightarrow \infty$ si $\epsilon \in (0, A/2)$.

2.1.9. Problema 9

Este problema ilustra lo que ocurre con un estimador insesgado cuando se le aplica una transformación no lineal. En el problema de estimación de nivel de DC en WGN, se elige estimar el parámetro $\theta = A^2$ como

$$\hat{\theta} = \left(\frac{1}{N} \sum_{n=0}^{N-1} x[n] \right)^2.$$

¿Este estimador es insesgado? ¿Qué ocurre cuando $N \rightarrow \infty$?

Solución Para verificar si el estimador es insesgado, hay que calcular su esperanza y ver si coincide con el valor verdadero del parámetro. Para hacerlo, se parte observando que

$$\begin{aligned}
\hat{\theta} &= \left(\frac{1}{N} \sum_{n=0}^{N-1} x[n] \right)^2 \\
&= \left(\frac{1}{N} \sum_{m=0}^{N-1} x[m] \right) \left(\frac{1}{N} \sum_{n=0}^{N-1} x[n] \right) \\
&= \frac{1}{N^2} \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} x[m]x[n] \\
&= \frac{1}{N^2} \sum_{n=0}^{N-1} x^2[n] + \frac{1}{N^2} \sum_{m=0}^{N-1} \sum_{\substack{n=0 \\ n \neq m}}^{N-1} x[m]x[n],
\end{aligned}$$

y por lo tanto,

$$\begin{aligned}
E(\hat{\theta}) &= \frac{1}{N^2} \sum_{n=0}^{N-1} E(x^2[n]) + \frac{1}{N^2} \sum_{m=0}^{N-1} \sum_{\substack{n=0 \\ n \neq m}}^{N-1} E(x[m]x[n]) \\
&\stackrel{(a)}{=} \frac{1}{N^2} \sum_{n=0}^{N-1} E(x^2[n]) + \frac{1}{N^2} \sum_{m=0}^{N-1} \sum_{\substack{n=0 \\ n \neq m}}^{N-1} E(x[m])E(x[n]) \\
&\stackrel{(b)}{=} \frac{1}{N^2} \sum_{n=0}^{N-1} (A^2 + \sigma^2) + \frac{1}{N^2} \sum_{m=0}^{N-1} \sum_{\substack{n=0 \\ n \neq m}}^{N-1} A^2 \\
&\stackrel{(c)}{=} \frac{1}{N^2} N(A^2 + \sigma^2) + \frac{1}{N^2} N(N-1)A^2 \\
&= A^2 + \frac{\sigma^2}{N},
\end{aligned}$$

donde en (a) se consideró que como el ruido $w[n]$ es blanco, $x[m]$ y $x[n]$ son no correlacionadas si $m \neq n$ y por lo tanto $E(x[m]x[n]) = E(x[m])E(x[n])$, en (b) se tuvo en cuenta que

$$\text{var}(x[n]) = E(x^2[n]) - E^2(x[n]) \quad \Rightarrow \quad E(x^2[n]) = E^2(x[n]) + \text{var}(x[n]) = A^2 + \sigma^2$$

y en (c) que la primer sumatoria tiene N sumandos y la segunda doble sumatoria tiene $N(N-1)$ sumandos. Se concluye que

$$E(\hat{\theta}) = A^2 + \frac{\sigma^2}{N} \neq A^2,$$

y por lo tanto $\hat{\theta}$ no es un estimador insesgado. Sin embargo, se cumple que

$$\lim_{N \rightarrow \infty} E(\hat{\theta}) = \lim_{N \rightarrow \infty} A^2 + \frac{\sigma^2}{N} = A^2,$$

es decir, $\hat{\theta}$ es insesgado con $N \rightarrow \infty$. Cuando esto ocurre, se dice que el estimador es *asintóticamente insesgado*.

El cálculo de $E(\hat{\theta})$ se puede realizar de forma mas directa considerando que $\hat{\theta} = \hat{A}^2$, con $\hat{A} \sim \mathcal{N}(A, \sigma^2/N)$, como se calculó en el problema de la sección 2.1.3. De esta forma

$$E(\hat{\theta}) = E(\hat{A}^2) \quad \text{y} \quad \text{var}(\hat{A}) = E(\hat{A}^2) - E^2(\hat{A}) \quad \Rightarrow \quad E(\hat{A}^2) = E^2(\hat{A}) + \text{var}(\hat{A}) = A^2 + \frac{\sigma^2}{N}.$$

2.1.10. Problema 10

En el problema de estimación de nivel de DC en WGN, se asume ahora que además del parámetro A también se desconoce el parámetro σ^2 . Se desea estimar el parámetro vectorial

$$\boldsymbol{\theta} = \begin{bmatrix} A \\ \sigma^2 \end{bmatrix}.$$

¿Es el estimador

$$\hat{\boldsymbol{\theta}} = \begin{bmatrix} \hat{A} \\ \hat{\sigma}^2 \end{bmatrix} = \begin{bmatrix} \frac{1}{N} \sum_{n=0}^{N-1} x[n] \\ \frac{1}{N-1} \sum_{n=0}^{N-1} (x[n] - \hat{A})^2 \end{bmatrix}$$

insesgado?

Solución Como ya se demostró en el problema de la sección 1.1.4, $E(\hat{A}) = A$, es decir, la media muestral \hat{A} es un estimador insesgado de A . Falta ver si se cumple que $E(\hat{\sigma}^2) = \sigma^2$. Se comienza observando que

$$E(\hat{\sigma}^2) = \frac{1}{N-1} \sum_{n=0}^{N-1} E \left[(x[n] - \hat{A})^2 \right] \quad (2.7)$$

La esperanza en la sumatoria se desarrolla como

$$E \left[(x[n] - \hat{A})^2 \right] = E(x^2[n]) - 2E(x[n]\hat{A}) + E(\hat{A}^2). \quad (2.8)$$

Como $x[n] \sim \mathcal{N}(A, \sigma^2)$, el primer sumando es

$$E(x^2[n]) = \text{var}(x[n]) + E^2(x[n]) = \sigma^2 + A^2.$$

Además, como $\hat{A} \sim \mathcal{N}(A, \sigma^2/N)$ (ver el problema de la sección 2.1.3), el tercer sumando es

$$E(\hat{A}^2) = \text{var}(\hat{A}) + E^2(\hat{A}) = \frac{\sigma^2}{N} + A^2.$$

El segundo sumando puede calcularse como sigue:

$$\begin{aligned} E(x[n]\hat{A}) &= E \left(x[n] \frac{1}{N} \sum_{m=0}^{N-1} x[m] \right) \\ &= \frac{1}{N} \sum_{m=0}^{N-1} E(x[n]x[m]) \\ &= \frac{1}{N} \left[E(x^2[n]) + \sum_{\substack{m=0 \\ m \neq n}}^{N-1} E(x[n])E(x[m]) \right] \\ &= \frac{1}{N} [(\sigma^2 + A^2) + (N-1)A^2] \\ &= \frac{\sigma^2}{N} + A^2. \end{aligned}$$

Sustituyendo estos resultados en la ecuación 2.8, se tiene que

$$E \left[(x[n] - \hat{A})^2 \right] = (\sigma^2 + A^2) - 2 \left(\frac{\sigma^2}{N} + A^2 \right) + \left(\frac{\sigma^2}{N} + A^2 \right) = \frac{N-1}{N} \sigma^2,$$

y finalmente, sustituyendo este resultado en la ecuación 2.7 se llega a que

$$E(\hat{\sigma}^2) = \frac{1}{N-1} \sum_{n=0}^{N-1} \frac{N-1}{N} \sigma^2 = \sigma^2.$$

Se concluye que $E(\hat{\boldsymbol{\theta}}) = \boldsymbol{\theta}$, es decir, el estimador vectorial es insesgado.

2.1.11. Problema 11

Dada una única observación $x[0]$ de distribución $\mathcal{U}[0, 1/\theta]$, se quiere estimar θ . Se asume que $\theta > 0$. Mostrar que para que un estimador $\hat{\theta} = g(x[0])$ sea insesgado, se tiene que cumplir que

$$\int_0^{1/\theta} g(u) du = 1.$$

Luego, mostrar que no existe una función g que cumpla esta condición para todo $\theta > 0$.

Solución La PDF de la observación es

$$x[0] \sim p(x[0]) = \mathcal{U}[0, 1/\theta] = \begin{cases} \theta, & 0 \leq x[0] \leq 1/\theta \\ 0, & \text{en otro caso} \end{cases},$$

y por lo tanto, la esperanza del estimador es

$$E(\hat{\theta}) = E(g(x[0])) = \int_{-\infty}^{\infty} g(x[0])p(x[0]) dx[0] = \int_0^{1/\theta} g(x[0])\theta dx[0] = \theta \int_0^{1/\theta} g(x[0]) dx[0]$$

Para que el estimador sea insesgado se tiene que cumplir que $E(\hat{\theta}) = \theta$, y por lo tanto

$$\theta \int_0^{1/\theta} g(x[0]) dx[0] = \theta \quad \Rightarrow \quad \int_0^{1/\theta} g(u) du = 1,$$

que es lo se quería demostrar.

A continuación, se demostrará que no se puede encontrar una función g que cumpla esta condición para todo θ . Para hacerlo, se considera θ_1 y θ_2 tal que $\theta_2 < \theta_1$. De esta forma, hay que encontrar una función g tal que,

$$\int_0^{1/\theta_1} g(u) du = 1, \quad \int_0^{1/\theta_2} g(u) du = 1.$$

Restando ambas ecuaciones, se tiene que

$$\int_0^{1/\theta_2} g(u) du - \int_0^{1/\theta_1} g(u) du = 0 \quad \Rightarrow \quad \int_{1/\theta_1}^{1/\theta_2} g(u) du = 0 \quad \forall \theta_2 < \theta_1.$$

La única forma que se cumpla esta condición es que $g(u) = 0$ para todo $u > 0$, función que produce un estimador sesgado,

$$E(\hat{\theta}) = E(g(x[0])) = E(0) = 0 \quad \forall x[0], \quad \forall \theta > 0.$$

Capítulo 3

Cota Inferior de Cramer-Rao

3.1. Cota Inferior de Cramer-Rao: parámetro escalar

La cota inferior de Cramer-Rao (*Cramer-Rao lower bond*, CRLB) es el valor mínimo que puede alcanzar la varianza de cualquier estimador.

3.1.1. Teorema

Hipótesis Sea el conjunto de datos \mathbf{x} con PDF $p(\mathbf{x}; \theta)$, donde θ es una constante determinística desconocida. Se asume que $p(\mathbf{x}; \theta)$ satisface la siguiente *condición de regularidad*:

$$E \left[\frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \theta} \right] = 0, \quad \forall \theta,$$

donde la esperanza se toma respecto a $p(\mathbf{x}; \theta)$.

Tesis La varianza de cualquier estimador insesgado $\hat{\theta}$ cumple que

$$\text{var}(\hat{\theta}) \geq \frac{1}{I(\theta)}, \quad (3.1)$$

donde

$$I(\theta) = -E \left[\frac{\partial^2 \ln p(\mathbf{x}; \theta)}{\partial \theta^2} \right], \quad (3.2)$$

función que es referida como la *información de Fisher* de los datos \mathbf{x} . En $I(\theta)$, la derivada se evalúa en el valor verdadero de θ y la esperanza se toma respecto a $p(\mathbf{x}; \theta)$.

Además, existe un estimador que alcanza la cota para todo θ si y solo si

$$\frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \theta} = I(\theta)(g(\mathbf{x}) - \theta) \quad (3.3)$$

para alguna función g . Este estimador, que es el estimador *insesgado de varianza mínima* (*minimum variance unbiased estimator*, MVU), es $\hat{\theta} = g(\mathbf{x})$ y su varianza es $1/I(\theta)$.

Demostración Previo a la demostración, se examinará la condición de regularidad,

$$E \left[\frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \theta} \right] = 0,$$

la cual se asume que se cumple. Observando que

$$\begin{aligned}
E \left[\frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \theta} \right] &= \int \frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \theta} p(\mathbf{x}; \theta) d\mathbf{x} \\
&= \int \frac{1}{p(\mathbf{x}; \theta)} \frac{\partial p(\mathbf{x}; \theta)}{\partial \theta} p(\mathbf{x}; \theta) d\mathbf{x} \\
&= \int \frac{\partial p(\mathbf{x}; \theta)}{\partial \theta} d\mathbf{x} \\
&= \frac{\partial}{\partial \theta} \int p(\mathbf{x}; \theta) d\mathbf{x} \\
&= \frac{\partial 1}{\partial \theta} \\
&= 0,
\end{aligned}$$

se concluye que la condición de regularidad se cumple siempre y cuando sea posible intercambiar el orden de integración y diferenciación. Esto en general puede hacerse excepto cuando el soporte de la PDF depende del parámetro θ (ver el problema de la sección 3.6.1).

Para la demostración del teorema, considérense las siguientes variables aleatorias:

$$y = \hat{\theta}(\mathbf{x}) - \theta, \quad z = \frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \theta}, \quad (3.4)$$

las cuales tienen media nula, y por ser $\hat{\theta}$ insesgado y z porque se cumple la condición de regularidad. Se demostrará que $E(yz) = 1$,

$$\begin{aligned}
E(yz) &= E \left[(\hat{\theta}(\mathbf{x}) - \theta) \frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \theta} \right] \\
&= E \left[\hat{\theta}(\mathbf{x}) \frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \theta} \right] - \theta E \left[\frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \theta} \right] \\
&\stackrel{(a)}{=} \int \hat{\theta}(\mathbf{x}) \frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \theta} p(\mathbf{x}; \theta) d\mathbf{x} \\
&= \int \hat{\theta}(\mathbf{x}) \frac{\partial p(\mathbf{x}; \theta)}{\partial \theta} d\mathbf{x} \\
&\stackrel{(b)}{=} \frac{\partial}{\partial \theta} \int \hat{\theta}(\mathbf{x}) p(\mathbf{x}; \theta) d\mathbf{x} \\
&= \frac{\partial}{\partial \theta} E[\hat{\theta}(\mathbf{x})] \\
&= \frac{\partial \theta}{\partial \theta} \\
&= 1,
\end{aligned}$$

donde en (a) se tuvo en cuenta que el segundo sumando es la condición de regularidad y por lo tanto es nulo y además se aplicó la definición de esperanza, y en (b) se intercambió el orden de integración y diferenciación considerando que $\hat{\theta}(\mathbf{x})$ no depende de θ . Como $E(yz) = 1$, también se cumple que $E^2(yz) = 1$, así que

$$\begin{aligned}
1 &= E^2(yz) \\
&\stackrel{(a)}{\leq} E(y^2)E(z^2) \\
&= E \left[\left(\hat{\theta}(\mathbf{x}) - \theta \right)^2 \right] E \left[\left(\frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \theta} \right)^2 \right] \\
&\stackrel{(b)}{=} -E \left[\left(\hat{\theta}(\mathbf{x}) - \theta \right)^2 \right] E \left[\frac{\partial^2 \ln p(\mathbf{x}; \theta)}{\partial \theta^2} \right] \\
&= \text{var}(\hat{\theta})I(\theta),
\end{aligned}$$

resultando en

$$\text{var}(\hat{\theta}) \geq \frac{1}{I(\theta)},$$

que es la primera parte de la tesis. En esta deducción, en (a) se empleó el análogo de la desigualdad de Cauchy-Schwarz para variables aleatorias, que indica que si y y z son variables aleatorias,

$$E^2(yz) \leq E(y^2)E(z^2). \quad (3.5)$$

Para demostrarlo (ver el capítulo 6 de [2]) se observa que si c es una constante

$$E[(cy - z)^2] = E[c^2y^2 - 2czy + z^2] = c^2E(y^2) - 2cE(yz) + E(z^2) \geq 0, \quad \forall c,$$

ya que la variable aleatoria $(cy - z)^2$ es no negativa y por lo tanto, su media también es no negativa. Por lo tanto, el discriminante del polinomio en c debe ser negativo o nulo,

$$\Delta = 4E^2(yz) - 4E(y^2)E(z^2) \leq 0 \quad \Leftrightarrow \quad E^2(yz) \leq E(y^2)E(z^2).$$

Además, en (b) se consideró que

$$E \left[\left(\frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \theta} \right)^2 \right] = -E \left[\frac{\partial^2 \ln p(\mathbf{x}; \theta)}{\partial \theta^2} \right]. \quad (3.6)$$

Esta igualdad proviene de la condición de regularidad. Efectivamente, la condición de regularidad indica que

$$\int \frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \theta} p(\mathbf{x}; \theta) d\mathbf{x} = 0, \quad \forall \theta.$$

Diferenciando respecto a θ se observa que

$$\begin{aligned} 0 &= \frac{\partial}{\partial \theta} \int \frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \theta} p(\mathbf{x}; \theta) d\mathbf{x} \\ &= \int \frac{\partial}{\partial \theta} \left[\frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \theta} p(\mathbf{x}; \theta) \right] d\mathbf{x} \\ &= \int \left[\frac{\partial^2 \ln p(\mathbf{x}; \theta)}{\partial \theta^2} p(\mathbf{x}; \theta) + \frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \theta} \frac{\partial p(\mathbf{x}; \theta)}{\partial \theta} \right] d\mathbf{x} \\ &\stackrel{(a)}{=} \int \left[\frac{\partial^2 \ln p(\mathbf{x}; \theta)}{\partial \theta^2} p(\mathbf{x}; \theta) + \frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \theta} \frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \theta} p(\mathbf{x}; \theta) \right] d\mathbf{x}, \end{aligned}$$

donde en (a) se tomó en cuenta que

$$\frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \theta} = \frac{1}{p(\mathbf{x}; \theta)} \frac{\partial p(\mathbf{x}; \theta)}{\partial \theta} \quad \Rightarrow \quad \frac{\partial p(\mathbf{x}; \theta)}{\partial \theta} = \frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \theta} p(\mathbf{x}; \theta). \quad (3.7)$$

Por lo tanto,

$$\int \left(\frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \theta} \right)^2 p(\mathbf{x}; \theta) d\mathbf{x} = - \int \frac{\partial^2 \ln p(\mathbf{x}; \theta)}{\partial \theta^2} p(\mathbf{x}; \theta) d\mathbf{x},$$

que es la ecuación 3.6.

Para demostrar la segunda parte de la tesis, se considera la condición de igualdad en la desigualdad de Cauchy-Schwarz (ecuación 3.5). En la situación del teorema, las variables aleatorias y y z , definidas en la ecuación 3.4 son funciones de los datos \mathbf{x} y del parámetro θ . La condición de igualdad en la ecuación 3.5 es que $cy = z$, donde c puede depender de θ pero no de los datos \mathbf{x} , ya que

$$E^2(yz) \stackrel{(a)}{=} E^2[c(\theta)y^2] = c^2(\theta)E^2(y^2)$$

y

$$E(y^2)E(z^2) \stackrel{(a)}{=} E(y^2)E[c^2(\theta)y^2] = c^2(\theta)E^2(y^2),$$

donde en (a) se substituyó $z = c(\theta)y$, obteniendo que $E^2(yz) = E(y^2)E(z^2)$. De esta forma, a partir de las definiciones de y y z , la condición de igualdad es

$$\frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \theta} = c(\theta) [\hat{\theta}(\mathbf{x}) - \theta].$$

Para llegar a la ecuación 3.3, hay que demostrar que $c(\theta) = I(\theta)$. Esto puede hacerse derivando respecto a θ ,

$$\frac{\partial^2 \ln p(\mathbf{x}; \theta)}{\partial \theta^2} = \frac{\partial c(\theta)}{\partial \theta} [\hat{\theta}(\mathbf{x}) - \theta] - c(\theta),$$

y tomando esperanza,

$$E \left[\frac{\partial^2 \ln p(\mathbf{x}; \theta)}{\partial \theta^2} \right] = \frac{\partial c(\theta)}{\partial \theta} \left\{ E [\hat{\theta}(\mathbf{x})] - \theta \right\} - c(\theta) \stackrel{(a)}{=} -c(\theta),$$

donde en (a) se tuvo en cuenta que el estimador es insesgado. Como el lado izquierdo de la igualdad es $-I(\theta)$ (ecuación 3.2), se concluye que $c(\theta) = I(\theta)$, finalizando la demostración.

En el caso en que la PDF de los datos está parametrizada por θ y se desee encontrar un estimador insesgado $\hat{\alpha}$ de $\alpha = g(\theta)$, puede demostrarse de forma análoga que la CRLB está dada por

$$\text{var}(\hat{\alpha}) = \frac{\left(\frac{\partial g(\theta)}{\partial \theta} \right)^2}{-E \left[\frac{\partial^2 \ln p(\mathbf{x}; \theta)}{\partial \theta^2} \right]}. \quad (3.8)$$

Ejemplo: parámetros de ruido de Laplace

Se consideran las observaciones $\{x[0], x[1], \dots, x[N-1]\}$ IID con distribución de Laplace,

$$p(x; \mu, \sigma) = \frac{1}{\sqrt{2}\sigma} \exp \left(-\frac{\sqrt{2}|x - \mu|}{\sigma} \right).$$

La PDF de las observaciones es

$$\begin{aligned} p(\mathbf{x}; \mu, \sigma) &= \prod_{n=0}^{N-1} \frac{1}{\sqrt{2}\sigma} \exp \left(-\frac{\sqrt{2}|x[n] - \mu|}{\sigma} \right) \\ &= \frac{1}{(\sqrt{2}\sigma)^N} \exp \left(-\frac{\sqrt{2}}{\sigma} \sum_{n=0}^{N-1} |x[n] - \mu| \right), \end{aligned}$$

por lo que la función de verosimilitud logarítmica es

$$\ln p(\mathbf{x}; \mu, \sigma) = -N \ln \sqrt{2} - N \ln \sigma - \frac{\sqrt{2}}{\sigma} \sum_{n=0}^{N-1} |x[n] - \mu|.$$

Supóngase que la media μ es conocida y se quiere encontrar el estimador MVU de la desviación estándar σ . La derivada respecto al parámetro es

$$\frac{\partial \ln p(\mathbf{x}; \sigma)}{\partial \sigma} = -\frac{N}{\sigma} + \frac{\sqrt{2}}{\sigma^2} \sum_{n=0}^{N-1} |x[n] - \mu|,$$

y observando que se puede factorizar de la forma de la ecuación 3.3 como

$$\frac{\partial \ln p(\mathbf{x}; \sigma)}{\partial \sigma} = \frac{N}{\sigma^2} \left(\frac{\sqrt{2}}{N} \sum_{n=0}^{N-1} |x[n] - \mu| - \sigma \right),$$

se concluye que el estimador MVU es

$$\hat{\sigma} = \frac{\sqrt{2}}{N} \sum_{n=0}^{N-1} |x[n] - \mu|$$

y la CRLB es

$$\text{var}(\hat{\sigma}) \geq \frac{\sigma^2}{N}.$$

En este caso, como se cumple la condición de la ecuación 3.3, el estimador es eficiente, es decir, alcanza la CRLB para todo valor del parámetro σ .

Supóngase ahora que la desviación estándar σ es conocida y se quiere estimar la media μ . La derivada de la función de verosimilitud logarítmica es

$$\begin{aligned} \frac{\partial \ln p(\mathbf{x}; \mu)}{\partial \mu} &= -\frac{\sqrt{2}}{\sigma} \sum_{n=0}^{N-1} \frac{\partial |x[n] - \mu|}{\partial \mu} \\ &= \frac{\sqrt{2}}{\sigma} \sum_{n=0}^{N-1} \text{sig}(x[n] - \mu), \end{aligned}$$

donde se tuvo en cuenta que $\partial |x[n] - \mu| / \partial \mu = -\text{sig}(x[n] - \mu)$. Esta función no parece poder factorizarse en la forma de la ecuación 3.3, y por lo tanto, mediante esta técnica no puede encontrarse un estimador MVU. Para determinar la CRLB, se deriva nuevamente,

$$\begin{aligned} \frac{\partial^2 \ln p(\mathbf{x}; \mu)}{\partial \mu^2} &= \frac{\sqrt{2}}{\sigma} \sum_{n=0}^{N-1} \frac{\partial \text{sig}(x[n] - \mu)}{\partial \mu} \\ &= -\frac{2\sqrt{2}}{\sigma} \sum_{n=0}^{N-1} \delta(x[n] - \mu), \end{aligned}$$

donde se tuvo en cuenta que $\partial \text{sig}(x[n] - \mu) / \partial \mu = -2\delta(x[n] - \mu)$. Tomando esperanza, se tiene que la información de Fisher es

$$\begin{aligned} I(\mu) &= -E \left[\frac{\partial^2 \ln p(\mathbf{x}; \mu)}{\partial \mu^2} \right] = \frac{2\sqrt{2}}{\sigma} \sum_{n=0}^{N-1} E[\delta(x[n] - \mu)] \\ &\stackrel{(a)}{=} \frac{2\sqrt{2}}{\sigma} \left(N \frac{1}{\sqrt{2}\sigma} \right) \\ &= \frac{2N}{\sigma^2} \end{aligned}$$

donde en (a) se consideró que

$$\begin{aligned} E[\delta(x[n] - \mu)] &= \int_{-\infty}^{\infty} \delta(x[n] - \mu) \frac{1}{\sqrt{2}\sigma} \exp \left(-\frac{\sqrt{2}|x[n] - \mu|}{\sigma} \right) dx[n] \\ &= \frac{1}{\sqrt{2}\sigma} \exp \left(-\frac{\sqrt{2}|x[n] - \mu|}{\sigma} \right) \Big|_{x[n]=\mu} \\ &= \frac{1}{\sqrt{2}\sigma}. \end{aligned}$$

Se concluye que la CLRB es

$$\text{var}(\hat{\mu}) \geq \frac{1}{I(\mu)} = \frac{\sigma^2}{2N}.$$

Ver además el problema de la sección 3.6.2.

3.1.2. Transformación de parámetros

En frecuencia ocurre en la práctica que el parámetro que se quiere estimar es función de otro parámetro fundamental. Si se desea estimar $\alpha = g(\theta)$, la CRLB es

$$\text{var}(\hat{\alpha}) \geq \frac{\left(\frac{\partial g}{\partial \theta}\right)^2}{-E\left[\frac{\partial^2 \ln p(\mathbf{x}; \theta)}{\partial \theta^2}\right]}$$

Demostración Se demostrará que

$$I^{-1}(\alpha) = I^{-1}(\theta) \left(\frac{\partial g}{\partial \theta}\right)^2, \quad (3.9)$$

y por lo tanto, se cumple que

$$\text{var}(\alpha) \geq \frac{1}{I(\alpha)} = \frac{\left(\frac{\partial g}{\partial \theta}\right)^2}{I(\theta)}.$$

Para mostrarlo, empleando la regla de la cadena, se cumple que

$$\frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \theta} = \frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \alpha} \frac{\partial \alpha}{\partial \theta}$$

y derivando nuevamente,

$$\begin{aligned} \frac{\partial^2 \ln p(\mathbf{x}; \theta)}{\partial \theta^2} &= \frac{\partial}{\partial \theta} \left[\frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \alpha} \right] \frac{\partial \alpha}{\partial \theta} + \frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \alpha} \frac{\partial}{\partial \theta} \left[\frac{\partial \alpha}{\partial \theta} \right] \\ &\stackrel{(a)}{=} \frac{\partial}{\partial \alpha} \left[\frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \alpha} \right] \frac{\partial \alpha}{\partial \theta} \frac{\partial \alpha}{\partial \theta} + \frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \alpha} \frac{\partial^2 \alpha}{\partial \theta^2} \\ &= \frac{\partial^2 \ln p(\mathbf{x}; \theta)}{\partial \alpha^2} \left(\frac{\partial \alpha}{\partial \theta}\right)^2 + \frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \alpha} \frac{\partial^2 \alpha}{\partial \theta^2}, \end{aligned}$$

donde en (a) se aplicó nuevamente la regla de la cadena en la derivada del primer sumando. Tomando esperanza en ambos lados de la igualdad, se tiene que

$$E\left[\frac{\partial^2 \ln p(\mathbf{x}; \theta)}{\partial \theta^2}\right] = E\left[\frac{\partial^2 \ln p(\mathbf{x}; \theta)}{\partial \alpha^2}\right] \left(\frac{\partial \alpha}{\partial \theta}\right)^2,$$

resultando en que

$$I(\theta) = I(\alpha) \left(\frac{\partial g}{\partial \theta}\right)^2,$$

que es lo que se quería demostrar.

3.2. Cota Inferior de Cramer-Rao: parámetro vectorial

En esta sección se deriva la CRLB para el caso de un parámetro vectorial $\boldsymbol{\theta}$, es decir, la PDF de los datos \mathbf{x} está caracterizada por un vector $\boldsymbol{\theta}$.

3.2.1. Teorema

Hipótesis Sea el conjunto de datos \mathbf{x} con PDF $p(\mathbf{x}; \boldsymbol{\theta})$, donde $\boldsymbol{\theta}$ es un vector $p \times 1$ constante desconocido. Se asume que $p(\mathbf{x}; \boldsymbol{\theta})$ satisface las *condiciones de regularidad*

$$E\left[\frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right] = \mathbf{0}, \quad \forall \boldsymbol{\theta},$$

donde la esperanza se toma respecto a $p(\mathbf{x}; \boldsymbol{\theta})$.

Tesis La matriz de covarianza de cualquier estimador insesgado $\hat{\theta}$ cumple que

$$\mathbf{C}_{\hat{\theta}} - \mathbf{I}^{-1}(\theta) \geq \mathbf{0}, \quad (3.10)$$

donde $\geq \mathbf{0}$ se interpreta como que la matriz es semidefinida positiva. La *matriz de información de Fisher* $\mathbf{I}(\theta)$ esta dada por

$$[\mathbf{I}(\theta)]_{ij} = -E \left[\frac{\partial^2 \ln p(\mathbf{x}; \theta)}{\partial \theta_i \partial \theta_j} \right], \quad i = 1, \dots, p, \quad j = 1, \dots, p, \quad (3.11)$$

donde las derivadas se evalúan en el valor verdadero de θ y la esperanza se toma respecto a $p(\mathbf{x}; \theta)$.

Además, existe un estimador insesgado que alcanza la cota, en el sentido de que $\mathbf{C}_{\hat{\theta}} = \mathbf{I}^{-1}(\theta)$ para todo θ si y solo si

$$\frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \theta} = \mathbf{I}(\theta)(\mathbf{g}(\mathbf{x}) - \theta) \quad (3.12)$$

para alguna función $\mathbf{g} \subset \mathbb{R}^p$ y alguna matriz \mathbf{I} $p \times p$. Este estimador, que es el estimador MVU, es $\hat{\theta} = \mathbf{g}(\mathbf{x})$ y su matriz de covarianza es $\mathbf{I}^{-1}(\theta)$.

Como los elementos de la diagonal de una matriz semidefinida positiva son no negativos, se cumple que

$$[\mathbf{C}_{\hat{\theta}} - \mathbf{I}^{-1}(\theta)]_{ii} \geq 0,$$

y por lo tanto,

$$\text{var}(\hat{\theta}_i) = [\mathbf{C}_{\hat{\theta}}]_{ii} \geq [\mathbf{I}^{-1}(\theta)]_{ii}. \quad (3.13)$$

Demostración Se considerará el caso mas general en que el parámetro a estimar es una función $\alpha = \mathbf{g}(\theta)$ del parámetro vectorial θ^1 . La PDF de los datos \mathbf{x} está caracterizada por θ . El parámetro θ tiene dimensiones $p \times 1$, α tiene dimensiones $r \times 1$ y $\mathbf{g} : \mathbb{R}^p \rightarrow \mathbb{R}^r$. Se tendrán en cuenta los estimadores insesgados,

$$E(\hat{\alpha}_i) = \alpha_i = [\mathbf{g}(\theta)]_i, \quad i = 1, \dots, r.$$

Se comienza notando que se cumple que

$$\int (\hat{\alpha}_i - \alpha_i) \frac{\partial \ln p(\mathbf{x}, \theta)}{\partial \theta_i} p(\mathbf{x}, \theta) d\mathbf{x} = \frac{\partial [\mathbf{g}(\theta)]_i}{\partial \theta_i}.$$

Esto es consecuencia de las condiciones de regularidad, ya que

$$\begin{aligned} \int (\hat{\alpha}_i - \alpha_i) \frac{\partial \ln p(\mathbf{x}, \theta)}{\partial \theta_i} p(\mathbf{x}, \theta) d\mathbf{x} &= \int \hat{\alpha}_i \frac{\partial \ln p(\mathbf{x}, \theta)}{\partial \theta_i} p(\mathbf{x}, \theta) d\mathbf{x} - \int \alpha_i \frac{\partial \ln p(\mathbf{x}, \theta)}{\partial \theta_i} p(\mathbf{x}, \theta) d\mathbf{x} \\ &\stackrel{(a)}{=} \int \hat{\alpha}_i \frac{\partial p(\mathbf{x}, \theta)}{\partial \theta_i} d\mathbf{x} - \alpha_i \int \frac{\partial \ln p(\mathbf{x}, \theta)}{\partial \theta_i} p(\mathbf{x}, \theta) d\mathbf{x} \\ &\stackrel{(b)}{=} \frac{\partial}{\partial \theta_i} \int \hat{\alpha}_i p(\mathbf{x}, \theta) d\mathbf{x} - \alpha_i E \left[\frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \theta_i} \right] \\ &\stackrel{(c)}{=} \frac{\partial E(\hat{\alpha}_i)}{\partial \theta_i} - 0 \\ &= \frac{\partial [\mathbf{g}(\theta)]_i}{\partial \theta_i}, \end{aligned}$$

donde en (a) se aplicó la igualdad equivalente de la ecuación 3.7 del caso escalar y además se consideró que α_i es una constante por lo que puede sacarse de la integral, en (b) se intercambié el orden de integración y diferenciación teniendo en cuenta que $\hat{\alpha}_i$ depende de los datos \mathbf{x} pero no de θ_i y en (c) se consideró que el segundo sumando es la condición de regularidad y por lo tanto, es nulo. Realizando un razonamiento análogo para $j \neq i$, puede verse que

$$\int (\hat{\alpha}_i - \alpha_i) \frac{\partial \ln p(\mathbf{x}, \theta)}{\partial \theta_j} p(\mathbf{x}, \theta) d\mathbf{x} = \frac{\partial [\mathbf{g}(\theta)]_i}{\partial \theta_j}.$$

¹No debe confundirse esta función \mathbf{g} , que es una transformación del parámetro θ , con la función \mathbf{g} de la ecuación 3.12, que es el estimador MVU. Si bien la notación puede ser confusa, se pretende seguir lo mas fielmente posible la empleada en [1]

Combinando estos dos resultados en notación matricial, se llega a que

$$\int (\hat{\alpha} - \alpha) \frac{\partial \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T p(\mathbf{x}, \boldsymbol{\theta}) d\mathbf{x} = \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}},$$

donde

$$\hat{\alpha} - \alpha = \begin{bmatrix} \hat{\alpha}_1 - \alpha_1 \\ \vdots \\ \hat{\alpha}_r - \alpha_r \end{bmatrix} \quad \frac{\partial \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \frac{\partial \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial \theta_1} \\ \vdots \\ \frac{\partial \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial \theta_p} \end{bmatrix} \quad \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \frac{\partial g_1(\boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial g_1(\boldsymbol{\theta})}{\partial \theta_2} & \cdots & \frac{\partial g_1(\boldsymbol{\theta})}{\partial \theta_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_r(\boldsymbol{\theta})}{\partial \theta_1} & \frac{\partial g_r(\boldsymbol{\theta})}{\partial \theta_2} & \cdots & \frac{\partial g_r(\boldsymbol{\theta})}{\partial \theta_p} \end{bmatrix}.$$

Premultiplicando por \mathbf{a}^T y postmultiplicando \mathbf{b} ambos lados de la igualdad, donde \mathbf{a} y \mathbf{b} son vectores arbitrarios de dimensiones $r \times 1$ y $p \times 1$ respectivamente, se obtiene que

$$\int \mathbf{a}^T (\hat{\alpha} - \alpha) \frac{\partial \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T \mathbf{b} p(\mathbf{x}, \boldsymbol{\theta}) d\mathbf{x} = \mathbf{a}^T \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathbf{b}, \quad (3.14)$$

donde ahora la ecuación es escalar. Definiendo las variables aleatorias escalares

$$y(\mathbf{x}) = \mathbf{a}^T (\hat{\alpha} - \alpha), \quad z(\mathbf{x}) = \frac{\partial \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T \mathbf{b} \quad (3.15)$$

se tiene que

$$\begin{aligned} y(\mathbf{x})z(\mathbf{x}) &= \mathbf{a}^T (\hat{\alpha} - \alpha) \frac{\partial \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T \mathbf{b} \\ y(\mathbf{x})y^T(\mathbf{x}) &= y^2(\mathbf{x}) = \mathbf{a}^T (\hat{\alpha} - \alpha)(\hat{\alpha} - \alpha)^T \mathbf{a} \\ z^T(\mathbf{x})z(\mathbf{x}) &= z^2(\mathbf{x}) = \mathbf{b}^T \frac{\partial \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T \mathbf{b}, \end{aligned}$$

donde se consideró que $y(\mathbf{x}) = y^T(\mathbf{x})$ y $z(\mathbf{x}) = z^T(\mathbf{x})$ por ser escalares. Tomando la esperanza se obtiene que

$$\begin{aligned} E[y(\mathbf{x})z(\mathbf{x})] &= \int \mathbf{a}^T (\hat{\alpha} - \alpha) \frac{\partial \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T \mathbf{b} p(\mathbf{x}, \boldsymbol{\theta}) d\mathbf{x} \stackrel{(a)}{=} \mathbf{a}^T \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathbf{b} \\ E[y^2(\mathbf{x})] &= E[\mathbf{a}^T (\hat{\alpha} - \alpha)(\hat{\alpha} - \alpha)^T \mathbf{a}] = \mathbf{a}^T E[(\hat{\alpha} - \alpha)(\hat{\alpha} - \alpha)^T] \mathbf{a} \stackrel{(b)}{=} \mathbf{a}^T \mathbf{C}_{\hat{\alpha}} \mathbf{a} \\ E[z^2(\mathbf{x})] &= E\left[\mathbf{b}^T \frac{\partial \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T \mathbf{b}\right] = \mathbf{b}^T E\left[\frac{\partial \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T\right] \mathbf{b} \stackrel{(c)}{=} \mathbf{b}^T \mathbf{I}(\boldsymbol{\theta}) \mathbf{b}, \end{aligned} \quad (3.16)$$

donde la igualdad en (a) proviene de la ecuación 3.14, en (b) se consideró que como $E(\hat{\alpha}) = \alpha$, $\mathbf{C}_{\hat{\alpha}} = E[(\hat{\alpha} - \alpha)(\hat{\alpha} - \alpha)^T]$ es la matriz de covarianza de $\hat{\alpha}$ y en (c) se consideró que como en el caso escalar, se cumple que

$$E\left[\frac{\partial \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial \theta_i} \frac{\partial \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial \theta_j}\right] = -E\left[\frac{\partial^2 \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial \theta_i \partial \theta_j}\right] = [\mathbf{I}(\boldsymbol{\theta})]_{ij}, \quad (3.17)$$

donde $\mathbf{I}(\boldsymbol{\theta})$ es la matriz de información de Fisher. Esta igualdad proviene de la condición de regularidad,

$$\int \frac{\partial \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial \theta_i} p(\mathbf{x}, \boldsymbol{\theta}) d\mathbf{x} = 0,$$

ya que diferenciando ambos lados de la igualdad respecto a θ_j , se observa que

$$\begin{aligned}
0 &= \frac{\partial}{\partial \theta_j} \int \frac{\partial \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial \theta_i} p(\mathbf{x}, \boldsymbol{\theta}) d\mathbf{x} \\
&= \int \frac{\partial}{\partial \theta_j} \left[\frac{\partial \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial \theta_i} p(\mathbf{x}, \boldsymbol{\theta}) \right] d\mathbf{x} \\
&= \int \left[\frac{\partial^2 \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} p(\mathbf{x}, \boldsymbol{\theta}) + \frac{\partial \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial \theta_i} \frac{\partial p(\mathbf{x}, \boldsymbol{\theta})}{\partial \theta_j} \right] d\mathbf{x} \\
&= \int \frac{\partial^2 \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} p(\mathbf{x}, \boldsymbol{\theta}) d\mathbf{x} + \int \frac{\partial \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial \theta_i} \frac{\partial \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial \theta_j} p(\mathbf{x}, \boldsymbol{\theta}) d\mathbf{x} \\
&= E \left[\frac{\partial^2 \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \right] + E \left[\frac{\partial \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial \theta_i} \frac{\partial \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial \theta_j} \right].
\end{aligned}$$

Finalmente, aplicando la desigualdad de Cauchy-Schwarz $E^2(yz) \leq E(y^2)E(z^2)$ (ecuación 3.5) en las ecuaciones 3.16 se llega a que

$$\left(\mathbf{a}^T \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathbf{b} \right)^2 \leq \mathbf{a}^T \mathbf{C}_{\hat{\alpha}} \mathbf{a} \mathbf{b}^T \mathbf{I}(\boldsymbol{\theta}) \mathbf{b}.$$

Teniendo en cuenta que \mathbf{b} es arbitrario, definase como

$$\mathbf{b} = \mathbf{I}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T \mathbf{a}. \quad (3.18)$$

De esta forma,

$$\left(\mathbf{a}^T \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathbf{I}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T \mathbf{a} \right)^2 \leq \mathbf{a}^T \mathbf{C}_{\hat{\alpha}} \mathbf{a} \left(\mathbf{a}^T \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} [\mathbf{I}^{-1}(\boldsymbol{\theta})]^T \mathbf{I}(\boldsymbol{\theta}) \mathbf{I}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T \mathbf{a} \right).$$

Considerando que la matriz $\mathbf{I}(\boldsymbol{\theta})$ es simétrica y definida positiva, ya que

$$\mathbf{u}^T E[\mathbf{v}\mathbf{v}^T] \mathbf{u} = E[\mathbf{u}^T \mathbf{v} \mathbf{v}^T \mathbf{u}] = E[(\mathbf{u}^T \mathbf{v})(\mathbf{u}^T \mathbf{v})^T] = E[(\mathbf{u}^T \mathbf{v})^2] > 0, \quad \forall \mathbf{u} > \mathbf{0},$$

también lo es $\mathbf{I}^{-1}(\boldsymbol{\theta})$, y por lo tanto $[\mathbf{I}^{-1}(\boldsymbol{\theta})]^T = \mathbf{I}^{-1}(\boldsymbol{\theta})$, por lo que

$$\left(\mathbf{a}^T \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathbf{I}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T \mathbf{a} \right)^2 \leq \mathbf{a}^T \mathbf{C}_{\hat{\alpha}} \mathbf{a} \left(\mathbf{a}^T \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathbf{I}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T \mathbf{a} \right)$$

y cancelando el término entre paréntesis, se tiene que

$$\mathbf{a}^T \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathbf{I}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T \mathbf{a} \leq \mathbf{a}^T \mathbf{C}_{\hat{\alpha}} \mathbf{a},$$

por lo que

$$\mathbf{a}^T \mathbf{C}_{\hat{\alpha}} \mathbf{a} - \mathbf{a}^T \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathbf{I}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T \mathbf{a} \geq 0,$$

resultando en

$$\mathbf{a}^T \left(\mathbf{C}_{\hat{\alpha}} - \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathbf{I}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T \right) \mathbf{a} \geq 0.$$

Como \mathbf{a} es arbitrario, la desigualdad se cumple para todo \mathbf{a} . Se concluye que la matriz entre paréntesis es semidefinida positiva,

$$\mathbf{C}_{\hat{\alpha}} - \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathbf{I}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T \geq \mathbf{0}. \quad (3.19)$$

En el caso en que no haya una transformación del parámetro, $\boldsymbol{\alpha} = \mathbf{g}(\boldsymbol{\theta}) = \boldsymbol{\theta}$ y $\frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{I}$ obteniéndose la ecuación 3.10, con lo que queda demostrada la primera parte del teorema.

Como se mostró en la sección 3.1.1, la condición de igualdad en la ecuación de Cauchy-Schwarz es que $y(\mathbf{x}) = c(\boldsymbol{\theta})z(\mathbf{x})$, donde $c(\boldsymbol{\theta})$ es una constante que puede depender de $\boldsymbol{\theta}$ pero no de los datos \mathbf{x} . A partir de las definiciones de $y(\mathbf{x})$ y $z(\mathbf{x})$ (ecuación 3.15), la condición es

$$\begin{aligned}\mathbf{a}^T(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) &= c(\boldsymbol{\theta}) \frac{\partial \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T \mathbf{b} \\ &\stackrel{(a)}{=} c(\boldsymbol{\theta}) \frac{\partial \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T \mathbf{I}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T \mathbf{a}\end{aligned}$$

donde en (a) se sustituyó \mathbf{b} a partir de la ecuación 3.18. Esta ecuación es escalar de la forma $\mathbf{a}^T \mathbf{u} = \mathbf{v}^T \mathbf{a}$, por lo que la única forma de que se cumpla para todo \mathbf{a} , ya que \mathbf{a} es arbitrario, es que $\mathbf{u} = \mathbf{v}$, y por lo tanto,

$$\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha} = c(\boldsymbol{\theta}) \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathbf{I}^{-1}(\boldsymbol{\theta}) \frac{\partial \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}.$$

En el caso en que $\boldsymbol{\alpha} = \mathbf{g}(\boldsymbol{\theta}) = \boldsymbol{\theta}$ y $\frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{I}$ se tiene que

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} = c(\boldsymbol{\theta}) \mathbf{I}^{-1}(\boldsymbol{\theta}) \frac{\partial \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}},$$

y por lo tanto, despejando, se obtiene que

$$\frac{\partial \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{1}{c(\boldsymbol{\theta})} \mathbf{I}(\boldsymbol{\theta}) (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}).$$

Para finalizar la prueba, se demostrará que $c(\boldsymbol{\theta}) = 1$ para obtener la ecuación 3.12. El elemento i -ésimo de este vector es

$$\frac{\partial \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial \theta_i} = \sum_{k=1}^p \frac{[\mathbf{I}(\boldsymbol{\theta})]_{ik}}{c(\boldsymbol{\theta})} (\hat{\theta}_k - \theta_k).$$

Diferenciando una vez mas, se ve que

$$\begin{aligned}\frac{\partial^2 \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} &= \frac{\partial}{\partial \theta_j} \left(\sum_{k=1}^p \frac{[\mathbf{I}(\boldsymbol{\theta})]_{ik}}{c(\boldsymbol{\theta})} (\hat{\theta}_k - \theta_k) \right) \\ &= \sum_{k=1}^p \frac{[\mathbf{I}(\boldsymbol{\theta})]_{ik}}{c(\boldsymbol{\theta})} \frac{\partial (\hat{\theta}_k - \theta_k)}{\partial \theta_j} + \sum_{k=1}^p \frac{\partial \left(\frac{[\mathbf{I}(\boldsymbol{\theta})]_{ik}}{c(\boldsymbol{\theta})} \right)}{\partial \theta_j} (\hat{\theta}_k - \theta_k) \\ &\stackrel{(a)}{=} -\frac{[\mathbf{I}(\boldsymbol{\theta})]_{ij}}{c(\boldsymbol{\theta})} + \sum_{k=1}^p \frac{\partial \left(\frac{[\mathbf{I}(\boldsymbol{\theta})]_{ik}}{c(\boldsymbol{\theta})} \right)}{\partial \theta_j} (\hat{\theta}_k - \theta_k),\end{aligned}$$

donde en (a) se tuvo en cuenta que $\partial(\hat{\theta}_k - \theta_k)/\partial \theta_j$ es 0 si $k \neq j$ y -1 si $k = j$. Tomando la esperanza, se tiene que

$$\begin{aligned}E \left[\frac{\partial^2 \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \right] &= E \left[-\frac{[\mathbf{I}(\boldsymbol{\theta})]_{ij}}{c(\boldsymbol{\theta})} + \sum_{k=1}^p \frac{\partial \left(\frac{[\mathbf{I}(\boldsymbol{\theta})]_{ik}}{c(\boldsymbol{\theta})} \right)}{\partial \theta_j} (\hat{\theta}_k - \theta_k) \right] \\ &= -\frac{[\mathbf{I}(\boldsymbol{\theta})]_{ij}}{c(\boldsymbol{\theta})} + \sum_{k=1}^p \frac{\partial \left(\frac{[\mathbf{I}(\boldsymbol{\theta})]_{ik}}{c(\boldsymbol{\theta})} \right)}{\partial \theta_j} [E(\hat{\theta}_k) - \theta_k] \\ &= -\frac{[\mathbf{I}(\boldsymbol{\theta})]_{ij}}{c(\boldsymbol{\theta})},\end{aligned}$$

ya que $E(\hat{\theta}_k) = \theta_k$. Como el lado izquierdo de la igualdad es $-[\mathbf{I}(\boldsymbol{\theta})]_{ij}$, se concluye que $c(\boldsymbol{\theta}) = 1$.

3.2.2. Ejemplo: nivel de DC en WGN

Se considera el problema de estimación del nivel de DC en WGN. El modelo de los datos es

$$x[n] = A + w[n], \quad n = 0, \dots, N-1,$$

donde los parámetros a estimar son el nivel A de DC y la varianza σ^2 del ruido $w[n] \sim \mathcal{N}(0, \sigma^2)$. El parámetro vectorial es $\boldsymbol{\theta} = [A \ \sigma^2]^T$ y por lo tanto, $p = 2$. La matriz 2×2 de información de Fisher es

$$I(\boldsymbol{\theta}) = \begin{bmatrix} -E \left[\frac{\partial^2 \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial A^2} \right] & -E \left[\frac{\partial^2 \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial A \partial \sigma^2} \right] \\ -E \left[\frac{\partial^2 \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial \sigma^2 \partial A} \right] & -E \left[\frac{\partial^2 \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial \sigma^2^2} \right] \end{bmatrix}.$$

La matriz de información de Fisher es simétrica y puede demostrarse que es definida positiva, como se hace en el problema 3.6.10. La PDF de los datos es

$$p(\mathbf{x}; \boldsymbol{\theta}) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - A)^2 \right]$$

por lo que la función de verosimilitud logarítmica es

$$\ln p(\mathbf{x}; \boldsymbol{\theta}) = -\frac{N}{2} \ln 2\pi - \frac{N}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - A)^2.$$

Las derivadas primeras son

$$\frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial A} = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} (x[n] - A) \quad \frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{n=0}^{N-1} (x[n] - A)^2$$

y las derivadas segundas son

$$\begin{aligned} \frac{\partial^2 \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial A^2} &= \frac{1}{\sigma^2} \sum_{n=0}^{N-1} (-1) = -\frac{N}{\sigma^2} \\ \frac{\partial^2 \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial A \partial \sigma^2} &= -\frac{1}{\sigma^4} \sum_{n=0}^{N-1} (x[n] - A) \\ \frac{\partial^2 \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \sigma^2^2} &= \frac{N}{2\sigma^4} - \frac{1}{\sigma^6} \sum_{n=0}^{N-1} (x[n] - A)^2. \end{aligned}$$

Al tomar las esperanzas y cambiar el signo se obtiene que la información de Fisher es

$$I(\boldsymbol{\theta}) = \begin{bmatrix} \frac{N}{\sigma^2} & 0 \\ 0 & \frac{N}{2\sigma^4} \end{bmatrix}.$$

Luego de invertirla se obtiene que la CRLB de los parámetros es

$$\begin{aligned} \text{var}(\hat{A}) &\geq \frac{\sigma^2}{N} \\ \text{var}(\hat{\sigma}^2) &\geq \frac{2\sigma^4}{N} \end{aligned}$$

Es fácil demostrar que la CRLB para \hat{A} es la misma que cuando σ^2 es conocido (ver por ejemplo el problema de la sección 3.6.8) y también que la CRLB para $\hat{\sigma}^2$ es la misma cuando A es conocido. Esto se debe a que la matriz de información de Fisher es diagonal en este caso. Esto en general no es cierto, como se ilustra en el ejemplo que sigue.

3.2.3. Ejemplo: ajuste de recta

Se considera el problema del ajuste de una recta: dadas las observaciones

$$x[n] = A + Bn + w[n], \quad n = 0, \dots, N-1,$$

donde $w[n]$ es WGN de media nula y varianza σ^2 , se quiere determinar la CRLB de la pendiente B y la intersección A . El parámetro vectorial es en este caso $\boldsymbol{\theta} = [A \ B]^T$, y hay que determinar la matriz de información de Fisher 2×2 ,

$$I(\boldsymbol{\theta}) = \begin{bmatrix} -E \left[\frac{\partial^2 \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial A^2} \right] & -E \left[\frac{\partial^2 \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial A \partial B} \right] \\ -E \left[\frac{\partial^2 \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial B \partial A} \right] & -E \left[\frac{\partial^2 \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial B^2} \right] \end{bmatrix}.$$

La PDF de cada muestra $x[n]$ es (ver el ejercicio de la sección 1.1.3)

$$p(x[n]; \boldsymbol{\theta}) = p_w(x[n] - A - Bn) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2\sigma^2} (x[n] - A - Bn)^2 \right\},$$

y como son independientes, se tiene que

$$\begin{aligned} p(\mathbf{x}; \boldsymbol{\theta}) &= \prod_{n=0}^{N-1} p(x[n]; \boldsymbol{\theta}) \\ &= \prod_{n=0}^{N-1} \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2\sigma^2} (x[n] - A - Bn)^2 \right\} \\ &= \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - A - Bn)^2 \right\}. \end{aligned}$$

Tomando el logaritmo, se llega a que

$$\ln p(\mathbf{x}; \boldsymbol{\theta}) = -\frac{N}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - A - Bn)^2.$$

Por lo tanto, las derivadas primeras son

$$\frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial A} = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} (x[n] - A - Bn) \quad \frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial B} = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} (x[n] - A - Bn)n$$

y las derivadas segundas son

$$\begin{aligned} \frac{\partial^2 \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial A^2} &= \frac{1}{\sigma^2} \sum_{n=0}^{N-1} (-1) = -\frac{N}{\sigma^2} \\ \frac{\partial^2 \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial A \partial B} &= -\frac{1}{\sigma^2} \sum_{n=0}^{N-1} n = -\frac{N(N-1)}{2\sigma^2} \\ \frac{\partial^2 \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial B^2} &= -\frac{1}{\sigma^2} \sum_{n=0}^{N-1} n^2 = -\frac{N(N-1)(2N-1)}{6\sigma^2}, \end{aligned}$$

donde se emplearon las identidades²

$$\sum_{n=0}^{N-1} n = \frac{N(N-1)}{2} \quad \sum_{n=0}^{N-1} n^2 = -\frac{N(N-1)(2N-1)}{6} \quad (3.20)$$

²Ver por ejemplo: https://en.wikipedia.org/wiki/List_of_mathematical_series#Sums_of_powers

Por lo tanto, la matriz de información de Fisher queda

$$\mathbf{I}(\boldsymbol{\theta}) = \frac{N}{\sigma^2} \begin{bmatrix} 1 & \frac{N-1}{2} \\ \frac{N-1}{2} & \frac{(N-1)(2N-1)}{6} \end{bmatrix}.$$

Para calcular la CRLB (ecuación 3.13), hay que invertir $\mathbf{I}(\boldsymbol{\theta})$. El determinante es

$$\det\left(\frac{\sigma^2}{N}\mathbf{I}(\boldsymbol{\theta})\right) = \frac{(N-1)(2N-1)}{6} - \frac{(N-1)^2}{4} = \frac{N-1}{12} [2(2N-1) - 3(N-1)] = \frac{(N-1)(N+1)}{12},$$

por lo que

$$\mathbf{I}^{-1}(\boldsymbol{\theta}) = \frac{12\sigma^2}{N(N-1)(N+1)} \begin{bmatrix} \frac{(N-1)(2N-1)}{6} & -\frac{N-1}{2} \\ -\frac{N-1}{2} & 1 \end{bmatrix},$$

resultando en

$$\mathbf{I}^{-1}(\boldsymbol{\theta}) = \frac{\sigma^2}{N} \begin{bmatrix} \frac{2(2N-1)}{N+1} & -\frac{6}{N+1} \\ -\frac{6}{N+1} & \frac{12}{N^2-1} \end{bmatrix}. \quad (3.21)$$

De aquí se concluye que la CRLB es

$$\text{var}(\hat{A}) \geq \frac{2(2N-1)\sigma^2}{N(N+1)} \quad \text{var}(\hat{B}) \geq \frac{12\sigma^2}{N(N^2-1)}.$$

Para averiguar si existe un estimador eficiente que alcance la CRLB, hay que ver si es posible factorizar la función de verosimilitud logarítmica como en la ecuación 3.12,

$$\frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{I}(\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}(\mathbf{x}) - \boldsymbol{\theta}),$$

que en este caso es

$$\begin{bmatrix} \frac{1}{\sigma^2} \sum_{n=0}^{N-1} (x[n] - A - Bn) \\ \frac{1}{\sigma^2} \sum_{n=0}^{N-1} (x[n] - A - Bn)n \end{bmatrix} = \frac{N}{\sigma^2} \begin{bmatrix} 1 & \frac{N-1}{2} \\ \frac{N-1}{2} & \frac{(N-1)(2N-1)}{6} \end{bmatrix} \begin{bmatrix} \hat{A} - A \\ \hat{B} - B \end{bmatrix}.$$

Se intentará resolver este sistema de ecuaciones calculado \hat{A} y \hat{B} , que no pueden depender de A y B . Desarrollando el sistema matricial, se obtiene el sistema de ecuaciones

$$\begin{cases} \frac{1}{\sigma^2} \sum_{n=0}^{N-1} (x[n] - A - Bn) &= \frac{N}{\sigma^2} \left[(\hat{A} - A) + \frac{N-1}{2} (\hat{B} - B) \right] \\ \frac{1}{\sigma^2} \sum_{n=0}^{N-1} (x[n] - A - Bn)n &= \frac{N}{\sigma^2} \left[\frac{N-1}{2} (\hat{A} - A) + \frac{(N-1)(2N-1)}{6} (\hat{B} - B) \right] \end{cases}.$$

Las sumatorias del lado izquierdo de la igualdad pueden desarrollarse como

$$\begin{aligned} \sum_{n=0}^{N-1} (x[n] - A - Bn) &= \sum_{n=0}^{N-1} x[n] - \sum_{n=0}^{N-1} A - \sum_{n=0}^{N-1} Bn = \sum_{n=0}^{N-1} x[n] - NA - \frac{N(N-1)}{2} B \\ \sum_{n=0}^{N-1} (x[n] - A - Bn)n &= \sum_{n=0}^{N-1} nx[n] - \sum_{n=0}^{N-1} An - \sum_{n=0}^{N-1} Bn^2 = \sum_{n=0}^{N-1} nx[n] - \frac{N(N-1)}{2} A - \frac{(N-1)(2N-1)}{6} B, \end{aligned}$$

donde nuevamente se emplearon las identidades de la ecuación 3.20. Teniendo esto en cuenta, desarrollando el lado derecho de la igualdad y cancelando términos, el sistema se reduce a

$$\begin{cases} \sum_{n=0}^{N-1} x[n] &= N \left[\hat{A} + \frac{N-1}{2} \hat{B} \right] \\ \sum_{n=0}^{N-1} nx[n] &= N \left[\frac{N-1}{2} \hat{A} + \frac{(N-1)(2N-1)}{6} \hat{B} \right] \end{cases}.$$

Despejando \hat{A} de la primera ecuación,

$$\hat{A} = -\frac{N-1}{2} \hat{B} + \frac{1}{N} \sum_{n=0}^{N-1} x[n] \quad (3.22)$$

y sustituyendo el resultado en la segunda, se tiene que

$$\begin{aligned} \frac{N-1}{2} \left[-\frac{N-1}{2} \hat{B} + \frac{1}{N} \sum_{n=0}^{N-1} x[n] \right] + \frac{(N-1)(2N-1)}{6} \hat{B} &= \frac{1}{N} \sum_{n=0}^{N-1} nx[n] \\ -\frac{N-1}{4} \hat{B} + \frac{1}{2N} \sum_{n=0}^{N-1} x[n] + \frac{2N-1}{6} \hat{B} &= \frac{1}{N(N-1)} \sum_{n=0}^{N-1} nx[n] \\ -\frac{N-1}{4} \hat{B} + \frac{2N-1}{6} \hat{B} &= -\frac{1}{2N} \sum_{n=0}^{N-1} x[n] + \frac{1}{N(N-1)} \sum_{n=0}^{N-1} nx[n] \\ \frac{-3(N-1) + 2(2N-1)}{12} \hat{B} &= -\frac{1}{2N} \sum_{n=0}^{N-1} x[n] + \frac{1}{N(N-1)} \sum_{n=0}^{N-1} nx[n] \\ \frac{N+1}{12} \hat{B} &= -\frac{1}{2N} \sum_{n=0}^{N-1} x[n] + \frac{1}{N(N-1)} \sum_{n=0}^{N-1} nx[n] \end{aligned}$$

resultando en

$$\hat{B} = -\frac{6}{N(N+1)} \sum_{n=0}^{N-1} x[n] + \frac{12}{N(N^2-1)} \sum_{n=0}^{N-1} nx[n].$$

Sustituyendo este resultado en la ecuación 3.22 para calcular \hat{A} se ve que

$$\begin{aligned} \hat{A} &= -\frac{N-1}{2} \left[-\frac{6}{N(N+1)} \sum_{n=0}^{N-1} x[n] + \frac{12}{N(N^2-1)} \sum_{n=0}^{N-1} nx[n] \right] + \frac{1}{N} \sum_{n=0}^{N-1} x[n] \\ &= \left[\frac{3(N-1)}{N(N+1)} \sum_{n=0}^{N-1} x[n] - \frac{6}{N(N+1)} \sum_{n=0}^{N-1} nx[n] \right] + \frac{1}{N} \sum_{n=0}^{N-1} x[n] \\ &= \frac{3(N-1) + (N+1)}{N(N+1)} \sum_{n=0}^{N-1} x[n] - \frac{6}{N(N+1)} \sum_{n=0}^{N-1} nx[n] \end{aligned}$$

resultando en

$$\hat{A} = \frac{2(2N-1)}{N(N+1)} \sum_{n=0}^{N-1} x[n] - \frac{6}{N(N+1)} \sum_{n=0}^{N-1} nx[n].$$

Se concluye que $[\hat{A} \ \hat{B}]^T$ es eficiente y por lo tanto, es el estimador MVU. Además, la matriz $\mathbf{I}^{-1}(\boldsymbol{\theta})$ de la ecuación 3.21, es la matriz de covarianza del estimador.

3.3. CRLB general para señales en WGN

Considérese el caso en que una señal determinística que depende de un parámetro desconocido θ se observa en WGN, es decir,

$$x[n] = s[n; \theta] + w[n] \quad n = 0, 1, \dots, N-1.$$

Es fácil demostrar (ver sección 3.5 de [1]) que la información de Fisher es

$$I(\theta) = E \left(\frac{\partial^2 \ln p(\mathbf{x}; \theta)}{\partial \theta^2} \right) = -\frac{1}{\sigma^2} \sum_{n=0}^{N-1} \left(\frac{\partial s[n; \theta]}{\partial \theta} \right)^2$$

y por lo tanto, la CRLB del parámetro es

$$\text{var}(\hat{\theta}) \geq \frac{\sigma^2}{\sum_{n=0}^{N-1} \left(\frac{\partial s[n; \theta]}{\partial \theta} \right)^2}. \quad (3.23)$$

3.3.1. Ejemplo: fase de una senoide

Se desea estimar la fase ϕ de una senoide en WGN,

$$x[n] = A \cos(2\pi f_0 n + \phi) + w[n], \quad n = 0, \dots, N-1,$$

donde la amplitud A y la frecuencia f_0 de la senoide son conocidas (ver el ejemplo de la sección 3.4.4 para el caso en que son desconocidos), así como la varianza σ^2 del ruido. Se calculará la CRLB empleando la ecuación 3.23. En este caso

$$s[n; \phi] = A \cos(2\pi f_0 n + \phi)$$

y por lo tanto

$$\frac{\partial s[n; \theta]}{\partial \theta} = -A \sin(2\pi f_0 n + \phi).$$

Además

$$\begin{aligned} \sum_{n=0}^{N-1} \left(\frac{\partial s[n; \theta]}{\partial \theta} \right)^2 &= A^2 \sum_{n=0}^{N-1} \sin^2(2\pi f_0 n + \phi) \\ &\stackrel{(a)}{=} A^2 \sum_{n=0}^{N-1} \left[\frac{1}{2} - \frac{1}{2} \cos(4\pi f_0 n + 2\phi) \right] \\ &\stackrel{(b)}{\approx} \frac{NA^2}{2}, \end{aligned}$$

donde en (a) se empleó la igualdad trigonométrica $2 \sin^2 \theta = 1 - \cos 2\theta$ y en (b) que

$$\frac{1}{N} \sum_{n=0}^{N-1} \cos(4\pi f_0 n + 2\phi) \approx 0,$$

si f_0 no está cercano a 0 ni a $1/2$, como se muestra en el problema 3.6.7. Sustituyendo el resultado en la ecuación 3.23, se concluye que

$$\text{var}(\hat{\phi}) \geq \frac{2\sigma^2}{NA^2}.$$

3.4. CRLB para el caso gaussiano general

3.4.1. Teorema

Se asume que $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}(\boldsymbol{\theta}), \mathbf{C}(\boldsymbol{\theta}))$, donde $\boldsymbol{\mu}(\boldsymbol{\theta})$ es el vector esperanza $N \times 1$ y $\mathbf{C}(\boldsymbol{\theta})$ es la matriz de covarianza $N \times N$, los cuales dependen del parámetro vectorial $\boldsymbol{\theta}$ de dimensión $p \times 1$. Se demostrará que la matriz de información de Fisher es

$$[\mathbf{I}(\boldsymbol{\theta})]_{ij} = \left[\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \theta_i} \right]^T \mathbf{C}^{-1}(\boldsymbol{\theta}) \left[\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \theta_j} \right] + \frac{1}{2} \text{tr} \left[\mathbf{C}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \theta_i} \mathbf{C}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \theta_j} \right], \quad (3.24)$$

donde

$$\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \theta_i} = \begin{bmatrix} \frac{\partial [\boldsymbol{\mu}(\boldsymbol{\theta})]_1}{\partial \theta_i} \\ \vdots \\ \frac{\partial [\boldsymbol{\mu}(\boldsymbol{\theta})]_N}{\partial \theta_i} \end{bmatrix} \quad \text{y} \quad \frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \theta_i} = \begin{bmatrix} \frac{\partial [\mathbf{C}(\boldsymbol{\theta})]_{11}}{\partial \theta_i} & \cdots & \frac{\partial [\mathbf{C}(\boldsymbol{\theta})]_{1N}}{\partial \theta_i} \\ \vdots & \ddots & \vdots \\ \frac{\partial [\mathbf{C}(\boldsymbol{\theta})]_{N1}}{\partial \theta_i} & \cdots & \frac{\partial [\mathbf{C}(\boldsymbol{\theta})]_{NN}}{\partial \theta_i} \end{bmatrix}.$$

Demostración Se calculará la matriz de información de Fisher usando la ecuación 3.17. La PDF del vector \mathbf{x} es

$$p(\mathbf{x}; \boldsymbol{\theta}) = \frac{1}{(2\pi)^{\frac{N}{2}} \det^{\frac{1}{2}}[\mathbf{C}(\boldsymbol{\theta})]} \exp \left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}(\boldsymbol{\theta}))^T \mathbf{C}^{-1}(\boldsymbol{\theta}) (\mathbf{x} - \boldsymbol{\mu}(\boldsymbol{\theta})) \right].$$

Tomando el logaritmo se obtiene que

$$\ln p(\mathbf{x}; \boldsymbol{\theta}) = -\frac{N}{2} \ln(2\pi) - \frac{1}{2} \ln \det[\mathbf{C}(\boldsymbol{\theta})] - \frac{1}{2} [(\mathbf{x} - \boldsymbol{\mu}(\boldsymbol{\theta}))^T \mathbf{C}^{-1}(\boldsymbol{\theta}) (\mathbf{x} - \boldsymbol{\mu}(\boldsymbol{\theta}))],$$

y diferenciando respecto a θ_k se llega a que

$$\frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \theta_k} = -\frac{1}{2} \frac{\partial \ln \det[\mathbf{C}(\boldsymbol{\theta})]}{\partial \theta_k} - \frac{1}{2} \frac{\partial}{\partial \theta_k} [(\mathbf{x} - \boldsymbol{\mu}(\boldsymbol{\theta}))^T \mathbf{C}^{-1}(\boldsymbol{\theta}) (\mathbf{x} - \boldsymbol{\mu}(\boldsymbol{\theta}))]. \quad (3.25)$$

Se demostrará que

$$\frac{\partial \ln \det[\mathbf{C}(\boldsymbol{\theta})]}{\partial \theta_k} = \text{tr} \left(\mathbf{C}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \theta_k} \right). \quad (3.26)$$

Para ello, se parte notando que

$$\frac{\partial \ln \det[\mathbf{C}(\boldsymbol{\theta})]}{\partial \theta_k} = \frac{1}{\det[\mathbf{C}(\boldsymbol{\theta})]} \frac{\partial \det[\mathbf{C}(\boldsymbol{\theta})]}{\partial \theta_k}. \quad (3.27)$$

Como $\det[\mathbf{C}(\boldsymbol{\theta})]$ depende de todos los elementos de $\mathbf{C}(\boldsymbol{\theta})$, es decir, depende de $[\mathbf{C}(\boldsymbol{\theta})]_{ij}$ para todo $i = 1, \dots, N$ y $j = 1, \dots, N$, se cumple que

$$\begin{aligned} \frac{\partial \det[\mathbf{C}(\boldsymbol{\theta})]}{\partial \theta_k} &= \sum_{i=1}^N \sum_{j=1}^N \frac{\partial \det[\mathbf{C}(\boldsymbol{\theta})]}{\partial [\mathbf{C}(\boldsymbol{\theta})]_{ij}} \frac{\partial [\mathbf{C}(\boldsymbol{\theta})]_{ij}}{\partial \theta_k} \\ &\stackrel{(a)}{=} \sum_{i=1}^N \sum_{j=1}^N \left[\frac{\partial \det[\mathbf{C}(\boldsymbol{\theta})]}{\partial \mathbf{C}(\boldsymbol{\theta})} \right]_{ij} \frac{\partial [\mathbf{C}(\boldsymbol{\theta})]_{ij}}{\partial \theta_k} \\ &\stackrel{(b)}{=} \text{tr} \left(\frac{\partial \det[\mathbf{C}(\boldsymbol{\theta})]}{\partial \mathbf{C}(\boldsymbol{\theta})} \frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \theta_k}^T \right) \\ &\stackrel{(c)}{=} \text{tr} \left(\frac{\partial \det[\mathbf{C}(\boldsymbol{\theta})]}{\partial \mathbf{C}(\boldsymbol{\theta})} \frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \theta_k} \right), \end{aligned} \quad (3.28)$$

donde en (a) se definió la matriz $N \times N$ $\partial \det[\mathbf{C}(\boldsymbol{\theta})]/\partial \mathbf{C}(\boldsymbol{\theta})$ cuyo elemento $[i, j]$ es $\partial \det[\mathbf{C}(\boldsymbol{\theta})]/\partial [\mathbf{C}(\boldsymbol{\theta})]_{ij}$, en (b) se usó la identidad

$$\text{tr}(\mathbf{A}\mathbf{B}^T) = \sum_{i=1}^N [\mathbf{A}\mathbf{B}^T]_{ii} = \sum_{i=1}^N \sum_{j=1}^N [\mathbf{A}]_{ij} [\mathbf{B}]_{ij}, \quad (3.29)$$

y en (c) que la matriz $\mathbf{C}(\boldsymbol{\theta})$ es simétrica, por lo que la matriz $\partial \mathbf{C}(\boldsymbol{\theta})/\partial \theta_k$ también es simétrica. Por definición del determinante, se tiene que

$$\det[\mathbf{C}(\boldsymbol{\theta})] = \sum_{i=1}^N [\mathbf{C}(\boldsymbol{\theta})]_{ii} [\mathbf{M}]_{ii},$$

donde \mathbf{M} es la matriz $N \times N$ de cofactores y j puede tomar cualquier valor entre 1 y N^3 . De esta forma,

$$\frac{\partial \det[\mathbf{C}(\boldsymbol{\theta})]}{\partial [\mathbf{C}(\boldsymbol{\theta})]_{ij}} = [\mathbf{M}]_{ij},$$

o

$$\frac{\partial \det[\mathbf{C}(\boldsymbol{\theta})]}{\partial \mathbf{C}(\boldsymbol{\theta})} = \mathbf{M}.$$

Considerando la identidad ⁴

$$\mathbf{C}^{-1}(\boldsymbol{\theta}) = \frac{\mathbf{M}^T}{\det[\mathbf{C}(\boldsymbol{\theta})]}$$

y sustituyendo \mathbf{M} en la ecuación anterior se obtiene que

$$\frac{\partial \det[\mathbf{C}(\boldsymbol{\theta})]}{\partial \mathbf{C}(\boldsymbol{\theta})} = \det[\mathbf{C}(\boldsymbol{\theta})] \mathbf{C}^{-1}(\boldsymbol{\theta}),$$

donde se consideró nuevamente que $\mathbf{C}(\boldsymbol{\theta})$ es simétrica y por lo tanto $\mathbf{C}^{-1}(\boldsymbol{\theta})$ también es simétrica. Sustituyendo este resultado en la ecuación 3.28 se tiene que

$$\frac{\partial \det[\mathbf{C}(\boldsymbol{\theta})]}{\partial \theta_k} = \det[\mathbf{C}(\boldsymbol{\theta})] \operatorname{tr} \left(\mathbf{C}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \theta_k} \right),$$

y finalmente, sustituyendo esta expresión en la ecuación 3.27 se obtiene la ecuación 3.26, que es lo que se quería demostrar.

Otra identidad que se necesitará es

$$\frac{\partial \mathbf{C}^{-1}(\boldsymbol{\theta})}{\partial \theta_k} = -\mathbf{C}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \theta_k} \mathbf{C}^{-1}(\boldsymbol{\theta}). \quad (3.30)$$

Para demostrarlo, se parte de la igualdad

$$\mathbf{C}^{-1}(\boldsymbol{\theta}) \mathbf{C}(\boldsymbol{\theta}) = \mathbf{I}.$$

Al diferenciar cada elemento de las matrices respecto a θ_k , el lado derecho es la matriz nula. Como el elemento $[i, j]$ de la matriz del lado izquierdo es

$$[\mathbf{C}^{-1}(\boldsymbol{\theta}) \mathbf{C}(\boldsymbol{\theta})]_{ij} = \sum_{n=1}^N [\mathbf{C}^{-1}(\boldsymbol{\theta})]_{in} [\mathbf{C}(\boldsymbol{\theta})]_{nj},$$

al derivar respecto a θ_k se tiene que

$$\begin{aligned} \frac{[\mathbf{C}^{-1}(\boldsymbol{\theta}) \mathbf{C}(\boldsymbol{\theta})]_{ij}}{\partial \theta_k} &= \sum_{n=1}^N \left([\mathbf{C}^{-1}(\boldsymbol{\theta})]_{in} \frac{\partial [\mathbf{C}(\boldsymbol{\theta})]_{nj}}{\partial \theta_k} + \frac{\partial [\mathbf{C}^{-1}(\boldsymbol{\theta})]_{in}}{\partial \theta_k} [\mathbf{C}(\boldsymbol{\theta})]_{nj} \right) \\ &= \sum_{n=1}^N [\mathbf{C}^{-1}(\boldsymbol{\theta})]_{in} \frac{\partial [\mathbf{C}(\boldsymbol{\theta})]_{nj}}{\partial \theta_k} + \sum_{n=1}^N \frac{\partial [\mathbf{C}^{-1}(\boldsymbol{\theta})]_{in}}{\partial \theta_k} [\mathbf{C}(\boldsymbol{\theta})]_{nj} \\ &= \left[\mathbf{C}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \theta_k} \right]_{ij} + \left[\frac{\partial \mathbf{C}^{-1}(\boldsymbol{\theta})}{\partial \theta_k} \mathbf{C}(\boldsymbol{\theta}) \right]_{ij}. \end{aligned}$$

Estos resultados combinados, en notación matricial, resultan en

$$\mathbf{C}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \theta_k} + \frac{\partial \mathbf{C}^{-1}(\boldsymbol{\theta})}{\partial \theta_k} \mathbf{C}(\boldsymbol{\theta}) = \mathbf{0}, \quad (3.31)$$

y despejando se obtiene la ecuación 3.30.

³Ver https://en.wikipedia.org/wiki/Laplace_expansion

⁴Ver [https://en.wikipedia.org/wiki/Minor_\(linear_algebra\)#Inverse_of_a_matrix](https://en.wikipedia.org/wiki/Minor_(linear_algebra)#Inverse_of_a_matrix)

Volviendo a la derivada de la función de verosimilitud logarítmica, dada en la ecuación 3.25, se ve que el primer sumando puede evaluarse a partir de la ecuación 3.26. Se evaluará a continuación el segundo sumando. Para hacerlo, se observa que $\mathbf{C}^{-1}(\boldsymbol{\theta})(\mathbf{x} - \boldsymbol{\mu}(\boldsymbol{\theta}))$ es un vector columna cuyo elemento i -ésimo está dado por

$$[\mathbf{C}^{-1}(\boldsymbol{\theta})(\mathbf{x} - \boldsymbol{\mu}(\boldsymbol{\theta}))]_i = \sum_{j=1}^N [\mathbf{C}^{-1}(\boldsymbol{\theta})]_{ij} (x[j] - [\boldsymbol{\mu}(\boldsymbol{\theta})]_j), \quad i = 1, \dots, N,$$

y por lo tanto,

$$\begin{aligned} (\mathbf{x} - \boldsymbol{\mu}(\boldsymbol{\theta}))^T \mathbf{C}^{-1}(\boldsymbol{\theta})(\mathbf{x} - \boldsymbol{\mu}(\boldsymbol{\theta})) &= \sum_{i=1}^N (x[i] - [\boldsymbol{\mu}(\boldsymbol{\theta})]_i) \sum_{j=1}^N [\mathbf{C}^{-1}(\boldsymbol{\theta})]_{ij} (x[j] - [\boldsymbol{\mu}(\boldsymbol{\theta})]_j) \\ &= \sum_{i=1}^N \sum_{j=1}^N (x[i] - [\boldsymbol{\mu}(\boldsymbol{\theta})]_i) [\mathbf{C}^{-1}(\boldsymbol{\theta})]_{ij} (x[j] - [\boldsymbol{\mu}(\boldsymbol{\theta})]_j). \end{aligned}$$

Por lo tanto,

$$\begin{aligned} \frac{\partial}{\partial \theta_k} [(\mathbf{x} - \boldsymbol{\mu}(\boldsymbol{\theta}))^T \mathbf{C}^{-1}(\boldsymbol{\theta})(\mathbf{x} - \boldsymbol{\mu}(\boldsymbol{\theta}))] &= \sum_{i=1}^N \sum_{j=1}^N (x[i] - [\boldsymbol{\mu}(\boldsymbol{\theta})]_i) [\mathbf{C}^{-1}(\boldsymbol{\theta})]_{ij} \left(-\frac{\partial [\boldsymbol{\mu}(\boldsymbol{\theta})]_j}{\partial \theta_k} \right) \\ &\quad + \sum_{i=1}^N \sum_{j=1}^N (x[i] - [\boldsymbol{\mu}(\boldsymbol{\theta})]_i) \frac{\partial [\mathbf{C}^{-1}(\boldsymbol{\theta})]_{ij}}{\partial \theta_k} (x[j] - [\boldsymbol{\mu}(\boldsymbol{\theta})]_j) \\ &\quad + \sum_{i=1}^N \sum_{j=1}^N \left(-\frac{\partial [\boldsymbol{\mu}(\boldsymbol{\theta})]_i}{\partial \theta_k} \right) [\mathbf{C}^{-1}(\boldsymbol{\theta})]_{ij} (x[j] - [\boldsymbol{\mu}(\boldsymbol{\theta})]_j) \\ &= -(\mathbf{x} - \boldsymbol{\mu}(\boldsymbol{\theta}))^T \mathbf{C}^{-1}(\boldsymbol{\theta}) \frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \theta_k} \\ &\quad + (\mathbf{x} - \boldsymbol{\mu}(\boldsymbol{\theta}))^T \frac{\partial \mathbf{C}^{-1}(\boldsymbol{\theta})}{\partial \theta_k} (\mathbf{x} - \boldsymbol{\mu}(\boldsymbol{\theta})) \\ &\quad - \frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})^T}{\partial \theta_k} \mathbf{C}^{-1}(\boldsymbol{\theta})(\mathbf{x} - \boldsymbol{\mu}(\boldsymbol{\theta})) \\ &\stackrel{(a)}{=} -2 \frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})^T}{\partial \theta_k} \mathbf{C}^{-1}(\boldsymbol{\theta})(\mathbf{x} - \boldsymbol{\mu}(\boldsymbol{\theta})) \\ &\quad + (\mathbf{x} - \boldsymbol{\mu}(\boldsymbol{\theta}))^T \frac{\partial \mathbf{C}^{-1}(\boldsymbol{\theta})}{\partial \theta_k} (\mathbf{x} - \boldsymbol{\mu}(\boldsymbol{\theta})), \end{aligned} \tag{3.32}$$

donde en (a) se notó que como los sumandos son escalares y por lo tanto su valor es igual al valor transpuesto, el primer y tercer sumando son iguales. Sustituyendo las ecuaciones 3.26 y 3.32 en la ecuación 3.25 se obtiene que

$$\frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \theta_k} = -\frac{1}{2} \text{tr} \left(\mathbf{C}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \theta_k} \right) + \frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})^T}{\partial \theta_k} \mathbf{C}^{-1}(\boldsymbol{\theta})(\mathbf{x} - \boldsymbol{\mu}(\boldsymbol{\theta})) - \frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}(\boldsymbol{\theta}))^T \frac{\partial \mathbf{C}^{-1}(\boldsymbol{\theta})}{\partial \theta_k} (\mathbf{x} - \boldsymbol{\mu}(\boldsymbol{\theta})) \tag{3.33}$$

Considerando que se aplicará la ecuación 3.17, y sea $\mathbf{y} = \mathbf{x} - \boldsymbol{\mu}(\boldsymbol{\theta})$, se tiene que

$$\begin{aligned} \frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \theta_k} \frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \theta_l} &= \frac{1}{4} \text{tr} \left(\mathbf{C}^{-1} \frac{\partial \mathbf{C}}{\partial \theta_k} \right) \text{tr} \left(\mathbf{C}^{-1} \frac{\partial \mathbf{C}}{\partial \theta_l} \right) - \frac{1}{2} \text{tr} \left(\mathbf{C}^{-1} \frac{\partial \mathbf{C}}{\partial \theta_k} \right) \frac{\partial \boldsymbol{\mu}^T}{\partial \theta_l} \mathbf{C}^{-1} \mathbf{y} \\ &\quad + \frac{1}{4} \text{tr} \left(\mathbf{C}^{-1} \frac{\partial \mathbf{C}}{\partial \theta_k} \right) \mathbf{y}^T \frac{\partial \mathbf{C}^{-1}}{\partial \theta_l} \mathbf{y} \\ &\quad - \frac{1}{2} \frac{\partial \boldsymbol{\mu}^T}{\partial \theta_k} \mathbf{C}^{-1} \mathbf{y} \text{tr} \left(\mathbf{C}^{-1} \frac{\partial \mathbf{C}}{\partial \theta_l} \right) + \frac{\partial \boldsymbol{\mu}^T}{\partial \theta_k} \mathbf{C}^{-1} \mathbf{y} \frac{\partial \boldsymbol{\mu}^T}{\partial \theta_l} \mathbf{C}^{-1} \mathbf{y} \\ &\quad + \frac{\partial \boldsymbol{\mu}^T}{\partial \theta_k} \mathbf{C}^{-1} \mathbf{y} \mathbf{y}^T \frac{\partial \mathbf{C}^{-1}}{\partial \theta_l} \mathbf{y} \\ &\quad + \frac{1}{4} \mathbf{y}^T \frac{\partial \mathbf{C}^{-1}}{\partial \theta_k} \mathbf{y} \text{tr} \left(\mathbf{C}^{-1} \frac{\partial \mathbf{C}}{\partial \theta_l} \right) - \frac{1}{2} \mathbf{y}^T \frac{\partial \mathbf{C}^{-1}}{\partial \theta_k} \mathbf{y} \frac{\partial \boldsymbol{\mu}^T}{\partial \theta_l} \mathbf{C}^{-1} \mathbf{y} \\ &\quad + \frac{1}{4} \mathbf{y}^T \frac{\partial \mathbf{C}^{-1}}{\partial \theta_k} \mathbf{y} \mathbf{y}^T \frac{\partial \mathbf{C}^{-1}}{\partial \theta_l} \mathbf{y}, \end{aligned}$$

donde se omitió la dependencia explícita de \mathbf{C} y $\boldsymbol{\mu}$ con $\boldsymbol{\theta}$. El paso siguiente es tomar la esperanza, teniendo en cuenta que la única variable aleatoria en la expresión es \mathbf{y} . Como $\mathbf{y} \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$, distribución que es simétrica, los momentos de orden impar son nulos⁵. Los momentos de orden impar corresponden al segundo, cuarto, sexto y octavo sumando. Considerando que el quinto sumando puede escribirse como

$$\frac{\partial \boldsymbol{\mu}^T}{\partial \theta_k} \mathbf{C}^{-1} \mathbf{y} \frac{\partial \boldsymbol{\mu}^T}{\partial \theta_l} \mathbf{C}^{-1} \mathbf{y} = \frac{\partial \boldsymbol{\mu}^T}{\partial \theta_k} \mathbf{C}^{-1} \mathbf{y} \left(\frac{\partial \boldsymbol{\mu}^T}{\partial \theta_l} \mathbf{C}^{-1} \mathbf{y} \right)^T = \frac{\partial \boldsymbol{\mu}^T}{\partial \theta_k} \mathbf{C}^{-1} \mathbf{y} \mathbf{y}^T \mathbf{C}^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \theta_l},$$

al tomar esperanza se tiene que

$$\begin{aligned} [\mathbf{I}(\boldsymbol{\theta})]_{kl} &= E \left[\frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \theta_k} \frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \theta_l} \right] \\ &= \frac{1}{4} \text{tr} \left(\mathbf{C}^{-1} \frac{\partial \mathbf{C}}{\partial \theta_k} \right) \text{tr} \left(\mathbf{C}^{-1} \frac{\partial \mathbf{C}}{\partial \theta_l} \right) + \frac{1}{4} \text{tr} \left(\mathbf{C}^{-1} \frac{\partial \mathbf{C}}{\partial \theta_k} \right) E \left(\mathbf{y}^T \frac{\partial \mathbf{C}^{-1}}{\partial \theta_l} \mathbf{y} \right) \\ &\quad + \frac{\partial \boldsymbol{\mu}^T}{\partial \theta_k} \mathbf{C}^{-1} E(\mathbf{y} \mathbf{y}^T) \mathbf{C}^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \theta_l} + \frac{1}{4} \text{tr} \left(\mathbf{C}^{-1} \frac{\partial \mathbf{C}}{\partial \theta_l} \right) E \left(\mathbf{y}^T \frac{\partial \mathbf{C}^{-1}}{\partial \theta_k} \mathbf{y} \right) \\ &\quad + \frac{1}{4} E \left(\mathbf{y}^T \frac{\partial \mathbf{C}^{-1}}{\partial \theta_k} \mathbf{y} \mathbf{y}^T \frac{\partial \mathbf{C}^{-1}}{\partial \theta_l} \mathbf{y} \right). \end{aligned}$$

Se evaluará a continuación el segundo y cuarto sumando. Para eso, se considera que para vectores $N \times 1$ \mathbf{y} , \mathbf{z} se cumple que

$$\mathbf{y}^T \mathbf{z} = \text{tr}(\mathbf{z} \mathbf{y}^T) \quad \Rightarrow \quad E(\mathbf{y}^T \mathbf{z}) = E[\text{tr}(\mathbf{z} \mathbf{y}^T)] = \text{tr}[E(\mathbf{z} \mathbf{y}^T)],$$

ya que $[E(\mathbf{A})]_{ij} = E([\mathbf{A}]_{ij})$, como de costumbre. Teniendo esto en cuenta, la esperanza en el segundo sumando puede calcularse de la siguiente forma:

$$E \left(\mathbf{y}^T \frac{\partial \mathbf{C}^{-1}}{\partial \theta_l} \mathbf{y} \right) = E \left[\text{tr} \left(\frac{\partial \mathbf{C}^{-1}}{\partial \theta_l} \mathbf{y} \mathbf{y}^T \right) \right] = \text{tr} \left[\frac{\partial \mathbf{C}^{-1}}{\partial \theta_l} E(\mathbf{y} \mathbf{y}^T) \right] = \text{tr} \left(\frac{\partial \mathbf{C}^{-1}}{\partial \theta_l} \mathbf{C} \right) \stackrel{(a)}{=} -\text{tr} \left(\mathbf{C}^{-1} \frac{\partial \mathbf{C}}{\partial \theta_l} \right),$$

donde en (a) se empleó la identidad de la ecuación 3.30 o equivalentemente, la ecuación 3.31. Aplicando el mismo razonamiento en el cuarto sumando, y sustituyendo $E(\mathbf{y} \mathbf{y}^T) = \mathbf{C}$ en el tercer sumando, se obtiene que

$$[\mathbf{I}(\boldsymbol{\theta})]_{kl} = -\frac{1}{4} \text{tr} \left(\mathbf{C}^{-1} \frac{\partial \mathbf{C}}{\partial \theta_k} \right) \text{tr} \left(\mathbf{C}^{-1} \frac{\partial \mathbf{C}}{\partial \theta_l} \right) + \frac{\partial \boldsymbol{\mu}^T}{\partial \theta_k} \mathbf{C}^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \theta_l} + \frac{1}{4} E \left(\mathbf{y}^T \frac{\partial \mathbf{C}^{-1}}{\partial \theta_k} \mathbf{y} \mathbf{y}^T \frac{\partial \mathbf{C}^{-1}}{\partial \theta_l} \mathbf{y} \right) \quad (3.34)$$

Falta evaluar el último sumando de esta expresión. Para hacerlo, se demostrará que [3]

$$E[\mathbf{y}^T \mathbf{A} \mathbf{y} \mathbf{y}^T \mathbf{B} \mathbf{y}] = \text{tr}(\mathbf{A} \mathbf{C}) \text{tr}(\mathbf{B} \mathbf{C}) + 2 \text{tr}(\mathbf{A} \mathbf{C} \mathbf{B} \mathbf{C}), \quad (3.35)$$

⁵Ver https://en.wikipedia.org/wiki/Isserlis'_theorem

donde \mathbf{C} es la matriz de covarianza de \mathbf{y} y \mathbf{A} y \mathbf{B} son matrices simétricas. Se parte observando que

$$\begin{aligned}
E[\mathbf{y}^T \mathbf{A} \mathbf{y} \mathbf{y}^T \mathbf{B} \mathbf{y}] &= E \left(\sum_{i,j} y[i] A_{ij} y[j] \sum_{m,n} y[m] B_{mn} y[n] \right) \\
&= \sum_{i,j,m,n} A_{ij} B_{mn} E(y[i] y[j] y[m] y[n]) \\
&\stackrel{(a)}{=} \sum_{i,j,m,n} A_{ij} B_{mn} (C_{ij} C_{mn} + C_{im} C_{jn} + C_{in} C_{jm}) \\
&= \sum_{i,j,m,n} A_{ij} B_{mn} C_{ij} C_{mn} + \sum_{i,j,m,n} A_{ij} B_{mn} C_{im} C_{jn} + \sum_{i,j,m,n} A_{ij} B_{mn} C_{in} C_{jm} \\
&= \sum_{i,j} A_{ij} C_{ij} \sum_{m,n} B_{mn} C_{mn} + \sum_{i,j,m,n} A_{ij} C_{im} B_{mn} C_{jn} + \sum_{i,j,m,n} A_{ij} C_{in} B_{mn} C_{jm}, \quad (3.36)
\end{aligned}$$

donde en (a) se empleó que el momento cuarto de una variable aleatoria \mathbf{y} normal multivariada de media nula y matriz de covarianza de elemento $[i, j]$ C_{ij} es

$$\begin{aligned}
E(y[1]y[2]y[3]y[4]) &= E(y[1]y[2])E(y[3]y[4]) + E(y[1]y[3])E(y[2]y[4]) + E(y[1]y[4])E(y[2]y[3]) \\
&= C_{12}C_{34} + C_{13}C_{24} + C_{14}C_{23}
\end{aligned}$$

(ver el ejemplo 7-9 de [2]). A partir de la identidad de la ecuación 3.29, se deduce que el primer sumando de la ecuación 3.36 es $\text{tr}(\mathbf{AC})\text{tr}(\mathbf{BC})$, ya que todas las matrices involucradas son simétricas. Se verá que los otros dos sumandos son $\text{tr}(\mathbf{ACBC})$, para lo cual se calculará $\text{tr}(\mathbf{ACBC})$. Observando que

$$[\mathbf{AC}]_{kl} = \sum_{j=1}^N A_{kj} C_{jl} \quad \text{y} \quad [\mathbf{BC}]_{st} = \sum_{n=1}^N B_{sn} C_{nt}$$

se obtiene que los elementos de la diagonal de \mathbf{ACBC} son

$$[\mathbf{ACBC}]_{ii} = \sum_{m=1}^N [\mathbf{AC}]_{im} [\mathbf{BC}]_{mi} = \sum_{m=1}^N \left(\sum_{j=1}^N A_{ij} C_{jm} \right) \left(\sum_{n=1}^N B_{mn} C_{ni} \right) = \sum_{j=1}^N \sum_{n=1}^N \sum_{m=1}^N A_{ij} C_{ni} B_{mn} C_{jm},$$

por lo que

$$\text{tr}(\mathbf{ACBC}) = \sum_{i=1}^N [\mathbf{ACBC}]_{ii} = \sum_{i=1}^N \sum_{j=1}^N \sum_{n=1}^N \sum_{m=1}^N A_{ij} C_{ni} B_{mn} C_{jm}.$$

Como las matrices son simétricas, considerando que $C_{ni} = C_{in}$, la expresión coincide con el tercer sumando de la ecuación 3.36. Además, haciendo el cambio de variable n por m y m por n y teniendo en cuenta la simetría de las matrices, también coincide con el segundo sumando de la ecuación 3.36, con lo que se concluye que se verifica la identidad de la ecuación 3.35. Usando esta identidad, se tiene que

$$\begin{aligned}
E \left(\mathbf{y}^T \frac{\partial \mathbf{C}^{-1}}{\partial \theta_k} \mathbf{y} \mathbf{y}^T \frac{\partial \mathbf{C}^{-1}}{\partial \theta_l} \mathbf{y} \right) &= \text{tr} \left(\frac{\partial \mathbf{C}^{-1}}{\partial \theta_k} \mathbf{C} \right) \text{tr} \left(\frac{\partial \mathbf{C}^{-1}}{\partial \theta_l} \mathbf{C} \right) + 2 \text{tr} \left(\frac{\partial \mathbf{C}^{-1}}{\partial \theta_k} \mathbf{C} \frac{\partial \mathbf{C}^{-1}}{\partial \theta_l} \mathbf{C} \right) \\
&= \text{tr} \left(\mathbf{C}^{-1} \frac{\partial \mathbf{C}}{\partial \theta_k} \right) \text{tr} \left(\mathbf{C}^{-1} \frac{\partial \mathbf{C}}{\partial \theta_l} \right) + 2 \text{tr} \left(\mathbf{C}^{-1} \frac{\partial \mathbf{C}}{\partial \theta_k} \mathbf{C}^{-1} \frac{\partial \mathbf{C}}{\partial \theta_l} \right),
\end{aligned}$$

donde en la segunda igualdad se empleó la identidad de la ecuación 3.31. Sustituyendo este resultado en la ecuación 3.34 se llega a que

$$[\mathbf{I}(\theta)]_{kl} = \frac{\partial \boldsymbol{\mu}^T}{\partial \theta_k} \mathbf{C}^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \theta_l} + \frac{1}{2} \text{tr} \left(\mathbf{C}^{-1} \frac{\partial \mathbf{C}}{\partial \theta_k} \mathbf{C}^{-1} \frac{\partial \mathbf{C}}{\partial \theta_l} \right),$$

que es lo que se quería demostrar.

En el caso en que el parámetro es escalar la información de Fisher se reduce a

$$I(\theta) = \frac{\partial \boldsymbol{\mu}(\theta)^T}{\partial \theta} \mathbf{C}^{-1}(\theta) \frac{\partial \boldsymbol{\mu}(\theta)}{\partial \theta} + \frac{1}{2} \text{tr} \left[\left(\mathbf{C}^{-1}(\theta) \frac{\partial \mathbf{C}(\theta)}{\partial \theta} \right)^2 \right]. \quad (3.37)$$

3.4.2. Ejemplo: parámetros de una señal en WGN

Asúmase que se desea estimar el parámetro escalar θ para el conjunto de datos

$$x[n] = s[n; \theta] + w[n] \quad n = 0, 1, \dots, N-1,$$

donde $w[n]$ es WGN. En este caso, $\mathbf{x} \sim \mathcal{N}(\mathbf{s}(\theta), \mathbf{C})$, donde $\mathbf{s}(\theta) = [s[0; \theta] \ s[1; \theta] \ \dots \ s[N-1; \theta]]^T$ y $\mathbf{C} = \sigma^2 \mathbf{I}$, y por lo tanto, la ecuación 3.37 se reduce a

$$\begin{aligned} I(\theta) &\stackrel{(a)}{=} \frac{1}{\sigma^2} \left[\frac{\partial \boldsymbol{\mu}(\theta)}{\partial \theta} \right]^T \left[\frac{\partial \boldsymbol{\mu}(\theta)}{\partial \theta} \right] \\ &= \frac{1}{\sigma^2} \sum_{n=0}^{N-1} \left(\frac{\partial s[n; \theta]}{\partial \theta} \right)^2 \\ &= \frac{1}{\sigma^2} \sum_{n=0}^{N-1} \left(\frac{\partial s[n; \theta]}{\partial \theta} \right)^2, \end{aligned}$$

donde en (a) se tuvo en cuenta que la matriz de covarianza es independiente de θ y por lo tanto, el segundo sumando de la ecuación 3.37 es nulo.

Generalizando este resultado para el caso de una señal con parámetro vectorial en WGN, la ecuación 3.24 queda

$$[\mathbf{I}(\boldsymbol{\theta})]_{ij} = \left[\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \theta_i} \right]^T \frac{\mathbf{I}}{\sigma^2} \left[\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \theta_j} \right],$$

que conduce a que los elementos de la matriz de información de Fisher son

$$[\mathbf{I}(\boldsymbol{\theta})]_{ij} = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} \frac{\partial s[n; \boldsymbol{\theta}]}{\partial \theta_i} \frac{\partial s[n; \boldsymbol{\theta}]}{\partial \theta_j}. \quad (3.38)$$

3.4.3. Ejemplo: nivel de DC aleatorio en WGN

Se consideran los datos

$$x[n] = A + w[n], \quad n = 0, \dots, N-1$$

donde $w[n]$ es WGN de media nula y varianza σ^2 y A , el nivel de DC, es una variable aleatoria normal de media nula y varianza σ_A^2 independiente de $w[n]$. La potencia de la señal σ_A^2 es el parámetro desconocido, y se quiere encontrar su CRLB.

En esta situación, $\mathbf{x} = [x[0] \ \dots \ x[N-1]]^T$ es un vector gaussiano de media nula y matriz de covarianza $N \times N$ cuyo elemento $[i, j]$ es

$$\begin{aligned} [\mathbf{C}(\sigma_A^2)]_{ij} &= E[x[i-1]x[j-1]] \\ &= E[(A + w[i-1])(A + w[j-1])] \\ &= E(A^2) + E(w[i-1]w[j-1]) + E(Aw[i-1]) + E(Aw[j-1]) \\ &= \sigma_A^2 + \sigma^2 \delta_{ij}. \end{aligned}$$

Se concluye que

$$\mathbf{C}(\sigma_A^2) = \sigma_A^2 \mathbf{1}\mathbf{1}^T + \sigma^2 \mathbf{I},$$

donde $\mathbf{1} = [1 \ \dots \ 1]^T$ de tamaño $N \times 1$. La CRLB se calcula a partir de la ecuación 3.37, en la cual el primer sumando es nulo ya que la media no depende del parámetro desconocido, así que

$$I(\sigma_A^2) = \frac{1}{2} \text{tr} \left[\left(\mathbf{C}^{-1}(\sigma_A^2) \frac{\partial \mathbf{C}(\sigma_A^2)}{\partial \sigma_A^2} \right)^2 \right] \quad \text{y} \quad \text{var}(\sigma_A^2) \geq \frac{1}{I(\sigma_A^2)}.$$

Por un lado, se ve que

$$\frac{\partial \mathbf{C}(\sigma_A^2)}{\partial \sigma_A^2} = \mathbf{1}\mathbf{1}^T.$$

Para calcular $\mathbf{C}^{-1}(\sigma_A^2)$ se usará la identidad de Woodbury, dada por la ecuación A.2 y demostrada en la sección A, que indica que

$$(\mathbf{A} + \mathbf{u}\mathbf{u}^T)^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{u}\mathbf{u}^T\mathbf{A}^{-1}}{1 + \mathbf{u}^T\mathbf{A}^{-1}\mathbf{u}}.$$

De esta forma,

$$\begin{aligned}\mathbf{C}^{-1}(\sigma_A^2) &= (\sigma^2\mathbf{I} + \sigma_A^2\mathbf{1}\mathbf{1}^T)^{-1} \\ &= \frac{\mathbf{I}}{\sigma^2} - \frac{\frac{\mathbf{I}}{\sigma^2}\sigma_A\mathbf{1}\sigma_A\mathbf{1}^T\frac{\mathbf{I}}{\sigma^2}}{1 + \sigma_A\mathbf{1}^T\frac{\mathbf{I}}{\sigma^2}\sigma_A\mathbf{1}} \\ &= \frac{\mathbf{I}}{\sigma^2} - \frac{\frac{\sigma_A^2}{\sigma^4}\mathbf{1}\mathbf{1}^T}{1 + N\frac{\sigma_A^2}{\sigma^2}} \\ &= \frac{\mathbf{I}}{\sigma^2} - \frac{\frac{\sigma_A^2}{\sigma^2}\mathbf{1}\mathbf{1}^T}{\sigma^2 + N\sigma_A^2},\end{aligned}$$

concluyendo que

$$\mathbf{C}^{-1}(\sigma_A^2) = \frac{1}{\sigma^2} \left(\mathbf{I} - \frac{\sigma_A^2}{\sigma^2 + N\sigma_A^2} \mathbf{1}\mathbf{1}^T \right).$$

Por lo tanto,

$$\begin{aligned}\mathbf{C}^{-1}(\sigma_A^2) \frac{\partial \mathbf{C}(\sigma_A^2)}{\partial \sigma_A^2} &= \frac{1}{\sigma^2} \left(\mathbf{I} - \frac{\sigma_A^2}{\sigma^2 + N\sigma_A^2} \mathbf{1}\mathbf{1}^T \right) \mathbf{1}\mathbf{1}^T \\ &\stackrel{(a)}{=} \frac{1}{\sigma^2} \left(\mathbf{1}\mathbf{1}^T - \frac{N\sigma_A^2}{\sigma^2 + N\sigma_A^2} \mathbf{1}\mathbf{1}^T \right) \\ &= \frac{1}{\sigma^2} \left(1 - \frac{N\sigma_A^2}{\sigma^2 + N\sigma_A^2} \right) \mathbf{1}\mathbf{1}^T \\ &= \frac{1}{\sigma^2} \left(\frac{\sigma^2}{\sigma^2 + N\sigma_A^2} \right) \mathbf{1}\mathbf{1}^T,\end{aligned}$$

resultando en

$$\mathbf{C}^{-1}(\sigma_A^2) \frac{\partial \mathbf{C}(\sigma_A^2)}{\partial \sigma_A^2} = \left(\frac{1}{\sigma^2 + N\sigma_A^2} \right) \mathbf{1}\mathbf{1}^T,$$

donde en (a) se notó que $\mathbf{1}\mathbf{1}^T\mathbf{1}\mathbf{1}^T = N\mathbf{1}\mathbf{1}^T$. Sustituyendo este resultado en la información de Fisher, se obtiene que

$$\begin{aligned}I(\sigma_A^2) &= \frac{1}{2} \text{tr} \left[\left(\frac{1}{\sigma^2 + N\sigma_A^2} \right)^2 \mathbf{1}\mathbf{1}^T\mathbf{1}\mathbf{1}^T \right] \\ &= \frac{1}{2} \text{tr} \left[\left(\frac{1}{\sigma^2 + N\sigma_A^2} \right)^2 N\mathbf{1}\mathbf{1}^T \right] \\ &= \frac{N}{2} \left(\frac{1}{\sigma^2 + N\sigma_A^2} \right)^2 \text{tr}(\mathbf{1}\mathbf{1}^T) \\ &= \frac{1}{2} \left(\frac{N}{\sigma^2 + N\sigma_A^2} \right)^2,\end{aligned}$$

y la CRLB es

$$\text{var}(\sigma_A^2) \geq 2 \left(\frac{\sigma^2 + N\sigma_A^2}{N} \right)^2 = 2 \left(\sigma_A^2 + \frac{\sigma^2}{N} \right)^2. \quad (3.39)$$

Se observa que incluso cuando $N \rightarrow \infty$, la CRLB no decrece a mas allá de $2\sigma_A^4$.

3.4.4. Ejemplo: estimación de los parámetros de una senoide

Un problema presente en diversas áreas de la ciencia es la estimación de los parámetros de una señal sinusoidal. Por lo tanto, se determinará la CRLB de la amplitud A , la frecuencia f_0 y la fase ϕ de una senoide en WGN. Se asume que se observan los datos

$$x[n] = A \cos(2\pi f_0 n + \phi) + w[n] \quad n = 0, 1, \dots, N-1,$$

con $A > 0$ y $0 < f_0 < 1/2$. Como en este caso, hay múltiples parámetros desconocidos, hay que calcular la matriz de información de Fisher empleando la ecuación 3.38,

$$[\mathbf{I}(\boldsymbol{\theta})]_{ij} = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} \frac{\partial s[n; \boldsymbol{\theta}]}{\partial \theta_i} \frac{\partial s[n; \boldsymbol{\theta}]}{\partial \theta_j},$$

con $\boldsymbol{\theta} = [A \ f_0 \ \phi]^T$ y $s[n; \boldsymbol{\theta}] = A \cos(2\pi f_0 n + \phi)$. En el cálculo de la CRLB se asumirá que f_0 no es cercano a 0 ni a $1/2$, ya que esto permite realizar ciertas simplificaciones en base a las aproximaciones

$$\begin{aligned} \frac{1}{N^{i+1}} \sum_{n=0}^{N-1} n^i \sin(4\pi f_0 n + 2\phi) &\approx 0 \\ \frac{1}{N^{i+1}} \sum_{n=0}^{N-1} n^i \cos(4\pi f_0 n + 2\phi) &\approx 0 \end{aligned}$$

para $i = 0, 1, 2$, las cuales se deducen en el problema 3.6.7 y en el apéndice B. Empleando estas aproximaciones y definiendo $\alpha = 2\pi f_0 n + \phi$, se tiene que

$$\begin{aligned} [\mathbf{I}(\boldsymbol{\theta})]_{11} &= \frac{1}{\sigma^2} \sum_{n=0}^{N-1} \left(\frac{\partial s[n; \boldsymbol{\theta}]}{\partial A} \right)^2 = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} \cos^2 \alpha \stackrel{(a)}{=} \frac{1}{\sigma^2} \sum_{n=0}^{N-1} \left(\frac{1}{2} + \frac{1}{2} \cos 2\alpha \right) \approx \frac{N}{2\sigma^2} \\ [\mathbf{I}(\boldsymbol{\theta})]_{12} &= \frac{1}{\sigma^2} \sum_{n=0}^{N-1} \frac{\partial s[n; \boldsymbol{\theta}]}{\partial A} \frac{\partial s[n; \boldsymbol{\theta}]}{\partial f_0} = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} (\cos \alpha)(-A 2\pi n \sin \alpha) \stackrel{(b)}{=} -\frac{\pi A}{\sigma^2} \sum_{n=0}^{N-1} n \sin 2\alpha \approx 0 \\ [\mathbf{I}(\boldsymbol{\theta})]_{13} &= \frac{1}{\sigma^2} \sum_{n=0}^{N-1} \frac{\partial s[n; \boldsymbol{\theta}]}{\partial A} \frac{\partial s[n; \boldsymbol{\theta}]}{\partial \phi} = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} (\cos \alpha)(-A \sin \alpha) \stackrel{(b)}{=} -\frac{A}{2\sigma^2} \sum_{n=0}^{N-1} \sin 2\alpha \approx 0 \\ [\mathbf{I}(\boldsymbol{\theta})]_{22} &= \frac{1}{\sigma^2} \sum_{n=0}^{N-1} \left(\frac{\partial s[n; \boldsymbol{\theta}]}{\partial f_0} \right)^2 = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} (-A 2\pi n \sin \alpha)^2 \stackrel{(c)}{=} \frac{4\pi^2 A^2}{\sigma^2} \sum_{n=0}^{N-1} n^2 \left(\frac{1}{2} - \frac{1}{2} \cos 2\alpha \right) \\ &\approx \frac{2\pi^2 A^2}{\sigma^2} \sum_{n=0}^{N-1} n^2 \stackrel{(d)}{=} \frac{2\pi^2 A^2 N(N-1)(2N-1)}{6\sigma^2} \\ [\mathbf{I}(\boldsymbol{\theta})]_{23} &= \frac{1}{\sigma^2} \sum_{n=0}^{N-1} \frac{\partial s[n; \boldsymbol{\theta}]}{\partial f_0} \frac{\partial s[n; \boldsymbol{\theta}]}{\partial \phi} = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} (-A 2\pi n \sin \alpha)(-A \sin \alpha) \stackrel{(c)}{=} \frac{2\pi A^2}{\sigma^2} \sum_{n=0}^{N-1} n \left(\frac{1}{2} - \frac{1}{2} \cos 2\alpha \right) \\ &\approx \frac{\pi A^2}{\sigma^2} \sum_{n=0}^{N-1} n \stackrel{(d)}{=} \frac{\pi A^2 N(N-1)}{2\sigma^2} \\ [\mathbf{I}(\boldsymbol{\theta})]_{33} &= \frac{1}{\sigma^2} \sum_{n=0}^{N-1} \left(\frac{\partial s[n; \boldsymbol{\theta}]}{\partial \phi} \right)^2 = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} (-A \sin \alpha)^2 \stackrel{(c)}{=} \frac{A^2}{\sigma^2} \sum_{n=0}^{N-1} \left(\frac{1}{2} - \frac{1}{2} \cos 2\alpha \right) \approx \frac{A^2 N}{2\sigma^2}, \end{aligned} \tag{3.40}$$

donde en (a), (b) y (c) se emplearon las identidades trigonométricas $2\cos^2 \theta = 1 + \cos 2\theta$, $2\sin \theta \cos \theta = \sin 2\theta$ y $2\sin^2 \theta = 1 - \cos 2\theta$, y en (d) se emplearon las identidades dadas en la ecuación 3.20. Se concluye que la matriz de información de Fisher resulta en

$$\mathbf{I}(\boldsymbol{\theta}) = \frac{N}{2\sigma^2} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{2\pi^2 A^2 (N-1)(2N-1)}{3} & \pi A^2 (N-1) \\ 0 & \pi A^2 (N-1) & A^2 \end{bmatrix}.$$

Para calcular la CRLB del estimador hay que invertir la matriz de información de Fisher. Teniendo en cuenta que la matriz es diagonal a bloques⁶, alcanza con invertir la submatriz 2×2

$$\mathbf{I}_{22}(\boldsymbol{\theta}) = \frac{A^2 N}{2\sigma^2} \begin{bmatrix} \frac{2\pi^2(N-1)(2N-1)}{3} & \pi(N-1) \\ \pi(N-1) & 1 \end{bmatrix}.$$

El determinante de la submatriz es

$$\begin{aligned} \det \left[\frac{2\sigma^2}{A^2 N} \mathbf{I}_{22}(\boldsymbol{\theta}) \right] &= \frac{2\pi^2(N-1)(2N-1)}{3} - \pi^2(N-1)^2 = \frac{\pi^2(N-1)}{3} [2(2N-1) - 3(N-1)] \\ &= \frac{\pi^2(N-1)(N+1)}{3} = \frac{\pi^2(N^2-1)}{3}, \end{aligned}$$

por lo que la submatriz inversa es

$$\begin{aligned} \mathbf{I}_{22}^{-1}(\boldsymbol{\theta}) &= \frac{6\sigma^2}{A^2 \pi^2 N(N^2-1)} \begin{bmatrix} 1 & -\pi(N-1) \\ -\pi(N-1) & \frac{2\pi^2(N-1)(2N-1)}{3} \end{bmatrix} \\ &= \frac{2\sigma^2}{N} \begin{bmatrix} \frac{3}{A^2 \pi^2 (N^2-1)} & -\frac{3}{A^2 \pi (N+1)} \\ -\frac{3}{A^2 \pi (N+1)} & \frac{2(2N-1)}{A^2 (N+1)} \end{bmatrix} \end{aligned}$$

y por lo tanto, la inversa de la matriz de información de Fisher es

$$\mathbf{I}^{-1}(\boldsymbol{\theta}) = \frac{2\sigma^2}{N} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{3}{A^2 \pi^2 (N^2-1)} & -\frac{3}{A^2 \pi (N+1)} \\ 0 & -\frac{3}{A^2 \pi (N+1)} & \frac{2(2N-1)}{A^2 (N+1)} \end{bmatrix}.$$

Se concluye que la CRLB de los parámetros es

$$\begin{aligned} \text{var}(\hat{A}) &\geq \frac{2\sigma^2}{N} \\ \text{var}(\hat{f}_0) &\geq \frac{6\sigma^2}{A^2 \pi^2 N(N^2-1)} = \frac{3}{\pi^2 \eta N(N^2-1)} \\ \text{var}(\hat{\phi}) &\geq \frac{4\sigma^2(2N-1)}{A^2 N(N+1)} = \frac{2(2N-1)}{\eta N(N+1)}, \end{aligned} \tag{3.41}$$

donde $\eta = A^2/(2\sigma^2)$ es la SNR. El problema de la estimación de la frecuencia de una senoide es de considerable interés. Se observa que la CRLB de la frecuencia decrece cuando decrece la relación señal a ruido. Además, la cota decrece como $1/N^3$, haciéndola muy sensible a la cantidad de observaciones.

3.5. CRLB asintótica para procesos WSS gaussianos

En ocasiones es difícil calcular la CRLB analíticamente usando la ecuación 3.24 debido a que se requiere la inversión de la matriz de covarianza. Se presentará a continuación una forma alternativa válida para procesos gaussianos estacionarios en sentido amplio (*wide sense stationary*, WSS) en el caso

⁶Ver https://en.wikipedia.org/wiki/Block_matrix#Block_diagonal_matrices

en que $N \rightarrow \infty$. En la práctica, esta forma brinda una buena aproximación en el caso en que la cantidad N de datos es mucho mayor que el tiempo de correlación del proceso. El tiempo de correlación es el retardo máximo de la función de autocorrelación del proceso (*autocorrelation function*, ACF) con un valor significativo.

3.5.1. Teorema

Se demostrará que los elementos de la matriz de información de Fisher cuando $N \rightarrow \infty$ son aproximadamente

$$[\mathbf{I}(\boldsymbol{\theta})]_{ij} = \frac{N}{2} \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{\partial \ln P_{xx}(f; \boldsymbol{\theta})}{\partial \theta_i} \frac{\partial \ln P_{xx}(f; \boldsymbol{\theta})}{\partial \theta_j} df, \quad (3.42)$$

donde $P_{xx}(f; \boldsymbol{\theta})$ es la densidad espectral de potencia (*power spectral density*, PSD) del proceso. También se asume que $x[n]$ tiene media nula.

Demostración El teorema de la descomposición de Wold (ver por ejemplo, el capítulo 1 de [4]) indica que cualquier proceso WSS gaussiano puede representarse como la salida de un filtro causal $h[n]$ cuya entrada $u[n]$ es ruido blanco gaussiano,

$$x[n] = \sum_{k=0}^{\infty} h[k]u[n-k],$$

con $h[0] = 1$. Una representación de esta forma se denomina modelo de media móvil, $\text{MA}(\infty)$ (*moving average*, MA), donde el ∞ se refiere al largo de la respuesta al impulso. Con esta representación, la PSD de $x[n]$ es

$$P_{xx}(f) = |H(f)|^2 \sigma_u^2, \quad (3.43)$$

donde σ_u^2 es la varianza de $u[n]$ y $H(f)$ es la respuesta en frecuencia del filtro,

$$H(f) = \sum_{k=0}^{\infty} h[k]e^{-j2\pi f k}.$$

Esto proviene de que la PSD de un proceso $x[n]$ obtenido de filtrar un proceso $u[n]$ WSS con PSD $P_{uu}(f)$ mediante un filtro de respuesta en frecuencia $H(f)$ es $P_{xx}(f) = |H(f)|^2 P_{uu}(f)$ (ver por ejemplo, el capítulo 3 de [5]), y como en este caso $u[n]$ es WGN, su función de autocorrelación es $r_{uu}[k] = \sigma_u^2 \delta[k]$ y la PSD es $P_{uu}(f) = \sigma_u^2$.

La derivación de la CRLB asintótica se basará en la siguiente aproximación,

$$\begin{aligned} x[n] &= \sum_{k=0}^n h[k]u[n-k] + \sum_{k=n+1}^{\infty} h[k]u[n-k] \\ &\approx \sum_{k=0}^n h[k]u[n-k], \end{aligned} \quad (3.44)$$

que es equivalente a considerar $u[n] = 0$ en $n < 0$. Con este modelo, las primeras muestras de $x[n]$ estarán pobremente representadas a menos que $h[k]$ sea pequeño para $k > n$, pero para valores grandes de n , mayores al largo de $h[k]$, la representación es buena. Por lo tanto, si las observaciones son $\{x[0], x[1], \dots, x[N-1]\}$ y N es mucho mayor que el largo de la respuesta al impulso $h[k]$, la mayoría de las muestras de $x[n]$ están bien representadas por el modelo de la ecuación 3.44. Además, como

$$r_{xx}[k] = \sigma_u^2 \sum_{n=0}^{\infty} h[n]h[n+k],$$

ya que $r_{xx}[k] = r_{uu}[k] * h[k] * k[-k]$ (ver el capítulo 3 de [5]), el largo de la correlación del proceso $x[n]$ coincide con el largo de la respuesta al impulso $h[k]$. Se concluye que como la deducción de la CRLB asintótica está basada en la aproximación de la ecuación 3.44, será una buena aproximación si el largo N de los datos es mucho mayor al largo de la correlación $r_{xx}[k]$ del proceso $x[n]$.

Para calcular la CRLB, hay que calcular los elementos de la matriz de información de Fisher. Para eso, se parte calculando la PDF de los datos \mathbf{x} . El modelo de los datos de la ecuación 3.44 en forma matricial es

$$\mathbf{x} = \mathbf{H}\mathbf{u},$$

donde $\mathbf{u} = [u[0] u[1] \dots u[N-1]]^T$, $\mathbf{x} = [x[0] x[1] \dots x[N-1]]^T$ y

$$\mathbf{H} = \begin{bmatrix} h[0] & 0 & \dots & 0 \\ h[1] & h[0] & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ h[N-1] & h[N-2] & \dots & h[0] \end{bmatrix}.$$

La matriz de covarianza de \mathbf{u} es

$$C_u = E(\mathbf{u}\mathbf{u}^T) = \sigma_u^2 \mathbf{I},$$

y por lo tanto, $\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \sigma_u^2 \mathbf{I})$. La matriz de covarianza de \mathbf{x} es

$$C_x = E[\mathbf{x}\mathbf{x}^T] = E[\mathbf{H}\mathbf{u}(\mathbf{H}\mathbf{u})^T] = \mathbf{H}E[\mathbf{u}\mathbf{u}^T]\mathbf{H}^T = \sigma_u^2 \mathbf{H}\mathbf{H}^T$$

y por lo tanto, la PDF de \mathbf{x} es $\mathcal{N}(\mathbf{0}, \sigma_u^2 \mathbf{H}\mathbf{H}^T)$, es decir,

$$p(\mathbf{x}; \boldsymbol{\theta}) = \frac{1}{(2\pi)^{\frac{N}{2}} \det^{\frac{1}{2}}(\sigma_u^2 \mathbf{H}\mathbf{H}^T)} \exp \left[-\frac{1}{2} \mathbf{x}^T (\sigma_u^2 \mathbf{H}\mathbf{H}^T)^{-1} \mathbf{x} \right].$$

Considerando que

$$\det(\sigma_u^2 \mathbf{H}\mathbf{H}^T) = \sigma_u^{2N} \det(\mathbf{H}) \det(\mathbf{H}^T) = \sigma_u^{2N} \det^2(\mathbf{H}) = \sigma_u^{2N} (h[0])^{2N} = \sigma_u^{2N},$$

donde en las identidades se emplearon propiedades del determinante⁷, que el determinante de \mathbf{H} es su traza por ser una matriz triangular y que $h[0] = 1$, y teniendo en cuenta además que

$$\mathbf{x}^T (\sigma_u^2 \mathbf{H}\mathbf{H}^T)^{-1} \mathbf{x} = \frac{1}{\sigma_u^2} \mathbf{x}^T (\mathbf{H}^T)^{-1} \mathbf{H}^{-1} \mathbf{x} = \frac{1}{\sigma_u^2} \mathbf{x}^T (\mathbf{H}^{-1})^T \mathbf{H}^{-1} \mathbf{x} = \frac{1}{\sigma_u^2} (\mathbf{H}^{-1} \mathbf{x})^T (\mathbf{H}^{-1} \mathbf{x}) = \frac{1}{\sigma_u^2} \mathbf{u}^T \mathbf{u},$$

la PDF de los datos queda

$$p(\mathbf{x}; \boldsymbol{\theta}) = \frac{1}{(2\pi\sigma_u^2)^{\frac{N}{2}}} \exp \left(-\frac{1}{2\sigma_u^2} \mathbf{u}^T \mathbf{u} \right),$$

y tomando el logaritmo se obtiene que

$$\ln p(\mathbf{x}; \boldsymbol{\theta}) = -\frac{N}{2} \ln 2\pi - \frac{N}{2} \ln \sigma_u^2 - \frac{1}{2\sigma_u^2} \mathbf{u}^T \mathbf{u}. \quad (3.45)$$

Aplicando la transformada de Fourier a la ecuación 3.44 y usando el teorema de la convolución de la transformada de Fourier (ver el capítulo 2 de [6]), se cumple que

$$X(f) \approx H(f)U(f), \quad (3.46)$$

donde

$$X(f) = \sum_{n=0}^{N-1} x[n] e^{-j2\pi f n} \quad \text{y} \quad U(f) = \sum_{n=0}^{N-1} u[n] e^{-j2\pi f n}$$

son las transformadas de Fourier de las secuencias truncadas. Además, se tiene que

$$\begin{aligned} \frac{1}{\sigma_u^2} \mathbf{u}^T \mathbf{u} &= \frac{1}{\sigma_u^2} \sum_{n=0}^{N-1} u^2[n] \\ &\stackrel{(a)}{=} \frac{1}{\sigma_u^2} \int_{-\frac{1}{2}}^{\frac{1}{2}} |U(f)|^2 df \\ &\stackrel{(b)}{\approx} \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{|X(f)|^2}{\sigma_u^2 |H(f)|^2} df \\ &\stackrel{(c)}{=} \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{|X(f)|^2}{P_{xx}(f)} df, \end{aligned} \quad (3.47)$$

⁷Ver https://en.wikipedia.org/wiki/Determinant#Properties_of_the_determinant

donde en (a) se empleó la identidad de Parseval (ver el capítulo 2 de [6]), en (b) se empleó la ecuación 3.46 y en (c) se empleó la ecuación 3.43.

Por otro lado, se observa que

$$\begin{aligned}\ln \sigma_u^2 &= \int_{-\frac{1}{2}}^{\frac{1}{2}} \ln \sigma_u^2 df \\ &\stackrel{(a)}{=} \int_{-\frac{1}{2}}^{\frac{1}{2}} \ln \frac{P_{xx}(f)}{|H(f)|^2} df \\ &= \int_{-\frac{1}{2}}^{\frac{1}{2}} \ln P_{xx}(f) df - \int_{-\frac{1}{2}}^{\frac{1}{2}} \ln |H(f)|^2 df,\end{aligned}$$

donde en (a) se empleó la ecuación 3.43. Además, como $|H(f)|^2 = H(f)H^*(f)$, se cumple que

$$\begin{aligned}\int_{-\frac{1}{2}}^{\frac{1}{2}} \ln |H(f)|^2 df &= \int_{-\frac{1}{2}}^{\frac{1}{2}} \ln H(f) df + \int_{-\frac{1}{2}}^{\frac{1}{2}} \ln H^*(f) df \\ &= \int_{-\frac{1}{2}}^{\frac{1}{2}} \ln H(f) df + \left(\int_{-\frac{1}{2}}^{\frac{1}{2}} \ln H(f) df \right)^* \\ &= 2 \operatorname{Re} \int_{-\frac{1}{2}}^{\frac{1}{2}} \ln H(f) df \\ &\stackrel{(a)}{=} 2 \operatorname{Re} \left[\frac{1}{2\pi j} \oint_C \ln H(z) z^{-1} dz \right] \\ &\stackrel{(b)}{=} 2 \operatorname{Re} \left[\mathcal{Z}^{-1} \{ \ln H(z) \} \Big|_{n=0} \right],\end{aligned}\tag{3.48}$$

donde $H(z)$ es función de transferencia del sistema, es decir, la transformada Z de $h[k]$ y C es la circunferencia unidad. Las identidades (a) y (b) provienen de igualar las transformadas de Fourier y Z inversas,

$$x[n] = \int_{-\frac{1}{2}}^{\frac{1}{2}} X(f) e^{-j2\pi f n} df = \frac{1}{2\pi j} \oint_C X(z) z^{n-1} dz,$$

evaluadas en $n = 0$,

$$x[0] = \int_{-\frac{1}{2}}^{\frac{1}{2}} X(f) df = \frac{1}{2\pi j} \oint_C X(z) z^{-1} dz.$$

Como $H(z)$ es a la función de transferencia de un filtro causal, converge fuera de un círculo de radio $r < 1$, ya que $H(z)$ debe converger en la circunferencia unidad para que exista la respuesta en frecuencia y el sistema sea estable. Asumiendo además que $H(z)$ es de fase mínima y por lo tanto todos sus ceros se encuentran dentro del círculo unidad, $\ln H(z)$ también converge fuera del círculo de radio $r < 1$, por lo que la secuencia correspondiente también es causal. El teorema del valor inicial de la transformada Z indica que para una secuencia $x[n]$ que es nula en $n < 0$ se cumple que

$$\lim_{z \rightarrow \infty} X(z) = x[0],$$

ya que

$$\lim_{z \rightarrow \infty} X(z) = \lim_{z \rightarrow \infty} \sum_{n=-\infty}^{\infty} x[n] z^{-n} = \lim_{z \rightarrow \infty} \sum_{n=0}^{\infty} x[n] z^{-n} = \lim_{z \rightarrow \infty} x[0] + x[1] z^{-1} + x[2] z^{-2} + \dots = x[0].$$

Por lo tanto,

$$\begin{aligned}\mathcal{Z}^{-1} \{ \ln H(z) \} \Big|_{n=0} &= \lim_{z \rightarrow \infty} \ln H(z) \\ &= \ln \lim_{z \rightarrow \infty} H(z) \\ &= \ln h[0] \\ &= 0,\end{aligned}$$

concluyendo que

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \ln |H(f)|^2 df = 0,$$

y finalmente que

$$\ln \sigma_u^2 = \int_{-\frac{1}{2}}^{\frac{1}{2}} \ln P_{xx}(f) df. \quad (3.49)$$

Sustituyendo las ecuaciones 3.47 y 3.49 en la ecuación 3.45 se obtiene la PDF logarítmica asintótica,

$$\ln p(\mathbf{x}; \boldsymbol{\theta}) = -\frac{N}{2} \ln 2\pi - \frac{N}{2} \int_{-\frac{1}{2}}^{\frac{1}{2}} \ln P_{xx}(f) df - \frac{1}{2} \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{|X(f)|^2}{P_{xx}(f)} df,$$

es decir,

$$\ln p(\mathbf{x}; \boldsymbol{\theta}) = -\frac{N}{2} \ln 2\pi - \frac{N}{2} \int_{-\frac{1}{2}}^{\frac{1}{2}} \left[\ln P_{xx}(f) + \frac{1}{N} \frac{|X(f)|^2}{P_{xx}(f)} \right] df. \quad (3.50)$$

Diferenciando respecto a θ_i y luego respecto a θ_j para determinar la CRLB, se tiene que

$$\frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \theta_i} = -\frac{N}{2} \int_{-\frac{1}{2}}^{\frac{1}{2}} \left[\frac{1}{P_{xx}(f)} - \frac{1}{N} \frac{|X(f)|^2}{P_{xx}^2(f)} \right] \frac{\partial P_{xx}(f)}{\partial \theta_i} df, \quad (3.51)$$

donde se intercambi6 el orden de diferenciación e integración. Además,

$$\begin{aligned} \frac{\partial^2 \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} = & -\frac{N}{2} \int_{-\frac{1}{2}}^{\frac{1}{2}} \left[\left(-\frac{1}{P_{xx}^2(f)} + \frac{2}{N} \frac{|X(f)|^2}{P_{xx}^3(f)} \right) \frac{\partial P_{xx}(f)}{\partial \theta_i} \frac{\partial P_{xx}(f)}{\partial \theta_j} \right. \\ & \left. + \left(\frac{1}{P_{xx}(f)} - \frac{1}{N} \frac{|X(f)|^2}{P_{xx}^2(f)} \right) \frac{\partial^2 P_{xx}(f)}{\partial \theta_i \partial \theta_j} \right] df, \end{aligned} \quad (3.52)$$

Al tomar la esperanza, aparece el término $E(|X(f)|^2/N)$. Se demostrará a continuación que este término es $P_{xx}(f)$ si N es grande. $|X(f)|^2/N$ se denomina *periodograma*, y es un estimador de la PSD de los datos. Se parte observando que

$$|X(f)|^2 = X(f)X^*(f) = \left(\sum_{m=0}^{N-1} x[m]e^{-j2\pi f m} \right) \left(\sum_{n=0}^{N-1} x[n]e^{-j2\pi f n} \right)^* = \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} x[m]x[n]e^{-j2\pi f(m-n)},$$

y por lo tanto,

$$\begin{aligned} E \left(\frac{1}{N} |X(f)|^2 \right) &= E \left(\frac{1}{N} \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} x[m]x[n]e^{-j2\pi f(m-n)} \right) \\ &= \frac{1}{N} \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} E(x[m]x[n])e^{-j2\pi f(m-n)} \\ &\stackrel{(a)}{=} \frac{1}{N} \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} r_{xx}[m-n]e^{-j2\pi f(m-n)} \\ &\stackrel{(b)}{=} \sum_{k=-(N-1)}^{N-1} \left(1 - \frac{|k|}{N} \right) r_{xx}[k]e^{-j2\pi f k}, \end{aligned}$$

donde en (a) se consider6 que el proceso $x[n]$ es WSS y por lo tanto, $r_{xx}[m, n] = r_{xx}[m-n]$ y en (b) se

tuvo en cuenta que

$$\begin{aligned}
\sum_{m=0}^{N-1} \sum_{n=0}^{N-1} g[m-n] &= \begin{matrix} g[0] + g[-1] + g[-2] + \dots + g[-N+2] + g[-N+1] \\ g[1] + g[0] + g[-1] + g[-2] + \dots + g[-N+2] + \\ g[2] + g[1] + g[0] + g[-1] + g[-2] + \dots + \\ \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \\ g[N-1] + \dots + g[2] + g[1] + g[0] \end{matrix} \\
&= \sum_{k=-(N-1)}^{N-1} (N - |k|)g[k].
\end{aligned}$$

Por lo tanto, cuando $N \rightarrow \infty$,

$$\begin{aligned}
\lim_{N \rightarrow \infty} E \left(\frac{1}{N} |X(f)|^2 \right) &= \lim_{N \rightarrow \infty} \sum_{k=-(N-1)}^{N-1} \left(1 - \frac{|k|}{N} \right) r_{xx}[k] e^{-j2\pi f k} \\
&= \sum_{k=-\infty}^{\infty} r_{xx}[k] e^{-j2\pi f k} \\
&= P_{xx}(f).
\end{aligned}$$

y al tomar la esperanza en 3.52 se obtiene que

$$\begin{aligned}
[\mathbf{I}(\boldsymbol{\theta})]_{ij} &= \frac{N}{2} \int_{-\frac{1}{2}}^{\frac{1}{2}} \left[\left(-\frac{1}{P_{xx}^2(f)} + \frac{2}{P_{xx}(f)} \right) \frac{\partial P_{xx}(f)}{\partial \theta_i} \frac{\partial P_{xx}(f)}{\partial \theta_j} + \left(\frac{1}{P_{xx}(f)} - \frac{1}{P_{xx}^2(f)} \right) \frac{\partial^2 P_{xx}(f)}{\partial \theta_i \partial \theta_j} \right] df \\
&= \frac{N}{2} \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{1}{P_{xx}^2(f)} \frac{\partial P_{xx}(f)}{\partial \theta_i} \frac{\partial P_{xx}(f)}{\partial \theta_j} df,
\end{aligned}$$

que se puede expresar como

$$[\mathbf{I}(\boldsymbol{\theta})]_{ij} = \frac{N}{2} \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{\partial \ln P_{xx}(f)}{\partial \theta_i} \frac{\partial \ln P_{xx}(f)}{\partial \theta_j} df,$$

que es lo que se quería demostrar.

3.5.2. Ejemplo: estimación de los parámetros de un modelo autorregresivo

En el procesamiento de señales de voz, un modelo importante del mecanismo de producción de la señal de voz es como un proceso autorregresivo (AR, *autoregressive*), o equivalentemente, de predicción lineal (LPC, *linear predictive coding*) [7]. En este modelo, la señal se considera como la salida un filtro causal todo polos excitado por WGN. El filtro todo polos modela el aparato fonador, mientras que el ruido de excitación modela el pasaje del aire a través de las restricciones del aparato fonador de forma de producir un sonido sordo, como el de una “s” por ejemplo. El efecto del filtro es colorear el ruido blanco para obtener una señal de salida con PSD con múltiples resonancias como ocurre con las señales de voz. El modelo AR es capaz de producir una gran variedad de PSD en función de la elección de los parámetros $\{a[1], a[2], \dots, a[p]\}$ del filtro AR y la varianza σ_u^2 de la excitación. La PSD implicada por el modelo AR es

$$P_{xx}(f; \boldsymbol{\theta}) = \frac{\sigma_u^2}{\left| 1 + \sum_{m=1}^p a[m] e^{-j2\pi f m} \right|^2}, \quad (3.53)$$

donde $\boldsymbol{\theta} = [a[1] a[2] \dots a[p] \sigma_u^2]^T$. En base a la observación de los datos $\{x[0], x[1], \dots, x[N-1]\}$, se obtiene la estimación $\hat{\boldsymbol{\theta}} = [\hat{a}[1] \hat{a}[2] \dots \hat{a}[p] \hat{\sigma}_u^2]^T$ de los parámetros, y la PSD estimada como

$$\hat{P}_{xx}(f; \boldsymbol{\theta}) = \frac{\hat{\sigma}_u^2}{\left| 1 + \sum_{m=1}^p \hat{a}[m] e^{-j2\pi f m} \right|^2}.$$

El objetivo de este ejemplo es estimar la CRLB de los parámetros $\hat{\boldsymbol{\theta}}$ del modelo AR. El cálculo exacto se obtiene a partir de la ecuación 3.24 y resulta ser considerablemente tedioso. En la práctica, la CRLB asintótica provee un resultado apropiado, incluso para largo de datos relativamente pequeños si los polos del modelo no están demasiado cerca del círculo unidad. Por lo tanto, se calculará la CRLB asintótica. El cálculo exacto de la CRLB de los parámetros de un proceso AR puede encontrarse en [8].

Para calcular la CRLB asintótica de los parámetros, se empleará la ecuación 3.42. El logaritmo de la PSD es

$$\ln P_{xx}(f; \boldsymbol{\theta}) = \ln \sigma_u^2 - \ln |A(f)|^2,$$

donde

$$A(f) = 1 + \sum_{m=1}^p a[m] e^{-j2\pi f m},$$

y las derivadas parciales son

$$\begin{aligned} \frac{\partial \ln P_{xx}(f; \boldsymbol{\theta})}{\partial a[k]} &= -\frac{1}{|A(f)|^2} \frac{\partial |A(f)|^2}{\partial a[k]} \\ &= -\frac{1}{|A(f)|^2} \frac{\partial A(f) A^*(f)}{\partial a[k]} \\ &= -\frac{1}{|A(f)|^2} \left[A(f) \frac{\partial A^*(f)}{\partial a[k]} + A^*(f) \frac{\partial A(f)}{\partial a[k]} \right] \\ &= -\frac{1}{|A(f)|^2} [A(f) e^{j2\pi f k} + A^*(f) e^{-j2\pi f k}], \end{aligned} \quad (3.54)$$

$$\frac{\partial \ln P_{xx}(f; \boldsymbol{\theta})}{\partial \sigma_u^2} = \frac{1}{\sigma_u^2}. \quad (3.55)$$

Por lo tanto, sustituyendo el resultado de la ecuación 3.54 en la ecuación 3.42, se obtiene que para $k = 1, 2, \dots, p$; $l = 1, 2, \dots, p$,

$$\begin{aligned} [\mathbf{I}(\boldsymbol{\theta})]_{kl} &= \frac{N}{2} \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{1}{|A(f)|^4} [A(f) e^{j2\pi f k} + A^*(f) e^{-j2\pi f k}] [A(f) e^{j2\pi f l} + A^*(f) e^{-j2\pi f l}] df, \\ &= \frac{N}{2} \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{1}{|A(f)|^4} [A^2(f) e^{j2\pi f(k+l)} + |A(f)|^2 e^{j2\pi f(k-l)} \\ &\quad + |A(f)|^2 e^{j2\pi f(l-k)} + A^{*2}(f) e^{-j2\pi f(k+l)}] df \\ &= \frac{N}{2} \int_{-\frac{1}{2}}^{\frac{1}{2}} \left[\frac{1}{A^{*2}(f)} e^{j2\pi f(k+l)} + \frac{1}{|A(f)|^2} e^{j2\pi f(k-l)} \right. \\ &\quad \left. + \frac{1}{|A(f)|^2} e^{j2\pi f(l-k)} + \frac{1}{A^2(f)} e^{-j2\pi f(k+l)} \right] df. \end{aligned}$$

Observando que como $A(-f) = A^*(f)$ se cumple que

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{1}{A^{*2}(f)} e^{j2\pi f(k+l)} df = - \int_{\frac{1}{2}}^{-\frac{1}{2}} \frac{1}{A^{*2}(-u)} e^{-j2\pi u(k+l)} du = \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{1}{A^2(u)} e^{-j2\pi u(k+l)} du$$

y

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{1}{|A(f)|^2} e^{j2\pi f(k+l)} df = - \int_{\frac{1}{2}}^{-\frac{1}{2}} \frac{1}{|A(-u)|^2} e^{-j2\pi u(k+l)} du = \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{1}{|A(u)|^2} e^{-j2\pi u(k+l)} du,$$

donde en ambas identidades se realizó el cambio de variable $f = -u$, y por lo tanto,

$$[\mathbf{I}(\boldsymbol{\theta})]_{kl} = N \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{1}{|A(f)|^2} e^{j2\pi f(k-l)} df + N \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{1}{A^{*2}(f)} e^{j2\pi f(k+l)} df.$$

La segunda integral es la transformada de Fourier inversa de $1/A^{*2}(f)$ evaluada en $n = k + l > 0$. Como

$$\mathcal{F}^{-1} \left\{ \frac{1}{A(f)} \right\} = h[n] \quad \Rightarrow \quad \mathcal{F}^{-1} \left\{ \frac{1}{A^*(f)} \right\} = h[-n] \quad \Rightarrow \quad \mathcal{F}^{-1} \left\{ \frac{1}{A^{*2}(f)} \right\} = h[-n] * h[-n]$$

y como $h[n]$ es causal, $h[-n]$ y $h[-n] * h[-n]$ son anticausales, es decir, valen cero en $n > 0$, por lo que al evaluar en $n = k + l > 0$, se concluye que la segunda integral es nula. Por lo tanto,

$$[\mathbf{I}(\boldsymbol{\theta})]_{kl} = N \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{1}{|A(f)|^2} e^{j2\pi f(k-l)} df = N \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{P_{xx}(f)}{\sigma_u^2} e^{j2\pi f(k-l)} df = \frac{N}{\sigma_u^2} \int_{-\frac{1}{2}}^{\frac{1}{2}} P_{xx}(f) e^{j2\pi f(k-l)} df,$$

y como la última integral es la transformada inversa de la PSD evaluada en $n = k - l$, se concluye que

$$[\mathbf{I}(\boldsymbol{\theta})]_{kl} = \frac{N}{\sigma_u^2} r_{xx}[k - l]. \quad (3.56)$$

Además, para $k = 1, 2, \dots, p$; $l = p + 1$, la ecuación 3.42 queda

$$\begin{aligned} [\mathbf{I}(\boldsymbol{\theta})]_{kl} &= \frac{N}{2} \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{\partial \ln P_{xx}(f; \boldsymbol{\theta})}{\partial a[k]} \frac{\partial \ln P_{xx}(f; \boldsymbol{\theta})}{\partial \sigma_u^2} df \\ &\stackrel{(a)}{=} -\frac{N}{2} \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{1}{\sigma_u^2} \frac{1}{|A(f)|^2} [A(f)e^{j2\pi f k} + A^*(f)e^{-j2\pi f k}] df \\ &= -\frac{N}{2\sigma_u^2} \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{1}{A^*(f)} e^{j2\pi f k} df - \frac{N}{2\sigma_u^2} \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{1}{A(f)} e^{-j2\pi f k} df \\ &\stackrel{(b)}{=} -\frac{N}{\sigma_u^2} \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{1}{A^*(f)} e^{j2\pi f k} df \\ &\stackrel{(c)}{=} 0, \end{aligned} \quad (3.57)$$

donde en (a) se emplearon los resultados de las ecuaciones 3.54 y 3.55, en (b) se notó que ambas integrales son iguales debido a que $A(-f) = A^*(f)$ y en (c) se consideró la anticausalidad de $\mathcal{F}^{-1}\{1/A^*(f)\}$ de forma análoga a lo razonado antes.

Finalmente, para $k = p + 1$; $l = p + 1$, la ecuación 3.42 queda

$$\begin{aligned} [\mathbf{I}(\boldsymbol{\theta})]_{kl} &= \frac{N}{2} \int_{-\frac{1}{2}}^{\frac{1}{2}} \left(\frac{\partial \ln P_{xx}(f; \boldsymbol{\theta})}{\partial \sigma_u^2} \right)^2 df \\ &= \frac{N}{2} \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{1}{\sigma_u^4} df \\ &= \frac{N}{2\sigma_u^4}. \end{aligned} \quad (3.58)$$

Combinando los resultados de las ecuaciones 3.56, 3.57 y 3.58, se obtiene que la matriz de información de Fisher es

$$\mathbf{I}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{N}{\sigma_u^2} \mathbf{R}_{xx} & \mathbf{0} \\ \mathbf{0}^T & \frac{N}{2\sigma_u^4} \end{bmatrix}, \quad (3.59)$$

donde $[\mathbf{R}_{xx}]_{ij} = r_{xx}[i - j]$ es la matriz $p \times p$ Toeplitz de autocorrelación del proceso $x[n]$ y $\mathbf{0}$ es un vector $p \times 1$ de ceros. Como la matriz es diagonal a bloques, la inversa es⁸

$$\mathbf{I}^{-1}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\sigma_u^2}{N} \mathbf{R}_{xx}^{-1} & \mathbf{0} \\ \mathbf{0}^T & \frac{2\sigma_u^4}{N} \end{bmatrix},$$

⁸Ver https://en.wikipedia.org/wiki/Block_matrix#Block_diagonal_matrices

por lo que la CRLB de los parámetros es

$$\begin{aligned}\text{var}(\hat{a}[k]) &\geq \frac{\sigma_u^2}{N} [\mathbf{R}_{xx}^{-1}]_{kk}, \quad k = 1, 2, \dots, p \\ \text{var}(\hat{\sigma}_u^2) &\geq \frac{2\sigma_u^4}{N}.\end{aligned}$$

Como ilustración, considérese el caso con $p = 1$. Como \mathbf{R}_{xx} es el escalar

$$\mathbf{R}_{xx} = r_{xx}[1 - 1] = r_{xx}[0],$$

se tiene que

$$\text{var}(\hat{a}[1]) \geq \frac{\sigma_u^2}{Nr_{xx}[0]}.$$

Se calculará a continuación $r_{xx}[0]$. La función de transferencia del filtro AR es

$$H(z) = \frac{1}{1 + a[1]z^{-1}} = \frac{X(z)}{U(z)},$$

por lo que el proceso AR es

$$x[n] + a[1]x[n-1] = u[n].$$

La autocorrelación en cero es

$$\begin{aligned}r_{xx}[0] &= E(x^2[n]) \\ &= E\{(-a[1]x[n-1] + u[n])^2\} \\ &= a^2[1]E(x^2[n-1]) - 2a[1]E(x[n-1]u[n]) + E(u^2[n]) \\ &= a^2[1]r_{xx}[0] + \sigma_u^2,\end{aligned}$$

donde en la última igualdad se notó que $x[n-1]$ y $u[n]$ son independientes debido a que como el proceso es causal, muestras pasadas de la salida no dependen de la muestra actual de la entrada, y por lo tanto, $E(x[n-1]u[n]) = E(x[n-1])E(u[n])$ y $E(u[n]) = 0$. Despejando, se llega a que

$$r_{xx}[0] = \frac{\sigma_u^2}{1 - a^2[1]},$$

y por lo tanto, la CRLB queda

$$\text{var}(\hat{a}[1]) \geq \frac{1}{N}(1 - a^2[1]). \quad (3.60)$$

Esto indica que es mas fácil estimar el parámetro del filtro cuando $|a[1]|$ es mas cercano a uno que a cero. Como el polo del filtro AR está en $-a[1]$, significa que parámetros de filtros de procesos con PSD con picos agudos son estimados mas fácilmente.

3.6. Problemas

3.6.1. Problema 1

Si $x[n]$ para $n = 0, 1, \dots, N-1$ son IID con PDF $\mathcal{U}[0, \theta]$, mostrar que la condición de regularidad no se cumple, o que

$$E \left[\frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \theta} \right] \neq 0, \quad \text{para todo } \theta > 0. \quad (3.61)$$

Por lo tanto, la CRLB no puede ser aplicada en este problema.

Solución Sea $p(x[n]; \theta)$ la densidad de probabilidad de la muestra $x[n]$. En el caso general en que las muestras son independientes, se cumple que

$$p(\mathbf{x}; \theta) = \prod_{n=0}^{N-1} p(x[n]; \theta),$$

Tomando el logaritmo, se obtiene

$$\ln p(\mathbf{x}; \theta) = \sum_{n=0}^{N-1} \ln p(x[n]; \theta),$$

y derivando respecto a θ se llega a que

$$\frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \theta} = \sum_{n=0}^{N-1} \frac{\partial \ln p(x[n]; \theta)}{\partial \theta}.$$

Finalmente, tomando la esperanza, se concluye que

$$E \left[\frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \theta} \right] = \sum_{n=0}^{N-1} E \left[\frac{\partial \ln p(x[n]; \theta)}{\partial \theta} \right].$$

En el caso particular en que las muestras son idénticamente distribuidas, la expresión anterior queda

$$E \left[\frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \theta} \right] = N E \left[\frac{\partial \ln p(x[n]; \theta)}{\partial \theta} \right], \quad \text{para cualquier } n = 0, \dots, N-1,$$

y por lo tanto, para demostrar que no se cumple la condición de regularidad (ecuación 3.61), alcanza con demostrar que

$$E \left[\frac{\partial \ln p(x[n]; \theta)}{\partial \theta} \right] \neq 0.$$

Como en este caso las muestras $x[n]$ son IID con $x[n] \sim \mathcal{U}[0, \theta]$, la PDF para todo n es

$$p(x[n]; \theta) = \begin{cases} 1/\theta, & 0 \leq x[n] \leq \theta \\ 0, & \text{en otro caso.} \end{cases}$$

Cuando se considera a la PDF como una función del parámetro desconocido para un valor fijo de los datos, como en la condición de regularidad, se denomina *función de verosimilitud*, y en este caso es

$$p(x[n]; \theta) = \begin{cases} 0, & \theta < x[n] \\ 1/\theta, & \theta \geq x[n], \end{cases}$$

la cual se muestra en la figura 3.1. Para $\theta > x[n]$ se tiene que $p(x[n]; \theta) = 1/\theta$ y por lo tanto,

$$E \left[\frac{\partial \ln p(x[n]; \theta)}{\partial \theta} \right] = E \left[\frac{\partial}{\partial \theta} \left(\ln \frac{1}{\theta} \right) \right] = E \left[\frac{\partial}{\partial \theta} (-\ln \theta) \right] = E \left[-\frac{1}{\theta} \right] = -\frac{1}{\theta} \neq 0.$$

3.6.2. Problema 2

Se considera el caso en que se observa una única muestra

$$x[0] = A + w[0],$$

donde $w[0]$ tiene una PDF $p(w[0])$ arbitraria y se desea estimar el parámetro A . Mostrar que la CRLB para A es

$$\text{var}(\hat{A}) \geq \left[\int_{-\infty}^{\infty} \frac{\left(\frac{dp(u)}{du} \right)^2}{p(u)} du \right]^{-1}.$$

Evaluar la CRLB en el caso en que la PDF es de Laplace, es decir,

$$p(w[0]) = \frac{1}{\sqrt{2}\sigma} \exp \left(-\frac{\sqrt{2}|w[0]|}{\sigma} \right), \quad (3.62)$$

y comparar el resultado con el caso gaussiano.

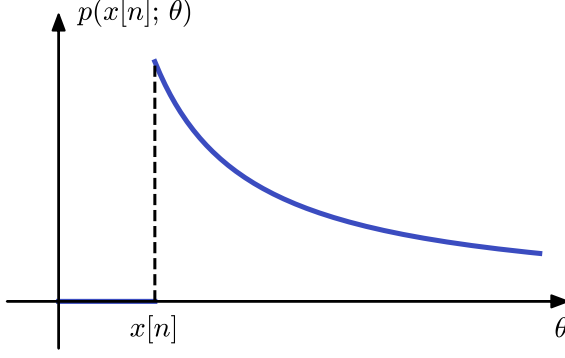


Figura 3.1: Función de verosimilitud cuando $x[n] \sim \mathcal{U}[0, \theta]$.

Solución La CRLB está dada por las ecuaciones 3.1 y 3.2, por lo tanto hay que calcular la información de Fisher. En este caso,

$$x[0] = A + w[0],$$

y como se obtuvo en el problema de la sección 1.1.3, la PDF de $x[0]$ en términos de la PDF de $w[0]$ es

$$p_{x[0]}(x[0]; A) = p_{w[0]}(x[0] - A) = p(x[0] - A).$$

Por lo tanto, hay que calcular

$$I(A) = -E \left[\frac{\partial^2 \ln p(x[0] - A)}{\partial A^2} \right].$$

Usando la regla de la cadena, se obtiene que la derivada primera es

$$\frac{\partial \ln p(x[0] - A)}{\partial A} = \frac{1}{p(x[0] - A)} \frac{\partial p(x[0] - A)}{\partial A} (-1), \quad (3.63)$$

y derivando nuevamente, se ve que

$$\begin{aligned} \frac{\partial^2 \ln p(x[0] - A)}{\partial A^2} &= - \left\{ \frac{-1}{p^2(x[0] - A)} \left[\frac{\partial p(x[0] - A)}{\partial A} \right]^2 (-1) + \frac{1}{p(x[0] - A)} \frac{\partial^2 p(x[0] - A)}{\partial A^2} (-1) \right\} \\ &= - \frac{1}{p^2(x[0] - A)} \left[\frac{\partial p(x[0] - A)}{\partial A} \right]^2 + \frac{1}{p(x[0] - A)} \frac{\partial^2 p(x[0] - A)}{\partial A^2}. \end{aligned} \quad (3.64)$$

Para obtener la información de Fisher, hay que tomar la esperanza. Se demostrará que la esperanza del segundo sumando en la ecuación 3.64 es nula,

$$\begin{aligned} E \left[\frac{1}{p(x[0] - A)} \frac{\partial^2 p(x[0] - A)}{\partial A^2} \right] &= \int \frac{1}{p(x[0] - A)} \frac{\partial^2 p(x[0] - A)}{\partial A^2} p(x[0] - A) dx[0] \\ &= \int \frac{\partial^2 p(x[0] - A)}{\partial A^2} dx[0] \\ &\stackrel{(a)}{=} \frac{\partial}{\partial A} \int \frac{\partial p(x[0] - A)}{\partial A} dx[0] \\ &\stackrel{(b)}{=} - \frac{\partial}{\partial A} \int \frac{\partial \ln p(x[0] - A)}{\partial A} p(x[0] - A) dx[0] \\ &= - \frac{\partial}{\partial A} E \left[\frac{\partial \ln p(x[0] - A)}{\partial A} \right] \\ &\stackrel{(c)}{=} - \frac{\partial 0}{\partial A} \\ &= 0, \end{aligned}$$

donde en (a) se intercambi6 el orden de una derivada y la integraci6n, en (b) se sustituy6 el integrando usando la ecuaci6n 3.63 y en (c) se observ6 que el t6rmino de la esperanza es la condici6n de regularidad

y por lo tanto, es nula. Por lo tanto, al tomar la esperanza en la ecuación 3.64, solo queda el primer sumando, y es

$$\begin{aligned}
E \left[\frac{\partial^2 \ln p(x[0] - A)}{\partial A^2} \right] &= -E \left\{ \frac{1}{p^2(x[0] - A)} \left[\frac{\partial p(x[0] - A)}{\partial A} \right]^2 \right\} \\
&= - \int \frac{1}{p^2(x[0] - A)} \left[\frac{\partial p(x[0] - A)}{\partial A} \right]^2 p(x[0] - A) dx[0] \\
&= - \int \frac{1}{p(x[0] - A)} \left[\frac{\partial p(x[0] - A)}{\partial A} \right]^2 dx[0] \\
&\stackrel{(a)}{=} - \int \frac{1}{p(x[0] - A)} \left[- \frac{\partial p(x[0] - A)}{\partial (x[0] - A)} \right]^2 dx[0] \\
&\stackrel{(b)}{=} - \int \frac{1}{p(u)} \left[\frac{\partial p(u)}{\partial u} \right]^2 du
\end{aligned} \tag{3.65}$$

donde en (a) se usó la regla de la cadena de derivación para obtener que

$$\frac{\partial p(x[0] - A)}{\partial A} = \frac{\partial p(x[0] - A)}{\partial (x[0] - A)} \frac{\partial (x[0] - A)}{\partial A} = - \frac{\partial p(x[0] - A)}{\partial (x[0] - A)}$$

y en (b) se realizó el cambio de variable $u = x[0] - A$ y por lo tanto, $du = dx[0]$ y además se eliminó el signo de menos dentro del cuadrado. De esta forma, se llegó a que

$$I(A) = -E \left[\frac{\partial^2 \ln p(x[0] - A)}{\partial A^2} \right] = \int \frac{1}{p(u)} \left[\frac{\partial p(u)}{\partial u} \right]^2 du,$$

y por lo tanto,

$$\text{var}(\hat{A}) \geq \frac{1}{I(A)} = \left[\int_{-\infty}^{\infty} \frac{\left(\frac{dp(u)}{du} \right)^2}{p(u)} du \right]^{-1}.$$

que es lo que se quería demostrar.

El mismo resultado puede obtenerse de forma un poco mas directa si se parte de la expresión alternativa de la información de Fisher dada por la igualdad de la ecuación 3.6,

$$I(A) = E \left[\left(\frac{\partial \ln p(x[0] - A)}{\partial A} \right)^2 \right].$$

Observando que la derivada primera ya se calculó en la ecuación 3.63, se tiene que

$$I(A) = E \left[\left(\frac{\partial \ln p(x[0] - A)}{\partial A} \right)^2 \right] = E \left\{ \frac{1}{p^2(x[0] - A)} \left[\frac{\partial p(x[0] - A)}{\partial A} \right]^2 \right\},$$

y notando que la expresión dentro de la esperanza es igual al primer sumando de la ecuación 3.64, la cual se calculó en la ecuación 3.65, se llega al mismo resultado.

En el caso en que la PDF de $w[0]$ es Laplaciana, se tiene que

$$p(u) = \frac{1}{\sqrt{2}\sigma} \exp \left(- \frac{\sqrt{2}u}{\sigma} \right), \quad u \geq 0, \text{ par},$$

y la derivada es

$$\frac{dp(u)}{du} = \frac{1}{\sqrt{2}\sigma} \left(- \frac{\sqrt{2}}{\sigma} \right) \exp \left(- \frac{\sqrt{2}u}{\sigma} \right) = - \frac{1}{\sigma^2} \exp \left(- \frac{\sqrt{2}u}{\sigma} \right), \quad u \geq 0,$$

la cual es impar, ya que la derivada de una función par es impar. Por lo tanto, usando el resultado anterior, se tiene que

$$\begin{aligned}
I(A) &= \int_{-\infty}^{\infty} \frac{1}{p(u)} \left[\frac{dp(u)}{du} \right]^2 du \\
&\stackrel{(a)}{=} 2 \int_0^{\infty} \frac{\sqrt{2}\sigma}{\exp\left(-\frac{\sqrt{2}u}{\sigma}\right)} \frac{1}{\sigma^4} \left[\exp\left(-\frac{\sqrt{2}u}{\sigma}\right) \right]^2 du \\
&= \frac{2\sqrt{2}}{\sigma^3} \int_0^{\infty} \exp\left(-\frac{\sqrt{2}u}{\sigma}\right) du \\
&= \frac{2\sqrt{2}}{\sigma^3} \left[-\frac{\sigma}{\sqrt{2}} \exp\left(-\frac{\sqrt{2}u}{\sigma}\right) \right]_0^{\infty} \\
&= -\frac{2}{\sigma^2} (0 - 1) \\
&= \frac{2}{\sigma^2},
\end{aligned}$$

donde en (a) se consideró que el integrando es una función par, ya que $1/p(u)$ es par por ser $p(u)$ par y el cuadrado de la derivada de $p(u)$ también es par por ser la derivada impar. De esta forma, se obtiene que

$$\text{var}(\hat{A}) \geq \frac{1}{I(A)} = \frac{\sigma^2}{2}.$$

Puede demostrarse que el parámetro σ en la PDF de Laplace (ecuación 3.62) es la desviación estándar, es decir, $\text{var}(w[0]) = \sigma^2$. En el ejemplo 3.2 de [1] se demuestra que si $w[0] \sim \mathcal{N}(0, \sigma^2)$, la CRLB de \hat{A} es

$$\text{var}(\hat{A}) \geq \sigma^2.$$

Se concluye que la CRLB es la mitad en el caso en que el ruido es de Laplace respecto al caso en que el ruido es gaussiano. Puede demostrarse que el ruido gaussiano es el que produce la mayor CRLB [9].

3.6.3. Problema 3

Se observan los datos $x[n] = Ar^n + w[n]$ para $n = 0, \dots, N-1$, donde $w[n]$ es WGN con varianza σ^2 y $r > 0$ es conocido. Encontrar la CRLB para A . Mostrar que existe un estimador eficiente y encontrar su varianza. Indicar que ocurre cuando $N \rightarrow \infty$ discutiendo según el valor de r .

Solución La CRLB del parámetro A puede calcularse directamente empleando la ecuación 3.23, pero se realizará el razonamiento completo ya que conviene tener calculada la derivada primera de la función de verosimilitud logarítmica para el estudio de la eficiencia.

La PDF de los datos es

$$p(\mathbf{x}; A) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - Ar^n)^2 \right\}$$

por lo que la función de verosimilitud logarítmica queda

$$\ln p(\mathbf{x}; A) = -\frac{N}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - Ar^n)^2.$$

Diferenciando una vez se obtiene que

$$\begin{aligned}
\frac{\partial \ln p(\mathbf{x}; A)}{\partial A} &= \frac{1}{\sigma^2} \sum_{n=0}^{N-1} (x[n] - Ar^n) r^n \\
&= \frac{1}{\sigma^2} \sum_{n=0}^{N-1} x[n] r^n - \frac{A}{\sigma^2} \sum_{n=0}^{N-1} r^{2n},
\end{aligned} \tag{3.66}$$

y diferenciando nuevamente se llega a que

$$\frac{\partial^2 \ln p(\mathbf{x}; A)}{\partial A^2} = -\frac{1}{\sigma^2} \sum_{n=0}^{N-1} r^{2n}.$$

Se concluye que la información de Fisher de A es

$$I(A) = -E \left[\frac{\partial^2 \ln p(\mathbf{x}; A)}{\partial A^2} \right] = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} r^{2n},$$

por lo que la CRLB es

$$\text{var}(\hat{A}) \geq \frac{\sigma^2}{\sum_{n=0}^{N-1} r^{2n}}.$$

Para estudiar si existe un estimador eficiente, hay que ver si es posible factorizar la ecuación 3.66 como la ecuación 3.3. Efectivamente, se observa que la ecuación 3.66 puede expresarse como

$$\frac{\partial \ln p(\mathbf{x}; A)}{\partial A} = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} r^{2n} \left(\frac{\sum_{n=0}^{N-1} x[n] r^n}{\sum_{n=0}^{N-1} r^{2n}} - A \right) = I(A)(g(\mathbf{x}) - A),$$

concluyendo que el estimador eficiente de A es

$$\hat{A} = \frac{\sum_{n=0}^{N-1} x[n] r^n}{\sum_{n=0}^{N-1} r^{2n}}$$

cuya varianza es $1/I(A)$.

Se estudiará a continuación el límite de la varianza cuando $N \rightarrow \infty$. Considerando que para $r > 0$

$$\sum_{n=0}^{N-1} r^{2n} = \begin{cases} \frac{1 - r^{2N}}{1 - r^2}, & \text{si } r \neq 1 \\ N, & \text{si } r = 1, \end{cases}$$

se tiene que

$$\text{var}(\hat{A}) = \frac{\sigma^2}{\sum_{n=0}^{N-1} r^{2n}} = \begin{cases} \frac{\sigma^2(1 - r^2)}{1 - r^{2N}}, & \text{si } r \neq 1 \\ \frac{\sigma^2}{N}, & \text{si } r = 1. \end{cases}$$

Por lo tanto, si $r \neq 1$, el límite es

$$\lim_{N \rightarrow \infty} \text{var}(\hat{A}) = \lim_{N \rightarrow \infty} \frac{\sigma^2(1 - r^2)}{1 - r^{2N}} = \begin{cases} \sigma^2(1 - r^2), & \text{si } 0 < r < 1 \\ 0, & \text{si } r > 1. \end{cases}$$

y como en el caso con $r = 1$ el límite es cero, se concluye que

$$\lim_{N \rightarrow \infty} \text{var}(\hat{A}) = \begin{cases} \sigma^2(1 - r^2), & \text{si } 0 < r < 1 \\ 0, & \text{si } r \geq 1. \end{cases}$$

3.6.4. Problema 4

Se observa $x[n] = r^n + w[n]$ para $n = 0, \dots, N-1$, donde $w[n]$ es WGN con varianza σ^2 y r es el parámetro desconocido a estimar. Encontrar la CRLB para A . Si existe un estimador eficiente, encontrar su varianza.

Solución La PDF de los datos es

$$p(\mathbf{x}; r) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - r^n)^2 \right\}$$

y la función de verosimilitud logarítmica queda

$$\ln p(\mathbf{x}; r) = -\frac{N}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - r^n)^2.$$

Diferenciando una vez se obtiene que

$$\begin{aligned} \frac{\partial \ln p(\mathbf{x}; r)}{\partial r} &= \frac{1}{\sigma^2} \sum_{n=0}^{N-1} (x[n] - r^n) n r^{n-1} \\ &= \frac{1}{\sigma^2} \sum_{n=0}^{N-1} n (x[n] r^{n-1} - r^{2n-1}). \end{aligned}$$

Puede verse que la derivada no puede factorizarse como

$$\frac{\partial \ln p(\mathbf{x}; r)}{\partial r} = I(r) (g(\mathbf{x}) - r)$$

y por lo tanto, no existe un estimador eficiente. Diferenciando nuevamente se llega a que

$$\frac{\partial^2 \ln p(\mathbf{x}; r)}{\partial r^2} = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} n [x[n](n-1)r^{n-2} - (2n-1)r^{2n-2}]$$

y tomando esperanza se obtiene que

$$\begin{aligned} E \left[\frac{\partial^2 \ln p(\mathbf{x}; r)}{\partial r^2} \right] &= \frac{1}{\sigma^2} \sum_{n=0}^{N-1} n [E(x[n])(n-1)r^{n-2} - (2n-1)r^{2n-2}] \\ &\stackrel{(a)}{=} \frac{1}{\sigma^2} \sum_{n=0}^{N-1} n [(n-1)r^{2n-2} - (2n-1)r^{2n-2}] \\ &= -\frac{1}{\sigma^2} \sum_{n=0}^{N-1} n^2 r^{2n-2}, \end{aligned}$$

donde en (a) se empleó que $E(x[n]) = r^n$. Se concluye que la CRLB es

$$\text{var}(\hat{r}) \geq \frac{\sigma^2}{\sum_{n=0}^{N-1} n^2 r^{2n-2}}.$$

Nuevamente el resultado se podría haber obtenido directamente usando la ecuación 3.23.

3.6.5. Problema 5

Se observa $x[n] = A + w[n]$ para $n = 0, \dots, N-1$, donde $\mathbf{w} = [w[0] w[1] \dots w[N-1]]^T \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$. Encontrar la CRLB para A . Si existe un estimador eficiente, encontrar su varianza.

Solución La PDF de los datos es

$$p(\mathbf{x}, A) = \frac{1}{(2\pi)^{\frac{N}{2}} \det^{\frac{1}{2}}(\mathbf{C})} \exp \left\{ -\frac{1}{2}(\mathbf{x} - A\mathbf{1})^T \mathbf{C}^{-1}(\mathbf{x} - A\mathbf{1}) \right\},$$

donde $\mathbf{1} = [1 \ 1 \ \dots \ 1]^T$ de tamaño $N \times 1$. Tomando el logaritmo, se tiene que

$$\ln p(\mathbf{x}, A) = -\frac{N}{2} \ln(2\pi) - \frac{1}{2} \ln \det(\mathbf{C}) - \frac{1}{2}(\mathbf{x} - A\mathbf{1})^T \mathbf{C}^{-1}(\mathbf{x} - A\mathbf{1}).$$

y diferenciando, se observa que

$$\begin{aligned} \frac{\partial \ln p(\mathbf{x}, A)}{\partial A} &= -\frac{1}{2} \frac{\partial}{\partial A} [(\mathbf{x} - A\mathbf{1})^T \mathbf{C}^{-1}(\mathbf{x} - A\mathbf{1})] \\ &= -\frac{1}{2} \frac{\partial}{\partial A} [\mathbf{x}^T \mathbf{C}^{-1} \mathbf{x} - A \mathbf{x}^T \mathbf{C}^{-1} \mathbf{1} - A \mathbf{1}^T \mathbf{C}^{-1} \mathbf{x} + A^2 \mathbf{1}^T \mathbf{C}^{-1} \mathbf{1}] \\ &= -\frac{1}{2} \frac{\partial}{\partial A} [\mathbf{x}^T \mathbf{C}^{-1} \mathbf{x} - 2A \mathbf{1}^T \mathbf{C}^{-1} \mathbf{x} + A^2 \mathbf{1}^T \mathbf{C}^{-1} \mathbf{1}] \\ &= -\frac{1}{2} [-2\mathbf{1}^T \mathbf{C}^{-1} \mathbf{x} + 2A \mathbf{1}^T \mathbf{C}^{-1} \mathbf{1}] \\ &= \mathbf{1}^T \mathbf{C}^{-1} \mathbf{x} - A \mathbf{1}^T \mathbf{C}^{-1} \mathbf{1}, \end{aligned}$$

que se puede expresar como

$$\frac{\partial \ln p(\mathbf{x}, A)}{\partial A} = \mathbf{1}^T \mathbf{C}^{-1} \mathbf{1} \left[\frac{\mathbf{1}^T \mathbf{C}^{-1} \mathbf{x}}{\mathbf{1}^T \mathbf{C}^{-1} \mathbf{1}} - A \right].$$

Comparando con la ecuación 3.3 se tiene que

$$\hat{A} = \frac{\mathbf{1}^T \mathbf{C}^{-1} \mathbf{x}}{\mathbf{1}^T \mathbf{C}^{-1} \mathbf{1}} \quad I(A) = \mathbf{1}^T \mathbf{C}^{-1} \mathbf{1}.$$

Se concluye que el estimador \hat{A} es eficiente y su varianza es $1/I(A)$.

3.6.6. Problema 6

Se consideran dos observaciones $x[0]$ y $x[1]$ independientes con PDF

$$x[0] \sim \mathcal{N}(\theta, 1) \quad x[1] \sim \begin{cases} \mathcal{N}(\theta, 1), & \text{si } \theta \geq 0 \\ \mathcal{N}(\theta, 2), & \text{si } \theta < 0. \end{cases}$$

Calcular la CRLB del parámetro θ .

Solución La PDF de cada observación es

$$p(x[0]; \theta) = \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2}(x[0] - \theta)^2 \right\} \quad p(x[1]; \theta) = \begin{cases} \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2}(x[1] - \theta)^2 \right\}, & \text{si } \theta \geq 0 \\ \frac{1}{\sqrt{4\pi}} \exp \left\{ -\frac{1}{4}(x[1] - \theta)^2 \right\}, & \text{si } \theta < 0. \end{cases}$$

Por lo tanto, la PDF de \mathbf{x} es

$$p(\mathbf{x}; \theta) = p(x[0]; \theta)p(x[1]; \theta) = \begin{cases} \frac{1}{2\pi} \exp \left\{ -\frac{1}{2}(x[0] - \theta)^2 - \frac{1}{2}(x[1] - \theta)^2 \right\}, & \text{si } \theta \geq 0 \\ \frac{1}{2\sqrt{2}\pi} \exp \left\{ -\frac{1}{2}(x[0] - \theta)^2 - \frac{1}{4}(x[1] - \theta)^2 \right\}, & \text{si } \theta < 0, \end{cases}$$

y al tomar el logaritmo se obtiene que

$$\ln p(\mathbf{x}; \theta) = \begin{cases} -\ln(2\pi) - \frac{1}{2}(x[0] - \theta)^2 - \frac{1}{2}(x[1] - \theta)^2, & \text{si } \theta \geq 0 \\ -\ln(2\sqrt{2}\pi) - \frac{1}{2}(x[0] - \theta)^2 - \frac{1}{4}(x[1] - \theta)^2, & \text{si } \theta < 0. \end{cases}$$

Diferenciando respecto al parámetro, se obtiene que

$$\frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \theta} = \begin{cases} (x[0] - \theta) + (x[1] - \theta), & \text{si } \theta \geq 0 \\ (x[0] - \theta) + \frac{1}{2}(x[1] - \theta), & \text{si } \theta < 0 \end{cases} = \begin{cases} x[0] + x[1] - 2\theta, & \text{si } \theta \geq 0 \\ x[0] + \frac{1}{2}x[1] - \frac{3}{2}\theta, & \text{si } \theta < 0, \end{cases}$$

y diferenciando nuevamente se llega a que

$$\frac{\partial^2 \ln p(\mathbf{x}; \theta)}{\partial \theta^2} = \begin{cases} -2, & \text{si } \theta \geq 0 \\ -\frac{3}{2}, & \text{si } \theta < 0. \end{cases}$$

Por lo tanto, la información de Fisher es

$$I(\theta) = -E \left[\frac{\partial^2 \ln p(\mathbf{x}; \theta)}{\partial \theta^2} \right] = \begin{cases} 2, & \text{si } \theta \geq 0 \\ \frac{3}{2}, & \text{si } \theta < 0, \end{cases}$$

concluyendo que la CBRL es

$$\text{var}(\hat{\theta}) \geq \frac{1}{I(\theta)} = \begin{cases} \frac{1}{2}, & \text{si } \theta \geq 0 \\ \frac{2}{3}, & \text{si } \theta < 0. \end{cases}$$

3.6.7. Problema 7

Probar que

$$\frac{1}{N} \sum_{n=0}^{N-1} \cos(4\pi f_0 n + 2\phi) \approx 0,$$

resultado empleado en el ejemplo 3.4. ¿Qué condiciones se requieren sobre f_0 para que el resultado se cumpla?

Solución Defínase $\alpha = 4\pi f_0$ y $\beta = 2\phi$ y se observe que

$$\begin{aligned} \sum_{n=0}^{N-1} \cos(\alpha n + \beta) &= \text{Re} \left[\sum_{n=0}^{N-1} e^{j(\alpha n + \beta)} \right] \\ &= \text{Re} \left[e^{j\beta} \sum_{n=0}^{N-1} e^{j\alpha n} \right] \\ &= \text{Re} \left[e^{j\beta} \frac{1 - e^{j\alpha N}}{1 - e^{j\alpha}} \right] \\ &= \text{Re} \left[e^{j\beta} \frac{e^{j\alpha N/2}}{e^{j\alpha/2}} \frac{e^{-j\alpha N/2} - e^{j\alpha N/2}}{e^{-j\alpha/2} - e^{j\alpha/2}} \right] \\ &= \text{Re} \left[e^{j\beta} e^{j\alpha(N-1)/2} \frac{\sin(\alpha N/2)}{\sin(\alpha/2)} \right] \\ &= \text{Re} \left\{ [\cos(\alpha(N-1)/2 + \beta) + j \sin(\alpha(N-1)/2 + \beta)] \frac{\sin(\alpha N/2)}{\sin(\alpha/2)} \right\} \\ &= \cos[\alpha(N-1)/2 + \beta] \frac{\sin(\alpha N/2)}{\sin(\alpha/2)}, \end{aligned}$$

y por lo tanto,

$$\frac{1}{N} \sum_{n=0}^{N-1} \cos(4\pi f_0 n + 2\phi) = \frac{\sin(2\pi f_0 N)}{N \sin(2\pi f_0)} \cos[2\pi f_0(N-1) + 2\phi].$$

En la figura 3.2 se grafica $|\sin(2\pi f_0 N)/(N \sin(2\pi f_0))|$ para $N = 20$. Esta función es periódica de período $1/2$. Como se observa en la figura, se cumple que

$$\left| \frac{1}{N} \sum_{n=0}^{N-1} \cos(4\pi f_0 n + 2\phi) \right| \leq \left| \frac{\sin(2\pi f_0 N)}{N \sin(2\pi f_0)} \right| \approx 0$$

si f_0 no está cerca a 0 o a $1/2$, y la aproximación es mejor para f_0 en torno a $1/4$. Además, la aproximación mejora al incrementar N .

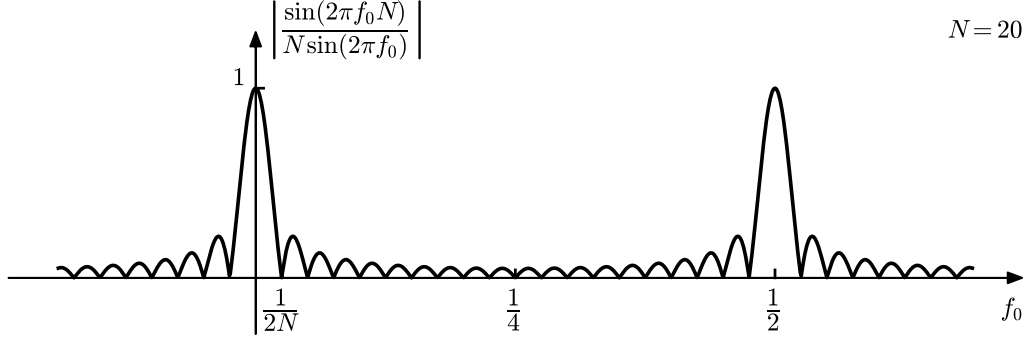


Figura 3.2: Gráfica de $|\sin(2\pi f_0 N)/(N \sin(2\pi f_0))|$ para $N = 20$. La función es periódica de período $1/2$ y los ceros se encuentran en $f_0 = k/(2N)$. Como se observa, la función toma valores pequeños si f_0 no está cerca de 0 o $1/2$. Además, la función decrece al incrementar N .

Notar además que

$$\begin{aligned} \sum_{n=0}^{N-1} \sin(\alpha n + \beta) &= \text{Im} \left[\sum_{n=0}^{N-1} e^{j(\alpha n + \beta)} \right] \\ &= \sin[\alpha(N-1)/2 + \beta] \frac{\sin(\alpha N/2)}{\sin(\alpha/2)}, \end{aligned}$$

3.6.8. Problema 8

Para el problema de estimación del nivel de DC en WGN (ver sección 1.1.4), calcular la CRLB empleando la expresión alternativa

$$\text{var}(\hat{\theta}) \geq \frac{1}{E \left[\left(\frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \theta} \right)^2 \right]},$$

que proviene de las ecuaciones 3.1, 3.2 y 3.6.

Solución En el problema de la estimación del nivel de DC en WGN, se tienen las observaciones

$$x[n] = A + w[n] \quad n = 0, 1, \dots, N-1,$$

donde $w[n]$ es WGN con varianza σ^2 y se quiere determinar la CRLB para A . Como se explica en el ejemplo 3.3 de [1], la PDF de los datos es

$$\begin{aligned} p(\mathbf{x}; A) &= \prod_{n=0}^{N-1} \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{1}{2\sigma^2} (x[n] - A)^2 \right] \\ &= \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - A)^2 \right], \end{aligned} \quad (3.67)$$

y al tomar el logaritmo, se obtiene que

$$\ln p(\mathbf{x}; A) = -\frac{N}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - A)^2.$$

Diferenciando respecto al parámetro, se llega a que

$$\frac{\partial \ln p(\mathbf{x}; A)}{\partial A} = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} (x[n] - A). \quad (3.68)$$

La información de Fisher es

$$\begin{aligned} I(A) &= E \left[\left(\frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \theta} \right)^2 \right] = E \left\{ \left[\frac{1}{\sigma^2} \sum_{n=0}^{N-1} (x[n] - A) \right]^2 \right\} \\ &= E \left\{ \frac{1}{\sigma^4} \left[\sum_{m=0}^{N-1} (x[m] - A) \right] \left[\sum_{n=0}^{N-1} (x[n] - A) \right] \right\} \\ &= E \left[\frac{1}{\sigma^4} \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} (x[m] - A)(x[n] - A) \right] \\ &= \frac{1}{\sigma^4} \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} E[(x[m] - A)(x[n] - A)] \\ &\stackrel{(a)}{=} \frac{1}{\sigma^4} \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} \sigma^2 \delta[m - n] \\ &= \frac{1}{\sigma^4} \sum_{n=0}^{N-1} \sigma^2 \\ &= \frac{N}{\sigma^2}, \end{aligned}$$

donde en (a) se empleó que

$$\begin{aligned} E[(x[m] - A)(x[n] - A)] &= \begin{cases} E[(x[n] - A)^2] = \sigma^2, & \text{si } m = n \\ E[(x[m] - A)] E[(x[n] - A)] = 0, & \text{si } m \neq n \end{cases} \\ &= \sigma^2 \delta[m - n]. \end{aligned}$$

Se concluye que

$$\text{var}(\hat{\theta}) \geq \frac{\sigma^2}{N}.$$

Otra forma de ver lo mismo empleando resultados anteriores, es notando que la ecuación 3.68 puede escribirse como

$$\begin{aligned} \frac{\partial \ln p(\mathbf{x}; A)}{\partial A} &= \frac{1}{\sigma^2} \sum_{n=0}^{N-1} (x[n] - A) \\ &= \frac{1}{\sigma^2} \left(\sum_{n=0}^{N-1} x[n] - NA \right) \\ &= \frac{1}{\sigma^2} (N\bar{\mathbf{x}} - NA) \\ &= \frac{N}{\sigma^2} (\bar{\mathbf{x}} - A), \end{aligned}$$

y al elevar al cuadrado y tomar esperanza, se tiene que

$$\begin{aligned}
 I(A) &= E \left[\left(\frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \theta} \right)^2 \right] = \frac{N^2}{\sigma^4} E \left[(\bar{\mathbf{x}} - A)^2 \right] \\
 &\stackrel{(a)}{=} \frac{N^2}{\sigma^4} \text{var}(\bar{\mathbf{x}}) \\
 &\stackrel{(b)}{=} \frac{N^2}{\sigma^4} \frac{\sigma^2}{N} \\
 &= \frac{N}{\sigma^2},
 \end{aligned}$$

donde en (a) y en (b) se empleó que $\bar{\mathbf{x}} \sim \mathcal{N}(A, \sigma^2/N)$, como se demostró en la sección 2.1.3.

3.6.9. Problema 9

Se observan dos muestras de un nivel de DC en ruido gaussiano *correlacionado*,

$$\begin{aligned}
 x[0] &= A + w[0] \\
 x[1] &= A + w[1],
 \end{aligned}$$

donde $\mathbf{w} = [w[0] \ w[1]]^T$ tiene media nula y matriz de covarianza

$$\mathbf{C} = \sigma^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}.$$

El parámetro ρ es el coeficiente de correlación entre $w[0]$ y $w[1]$. Calcular la CRLB de A y comparar con el caso en que $w[n]$ es WGN, es decir, cuando $\rho = 0$. Además explicar que ocurre cuando $\rho \rightarrow \pm 1$. Finalmente, comentar sobre la propiedad de aditividad de la información de Fisher para observaciones no independientes.

Solución La CRLB para el caso general de ruido gaussiano ya fue resuelto en el problema de la sección 3.6.5, por lo que podrían emplearse los resultados obtenidos allí. También podría obtenerse la información de Fisher de forma inmediata a partir de la ecuación 3.37, pero se harán las cuentas nuevamente para este caso particular.

La PDF de los datos es

$$p(\mathbf{x}, A) = \frac{1}{2\pi\sqrt{\det(\mathbf{C})}} \exp \left\{ -\frac{1}{2}(\mathbf{x} - A\mathbf{1})^T \mathbf{C}^{-1}(\mathbf{x} - A\mathbf{1}) \right\},$$

donde $\mathbf{1} = [1 \ 1]^T$. El determinante de la matriz de correlación es

$$\det(\mathbf{C}) = \sigma^4(1 - \rho^2)$$

y la inversa de la matriz de correlación es

$$\mathbf{C}^{-1} = \frac{1}{\sigma^2(1 - \rho^2)} \begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix}.$$

Además,

$$\mathbf{C}^{-1}(\mathbf{x} - A\mathbf{1}) = \frac{1}{\sigma^2(1 - \rho^2)} \begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix} \begin{bmatrix} x[0] - A \\ x[1] - A \end{bmatrix} = \frac{1}{\sigma^2(1 - \rho^2)} \begin{bmatrix} (x[0] - A) - \rho(x[1] - A) \\ -\rho(x[0] - A) + (x[1] - A) \end{bmatrix}$$

y

$$\begin{aligned}
 (\mathbf{x} - A\mathbf{1})^T \mathbf{C}^{-1}(\mathbf{x} - A\mathbf{1}) &= \frac{1}{\sigma^2(1 - \rho^2)} \begin{bmatrix} x[0] - A & x[1] - A \end{bmatrix} \begin{bmatrix} (x[0] - A) - \rho(x[1] - A) \\ -\rho(x[0] - A) + (x[1] - A) \end{bmatrix} \\
 &= \frac{(x[0] - A)^2 - 2\rho(x[0] - A)(x[1] - A) + (x[1] - A)^2}{\sigma^2(1 - \rho^2)}.
 \end{aligned}$$

Por lo tanto, la PDF de los datos en este caso queda

$$p(\mathbf{x}, A) = \frac{1}{2\pi\sigma^2\sqrt{(1-\rho^2)}} \exp \left\{ -\frac{(x[0]-A)^2 - 2\rho(x[0]-A)(x[1]-A) + (x[1]-A)^2}{2\sigma^2(1-\rho^2)} \right\},$$

y tomando el logaritmo, se obtiene que

$$\ln p(\mathbf{x}, A) = -\ln 2\pi\sigma^2\sqrt{(1-\rho^2)} - \frac{(x[0]-A)^2 - 2\rho(x[0]-A)(x[1]-A) + (x[1]-A)^2}{2\sigma^2(1-\rho^2)}.$$

Al diferenciar respecto al parámetro se observa que

$$\begin{aligned} \frac{\partial \ln p(\mathbf{x}, A)}{\partial A} &= -\frac{-2(x[0]-A) - 2\rho[-(x[0]-A) - (x[1]-A)] - 2(x[1]-A)}{2\sigma^2(1-\rho^2)} \\ &= \frac{x[0] + x[1] - 2A - \rho(x[0] + x[1] - 2A)}{\sigma^2(1-\rho^2)} \\ &= \frac{(x[0] + x[1] - 2A)(1-\rho)}{\sigma^2(1-\rho^2)}, \end{aligned}$$

que puede expresarse como

$$\frac{\partial \ln p(\mathbf{x}, A)}{\partial A} = \frac{2}{\sigma^2(1+\rho)} \left(\frac{x[0] + x[1]}{2} - A \right),$$

donde se tuvo en cuenta que $(1-\rho^2) = (1+\rho)(1-\rho)$. Como la derivada de la función de verosimilitud se factorizó como indica la ecuación 3.3, se concluye que el estimador media muestral

$$\hat{A} = \frac{x[0] + x[1]}{2},$$

es eficiente y la información de Fisher es

$$I(A) = \frac{2}{\sigma^2(1+\rho)}. \quad (3.69)$$

Por lo tanto, la CRLB de A es

$$\text{var}(\hat{A}) \geq \frac{\sigma^2}{2}(1+\rho).$$

El coeficiente de correlación entre dos variables aleatorias x e y es una medida de la dependencia entre las variables aleatorias (ver el capítulo 6 de [2]) y se define como

$$\rho_{xy} = \frac{C_{xy}}{\sigma_x \sigma_y}.$$

Se cumple que $-1 \leq \rho_{xy} \leq 1$. En el caso en que $\rho_{xy} = 0$, la covarianza es nula y las variables son no correlacionadas, ya que

$$C_{xy} = E(xy) - E(x)E(y) = 0 \quad \Rightarrow \quad E(xy) = E(x)E(y).$$

En el caso en que $\rho_{xy} = 1$, las variables aleatorias tienen dependencia lineal, $y = ax$, donde a es una constante. De esta forma,

$$\rho_{xy} = \frac{C_{xy}}{\sigma_x \sigma_y} = \frac{E(xy) - E(x)E(y)}{\sigma_x \sigma_y} = \frac{aE(x^2) - aE^2(x)}{a\sigma_x^2} = \frac{a\sigma_x^2}{a\sigma_x^2} = 1,$$

donde se empleó que si $y = ax$ se cumple que $\sigma_y = a\sigma_x$. En el caso en que las variables aleatorias tienen la misma varianza y $\rho = 1$, se cumple que $y = x$, y en el caso en que $\rho = -1$, se cumple que $y = -x$. Volviendo a la CRLB de A , en el caso en que $\rho = 0$ se cumple que

$$\text{var}(\hat{A}) \geq \frac{\sigma^2}{2},$$

que es la CRLB calculada previamente para el caso de WGN con $N = 2$. En el caso en que $\rho \rightarrow 1$

$$\text{var}(\hat{A}) \geq \sigma^2,$$

que es la CRLB al observar una única muestra. Esto se debe a que como $\rho \rightarrow 1$, $w[1] = w[0]$ y por lo tanto, se dispone de una única muestra independiente. En el caso en que $\rho \rightarrow -1$, se cumple que

$$\text{var}(\hat{A}) \geq 0,$$

y de hecho, la media muestral es

$$\hat{A} = \frac{x[0] + x[1]}{2} = \frac{A + w[0] + A - w[0]}{2} = A,$$

es decir, el valor exacto del parámetro para cualquier realización de las muestras de ruido.

La información de Fisher en el caso de una sola muestra es

$$i(A) = \frac{1}{\sigma^2}.$$

La propiedad de aditividad de la información de Fisher se cumple en el caso de muestras no correlacionadas. Efectivamente, de la ecuación 3.69 se observa que en el caso en que $\rho = 0$, se cumple que

$$I(A) = \frac{2}{\sigma^2} = 2i(A).$$

Además, se observa que

$$i(A) \leq I(A) \leq \infty,$$

donde la cota inferior se da en el caso en que $\rho = 1$ (muestras completamente correlacionadas) y el límite superior se da en el caso en que $\rho = -1$. Cuando las muestras están correlacionadas con $0 \leq \rho \leq 1$ se cumple que

$$i(A) \leq I(A) \leq 2i(A),$$

y en general, si las muestras están correlacionadas con $\rho > 0$ se cumple que

$$I(A) \leq Ni(A).$$

3.6.10. Problema 10

Usando la ecuación 3.17, probar que la matriz de información de Fisher es semidefinida positiva para todo θ . En la práctica se asumirá que es positiva definida y por lo tanto invertible, aunque no sea siempre el caso. Considerar el modelo de los datos del problema de la sección 3.6.3 con la modificación de que ahora r también es desconocido. Calcular la matriz de información de Fisher de $\theta = [Ar]^T$. ¿Existen valores de θ tal que $\mathbf{I}(\theta)$ no es definida positiva?

Solución Una matriz \mathbf{M} $n \times n$ es semidefinida positiva si se cumple que $\mathbf{a}^T \mathbf{M} \mathbf{a} \geq 0$ para todo $\mathbf{a} \neq 0$ $n \times 1$. Como indica la ecuación 3.17,

$$[\mathbf{I}(\theta)]_{ij} = E \left[\frac{\partial \ln p(\mathbf{x}, \theta)}{\partial \theta_i} \frac{\partial \ln p(\mathbf{x}, \theta)}{\partial \theta_j} \right],$$

que en notación matricial se puede expresar como

$$\mathbf{I}(\theta) = E \left[\frac{\partial \ln p(\mathbf{x}, \theta)}{\partial \theta} \frac{\partial \ln p(\mathbf{x}, \theta)}{\partial \theta}^T \right].$$

Por lo tanto,

$$\begin{aligned}
\mathbf{a}^T \mathbf{I}(\boldsymbol{\theta}) \mathbf{a} &= \mathbf{a}^T E \left[\frac{\partial \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T \right] \mathbf{a} \\
&= E \left[\left(\mathbf{a}^T \frac{\partial \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) \left(\frac{\partial \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T \mathbf{a} \right) \right] \\
&\stackrel{(a)}{=} E \left[\left(\mathbf{a}^T \frac{\partial \ln p(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^2 \right] \\
&\geq 0,
\end{aligned}$$

donde en (a) se consideró que los valores entre paréntesis curvos son escalares y por lo tanto su valor es igual a su valor transpuesto. Se concluye que $\mathbf{I}(\boldsymbol{\theta})$ es semidefinida positiva para todo $\boldsymbol{\theta}$

Se calculará la matriz de información de Fisher del problema de la sección 3.6.3 considerando ahora que r también es desconocido. El parámetro ahora es el vector $\boldsymbol{\theta} = [A r]^T$. La función de verosimilitud logarítmica de los datos es (ver sección 3.6.3)

$$\ln p(\mathbf{x}; A) = -\frac{N}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - Ar^n)^2.$$

Diferenciando una vez respecto a A se obtiene que (ecuación 3.66)

$$\frac{\partial \ln p(\mathbf{x}; A)}{\partial A} = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} (x[n] - Ar^n)r^n$$

y diferenciando nuevamente se llega a que

$$\frac{\partial^2 \ln p(\mathbf{x}; A)}{\partial A^2} = -\frac{1}{\sigma^2} \sum_{n=0}^{N-1} r^{2n}.$$

Tomando esperanza y cambiando el signo se concluye que

$$[\mathbf{I}(\boldsymbol{\theta})]_{11} = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} r^{2n}.$$

Además, diferenciando 3.66 respecto a r se observa que

$$\frac{\partial^2 \ln p(\mathbf{x}; A)}{\partial A \partial r} = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} [-Anr^{n-1}r^n + (x[n] - Ar^n)nr^{n-1}],$$

y al tomar el opuesto de la esperanza se ve que

$$\begin{aligned}
[\mathbf{I}(\boldsymbol{\theta})]_{12} &= [\mathbf{I}(\boldsymbol{\theta})]_{21} = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} [Anr^{n-1}r^n - E(x[n] - Ar^n)nr^{n-1}] \\
&= \frac{A}{\sigma^2} \sum_{n=0}^{N-1} nr^{2n-1}.
\end{aligned}$$

Las derivadas respecto a r son

$$\begin{aligned}
\frac{\partial \ln p(\mathbf{x}; A)}{\partial r} &= \frac{1}{\sigma^2} \sum_{n=0}^{N-1} (x[n] - Ar^n)Anr^{n-1} \\
\frac{\partial^2 \ln p(\mathbf{x}; A)}{\partial r^2} &= \frac{1}{\sigma^2} \sum_{n=0}^{N-1} [-Anr^{n-1}Anr^{n-1} + (x[n] - Ar^n)An(n-1)r^{n-2}]
\end{aligned}$$

y al tomar esperanza y cambiar el signo se obtiene que

$$[\mathbf{I}(\boldsymbol{\theta})]_{22} = \frac{A^2}{\sigma^2} \sum_{n=0}^{N-1} n^2 r^{2n-2}.$$

Se concluye que la matriz de información de Fisher es

$$\mathbf{I}(\boldsymbol{\theta}) = \frac{1}{\sigma^2} \begin{bmatrix} \sum_{n=0}^{N-1} r^{2n} & A \sum_{n=0}^{N-1} n r^{2n-1} \\ A \sum_{n=0}^{N-1} n r^{2n-1} & A^2 \sum_{n=0}^{N-1} n^2 r^{2n-2} \end{bmatrix}.$$

En el caso en que $A = 0$,

$$\mathbf{I}(\boldsymbol{\theta}) = \frac{1}{\sigma^2} \begin{bmatrix} \sum_{n=0}^{N-1} r^{2n} & 0 \\ 0 & 0 \end{bmatrix},$$

es una matriz con determinante nulo, y como el determinante es el producto de los valores propios, tiene algún valor propio nulo, por lo que es semidefinida positiva. Otra forma de verlo es notar que $\mathbf{a}^T \mathbf{I}(\boldsymbol{\theta}) \mathbf{a} = 0$ para cualquier $\mathbf{a} = [0 \ c]^T$, con c una constante arbitraria.

3.6.11. Problema 11

3.6.12

Para una matriz de información Fisher 2×2

$$\mathbf{I}(\boldsymbol{\theta}) = \begin{bmatrix} a & b \\ b & c \end{bmatrix},$$

definida positiva, mostrar que

$$[\mathbf{I}^{-1}(\boldsymbol{\theta})]_{11} = \frac{c}{ac - b^2} \geq \frac{1}{a} = \frac{1}{[\mathbf{I}(\boldsymbol{\theta})]_{11}}.$$

Solución El determinante de la matriz de información de Fisher es $\det[\mathbf{I}(\boldsymbol{\theta})] = ac - b^2 > 0$ por ser una matriz definida positiva. Por lo tanto, el elemento 11 de la matriz inversa es

$$[\mathbf{I}^{-1}(\boldsymbol{\theta})]_{11} = \frac{c}{ac - b^2} = \frac{1}{a - \frac{b^2}{c}} \stackrel{(a)}{\geq} \frac{1}{a} = \frac{1}{[\mathbf{I}(\boldsymbol{\theta})]_{11}},$$

donde en (a) se consideró que $b^2/c \geq 0$, ya que los elementos de la diagonal de una matriz definida positiva son positivos. Se cumple entonces que

$$[\mathbf{I}^{-1}(\boldsymbol{\theta})]_{11} \geq \frac{1}{[\mathbf{I}(\boldsymbol{\theta})]_{11}}.$$

Teniendo en cuenta que $1/[\mathbf{I}(\boldsymbol{\theta})]_{11}$ es la CRLB de θ_1 en el caso de un solo parámetro desconocido y $[\mathbf{I}^{-1}(\boldsymbol{\theta})]_{11}$ es la CRLB de θ_1 en el caso de dos parámetros desconocidos, $\boldsymbol{\theta} = [\theta_1 \ \theta_2]^T$, se concluye que la CRLB se incrementa cuando se incrementa la cantidad de parámetros desconocidos a estimar. La igualdad se da cuando $b = 0$, es decir, cuando la matriz de información de Fisher es diagonal, ya que la inversa de una matriz diagonal es diagonal y los elementos de la diagonal son los inversos de los elementos de la matriz original. En ese caso, parámetros adicionales no afectan la CRLB.

3.6.12. Problema 12

Probar que

$$[\mathbf{I}^{-1}(\boldsymbol{\theta})]_{ii} \geq \frac{1}{[\mathbf{I}(\boldsymbol{\theta})]_{ii}}.$$

Esto es generaliza el resultado del problema de la sección . Además, establece otra cota inferior a la varianza, aunque típicamente no es alcanzable. ¿Bajo que condiciones se alcanza la nueva cota? Sugerencia: aplicar la desigualdad de Cauchy-Schwarz a $\mathbf{e}_i^T \sqrt{\mathbf{I}(\boldsymbol{\theta})} \sqrt{\mathbf{I}^{-1}(\boldsymbol{\theta})} \mathbf{e}_i$, donde \mathbf{e}_i son vectores con todos ceros excepto en la posición i -ésima, que tienen valor 1. La raíz cuadrada de una matriz definida positiva \mathbf{A} se define como la matriz con los mismos vectores propios de \mathbf{A} pero cuyos valores propios son la raíz cuadrada de los valores propios de \mathbf{A} .

Solución En la demostración se empleará la desigualdad de Cauchy-Schwarz⁹, que indica que

$$(\mathbf{a}^T \mathbf{b})^2 \leq (\mathbf{a}^T \mathbf{a})(\mathbf{b}^T \mathbf{b}) \quad (3.70)$$

donde \mathbf{a} y \mathbf{b} son de tamaño $n \times 1$. La igualdad se da si $\mathbf{a} = c\mathbf{b}$, donde c una constante. Se considerará la desigualdad de Cauchy-Schwarz con

$$\mathbf{a} = \sqrt{\mathbf{I}(\boldsymbol{\theta})} \mathbf{e}_i, \quad \mathbf{b} = \sqrt{\mathbf{I}^{-1}(\boldsymbol{\theta})} \mathbf{e}_i.$$

Previo a la demostración se realizarán algunas consideraciones respecto a la raíz cuadrada de una matriz. La raíz cuadrada de una matriz definida positiva \mathbf{A} se define como la matriz con los mismos vectores propios de \mathbf{A} pero cuyos valores propios son la raíz cuadrada de los valores propios de \mathbf{A} .

- Esta definición tiene sentido, ya que una matriz cuadrada definida positiva puede descomponerse como $\mathbf{A} = \mathbf{Q} \boldsymbol{\Lambda} \mathbf{Q}^{-1}$, donde \mathbf{Q} es una matriz cuyas columnas son los vectores propios de \mathbf{A} y $\boldsymbol{\Lambda}$ es una matriz diagonal cuyos elementos en la diagonal son los valores propios de \mathbf{A} . De esta forma

$$\sqrt{\mathbf{A}} \sqrt{\mathbf{A}} = \left(\sqrt{\mathbf{Q} \boldsymbol{\Lambda} \mathbf{Q}^{-1}} \right) \left(\sqrt{\mathbf{Q} \boldsymbol{\Lambda} \mathbf{Q}^{-1}} \right) \stackrel{(a)}{=} \left(\mathbf{Q} \sqrt{\boldsymbol{\Lambda}} \mathbf{Q}^{-1} \right) \left(\mathbf{Q} \sqrt{\boldsymbol{\Lambda}} \mathbf{Q}^{-1} \right) = \mathbf{Q} \sqrt{\boldsymbol{\Lambda}} \sqrt{\boldsymbol{\Lambda}} \mathbf{Q}^{-1} = \mathbf{Q} \boldsymbol{\Lambda} \mathbf{Q}^{-1} = \mathbf{A},$$

donde en (a) se empleó la definición de la raíz cuadrada de una matriz con $\sqrt{\boldsymbol{\Lambda}}$ una matriz diagonal con elementos la raíz cuadrada de los elementos de $\boldsymbol{\Lambda}$.

- Se cumple que $\mathbf{A}^{-1} = \mathbf{Q} \boldsymbol{\Lambda}^{-1} \mathbf{Q}^{-1}$, ya que

$$\mathbf{A}^{-1} = (\mathbf{Q} \boldsymbol{\Lambda} \mathbf{Q}^{-1})^{-1} = (\mathbf{Q}^{-1})^{-1} \boldsymbol{\Lambda}^{-1} \mathbf{Q}^{-1} = \mathbf{Q} \boldsymbol{\Lambda}^{-1} \mathbf{Q}^{-1}.$$

- Se cumple que $\sqrt{\mathbf{A}^{-1}} = (\sqrt{\mathbf{A}})^{-1}$, ya que

$$\sqrt{\mathbf{A}^{-1}} = \sqrt{\mathbf{Q} \boldsymbol{\Lambda}^{-1} \mathbf{Q}^{-1}} = \mathbf{Q} \sqrt{\boldsymbol{\Lambda}^{-1}} \mathbf{Q}^{-1} \stackrel{(a)}{=} \mathbf{Q} \left(\sqrt{\boldsymbol{\Lambda}} \right)^{-1} \mathbf{Q}^{-1} = \left(\sqrt{\mathbf{A}} \right)^{-1},$$

donde en (a) se consideró que la matriz $\boldsymbol{\Lambda}$ es diagonal y las operaciones raíz cuadrada e inversión son conmutativas en los números reales.

- En el caso en que la matriz \mathbf{A} es simétrica definida positiva, se cumple además que:

- la descomposición es $\mathbf{A} = \mathbf{Q} \boldsymbol{\Lambda} \mathbf{Q}^T$, donde la matriz \mathbf{Q} de los vectores propios es ortogonal, $\mathbf{Q}^T = \mathbf{Q}^{-1}$.
- se cumple que $\sqrt{\mathbf{A}}$ es simétrica, ya que

$$\sqrt{\mathbf{A}}^T = (\mathbf{Q} \sqrt{\boldsymbol{\Lambda}} \mathbf{Q}^T)^T = (\mathbf{Q}^T)^T \sqrt{\boldsymbol{\Lambda}}^T \mathbf{Q}^T = \mathbf{Q} \sqrt{\boldsymbol{\Lambda}} \mathbf{Q}^T = \sqrt{\mathbf{A}}.$$

Teniendo en cuenta estas consideraciones, aplicando la desigualdad de Cauchy-Schwarz con los vectores \mathbf{a} y \mathbf{b} definidos previamente, se tiene que

$$\left(\mathbf{e}_i^T \sqrt{\mathbf{I}(\boldsymbol{\theta})}^T \sqrt{\mathbf{I}^{-1}(\boldsymbol{\theta})} \mathbf{e}_i \right)^2 \leq \left(\mathbf{e}_i^T \sqrt{\mathbf{I}(\boldsymbol{\theta})}^T \sqrt{\mathbf{I}(\boldsymbol{\theta})} \mathbf{e}_i \right) \left(\mathbf{e}_i^T \sqrt{\mathbf{I}^{-1}(\boldsymbol{\theta})}^T \sqrt{\mathbf{I}^{-1}(\boldsymbol{\theta})} \mathbf{e}_i \right)$$

⁹Ver https://en.wikipedia.org/wiki/Cauchy-Schwarz_inequality

Desarrollando el término entre paréntesis del lado izquierdo de la desigualdad se ve que

$$\mathbf{e}_i^T \sqrt{\mathbf{I}(\boldsymbol{\theta})}^T \sqrt{\mathbf{I}^{-1}(\boldsymbol{\theta})} \mathbf{e}_i \stackrel{(a)}{=} \mathbf{e}_i^T \sqrt{\mathbf{I}(\boldsymbol{\theta})} \sqrt{\mathbf{I}(\boldsymbol{\theta})}^{-1} \mathbf{e}_i = \mathbf{e}_i^T \mathbf{e}_i = 1, \quad \forall i = 1, \dots, p,$$

donde en (a) se consideró que se puede cambiar el orden de las operaciones de inversión y raíz cuadrada y que como $\mathbf{I}(\boldsymbol{\theta})$ es simétrica, $\sqrt{\mathbf{I}(\boldsymbol{\theta})}$ también es simétrica, como se mencionó arriba. Además, el lado derecho de la desigualdad es

$$\begin{aligned} \left(\mathbf{e}_i^T \sqrt{\mathbf{I}(\boldsymbol{\theta})}^T \sqrt{\mathbf{I}(\boldsymbol{\theta})} \mathbf{e}_i \right) \left(\mathbf{e}_i^T \sqrt{\mathbf{I}^{-1}(\boldsymbol{\theta})}^T \sqrt{\mathbf{I}^{-1}(\boldsymbol{\theta})} \mathbf{e}_i \right) &= \left(\mathbf{e}_i^T \sqrt{\mathbf{I}(\boldsymbol{\theta})} \sqrt{\mathbf{I}(\boldsymbol{\theta})} \mathbf{e}_i \right) \left(\mathbf{e}_i^T \sqrt{\mathbf{I}^{-1}(\boldsymbol{\theta})} \sqrt{\mathbf{I}^{-1}(\boldsymbol{\theta})} \mathbf{e}_i \right) \\ &= (\mathbf{e}_i^T \mathbf{I}(\boldsymbol{\theta}) \mathbf{e}_i) (\mathbf{e}_i^T \mathbf{I}^{-1}(\boldsymbol{\theta}) \mathbf{e}_i) \\ &= [\mathbf{I}(\boldsymbol{\theta})]_{ii} [\mathbf{I}^{-1}(\boldsymbol{\theta})]_{ii}. \end{aligned}$$

Combinando ambos resultados, se tiene que

$$1 \leq [\mathbf{I}(\boldsymbol{\theta})]_{ii} [\mathbf{I}^{-1}(\boldsymbol{\theta})]_{ii},$$

es decir,

$$\frac{1}{[\mathbf{I}(\boldsymbol{\theta})]_{ii}} \leq [\mathbf{I}^{-1}(\boldsymbol{\theta})]_{ii}.$$

El resultado indica que la CRLB de cada parámetro individual considerando al resto de los parámetros conocidos es inferior que la CRLB de cada parámetro considerando al resto de los parámetros desconocidos. La CRLB se incrementa a medida de que se incrementa la cantidad de parámetros desconocidos.

La igualdad se da si $\mathbf{a} = c\mathbf{b}$, es decir,

$$\sqrt{\mathbf{I}(\boldsymbol{\theta})} \mathbf{e}_i = c_i \sqrt{\mathbf{I}^{-1}(\boldsymbol{\theta})} \mathbf{e}_i \Rightarrow \sqrt{\mathbf{I}(\boldsymbol{\theta})} \sqrt{\mathbf{I}(\boldsymbol{\theta})} \mathbf{e}_i = c_i \mathbf{e}_i \Rightarrow \mathbf{I}(\boldsymbol{\theta}) \mathbf{e}_i = c_i \mathbf{e}_i \quad \forall i = 1, \dots, p.$$

Esto significa que $\mathbf{I}(\boldsymbol{\theta}) = \text{diag}(c_1, \dots, c_p)$, es decir, en el caso en que la matriz de información de Fisher es diagonal, la CRLB de los parámetros desconocidos individuales es igual a la CRLB de los parámetros desconocidos combinados.

3.6.13. Problema 13

Se considera el problema de ajuste de una curva, que es una generalización del problema de ajuste de una recta de la sección 3.2.3. Ahora, el modelo de los datos es

$$x[n] = \sum_{k=0}^{p-1} A_k n^k + w[n]$$

para $n = 0, \dots, N-1$. Como antes, $w[n]$ es WGN con varianza σ^2 . Se desea estimar $\{A_0, \dots, A_{p-1}\}$. Encontrar la matriz de información de Fisher para este problema.

Solución Como en este caso se trata de un señal con parámetro vectorial en WGN, la matriz de información de Fisher puede obtenerse directamente a partir de la ecuación 3.38 con

$$s[n; \boldsymbol{\theta}] = \sum_{k=0}^{p-1} A_k n^k.$$

Por lo tanto,

$$\begin{aligned} [\mathbf{I}(\boldsymbol{\theta})]_{ij} &= \frac{1}{\sigma^2} \sum_{n=0}^{N-1} \left(\frac{\partial}{\partial A_{i-1}} \sum_{k=0}^{p-1} A_k n^k \right) \left(\frac{\partial}{\partial A_{j-1}} \sum_{k=0}^{p-1} A_k n^k \right) \\ &= \frac{1}{\sigma^2} \sum_{n=0}^{N-1} n^{i-1} n^{j-1}, \end{aligned}$$

concluyendo que

$$[\mathbf{I}(\boldsymbol{\theta})]_{ij} = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} n^{i+j-2}, \quad i = 1, \dots, p, \quad j = 1, \dots, p.$$

3.6.14. Problema 14

Se considera la situación del ejemplo de la sección 3.4.3, en el cual los datos son

$$x[n] = A + w[n], \quad n = 0, \dots, N-1$$

donde $w[n]$ es WGN de media nula y varianza σ^2 y A , el nivel de DC, es una variable aleatoria normal de media nula y varianza σ_A^2 independiente de $w[n]$. La potencia de la señal σ_A^2 es el parámetro desconocido, y se considera el estimador $\hat{\sigma}^2 = (\hat{A})^2$, donde \hat{A} es la media muestral. Asíumase que se observa un conjunto de datos dado en el cual la realización de la variable aleatoria A tiene valor A_0 . Mostrar que $\hat{A} \rightarrow A_0$ con $N \rightarrow \infty$ probando que

$$E(\hat{A}|A = A_0) = A_0, \quad \text{var}(\hat{A}|A = A_0) = \frac{\sigma^2}{N}.$$

De esta forma, $\hat{\sigma}^2 \rightarrow A_0^2$ con $N \rightarrow \infty$ para la realización $A = A_0$. Luego, encontrar la varianza de $\hat{\sigma}^2$ cuando $N \rightarrow \infty$ determinando $\text{var}(A^2)$, donde $A \sim \mathcal{N}(0, \sigma_A^2)$ y comparar con la CRLB. Explicar porque σ_A^2 no puede ser estimado sin error incluso aunque $N \rightarrow \infty$.

Solución El caso en que variable aleatoria A está condicionada a tomar el valor A_0 es equivalente al caso en que el parámetro es determinístico, y la media y la varianza ya se calcularon en el problema 1.1.4 y son respectivamente

$$E(\hat{A}|A = A_0) = A_0, \quad \text{var}(\hat{A}|A = A_0) = \frac{\sigma^2}{N}.$$

Se concluye que como $\text{var}(\hat{A}|A = A_0) = \sigma^2/N \rightarrow 0$ con $N \rightarrow \infty$, $\hat{A} \rightarrow A_0$. Además, $\hat{\sigma}^2 = (\hat{A})^2 \rightarrow A_0^2$ con $N \rightarrow \infty$.

Se calculará $\text{var}(\hat{\sigma}^2) = \text{var}[(\hat{A})^2] = E[(\hat{A})^4] - E^2[(\hat{A})^2]$. El momento segundo $E[(\hat{A})^2]$ de la media muestral es

$$\begin{aligned} E[(\hat{A})^2] &= E\left(\frac{1}{N} \sum_{n=0}^{N-1} x[n]\right)^2 \\ &= \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} E(x[i]x[j]) \\ &\stackrel{(a)}{=} \frac{1}{N^2} [NE(x^2[i]) + N(N-1)E(x[i]x[j])] \end{aligned}$$

donde en (a) se consideró que en la doble sumatoria hay N sumandos con $i = j$ y $N(N-1)$ sumandos con $i \neq j$ y además que las variables aleatorias $x[n]$ son idénticamente distribuidas. Por lo tanto, al tomar el límite se tiene que

$$\lim_{N \rightarrow \infty} E^2[(\hat{A})^2] = E^2(x[i]x[j]), \quad i \neq j.$$

El momento cuarto $E[(\hat{A})^4]$ de la media muestral se calculó en el apéndice C. De la ecuación C.1 se deduce que el límite con $N \rightarrow \infty$ es

$$\lim_{N \rightarrow \infty} E[(\hat{A})^4] = E(x[i]x[j]x[k]x[l]),$$

donde i, j, k, l son todos distintos. Por lo tanto,

$$\lim_{N \rightarrow \infty} \text{var}(\hat{\sigma}^2) = E(x[i]x[j]x[k]x[l]) - E^2(x[i]x[j]).$$

Además,

$$E(x[i]x[j]x[k]x[l]) = E[(A + w[i])(A + w[j])(A + w[k])(A + w[l])] \stackrel{(a)}{=} E(A^4) \stackrel{(b)}{=} 3\sigma_A^4,$$

donde en (a) se consideró que como $w[n]$ es WGN, las variables aleatorias $w[i], w[j], w[k], w[l]$ son independientes y además tienen media nula, y en (b) se empleó que el momento cuarto de una variable

aleatoria normal de media nula y varianza σ_A^2 es $3\sigma_A^4$ (ver el capítulo 5 de [2]). De la misma forma, se tiene que

$$E(x[i]x[j]) = E[(A + w[i])(A + w[j])] = E(A^2) = \sigma_A^2.$$

Se concluye que

$$\lim_{N \rightarrow \infty} \text{var}(\hat{\sigma}^2) = 3\sigma_A^4 - \sigma_A^4 = 2\sigma_A^4.$$

En el ejemplo de la sección 3.4.3 se calculó la CRLB de $\hat{\sigma}^2$, la cual está dada por la ecuación 3.39 y es $2\sigma_A^4$ cuando $N \rightarrow \infty$. Por lo tanto, el estimador $(\hat{A})^2$ alcanza la CRLB, y por lo tanto, es eficiente cuando $N \rightarrow \infty$.

Se observa que $\hat{\sigma}^2$ no puede ser estimado sin error incluso aunque $N \rightarrow \infty$. Esto se debe a que el promediado con $N \rightarrow \infty$ cancela el efecto del ruido ($\hat{A} \rightarrow A_0$), pero se cuenta solo con una realización de A .

3.6.15. Problema 15

Se observan las muestras gaussianas bivariadas independientes $\{\mathbf{x}[0], \mathbf{x}[1], \dots, \mathbf{x}[N-1]\}$. Cada observación es un vector 2×1 con distribución $\mathbf{x}[n] \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$ con

$$\mathbf{C} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}.$$

Calcular la CRLB del coeficiente de correlación ρ . Sugerencia: emplear la ecuación 3.37

Solución Se resolverá el problema de dos formas. La primera es realizando todas la cuentas del cálculo de las derivadas de la función de verosimilitud para obtener la información de Fisher dada por la ecuación 3.2, y la segunda es empleando la ecuación 3.37 de la CRLB para el caso gaussiano general.

Para calcular la función de verosimilitud, se parte de la PDF de una única observación $\mathbf{x}[n]$, que por ser bivariada es

$$p(\mathbf{x}[n]; \rho) = \frac{1}{(2\pi) \det^{\frac{1}{2}}(\mathbf{C})} \exp \left\{ -\frac{1}{2} \mathbf{x}[n]^T \mathbf{C}^{-1} \mathbf{x}[n] \right\},$$

y como las observaciones son independientes, la PDF conjunta de los datos es

$$p(\mathbf{x}; \rho) = \frac{1}{(2\pi)^N \det^{\frac{N}{2}}(\mathbf{C})} \exp \left\{ -\frac{1}{2} \sum_{n=0}^{N-1} \mathbf{x}[n]^T \mathbf{C}^{-1} \mathbf{x}[n] \right\}.$$

Tomando el logaritmo, se obtiene la función de verosimilitud,

$$\ln p(\mathbf{x}; \rho) = -N \ln(2\pi) - \frac{N}{2} \ln \det(\mathbf{C}) - \frac{1}{2} \sum_{n=0}^{N-1} \mathbf{x}[n]^T \mathbf{C}^{-1} \mathbf{x}[n],$$

que al derivar respecto al parámetro queda

$$\frac{\partial \ln p(\mathbf{x}; \rho)}{\partial \rho} = -\frac{N}{2} \frac{\partial}{\partial \rho} [\ln \det(\mathbf{C})] - \frac{1}{2} \sum_{n=0}^{N-1} \frac{\partial}{\partial \rho} (\mathbf{x}[n]^T \mathbf{C}^{-1} \mathbf{x}[n]).$$

Teniendo en cuenta que el determinante de \mathbf{C} y la matriz de correlación inversa son

$$\det(\mathbf{C}) = 1 - \rho^2, \quad \mathbf{C}^{-1} = \frac{1}{1 - \rho^2} \begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix},$$

se tiene que

$$\mathbf{C}^{-1} \mathbf{x}[n] = \frac{1}{1 - \rho^2} \begin{bmatrix} x_0[n] - \rho x_1[n] \\ -\rho x_0[n] + x_1[n] \end{bmatrix} \quad \text{y} \quad \mathbf{x}[n]^T \mathbf{C}^{-1} \mathbf{x}[n] = \frac{x_0^2[n] + x_1^2[n] - 2\rho x_0[n]x_1[n]}{1 - \rho^2},$$

donde $\mathbf{x}[n] = [x_0[n] \ x_1[n]]^T$. Por lo tanto,

$$\begin{aligned}
\frac{\partial \ln p(\mathbf{x}; \rho)}{\partial \rho} &= -\frac{N}{2} \frac{\partial}{\partial \rho} [\ln(1 - \rho^2)] - \frac{1}{2} \sum_{n=0}^{N-1} \frac{\partial}{\partial \rho} \left(\frac{x_0^2[n] + x_1^2[n] - 2\rho x_0[n]x_1[n]}{1 - \rho^2} \right) \\
&= -\frac{N}{2} \frac{1}{(1 - \rho^2)} (-2\rho) - \frac{1}{2} \sum_{n=0}^{N-1} \frac{-2x_0[n]x_1[n](1 - \rho^2) - (x_0^2[n] + x_1^2[n] - 2\rho x_0[n]x_1[n])(-2\rho)}{(1 - \rho^2)^2} \\
&= \frac{N\rho}{(1 - \rho^2)} + \sum_{n=0}^{N-1} \frac{x_0[n]x_1[n] + \rho^2 x_0[n]x_1[n] - \rho(x_0^2[n] + x_1^2[n])}{(1 - \rho^2)^2}. \tag{3.71}
\end{aligned}$$

Derivando nuevamente, se tiene que

$$\begin{aligned}
\frac{\partial^2 \ln p(\mathbf{x}; \rho)}{\partial \rho^2} &= \frac{N[(1 - \rho^2) - \rho(-2\rho)]}{(1 - \rho^2)^2} + \sum_{n=0}^{N-1} \frac{(2\rho x_0[n]x_1[n] - x_0^2[n] - x_1^2[n])(1 - \rho^2)^2}{(1 - \rho^2)^4} \\
&\quad - \sum_{n=0}^{N-1} \frac{[x_0[n]x_1[n] + \rho^2 x_0[n]x_1[n] - \rho(x_0^2[n] + x_1^2[n])]2(1 - \rho^2)(-2\rho)}{(1 - \rho^2)^4} \\
&= \frac{N(1 + \rho^2)}{(1 - \rho^2)^2} + \sum_{n=0}^{N-1} \frac{(2\rho x_0[n]x_1[n] - x_0^2[n] - x_1^2[n])(1 - \rho^2)^2}{(1 - \rho^2)^4} \\
&\quad + \sum_{n=0}^{N-1} \frac{[x_0[n]x_1[n] + \rho^2 x_0[n]x_1[n] - \rho(x_0^2[n] + x_1^2[n])]4\rho(1 - \rho^2)}{(1 - \rho^2)^4}
\end{aligned}$$

y al tomar esperanza, se obtiene que

$$\begin{aligned}
E \left[\frac{\partial^2 \ln p(\mathbf{x}; \rho)}{\partial \rho^2} \right] &= \frac{N(1 + \rho^2)}{(1 - \rho^2)^2} + \sum_{n=0}^{N-1} \frac{[2\rho E(x_0[n]x_1[n]) - E(x_0^2[n]) - E(x_1^2[n])](1 - \rho^2)^2}{(1 - \rho^2)^4} \\
&\quad + \sum_{n=0}^{N-1} \frac{\{E(x_0[n]x_1[n]) + \rho^2 E(x_0[n]x_1[n]) - \rho[E(x_0^2[n]) + E(x_1^2[n])]\}4\rho(1 - \rho^2)}{(1 - \rho^2)^4} \\
&\stackrel{(a)}{=} \frac{N(1 + \rho^2)}{(1 - \rho^2)^2} + \sum_{n=0}^{N-1} \frac{(2\rho^2 - 2)(1 - \rho^2)^2 + (\rho + \rho^3 - 2\rho)4\rho(1 - \rho^2)}{(1 - \rho^2)^4} \\
&= \frac{N(1 + \rho^2)}{(1 - \rho^2)^2} + N \frac{(2\rho^2 - 2)(1 - \rho^2)^2 - 4\rho^2(1 - \rho^2)^2}{(1 - \rho^2)^4} \\
&= \frac{N(1 + \rho^2)}{(1 - \rho^2)^2} + \frac{N(2\rho^2 - 2 - 4\rho^2)}{(1 - \rho^2)^2} \\
&= \frac{N(1 + \rho^2)}{(1 - \rho^2)^2} - \frac{2N(1 + \rho^2)}{(1 - \rho^2)^2} \\
&= -\frac{N(1 + \rho^2)}{(1 - \rho^2)^2},
\end{aligned}$$

donde en (a) se empleó que $E(x_0[n]x_1[n]) = \rho$ y $E(x_0^2[n]) = E(x_1^2[n]) = 1$, datos que se obtienen de la matriz de covarianza \mathbf{C} . Se concluye que la información de Fisher es

$$I(\rho) = \frac{N(1 + \rho^2)}{(1 - \rho^2)^2},$$

y la CRLB es

$$\text{var}(\hat{\rho}) \geq \frac{(1 - \rho^2)^2}{N(1 + \rho^2)}.$$

Se resolverá nuevamente el problema empleando la ecuación 3.37 de la CRLB para el caso gaussiano general. Para hacerlo, se comienza observando que como las variables aleatorias $\mathbf{x}[n]$ son IID, la información de Fisher $I(\rho)$ de las N variables aleatorias es la suma de la información de Fisher $i(\rho)$ de cada

variable individual, es decir, $I(\rho) = Ni(\rho)$. En la sección 3.4 se dedujo la información de Fisher de variables aleatorias gaussianas correlacionadas en WGN. De la ecuación 3.37 se obtiene que la información de Fisher de una observación $\mathbf{x}[n]$ es

$$i(\rho) = \frac{1}{2} \text{tr} \left[\left(\mathbf{C}^{-1}(\rho) \frac{\partial \mathbf{C}(\rho)}{\partial \rho} \right)^2 \right],$$

donde se tuvo en cuenta que en este caso $\mathbf{x}[n]$ tiene media nula y por lo tanto, $\partial \boldsymbol{\mu}(\rho)/\partial \rho = \mathbf{0}$. Como

$$\mathbf{C}^{-1}(\rho) = \frac{1}{1-\rho^2} \begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix} \quad y \quad \frac{\partial \mathbf{C}(\rho)}{\partial \rho} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

se tiene que

$$\mathbf{C}^{-1}(\rho) \frac{\partial \mathbf{C}(\rho)}{\partial \rho} = \frac{1}{1-\rho^2} \begin{bmatrix} -\rho & 1 \\ 1 & -\rho \end{bmatrix} \quad y \quad \left(\mathbf{C}^{-1}(\rho) \frac{\partial \mathbf{C}(\rho)}{\partial \rho} \right)^2 = \frac{1}{(1-\rho^2)^2} \begin{bmatrix} \rho^2 + 1 & -2\rho \\ -2\rho & \rho^2 + 1 \end{bmatrix}.$$

Por lo tanto,

$$i(\rho) = \frac{1}{2} \text{tr} \left[\left(\mathbf{C}^{-1}(\rho) \frac{\partial \mathbf{C}(\rho)}{\partial \rho} \right)^2 \right] = \frac{1}{2} \times \frac{2(\rho^2 + 1)}{(1-\rho^2)^2} = \frac{\rho^2 + 1}{(1-\rho^2)^2},$$

resultando en

$$I(\rho) = Ni(\rho) = \frac{N(1+\rho^2)}{(1-\rho^2)^2}.$$

3.6.16. Problema 16

Se desea estimar la potencia total P_0 de un proceso aleatorio WSS, cuya PSD está dada por

$$P_{xx}(f) = P_0 Q(f),$$

donde

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} Q(f) df = 1$$

y $Q(f)$ es conocido. Si se dispone de N observaciones, encontrar la CRLB de la potencia total empleando la expresión exacta dada por la ecuación 3.37 y también empleando la aproximación asintótica dada por la ecuación 3.42. Compara los resultados.

Solución Como indica la ecuación 3.37 y asumiendo que el proceso es gaussiano de media nula, la información de Fisher de P_0 está dada por

$$I(P_0) = \frac{1}{2} \text{tr} \left[\left(\mathbf{C}^{-1}(P_0) \frac{\partial \mathbf{C}(P_0)}{\partial P_0} \right)^2 \right].$$

Para obtener la matriz $N \times N$ de covarianza $\mathbf{C}(P_0)$, se considera que por ser un proceso de media nula, la covarianza coincide con la autocorrelación,

$$[\mathbf{C}(P_0)]_{ij} = r_{xx}[|i-j|]$$

y la autocorrelación es

$$r_{xx}[k] = \mathcal{F}^{-1}\{P_{xx}(f)\} = \mathcal{F}^{-1}\{P_0 Q(f)\} = P_0 \mathcal{F}^{-1}\{Q(f)\} = P_0 q[k],$$

con $q[k]$ conocido. Por lo tanto,

$$[\mathbf{C}(P_0)]_{ij} = P_0 q[|i-j|]$$

o

$$\mathbf{C}(P_0) = P_0 \mathbf{C}_q,$$

donde se definió la matriz $N \times N$ $[\mathbf{C}_q]_{ij} = q[|i - j|]$. Sustituyendo este resultado en la información de Fisher, se tiene que

$$\begin{aligned} I(P_0) &= \frac{1}{2} \text{tr} \left[\left(\frac{1}{P_0} \mathbf{C}_q^{-1} \mathbf{C}_q \right)^2 \right] \\ &= \frac{1}{2} \text{tr} \left(\frac{1}{P_0^2} \mathbf{I} \right) \\ &= \frac{N}{2P_0^2}. \end{aligned}$$

Se concluye que la CRLB de P_0 es

$$\text{var}(\hat{P}_0) \geq \frac{2P_0^2}{N}.$$

Se calculará a continuación la información de Fisher empleando la aproximación asintótica dada por la ecuación 3.42, que en este caso queda

$$\begin{aligned} I(\theta) &= \frac{N}{2} \int_{-\frac{1}{2}}^{\frac{1}{2}} \left(\frac{\partial \ln P_{xx}(f; P_0)}{\partial P_0} \right)^2 df \\ &= \frac{N}{2} \int_{-\frac{1}{2}}^{\frac{1}{2}} \left(\frac{\partial \ln P_0 Q(f)}{\partial P_0} \right)^2 df \\ &= \frac{N}{2} \int_{-\frac{1}{2}}^{\frac{1}{2}} \left[\frac{\partial}{\partial P_0} (\ln P_0 + \ln Q(f)) \right]^2 df \\ &= \frac{N}{2} \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{1}{P_0^2} df, \end{aligned}$$

resultando en

$$I(\theta) = \frac{N}{2P_0^2}.$$

En este ejemplo, la CRLB exacta y la asintótica coinciden, pero esto en general no ocurre.

3.6.17. Problema 17

Si en el ejemplo de la estimación de los parámetros de una senoide de la sección 3.4.4, los datos se observan en el intervalo $n = -M, \dots, M$, calcular la matriz de información de Fisher. ¿Cuál es la CRLB de los parámetros de la senoide? Se pueden aplicar la mismas aproximaciones que en el ejemplo asumiendo que M es grande. Comparar los resultados con los del ejemplo.

Solución Los elementos de la matriz de información de Fisher están dados en la ecuación 3.40, pero ahora las sumatorias son de $n = -M$ a $n = M$. Se reescriben a continuación considerando esto:

$$\begin{aligned}
[\mathbf{I}(\boldsymbol{\theta})]_{11} &= \frac{1}{\sigma^2} \sum_{n=-M}^M \left(\frac{1}{2} + \frac{1}{2} \cos 2\alpha \right) \approx \frac{2M+1}{2\sigma^2} \\
[\mathbf{I}(\boldsymbol{\theta})]_{12} &= -\frac{\pi A}{\sigma^2} \sum_{n=-M}^M n \sin 2\alpha \approx 0 \\
[\mathbf{I}(\boldsymbol{\theta})]_{13} &= -\frac{A}{2\sigma^2} \sum_{n=-M}^M \sin 2\alpha \approx 0 \\
[\mathbf{I}(\boldsymbol{\theta})]_{22} &\approx \frac{2\pi^2 A^2}{\sigma^2} \sum_{n=-M}^M n^2 = \frac{4\pi^2 A^2}{\sigma^2} \sum_{n=1}^M n^2 = \frac{4\pi^2 A^2}{\sigma^2} \frac{M(M+1)(2M+1)}{6} \\
[\mathbf{I}(\boldsymbol{\theta})]_{23} &\approx \frac{\pi A^2}{\sigma^2} \sum_{n=-M}^M n = 0 \\
[\mathbf{I}(\boldsymbol{\theta})]_{33} &= \frac{A^2}{\sigma^2} \sum_{n=-M}^M \left(\frac{1}{2} - \frac{1}{2} \cos 2\alpha \right) \approx \frac{A^2(2M+1)}{2\sigma^2},
\end{aligned}$$

Ahora, el elemento $[\mathbf{I}(\boldsymbol{\theta})]_{23}$ es nulo y por lo tanto, la matriz de información de Fisher es diagonal. La CRLB es ahora

$$\begin{aligned}
\text{var}(\hat{A}) &\geq \frac{2\sigma^2}{2M+1} \\
\text{var}(\hat{f}_0) &\geq \frac{6\sigma^2}{4A^2\pi^2 M(M+1)(2M+1)} \\
\text{var}(\hat{\phi}) &\geq \frac{2\sigma^2}{A^2(2M+1)}.
\end{aligned}$$

En este caso, el número de observaciones es $2M+1$. Para comparar con el ejemplo de la sección 3.4.4, donde el número de observaciones es N , se establece $2M+1 = N$. Al hacerlo, la CRLB queda

$$\begin{aligned}
\text{var}(\hat{A}) &\geq \frac{2\sigma^2}{N} \\
\text{var}(\hat{f}_0) &\geq \frac{6\sigma^2}{4A^2\pi^2 \left(\frac{N-1}{2}\right) \left(\frac{N+1}{2}\right) N} = \frac{6\sigma^2}{A^2\pi^2 N(N^2-1)} \\
\text{var}(\hat{\phi}) &\geq \frac{2\sigma^2}{A^2 N}.
\end{aligned}$$

Se observa comparando con la ecuación 3.41 que la CRLB de los parámetros A y f_0 no cambia, mientras que la de ϕ se reduce.

3.6.18. Problema 20

En el ejemplo de la sección 3.5.2 se calculó la CRLB de los parámetros del filtro de un proceso AR(1). Asumiendo que σ_u^2 es conocido, calcular la CRLB asintótica de $P_{xx}(f)$ para una frecuencia dada empleando la ecuación 3.8 de la CRLB de un parámetro transformado y la ecuación 3.59 de la matriz de información de Fisher de los parámetros del proceso. Para $a[1] = -0.9$, $\sigma_u^2 = 1$, y $N = 100$, graficar la CRLB en función de la frecuencia. Explicar los resultados.

Solución Se quiere calcular la CRLB de la PSD $P_{xx}(f)$ de un proceso AR(1). La PSD de un proceso AR(1) es

$$P_{xx}(f) = \frac{\sigma_u^2}{|1 + a[1]e^{-j2\pi f}|^2}.$$

Teniendo en cuenta que σ_u^2 es conocido, si se dispone de un estimador de $a[1]$, el estimador de $P_{xx}(f)$ es

$$\hat{P}_{xx}(f) = \frac{\sigma_u^2}{|1 + \hat{a}[1]e^{-j2\pi f}|^2}.$$

En el ejemplo de la sección 3.5.2 se calculó la información de Fisher y la CRLB de $a[1]$, que es

$$I(a[1]) = \frac{N}{1 - a^2[1]},$$

como indica la ecuación 3.60. A partir de la ecuación 3.8 de la CRLB de un parámetro transformado, se tiene que la CRLB de $P_{xx}(f)$ es

$$\text{var}(\hat{P}_{xx}(f)) \geq \frac{\left(\frac{\partial P_{xx}(f)}{\partial a[1]}\right)^2}{I(a[1])}.$$

Como

$$\begin{aligned} \frac{\partial P_{xx}(f)}{\partial a[1]} &= \frac{\partial}{\partial a[1]} \left[\frac{\sigma_u^2}{(1 + a[1]e^{-j2\pi f})(1 + a[1]e^{-j2\pi f})^*} \right] \\ &= \frac{\partial}{\partial a[1]} \left[\frac{\sigma_u^2}{(1 + a[1]e^{-j2\pi f})(1 + a[1]e^{j2\pi f})} \right] \\ &= -\frac{\sigma_u^2 [e^{-j2\pi f}(1 + a[1]e^{j2\pi f}) + (1 + a[1]e^{-j2\pi f})e^{j2\pi f}]}{(1 + a[1]e^{-j2\pi f})^2 (1 + a[1]e^{j2\pi f})^2} \\ &= -\frac{\sigma_u^2 (e^{-j2\pi f} + a[1] + e^{j2\pi f} + a[1])}{|1 + a[1]e^{-j2\pi f}|^4} \\ &= -\frac{2\sigma_u^2 (a[1] + \cos(2\pi f))}{|1 + a[1]e^{-j2\pi f}|^4}, \end{aligned}$$

se tiene que

$$\begin{aligned} \text{var}(\hat{P}_{xx}(f)) &\geq \frac{\frac{4\sigma_u^4 (a[1] + \cos(2\pi f))^2}{|1 + a[1]e^{-j2\pi f}|^8}}{\frac{N}{1 - a^2[1]}} \\ &= \frac{4\sigma_u^4 (1 - a^2[1]) (a[1] + \cos(2\pi f))^2}{N |1 + a[1]e^{-j2\pi f}|^8}. \end{aligned}$$

En el caso en que $a[1] = -0.9$, $\sigma_u^2 = 1$, y $N = 100$, la cota es

$$\text{var}(\hat{P}_{xx}(f)) \geq \frac{4 \times 0.19 (-0.9 + \cos(2\pi f))^2}{100 |1 - 0.9e^{-j2\pi f}|^8} = \frac{0.0076 (-0.9 + \cos(2\pi f))^2}{|1 - 0.9e^{-j2\pi f}|^8},$$

la cual se muestra en la figura 3.3 junto con $P_{xx}(f)$ para los mismos valores. Debido a la gran sensibilidad de la PSD ante variaciones de $a[1]$ en torno a frecuencias cercanas a cero, la CRLB es alta en continua. Por ejemplo, en $f = 0$ se cumple que

$$\text{var}(\hat{P}_{xx}(0)) \geq \frac{4\sigma_u^4 (1 - a^2[1]) (a[1] + 1)^2}{N (1 + a[1])^8} = \frac{4\sigma_u^4 (1 - a[1]) (1 + a[1])^3}{N (1 + a[1])^8} = \frac{4\sigma_u^4 (1 - a[1])}{N (1 + a[1])^5} = 7600.$$

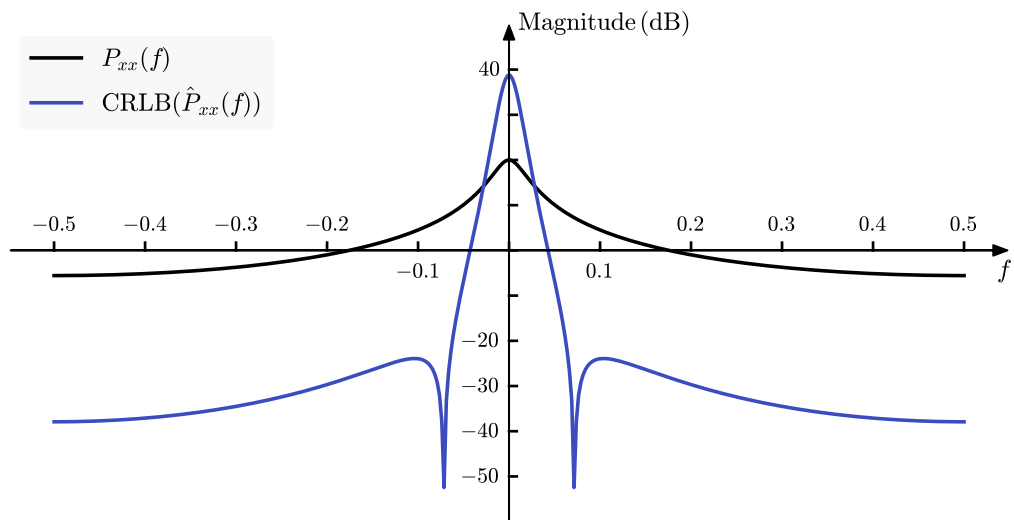


Figura 3.3: $P_{xx}(f)$ y $\text{CRLB}(\hat{P}_{xx}(f))$. La $\text{CRLB}(\hat{P}_{xx}(f))$ es mas alta cerca de continua debido a la gran sensibilidad de $P_{xx}(f)$ ante pequeñas variaciones de $a[1]$ en torno a las bajas frecuencias.

Capítulo 4

Modelos lineales

4.1. Introducción

En muchos problemas en procesamiento de señales, determinar el estimador MVU es difícil y no siempre puede hacerse. Sin embargo hay un modelo de datos que siempre permite determinar el estimador MVU. Este modelo es el modelo lineal de los datos con el parámetro. Si es posible expresar los datos como un modelo lineal, el estimador óptimo, que además es eficiente, se obtiene inmediatamente.

4.2. Definición y propiedades

En el problema de ajuste de una recta explicado en la sección 3.2.3, el modelo de los datos es

$$x[n] = A + Bn + w[n], \quad n = 0, \dots, N-1,$$

y el objetivo es determinar los estimadores MVU de A y B . En notación matricial, este modelo puede expresarse como

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}, \quad (4.1)$$

donde

$$\begin{aligned} \mathbf{x} &= [x[0] \ x[1] \ \dots \ x[N-1]]^T \\ \mathbf{w} &= [w[0] \ w[1] \ \dots \ w[N-1]]^T \\ \boldsymbol{\theta} &= [A \ B]^T \end{aligned}$$

y

$$\mathbf{H} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & N-1 \end{bmatrix}.$$

La matriz \mathbf{H} , de dimensiones $N \times 2$ en este caso, es conocida y se denomina *matriz de observación*. El vector de ruido tiene PDF $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$. En estas condiciones, el modelo de los datos de la ecuación 4.1 es referido como *modelo lineal*.

Como se explicó en el capítulo 3, en ocasiones es posible obtener el estimador MVU si se verifica la condición de igualdad en el teorema de CRLB. Como indica la ecuación 3.12, $\hat{\boldsymbol{\theta}} = \mathbf{g}(\mathbf{x})$ es el estimador MVU eficiente si

$$\frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{I}(\boldsymbol{\theta})(\mathbf{g}(\mathbf{x}) - \boldsymbol{\theta}).$$

Se determinará si esta condición se satisface en el modelo lineal de la ecuación 4.1. La PDF de los datos

es

$$\begin{aligned} p(\mathbf{x}, \boldsymbol{\theta}) &= \frac{1}{(2\pi)^{\frac{N}{2}} \det^{\frac{1}{2}}(\sigma^2 \mathbf{I})} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T (\sigma^2 \mathbf{I})^{-1} (\mathbf{x} - \mathbf{H}\boldsymbol{\theta}) \right\} \\ &= \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T (\mathbf{x} - \mathbf{H}\boldsymbol{\theta}) \right\}, \end{aligned} \quad (4.2)$$

y al tomar logaritmo y derivar respecto al parámetro, se obtiene que

$$\begin{aligned} \frac{\partial p(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= \frac{\partial}{\partial \boldsymbol{\theta}} \left[-\frac{N}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T (\mathbf{x} - \mathbf{H}\boldsymbol{\theta}) \right] \\ &= -\frac{1}{2\sigma^2} \frac{\partial}{\partial \boldsymbol{\theta}} [(\mathbf{x}^T - \boldsymbol{\theta}^T \mathbf{H}^T)(\mathbf{x} - \mathbf{H}\boldsymbol{\theta})] \\ &= -\frac{1}{2\sigma^2} \frac{\partial}{\partial \boldsymbol{\theta}} [\mathbf{x}^T \mathbf{x} - \mathbf{x}^T \mathbf{H}\boldsymbol{\theta} - \boldsymbol{\theta}^T \mathbf{H}^T \mathbf{x} + \boldsymbol{\theta}^T \mathbf{H}^T \mathbf{H}\boldsymbol{\theta}] \\ &= -\frac{1}{2\sigma^2} \frac{\partial}{\partial \boldsymbol{\theta}} [\mathbf{x}^T \mathbf{x} - 2\mathbf{x}^T \mathbf{H}\boldsymbol{\theta} + \boldsymbol{\theta}^T \mathbf{H}^T \mathbf{H}\boldsymbol{\theta}]. \end{aligned}$$

Para calcular la derivada, defínase $\mathbf{b}^T = \mathbf{x}^T \mathbf{H}$ y $\mathbf{A} = \mathbf{H}^T \mathbf{H}$ simétrica, y teniendo en cuenta que se cumple que

$$\frac{\partial \mathbf{b}^T \boldsymbol{\theta}}{\partial \boldsymbol{\theta}} = \mathbf{b}, \quad \frac{\partial \boldsymbol{\theta}^T \mathbf{A} \boldsymbol{\theta}}{\partial \boldsymbol{\theta}} = 2\mathbf{A}\boldsymbol{\theta},$$

como se explica en el apéndice D, se obtiene que

$$\frac{\partial p(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = -\frac{1}{2\sigma^2} [-2\mathbf{H}^T \mathbf{x} + 2\mathbf{H}^T \mathbf{H}\boldsymbol{\theta}] = \frac{1}{\sigma^2} [\mathbf{H}^T \mathbf{x} - \mathbf{H}^T \mathbf{H}\boldsymbol{\theta}],$$

y asumiendo que $\mathbf{H}^T \mathbf{H}$ es invertible, se llega a que

$$\frac{\partial p(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{\mathbf{H}^T \mathbf{H}}{\sigma^2} [(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x} - \boldsymbol{\theta}],$$

ecuación que está factorizada en la forma de la ecuación 3.12 con

$$\hat{\boldsymbol{\theta}} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x} \quad (4.3)$$

$$\mathbf{I}(\boldsymbol{\theta}) = \frac{\mathbf{H}^T \mathbf{H}}{\sigma^2}. \quad (4.4)$$

Se concluye que el estimador de $\boldsymbol{\theta}$ está dado por la ecuación 4.3, y su matriz de covarianza es

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}} = \mathbf{I}^{-1}(\boldsymbol{\theta}) = \sigma^2 (\mathbf{H}^T \mathbf{H})^{-1}.$$

Además, el estimador es eficiente y alcanza la CRLB.

Si bien el razonamiento se realizó en base al problema del ajuste de una recta, el resultado es general cuando los datos se pueden modelar de forma lineal, como indica el siguiente teorema.

4.3. Teorema: estimador MVU para el modelo lineal

Hipótesis Los datos observados pueden modelarse como

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w},$$

donde \mathbf{x} de dimensiones $N \times 1$ es el *vector de observaciones*, \mathbf{H} de dimensiones $N \times p$ es la *matriz de observación*, conocida, con $N > p$ y rango p , $\boldsymbol{\theta}$ de dimensiones $p \times 1$ es el *vector de parámetros* a estimar y \mathbf{w} de dimensiones $N \times 1$ es el vector de ruido con PDF $\mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$.

Tesis El estimador MVU es

$$\hat{\boldsymbol{\theta}} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x}$$

y la matriz de covarianza de $\hat{\theta}$ es

$$\mathbf{C}_{\hat{\theta}} = \sigma^2 (\mathbf{H}^T \mathbf{H})^{-1}. \quad (4.5)$$

Además, el estimador MVU es eficiente ya que alcanza la CRLB para todo θ .

Para confirmar que el estimador es insesgado, se observa que

$$\begin{aligned}\hat{\boldsymbol{\theta}} &= (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x} \\ &\stackrel{(a)}{=} (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T (\mathbf{H} \boldsymbol{\theta} + \mathbf{w}) \\ &= \boldsymbol{\theta} + (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{w}\end{aligned}$$

donde en (a) se substituyó el modelo lineal dado por la ecuación 4.1, y al tomar esperanza, se tiene que

$$\begin{aligned} E(\hat{\boldsymbol{\theta}}) &= E\left(\boldsymbol{\theta} + (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{w}\right) \\ &= \boldsymbol{\theta} + (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T E(\mathbf{w}) \\ &\stackrel{(b)}{=} \boldsymbol{\theta} \end{aligned}$$

donde en (b) se tuvo en cuenta que el ruido es de media nula, $E(\mathbf{w}) = \mathbf{0}$. Además, el estimador queda completamente especificado estadísticamente. Efectivamente, como $\hat{\boldsymbol{\theta}}$ es una transformación lineal del vector gaussiano \mathbf{x} ,

$$\hat{\boldsymbol{\theta}} \sim \mathcal{N}(\boldsymbol{\theta}, \sigma^2 (\mathbf{H}^T \mathbf{H})^{-1}).$$

4.4. Ejemplos

4.4.1. Ajuste de curva

Una extensión al problema del ajuste de una recta presentado en la sección 3.2.3 es el ajuste de una curva, en el cual el modelo de los datos es

$$x[n] = \sum_{k=0}^{p-1} A_k n^k + w[n], \quad n = 0, \dots, N-1,$$

donde $w[n]$ es WGN de varianza σ^2 y se quiere determinar los p parámetros A_k , $k = 0, \dots, p-1$. Al igual que en el problema del ajuste de una recta explicado al comienzo de este capítulo, el modelo de los datos es lineal,

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$$

donde

$$\mathbf{H} = \begin{bmatrix} 0^0 & 0^1 & 0^2 & \dots & 0^{p-1} \\ 1^0 & 1^1 & 1^2 & \dots & 1^{p-1} \\ 2^0 & 2^1 & 2^2 & \dots & 2^{p-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ (N-1)^0 & (N-1)^1 & (N-1)^2 & \dots & (N-1)^{p-1} \end{bmatrix} \quad y \quad \theta = \begin{bmatrix} A_0 \\ A_1 \\ \vdots \\ A_{p-1} \end{bmatrix}.$$

La matriz \mathbf{H} tiene dimensiones $N \times p$ con $\mathbf{H}_{ij} = (i-1)^{j-1}$. Una matriz con esta estructura se denomina matriz de Vandermonde. El estimador y su matriz de covarianza están dados por las ecuaciones 4.3 y 4.5. La matriz de información de Fisher para este caso se calculó en el problema de la sección 3.6.13. A continuación se muestra una simulación por computadora para el caso con $p = 4$ donde los coeficientes a estimar son

$$A_0 = 1, \quad A_1 = 0.73, \quad A_2 = -1.6 \times 10^{-2}, \quad A_3 = 1 \times 10^{-4},$$

la cantidad de muestras es $N = 100$ y el ruido tiene varianza $\sigma^2 = 10$. La varianza de los coeficientes son los elementos de la diagonal de la matriz de covarianza (ecuación 4.5), y con estos valores es

$$\text{var}(\hat{A}_0) \approx 1.49, \quad \text{var}(\hat{A}_1) \approx 1.15 \times 10^{-02}, \quad \text{var}(\hat{A}_2) \approx 6.36 \times 10^{-06}, \quad \text{var}(\hat{A}_3) \approx 2.80 \times 10^{-10}$$

En la figura 4.1 se muestra el resultado de una realización. Las estimaciones de los parámetros en esa

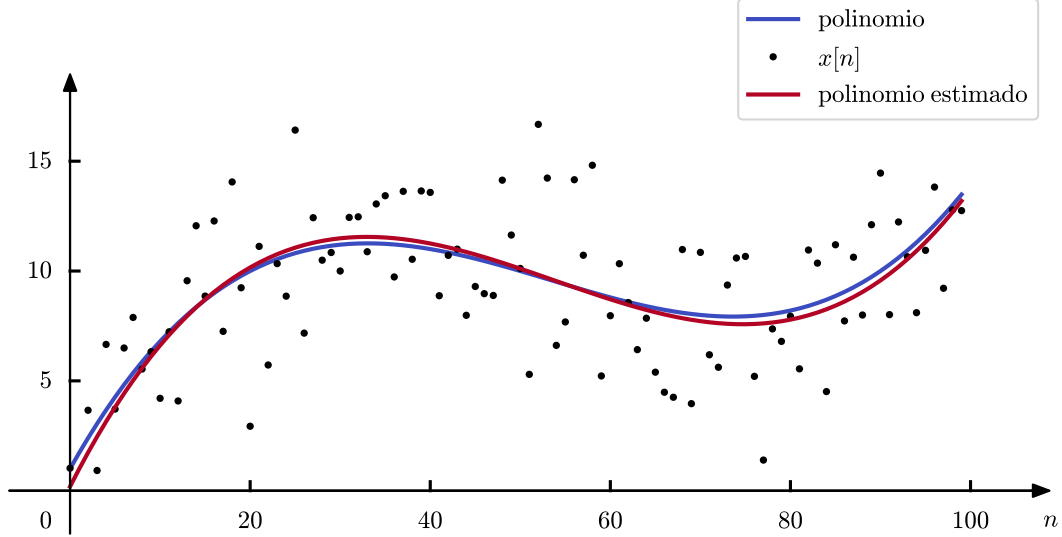


Figura 4.1: Ajuste de curva: estimación de los coeficientes de un polinomio de grado 3 ($p = 4$). En la figura se muestra el polinomio original, las $N = 100$ muestras ruidosas con potencia de ruido $\sigma^2 = 10$ y el polinomio estimado.

realización son

$$\hat{A}_0 \approx 0.188, \quad \hat{A}_1 \approx 0.808, \quad \hat{A}_2 \approx -1.77 \times 10^{-2}, \quad \hat{A}_3 \approx 1.09 \times 10^{-4}.$$

4.4.2. Análisis de Fourier

Muchas señales tienen un comportamiento periódico. La presencia de componentes sinusoidales puede ser detectada mediante el análisis de Fourier, en el cual grandes coeficientes de Fourier indican la presencia de fuertes componentes sinusoidales. Este ejemplo muestra que el análisis de Fourier es en realidad la estimación de los parámetros de un modelo lineal. Considérese el modelo de datos que consiste en sinusoides en WGN:

$$x[n] = \sum_{k=1}^M a_k \cos\left(\frac{2\pi kn}{N}\right) + \sum_{k=1}^M b_k \sin\left(\frac{2\pi kn}{N}\right) + w[n], \quad n = 0, \dots, N-1, \quad (4.6)$$

donde $w[n]$ es WGN. Se asume que las frecuencias de las sinusoides están relacionadas armónicamente, siendo múltiplos de la frecuencia fundamental $f_1 = 1/N$, es decir, $f_k = k/N$. El objetivo es estimar las amplitudes a_k y b_k de los cosenos y senos a partir de las observaciones ruidosas $x[n]$. Los datos pueden expresarse en el modelo lineal de la ecuación 4.1 con

$$\boldsymbol{\theta} = [a_1 \ a_2 \ \dots \ a_M \ b_1 \ b_2 \ \dots \ b_M]^T \quad (4.7)$$

y

$$\mathbf{H} = \begin{bmatrix} 1 & \dots & 1 & 0 & \dots & 0 \\ \cos\left(\frac{2\pi}{N}\right) & \dots & \cos\left(\frac{2\pi M}{N}\right) & \sin\left(\frac{2\pi}{N}\right) & \dots & \sin\left(\frac{2\pi M}{N}\right) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \cos\left[\frac{2\pi(N-1)}{N}\right] & \dots & \cos\left[\frac{2\pi M(N-1)}{N}\right] & \sin\left[\frac{2\pi(N-1)}{N}\right] & \dots & \sin\left[\frac{2\pi M(N-1)}{N}\right] \end{bmatrix}. \quad (4.8)$$

$\boldsymbol{\theta}$ es de dimensión $2M \times 1$ y \mathbf{H} es de dimensión $N \times 2M$, donde $p = 2M$. Por lo tanto, para que \mathbf{H} satisfaga la condición $N > p$ se requiere que $M < N/2$. El cálculo del estimador MVU se simplifica notando que las columnas de \mathbf{H} son ortogonales. Si se representa \mathbf{H} en columnas,

$$\mathbf{H} = [\mathbf{h}_1 \quad \mathbf{h}_2 \quad \cdots \quad \mathbf{h}_{2M}]$$

donde \mathbf{h}_i es la columna i -ésima de \mathbf{H} , se cumple que

$$\mathbf{h}_i^T \mathbf{h}_j = 0 \quad \text{para } i \neq j.$$

Considerando esta propiedad,

$$\mathbf{H}^T \mathbf{H} = \begin{bmatrix} \mathbf{h}_1^T \\ \mathbf{h}_2^T \\ \vdots \\ \mathbf{h}_{2M}^T \end{bmatrix} [\mathbf{h}_1 \quad \mathbf{h}_2 \quad \cdots \quad \mathbf{h}_{2M}] = \begin{bmatrix} \mathbf{h}_1^T \mathbf{h}_1 & \mathbf{h}_1^T \mathbf{h}_2 & \cdots & \mathbf{h}_1^T \mathbf{h}_{2M} \\ \mathbf{h}_2^T \mathbf{h}_1 & \mathbf{h}_2^T \mathbf{h}_2 & \cdots & \mathbf{h}_2^T \mathbf{h}_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{h}_{2M}^T \mathbf{h}_1 & \mathbf{h}_{2M}^T \mathbf{h}_2 & \cdots & \mathbf{h}_{2M}^T \mathbf{h}_{2M} \end{bmatrix}$$

es una matriz diagonal y por lo tanto, fácil de invertir. La ortogonalidad de las columnas proviene de las relaciones

$$\begin{aligned} \sum_{n=0}^{N-1} \cos\left(\frac{2\pi in}{N}\right) \cos\left(\frac{2\pi jn}{N}\right) &= \frac{N}{2} \delta_{ij} \\ \sum_{n=0}^{N-1} \sin\left(\frac{2\pi in}{N}\right) \sin\left(\frac{2\pi jn}{N}\right) &= \frac{N}{2} \delta_{ij} \\ \sum_{n=0}^{N-1} \cos\left(\frac{2\pi in}{N}\right) \sin\left(\frac{2\pi jn}{N}\right) &= 0 \quad \forall i, j, \end{aligned}$$

válidas para $i, j = 1, \dots, M < N/2$, que se demuestran en el problema de la sección 4.6.5. A partir de estas relaciones, se obtiene que

$$\mathbf{H}^T \mathbf{H} = \begin{bmatrix} \frac{N}{2} & 0 & \cdots & 0 \\ 0 & \frac{N}{2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{N}{2} \end{bmatrix} = \frac{N}{2} \mathbf{I},$$

por lo que el estimador MVU de las amplitudes es

$$\hat{\boldsymbol{\theta}} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x} = \frac{2}{N} \mathbf{I} \mathbf{H}^T \mathbf{x} = \frac{2}{N} \begin{bmatrix} \mathbf{h}_1^T \\ \mathbf{h}_2^T \\ \vdots \\ \mathbf{h}_{2M}^T \end{bmatrix} \mathbf{x} = \begin{bmatrix} \frac{2}{N} \mathbf{h}_1^T \mathbf{x} \\ \frac{2}{N} \mathbf{h}_2^T \mathbf{x} \\ \vdots \\ \frac{2}{N} \mathbf{h}_{2M}^T \mathbf{x} \end{bmatrix}.$$

Finalmente

$$\begin{aligned} \hat{a}_k &= \frac{2}{N} \sum_{n=0}^{N-1} x[n] \cos\left(\frac{2\pi kn}{N}\right) \\ \hat{b}_k &= \frac{2}{N} \sum_{n=0}^{N-1} x[n] \sin\left(\frac{2\pi kn}{N}\right). \end{aligned} \tag{4.9}$$

Se concluye que \hat{a}_k y \hat{b}_k son los coeficientes de Fourier de $x[n]$, como se muestra en el apéndice E (ver la ecuación E.2). La matriz de covarianza es

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}} = \sigma^2 (\mathbf{H}^T \mathbf{H})^{-1} = \frac{2\sigma^2}{N} \mathbf{I}. \tag{4.10}$$

Como la matriz de covarianza es diagonal, los estimadores son independientes.

4.4.3. Identificación de un sistema

En ocasiones es de interés identificar un sistema desconocido. Un modelo común del sistema es como un filtro de respuesta al impulso finita (FIR, *finite impulse response*), como se muestra en la figura 4.2, y el objetivo es estimar los coeficientes $h[k]$ del filtro. La estrategia es alimentar al sistema con una entrada $u[n]$ conocida y observar la salida, que idealmente será $\sum_{k=0}^{p-1} h[k]u[n-k]$, y a partir de ésta, estimar los coeficientes del filtro, o equivalentemente, la respuesta al impulso. En la práctica, la salida observada

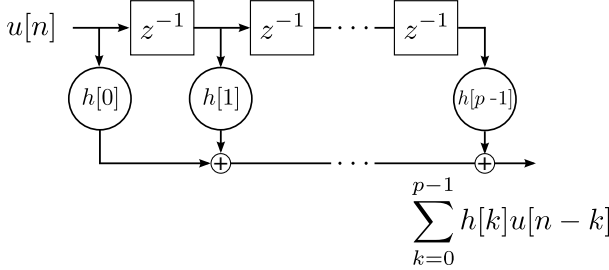


Figura 4.2: Modelo de un sistema desconocido como filtro FIR.

es ruidosa, por lo que un modelo mas apropiado es como en la figura 4.3. Supóngase que la entrada es

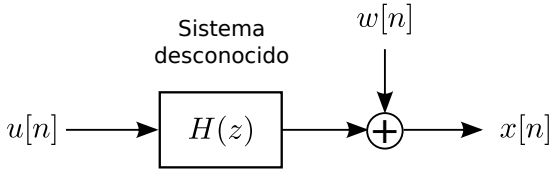


Figura 4.3: Modelo del sistema con la salida corrompida con ruido.

provista en $n = 0, \dots, N-1$ y la salida se observa en el mismo intervalo,

$$x[n] = \sum_{k=0}^{p-1} h[k]u[n-k] + w[n] \quad n = 0, \dots, N-1,$$

donde se asume que $u[n] = 0$ si $n < 0$. En notación matricial, las ecuaciones se pueden expresar como

$$\mathbf{x} = \underbrace{\begin{bmatrix} u[0] & 0 & \cdots & 0 \\ u[1] & u[0] & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ u[p-1] & u[p-2] & \cdots & u[0] \\ \vdots & \vdots & & \vdots \\ u[N-1] & u[N-2] & \cdots & u[N-p] \end{bmatrix}}_{\mathbf{H}} \underbrace{\begin{bmatrix} h[0] \\ h[1] \\ \vdots \\ h[p-1] \end{bmatrix}}_{\boldsymbol{\theta}} + \mathbf{w}.$$

Asumiendo que el ruido $w[n]$ es WGN, las ecuaciones tienen la forma del modelo lineal de la ecuación 4.1, por lo que el estimador MVU de la respuesta al impulso es

$$\hat{\boldsymbol{\theta}} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x},$$

y la matriz de covarianza del estimador es

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}} = \sigma^2 (\mathbf{H}^T \mathbf{H})^{-1}.$$

Se comenzará calculando $\mathbf{H}^T \mathbf{H}$. Para hacerlo, se parte notando que la columna j -ésima de \mathbf{H} es

$$[\mathbf{H}]_j = [\underbrace{0 \dots 0}_{j-1} u[0] \dots u[N-j]]^T, \quad 1 \leq j \leq p,$$

por lo que el elemento (i, j) de $\mathbf{H}^T \mathbf{H}$, con $i \geq j$, es

$$\begin{aligned}
[\mathbf{H}^T \mathbf{H}]_{ij} &= [\mathbf{H}]_i^T [\mathbf{H}]_j = \underbrace{[0 \dots 0]_{i-1}}_{i-1} u[0] \dots u[N-i] \left[\begin{array}{c} 0 \\ \vdots \\ 0 \\ u[0] \\ \vdots \\ u[N-j] \end{array} \right] \Bigg\} j-1 \\
&= \underbrace{0 + \dots + 0}_{i-1} + u[0]u[i-j] + u[1]u[i-j+1] + \dots + u[N-i]u[N-j] \\
&= \sum_{n=0}^{N-i} u[n]u[n+i-j].
\end{aligned}$$

Para valores arbitrarios de (i, j) , esto se generaliza como

$$[\mathbf{H}^T \mathbf{H}]_{ij} = \sum_{n=0}^{N-\max(i, j)} u[n]u[n+|i-j|]. \quad (4.11)$$

Notar que la matriz $\mathbf{H}^T \mathbf{H}$ es simétrica. Como se muestra en el problema de la sección 4.6.7, en el caso en que $N \gg p$ se cumple que

$$[\mathbf{H}^T \mathbf{H}]_{ij} \approx \sum_{n=0}^{N-1-|i-j|} u[n]u[n+|i-j|], \quad (4.12)$$

que es la *función de correlación* de la secuencia determinística $u[n]$. De esta forma,

$$\mathbf{H}^T \mathbf{H} = N \begin{bmatrix} r_{uu}[0] & r_{uu}[1] & r_{uu}[2] & \dots & r_{uu}[p-1] \\ r_{uu}[1] & r_{uu}[0] & r_{uu}[1] & \dots & r_{uu}[p-2] \\ r_{uu}[2] & r_{uu}[1] & r_{uu}[0] & \dots & r_{uu}[p-3] \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{uu}[p-1] & r_{uu}[p-2] & r_{uu}[p-3] & \dots & r_{uu}[0] \end{bmatrix} = N \mathbf{R}_{uu} \quad (4.13)$$

donde

$$r_{uu}[k] = \frac{1}{N} \sum_{n=0}^{N-1-k} u[n]u[n+k] \quad (4.14)$$

es la función de autocorrelación muestral de $u[n]$. El motivo de realizar la aproximación es que de esta forma, la matriz $\mathbf{H}^T \mathbf{H}$ además de ser simétrica, es una matriz Toeplitz de autocorrelación, lo que simplifica los resultados. Esto es

$$[\mathbf{H}^T \mathbf{H}]_{ij} = [\mathbf{H}^T \mathbf{H}]_{ji} \quad (\text{simétrica}) \quad [\mathbf{H}^T \mathbf{H}]_{ij} = [\mathbf{H}^T \mathbf{H}]_{i+k, j+k} \quad (\text{Toeplitz}).$$

Por otro lado,

$$\mathbf{H}^T \mathbf{x} = \begin{bmatrix} u[0] & u[1] & \dots & u[p-1] & \dots & u[N-1] \\ 0 & u[0] & \dots & u[p-2] & \dots & u[N-2] \\ \vdots & \vdots & \ddots & \vdots & & \vdots \\ 0 & 0 & \dots & u[0] & \dots & u[N-p] \end{bmatrix} \begin{bmatrix} x[0] \\ x[1] \\ \vdots \\ x[N-1] \end{bmatrix}$$

por lo que el elemento $i+1$ de $\mathbf{H}^T \mathbf{x}$, $i = 0, \dots, p-1$ es

$$\begin{aligned}
[\mathbf{H}^T \mathbf{x}]_{i+1} &= \sum_{n=0}^{N-i} u[n]x[n+i] \\
&= N r_{ux}[i],
\end{aligned}$$

donde

$$r_{ux}[k] = \frac{1}{N} \sum_{n=0}^{N-k} u[n]x[n+k] \quad (4.15)$$

es la correlación cruzada muestral entre $u[n]$ y $x[n]$. Finalmente, $\mathbf{H}^T \mathbf{x}$ queda

$$\mathbf{H}^T \mathbf{x} = N \begin{bmatrix} r_{ux}[0] \\ r_{ux}[1] \\ \vdots \\ r_{ux}[p-1] \end{bmatrix} = N \mathbf{r}_{ux}, \quad (4.16)$$

donde \mathbf{r}_{ux} es vector de correlación cruzada entre $u[n]$ y $x[n]$. Sustituyendo las ecuaciones 4.13 y 4.16 en las ecuaciones 4.3 y 4.5 del estimador óptimo y la matriz de covarianza del modelo lineal se llega a que

$$\hat{\mathbf{h}} = \mathbf{R}_{uu}^{-1} \mathbf{r}_{ux} \quad (4.17)$$

$$\mathbf{C}_{\hat{h}} = \frac{\sigma^2}{N} \mathbf{R}_{uu}^{-1}. \quad (4.18)$$

Los coeficientes de varianza mínima del filtro se obtienen multiplicando la matriz de autocorrelación inversa de la entrada con la correlación cruzada entre la entrada y la salida. Este resultado coincide con los coeficientes del filtro de Wiener, y en ese contexto, el sistema de ecuaciones 4.17 se denomina *ecuaciones de Wiener-Hopf* (ver por ejemplo el capítulo 7 de [5]).

La ecuación 4.18 indica que la varianza de los coeficientes óptimos $\hat{\mathbf{h}}$ dependen de la señal de prueba $u[n]$ a través de \mathbf{R}_{uu} . Es de interés encontrar la señal de prueba que produzca la mínima varianza de los coeficientes. Se mostrará a continuación que la señal de prueba que minimiza la varianza es ruido no correlacionado. La varianza del coeficiente i -ésimo $\hat{h}[i]$ es

$$\text{var}(\hat{h}[i]) = [\mathbf{C}_{\hat{h}}]_{ii} = \mathbf{e}_i^T \mathbf{C}_{\hat{h}} \mathbf{e}_i \quad \text{con} \quad \mathbf{e}_i = [0 \ 0 \ \dots \ 0 \ \underbrace{1}_{\text{posición } i} \ 0 \ \dots \ 0]^T.$$

Como $\mathbf{C}_{\hat{h}}$ es simétrica y definida positiva, también lo es $\mathbf{C}_{\hat{h}}^{-1}$, por lo que puede factorizarse aplicando la descomposición de Cholesky¹ como

$$\mathbf{C}_{\hat{h}}^{-1} = \mathbf{D}^T \mathbf{D},$$

donde \mathbf{D} es una matriz $p \times p$ invertible². Teniendo esto en cuenta y notando que

$$\mathbf{e}_i^T \mathbf{I} \mathbf{e}_i = 1 \quad \Rightarrow \quad \mathbf{e}_i^T \mathbf{D}^T \mathbf{D}^{T-1} \mathbf{e}_i = 1 \quad \Rightarrow \quad (\mathbf{e}_i^T \mathbf{D}^T \mathbf{D}^{T-1} \mathbf{e}_i)^2 = 1,$$

se definen los vectores

$$\boldsymbol{\xi}_1 = \mathbf{D} \mathbf{e}_i, \quad \boldsymbol{\xi}_2 = \mathbf{D}^{T-1} \mathbf{e}_i,$$

y la igualdad se puede expresar como

$$(\boldsymbol{\xi}_1^T \boldsymbol{\xi}_2)^2 = 1.$$

Aplicando la desigualdad de Cauchy-Schwarz (ecuación 3.70), se tiene que

$$(\boldsymbol{\xi}_1^T \boldsymbol{\xi}_1) (\boldsymbol{\xi}_2^T \boldsymbol{\xi}_2) \geq (\boldsymbol{\xi}_1^T \boldsymbol{\xi}_2)^2 = 1,$$

y sustituyendo, se llega a que

$$\begin{aligned} (\boldsymbol{\xi}_1^T \boldsymbol{\xi}_1) (\boldsymbol{\xi}_2^T \boldsymbol{\xi}_2) &= (\mathbf{e}_i^T \mathbf{D}^T \mathbf{D} \mathbf{e}_i) (\mathbf{e}_i^T \mathbf{D}^{T-1} \mathbf{D}^{T-1} \mathbf{e}_i) \\ &= (\mathbf{e}_i^T \mathbf{D}^T \mathbf{D} \mathbf{e}_i) (\mathbf{e}_i^T \mathbf{D}^{-1} \mathbf{D}^{T-1} \mathbf{e}_i) \\ &= (\mathbf{e}_i^T \mathbf{C}_{\hat{h}}^{-1} \mathbf{e}_i) (\mathbf{e}_i^T \mathbf{C}_{\hat{h}} \mathbf{e}_i) \\ &\geq 1. \end{aligned}$$

¹Ver por ejemplo https://en.wikipedia.org/wiki/Cholesky_decomposition.

²Una matriz de covarianza es semidefinida-positiva. A una matriz simétrica y semi-definida positiva también puede aplicarse la descomposición de Cholesky si se permite que elementos de la diagonal de \mathbf{D} sean nulos, lo que hace que \mathbf{D} no sea invertible. Es por esto que se asume aquí que $\mathbf{C}_{\hat{h}}$ es definida positiva.

Finalmente, se obtiene que

$$\mathbf{e}_i^T \mathbf{C}_{\hat{h}} \mathbf{e}_i = \text{var}(h[i]) \geq \frac{1}{\mathbf{e}_i^T \mathbf{C}_{\hat{h}}^{-1} \mathbf{e}_i}.$$

Se observa que la varianza es mínima cuando la desigualdad de Cauchy-Schwarz se cumple con igualdad. Esto sucede cuando ξ_1 y ξ_2 son colineales, es decir

$$\xi_1 = c\xi_2,$$

donde c es una constante. Por lo tanto, la condición para que la varianza de todos los coeficientes sea mínima es que

$$\mathbf{D}\mathbf{e}_i = c_i \mathbf{D}^T \mathbf{e}_i,$$

o equivalentemente,

$$\mathbf{D}^T \mathbf{D} \mathbf{e}_i = c_i \mathbf{e}_i, \quad i = 1, 2, \dots, p,$$

donde c_i son constantes, $i = 1, \dots, p$. Teniendo en cuenta que $\mathbf{D}^T \mathbf{D} = \mathbf{C}_{\hat{h}}^{-1}$, la condición queda

$$\mathbf{C}_{\hat{h}}^{-1} \mathbf{e}_i = c_i \mathbf{e}_i, \quad i = 1, 2, \dots, p,$$

o en palabras, la columna i -ésima de $\mathbf{C}_{\hat{h}}^{-1}$ tiene solo un elemento no nulo c_i en la posición i . Se concluye que la matriz de covarianza inversa $\mathbf{C}_{\hat{h}}^{-1}$ tiene que ser diagonal. Por la ecuación 4.18

$$\mathbf{C}_{\hat{h}}^{-1} = \frac{N}{\sigma^2} \mathbf{R}_{uu},$$

indicando que la matriz de autocorrelación de la entrada debe ser diagonal,

$$r_{uu}[k] = 0 \quad \forall k \neq 0,$$

es decir,

$$\mathbf{R}_{uu} = r_{uu}[0] \mathbf{I},$$

lo cual se cumple si $u[n]$ es ruido no correlacionado. En esas condiciones, el estimador queda (ecuación 4.17)

$$\hat{\mathbf{h}} = \mathbf{R}_{uu}^{-1} \mathbf{r}_{ux} = \frac{1}{r_{uu}[0]} \mathbf{r}_{ux},$$

resultando en

$$\hat{h}[i] = \frac{r_{ux}[i]}{r_{uu}[0]}, \quad i = 1, 2, \dots, p. \quad (4.19)$$

Además, de la ecuación 4.18 se obtiene que la matriz de covarianza es

$$\mathbf{C}_{\hat{h}} = \frac{\sigma^2}{Nr_{uu}[0]} \mathbf{I}$$

de lo cual se deduce que por ser una matriz diagonal, los coeficientes son independientes con varianza

$$\text{var}(\hat{h}[i]) = \frac{\sigma^2}{Nr_{uu}[0]}, \quad i = 1, 2, \dots, p.$$

Por una deducción en el dominio de la frecuencia del problema de identificación de un sistema, ver el problema de la sección 4.6.8.

Simulación por computadora

A continuación se muestran los resultados de una simulación por computadora de la estimación de un sistema. El sistema desconocido es un filtro de respuesta al impulso finita (*finite impulse response*, FIR) de orden $p = 11$ causal de respuesta al impulso triangular, la cual se muestra en la figura 4.4. Se usan dos tipos de señales de prueba de largo $N = 1000$ muestras. La señal $u_1[n]$ es ruido blanco gaussiano y la señal $u_2[n]$ es ruido coloreado. Ambas señales tienen potencia unidad.

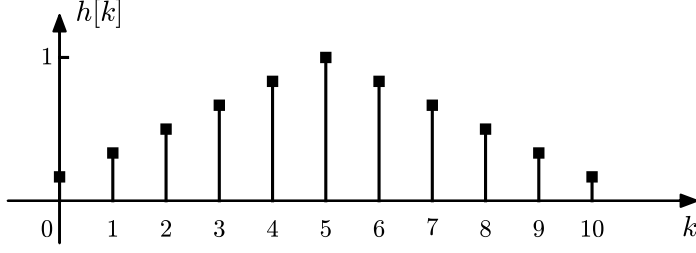


Figura 4.4: Respuesta al impulso del filtro a estimar.

Ruido coloreado $u_2[n]$ El ruido coloreado $u_2[n]$ usado como señal de prueba es un proceso AR(2),

$$u_2[n] + a_1 u_2[n-1] + a_2 u_2[n-2] = v[n],$$

con $a_1 = -0.8$ y $a_2 = 0.64$, donde $v[n]$ es ruido blanco gaussiano. Hay que tener en cuenta que la relación de potencias entre $v[n]$ y $u_2[n]$ es (ver el capítulo 1 de [4])

$$\sigma_{u_2}^2 = \frac{1 + a_2}{(1 - a_2)[(1 + a_2)^2 - a_1^2]} \sigma_v^2,$$

por lo que $v[n]$ debe tener la potencia adecuada para que $u_2[n]$ tenga potencia unidad. La PSD $S_{u_2 u_2}(\omega)$ del proceso $u_2[n]$ es

$$S_{u_2 u_2}(\omega) = \frac{\sigma_v^2}{|1 + a_1 e^{-j\omega} + a_2 e^{-2j\omega}|^2},$$

y la función de autocorrelación puede obtenerse mediante la transformada inversa de Fourier de la PSD (calculada por computadora) o mediante las ecuaciones de Yule-Walker. Con los parámetros elegidos, la señal $u_2[n]$ tiene PSD y función de autocorrelación mostradas en la figura 4.5.

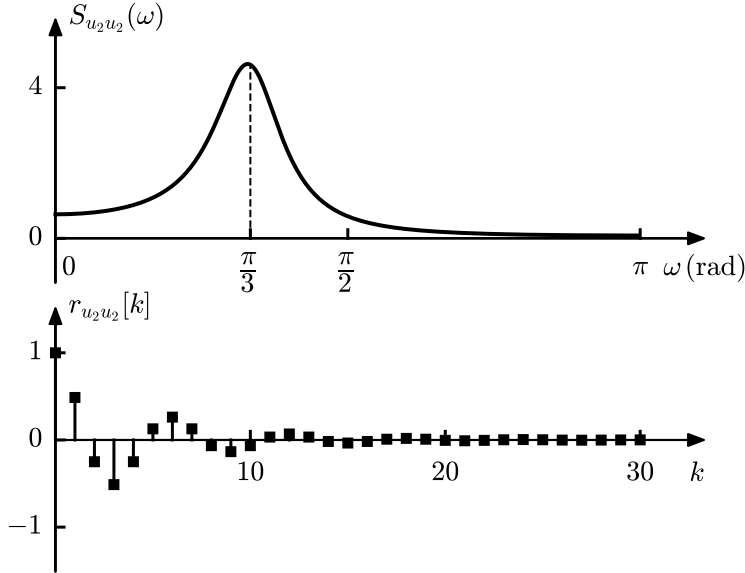


Figura 4.5: PSD $S_{u_2 u_2}(\omega)$ y función de autocorrelación $r_{u_2 u_2}[k]$ de la señal $u_2[n]$. El proceso es pasabanda con frecuencia de resonancia de $\pi/3$ radianes.

Salidas del sistema desconocido Los procesos $u_1[n]$ y $u_2[n]$ se emplean como entradas al sistema desconocido, produciendo la salidas $x_1[n]$ y $x_2[n]$ respectivamente, las cuales se observan contaminadas con ruido (WGN) $w[n]$. La SNR de las señales observadas es 0 dB, es decir, la potencia σ_w^2 del ruido es

igual a la potencia $\sigma_{x_i}^2$ de las salidas $x_i[n]$, $i = 1, 2$. La potencia de las salidas es

$$\begin{aligned}
\sigma_{x_i}^2 &\stackrel{(a)}{=} E(x_i^2[n]) \\
&= E\left[\left(\sum_{k=0}^{p-1} h[k]u_i[n-k]\right)^2\right] \\
&= E\left[\left(\sum_{k=0}^{p-1} h[k]u_i[n-k]\right)\left(\sum_{l=0}^{p-1} h[l]u_i[n-l]\right)\right] \\
&= E\left(\sum_{k=0}^{p-1} \sum_{l=0}^{p-1} h[k]h[l]u_i[n-k]u_i[n-l]\right) \\
&= \sum_{k=0}^{p-1} \sum_{l=0}^{p-1} h[k]h[l]E(u_i[n-k]u_i[n-l]) \\
&\stackrel{(b)}{=} \sum_{k=0}^{p-1} \sum_{l=0}^{p-1} h[k]h[l]\sigma_{u_i}^2 \delta_{kl} \\
&= \sigma_{u_i}^2 \sum_{k=0}^{p-1} h^2[k],
\end{aligned}$$

donde en (a) se consideró que $E(x_i[n]) = 0$ y en (b) se tuvo en cuenta que $E(u_i[n-k]u_i[n-l]) = \sigma_{u_i}^2$ si $k = l$ y cero en otro caso. En la figura 4.6 se muestran las primeras 100 muestras de una realización de los

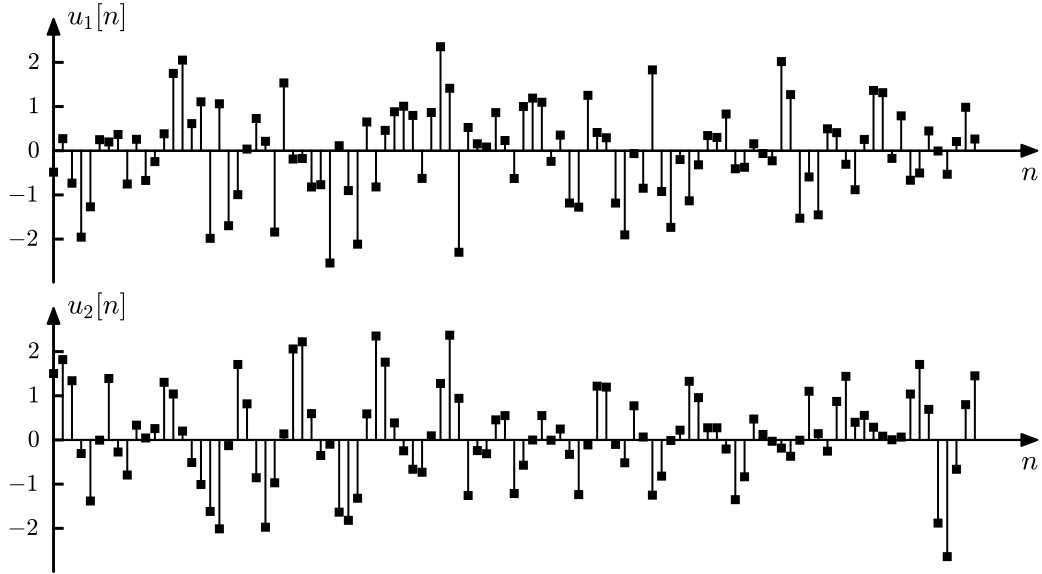


Figura 4.6: Primeras 100 muestras de una realización de los procesos $u_1[n]$ y $u_2[n]$ de entrada. $u_1[n]$ es WGN y $u_2[n]$ es un proceso AR(2). Se observa la naturaleza pseudo-periódica de $u_2[n]$, acorde a su PSD.

procesos de entrada $u_1[n]$ y $u_2[n]$ y en la figura 4.7 se muestran las primeras 100 muestras de las salidas $x_1[n]$ y $x_2[n]$ y las salidas observadas contaminadas con ruido $x_1[n] + w[n]$ y $x_2[n] + w[n]$.

Identificación del sistema Se asume que se conoce el orden p del filtro FIR desconocido. En el caso en que el proceso de entrada es WGN, como $u_1[n]$, esto no implica una pérdida de generalidad, ya que los estimadores de cada coeficiente son independientes y de igual varianza, y por lo tanto, se puede usar un

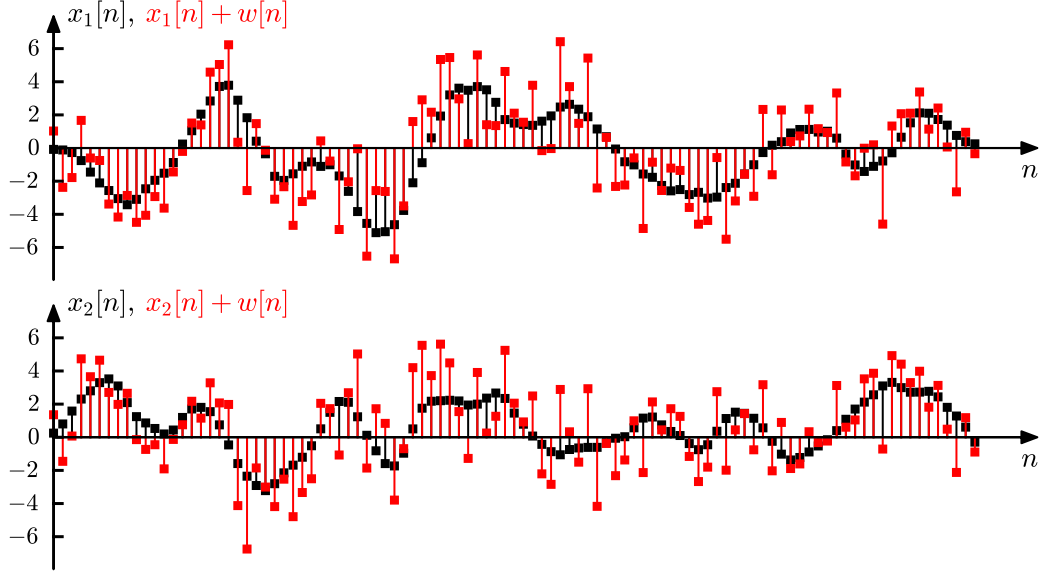


Figura 4.7: Primeras 100 muestras de una realización de las salidas $x_1[n]$ y $x_2[n]$ del sistema desconocido. $u_1[n]$ es WGN y $u_2[n]$ es un proceso AR(2). Se observa la naturaleza pseudo-periódica de $u_2[n]$, acorde a su PSD.

valor grande de p sin cambiar los resultados. En este caso, la estimación de los coeficientes que superen el orden del sistema desconocido, tendrán un valor cercano a cero.

Para identificar el sistema, se calculan las primeras p muestras de la autocorrelación muestral de la entrada $u_i[n]$ y la correlación cruzada muestral entre la entrada $u_i[n]$ y la salida contaminada con ruido $x_i[n] + w[n]$ empleando las ecuaciones 4.14 y 4.15 respectivamente. Con la autocorrelación se construye la matriz Toeplitz de autocorrelación de tamaño p , y se obtiene el estimador del sistema resolviendo el sistema de ecuaciones 4.17. El resultado de la estimación para una realización de cada tipo de entrada se muestra en la figura 4.8. Como se observa en la figura, el error en la estimación es considerablemente

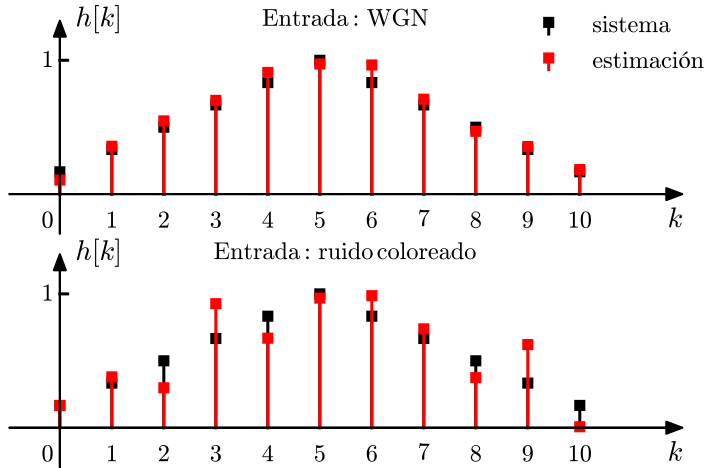


Figura 4.8: Resultado de la estimación de los coeficientes $h[k]$ del sistema desconocido empleando ruido blanco y ruido coloreado como señales de prueba.

grande debido a que si bien el número de muestras $N = 1000$ de las señales de prueba es grande, la relación señal a ruido es mala. En el caso particular de las realizaciones mostradas en las figuras, la estimación usando WGN parece un poco mejor que la estimación empleando ruido coloreado. Esto es

acorde a lo concluido previamente de que la entrada que produce estimadores de mínima varianza es WGN.

Finalmente, se calcula la CRLB de los estimadores. Para hacerlo, se calcula la matriz de covarianza empleando la ecuación 4.18, y los elementos de la diagonal de la matriz son la varianza de cada coeficiente. También se calcula la varianza práctica de los estimadores de cada coeficiente. Para esto, se repite el experimento y se calcula el error cuadrático (*square error*, SE) de cada coeficiente en cada experimento como

$$SE(\hat{h}[i]) = (\hat{h}[i] - h[i])^2.$$

Promediando el error cuadrático en todos los experimentos, se obtiene la varianza o el error cuadrático medio. En la figura 4.9 se muestra la CRLB y la varianza práctica repitiendo el experimento 10000 veces. Como se observa en la figura y se dedujo analíticamente, los estimadores son eficientes en el sentido de

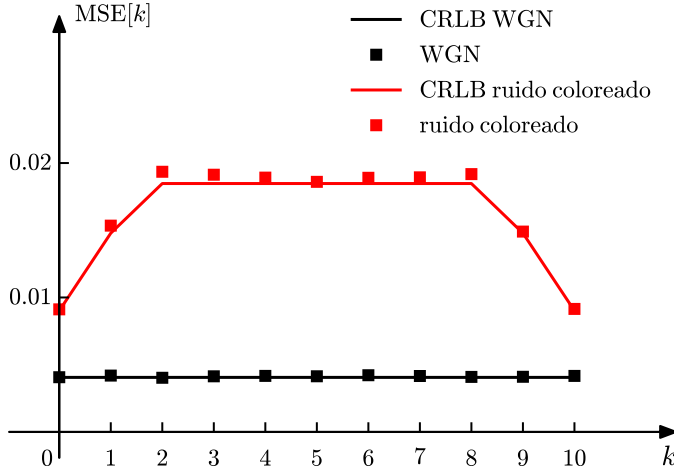


Figura 4.9: Comparación de la CRLB y la varianza práctica de los estimadores de cada coeficiente del sistema desconocido. En este problema, los estimadores son eficientes. Además, la entrada que produce estimadores de varianza mínima es WGN.

que alcanzan la CRLB. Adicionalmente, también se observa que la estimación empleando WGN produce estimadores de varianza menor que la producida empleando ruido coloreado.

4.5. Extensión del modelo lineal

Una formulación mas general del modelo lineal admite ruido que no es blanco. El *modelo lineal general* asume que el ruido es

$$\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}),$$

donde \mathbf{C} es una matriz genérica de covarianza y no una matriz identidad escalada. La determinación del estimador MVU puede hacerse de forma análoga al caso con ruido blanco (ver el problema de la sección 4.6.9). A continuación se realizará la deducción empleando un enfoque de blanqueado. Asumiendo que la matriz \mathbf{C} es definida positiva además de simétrica, \mathbf{C}^{-1} también es simétrica y definida positiva y por lo tanto, mediante la descomposición de Cholesky puede factorizarse como

$$\mathbf{C}^{-1} = \mathbf{D}^T \mathbf{D},$$

donde \mathbf{D} es una matriz $N \times N$ invertible. La matriz \mathbf{D} actúa como una transformación de blanqueado sobre el ruido, ya que si se considera $\mathbf{w}' = \mathbf{D}\mathbf{w}$, su esperanza es

$$E(\mathbf{w}') = \mathbf{D}E(\mathbf{w}) = \mathbf{0}$$

y su varianza es

$$\begin{aligned}
E(\mathbf{w}'\mathbf{w}'^T) &= E[(\mathbf{D}\mathbf{w})(\mathbf{D}\mathbf{w})^T] \\
&= \mathbf{D}E(\mathbf{w}\mathbf{w}^T)\mathbf{D}^T \\
&= \mathbf{D}\mathbf{C}\mathbf{D}^T \\
&= \mathbf{D}(\mathbf{D}^T\mathbf{D})^{-1}\mathbf{D}^T \\
&= \mathbf{D}\mathbf{D}^{-1}\mathbf{D}^{T^{-1}}\mathbf{D}^T \\
&= \mathbf{I},
\end{aligned}$$

y por lo tanto,

$$\mathbf{w}' = \mathbf{D}\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}).$$

Como consecuencia, si se transforma el modelo generalizado

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$$

a

$$\begin{aligned}
\mathbf{x}' &= \mathbf{D}\mathbf{x} \\
&= \mathbf{D}\mathbf{H}\boldsymbol{\theta} + \mathbf{D}\mathbf{w} \\
&= \mathbf{H}'\boldsymbol{\theta} + \mathbf{w}',
\end{aligned}$$

resulta en el modelo lineal usual con ruido blanco. De la ecuación 4.3, el estimador MVU para $\boldsymbol{\theta}$ es

$$\begin{aligned}
\hat{\boldsymbol{\theta}} &= (\mathbf{H}'^T\mathbf{H}')^{-1}\mathbf{H}'^T\mathbf{x}' \\
&= (\mathbf{H}^T\mathbf{D}^T\mathbf{D}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{D}^T\mathbf{D}\mathbf{x}
\end{aligned}$$

resultando en

$$\hat{\boldsymbol{\theta}} = (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}^{-1}\mathbf{x}. \quad (4.20)$$

De la ecuación 4.5, se obtiene que la matriz de covarianza del estimador es

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}} = (\mathbf{H}'^T\mathbf{H}')^{-1}$$

o finalmente,

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}} = (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1} \quad (4.21)$$

Si $\mathbf{C} = \sigma^2\mathbf{I}$ se obtiene el resultado previo para el caso con WGN.

4.5.1. Ejemplo: nivel de DC en ruido coloreado

Se generalizará a continuación el ejemplo la estimación del nivel de continua al caso en el que el ruido es coloreado. Sea

$$x[n] = A + w[n], \quad n = 0, \dots, N-1,$$

donde $w[n]$ es ruido coloreado gaussiano con matriz de covarianza \mathbf{C} de dimensiones $N \times N$, y se quiere determinar el estimador MVU de A y su varianza. Representando el modelo en forma matricial, se observa que es lineal,

$$\mathbf{x} = \mathbf{1}A + \mathbf{w}.$$

Como en este caso

$$\mathbf{H} = \mathbf{1} = [1 \ 1 \ \dots \ 1]^T,$$

de la ecuación 4.20 surge que el estimador MVU para el nivel de DC es

$$\begin{aligned}
\hat{A} &= (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}^{-1}\mathbf{x} \\
&= \frac{\mathbf{1}^T\mathbf{C}^{-1}\mathbf{x}}{\mathbf{1}^T\mathbf{C}^{-1}\mathbf{1}},
\end{aligned}$$

donde en la segunda igualdad se consideró que con $\mathbf{H} = \mathbf{1}$, $\mathbf{H}^T \mathbf{C}^{-1} \mathbf{H}$ y $\mathbf{H}^T \mathbf{C}^{-1} \mathbf{x}$ son escalares, y de la ecuación 4.21, se obtiene que la varianza es

$$\begin{aligned} \text{var}(\hat{A}) &= (\mathbf{H}^T \mathbf{C}^{-1} \mathbf{H})^{-1} \\ &= \frac{1}{\mathbf{1}^T \mathbf{C}^{-1} \mathbf{1}}. \end{aligned}$$

Notar que si el ruido es blanco, $\mathbf{C} = \sigma^2 \mathbf{I}$, y el estimador MVU es la media muestral,

$$\hat{A} = \frac{\mathbf{1}^T \left(\frac{1}{\sigma^2} \mathbf{I} \right) \mathbf{x}}{\mathbf{1}^T \left(\frac{1}{\sigma^2} \mathbf{I} \right) \mathbf{1}} = \frac{\mathbf{1}^T \mathbf{x}}{\mathbf{1}^T \mathbf{1}} = \frac{\sum_{n=0}^{N-1} x[n]}{N},$$

con varianza

$$\text{var}(\hat{A}) = \frac{1}{\mathbf{1}^T \left(\frac{1}{\sigma^2} \mathbf{I} \right) \mathbf{1}} = \frac{\sigma^2}{\mathbf{1}^T \mathbf{1}} = \frac{\sigma^2}{N},$$

como se dedujo previamente.

Una interpretación interesante del estimador MVU puede extraerse considerando la factorización de \mathbf{C}^{-1} como $\mathbf{D}^T \mathbf{D}$. Previamente se mostró que la matriz \mathbf{D} blanquea el ruido. Usando la factorización, el estimador MVU se expresa como

$$\hat{A} = \frac{\mathbf{1}^T \mathbf{D}^T \mathbf{D} \mathbf{x}}{\mathbf{1}^T \mathbf{D}^T \mathbf{D} \mathbf{1}} = \frac{(\mathbf{D} \mathbf{1})^T \mathbf{x}'}{(\mathbf{D} \mathbf{1})^T (\mathbf{D} \mathbf{1})} = \sum_{n=0}^{N-1} d_n x'[n], \quad (4.22)$$

con

$$d_n = \frac{[\mathbf{D} \mathbf{1}]_n}{(\mathbf{D} \mathbf{1})^T (\mathbf{D} \mathbf{1})} = \frac{[\mathbf{D} \mathbf{1}]_n}{\sum_{n=0}^{N-1} [\mathbf{D} \mathbf{1}]_n^2}. \quad (4.23)$$

De acuerdo a este resultado, los datos son blanqueados obteniendo $x'[n]$ y luego se promedian de forma ponderada empleando los pesos d_n del blanqueado. El blanqueado tiene el efecto de decorrelacionar e igualar las varianzas de las muestras de ruido. Ver los problemas de la sección 4.6.10 y 4.6.11.

4.6. Problemas

4.6.1. Problema 1

Se desea estimar las amplitudes de exponenciales en ruido. Los datos observados son

$$x[n] = \sum_{i=1}^p A_i r_i^n + w[n], \quad n = 0, \dots, N-1,$$

donde $w[n]$ es WGN con varianza σ^2 y los valores r_i son conocidos. Encontrar el estimador MVU de las amplitudes y también su covarianza. Evaluar el resultado en el caso en que $p = 2$, $r_1 = 1$, $r_2 = -1$, y N es par.

Solución Los datos se pueden expresar como un modelo lineal,

$$\mathbf{x} = \mathbf{H} \mathbf{A} + \mathbf{w},$$

con $\mathbf{A} = [A_1 \ A_2 \ \dots \ A_p]^T$ y

$$\mathbf{H} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ r_1 & r_2 & \dots & r_p \\ r_1^2 & r_2^2 & \dots & r_p^2 \\ \vdots & \vdots & \dots & \vdots \\ r_1^{N-1} & r_2^{N-1} & \dots & r_p^{N-1} \end{bmatrix}.$$

El estimador y su matriz de covarianza están dados por las ecuaciones 4.3 y 4.5 respectivamente, y son

$$\hat{\mathbf{A}} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x}, \quad \mathbf{C}_{\hat{\mathbf{A}}} = \sigma^2 (\mathbf{H}^T \mathbf{H})^{-1},$$

donde en este caso, el elemento (i, j) de la matriz $\mathbf{H}^T \mathbf{H}$ es

$$[\mathbf{H}^T \mathbf{H}]_{i,j} = \mathbf{h}_i^T \mathbf{h}_j = [1 \quad r_i \quad r_i^2 \quad \dots \quad r_i^{N-1}] \begin{bmatrix} 1 \\ r_j \\ r_j^2 \\ \vdots \\ r_j^{N-1} \end{bmatrix} = \sum_{n=0}^{N-1} (r_i r_j)^n = \frac{1 - (r_i r_j)^N}{1 - r_i r_j},$$

donde \mathbf{h}_k es la columna k -ésima de \mathbf{H} .

En el caso en que $p = 2$, $r_1 = 1$, $r_2 = -1$, y N es par, se tiene que

$$\mathbf{H}^T \mathbf{H} = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & -1 & 1 & \dots & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} N & 0 \\ 0 & N \end{bmatrix} = N \mathbf{I}.$$

Además,

$$\mathbf{H}^T \mathbf{x} = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & -1 & 1 & \dots & -1 \end{bmatrix} \begin{bmatrix} x[0] \\ x[1] \\ \vdots \\ x[N-1] \end{bmatrix} = \begin{bmatrix} \sum_{n=0}^{N-1} x[n] \\ \sum_{n=0}^{N-1} (-1)^n x[n] \end{bmatrix}.$$

Por lo tanto, el estimador es

$$\hat{\mathbf{A}} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x} = \frac{1}{N} \mathbf{I} \begin{bmatrix} \sum_{n=0}^{N-1} x[n] \\ \sum_{n=0}^{N-1} (-1)^n x[n] \end{bmatrix} = \begin{bmatrix} \frac{1}{N} \sum_{n=0}^{N-1} x[n] \\ \frac{1}{N} \sum_{n=0}^{N-1} (-1)^n x[n] \end{bmatrix},$$

y la matriz de covarianza es

$$\mathbf{C}_{\hat{\mathbf{A}}} = \sigma^2 (\mathbf{H}^T \mathbf{H})^{-1} = \frac{\sigma^2}{N} \mathbf{I}.$$

4.6.2. Problema 2

Probar que la inversa de la matriz $\mathbf{H}^T \mathbf{H}$ existe si y solo si las columnas de \mathbf{H} son linealmente independientes. Sugerencia: el problema es equivalente a probar que $\mathbf{H}^T \mathbf{H}$ es definida positiva y por lo tanto invertible si y solo si las columnas son linealmente independientes.

Solución Si las columnas de \mathbf{H} son linealmente independientes, se cumple por definición que

$$\mathbf{H} \mathbf{x} = \sum_{i=1}^p x_i \mathbf{h}_i \neq 0, \quad \forall \mathbf{x} \neq 0,$$

y por lo tanto $\|\mathbf{H} \mathbf{x}\|^2 > 0$ para todo $\mathbf{x} \neq 0$. Teniendo esto en cuenta, se cumple que

$$\|\mathbf{H} \mathbf{x}\|^2 = (\mathbf{H} \mathbf{x})^T (\mathbf{H} \mathbf{x}) = \mathbf{x}^T \mathbf{H}^T \mathbf{H} \mathbf{x} > 0, \quad \forall \mathbf{x} \neq 0,$$

que por definición significa que $\mathbf{H}^T \mathbf{H}$ es definida positiva y por lo tanto, invertible.

4.6.3. Problema 3

Considérese la matriz de observación

$$\mathbf{H} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 + \epsilon \end{bmatrix}$$

donde ϵ es pequeño. Calcular $(\mathbf{H}^T \mathbf{H})^{-1}$ y verificar que ocurre cuando $\epsilon \rightarrow 0$. Si $\mathbf{x} = [2 \ 2 \ 2]^T$, encontrar el estimador MVU y describir que ocurre cuando $\epsilon \rightarrow 0$.

Solución

$$\mathbf{H}^T \mathbf{H} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 + \epsilon \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 + \epsilon \end{bmatrix} = \begin{bmatrix} 3 & 3 + \epsilon \\ 3 + \epsilon & 3 + 2\epsilon + \epsilon^2 \end{bmatrix}$$

Para calcular la matriz inversa, se comienza calculando el determinante, que es

$$\det(\mathbf{H}^T \mathbf{H}) = 3(3 + 2\epsilon + \epsilon^2) - (3 + \epsilon)^2 = (9 + 6\epsilon + 3\epsilon^2) - (9 + 6\epsilon + \epsilon^2) = 2\epsilon^2,$$

por lo que la matriz inversa es

$$(\mathbf{H}^T \mathbf{H})^{-1} = \frac{1}{2\epsilon^2} \begin{bmatrix} 3 + 2\epsilon + \epsilon^2 & -3 - \epsilon \\ -3 - \epsilon & 3 \end{bmatrix}.$$

Cuando $\epsilon \rightarrow \infty$, todos los elementos tienden a ∞ .

El estimador MVU es (ecuación 4.3)

$$\hat{\boldsymbol{\theta}} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x},$$

y en este caso,

$$\mathbf{H}^T \mathbf{x} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 + \epsilon \end{bmatrix} \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 6 \\ 6 + 2\epsilon \end{bmatrix}.$$

Por lo tanto,

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= \frac{1}{2\epsilon^2} \begin{bmatrix} 3 + 2\epsilon + \epsilon^2 & -3 - \epsilon \\ -3 - \epsilon & 3 \end{bmatrix} \begin{bmatrix} 6 \\ 6 + 2\epsilon \end{bmatrix} \\ &= \frac{1}{\epsilon^2} \begin{bmatrix} 3 + 2\epsilon + \epsilon^2 & -3 - \epsilon \\ -3 - \epsilon & 3 \end{bmatrix} \begin{bmatrix} 3 \\ 3 + \epsilon \end{bmatrix} \\ &= \frac{1}{\epsilon^2} \begin{bmatrix} (9 + 6\epsilon + 3\epsilon^2) - (9 + 6\epsilon + \epsilon^2) \\ -(9 + 3\epsilon) + (9 + 3\epsilon) \end{bmatrix} \\ &= \begin{bmatrix} 2 \\ 0 \end{bmatrix}. \end{aligned}$$

Se observa que incluso aunque $\epsilon \rightarrow 0$, $\hat{\boldsymbol{\theta}} = [2 \ 0]^T$. Esto se debe a que \mathbf{x} pertenece al subespacio generado por la primer columna de \mathbf{H} , la cual no depende de ϵ . Mas precisamente, el modelo lineal es $\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$ y \mathbf{x} en este caso es la primer columna de \mathbf{H} multiplicada por 2.

4.6.4. Problema 4

En el modelo lineal se desea estimar la señal $\mathbf{s} = \mathbf{H}\boldsymbol{\theta}$. Si se encuentra un estimador $\boldsymbol{\theta}$ MVU, la señal puede ser estimada como $\hat{\mathbf{s}} = \mathbf{H}\hat{\boldsymbol{\theta}}$. ¿Cuál es la PDF de $\hat{\mathbf{s}}$? Aplicar los resultados al modelo lineal del ejemplo de la sección 4.4.2.

Solución En el modelo lineal, $\hat{\boldsymbol{\theta}} \sim \mathcal{N}(\boldsymbol{\theta}, \mathbf{C}_{\hat{\boldsymbol{\theta}}})$, con $\mathbf{C}_{\hat{\boldsymbol{\theta}}} = \sigma^2(\mathbf{H}^T \mathbf{H})^{-1}$. Como los elementos de $\hat{\mathbf{s}}$ son una combinación lineal de variables aleatorias gaussianas, también son gaussianos, y por lo tanto, para determinar su PDF, alcanza con encontrar la media y la matriz de covarianza. La media es

$$E(\hat{\mathbf{s}}) = H E(\hat{\boldsymbol{\theta}}) = H \boldsymbol{\theta},$$

y la matriz de covarianza es

$$\begin{aligned} \mathbf{C}_{\hat{\mathbf{s}}} &= E\{[\hat{\mathbf{s}} - E(\hat{\mathbf{s}})][\hat{\mathbf{s}} - E(\hat{\mathbf{s}})]^T\} \\ &= E\{[\mathbf{H}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})][\mathbf{H}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})]^T\} \\ &= \mathbf{H} E[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T] \mathbf{H}^T \\ &= \mathbf{H} \mathbf{C}_{\hat{\boldsymbol{\theta}}} \mathbf{H}^T, \end{aligned}$$

por lo que se concluye que

$$\hat{\mathbf{s}} = \mathbf{H} \hat{\boldsymbol{\theta}} \sim \mathcal{N}(H \boldsymbol{\theta}, \sigma^2 \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T).$$

En el ejemplo del análisis de Fourier de la sección 4.4.2, el vector de parámetros $\boldsymbol{\theta}$ esta dado por la ecuación 4.7 y la matriz de observación \mathbf{H} está dada en la ecuación 4.8. Por lo tanto,

$$\hat{s}[n] = \sum_{k=1}^M \hat{a}_k \cos\left(\frac{2\pi kn}{N}\right) + \sum_{k=1}^M \hat{b}_k \sin\left(\frac{2\pi kn}{N}\right),$$

donde los estimadores MVU están dados por la ecuación 4.9. La media de $\hat{\mathbf{s}}$ es $\mathbf{H} \boldsymbol{\theta}$, y teniendo en cuenta que la matriz de covarianza del estimador $\hat{\boldsymbol{\theta}}$ está dada por la ecuación 4.21, la matriz de covarianza de $\hat{\mathbf{s}}$ es

$$\mathbf{C}_{\hat{\mathbf{s}}} = \mathbf{H} \mathbf{C}_{\hat{\boldsymbol{\theta}}} \mathbf{H}^T = \mathbf{H} \left(\frac{2\sigma^2}{N} \mathbf{I} \right) \mathbf{H}^T = \frac{2\sigma^2}{N} \mathbf{H} \mathbf{H}^T.$$

Se concluye que la PDF de $\hat{\mathbf{s}}$ es

$$\hat{\mathbf{s}} \sim \mathcal{N}\left(\mathbf{H} \boldsymbol{\theta}, \frac{2\sigma^2}{N} \mathbf{H} \mathbf{H}^T\right).$$

4.6.5. Problema 5

Probar que para $k, l = 1, 2, \dots, M < N/2$

$$\sum_{n=0}^{N-1} \cos\left(\frac{2\pi kn}{N}\right) \cos\left(\frac{2\pi ln}{N}\right) = \frac{N}{2} \delta_{kl}$$

usando la identidad trigonométrica

$$\cos \omega_1 \cos \omega_2 = \frac{1}{2} \cos(\omega_1 + \omega_2) + \frac{1}{2} \cos(\omega_1 - \omega_2) \quad (4.24)$$

y observando que

$$\sum_{n=0}^{N-1} \cos \alpha n = \operatorname{Re} \left(\sum_{n=0}^{N-1} e^{j\alpha n} \right).$$

Solución Usando la identidad de la ecuación 4.24, se cumple que

$$\begin{aligned} \sum_{n=0}^{N-1} \cos\left(\frac{2\pi kn}{N}\right) \cos\left(\frac{2\pi ln}{N}\right) &= \sum_{n=0}^{N-1} \left\{ \frac{1}{2} \cos\left[\frac{2\pi(k+l)n}{N}\right] + \frac{1}{2} \cos\left[\frac{2\pi(k-l)n}{N}\right] \right\} \\ &= \frac{1}{2} \sum_{n=0}^{N-1} \cos\left[\frac{2\pi(k+l)n}{N}\right] + \frac{1}{2} \sum_{n=0}^{N-1} \cos\left[\frac{2\pi(k-l)n}{N}\right] \end{aligned} \quad (4.25)$$

Además,

$$\begin{aligned}
\sum_{n=0}^{N-1} \cos\left(\frac{2\pi Mn}{N}\right) &= \operatorname{Re} \left[\sum_{n=0}^{N-1} \exp\left(\frac{j2\pi Mn}{N}\right) \right] \\
&\stackrel{(a)}{=} \operatorname{Re} \left[\frac{1 - e^{j2\pi M}}{1 - e^{j2\pi M/N}} \right] \\
&= \operatorname{Re} \left[\frac{e^{j\pi M}}{e^{j\pi M/N}} \frac{e^{j\pi M} - e^{-j\pi M}}{e^{j\pi M/N} - e^{-j\pi M/N}} \right] \\
&= \operatorname{Re} \left[e^{j\pi M \frac{N-1}{N}} \frac{\sin(\pi M)}{\sin(\frac{\pi M}{N})} \right] \\
&= \cos \left[\frac{\pi M(N-1)}{N} \right] \frac{\sin(\pi M)}{\sin(\frac{\pi M}{N})} \\
&\stackrel{(b)}{=} \begin{cases} N & \text{si } M = 0 \\ 0 & \text{si } M \in \mathbb{Z}, 0 < |M| < N \end{cases} ,
\end{aligned}$$

donde en (a) se empleó el resultado de la suma finita de una serie geométrica, que indica que³

$$\sum_{n=0}^{N-1} r^n = \frac{1 - r^N}{1 - r}, \quad r \neq 1, \quad (4.26)$$

y en (b) se empleó la regla de L'Hôpital⁴ para evaluar la expresión en $M = 0$,

$$\lim_{M \rightarrow 0} \frac{\sin(\pi M)}{\sin(\frac{\pi M}{N})} = \lim_{M \rightarrow 0} \frac{\pi \cos(\pi M)}{\frac{\pi}{N} \sin(\frac{\pi M}{N})} = N.$$

Este resultado indica que la primera sumatoria de la ecuación 4.25 es nula para todo k, l y la segunda sumatoria es nula excepto cuando $k = l$, caso en que vale N . Se concluye que

$$\sum_{n=0}^{N-1} \cos\left(\frac{2\pi kn}{N}\right) \cos\left(\frac{2\pi ln}{N}\right) = \frac{N}{2} \delta_{kl}$$

4.6.6. Problema 6

Supóngase que en el ejemplo del análisis de Fourier de la sección 4.4.2 la señal consiste solo en un componente sinusoidal en $f_k = k/N$. El modelo dado por la ecuación 4.6 es entonces

$$x[n] = a_k \cos(2\pi f_k n) + b_k \sin(2\pi f_k n) + w[n], \quad n = 0, \dots, N-1.$$

Usando la identidad $A \cos \omega + B \sin \omega = \sqrt{A^2 + B^2} \cos(\omega - \phi)$, donde $\phi = \arctan(B/A)$, el modelo se puede escribir como

$$x[n] = \sqrt{a_k^2 + b_k^2} \cos(2\pi f_k n - \phi) + w[n].$$

Un estimador MVU se usa para a_k y b_k , de forma que la potencia estimada de la sinusoides es

$$\hat{P} = \frac{\hat{a}_k^2 + \hat{b}_k^2}{2}.$$

Una medida de la detectabilidad es $E^2(\hat{P})/\operatorname{var}(\hat{P})$. Comparar la medida cuando una sinusoides está presente con el caso en que solo hay ruido o $a_k = b_k = 0$. ¿Es posible emplear la potencia estimada para determinar si una señal está presente?

³Ver https://en.wikipedia.org/wiki/Geometric_series, por ejemplo

⁴Ver https://en.wikipedia.org/wiki/L'Hôpital's_rule

Solución Primero se mostrará la identidad $A \cos \omega + B \sin \omega = \sqrt{A^2 + B^2} \cos(\omega - \phi)$, donde $\phi = \arctan(B/A)$. Para hacerlo, elíjanse los valores C y ϕ de forma que se cumpla que

$$A = C \cos \phi \quad \text{y} \quad B = C \sin \phi.$$

Para que esto ocurra, ϕ debe ser

$$\frac{B}{A} = \frac{\sin \phi}{\cos \phi} = \tan \phi \quad \Rightarrow \quad \phi = \arctan\left(\frac{B}{A}\right),$$

y C debe ser

$$A^2 + B^2 = C^2(\cos^2 \phi + \sin^2 \phi) = C^2 \quad \Rightarrow \quad C = \sqrt{A^2 + B^2}.$$

De esta forma, se tiene que

$$\begin{aligned} A \cos \omega + B \sin \omega &= C \cos \phi \cos \omega + C \sin \phi \sin \omega \\ &\stackrel{(a)}{=} C \cos(\omega - \phi) \\ &= \sqrt{A^2 + B^2} \cos \left[\omega - \arctan\left(\frac{B}{A}\right) \right], \end{aligned}$$

donde en (a) se empleó la identidad trigonométrica $\cos(\alpha \pm \beta) = \cos \alpha \cos \beta \mp \sin \alpha \sin \beta$.

Se calculará a continuación el cociente $E^2(\hat{P})/\text{var}(\hat{P})$. Para hacerlo, se comienza calculando $E(\hat{P})$,

$$\begin{aligned} E(\hat{P}) &= E\left(\frac{\hat{a}_k^2 + \hat{b}_k^2}{2}\right) \\ &= \frac{E(\hat{a}_k^2) + E(\hat{b}_k^2)}{2} \\ &\stackrel{(a)}{=} \frac{\text{var}(\hat{a}_k) + E^2(\hat{a}_k) + \text{var}(\hat{b}_k) + E^2(\hat{b}_k)}{2} \\ &\stackrel{(b)}{=} \frac{\frac{2\sigma^2}{N} + a_k^2 + \frac{2\sigma^2}{N} + b_k^2}{2} \\ &= \frac{a_k^2 + b_k^2}{2} + \frac{2\sigma^2}{N} \\ &= P + \frac{2\sigma^2}{N}, \end{aligned} \tag{4.27}$$

donde en (a) se empleó que $\text{var}(x) = E(x^2) - E^2(x)$ y en (b) se empleó que los estimadores en el modelo lineal son insesgados y además que la varianza de todos los estimadores en el análisis de Fourier es $2\sigma^2/N$, dadas por la ecuación 4.10. Se calcula ahora $\text{var}(\hat{P})$,

$$\begin{aligned} \text{var}(\hat{P}) &= \text{var}\left(\frac{\hat{a}_k^2 + \hat{b}_k^2}{2}\right) \\ &\stackrel{(a)}{=} \frac{\text{var}(\hat{a}_k^2) + \text{var}(\hat{b}_k^2)}{4}, \end{aligned} \tag{4.28}$$

donde en (a) se empleó que los estimadores \hat{a}_k y \hat{b}_k son no correlacionados, ya que su matriz de covarianza es una matriz identidad escalada, como indica la ecuación 4.10. Para calcular la varianzas de \hat{a}_k^2 y \hat{b}_k^2 , se observa que si x es una variable aleatoria con PDF $\mathcal{N}(\mu, \sigma^2)$ se cumple que

$$\begin{aligned} \text{var}(x^2) &= E(x^4) - E^2(x^2) \\ &\stackrel{(a)}{=} (\mu^4 + 6\sigma^2\mu^2 + 3\sigma^4) - (\mu^2 + \sigma^2)^2 \\ &= (\mu^4 + 6\sigma^2\mu^2 + 3\sigma^4) - (\mu^4 + 2\mu^2\sigma^2 + \sigma^4) \\ &= 4\sigma^2\mu^2 + 2\sigma^4, \end{aligned}$$

donde en (a) se empleó que el momento segundo y cuarto de una variable aleatoria $x \sim \mathcal{N}(\mu, \sigma^2)$ son respectivamente⁵ $E(x^2) = \mu^2 + \sigma^2$ y $E(x^4) = \mu^4 + 6\sigma^2\mu^2 + 3\sigma^4$. Usando este resultado y teniendo en cuenta que $\hat{a}_k^2 \sim \mathcal{N}(a_k, 2\sigma^2/N)$, se cumple que

$$\text{var}(\hat{a}_k^2) = \frac{8\sigma^2 a_k^2}{N} + \frac{8\sigma^4}{N^2},$$

y lo análogo con \hat{b}_k^2 . Sustituyendo los resultados en la ecuación 4.28, se obtiene que

$$\begin{aligned} \text{var}(\hat{P}) &= \frac{\frac{8\sigma^2 a_k^2}{N} + \frac{8\sigma^4}{N^2} + \frac{8\sigma^2 b_k^2}{N} + \frac{8\sigma^4}{N^2}}{4} \\ &= \frac{2\sigma^2 a_k^2}{N} + \frac{2\sigma^2 b_k^2}{N} + \frac{4\sigma^4}{N^2} \\ &= \frac{2\sigma^2}{N} \left(a_k^2 + b_k^2 + \frac{2\sigma^2}{N} \right) \\ &= \frac{2\sigma^2}{N} \left(2P + \frac{2\sigma^2}{N} \right). \end{aligned} \tag{4.29}$$

A partir de las ecuaciones 4.27 y 4.29 se concluye que

$$\frac{E^2(\hat{P})}{\text{var}(\hat{P})} = \frac{\left(P + \frac{2\sigma^2}{N} \right)^2}{\frac{2\sigma^2}{N} \left(2P + \frac{2\sigma^2}{N} \right)}.$$

Para establecer si \hat{P} sirve como medida de detección de la presencia de la señal, se observa que ante la ausencia de la señal, $a_k = b_k = 0$ y por lo tanto $P = 0$. En ese caso, se cumple que

$$\frac{E^2(\hat{P})}{\text{var}(\hat{P})} = 1.$$

Si la señal está presente, supóngase que $P \gg 4\sigma^2/N$, lo cual en general es cierto si la cantidad de muestras N es suficientemente grande. En ese caso, se cumple que

$$\frac{E^2(\hat{P})}{\text{var}(\hat{P})} \approx \frac{P^2}{4P\sigma^2} = \frac{P}{4\sigma^2} \gg 1,$$

concluyendo que la senoide es fácilmente detectable.

4.6.7. Problema 7

Verificar que se cumple la aproximación de la ecuación 4.12 permitiendo que los límites de la sumatoria de la ecuación 4.11 vayan desde $n = -\infty$ a $n = \infty$ y considerando que $u[n] = 0$ si $n < 0$ y $n > N - 1$.

Solución La ecuación 4.11 indica que

$$[\mathbf{H}^T \mathbf{H}]_{ij} = \sum_{n=0}^{N-\max(i,j)} u[n]u[n+|i-j|],$$

y se quiere demostrar que se cumple que (ecuación 4.12)

$$[\mathbf{H}^T \mathbf{H}]_{ij} \approx \sum_{n=0}^{N-1-|i-j|} u[n]u[n+|i-j|].$$

⁵Ver por ejemplo https://en.wikipedia.org/wiki/Normal_distribution#Moments.

Esta aproximación se obtiene agregando sumandos a la ecuación 4.11 permitiendo que los límites de la sumatoria vayan desde $n = -\infty$ a $n = \infty$. Efectivamente,

$$\sum_{n=-\infty}^{\infty} u[n]u[n+|i-j|] = \sum_{n=0}^{N-1-|i-j|} u[n]u[n+|i-j|],$$

ya que $u[n] \neq 0$ desde $n = 0$ a $n = N-1$ y $u[n+|i-j|] \neq 0$ desde $n = -|i-j|$ a $n = N-1-|i-j|$.

Para verificar que la aproximación es válida, primero se observa que las ecuaciones 4.11 y 4.12 coinciden si $i = 1$ o $j = 1$. Efectivamente, supóngase que $j = 1$. En ese caso, el límite superior de la sumatoria de la ecuación 4.11 es $N - \max(i, j) = N - i$ y el límite superior de la sumatoria de la ecuación 4.12 es $N - 1 - |i - j| = N - 1 - (i - 1) = N - i$, concluyendo que ambas ecuaciones coinciden. También se observa que si $i \neq 1$ y $j \neq 1$,

$$\sum_{n=0}^{N-1-|i-j|} u[n]u[n+|i-j|] = \sum_{n=0}^{N-\max(i, j)} u[n]u[n+|i-j|] + \sum_{n=N-\max(i, j)+1}^{N-1-|i-j|} u[n]u[n+|i-j|].$$

La segunda sumatoria del lado derecho de la igualdad tiene un número de sumandos de

$$\# \text{ sumandos} = (N - 1 - |i - j|) - (N - \max(i, j) + 1) + 1 = \max(i, j) - |i - j| - 1.$$

Teniendo en cuenta que $i, j = 1, \dots, p$, esta cantidad es máxima cuando $i = j = p$,

$$\# \text{ sumandos máxima} = p - 1.$$

Por lo tanto, la ecuación 4.11 tiene como mínimo $N - p + 1$ sumandos y en la aproximación de la ecuación 4.12 se agregan a lo sumo $p - 1$ sumandos. Se concluye que si $N \gg p$, la aproximación de la ecuación 4.12 es buena.

4.6.8. Problema 8

Se desea estimar la respuesta en frecuencia $H(f)$ de un sistema lineal causal a partir de la observación de los procesos de entrada y salida. Se asume que tanto la entrada $u[n]$ y la salida $x[n]$ son procesos WSS. Demostrar que la densidad espectral de potencia cruzada entre la entrada y la salida es

$$P_{ux}(f) = H(f)P_{uu}(f),$$

donde $P_{ux}(f) = \mathcal{F}\{r_{ux}[k]\}$ y $r_{ux}[k] = E(u[n]x[n+k])$. Si el proceso de entrada es ruido blanco con PSD $P_{uu}(f) = \sigma^2$, proponer un estimador para la respuesta en frecuencia del sistema. ¿Es posible estimar la respuesta en frecuencia si $P_{uu}(f)$ es cero en una banda de frecuencias? ¿Cómo se relaciona el estimador con la ecuación 4.19 para el caso en que el sistema es un filtro FIR?

Solución La demostración que se incluye a continuación está basada en la desarrollada en el capítulo 3 de [5]. En el dominio del tiempo, la salida del sistema es

$$x[n] = (h * u)[n] = \sum_{l=0}^{\infty} h[l]u[n-l], \quad (4.30)$$

donde se empleó la hipótesis de que el sistema es causal y por lo tanto, $h[k] = 0$ si $k < 0$. Por definición, la correlación cruzada entre la entrada y la salida es

$$r_{ux}[k] = E(u[n]x[n+k]) \quad (4.31)$$

Sustituyendo la ecuación 4.30 se obtiene que

$$\begin{aligned} r_{ux}[k] &= E\left(u[n] \sum_{l=0}^{\infty} h[l]u[n+k-l]\right) \\ &= \sum_{l=0}^{\infty} h[l]E(u[n]u[n+k-l]) \\ &= \sum_{l=0}^{\infty} h[l]r_{uu}[k-l], \end{aligned}$$

es decir,

$$r_{ux}[k] = (h * r_{uu})[k].$$

Este resultado indica que si un proceso $u[n]$ WSS de entrada a un sistema lineal $h[k]$ produce la salida $x[n]$, la salida del mismo sistema cuando la entrada es la autocorrelación r_{uu} de $u[n]$, es la correlación cruzada $r_{ux}[k]$ entre la entrada y la salida, como se muestra en la figura 4.10. Aplicando al transformada

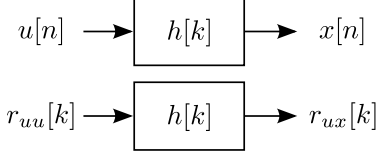


Figura 4.10: En un sistema lineal $h[k]$, si la salida es el proceso $x[n]$ cuando la entrada es el proceso $u[n]$ WSS, la salida cuando la entrada es la autocorrelación de $u[n]$ es la correlación cruzada entre la entrada y la salida.

de Fourier, se obtiene la densidad espectral de potencia cruzada entre la entrada y la salida,

$$P_{ux}(f) = H(f)P_{uu}(f).$$

Si la entrada es ruido blanco con PSD $P_{uu}(f) = \sigma^2$

$$P_{ux}(e^{j\omega}) = H(e^{j\omega})\sigma^2,$$

por lo que un estimador de la respuesta en frecuencia del filtro es

$$\hat{H}(e^{j\omega}) = \frac{P_{ux}(e^{j\omega})}{\sigma^2} = \frac{P_{ux}(e^{j\omega})}{r_{uu}[0]},$$

ya que $r_{uu}[k] = \sigma^2\delta[k]$ y por lo tanto, $r_{uu}[0] = \sigma^2$. Aplicando la antitransformada de Fourier, se obtiene la expresión en el dominio del tiempo, que es

$$\hat{\mathbf{h}} = \frac{\mathbf{r}_{ux}}{r_{uu}[0]},$$

coincidiendo con la ecuación 4.19.

Si $P_{uu}(f) = 0$ en una banda de frecuencias, $P_{ux}(f) = 0$ en la misma banda de frecuencias independientemente de $H(f)$, y por lo tanto, no se dispone de información sobre $H(f)$ en esa banda de frecuencias.

4.6.9. Problema 9

Derivar las ecuaciones 4.20 y 4.21 del estimador MVU y la covarianza del estimador de la extensión del modelo lineal de forma análoga a la realizada en la sección 4.2.

Solución Los datos observados pueden modelarse como

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w},$$

donde $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$. El vector \mathbf{x} de observaciones es de dimensiones $N \times 1$, la matriz \mathbf{H} conocida es de dimensiones $N \times p$ con $N > p$ y rango p , el vector de parámetros a estimar $\boldsymbol{\theta}$ tiene dimensiones $p \times 1$, \mathbf{w} es ruido coloreado de dimensiones $N \times 1$ con matriz de covarianza \mathbf{C} de dimensiones $N \times N$.

Se verá si es posible factorizar la función de verosimilitud logarítmica como

$$\frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{I}(\boldsymbol{\theta})(\mathbf{g}(\mathbf{x}) - \boldsymbol{\theta}).$$

En caso afirmativo, el estimador $\hat{\boldsymbol{\theta}} = \mathbf{g}(\mathbf{x})$ es MVU y eficiente. La PDF de los datos es

$$p(\mathbf{x}; \boldsymbol{\theta}) = \frac{1}{(2\pi)^{\frac{N}{2}} \det^{\frac{1}{2}}(\mathbf{C})} \exp \left[-\frac{1}{2}(\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T \mathbf{C}^{-1}(\mathbf{x} - \mathbf{H}\boldsymbol{\theta}) \right],$$

y al tomar logaritmo se tiene que

$$\ln p(\mathbf{x}; \boldsymbol{\theta}) = -\frac{N}{2} \ln(2\pi) - \frac{1}{2} \det(\mathbf{C}) - \frac{1}{2} (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T \mathbf{C}^{-1} (\mathbf{x} - \mathbf{H}\boldsymbol{\theta}).$$

Luego, derivando respecto al parámetro, se obtiene que

$$\begin{aligned} \frac{\partial p(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= -\frac{1}{2} \frac{\partial}{\partial \boldsymbol{\theta}} [(\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T \mathbf{C}^{-1} (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})] \\ &= -\frac{1}{2} \frac{\partial}{\partial \boldsymbol{\theta}} [(\mathbf{x}^T - \boldsymbol{\theta}^T \mathbf{H}^T) \mathbf{C}^{-1} (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})] \\ &= -\frac{1}{2} \frac{\partial}{\partial \boldsymbol{\theta}} [\mathbf{x}^T \mathbf{C}^{-1} \mathbf{x} - \mathbf{x}^T \mathbf{C}^{-1} \mathbf{H}\boldsymbol{\theta} - \boldsymbol{\theta}^T \mathbf{H}^T \mathbf{C}^{-1} \mathbf{x} + \boldsymbol{\theta}^T \mathbf{H}^T \mathbf{C}^{-1} \mathbf{H}\boldsymbol{\theta}] \\ &= -\frac{1}{2} \frac{\partial}{\partial \boldsymbol{\theta}} [\mathbf{x}^T \mathbf{C}^{-1} \mathbf{x} - 2\mathbf{x}^T \mathbf{C}^{-1} \mathbf{H}\boldsymbol{\theta} + \boldsymbol{\theta}^T \mathbf{H}^T \mathbf{C}^{-1} \mathbf{H}\boldsymbol{\theta}]. \end{aligned}$$

Para calcular la derivada, se procede de forma análoga a la realizada en la sección 4.2 definiendo el vector $\mathbf{b}^T = \mathbf{x}^T \mathbf{C}^{-1} \mathbf{H}$ y la matriz $\mathbf{A} = \mathbf{H}^T \mathbf{C}^{-1} \mathbf{H}$ simétrica, y teniendo en cuenta que se cumple que

$$\frac{\partial \mathbf{b}^T \boldsymbol{\theta}}{\partial \boldsymbol{\theta}} = \mathbf{b}, \quad \frac{\partial \boldsymbol{\theta}^T \mathbf{A} \boldsymbol{\theta}}{\partial \boldsymbol{\theta}} = 2\mathbf{A}\boldsymbol{\theta},$$

como se explica en el apéndice D, se obtiene que

$$\frac{\partial p(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = -\frac{1}{2} [-2\mathbf{H}^T \mathbf{C}^{-1} \mathbf{x} + 2\mathbf{H}^T \mathbf{C}^{-1} \mathbf{H}\boldsymbol{\theta}] = [\mathbf{H}^T \mathbf{C}^{-1} \mathbf{x} - \mathbf{H}^T \mathbf{C}^{-1} \mathbf{H}\boldsymbol{\theta}],$$

y asumiendo que $\mathbf{H}^T \mathbf{C}^{-1} \mathbf{H}$ es invertible, se llega a que

$$\frac{\partial p(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{H}^T \mathbf{C}^{-1} \mathbf{H} [(\mathbf{H}^T \mathbf{C}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C}^{-1} \mathbf{x} - \boldsymbol{\theta}],$$

ecuación que está factorizada en la forma de la ecuación 3.12 con

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= (\mathbf{H}^T \mathbf{C}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C}^{-1} \mathbf{x} \\ \mathbf{I}(\boldsymbol{\theta}) &= \mathbf{H}^T \mathbf{C}^{-1} \mathbf{H}. \end{aligned}$$

La matriz de covarianza del estimador es

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}} = \mathbf{I}^{-1}(\boldsymbol{\theta}) = (\mathbf{H}^T \mathbf{C}^{-1} \mathbf{H})^{-1}.$$

Además, el estimador es eficiente y alcanza la CRLB. Observar que el resultado coincide con el obtenido mediante el enfoque de blanqueado de la sección 4.5.

4.6.10. Problema 10

Si en el ejemplo de la sección 4.5.1 de la estimación del nivel de DC en ruido colorado las muestras de ruido son no correlacionadas pero con varianza distinta,

$$\mathbf{C} = \text{diag}(\sigma_0^2, \dots, \sigma_{N-1}^2),$$

calcular d_n e interpretar los resultados. ¿Qué pasaría con \hat{A} si una única varianza σ_n^2 fuera cero?

Solución Para calcular d_n (ecuación 4.23), hay que factorizar \mathbf{C}^{-1} como $\mathbf{D}^T \mathbf{D}$. Como la matriz de covarianza es diagonal, se cumple que

$$\mathbf{C}^{-1} = \text{diag}\left(\frac{1}{\sigma_0^2}, \dots, \frac{1}{\sigma_{N-1}^2}\right)$$

y por lo tanto

$$\mathbf{D} = \text{diag}\left(\frac{1}{\sigma_0}, \dots, \frac{1}{\sigma_{N-1}}\right).$$

Además,

$$\mathbf{D}\mathbf{1} = \left[\frac{1}{\sigma_0} \frac{1}{\sigma_1} \dots \frac{1}{\sigma_{N-1}} \right]^T$$

por lo que $[\mathbf{D}\mathbf{1}]_n = 1/\sigma_n$. Sustituyendo este resultado en la ecuación 4.23, se obtiene que

$$d_n = \frac{\frac{1}{\sigma_n}}{\sum_{n=0}^{N-1} \frac{1}{\sigma_n^2}}.$$

Los datos blanqueados $\mathbf{x}' = \mathbf{D}\mathbf{x}$ quedan

$$x'[n] = \frac{x[n]}{\sigma_n},$$

y se cumple que

$$\text{var}(x'[n]) = \frac{1}{\sigma_n^2} \text{var}(x[n]) = \frac{\sigma_n^2}{\sigma_n^2} = 1.$$

Como en este caso las muestras son no correlacionadas, el blanqueado lo único que hace es igualar las varianzas de todas las muestras al valor uno. Finalmente, sustituyendo estos resultados en la ecuación 4.22, el estimador MVU queda

$$\hat{A} = \frac{\sum_{n=0}^{N-1} \frac{1}{\sigma_n} x'[n]}{\sum_{n=0}^{N-1} \frac{1}{\sigma_n^2}} = \frac{\sum_{n=0}^{N-1} \frac{1}{\sigma_n^2} x[n]}{\sum_{n=0}^{N-1} \frac{1}{\sigma_n^2}}.$$

Observar que el estimador le da mas importancia, a través del peso, a las muestras de menor varianza.

Considérese el caso en que la varianza de alguna muestra es nula, digamos, la de la muestra m -ésima. Como $\sigma_m^2 \rightarrow 0$, el estimador es

$$\hat{A} = \frac{\sum_{n=0}^{N-1} \frac{1}{\sigma_n^2} x[n]}{\sum_{n=0}^{N-1} \frac{1}{\sigma_n^2}} \rightarrow \frac{\frac{1}{\sigma_m^2} x[m]}{\frac{1}{\sigma_m^2}} = x[m],$$

como es de esperarse, ya que la muestra $x[m]$, al tener ruido de varianza nula, no tiene error y por lo tanto, toma el valor del nivel de DC exacto.

4.6.11. Problema 11

Considérese el modelo de datos descrito en el problema de la sección 3.6.9. Encontrar el estimador MVU de A y su varianza. ¿Porque el estimador es independiente de ρ ? ¿Qué ocurre cuando $\rho \rightarrow \pm 1$?

Solución El modelo de los datos en notación matricial es

$$\mathbf{x} = \mathbf{H}A + \mathbf{w}, \quad \mathbf{H} = [1 \ 1]^T,$$

y el ruido \mathbf{w} es coloreado con matriz de covarianza

$$\mathbf{C} = \sigma^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}.$$

Por tratarse del problema de estimación del nivel de DC en ruido coloreado, el estimador y su varianza fueron calculados en el ejemplo de la sección 4.5.1, y son

$$\hat{A} = \frac{\mathbf{1}^T \mathbf{C}^{-1} \mathbf{x}}{\mathbf{1}^T \mathbf{C}^{-1} \mathbf{1}}, \quad \text{var}(\hat{A}) = \frac{1}{\mathbf{1}^T \mathbf{C}^{-1} \mathbf{1}}.$$

La matriz de covarianza inversa es

$$\mathbf{C}^{-1} = \frac{1}{\sigma^2(1-\rho^2)} \begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix},$$

y por lo tanto

$$\begin{aligned} \mathbf{1}^T \mathbf{C}^{-1} \mathbf{x} &= \begin{bmatrix} 1 & 1 \end{bmatrix} \frac{1}{\sigma^2(1-\rho^2)} \begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix} \begin{bmatrix} x[0] \\ x[1] \end{bmatrix} \\ &= \frac{1}{\sigma^2(1-\rho^2)} \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} x[0] - \rho x[1] \\ -\rho x[0] + x[1] \end{bmatrix} \\ &= \frac{x[0] - \rho x[1] - \rho x[0] + x[1]}{\sigma^2(1-\rho^2)} \\ &= \frac{(1-\rho)(x[0] + x[1])}{\sigma^2(1-\rho^2)}. \end{aligned}$$

Sustituyendo en el resultado $x[0]$ y $x[1]$ por 1, se tiene que

$$\mathbf{1}^T \mathbf{C}^{-1} \mathbf{1} = \frac{2(1-\rho)}{\sigma^2(1-\rho^2)}.$$

Por lo tanto,

$$\hat{A} = \frac{(1-\rho)(x[0] + x[1])}{2(1-\rho)} = \frac{x[0] + x[1]}{2} = \bar{\mathbf{x}},$$

y

$$\text{var}(\hat{A}) = \frac{\sigma^2(1-\rho^2)}{2(1-\rho)} = \frac{\sigma^2}{2}(1+\rho).$$

Para determinar porque el estimador no depende de ρ , se observa que \mathbf{H} es un vector propio de la matriz de covarianza \mathbf{C} . Efectivamente,

$$\mathbf{C}\mathbf{H} = \sigma^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1+\rho \\ 1+\rho \end{bmatrix} = \sigma^2(1+\rho) \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \sigma^2(1+\rho)\mathbf{H},$$

y por lo tanto se cumple que $\mathbf{C}\mathbf{H} = \lambda\mathbf{H}$, donde el valor propio correspondiente al vector propio \mathbf{H} es $\lambda = \sigma^2(1+\rho)$. De esta forma, \mathbf{H} también es un vector propio de \mathbf{C}^{-1} con valor propio $1/\lambda$, es decir, se cumple que $\mathbf{C}^{-1}\mathbf{H} = (1/\lambda)\mathbf{H}$ y el estimador se puede expresar como (ecuación 4.20)

$$\begin{aligned} \hat{A} &= (\mathbf{H}^T \mathbf{C}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C}^{-1} \mathbf{x} \\ &= (\mathbf{H}^T \mathbf{C}^{-1} \mathbf{H})^{-1} (\mathbf{C}^{-1} \mathbf{H})^T \mathbf{x} \\ &= \left(\mathbf{H}^T \frac{1}{\lambda} \mathbf{H} \right)^{-1} \left(\frac{1}{\lambda} \mathbf{H} \right)^T \mathbf{x} \\ &= \lambda (\mathbf{H}^T \mathbf{H})^{-1} \frac{1}{\lambda} \mathbf{H}^T \mathbf{x} \\ &= (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x}, \end{aligned}$$

que no depende de la matriz de covarianza \mathbf{C} y por lo tanto, tampoco del coeficiente de correlación ρ . En el caso de estimación del nivel de DC en ruido coloreado, esto ocurre siempre que el vector $\mathbf{1}$ sea vector propio de la matriz de covarianza \mathbf{C} .

En el problema de la sección 3.6.9 se explica qué ocurre cuando $\rho \rightarrow \pm 1$.

4.6.12. Problema 12

En este problema se investiga la estimación de una función lineal de $\boldsymbol{\theta}$. Siendo este nuevo parámetro $\boldsymbol{\alpha} = \mathbf{A}\boldsymbol{\theta}$, donde \mathbf{A} es una matriz conocida $r \times p$ con $r < p$ y rango r , mostrar que el estimador MVU es

$$\hat{\boldsymbol{\alpha}} = \mathbf{A}\hat{\boldsymbol{\theta}},$$

donde $\hat{\boldsymbol{\theta}}$ es el estimador MVU de $\boldsymbol{\theta}$. Además, calcular la matriz de covarianza. Sugerencia: reemplazar \mathbf{x} por

$$\mathbf{x}' = \mathbf{A}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x},$$

donde \mathbf{x}' tiene tamaño $r \times 1$. Esto resulta en el *modelo lineal reducido*. Se puede mostrar que \mathbf{x}' contiene toda la información sobre $\boldsymbol{\theta}$, por lo que es posible basar el estimador en este conjunto de datos de menor dimensión.

Solución En el modelo lineal, el estimador MVU es (ecuación 4.3)

$$\hat{\boldsymbol{\theta}} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x},$$

por lo que el estimador de $\boldsymbol{\alpha}$ es

$$\hat{\boldsymbol{\alpha}} = \mathbf{A} \hat{\boldsymbol{\theta}} = \mathbf{A}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x}.$$

Una forma de verificar que $\hat{\boldsymbol{\alpha}}$ es el estimador MVU de $\boldsymbol{\alpha}$, es multiplicar el modelo lineal $\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$ por $\mathbf{A}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T$. Esto conduce a

$$\begin{aligned} \mathbf{A}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x} &= \mathbf{A}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{H} \boldsymbol{\theta} + \mathbf{A}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{w} \\ &= \mathbf{A} \boldsymbol{\theta} + \mathbf{A}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{w} \\ &= \boldsymbol{\alpha} + \mathbf{A}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{w}, \end{aligned}$$

que es un nuevo modelo lineal,

$$\mathbf{x}' = \mathbf{H}' \boldsymbol{\alpha} + \mathbf{w}',$$

donde $\mathbf{x}' = \mathbf{A}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x}$ y $\mathbf{w}' = \mathbf{A}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{w}$ tienen tamaño $r \times 1$, y $\mathbf{H}' = \mathbf{I}$ de tamaño $r \times r$. En este modelo lineal, el ruido \mathbf{w}' es coloreado con matriz de covarianza

$$\begin{aligned} \mathbf{C}_{\mathbf{w}'} &= E(\mathbf{w}' \mathbf{w}'^T) \\ &= E \left[\mathbf{A}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{w} \mathbf{w}^T \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{A}^T \right] \\ &\stackrel{(a)}{=} \mathbf{A}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T E(\mathbf{w} \mathbf{w}^T) \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{A}^T \\ &\stackrel{(b)}{=} \sigma^2 \mathbf{A}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{A}^T, \end{aligned}$$

donde en (a) se tuvo en cuenta que $(\mathbf{H}^T \mathbf{H})^{-1}$ es una matriz simétrica y en (b) que el ruido en el modelo lineal convencional es blanco, y por lo tanto $E(\mathbf{w} \mathbf{w}^T) = \mathbf{C}_w = \sigma^2 \mathbf{I}$. Finalmente, eliminando una matriz y su inversa se obtiene que

$$\mathbf{C}_{\mathbf{w}'} = \sigma^2 \mathbf{A}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{A}^T.$$

El estimador MVU está dado por la ecuación 4.20 del modelo lineal extendido, y es

$$\begin{aligned} \hat{\boldsymbol{\alpha}} &= (\mathbf{H}'^T \mathbf{C}_{\mathbf{w}'}^{-1} \mathbf{H}')^{-1} \mathbf{H}'^T \mathbf{C}_{\mathbf{w}'}^{-1} \mathbf{x}' \\ &\stackrel{(a)}{=} \mathbf{x}' \\ &= \mathbf{A}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x}, \end{aligned}$$

donde en (a) se tuvo en cuenta que $\mathbf{H}' = \mathbf{I}$. Esto es lo que se quería demostrar. La matriz de covarianza del estimador está dada por la ecuación 4.21, y es

$$\begin{aligned} \mathbf{C}_{\hat{\boldsymbol{\alpha}}} &= (\mathbf{H}'^T \mathbf{C}_{\mathbf{w}'}^{-1} \mathbf{H}')^{-1} \\ &= \mathbf{C}_{\mathbf{w}'} \\ &= \sigma^2 \mathbf{A}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{A}^T. \end{aligned}$$

4.6.13. Problema 13

En la práctica a veces se encuentra el “modelo lineal” $\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$ pero con la matriz \mathbf{H} compuesta de variables aleatorias. Supóngase que se ignora esta diferencia y se emplea el estimador usual

$$\hat{\boldsymbol{\theta}} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x},$$

donde se asume que la realización particular de \mathbf{H} es conocida. Mostrar que si \mathbf{H} y \mathbf{w} son independientes, la media y la covarianza de $\hat{\boldsymbol{\theta}}$ son

$$\begin{aligned} E(\hat{\boldsymbol{\theta}}) &= \boldsymbol{\theta} \\ \mathbf{C}_{\hat{\boldsymbol{\theta}}} &= \sigma^2 E_H [(\mathbf{H}^T \mathbf{H})^{-1}], \end{aligned}$$

donde E_H denota la esperanza respecto a la PDF de \mathbf{H} . ¿Qué sucede si no se asume la hipótesis de independencia?

Solución Para calcular la esperanza, se observa que el estimador se puede escribir como

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x} \\ &= (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T (\mathbf{H}\boldsymbol{\theta} + \mathbf{w}) \\ &= \boldsymbol{\theta} + (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{w}, \end{aligned} \tag{4.32}$$

y por lo tanto, la esperanza es

$$\begin{aligned} E(\hat{\boldsymbol{\theta}}) &= \boldsymbol{\theta} + E[(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{w}] \\ &\stackrel{(a)}{=} \boldsymbol{\theta} + E_H [(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T] E(\mathbf{w}) \\ &\stackrel{(b)}{=} \boldsymbol{\theta}, \end{aligned}$$

donde en (a) se empleó que \mathbf{H} y \mathbf{w} son variables aleatorias independientes y en (b) que $E(\mathbf{w}) = 0$ por ser ruido blanco. Se concluye que el estimador habitual es insesgado aunque los elementos de \mathbf{H} sean variables aleatorias. La covarianza del estimador es

$$\begin{aligned} \mathbf{C}_{\hat{\boldsymbol{\theta}}} &= E[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T] \\ &\stackrel{(a)}{=} E[(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{w} \mathbf{w}^T \mathbf{H} (\mathbf{H}^T \mathbf{H})^{-1}] \\ &\stackrel{(b)}{=} E_H E[(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{w} \mathbf{w}^T \mathbf{H} (\mathbf{H}^T \mathbf{H})^{-1}] \\ &\stackrel{(c)}{=} E_H [(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T E(\mathbf{w} \mathbf{w}^T) \mathbf{H} (\mathbf{H}^T \mathbf{H})^{-1}] \\ &\stackrel{(d)}{=} E_H [(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T (\sigma^2 \mathbf{I}) \mathbf{H} (\mathbf{H}^T \mathbf{H})^{-1}] \\ &= \sigma^2 E_H [(\mathbf{H}^T \mathbf{H})^{-1}], \end{aligned}$$

donde en (a) se empleó la identidad de la ecuación 4.32 y en las identidades (b), (c) y (d) se consideró la independencia de \mathbf{H} y \mathbf{w} . Para aclarar la notación utilizada, se realiza a continuación el mismo cálculo con las integrales explícitas de la esperanza,

$$\begin{aligned} E[(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{w} \mathbf{w}^T \mathbf{H} (\mathbf{H}^T \mathbf{H})^{-1}] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{w} \mathbf{w}^T \mathbf{H} (\mathbf{H}^T \mathbf{H})^{-1} p(\mathbf{H}, \mathbf{w}) d\mathbf{H} d\mathbf{w} \\ &\stackrel{(a)}{=} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{w} \mathbf{w}^T \mathbf{H} (\mathbf{H}^T \mathbf{H})^{-1} p(\mathbf{H}) p(\mathbf{w}) d\mathbf{H} d\mathbf{w} \\ &= \int_{-\infty}^{\infty} (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \left(\int_{-\infty}^{\infty} \mathbf{w} \mathbf{w}^T p(\mathbf{w}) d\mathbf{w} \right) \mathbf{H} (\mathbf{H}^T \mathbf{H})^{-1} p(\mathbf{H}) d\mathbf{H} \\ &\stackrel{(b)}{=} \int_{-\infty}^{\infty} (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T (\sigma^2 \mathbf{I}) \mathbf{H} (\mathbf{H}^T \mathbf{H})^{-1} p(\mathbf{H}) d\mathbf{H} \\ &= \sigma^2 \int_{-\infty}^{\infty} (\mathbf{H}^T \mathbf{H})^{-1} p(\mathbf{H}) d\mathbf{H} \\ &= \sigma^2 E_H [(\mathbf{H}^T \mathbf{H})^{-1}], \end{aligned}$$

donde en (a) se consideró que \mathbf{H} y \mathbf{w} son independientes y por lo tanto, la PDF conjunta es el producto de las PFD, $p(\mathbf{H}, \mathbf{w}) = p(\mathbf{H})p(\mathbf{w})$ y en (b) se tuvo en cuenta que la integral entre paréntesis es $E(\mathbf{w}\mathbf{w}^T) = \mathbf{C} = \sigma^2\mathbf{I}$.

En el caso en que \mathbf{H} y \mathbf{w} no son independientes, se cumple que

$$\begin{aligned} E(\hat{\boldsymbol{\theta}}) &= \boldsymbol{\theta} + E[(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{w}] \\ \mathbf{C}_{\hat{\boldsymbol{\theta}}} &= E[(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{w}\mathbf{w}^T\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}]. \end{aligned}$$

Como el término $E[(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{w}]$ en la esperanza del estimador no necesariamente es nulo, el estimador podría no ser insesgado.

4.6.14. Problema 14

Supóngase que se observa una señal en ruido que está sujeta a desvanecerse. Se modela el proceso de desvanecimiento como que la señal está presente o no está presente. Como ilustración sencilla, se considera el caso de la estimación del nivel de DC en WGN, o $x[n] = A + w[n]$, $n = 0, \dots, N-1$. Cuando la señal se desvanece, el modelo de los datos es

$$x[n] = \begin{cases} A + w[n] & n = 0, \dots, M-1 \\ w[n] & n = M, \dots, N-1 \end{cases}$$

donde la probabilidad de desvanecimiento es ϵ . Asumiendo que se conoce cuando la señal experimentó desvanecimiento, usar los resultados del problema anterior para determinar un estimador de A y también su varianza. Comparar los resultados en el caso en que no se produce desvanecimiento.

Solución Asíumase que se observa una realización en donde se produce desvanecimiento en la muestra M . Como se mostró en el problema anterior, si el proceso de desvanecimiento y el ruido son independientes, el estimador usual del modelo lineal es insesgado. Por lo tanto, el estimador es

$$\hat{A} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x}, \quad \text{con} \quad \mathbf{H} = \begin{bmatrix} 1 & \dots & 1 & 0 & \dots & 0 \\ 0 & \dots & 0 & 1 & \dots & 1 \end{bmatrix}$$

es decir, \mathbf{H} es un vector con M unos seguidos de $N-M$ ceros. De esta forma,

$$\hat{A} = \frac{1}{M} \sum_{n=0}^{M-1} x[n].$$

Del problema anterior se tiene que la varianza es

$$\begin{aligned} \text{var}(\hat{A}) &= \sigma^2 E_H[(\mathbf{H}^T\mathbf{H})^{-1}] \\ &\stackrel{(a)}{=} \sigma^2 \left[\frac{1}{N}(1-\epsilon) + \frac{1}{M}\epsilon \right] \\ &= \frac{\sigma^2}{N} \left[1 - \epsilon + \frac{N}{M}\epsilon \right] \\ &= \frac{\sigma^2}{N} \left[1 + \left(\frac{N}{M} - 1 \right) \epsilon \right] \\ &\stackrel{(b)}{\geq} \frac{\sigma^2}{N}. \end{aligned}$$

donde en (a) la esperanza es el valor que toma $(\mathbf{H}^T\mathbf{H})^{-1}$ en el caso en que no hay desvanecimiento por la probabilidad de que no haya desvanecimiento sumado al valor que toma cuando hay desvanecimiento por la probabilidad de que haya desvanecimiento y la desigualdad en (b) proviene de que $N \geq M$. Notar que la igualdad se da en el caso en que $M = N$ o que $\epsilon = 0$, que es que no hay desvanecimiento.

Capítulo 5

Estimación insesgada de varianza mínima general

5.1. Factorización de Neyman-Fisher

El teorema de la factorización de Neyman-Fisher es una herramienta poderosa para encontrar estadísticos suficientes de un parámetro.

5.1.1. Teorema

Si la PDF $p(\mathbf{x}; \theta)$ de los datos puede ser factorizada como

$$p(\mathbf{x}; \theta) = g(T(\mathbf{x}), \theta)h(\mathbf{x}), \quad (5.1)$$

donde g es una función que depende de \mathbf{x} solo a través de $T(\mathbf{x})$ y h es una función que depende solo de los datos \mathbf{x} , se cumple que $T(\mathbf{x})$ es un estadístico suficiente de θ . Recíprocamente, si $T(\mathbf{x})$ es un estadístico suficiente de θ , la PDF de los datos puede ser factorizada como en la ecuación 5.1.

Demostración Para la demostración, se realizarán primero dos observaciones:

- Considérese la PDF conjunta $p(\mathbf{x}, T(\mathbf{x}); \theta)$. Para evaluar esta PDF, debe tenerse en cuenta que $T(\mathbf{x})$ es funcionalmente dependiente de \mathbf{x} . Por lo tanto, la PDF conjunta debe ser nula en todo $\mathbf{x} = \mathbf{x}_0$, $T(\mathbf{x}) = T_0$ a menos que $T(\mathbf{x}_0) = T_0$. Esto implica que

$$p(\mathbf{x}, T(\mathbf{x}); \theta) = p(\mathbf{x}; \theta)\delta(T(\mathbf{x}) - T_0). \quad (5.2)$$

Una analogía es por ejemplo el caso en que una variable aleatoria x escalar toma siempre el mismo valor, digamos c . En ese caso, x tiene una distribución de probabilidad degenerada¹

$$F_c(x) = \begin{cases} 0, & x < c \\ 1, & x \geq c \end{cases},$$

por lo que la PDF es $p(x) = \delta(x - c)$. Otra analogía mas cercana a la situación de interés es cuando se tienen dos variables aleatorias escalares x y y tales que $x = y$ siempre. La PDF conjunta es $p(x, y) = p(x)\delta(y - x)$, una PDF degenerada de dos variables.

- Una forma de representar la PDF de una función de varias variables es la siguiente: si $y = g(\mathbf{x})$, donde \mathbf{x} es una variable aleatoria vectorial, la PDF puede expresarse como

$$p(y) = \int p(\mathbf{x})\delta(y - g(\mathbf{x})) d\mathbf{x}. \quad (5.3)$$

¹Ver: https://en.wikipedia.org/wiki/Degenerate_distribution

Se mostrará para el caso escalar que esto es equivalente a la ecuación usual de la PDF de una variable aleatoria transformada. Si x es una variable aleatoria escalar, la PDF de la variable transformada $y = g(x)$ es (ver el capítulo 5 de [2])

$$p_y(y) = \sum_{i=1}^k \frac{p_x(x_i)}{|g'(x_i)|}, \quad (5.4)$$

donde $\{x_1, x_2, \dots, x_k\}$ son las soluciones de $y = g(x)$, es decir, $y = g(x_1) = g(x_2) = \dots = g(x_k)$. Para demostrar la equivalencia con la ecuación 5.3, se verá primero que

$$\delta(y - g(x)) = \sum_{i=1}^k \frac{\delta(x - x_i)}{|g'(x_i)|}. \quad (5.5)$$

Un resultado mas general, es que

$$\int_{-\infty}^{\infty} \delta(g(x)) f(x) dx = \sum_{i=1}^k \frac{f(x_i)}{|g'(x_i)|}.$$

Efectivamente²,

$$\begin{aligned} \int_{-\infty}^{\infty} \delta(g(x)) f(x) dx &\stackrel{(a)}{=} \sum_{i=1}^k \int_{x_i - \epsilon_i}^{x_i + \epsilon_i} \delta(g(x)) f(x) dx \\ &\stackrel{(b)}{=} \sum_{i=1}^k \int_{g(x_i - \epsilon_i)}^{g(x_i + \epsilon_i)} \delta(u) f(g^{-1}(u)) \frac{1}{g'(g^{-1}(u))} du \\ &\stackrel{(c)}{=} \sum_{i=1}^k \frac{f(x_i)}{g'(x_i)} \int_{g(x_i - \epsilon_i)}^{g(x_i + \epsilon_i)} \delta(u) du \\ &\stackrel{(d)}{=} \sum_{i=1}^k \frac{f(x_i)}{|g'(x_i)|}, \end{aligned}$$

donde en (a) los x_i son las raíces de g y se consideró que $\delta(g(x)) = 0$ excepto cuando $g(x) = 0$ y por lo tanto, es equivalente considerar el dominio de integración solo en torno a los puntos x_i donde $g(x_i) = 0$, en (b) se realizó el cambio de variable $u = g(x)$ y por lo tanto $x = g^{-1}(u)$ y $dx = [g^{-1}(u)]' du = du/g'(g^{-1}(u))$, donde en la última igualdad se empleó el teorema de la función inversa³, en (c) se evaluó la función que multiplica a la delta en $u = g(x_i)$, que es el único punto en donde $\delta(u)$ es distinto de cero en el intervalo de integración, y en (d) se consideró que si $g'(x_i) < 0$, $g(x_i - \epsilon_i) > g(x_i + \epsilon_i)$, y hay que invertir los límites de integración y cambiar el signo de la integral. Además, se observa que

$$\begin{aligned} \sum_{i=1}^k \frac{f(x_i)}{|g'(x_i)|} &= \sum_{i=1}^k \int_{-\infty}^{\infty} \frac{f(x) \delta(x - x_i)}{|g'(x_i)|} dx \\ &= \int_{-\infty}^{\infty} \left(\sum_{i=1}^k \frac{\delta(x - x_i)}{|g'(x_i)|} \right) f(x) dx, \end{aligned}$$

concluyendo que

$$\delta(g(x)) = \sum_{i=1}^k \frac{\delta(x - x_i)}{|g'(x_i)|},$$

²Esta prueba está basada en <https://math.stackexchange.com/questions/276583/dirac-delta-function-of-a-function>.

³Ver https://en.wikipedia.org/wiki/Inverse_function_theorem, por ejemplo.

que es la ecuación 5.5. Con esto, es posible mostrar que la ecuación 5.3 es equivalente a la ecuación de la PDF de una variable aleatoria transformada. Basta sustituir la ecuación 5.5 en la ecuación 5.3, resultando en

$$\begin{aligned}
p(y) &= \int_{-\infty}^{\infty} p(x) \delta(y - g(x)) dx \\
&= \int_{-\infty}^{\infty} p(x) \sum_{i=1}^k \frac{\delta(x - x_i)}{|g'(x_i)|} dx \\
&= \sum_{i=1}^k \frac{p(x_i)}{|g'(x_i)|} \int_{-\infty}^{\infty} \delta(x - x_i) dx \\
&= \sum_{i=1}^k \frac{p(x_i)}{|g'(x_i)|},
\end{aligned}$$

que es lo que se quería mostrar.

Se comenzará la demostración del teorema de la factorización de Neyman-Fisher probando que $T(\mathbf{x})$ es un estadístico suficiente si la factorización puede realizarse. Hay que demostrar que si la factorización se cumple, la probabilidad condicional de los datos dado el estadístico no depende del parámetro. Por el teorema de la probabilidad total, la probabilidad condicional es

$$\begin{aligned}
p(\mathbf{x} | T(\mathbf{x}) = T_0; \theta) &= \frac{p(\mathbf{x}, T(\mathbf{x}) = T_0; \theta)}{p(T(\mathbf{x}) = T_0; \theta)} \\
&= \frac{p(\mathbf{x}; \theta) \delta(T(\mathbf{x}) - T_0)}{p(T(\mathbf{x}) = T_0; \theta)},
\end{aligned}$$

donde en la segunda igualdad se empleó la observación realizada arriba. Aplicando la factorización (ecuación 5.1), que se asume que se cumple por hipótesis, se tiene que

$$\begin{aligned}
p(\mathbf{x} | T(\mathbf{x}) = T_0; \theta) &= \frac{g(T(\mathbf{x}), \theta) h(\mathbf{x}) \delta(T(\mathbf{x}) - T_0)}{p(T(\mathbf{x}) = T_0; \theta)} \\
&= \frac{g(T(\mathbf{x}) = T_0, \theta) h(\mathbf{x}) \delta(T(\mathbf{x}) - T_0)}{p(T(\mathbf{x}) = T_0; \theta)},
\end{aligned}$$

donde en la segunda igualdad se aplicó la propiedad de la delta que indica que $g(x) \delta(x - x_i) = g(x_i) \delta(x - x_i)$. Además, el denominador es

$$\begin{aligned}
p(T(\mathbf{x}) = T_0; \theta) &\stackrel{(a)}{=} \int p(\mathbf{x}; \theta) \delta(T(\mathbf{x}) - T_0) d\mathbf{x} \\
&\stackrel{(b)}{=} \int g(T(\mathbf{x}) = T_0, \theta) h(\mathbf{x}) \delta(T(\mathbf{x}) - T_0) d\mathbf{x} \\
&\stackrel{(c)}{=} g(T(\mathbf{x}) = T_0, \theta) \int h(\mathbf{x}) \delta(T(\mathbf{x}) - T_0) d\mathbf{x},
\end{aligned}$$

donde en (a) se empleó la ecuación 5.3, en (b) se aplicó nuevamente la factorización de la ecuación 5.1 y en (c) se consideró que la integral es cero excepto en la superficie en \mathbb{R}^N donde $T(\mathbf{x}) = T_0$, y g es constante en dicha superficie. Sustituyendo el resultado en la ecuación de la esperanza condicional, se obtiene que

$$p(\mathbf{x} | T(\mathbf{x}) = T_0; \theta) = \frac{h(\mathbf{x}) \delta(T(\mathbf{x}) - T_0)}{\int h(\mathbf{x}) \delta(T(\mathbf{x}) - T_0) d\mathbf{x}},$$

que no depende de θ , como se quería demostrar. Se concluye que $T(\mathbf{x})$ es un estadístico suficiente.

Se probará ahora el recíproco: si $T(\mathbf{x})$ es un estadístico suficiente, la factorización puede realizarse. La PDF conjunta es

$$p(\mathbf{x}, T(\mathbf{x}) = T_0; \theta) = p(\mathbf{x} | T(\mathbf{x}) = T_0; \theta) p(T(\mathbf{x}) = T_0; \theta). \quad (5.6)$$

Como se observó previamente, el lado derecho de la igualdad puede expresarse como

$$p(\mathbf{x}, T(\mathbf{x}) = T_0; \theta) = p(\mathbf{x}; \theta) \delta(T(\mathbf{x}) - T_0).$$

Además, como $T(\mathbf{x})$ es un estadístico suficiente, la PDF condicional no depende del parámetro θ y por lo tanto es $p(\mathbf{x} | T(\mathbf{x}) = T_0)$. Sustituyendo estos resultados en la ecuación 5.6, se tiene que

$$p(\mathbf{x}; \theta) \delta(T(\mathbf{x}) - T_0) = p(\mathbf{x} | T(\mathbf{x}) = T_0) p(T(\mathbf{x}) = T_0; \theta). \quad (5.7)$$

La PDF condicional es nula excepto en la superficie de \mathbb{R}^N donde $T(\mathbf{x}) = T_0$, y por lo tanto,

$$p(\mathbf{x} | T(\mathbf{x}) = T_0) = w(\mathbf{x}) \delta(T(\mathbf{x}) - T_0)$$

con

$$\int w(\mathbf{x}) \delta(T(\mathbf{x}) - T_0) d\mathbf{x} = 1 \quad (5.8)$$

para que se trate de una densidad de probabilidad. Sustituyendo en la ecuación 5.7, se obtiene que

$$p(\mathbf{x}; \theta) \delta(T(\mathbf{x}) - T_0) = w(\mathbf{x}) \delta(T(\mathbf{x}) - T_0) p(T(\mathbf{x}) = T_0; \theta),$$

y definiendo

$$w(\mathbf{x}) = \frac{h(\mathbf{x})}{\int h(\mathbf{x}) \delta(T(\mathbf{x}) - T_0) d\mathbf{x}}$$

de forma que se cumpla la condición de la ecuación 5.8 se obtiene que

$$p(\mathbf{x}; \theta) \delta(T(\mathbf{x}) - T_0) = \frac{h(\mathbf{x}) \delta(T(\mathbf{x}) - T_0)}{\int h(\mathbf{x}) \delta(T(\mathbf{x}) - T_0) d\mathbf{x}} p(T(\mathbf{x}) = T_0; \theta)$$

o

$$p(\mathbf{x}; \theta) = g(T(\mathbf{x}) = T_0; \theta) h(\mathbf{x})$$

con

$$g(T(\mathbf{x}) = T_0; \theta) = \frac{p(T(\mathbf{x}) = T_0; \theta)}{\int h(\mathbf{x}) \delta(T(\mathbf{x}) - T_0) d\mathbf{x}},$$

que es lo que se quería demostrar. Además, en principio, la PDF del estadístico suficiente puede obtenerse de la factorización como

$$p(T(\mathbf{x}) = T_0; \theta) = g(T(\mathbf{x}) = T_0; \theta) \int h(\mathbf{x}) \delta(T(\mathbf{x}) - T_0) d\mathbf{x}.$$

5.1.2. Ejemplo: nivel de DC en WGN

Se buscará un estadístico suficiente para el problema de la estimación del nivel de DC en WGN. Como se vió previamente, la PDF de los datos es (ecuación 3.67)

$$p(\mathbf{x}; A) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - A)^2 \right].$$

Para demostrar que existe una factorización, se observa que el exponente se puede escribir como

$$\sum_{n=0}^{N-1} (x[n] - A)^2 = \sum_{n=0}^{N-1} (x^2[n] - 2Ax[n] + A^2) = \sum_{n=0}^{N-1} x^2[n] - 2A \sum_{n=0}^{N-1} x[n] + NA^2,$$

por lo que la PDF se factoriza como

$$p(\mathbf{x}; A) = \underbrace{\frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left[-\frac{1}{2\sigma^2} \left(NA^2 - 2A \sum_{n=0}^{N-1} x[n] \right) \right]}_{g(T(\mathbf{x}), A)} \underbrace{\exp \left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} x^2[n] \right]}_{h(\mathbf{x})}.$$

Se concluye que $T(\mathbf{x}) = \sum_{n=0}^{N-1} x[n]$ es un estadístico suficiente de A . Notar que $T'(\mathbf{x}) = 2 \sum_{n=0}^{N-1} x[n]$ también es un estadístico suficiente de A , y de hecho, cualquier función uno a uno de $\sum_{n=0}^{N-1} x[n]$ también es un estadístico suficiente. Por lo tanto, los estadísticos suficientes son únicos a menos de transformaciones inyectivas.

5.1.3. Ejemplo: fase de una senoide

Se considera el problema de la estimación de la fase ϕ de una senoide en WGN,

$$x[n] = A \cos(2\pi f_0 n + \phi) + w[n], \quad n = 0, \dots, N-1,$$

donde la amplitud A y la frecuencia f_0 de la senoide son conocidas, así como la varianza σ^2 del ruido. La PDF de los datos es

$$p(\mathbf{x}, \phi) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} [x[n] - A \cos(2\pi f_0 n + \phi)]^2 \right\}.$$

El exponente se puede expandir como

$$\begin{aligned} \sum_{n=0}^{N-1} [x[n] - A(\cos 2\pi f_0 n + \phi)]^2 &= \sum_{n=0}^{N-1} x^2[n] - 2A \sum_{n=0}^{N-1} x[n] \cos(2\pi f_0 n + \phi) + A^2 \sum_{n=0}^{N-1} \cos^2(2\pi f_0 n + \phi) \\ &\stackrel{(a)}{=} \sum_{n=0}^{N-1} x^2[n] - 2A \left(\sum_{n=0}^{N-1} x[n] \cos 2\pi f_0 n \right) \cos \phi \\ &\quad + 2A \left(\sum_{n=0}^{N-1} x[n] \sin 2\pi f_0 n \right) \sin \phi + A^2 \sum_{n=0}^{N-1} \cos^2(2\pi f_0 n + \phi), \end{aligned}$$

donde en (a) se empleó la identidad trigonométrica $\cos(\alpha + \beta) = \cos \alpha \cos \beta - \sin \alpha \sin \beta$. En este ejemplo, no parece que la PDF se pueda factorizar como requiere el teorema de Neyman-Fisher, y por lo tanto, no existe un estadístico suficiente *individual*. Sin embargo, la PDF se puede factorizar como

$$\begin{aligned} p(\mathbf{x}, \phi) &= \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left\{ -\frac{1}{2\sigma^2} \left[A^2 \sum_{n=0}^{N-1} \cos^2(2\pi f_0 n + \phi) - 2AT_1(\mathbf{x}) \cos \phi + 2AT_2(\mathbf{x}) \sin \phi \right] \right\} \\ &\quad \underbrace{\hspace{15em}}_{g(T_1(\mathbf{x}), T_2(\mathbf{x}), \phi)} \\ &\quad \times \underbrace{\exp \left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} x^2[n] \right]}_{h(\mathbf{x})}, \end{aligned}$$

donde

$$\begin{aligned} T_1(\mathbf{x}) &= \sum_{n=0}^{N-1} x[n] \cos 2\pi f_0 n \\ T_2(\mathbf{x}) &= \sum_{n=0}^{N-1} x[n] \sin 2\pi f_0 n. \end{aligned} \tag{5.9}$$

En este caso, una generalización del teorema de Neyman-Fisher indica que los estadísticos $T_1(\mathbf{x})$ y $T_2(\mathbf{x})$ son *conjuntamente suficientes* para la estimación de ϕ . Sin embargo, no existe un estadístico suficiente individual en este caso.

Estadísticos conjuntamente suficientes El concepto de estadísticos conjuntamente suficientes es una extensión de la definición de estadístico suficiente. Se dice que los r estadísticos $T_1(\mathbf{x}), \dots, T_r(\mathbf{x})$ son conjuntamente suficientes si la PDF condicional $p(\mathbf{x} | T_1(\mathbf{x}), \dots, T_r(\mathbf{x}); \theta)$ de los datos no depende de θ . La generalización del teorema de la factorización de Neyman-Fisher indica que si $p(\mathbf{x}; \theta)$ puede ser factorizada como

$$p(\mathbf{x}; \theta) = g(T_1(\mathbf{x}), \dots, T_r(\mathbf{x}), \theta) h(\mathbf{x})$$

se cumple que $\{T_1(\mathbf{x}), \dots, T_r(\mathbf{x})\}$ son estadísticos suficientes de θ .

5.2. Uso del estadístico suficiente para encontrar el estimador MVU

Asumiendo que se ha encontrado un estadístico suficiente $T(\mathbf{x})$ del parámetro θ , se puede usar el teorema de Rao-Blackwell-Lehmann-Scheffe (RBLS) para encontrar el estimador MVU. El enfoque se ilustra mediante un ejemplo.

5.2.1. Ejemplo: nivel de DC en WGN

En el ejemplo de la sección 5.1.2 se encontró un estadístico suficiente para el nivel A de DC y se quiere encontrar el estimador \hat{A} MVU. Hay dos procedimientos distintos para encontrar \hat{A} en base al estadístico suficiente $T(\mathbf{x}) = \sum_{n=0}^{N-1} x[n]$:

1. Encontrar cualquier estimador insesgado de A , por ejemplo, $\check{A} = x[0]$, y determinar $\hat{A} = E(\check{A} | T)$. La esperanza se toma respecto a $p(\check{A} | T)$.
2. Encontrar alguna función g de forma que $\hat{A} = g(T)$ sea un estimador insesgado de A .

Para hacerlo mediante el primer procedimiento, sea $\check{A} = x[0]$, y hay que determinar $\hat{A} = E(x[0] | \sum_{n=0}^{N-1} x[n])$. Para calcular la esperanza condicional se necesitará la siguiente propiedad la PDF gaussiana condicional: si $[x \ y]^T$ es un vector gaussiano de media $\boldsymbol{\mu} = [E(x) \ E(y)]^T$ y matriz de covarianza

$$\mathbf{C} = \begin{bmatrix} \text{var}(x) & \text{cov}(x, y) \\ \text{cov}(y, x) & \text{var}(y) \end{bmatrix},$$

se puede demostrar que (ver el apéndice F)

$$\begin{aligned} E(x | y) &= \int_{-\infty}^{\infty} xp(x | y)dx \\ &= E(x) + \frac{\text{cov}(x, y)}{\text{var}(y)}(y - E(y)). \end{aligned} \quad (5.10)$$

Con $x = x[0]$ y $y = \sum_{n=0}^{N-1} x[n]$ se observa que

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x[0] \\ \sum_{n=0}^{N-1} x[n] \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & 0 & \dots & 1 \\ 1 & 1 & 1 & \dots & 1 \end{bmatrix}}_{\mathbf{L}} \begin{bmatrix} x[0] \\ x[1] \\ \vdots \\ x[N-1] \end{bmatrix}.$$

Esto implica que la PDF de $[x, y]^T$ es $\mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$ por tratarse de una transformación lineal de un vector gaussiano, donde la media es

$$\boldsymbol{\mu} = E(\mathbf{L}\mathbf{x}) = \mathbf{L}E(\mathbf{x}) = \mathbf{L}\mathbf{A}\mathbf{1} = \begin{bmatrix} A \\ NA \end{bmatrix},$$

y la matriz de covarianza es

$$\begin{aligned} \mathbf{C} &= E \{ [\mathbf{L}\mathbf{x} - E(\mathbf{L}\mathbf{x})][\mathbf{L}\mathbf{x} - E(\mathbf{L}\mathbf{x})]^T \} \\ &= E \{ \mathbf{L}[\mathbf{x} - E(\mathbf{x})][\mathbf{x} - E(\mathbf{x})]^T \mathbf{L}^T \} \\ &= \mathbf{L}E \{ [\mathbf{x} - E(\mathbf{x})][\mathbf{x} - E(\mathbf{x})]^T \} \mathbf{L}^T \\ &= \mathbf{L}(\sigma^2 \mathbf{I}) \mathbf{L}^T \\ &= \sigma^2 \mathbf{L}\mathbf{L}^T \\ &= \sigma^2 \begin{bmatrix} 1 & 1 \\ 1 & N \end{bmatrix}. \end{aligned}$$

Por lo tanto, aplicando el resultado de la ecuación 5.10 con $E(x) = A$, $E(y) = N$, $\text{cov}(x, y) = \sigma^2$ y $\text{var}(y) = N\sigma^2$

$$\begin{aligned}\hat{A} &= E\left(x[0] \left| \sum_{n=0}^{N-1} x[n] \right.\right) \\ &= A + \frac{\sigma^2}{N\sigma^2} \left(\sum_{n=0}^{N-1} x[n] - NA \right) \\ &= \frac{1}{N} \sum_{n=0}^{N-1} x[n],\end{aligned}$$

se obtiene la media muestral, que es el estimador MVU. Este enfoque, que requiere el cálculo de la media condicional, suele ser matemáticamente intratable.

El segundo enfoque consiste en encontrar una función g de forma que

$$\hat{A} = g\left(\sum_{n=0}^{N-1} x[n]\right)$$

sea un estimador insesgado de A . Como en este caso $E(\sum_{n=0}^{N-1} x[n]) = NA$, eligiendo $g(x) = x/N$ se obtiene

$$\hat{A} = \frac{1}{N} \sum_{n=0}^{N-1} x[n]$$

como el estimador MVU. Este enfoque suele ser mucho mas fácil de aplicar y es el que se emplea generalmente en la práctica.

5.2.2. Teorema de Rao-Blackwell-Lehmann-Scheffe

Si $\check{\theta}$ es un estimador insesgado de θ y $T(\mathbf{x})$ es un estadístico suficiente de θ , se cumple que $\hat{\theta} = E(\check{\theta} | T(\mathbf{x}))$ es

1. un estimador válido de θ , en el sentido de que no depende de θ .
2. insesgado
3. de menor varianza que $\check{\theta}$, para todo θ

Además, si el estadístico es *completo*, $\hat{\theta}$ es el estimador MVU.

Demostración Para demostrar (1), se observa que

$$\begin{aligned}\hat{\theta} &= E(\check{\theta} | T(\mathbf{x})) \\ &= \int \check{\theta}(\mathbf{x}) p(\mathbf{x} | T(\mathbf{x}); \theta) d\mathbf{x},\end{aligned}\tag{5.11}$$

donde la segunda igualdad proviene de que $E(g(x)) = \int g(x)p(x)dx$. Por la definición de estadístico suficiente, $p(\mathbf{x} | T(\mathbf{x}); \theta)$ no depende de θ , depende únicamente de \mathbf{x} y $T(\mathbf{x})$. De esta forma, al integrar en \mathbf{x} , el resultado será únicamente una función de T y como consecuencia, solo de \mathbf{x} . Se concluye que el estimador es válido por no depender de θ .

La afirmación (2) indica que $E(\hat{\theta}) = \theta$. Para demostrarlo, se considera que, como se dedujo previamente, $\hat{\theta}$ es únicamente función de T , y por lo tanto,

$$\begin{aligned} E(\hat{\theta}) &= \int \hat{\theta}(T) p(T(\mathbf{x}); \theta) dT \\ &\stackrel{(a)}{=} \iint \check{\theta}(\mathbf{x}) p(\mathbf{x} | T(\mathbf{x}); \theta) d\mathbf{x} p(T(\mathbf{x}); \theta) dT \\ &\stackrel{(b)}{=} \int \check{\theta}(\mathbf{x}) \int p(\mathbf{x}, T(\mathbf{x}); \theta) dT d\mathbf{x} \\ &\stackrel{(c)}{=} \int \check{\theta}(\mathbf{x}) p(\mathbf{x}; \theta) d\mathbf{x} \\ &= E(\check{\theta}), \end{aligned}$$

donde en (a) se uso la ecuación 5.11, en (b) se tuvo en cuenta que $p(x|y)p(y) = p(x, y)$ y en (c) que $\int p(x, y) dy = p(x)$. Como $\check{\theta}$ es insesgado por hipótesis, se cumple que

$$E(\hat{\theta}) = E(\check{\theta}) = \theta,$$

concluyendo que $\hat{\theta}$ es insesgado.

Para probar (3) se observa que

$$\begin{aligned} \text{var}(\check{\theta}) &= E[(\check{\theta} - E(\check{\theta}))^2] \\ &= E[(\check{\theta} - \hat{\theta} + \hat{\theta} - \theta)^2] \\ &= E[(\check{\theta} - \hat{\theta})^2] + 2E[(\check{\theta} - \hat{\theta})(\hat{\theta} - \theta)] + E[(\hat{\theta} - \theta)^2]. \end{aligned}$$

Se demostrará que el término cruzado $E[(\check{\theta} - \hat{\theta})(\hat{\theta} - \theta)]$ es nulo. Por definición, $\hat{\theta}$ es únicamente función de T y por lo tanto, la esperanza del término cruzado es respecto a la PDF conjunta de $\check{\theta}$ y T . De esta forma,

$$E_{T, \check{\theta}}[(\check{\theta} - \hat{\theta})(\hat{\theta} - \theta)] = E_T E_{\check{\theta}|T}[(\check{\theta} - \hat{\theta})(\hat{\theta} - \theta)].$$

Pero

$$\begin{aligned} E_{\check{\theta}|T}[(\check{\theta} - \hat{\theta})(\hat{\theta} - \theta)] &\stackrel{(a)}{=} E_{\check{\theta}|T}[(\check{\theta} - \hat{\theta})](\hat{\theta} - \theta) \\ &\stackrel{(b)}{=} [E_{\check{\theta}|T}(\check{\theta} | T) - \hat{\theta}](\hat{\theta} - \theta) \\ &\stackrel{(c)}{=} (\hat{\theta} - \hat{\theta})(\hat{\theta} - \theta) \\ &= 0, \end{aligned}$$

donde en (a) y (b) se empleó que como $\hat{\theta}$ es función de T se cumple que $E(\hat{\theta} | T) = \hat{\theta}$ y en (c) la definición de $\hat{\theta}$. Finalmente, se tiene que

$$\begin{aligned} \text{var}(\check{\theta}) &= E[(\check{\theta} - \hat{\theta})^2] + E[(\hat{\theta} - \theta)^2] \\ &= E[(\check{\theta} - \hat{\theta})^2] + \text{var}(\hat{\theta}) \\ &\geq \text{var}(\hat{\theta}). \end{aligned}$$

Falta ver que si $T(\mathbf{x})$ es completo, $\hat{\theta} = E(\check{\theta} | T(\mathbf{x}))$ es el estimador MVU. Un estadístico es completo si existe solo una función del estadístico que lo hace insesgado. $E(\check{\theta} | T(\mathbf{x}))$ es un estimador insesgado y de menor varianza que el estimador $\check{\theta}$. Como $E(\check{\theta} | T(\mathbf{x}))$ es una función únicamente de $T(\mathbf{x})$ ya que

$$\hat{\theta} = E(\check{\theta} | T(\mathbf{x})) = \int \check{\theta} p(\check{\theta} | T(\mathbf{x})) d\check{\theta} = g(T(\mathbf{x})),$$

y si $T(\mathbf{x})$ es completo, existe solo una función de T que es un estimador insesgado, $\hat{\theta}$ es único independientemente del estimador $\tilde{\theta}$ insesgado que se elija. Es decir, cualquier estimador $\tilde{\theta}$ insesgado se mapea en el mismo estimador $\hat{\theta}$. Como la varianza de $\hat{\theta}$ es menor que la de cualquier estimador $\tilde{\theta}$ insesgado, se concluye que $\hat{\theta}$ debe ser el estimador MVU.

Se puede demostrar que un estadístico suficiente es completo si se cumple que

$$\int_{-\infty}^{\infty} v(T)p(T; \theta) dT = 0, \quad \forall \theta$$

únicamente para la función v nula o $v(T) = 0$ para todo T (ver [2]).

5.2.3. Ejemplo: media de ruido uniforme

Se observan los datos

$$x[n] = w[n], \quad n = 0, \dots, N-1,$$

donde $w[n]$ es ruido IID con PDF $\mathcal{U}[0, \beta]$ con $\beta > 0$. Se desea encontrar el estimador de la media $\theta = \beta/2$. En este caso no puede emplearse la estrategia de intentar calcular la CRLB y encontrar un estimador eficiente debido a que la PDF no cumple las condiciones de regularidad, como se mostró en el problema de la sección 3.6.1. Un estimador evidente a considerar es la media muestral,

$$\hat{\theta} = \frac{1}{N} \sum_{n=0}^{N-1} x[n].$$

La media muestral es insesgada (ver el problema de la sección 2.1.2) y tiene varianza

$$\text{var}(\hat{\theta}) = \text{var}\left(\frac{1}{N} \sum_{n=0}^{N-1} x[n]\right) \stackrel{(a)}{=} \frac{1}{N^2} \sum_{n=0}^{N-1} \text{var}(x[n]) \stackrel{(b)}{=} \frac{1}{N^2} N \text{var}(x[n]) \stackrel{(c)}{=} \frac{\beta^2}{12N}, \quad (5.12)$$

donde en (a) se consideró que las muestras son independientes, en (b) que son idénticamente distribuidas y en (c) que

$$\begin{aligned} \text{var}(x[n]) &= E(x^2[n]) - E^2(x[n]) \\ &= \int_{-\infty}^{\infty} x^2[n]p(x[n]; \beta) dx[n] - \frac{\beta^2}{4} \\ &= \frac{1}{\beta} \int_0^{\beta} x^2[n] dx[n] - \frac{\beta^2}{4} \\ &= \frac{1}{\beta} \left(\frac{x^3[n]}{3} \Big|_0^{\beta} \right) - \frac{\beta^2}{4} \\ &= \frac{\beta^2}{3} - \frac{\beta^2}{4} \\ &= \frac{\beta^2}{12}. \end{aligned}$$

Para determinar si la media muestral es el estimador MVU, se seguirá el procedimiento basado en estadísticos suficientes, que consiste en emplear la factorización de Neyman-Fisher para encontrar un estadístico suficiente $T(\mathbf{x})$, determinar si $T(\mathbf{x})$ es completo, y encontrar una función g del estadístico suficiente que conduzca a un estimador $\hat{\theta} = g(T(\mathbf{x}))$ insesgado.

Como $x[n] \sim \mathcal{U}[0, \beta]$,

$$p(x[n]; \theta) = \frac{1}{\beta} [u(x[n]) - u(x[n] - \beta)],$$

donde $\beta = 2\theta$ y $u(x)$ es la función escalón definida como

$$u(x) = \begin{cases} 1, & x > 0 \\ 0, & x < 0. \end{cases}$$

Por lo tanto, la PDF de los datos es

$$p(\mathbf{x}; \theta) = \frac{1}{\beta^N} \prod_{n=0}^{N-1} [u(x[n]) - u(x[n] - \beta)].$$

Esta PDF es no nula solo en $0 < x[n] < \beta$ para todos los $x[n]$, es decir

$$p(\mathbf{x}; \theta) = \begin{cases} \frac{1}{\beta^N}, & 0 < x[n] < \beta, \quad n = 0, \dots, N-1 \\ 0, & \text{en otro caso,} \end{cases}$$

que alternativamente, se puede escribir como

$$p(\mathbf{x}; \theta) = \begin{cases} \frac{1}{\beta^N}, & \text{máx } x[n] < \beta, \text{ mín } x[n] > 0 \\ 0, & \text{en otro caso,} \end{cases}$$

o empleando la función escalón, como

$$p(\mathbf{x}; \theta) = \underbrace{\frac{1}{\beta^N} u(\beta - \text{máx } x[n])}_{g(T(\mathbf{x}), \theta)} \underbrace{u(\text{mín } x[n])}_{h(\mathbf{x})}.$$

Por el teorema de la factorización de Neyman-Fisher $T(\mathbf{x}) = \text{máx } x[n]$ es un estadístico suficiente de θ . Se asumirá por el momento que además es completo, pero se demostrará mas adelante.

El próximo paso es determinar una función del estadístico que lo haga insesgado. Para esto, se necesita calcular la esperanza de $T = \text{máx } x[n]$. El estadístico T se denomina *estadístico de orden* (ver el capítulo 7 de [2], por ejemplo). Se comenzará calculando la PDF. Una forma de hacerlo, es calcular la distribución de probabilidad y luego derivar para obtener la PDF. La distribución de probabilidad es

$$\begin{aligned} \Pr\{T \leq \xi\} &= \Pr\{x[0] \leq \xi, x[1] \leq \xi, \dots, x[N-1] \leq \xi\} \\ &= \prod_{n=0}^{N-1} \Pr\{x[n] \leq \xi\} \\ &= \Pr\{x[n] \leq \xi\}^N, \end{aligned}$$

donde se consideró que las variables aleatorias $x[n]$ son IID. La PDF de T es por lo tanto

$$\begin{aligned} p_T(\xi) &= \frac{d \Pr\{T \leq \xi\}}{d\xi} \\ &= N \Pr\{x[n] \leq \xi\}^{N-1} \frac{d \Pr\{x[n] \leq \xi\}}{d\xi} \\ &\stackrel{(a)}{=} N \Pr\{x[n] \leq \xi\}^{N-1} p_{x[n]}(\xi), \end{aligned} \tag{5.13}$$

donde en (a) se notó que $d \Pr\{x[n] \leq \xi\}/d\xi$ es la PDF de $x[n]$, dada por

$$p_{x[n]}(\xi; \theta) = \begin{cases} \frac{1}{\beta}, & 0 < \xi < \beta \\ 0, & \text{en otro caso.} \end{cases} \tag{5.14}$$

Integrando, se obtiene la distribución de $x[n]$,

$$\Pr\{x[n] \leq \xi\} = \int_{-\infty}^{\xi} \frac{1}{\beta} du = \int_0^{\xi} \frac{1}{\beta} du = \frac{1}{\beta} u \Big|_0^{\xi} = \frac{1}{\beta} (\xi - 0) = \frac{\xi}{\beta}, \quad 0 < \xi < \beta$$

resultando en

$$\Pr\{x[n] \leq \xi\} = \begin{cases} 0, & \xi < 0 \\ \frac{\xi}{\beta}, & 0 < \xi < \beta \\ 1, & \xi > \beta. \end{cases} \tag{5.15}$$

Sustituyendo las ecuaciones 5.14 y 5.15 en la ecuación 5.13, se concluye que

$$p_T(\xi) = \begin{cases} 0, & \xi < 0 \\ N \left(\frac{\xi}{\beta} \right)^{N-1} \frac{1}{\beta}, & 0 < \xi < \beta \\ 0, & \xi > \beta. \end{cases} \quad (5.16)$$

Por lo tanto, la esperanza del estadístico T es

$$\begin{aligned} E(T) &= \int_{-\infty}^{\infty} \xi p_T(\xi) d\xi \\ &= \int_0^{\beta} \xi N \left(\frac{\xi}{\beta} \right)^{N-1} \frac{1}{\beta} d\xi \\ &= \frac{N}{\beta^N} \int_0^{\beta} \xi^N d\xi \\ &= \frac{N}{\beta^N} \frac{\xi^{N+1}}{N+1} \Big|_0^{\beta} \\ &= \frac{N}{N+1} \beta \\ &= \frac{2N}{N+1} \theta. \end{aligned}$$

Como la esperanza del estadístico es proporcional al valor del parámetro, para hacerlo insesgado alcanza con multiplicarlo por el inverso de la constante de proporcionalidad. Por lo tanto, el estimador MVU es

$$\hat{\theta} = \frac{N+1}{2N} T = \frac{N+1}{2N} \max x[n].$$

Se observa que, de forma contraintuitiva, la media muestral no es el estimador MVU de la media para ruido con distribución uniforme. Es de interés comparar la varianza mínima con la de la media muestral. La varianza mínima es

$$\text{var}(\hat{\theta}) = \left(\frac{N+1}{2N} \right)^2 \text{var}(T),$$

donde

$$\text{var}(T) = E(T^2) - E^2(T)$$

y

$$\begin{aligned} E(T^2) &= \int_{-\infty}^{\infty} \xi^2 p_T(\xi) d\xi \\ &= \int_0^{\beta} \xi^2 N \left(\frac{\xi}{\beta} \right)^{N-1} \frac{1}{\beta} d\xi \\ &= \frac{N}{\beta^N} \int_0^{\beta} \xi^{N+1} d\xi \\ &= \frac{N}{\beta^N} \frac{\xi^{N+2}}{N+2} \Big|_0^{\beta} \\ &= \frac{N}{N+2} \beta^2. \end{aligned}$$

Combinando los resultados, se obtiene que

$$\text{var}(T) = \frac{N}{N+2} \beta^2 - \left(\frac{N}{N+1} \beta \right)^2 = \frac{N\beta^2}{(N+1)^2(N+2)},$$

por lo que la varianza mínima es

$$\text{var}(\hat{\theta}) = \left(\frac{N+1}{2N} \right)^2 \frac{N\beta^2}{(N+1)^2(N+2)} = \frac{\beta^2}{4N(N+2)}.$$

Comparando este resultado con la varianza de la media muestral dada por la ecuación 5.12, se observa que la varianza mínima es menor que la de la media muestral para $N \geq 2$ (para $N = 1$ son iguales ya que los estimadores son iguales). La diferencia es sustancial, ya que la varianza mínima decrece con N como $1/N^2$ mientras que la varianza de la media muestral decrece como $1/N$. En la figura 5.1 se compara gráficamente la diferencia en la varianza de ambos estimadores. Se muestran histogramas de los estimadores empleando $N = 100$ muestras en 10000 experimentos.

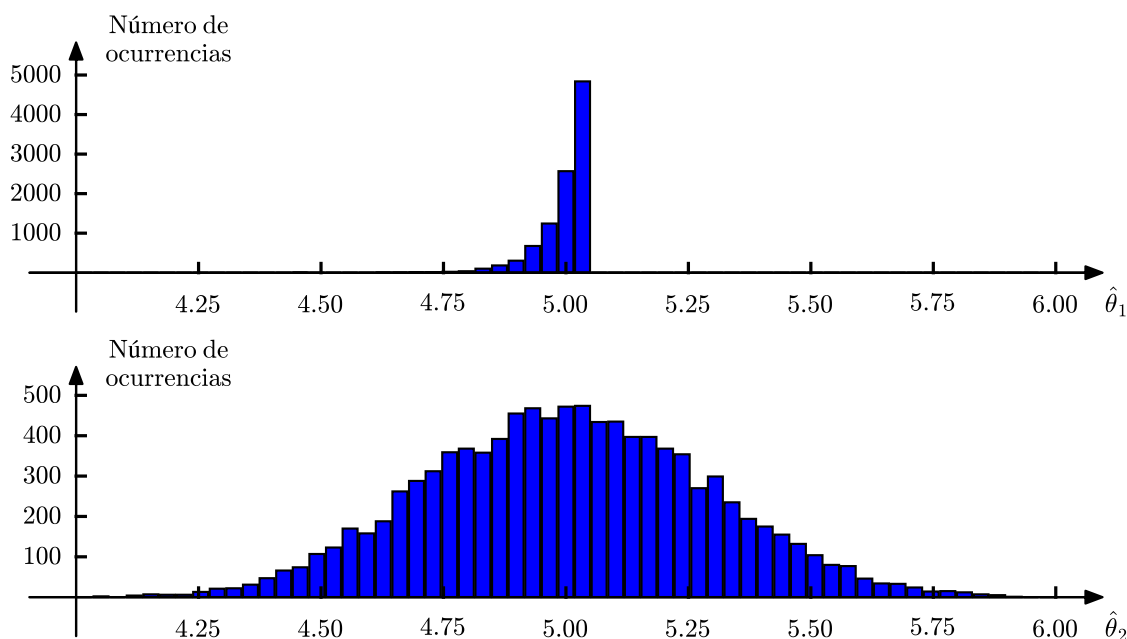


Figura 5.1: Histogramas de los estimadores $\hat{\theta}_1 = (N+1)\text{máx } x[n]/(2N)$ y la media muestral $\hat{\theta}_2$. Se calculó la media de $N = 100$ muestras de ruido uniforme $\mathcal{U}(0, 10)$ empleando ambos estimadores 10000 veces. Se observa que la dispersión de la media muestral respecto a la media $\beta/2 = 5$ es mucho mayor.

Falta verificar que el estadístico $T = \text{máx } x[n]$ es completo. Para hacerlo, se demostrará que dada una función $g(\xi)$, si se cumple que $E[g(T)] = 0$ para todo valor del parámetro, necesariamente $g(\xi) = 0$ para todo ξ (esta demostración está basada en el ejemplo 6.2.23 en [10]). Efectivamente, considérese la función $g(\xi)$ tal que

$$E[g(T)] = \int_{-\infty}^{\infty} g(\xi) p_T(\xi; \beta) d\xi = 0, \quad \forall \beta > 0.$$

Como $E[g(T)]$ es la función nula, y por lo tanto constante para todo β se cumple que

$$\frac{d}{d\beta} E[g(T)] = 0.$$

Además,

$$\begin{aligned}
\frac{d}{d\beta} E[g(T)] &= \frac{d}{d\beta} \int_{-\infty}^{\infty} g(\xi) p_T(\xi; \beta) d\xi = 0 \\
&\stackrel{(a)}{=} \frac{d}{d\beta} \int_0^{\beta} g(\xi) N \frac{\xi^{N-1}}{\beta^N} d\xi \\
&= \frac{d}{d\beta} \left[\frac{1}{\beta^N} \int_0^{\beta} g(\xi) N \xi^{N-1} d\xi \right] \\
&\stackrel{(b)}{=} \left(\frac{d}{d\beta} \frac{1}{\beta^N} \right) \int_0^{\beta} g(\xi) N \xi^{N-1} d\xi + \frac{1}{\beta^N} \frac{d}{d\beta} \int_0^{\beta} g(\xi) N \xi^{N-1} d\xi \\
&\stackrel{(c)}{=} -\frac{N}{\beta^{N+1}} \int_0^{\beta} g(\xi) N \xi^{N-1} d\xi + \frac{1}{\beta^N} g(\beta) N \beta^{N-1} d\xi \\
&= -\frac{N}{\beta} \int_0^{\beta} g(\xi) N \frac{\xi^{N-1}}{\beta^N} d\xi + \frac{N}{\beta} g(\beta) \\
&\stackrel{(d)}{=} -\frac{N}{\beta} E[g(T)] + \frac{N}{\beta} g(\beta) \\
&\stackrel{(e)}{=} \frac{N}{\beta} g(\beta),
\end{aligned}$$

donde en (a) se substituyó la PDF de T dada por la ecuación 5.16, en (b) se empleó la regla de derivación de un producto, en (c) se empleó el teorema fundamental del cálculo para determinar la derivada del segundo sumando⁴, en (d) se notó que la integral es $E[g(T)]$ y en e que $E[g(T)] = 0$. Como

$$\frac{d}{d\beta} E[g(T)] = 0 = \frac{N}{\beta} g(\beta), \quad \forall \beta > 0,$$

se concluye que $g(\beta) = 0$ para todo $\beta > 0$, y por lo tanto, el estadístico T es completo.

5.3. Extensión a un parámetro vectorial

En lo que sigue, se extienden los resultados en la búsqueda del estimador MVU en el caso en que el parámetro es un vector $p \times 1$. La cantidad de estadísticos suficientes puede ser mayor que el número de parámetros ($r > p$), la misma cantidad ($r = p$) o incluso menor que el número de parámetros ($r < p$). La situación deseable es que el número de estadísticos suficientes sea igual al número de parámetros. De esta forma, podría buscarse una transformación del estadístico suficiente que lo haga insesgado.

Definición Un vector estadístico $\mathbf{T}(\mathbf{x}) = [T_1(\mathbf{x}) T_2(\mathbf{x}) \dots T_r(\mathbf{x})]^T$ se dice suficiente para la estimación del parámetro $\boldsymbol{\theta}$ si la PDF de los datos condicionada al estadístico $p(\mathbf{x} | \mathbf{T}(\mathbf{x}); \boldsymbol{\theta})$ no depende de $\boldsymbol{\theta}$, es decir, $p(\mathbf{x} | \mathbf{T}(\mathbf{x}); \boldsymbol{\theta}) = p(\mathbf{x} | \mathbf{T}(\mathbf{x}))$. En general, existen muchos vectores $\mathbf{T}(\mathbf{x})$ que cumplen dicha condición, pero es de interés el *estadístico suficiente mínimo*, que es el de menor dimensión.

5.3.1. Teorema: Factorización de Neyman-Fisher (parámetro vectorial)

Si la PDF $p(\mathbf{x}; \boldsymbol{\theta})$ de los datos puede ser factorizada como

$$p(\mathbf{x}; \boldsymbol{\theta}) = g(\mathbf{T}(\mathbf{x}), \boldsymbol{\theta}) h(\mathbf{x}), \quad (5.17)$$

donde g es una función que depende de \mathbf{x} solo a través de $\mathbf{T}(\mathbf{x})$, un estadístico $r \times 1$, y también de $\boldsymbol{\theta}$, y h es una función que depende solo de \mathbf{x} , se cumple que $\mathbf{T}(\mathbf{x})$ es un estadístico suficiente de $\boldsymbol{\theta}$. Recíprocamente, si $\mathbf{T}(\mathbf{x})$ es un estadístico suficiente de $\boldsymbol{\theta}$, la PDF puede ser factorizada como la ecuación 5.17.

⁴Ver https://en.wikipedia.org/wiki/Fundamental_theorem_of_calculus por ejemplo.

5.3.2. Ejemplo: estimación de los parámetros de una senoide

Se considera el caso de una senoide en WGN,

$$x[n] = A \cos 2\pi f_0 n + w[n], \quad n = 0, \dots, N-1,$$

donde la amplitud A , la frecuencia f_0 y la varianza σ^2 del ruido son desconocidas. El parámetro vectorial desconocido es entonces $\boldsymbol{\theta} = [A f_0 \sigma^2]^T$. La PDF es

$$p(\mathbf{x}, \boldsymbol{\theta}) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - A \cos 2\pi f_0 n)^2 \right].$$

Expandiendo el exponente, se tiene que

$$\sum_{n=0}^{N-1} (x[n] - A \cos 2\pi f_0 n)^2 = \sum_{n=0}^{N-1} x^2[n] - 2A \sum_{n=0}^{N-1} x[n] \cos 2\pi f_0 n + A^2 \sum_{n=0}^{N-1} \cos^2 2\pi f_0 n.$$

Como la sumatoria del segundo sumando depende tanto de los datos \mathbf{x} como del parámetro desconocido f_0 , la PDF no puede factorizarse como la ecuación 5.17. Sin embargo, si la frecuencia f_0 fuera conocida y el parámetro desconocido fuera $\boldsymbol{\theta} = [A \sigma^2]^T$, la PDF se puede factorizar como

$$p(\mathbf{x}, \boldsymbol{\theta}) = \underbrace{\frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left[-\frac{1}{2\sigma^2} \left(\sum_{n=0}^{N-1} x^2[n] - 2A \sum_{n=0}^{N-1} x[n] \cos 2\pi f_0 n + A^2 \sum_{n=0}^{N-1} \cos^2 2\pi f_0 n \right) \right]}_{g(T(\mathbf{x}), \boldsymbol{\theta})} \times \underbrace{1}_{h(\mathbf{x})},$$

donde

$$\mathbf{T}(\mathbf{x}) = \begin{bmatrix} \sum_{n=0}^{N-1} x[n] \cos^2 2\pi f_0 n \\ \sum_{n=0}^{N-1} x^2[n] \end{bmatrix}.$$

Por lo tanto, $\mathbf{T}(\mathbf{x})$ es un estadístico suficiente para A y σ^2 , pero solo en el caso en que la frecuencia f_0 es conocida.

5.3.3. Ejemplo: Nivel de DC en WGN con potencia del ruido desconocida

Sea

$$x[n] = A + w[n], \quad n = 0, \dots, N-1,$$

donde la amplitud A y la varianza σ^2 del ruido son desconocidas. Observando que se trata del caso de la senoide en WGN del ejemplo anterior con $f_0 = 0$, se tiene que el estadístico suficiente para el parámetro $\boldsymbol{\theta} = [A \sigma^2]^T$ es

$$\mathbf{T}(\mathbf{x}) = \begin{bmatrix} \sum_{n=0}^{N-1} x[n] \\ \sum_{n=0}^{N-1} x^2[n] \end{bmatrix}.$$

En el ejemplo de la sección 5.1.2, se encontró que $\sum_{n=0}^{N-1} x[n]$ es un estadístico suficiente de A cuando σ^2 es conocido. También puede verse que $\sum_{n=0}^{N-1} x^2[n]$ es un estadístico suficiente de σ^2 cuando A es conocido. Esto sin embargo, no siempre se cumple. Como se encontró dos estadísticos suficientes para estimar dos parámetros, se debería poder encontrar un estimador MVU en este ejemplo.

5.3.4. Teorema: Rao-Blackwell-Lehmann-Scheffe (parámetro vectorial)

Si $\tilde{\theta}$ es un estimador insesgado de θ y $\mathbf{T}(\theta)$ es un estadístico suficiente $r \times 1$ para θ , se cumple que $\hat{\theta} = E(\tilde{\theta} | \mathbf{T}(\mathbf{x}))$ es

1. un estimador válido de θ , en el sentido de que no depende de θ .
2. insesgado
3. de menor varianza que $\tilde{\theta}$, para todo θ

Además, si el estadístico es completo, $\hat{\theta}$ es el estimador MVU. En el caso vectorial, completitud significa que para una función $r \times 1$ arbitraria $\mathbf{v}(\mathbf{T})$ de \mathbf{T} , si se cumple que

$$E(\mathbf{v}(\mathbf{T})) = \int \mathbf{v}(\mathbf{T}) p(\mathbf{T}; \theta) d\mathbf{T} = \mathbf{0}, \quad \forall \theta$$

se debe cumplir que

$$\mathbf{v}(\mathbf{T}) = \mathbf{0} \quad \forall \mathbf{T}.$$

En general, el cálculo de $E(\tilde{\theta} | \mathbf{T}(\mathbf{x}))$ es difícil. Para poder determinar el estimador MVU sin tener que evaluar $E(\tilde{\theta} | \mathbf{T}(\mathbf{x}))$, la dimensión del estadístico suficiente debe ser igual a la dimensión del parámetro desconocido, es decir, se debe cumplir que $r = p$. En ese caso, solo se necesita encontrar una función \mathbf{g} de dimensión p tal que

$$E(\mathbf{g}(\mathbf{T})) = \theta$$

donde \mathbf{T} es el estadístico suficiente.

5.3.5. Ejemplo: Nivel de DC en WGN con potencia del ruido desconocida (continuación)

En este problema se encontró que el estadístico suficiente es

$$\mathbf{T}(\mathbf{x}) = \begin{bmatrix} T_1(\mathbf{x}) \\ T_2(\mathbf{x}) \end{bmatrix} = \begin{bmatrix} \sum_{n=0}^{N-1} x[n] \\ \sum_{n=0}^{N-1} x^2[n] \end{bmatrix}.$$

Se intentará encontrar una función \mathbf{g} que cumpla que $E(\mathbf{g}(\mathbf{T})) = \theta$. En el caso de encontrarla, $\hat{\theta} = \mathbf{g}(\mathbf{T})$ es el estimador MVU. Para hacerlo, se parte calculando la media del estadístico suficiente,

$$E(\mathbf{T}(\mathbf{x})) = \begin{bmatrix} NA \\ NE(x^2[n]) \end{bmatrix} = \begin{bmatrix} NA \\ N(\sigma^2 + A^2) \end{bmatrix}.$$

Teniendo esto en cuenta, una transformación razonable es

$$\mathbf{g}(\mathbf{T}(\mathbf{x})) = \begin{bmatrix} \frac{1}{N} T_1(\mathbf{x}) \\ \frac{1}{N} T_2(\mathbf{x}) - \left[\frac{1}{N} T_1(\mathbf{x}) \right]^2 \end{bmatrix} = \begin{bmatrix} \bar{x} \\ \frac{1}{N} \sum_{n=0}^{N-1} x^2[n] - \bar{x}^2 \end{bmatrix}.$$

Claramente, el sesgo del primer componente se elimina dividiendo entre N . Por otro lado, al dividir entre N el segundo componente se obtiene un estimador del segundo momento, $E(x^2[n]) = \sigma^2 + A^2$, por lo que hay que restar una estimación de la media al cuadrado. De esta forma, se tiene que la esperanza del primer componente del estadístico transformado es

$$E(\bar{x}) = A,$$

y la del segundo componente es

$$\begin{aligned} E\left(\frac{1}{N} \sum_{n=0}^{N-1} x^2[n] - \bar{x}^2\right) &= E(x^2[n]) - E(\bar{x}^2) \\ &\stackrel{(a)}{=} (\sigma^2 + A^2) - \left(A^2 + \frac{\sigma^2}{N}\right) \\ &= \frac{N-1}{N} \sigma^2, \end{aligned}$$

donde en (a) se consideró que $\bar{x} \sim \mathcal{N}(A, \sigma^2/N)$ (ver el problema de la sección 2.1.3) y por lo tanto, $E(\bar{x}^2) = A^2 + \sigma^2/N$. El resultado indica que el segundo componente del estadístico transformado es sesgado, pero el sesgo se elimina multiplicando por $N/(N-1)$. Finalmente, la transformación apropiada es

$$\mathbf{g}(\mathbf{T}(\mathbf{x})) = \begin{bmatrix} \frac{1}{N} T_1(\mathbf{x}) \\ \frac{1}{N-1} \left\{ T_2(\mathbf{x}) - N \left[\frac{1}{N} T_1(\mathbf{x}) \right]^2 \right\} \end{bmatrix} = \begin{bmatrix} \bar{x} \\ \frac{1}{N-1} \left(\sum_{n=0}^{N-1} x^2[n] - N\bar{x}^2 \right) \end{bmatrix}.$$

Notando que

$$\sum_{n=0}^{N-1} (x[n] - \bar{x})^2 = \sum_{n=0}^{N-1} x^2[n] - 2\bar{x} \sum_{n=0}^{N-1} x[n] + N\bar{x}^2 = \sum_{n=0}^{N-1} x^2[n] - N\bar{x}^2,$$

el estimador se puede escribir como

$$\hat{\boldsymbol{\theta}} = \begin{bmatrix} \bar{x} \\ \frac{1}{N-1} \sum_{n=0}^{N-1} (x[n] - \bar{x})^2 \end{bmatrix},$$

y es el estimador MVU para $\boldsymbol{\theta} = [A \ \sigma^2]$. En resumen, se obtuvo que cuando el parámetro desconocido es $\boldsymbol{\theta} = [A \ \sigma^2]$, el estimador MVU de la media es la media muestral \bar{x} y el estimador de la varianza es la *varianza muestral* ($\hat{\sigma}^2$). Se puede demostrar que las variables aleatorias \bar{x} y $\hat{\sigma}^2$ son independientes (ver el ejemplo 8.20 de [2]) con PDF (ver sección la 7.2 de [2])

$$\begin{aligned} \bar{x} &\sim \mathcal{N}\left(A, \frac{\sigma^2}{N}\right) \\ \frac{(N-1)\hat{\sigma}^2}{\sigma^2} &\sim \chi_{N-1}^2, \end{aligned}$$

Como la varianza de una variable aleatoria χ_k^2 es $2k$, la varianza de $\hat{\sigma}^2$ es

$$\text{var}(\hat{\sigma}^2) = \frac{\sigma^4}{(N-1)^2} \times 2(N-1) = \frac{2\sigma^4}{N-1}.$$

Se concluye que la matriz de covarianza del estimador es

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}} = \begin{bmatrix} \frac{\sigma^2}{N} & 0 \\ 0 & \frac{2\sigma^4}{N-1} \end{bmatrix},$$

y puede demostrarse que la CRLB es (ver el ejemplo 3.6 de [1])

$$\mathbf{I}^{-1}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\sigma^2}{N} & 0 \\ 0 & \frac{2\sigma^4}{N} \end{bmatrix}.$$

Estos resultados muestran que el estimador MVU encontrado no alcanza la CRLB y por lo tanto, no es eficiente. Por esta misma razón, no podría haberse encontrado el estimador MVU mediante el estudio de la CRLB, es decir, intentando encontrar un estimador que alcance la cota.

5.4. Problemas

5.4.1. Problema 1

Se consideran los datos IID $x[n] \sim \mathcal{N}(0, \sigma^2)$ con $n = 0, \dots, N-1$, y el parámetro desconocido es σ^2 . Probar que $\sum_{n=0}^{N-1} x^2[n]$ es un estadístico suficiente de σ^2 empleando la definición que indica que $p(\mathbf{x} | \sum_{n=0}^{N-1} x^2[n] = T_0; \sigma^2)$ no depende de σ^2 . Sugerencia: Notar que la PDF de $s = \sum_{n=0}^{N-1} x^2[n]/\sigma^2$ es chi-cuadrado con N grados de libertad (ver ecuaciones 2.1 y 2.2),

$$p(s) = \begin{cases} \frac{s^{\frac{N}{2}-1}}{2^{\frac{N}{2}} \Gamma(N/2)} \exp\left(-\frac{s}{2}\right), & s \geq 0 \\ 0, & s < 0. \end{cases}$$

Solución Sea

$$T(\mathbf{x}) = \sum_{n=0}^{N-1} x^2[n],$$

y se quiere probar que $p(\mathbf{x} | T(\mathbf{x}) = T_0; \sigma^2)$ no depende de σ^2 . Por el teorema de la probabilidad total, la probabilidad condicional es

$$p(\mathbf{x} | T(\mathbf{x}) = T_0; \sigma^2) = \frac{p(\mathbf{x}, T(\mathbf{x}) = T_0; \sigma^2)}{p(T(\mathbf{x}) = T_0; \sigma^2)}. \quad (5.18)$$

La probabilidad de los datos es

$$p(\mathbf{x}; \sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left(-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} x^2[n]\right) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{T(\mathbf{x})}{2\sigma^2}\right]$$

por lo que la probabilidad conjunta es (ver la ecuación 5.2)

$$\begin{aligned} p(\mathbf{x}, T(\mathbf{x}) = T_0; \sigma^2) &= p(\mathbf{x}; \sigma^2) \delta(T(\mathbf{x}) - T_0) \\ &= \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{T(\mathbf{x})}{2\sigma^2}\right] \delta(T(\mathbf{x}) - T_0). \end{aligned} \quad (5.19)$$

Para calcular $p(T(\mathbf{x}) = T_0; \sigma^2)$, defínase la variable aleatoria

$$s = \sum_{n=0}^{N-1} \frac{x^2[n]}{\sigma^2}.$$

Como $x[n] \sim \mathcal{N}(0, \sigma^2)$, $x[n]/\sigma \sim \mathcal{N}(0, 1)$, por lo que s es la suma del cuadrado N variables aleatorias IID con PDF $\mathcal{N}(0, 1)$, y puede demostrarse que tiene PDF chi-cuadrado con N grados de libertad (ver por ejemplo, el capítulo 7 de [2]), es decir, $s \sim \chi^2(N)$ dada por la ecuación 2.1. Además,

$$T(\mathbf{x}) = \sum_{n=0}^{N-1} x^2[n] = \sigma^2 s,$$

y aplicando la ecuación 5.4 de la PDF de una variable aleatoria transformada con $g(s) = \sigma^2$ se tiene que

$$\begin{aligned} p(T(\mathbf{x}); \sigma^2) &= \frac{1}{\sigma^2} p_s\left(\frac{T(\mathbf{x})}{\sigma^2}\right) \\ &\stackrel{(a)}{=} \frac{\left[\frac{T(\mathbf{x})}{\sigma^2}\right]^{\frac{N}{2}-1}}{\sigma^2 2^{\frac{N}{2}} \Gamma(N/2)} \exp\left[-\frac{T(\mathbf{x})}{2\sigma^2}\right] \\ &= \frac{[T(\mathbf{x})]^{\frac{N}{2}-1}}{\sigma^2 (\sigma^2)^{\frac{N}{2}-1} 2^{\frac{N}{2}} \Gamma(N/2)} \exp\left[-\frac{T(\mathbf{x})}{2\sigma^2}\right] \\ &= \frac{[T(\mathbf{x})]^{\frac{N}{2}-1}}{(2\sigma^2)^{\frac{N}{2}} \Gamma(N/2)} \exp\left[-\frac{T(\mathbf{x})}{2\sigma^2}\right], \end{aligned} \quad (5.20)$$

donde en (a) se substituyó la ecuación de la PDF $\chi^2(N)$ evaluada en $s = T(\mathbf{x})/\sigma^2$. Sustituyendo las ecuaciones 5.19 y 5.20 en la ecuación 5.18 con $T(\mathbf{x}) = T_0$, se obtiene que

$$\begin{aligned} p(\mathbf{x} | T(\mathbf{x}) = T_0; \sigma^2) &= \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{T_0}{2\sigma^2}\right] \delta(T(\mathbf{x}) - T_0) \\ &= \frac{T_0^{\frac{N}{2}-1}}{(2\sigma^2)^{\frac{N}{2}} \Gamma(N/2)} \exp\left[-\frac{T_0}{2\sigma^2}\right] \\ &= \frac{\Gamma(N/2)}{\pi^{\frac{N}{2}} T_0^{\frac{N}{2}-1}} \delta(T(\mathbf{x}) - T_0), \end{aligned}$$

que es independiente de σ^2 . Se concluye que como la probabilidad condicional es independiente de σ^2 , $T(\mathbf{x}) = \sum_{n=0}^{N-1} x^2[n]$ es un estadístico suficiente de σ^2 .

5.4.2. Problema 2

Las observaciones IID $x[n]$ con $n = 0, \dots, N-1$ tienen PDF Rayleigh

$$p(x[n], \sigma^2) = \begin{cases} \frac{x[n]}{\sigma^2} \exp\left(-\frac{1}{2} \frac{x^2[n]}{\sigma^2}\right), & x[n] \geq 0 \\ 0, & x[n] < 0. \end{cases}$$

Encontrar un estadístico suficiente para σ^2 .

Solución La PDF de los datos es

$$\begin{aligned} p(\mathbf{x}; \sigma^2) &= \prod_{n=0}^{N-1} \frac{x[n]}{\sigma^2} \exp\left(-\frac{1}{2} \frac{x^2[n]}{\sigma^2}\right) u(x[n]) \\ &= \underbrace{\frac{1}{\sigma^{2N}} \exp\left(-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} x^2[n]\right)}_{g(T(\mathbf{x}), \theta)} \underbrace{u(\min x[n]) \prod_{n=0}^{N-1} x[n]}_{h(\mathbf{x})}, \end{aligned}$$

concluyendo que el estadístico suficiente es

$$T(\mathbf{x}) = \sum_{n=0}^{N-1} x^2[n].$$

5.4.3. Problema 3

Las observaciones IID $x[n]$ con $n = 0, \dots, N-1$ tienen PDF exponencial

$$p(x[n], \lambda) = \begin{cases} \lambda \exp(-\lambda x[n]), & x[n] \geq 0 \\ 0, & x[n] < 0. \end{cases}$$

Encontrar un estadístico suficiente para λ .

Solución La PDF de los datos es

$$\begin{aligned} p(\mathbf{x}; \lambda) &= \prod_{n=0}^{N-1} \lambda \exp(-\lambda x[n]) u(x[n]) \\ &= \underbrace{\lambda^N \exp\left(-\lambda \sum_{n=0}^{N-1} x[n]\right)}_{g(T(\mathbf{x}), \theta)} \underbrace{u(\min x[n])}_{h(\mathbf{x})}, \end{aligned}$$

concluyendo que el estadístico suficiente es

$$T(\mathbf{x}) = \sum_{n=0}^{N-1} x[n].$$

5.4.4. Problema 4

Las observaciones IID $x[n]$ con $n = 0, \dots, N-1$ tienen PDF $\mathcal{N}(\theta, \theta)$ con $\theta > 0$. Encontrar un estadístico suficiente para θ .

Solución La PDF de una muestra es

$$p(x[n]; \theta) = \frac{1}{\sqrt{2\pi\theta}} \exp \left[-\frac{(x[n] - \theta)^2}{2\theta} \right],$$

por lo que la PDF de los datos es

$$p(\mathbf{x}, \theta) = \frac{1}{(2\pi\theta)^{\frac{N}{2}}} \exp \left[-\frac{1}{2\theta} \sum_{n=0}^{N-1} (x[n] - \theta)^2 \right].$$

Desarrollando el exponente como

$$\sum_{n=0}^{N-1} (x[n] - \theta)^2 = \sum_{n=0}^{N-1} x^2[n] - 2\theta \sum_{n=0}^{N-1} x[n] + N\theta^2$$

la PDF se puede expresar como

$$\begin{aligned} p(\mathbf{x}, \theta) &= \frac{1}{(2\pi\theta)^{\frac{N}{2}}} \exp \left(-\frac{1}{2\theta} \sum_{n=0}^{N-1} x^2[n] + \sum_{n=0}^{N-1} x[n] - \frac{N\theta}{2} \right) \\ &= \underbrace{\frac{1}{(2\pi\theta)^{\frac{N}{2}}} \exp \left(-\frac{1}{2\theta} \sum_{n=0}^{N-1} x^2[n] - \frac{N\theta}{2} \right)}_{g(T(\mathbf{x}), \theta)} \underbrace{\exp \left(\sum_{n=0}^{N-1} x[n] \right)}_{h(\mathbf{x})}, \end{aligned}$$

concluyendo que el estadístico suficiente es

$$T(\mathbf{x}) = \sum_{n=0}^{N-1} x^2[n].$$

5.4.5. Problema 5

Las observaciones IID $x[n]$ con $n = 0, \dots, N-1$ tienen PDF $\mathcal{U}[-\theta, \theta]$ con $\theta > 0$. Encontrar un estadístico suficiente para θ .

Solución La PDF de una muestra es

$$p(x[n]; \theta) = \frac{1}{2\theta} [u(x[n] + \theta) - u(x[n] - \theta)],$$

por lo que la PDF de los datos es

$$p(\mathbf{x}; \theta) = \frac{1}{(2\theta)^N} \prod_{n=0}^{N-1} [u(x[n] + \theta) - u(x[n] - \theta)].$$

Este producto es nulo a menos que $-\theta < x[n] < \theta$ para todo $x[n]$, o equivalentemente, $\max x[n] < \theta$ y $\min x[n] > -\theta$, o equivalentemente, $\max |x[n]| < \theta$. De esta forma, se obtiene que

$$p(\mathbf{x}, \theta) = \underbrace{\frac{1}{(2\theta)^N} u(\theta - \max |x[n]|)}_{g(T(\mathbf{x}), \theta)} \times \underbrace{1}_{h(\mathbf{x})},$$

concluyendo que el estadístico suficiente es

$$T(\mathbf{x}) = \max |x[n]|.$$

5.4.6. Problema 6

Si se observan las muestras $x[n] = A + w[n]$ con $n = 0, \dots, N-1$, donde $w[n]$ es WGN con varianza σ^2 . Encontrar el estimador MVU para σ^2 asumiendo que A es conocido. Asíumase que el estadístico suficiente es completo.

Solución La PDF de los datos es

$$p(\mathbf{x}, \sigma^2) = \underbrace{\frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - A)^2 \right]}_{g(T(\mathbf{x}), \sigma^2)} \times \underbrace{1}_{h(\mathbf{x})},$$

por lo que un estadístico suficiente es

$$T(\mathbf{x}) = \sum_{n=0}^{N-1} (x[n] - A)^2.$$

La media del estadístico es

$$E(T(\mathbf{x})) = \sum_{n=0}^{N-1} E[(x[n] - A)^2] = N\sigma^2,$$

por lo que al dividir entre N se transforma el estadístico suficiente en insesgado. Se concluye que

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{n=0}^{N-1} (x[n] - A)^2,$$

es el estimador MVU.

5.4.7. Problema 7

Se considera el problema de la estimación de la frecuencia de una senoide en WGN,

$$x[n] = \cos 2\pi f_0 n + w[n], \quad n = 0, \dots, N-1,$$

donde $w[n]$ es WGN con varianza σ^2 conocida. Mostrar que no es posible encontrar un estadístico suficiente para la frecuencia f_0 .

Solución La PDF de los datos es (ver el ejemplo de la sección 5.3.2)

$$\begin{aligned} p(\mathbf{x}, \theta) &= \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - \cos 2\pi f_0 n)^2 \right] \\ &= \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left(\sum_{n=0}^{N-1} x^2[n] + 2 \sum_{n=0}^{N-1} x[n] \cos 2\pi f_0 n + \sum_{n=0}^{N-1} \cos^2 2\pi f_0 n \right). \end{aligned}$$

Debido al término $\sum_{n=0}^{N-1} x[n] \cos 2\pi f_0 n$, la PDF parece no poder ser factorizada como indica la ecuación 5.1, ya que en dicho término no puede separarse los datos del parámetro. Dicho término no puede ser un estadístico porque depende del parámetro f_0 a estimar.

5.4.8. Problema 8

Se considera el problema de la estimación de una constante amortiguada en WGN,

$$x[n] = r^n + w[n], \quad n = 0, \dots, N-1,$$

donde $w[n]$ es WGN con varianza σ^2 conocida. Mostrar que no es posible encontrar un estadístico suficiente para la constante r .

Solución La PDF de los datos es

$$\begin{aligned} p(\mathbf{x}, \theta) &= \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - r^n)^2 \right] \\ &= \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left(\sum_{n=0}^{N-1} x^2[n] + 2 \sum_{n=0}^{N-1} x[n]r^n + \sum_{n=0}^{N-1} r^{2n} \right). \end{aligned}$$

Al igual que en el problema anterior, en el término $\sum_{n=0}^{N-1} x[n]r^n$ no puede separarse los datos del parámetro, y por lo tanto, la PDF no puede ser factorizada como indica la ecuación 5.1.

5.4.9. Problema 9

Se asume que $x[n]$ es el resultado de un ensayo Bernoulli (tirar una moneda) con

$$\begin{aligned} \Pr\{x[n] = 1\} &= \theta \\ \Pr\{x[n] = 0\} &= 1 - \theta \end{aligned}$$

y se realizan N observaciones IID. Asumiendo que el teorema de la factorización de Neyman-Fisher se cumple también para variables aleatorias discretas, encontrar un estadístico suficiente para θ . Luego, asumiendo completitud del estadístico, encontrar el estimador MVU de θ .

Solución La PDF de una muestra es

$$\Pr\{x[n]; \theta\} = \theta^{x[n]}(1 - \theta)^{1-x[n]},$$

por lo que la PDF de los datos es

$$\begin{aligned} \Pr\{\mathbf{x}; \theta\} &= \prod_{n=0}^{N-1} \theta^{x[n]}(1 - \theta)^{1-x[n]} \\ &= \theta^{\sum_{n=0}^{N-1} x[n]} (1 - \theta)^{\sum_{n=0}^{N-1} (1-x[n])} \\ &= \underbrace{\theta^{\sum_{n=0}^{N-1} x[n]} (1 - \theta)^{N - \sum_{n=0}^{N-1} x[n]}}_{g(T(\mathbf{x}), \theta)} \times \underbrace{1}_{h(\mathbf{x})}. \end{aligned}$$

Se concluye que un estadístico suficiente es

$$T(\mathbf{x}) = \sum_{n=0}^{N-1} x[n].$$

Observando que la esperanza del estadístico suficiente es

$$E[T(\mathbf{x})] = \sum_{n=0}^{N-1} E(x[n]) = NE(x[n]) \stackrel{(a)}{=} N\theta,$$

donde en (a) se tuvo en cuenta que la esperanza de una muestra es $E(x[n]) = 1 \times \theta + 0 \times (1 - \theta) = \theta$. Por lo tanto, si se divide el estadístico suficiente entre N se hace insesgado, y el estimador MVU es

$$\hat{\theta} = \frac{1}{N} \sum_{n=0}^{N-1} x[n] = \bar{x},$$

que es la media muestral.

5.4.10. Problema 10

En el problema de la estimación de la fase de una senoide en WGN (ver el ejemplo de la sección 5.1.3) se propone el siguiente estimador

$$\hat{\phi} = -\arctan \left[\frac{T_2(\mathbf{x})}{T_1(\mathbf{x})} \right].$$

Mostrar que si la SNR es grande o $\sigma^2 \rightarrow 0$, para lo cual

$$\begin{aligned} T_1(\mathbf{x}) &\approx \sum_{n=0}^{N-1} A \cos(2\pi f_0 n + \phi) \cos 2\pi f_0 n \\ T_2(\mathbf{x}) &\approx \sum_{n=0}^{N-1} A \cos(2\pi f_0 n + \phi) \sin 2\pi f_0 n, \end{aligned} \quad (5.21)$$

el estimador cumple que $\hat{\phi} \approx \phi$. Emplear las aproximaciones que se consideren apropiadas. ¿Es el estimador propuesto el estimador MVU?

Solución El modelo de los datos en este problema es

$$x[n] = A \cos(2\pi f_0 n + \phi) + w[n], \quad n = 0, \dots, N-1.$$

En el caso en que la SNR es grande o $\sigma^2 \rightarrow 0$, se puede despreciar el ruido, y por lo tanto,

$$x[n] \approx A \cos(2\pi f_0 n + \phi), \quad n = 0, \dots, N-1.$$

Sustituyendo $x[n]$ por este valor en los estadísticos conjuntamente suficientes de ϕ dados por la ecuación 5.9, se obtiene que se pueden aproximar como la ecuación 5.21.

Considerando las identidades trigonométricas

$$\begin{aligned} 2 \cos \theta \cos \varphi &= \cos(\theta - \varphi) + \cos(\theta + \varphi) \\ 2 \cos \theta \sin \varphi &= \sin(\theta + \varphi) - \sin(\theta - \varphi), \end{aligned}$$

se tiene que

$$\begin{aligned} \cos(2\pi f_0 n + \phi) \cos 2\pi f_0 n &= \frac{\cos \phi + \cos(4\pi f_0 n + \phi)}{2} \\ \cos(2\pi f_0 n + \phi) \sin 2\pi f_0 n &= \frac{\sin(4\pi f_0 n + \phi) - \sin \phi}{2}, \end{aligned}$$

por lo que la ecuación 5.21 se puede escribir como

$$\begin{aligned} T_1(\mathbf{x}) &\approx \frac{NA}{2} \cos \phi + \frac{A}{2} \sum_{n=0}^{N-1} \cos(4\pi f_0 n + \phi) \\ T_2(\mathbf{x}) &\approx \frac{A}{2} \sum_{n=0}^{N-1} \sin(4\pi f_0 n + \phi) - \frac{NA}{2} \sin \phi. \end{aligned}$$

De esta forma, se tiene que

$$\begin{aligned}
\frac{T_2(\mathbf{x})}{T_1(\mathbf{x})} &\approx \frac{\sum_{n=0}^{N-1} \sin(4\pi f_0 n + \phi) - N \sin \phi}{N \cos \phi + \sum_{n=0}^{N-1} \cos(4\pi f_0 n + \phi)} \\
&= \frac{\frac{1}{N} \sum_{n=0}^{N-1} \sin(4\pi f_0 n + \phi) - \sin \phi}{\cos \phi + \frac{1}{N} \sum_{n=0}^{N-1} \cos(4\pi f_0 n + \phi)} \\
&\stackrel{(a)}{\approx} -\frac{\sin \phi}{\cos \phi} \\
&= \tan(-\phi),
\end{aligned}$$

donde en (a) se empleó el resultado del problema de la sección 3.6.7 para despreciar las sumatorias. Por lo tanto,

$$\hat{\phi} = -\arctan \left[\frac{T_2(\mathbf{x})}{T_1(\mathbf{x})} \right] \approx -\arctan [\tan(-\phi)] = \phi,$$

que es lo que se quería demostrar.

El estimador no es MVU ya que ni siquiera es insesgado. Por un lado, se cumple que $\hat{\phi} \approx \phi$ solo si $\sigma^2 \rightarrow 0$. Además, si se pudiera asumir que

$$\begin{aligned}
E[T_1(\mathbf{x})] &= \frac{NA}{2} \cos \phi \\
E[T_2(\mathbf{x})] &= -\frac{NA}{2} \sin \phi,
\end{aligned}$$

que solo es cierto a altas SNR, se observa que

$$\begin{aligned}
E(\hat{\phi}) &= E \left\{ -\arctan \left[\frac{T_2(\mathbf{x})}{T_1(\mathbf{x})} \right] \right\} \\
&\neq -\arctan \left\{ \frac{E[T_2(\mathbf{x})]}{E[T_1(\mathbf{x})]} \right\} \\
&= \phi.
\end{aligned}$$

5.4.11. Problema 11

En el problema de la estimación del nivel de DC en WGN (ver el ejemplo de la sección 5.1.2) se desea estimar $\theta = 2A + 1$ en lugar de A . Encontrar un estimador MVU para θ . ¿Y si se deseara encontrar el estimador MVU del parámetro $\theta = A^3$? Sugerencia: reparametrizar la PDF en términos de θ y seguir el enfoque habitual.

Solución Se quiere calcular el estimador MVU de $\theta = 2A + 1$. Para hacerlo, se buscará un estadístico suficiente de θ y luego una función del estadístico que lo haga insesgado. La PDF de los datos es

$$p(\mathbf{x}; A) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - A)^2 \right],$$

y sustituyendo $A = (\theta - 1)/2$ se tiene que

$$p(\mathbf{x}; \theta) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} \left(x[n] - \frac{\theta - 1}{2} \right)^2 \right].$$

Observando que la sumatoria del exponente se puede desarrollar como

$$\sum_{n=0}^{N-1} \left(x[n] - \frac{\theta - 1}{2} \right)^2 = \sum_{n=0}^{N-1} x^2[n] - (\theta - 1) \sum_{n=0}^{N-1} x[n] + \frac{N(\theta - 1)^2}{4},$$

la PDF puede ser factorizada como

$$p(\mathbf{x}; \theta) = \underbrace{\frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left\{ -\frac{1}{2\sigma^2} \left[\frac{N(\theta - 1)^2}{4} - (\theta - 1) \sum_{n=0}^{N-1} x[n] \right] \right\}}_{g(T(\mathbf{x}), \theta)} \underbrace{\exp \left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} x^2[n] \right]}_{h(\mathbf{x})}.$$

Por lo tanto, un estadístico suficiente para θ es

$$T(\mathbf{x}) = \sum_{n=0}^{N-1} x[n].$$

Como $T(\mathbf{x})$ es completo, falta que encontrar una función g tal que $E[g(T(\mathbf{x}))] = \theta = 2A + 1$. Teniendo en cuenta que $E(T(\mathbf{x})) = NA$, se considera la función $g(x) = 1 + 2x/N$. De esta forma,

$$\hat{\theta} = 1 + \frac{2}{N} \sum_{n=0}^{N-1} x[n].$$

Como $E(\hat{\theta}) = 2A + 1$, $\hat{\theta}$ es el estimador MVU.

Se buscará ahora el estimador MVU de $\theta = A^3$ siguiendo el mismo enfoque. Sustituyendo $A = \theta^{\frac{1}{3}}$ en la PDF de los datos, se tiene que

$$p(\mathbf{x}; \theta) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - \theta^{\frac{1}{3}})^2 \right],$$

y desarrollando el exponente, se ve que la PDF se puede factorizar como

$$p(\mathbf{x}; \theta) = \underbrace{\frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left[-\frac{1}{2\sigma^2} \left(\theta^{\frac{2}{3}} - 2A\theta^{\frac{1}{3}} \sum_{n=0}^{N-1} x[n] \right) \right]}_{g(T(\mathbf{x}), \theta)} \underbrace{\exp \left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} x^2[n] \right]}_{h(\mathbf{x})},$$

por lo que $T(\mathbf{x}) = \sum_{n=0}^{N-1} x[n]$ es un estadístico suficiente para θ . Para hacerlo insesgado hay que encontrar una función g tal que $E[g(T(\mathbf{x}))] = A^3$, o equivalentemente, una función h tal que $E[h(\bar{x})] = A^3$, ya que \bar{x} también es un estadístico suficiente para θ . Para intentar deducir la función h se calculará $E(\bar{x}^3)$. Teniendo en cuenta que el momento tercero de una variable aleatoria $x \sim \mathcal{N}(\mu, \sigma^2)$ es $E(x^3) = \mu^3 + 3\mu\sigma^2$ y que $\bar{x} \sim \mathcal{N}(A, \sigma^2/N)$, se cumple que

$$E(\bar{x}^3) = A^3 + 3A \frac{\sigma^2}{N}.$$

Esto indica que si se considerara a \bar{x}^3 como estimador de A^3 , habría un sesgo de $3A\sigma^2/N$. Esto motiva a considerar como función h a

$$h(x) = x^3 - 3x \frac{\sigma^2}{N}.$$

De esta forma,

$$\hat{\theta} = h(\bar{x}) = \bar{x}^3 - 3\bar{x} \frac{\sigma^2}{N},$$

con

$$E(\hat{\theta}) = E(\bar{x}^3) - 3E(\bar{x}) \frac{\sigma^2}{N} = \left(A^3 + 3A \frac{\sigma^2}{N} \right) - 3A \frac{\sigma^2}{N} = A^3,$$

concluyendo que la función h propuesta hace al estadístico suficiente \bar{x} insesgado y por lo tanto, $\hat{\theta}$ es el estimador MVU.

5.4.12. Problema 12

Este problema examina la idea de que los estadísticos suficientes son únicos a menos de transformaciones uno a uno. Considérese el problema de la estimación del nivel de DC en WGN del ejemplo de la sección 5.1.2 pero con los estadísticos suficientes

$$T_1(\mathbf{x}) = \sum_{n=0}^{N-1} x[n]$$

$$T_2(\mathbf{x}) = \left(\sum_{n=0}^{N-1} x[n] \right)^3.$$

Encontrar funciones g_1 y g_2 tales que se satisfaga la condición de insesgado, o

$$E[g_1(T_1(\mathbf{x}))] = A$$

$$E[g_2(T_2(\mathbf{x}))] = A,$$

haciendo que los estadísticos suficientes transformados sean los estimadores MVU. ¿Es la transformación de T_1 a T_2 uno a uno? ¿Qué pasaría si se considerara el estadístico $T_3(\mathbf{x}) = (\sum_{n=0}^{N-1} x[n])^2$?

Solución Puede verse que las funciones g_1 y g_2 que hacen a los estadísticos insesgados son

$$g_1(x) = \frac{x}{N}, \quad g_2(x) = \frac{x^{1/3}}{N},$$

y la transformación de T_1 a T_2 es uno a uno. En el caso del estadístico $T_3(\mathbf{x})$, la transformación de T_1 a T_3 no es uno a uno, ya que las secuencias $x[n]$ y $-x[n]$ para todo $n = 0, \dots, N-1$, producen valores distintos de $T_1(\mathbf{x})$ pero valores iguales de $T_3(\mathbf{x})$, y por lo tanto, $T_3(\mathbf{x})$ no es un estadístico suficiente. Se podría considerar la función $g_3(x) = \sqrt{x}/N$ para intentar hacer a $T_3(\mathbf{x})$ insesgado, pero en ese caso,

$$E[g_3(T_3(\mathbf{x}))] = E\left[\frac{\sqrt{(N\bar{x})^2}}{N}\right] = E[|\bar{x}|] = \begin{cases} A, & A \geq 0 \\ -A, & A < 0 \end{cases} \neq A \quad \forall A.$$

5.4.13. Problema 13

En este problema se deriva el estimador MVU para un ejemplo que no se ha trabajado previamente. Si se realizan N observaciones IID de acuerdo a la PDF

$$p(x[n]; \theta) = \begin{cases} \exp[-(x[n] - \theta)], & x[n] > \theta \\ 0, & x[n] < \theta, \end{cases}$$

encontrar el estimador MVU de θ . Notar que θ representa el mínimo valor que puede alcanzar $x[n]$. Asumir que el estadístico suficiente es completo.

Solución Para encontrar el estimador MVU, se buscará un estadístico suficiente de θ y luego una transformación que lo haga insesgado. La PDF de una muestra se puede escribir como

$$p(x[n]; \theta) = u(x[n] - \theta) \exp[-(x[n] - \theta)].$$

Por lo tanto, la PDF de los datos es

$$p(\mathbf{x}; \theta) = \prod_{n=0}^{N-1} u(x[n] - \theta) \exp[-(x[n] - \theta)]$$

$$= u(\min x[n] - \theta) \exp\left[-\sum_{n=0}^{N-1} (x[n] - \theta)\right]$$

$$= u(\min x[n] - \theta) \exp\left(N\theta - \sum_{n=0}^{N-1} x[n]\right),$$

que se puede factorizar como

$$p(\mathbf{x}; \theta) = \underbrace{u(\min x[n] - \theta) \exp(N\theta)}_{g(T(\mathbf{x}), \theta)} \underbrace{\exp\left(-\sum_{n=0}^{N-1} x[n]\right)}_{h(\mathbf{x})},$$

concluyendo que un estadístico suficiente de θ es

$$T(\mathbf{x}) = \min x[n].$$

Para buscar una transformación que lo haga insesgado, se procede de forma análoga al procedimiento del ejemplo de la sección 5.2.3, el cual implica calcular la media de $T = \min x[n]$, y por lo tanto, su densidad de probabilidad. La distribución de probabilidad de T es

$$\begin{aligned} \Pr\{T \leq \xi\} &= 1 - \Pr\{T > \xi\} \\ &= 1 - \Pr\{x[0] > \xi, x[1] > \xi, \dots, x[N-1] > \xi\} \\ &= 1 - \Pr\{x[n] > \xi\}^N \\ &= 1 - (1 - \Pr\{x[n] \leq \xi\})^N, \end{aligned}$$

y derivando, se obtiene la densidad de probabilidad,

$$\begin{aligned} p_T(\xi) &= \frac{d\Pr\{T \leq \xi\}}{d\xi} \\ &= -N(1 - \Pr\{x[n] \leq \xi\})^{N-1} \left(-\frac{d\Pr\{x[n] \leq \xi\}}{d\xi} \right) \\ &= N(1 - \Pr\{x[n] \leq \xi\})^{N-1} p_{x[n]}(\xi). \end{aligned} \tag{5.22}$$

La distribución de probabilidad de $x[n]$ se obtiene integrando la densidad de probabilidad, y para $\xi > \theta$ es

$$\Pr\{x[n] \leq \xi\} = \int_{-\infty}^{\xi} p_{x[n]}(u; \theta) du = \int_{\theta}^{\xi} e^{-(u-\theta)} du = -e^{-(u-\theta)} \Big|_{\theta}^{\xi} = -[e^{-(\xi-\theta)} - 1] = 1 - e^{-(\xi-\theta)},$$

resultando en

$$\Pr\{x[n] \leq \xi\} = \begin{cases} 1 - e^{-(\xi-\theta)}, & \xi > \theta \\ 0, & \xi < \theta. \end{cases}$$

Sustituyendo el resultado en la ecuación de la densidad de probabilidad de T se tiene que para $\xi > \theta$,

$$p_T(\xi) = N\{1 - [1 - e^{-(\xi-\theta)}]\}^{N-1} e^{-(\xi-\theta)} = N e^{-(\xi-\theta)(N-1)} e^{-(\xi-\theta)} = N e^{-N(\xi-\theta)},$$

es decir,

$$p_T(\xi) = \begin{cases} N e^{-N(\xi-\theta)}, & \xi > \theta \\ 0, & \xi < \theta. \end{cases}$$

Una vez obtenida la PDF de T es posible calcular su media,

$$\begin{aligned} E(T) &= \int_{-\infty}^{\infty} \xi p_T(\xi) d\xi \\ &= \int_{\theta}^{\infty} \xi N e^{-N(\xi-\theta)} d\xi \\ &\stackrel{(a)}{=} -\xi e^{-N(\xi-\theta)} \Big|_{\theta}^{\infty} + \int_{\theta}^{\infty} e^{-N(\xi-\theta)} d\xi \\ &= \theta + \left[-\frac{1}{N} e^{-N(\xi-\theta)} \Big|_{\theta}^{\infty} \right] \\ &= \theta + \frac{1}{N}, \end{aligned}$$

donde en (a) se aplicó integración por partes $\int u dv = uv - \int v du$ con $u = \xi$ y $dv = Ne^{-N(\xi-\theta)}$ y por lo tanto, $du = 1$ y $v = -e^{-N(\xi-\theta)}$. Se concluye que

$$E(\min x[n]) = \theta + \frac{1}{N},$$

obteniendo que

$$\hat{\theta} = \min x[n] - \frac{1}{N}$$

es insesgado y por lo tanto, el estimador MVU.

5.4.14. Problema 14

Considérese una variable aleatoria x con PDF de la forma

$$p(x; \theta) = \exp[A(\theta)B(x) + C(x) + D(\theta)].$$

Una PDF de este tipo se dice que pertenece a la *familia exponencial escalar*⁵ de PDF. Muchas propiedades útiles están asociadas a esta familia, y en particular, varias vinculadas a estadísticos suficientes, lo cual se explora en este problema y en el siguiente. Mostrar que las siguientes PDF pertenecen a dicha familia.

a. Gaussiana

$$p(x; \mu) = \frac{1}{\sqrt{2\pi}} \exp \left[-\frac{1}{2}(x - \mu)^2 \right]$$

b. Rayleigh

$$p(x; \sigma^2) = \begin{cases} \frac{x}{\sigma^2} \exp \left[-\frac{1}{2} \frac{x^2}{\sigma^2} \right], & x > 0 \\ 0, & x < 0 \end{cases}$$

c. Exponencial

$$p(x; \lambda) = \begin{cases} \lambda \exp[-\lambda x], & x > 0 \\ 0, & x < 0 \end{cases}$$

d. Laplace

$$p(x; \sigma) = \frac{1}{\sqrt{2}\sigma} \exp \left[-\frac{\sqrt{2}|x - \mu|}{\sigma} \right]$$

Solución

a. Gaussiana

$$\begin{aligned} p(x; \mu) &= \frac{1}{\sqrt{2\pi}} \exp \left[-\frac{1}{2}(x - \mu)^2 \right] \\ &= \exp \left[\ln \left(\frac{1}{\sqrt{2\pi}} \right) \right] \exp \left[-\frac{1}{2}(x^2 - 2x\mu + \mu^2) \right] \\ &= \exp \left[-\frac{1}{2} \ln 2\pi \right] \exp \left[-\frac{1}{2}x^2 + x\mu - \frac{1}{2}\mu^2 \right], \end{aligned}$$

resultando en

$$p(x; \mu) = \exp \left[\mu x - \frac{1}{2}x^2 - \frac{1}{2}(\mu^2 + \ln 2\pi) \right],$$

y por lo tanto, es de la familia exponencial con

$$A(\mu)B(x) = \mu x, \quad C(x) = -\frac{1}{2}x^2, \quad D(\mu) = -\frac{1}{2}(\mu^2 + \ln 2\pi).$$

⁵Ver https://en.wikipedia.org/wiki/Exponential_family

b. Rayleigh

$$\begin{aligned} p(x; \sigma^2) &= u(x) \frac{x}{\sigma^2} \exp \left[-\frac{1}{2} \frac{x^2}{\sigma^2} \right] \\ &= \exp [\ln xu(x) - \ln \sigma^2] \exp \left[-\frac{1}{2} \frac{x^2}{\sigma^2} \right], \end{aligned}$$

resultando en

$$p(x; \sigma^2) = \exp \left[-\frac{1}{2} \frac{x^2}{\sigma^2} + \ln xu(x) - \ln \sigma^2 \right]$$

y por lo tanto, es de la familia exponencial con

$$A(\sigma^2)B(x) = -\frac{1}{2} \frac{x^2}{\sigma^2}, \quad C(x) = \ln xu(x), \quad D(\sigma^2) = -\ln \sigma^2.$$

c. Exponencial

$$\begin{aligned} p(x; \lambda) &= u(x) \lambda \exp [-\lambda x] \\ &= \exp [\ln u(x) + \ln \lambda] \exp [-\lambda x], \end{aligned}$$

resultando en

$$p(x; \lambda) = \exp [-\lambda x + \ln u(x) + \ln \lambda],$$

y por lo tanto, es de la familia exponencial con

$$A(\lambda)B(x) = -\lambda x, \quad C(x) = \ln u(x), \quad D(\lambda) = \ln \lambda.$$

d. Laplace

$$\begin{aligned} p(x; \sigma) &= \frac{1}{\sqrt{2}\sigma} \exp \left[-\frac{\sqrt{2}|x - \mu|}{\sigma} \right] \\ &= \exp [\ln \sqrt{2}\sigma] \exp \left[-\frac{\sqrt{2}|x - \mu|}{\sigma} \right] \end{aligned}$$

resultando en

$$p(x; \sigma) = \exp \left[\ln \sqrt{2}\sigma - \frac{\sqrt{2}|x - \mu|}{\sigma} \right]$$

y por lo tanto, es de la familia exponencial con

$$A(\sigma)B(x) = \frac{\sqrt{2}}{\sigma} |x - \mu|, \quad C(x) = 0, \quad D(\sigma) = \ln \sqrt{2}\sigma.$$

5.4.15. Problema 15

Si se observan las muestras $x[n]$ para $n = 0, \dots, N-1$, las cuales son IID con PDF de la familia exponencial, mostrar que

$$T(\mathbf{x}) = \sum_{n=0}^{N-1} B(x[n])$$

es un estadístico suficiente para θ . Luego, aplicar este resultado para determinar un estadístico suficiente para las PDF Gaussiana, Rayleigh, exponencial y Laplace del problema anterior. Finalmente, asumiendo que los estadísticos suficientes son completos, encontrar un estimador en cada caso si es posible.

Solución La PDF de cada muestra $x[n]$ pertenece a la familia de PDFs exponenciales, y por lo tanto es de la forma

$$p(x[n]; \theta) = \exp[A(\theta)B(x[n]) + C(x[n]) + D(\theta)].$$

Como las muestras son IID, la PDF de los datos es

$$\begin{aligned} p(x[n]; \theta) &= \prod_{n=0}^{N-1} \exp[A(\theta)B(x[n]) + C(x[n]) + D(\theta)] \\ &= \exp \left\{ \sum_{n=0}^{N-1} [A(\theta)B(x[n]) + C(x[n]) + D(\theta)] \right\} \\ &= \exp \left[A(\theta) \sum_{n=0}^{N-1} B(x[n]) + \sum_{n=0}^{N-1} C(x[n]) + ND(\theta) \right], \end{aligned}$$

que puede factorizarse como

$$p(x[n]; \theta) = \underbrace{\exp \left[A(\theta) \sum_{n=0}^{N-1} B(x[n]) + ND(\theta) \right]}_{g(T(\mathbf{x}), \theta)} \underbrace{\exp \left[\sum_{n=0}^{N-1} C(x[n]) \right]}_{h(\mathbf{x})},$$

y por lo tanto,

$$T(\mathbf{x}) = \sum_{n=0}^{N-1} B(x[n])$$

es un estadístico suficiente para θ . A partir de este resultado y empleando el resultado del problema anterior se tiene que los estadísticos suficientes son

a. Gaussiana

$$T(\mathbf{x}) = \sum_{n=0}^{N-1} x[n]$$

b. Rayleigh

$$T(\mathbf{x}) = \sum_{n=0}^{N-1} x^2[n]$$

c. Exponencial

$$T(\mathbf{x}) = \sum_{n=0}^{N-1} x[n]$$

d. Laplace

$$T(\mathbf{x}) = \sum_{n=0}^{N-1} |x[n] - \mu|$$

Se buscará a partir del estadístico suficiente el estimador MVU en cada caso.

a. Gaussiana

$$\hat{\mu} = \frac{1}{N} \sum_{n=0}^{N-1} x[n],$$

como ya se vió previamente.

b. Rayleigh. La esperanza del estadístico suficiente es

$$E[T(\mathbf{x})] = NE(x^2[n]).$$

Para completar el resultado, hay que calcular el segundo momento,

$$\begin{aligned}
E(x^2) &= \int_{-\infty}^{\infty} x^2 p(x; \sigma^2) dx \\
&= \frac{1}{\sigma^2} \int_0^{\infty} x^3 e^{-x^2/2\sigma^2} dx \\
&\stackrel{(a)}{=} \frac{1}{2\sigma^2} \int_0^{\infty} y e^{-y/2\sigma^2} dy \\
&\stackrel{(b)}{=} -y e^{-y/2\sigma^2} \Big|_0^{\infty} + \int_0^{\infty} e^{-y/2\sigma^2} dy \\
&= -2\sigma^2 e^{-y/2\sigma^2} \Big|_0^{\infty} \\
&= 2\sigma^2,
\end{aligned}$$

donde en (a) se realizó el cambio de variable $y = x^2$, $dy = 2x dx$, y en (b) se aplicó integración por partes $\int u dv = uv - \int v du$ con $u = y$ y $dv = (1/2\sigma^2)e^{-y/2\sigma^2}$ y por lo tanto, $du = 1$ y $v = -e^{-y/2\sigma^2}$. Sustituyendo este resultado en la esperanza del estadístico suficiente se obtiene que

$$E[T(\mathbf{x})] = NE(x^2[n]) = 2N\sigma^2,$$

y por lo tanto, el estimador MVU de σ^2 es

$$\hat{\sigma}^2 = \frac{1}{2N} \sum_{n=0}^{N-1} x^2[n].$$

c. Exponencial. La esperanza del estadístico es

$$E[T(\mathbf{x})] = NE(x[n]),$$

con

$$\begin{aligned}
E(x) &= \int_0^{\infty} x \lambda e^{-\lambda x} dx \\
&\stackrel{(a)}{=} -x e^{-\lambda x} \Big|_0^{\infty} + \int_0^{\infty} e^{-\lambda x} dx \\
&= -\frac{1}{\lambda} e^{-\lambda x} \Big|_0^{\infty} \\
&= \frac{1}{\lambda},
\end{aligned}$$

donde en (a) se aplicó integración por partes $\int u dv = uv - \int v du$ con $u = x$ y $dv = \lambda e^{-\lambda x}$ y por lo tanto, $du = 1$ y $v = -e^{-\lambda x}$. De esta forma, se tiene que

$$E[T(\mathbf{x})] = NE(x[n]) = \frac{N}{\lambda}.$$

En este caso, no parece evidente una transformación del estadístico que lo haga insesgado. Se podría pensar en

$$\hat{\lambda} = \frac{N}{T(\mathbf{x})},$$

pero

$$E[\hat{\lambda}] = E\left[\frac{N}{T(\mathbf{x})}\right] \neq \frac{N}{E[T(\mathbf{x})]} = \lambda.$$

Sin embargo, si se reparametriza la PDF con $\theta = 1/\lambda$, se tiene que

$$p(x[n]) = \frac{u(x[n])}{\theta} e^{-x[n]/\theta}$$

y

$$p(\mathbf{x}, \theta) = \underbrace{\frac{1}{\theta^N} \exp\left(-\frac{1}{\theta} \sum_{n=0}^{N-1} x[n]\right)}_{g(T(\mathbf{x}), \theta)} \underbrace{u(\min x[n])}_{h(\mathbf{x})},$$

por lo que un estadístico de θ es

$$T(\mathbf{x}) = \sum_{n=0}^{N-1} x[n].$$

y

$$E[T(\mathbf{x})] = \frac{N}{\lambda} = N\theta$$

y por lo tanto,

$$\hat{\theta} = \frac{1}{N} \sum_{n=0}^{N-1} x[n]$$

es el estimador MVU de θ .

d. Laplace. La esperanza del estadístico es

$$E[T(\mathbf{x})] = NE(|x[n] - \mu|)$$

con

$$\begin{aligned} E(|u - \mu|) &= \int_{-\infty}^{\infty} |u - \mu| \frac{1}{\sqrt{2}\sigma} e^{-\sqrt{2}|u-\mu|/\sigma} du \\ &\stackrel{(a)}{=} \frac{1}{\sqrt{2}\sigma} \int_{-\infty}^{\infty} |x| e^{-\sqrt{2}|x|/\sigma} dx \\ &\stackrel{(b)}{=} \frac{\sqrt{2}}{\sigma} \int_0^{\infty} x e^{-\sqrt{2}x/\sigma} dx \\ &\stackrel{(c)}{=} -x e^{-\sqrt{2}x/\sigma} \Big|_0^{\infty} + \int_0^{\infty} e^{-\sqrt{2}x/\sigma} dx \\ &= -\frac{\sigma}{\sqrt{2}} e^{-\sqrt{2}x/\sigma} \Big|_0^{\infty} \\ &= \frac{\sigma}{\sqrt{2}}, \end{aligned}$$

donde en (a) se realizó el cambio de variable $x = u - \mu$, en (b) se tuvo en cuenta que el integrando es una función par y en (c) se aplicó integración por partes $\int u dv = uv - \int v du$ con $u = x$ y $dv = (\sqrt{2}/\sigma) e^{-\sqrt{2}x/\sigma} dx$ y por lo tanto, $du = dx$ y $v = -e^{-\sqrt{2}x/\sigma}$. Por lo tanto,

$$E[T(\mathbf{x})] = N \frac{\sigma}{\sqrt{2}},$$

resultando en que

$$\hat{\sigma} = \frac{\sqrt{2}}{N} \sum_{n=0}^{N-1} |x[n] - \mu|$$

es el estimador MVU.

5.4.16. Problema 16

En este problema se da un ejemplo donde la cantidad de estadísticos suficientes es menor que el número de parámetros a estimar. Asíumase que se observa $x[n]$ para $n = 0, \dots, N-1$ con $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$

con

$$\boldsymbol{\mu} = \begin{bmatrix} N\mu \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad y \quad \mathbf{C} = \begin{bmatrix} N-1+\sigma^2 & -\mathbf{1}^T \\ -\mathbf{1} & \mathbf{I} \end{bmatrix}.$$

La matriz \mathbf{C} de covarianza tiene dimensiones

$$\begin{bmatrix} 1 \times 1 & 1 \times (N-1) \\ (N-1) \times 1 & (N-1) \times (N-1) \end{bmatrix}.$$

Mostrar que con $\boldsymbol{\theta} = [\mu \sigma^2]$,

$$p(\mathbf{x}; \boldsymbol{\theta}) = \frac{1}{(2\pi)^{\frac{N}{2}} \sigma} \exp \left\{ -\frac{1}{2} \left[\frac{N^2}{\sigma^2} (\bar{x} - \mu)^2 + \sum_{n=1}^{N-1} x^2[n] \right] \right\}.$$

¿Cuales son los estadísticos suficientes para $\boldsymbol{\theta}$? Sugerencia: se necesita encontrar la inversa y el determinante de una matriz en bloques y emplear la identidad de Woodbury.

Solución La PDF de los datos es

$$p(\mathbf{x}, \boldsymbol{\theta}) = \frac{1}{(2\pi)^{\frac{N}{2}} \sqrt{\det(\mathbf{C})}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \mathbf{C}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\}. \quad (5.23)$$

Para continuar el desarrollo es necesario calcular el determinante y la inversa de la matriz \mathbf{C} de covarianza. Como se indica en el apéndice G el determinante de una matriz en bloques es (ecuación G.1)

$$\det \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} = \det(\mathbf{D}) \times \det(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})$$

(no confundir a \mathbf{C} con la matriz de covarianza en esta identidad). En el caso que se quiere resolver, $\mathbf{A} = N-1+\sigma^2$, $\mathbf{B} = -\mathbf{1}^T$, $\mathbf{C} = -\mathbf{1}$ y $\mathbf{D} = \mathbf{I}$ y por lo tanto,

$$\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C} = N-1+\sigma^2 - \mathbf{1}^T \mathbf{I} \mathbf{1} = N-1+\sigma^2 - \mathbf{1}^T \mathbf{1} = N-1+\sigma^2 - (N-1) = \sigma^2,$$

resultando en que el determinante de la matriz \mathbf{C} de covarianza es

$$\det(\mathbf{C}) = \det(\mathbf{I}) \times \det(\sigma^2) = \sigma^2. \quad (5.24)$$

Para calcular la inversa de la matriz de covarianza se usa la ecuación G.2 del apéndice G, que indica que

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} (\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} & -(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1} \\ -\mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} & \mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1} \end{bmatrix}.$$

En este caso se tiene que

$$\begin{aligned} (\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} &= \frac{1}{\sigma^2}, & -(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1} &= -\frac{1}{\sigma^2} \mathbf{1}^T \\ -\mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} &= -\frac{1}{\sigma^2} \mathbf{1}, & \mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1} &= \mathbf{I} + \frac{1}{\sigma^2} \mathbf{1}\mathbf{1}^T, \end{aligned}$$

resultando en que la matriz de covarianza inversa es

$$\mathbf{C} = \frac{1}{\sigma^2} \begin{bmatrix} 1 & \mathbf{1}^T \\ \mathbf{1} & \sigma^2 \mathbf{I} + \mathbf{1}\mathbf{1}^T \end{bmatrix}.$$

A continuación, se calculará la expresión del exponente de la PDF. Para eso, se observa que

$$\mathbf{C}(\mathbf{x} - \boldsymbol{\mu}) = \frac{1}{\sigma^2} \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & \sigma^2 + 1 & 1 & \dots & 1 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 1 \\ 1 & \dots & \dots & 1 & \sigma^2 + 1 \end{bmatrix} \begin{bmatrix} x[0] - N\mu \\ x[1] \\ \vdots \\ x[N-1] \end{bmatrix} = \frac{1}{\sigma^2} \begin{bmatrix} \sum_{n=0}^{N-1} x[n] - N\mu \\ \sum_{n=0}^{N-1} x[n] + x[1]\sigma^2 - N\mu \\ \vdots \\ \sum_{n=0}^{N-1} x[n] + x[N-1]\sigma^2 - N\mu \end{bmatrix}.$$

Luego,

$$(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{C}(\mathbf{x} - \boldsymbol{\mu}) = [x[0] - N\mu \quad x[1] \quad \dots \quad x[N-1]] \frac{1}{\sigma^2} \begin{bmatrix} \sum_{n=0}^{N-1} x[n] - N\mu \\ \sum_{n=0}^{N-1} x[n] + x[1]\sigma^2 - N\mu \\ \vdots \\ \sum_{n=0}^{N-1} x[n] + x[N-1]\sigma^2 - N\mu \end{bmatrix}$$

resulta en

$$\begin{aligned} (\mathbf{x} - \boldsymbol{\mu})^T \mathbf{C}(\mathbf{x} - \boldsymbol{\mu}) &= \frac{1}{\sigma^2} \left[(x[0] - N\mu) \left(\sum_{n=0}^{N-1} x[n] - N\mu \right) + \sum_{m=1}^{N-1} x[m] \left(\sum_{n=0}^{N-1} x[n] + x[m]\sigma^2 - N\mu \right) \right] \\ &= \frac{1}{\sigma^2} \left[x[0] \sum_{n=0}^{N-1} x[n] - N\mu x[0] - N\mu \sum_{n=0}^{N-1} x[n] + N^2 \mu^2 \right. \\ &\quad \left. + \sum_{m=1}^{N-1} x[m] \sum_{n=0}^{N-1} x[n] + \sigma^2 \sum_{m=1}^{N-1} x^2[m] - N\mu \sum_{m=1}^{N-1} x[m] \right] \\ &\stackrel{(a)}{=} \frac{1}{\sigma^2} \left[-N\mu \sum_{n=0}^{N-1} x[n] + N^2 \mu^2 + \sum_{m=0}^{N-1} x[m] \sum_{n=0}^{N-1} x[n] + \sigma^2 \sum_{m=1}^{N-1} x^2[m] - N\mu \sum_{m=0}^{N-1} x[m] \right] \\ &= \frac{1}{\sigma^2} \left[N^2 \mu^2 - 2N\mu \sum_{n=0}^{N-1} x[n] + \left(\sum_{n=0}^{N-1} x[n] \right)^2 + \sigma^2 \sum_{n=1}^{N-1} x^2[n] \right], \end{aligned}$$

donde en (a) se agrupó en una única sumatoria el primer sumando con el quinto sumando y el segundo sumando con el último sumando. Luego, operando, se tiene que

$$\begin{aligned} (\mathbf{x} - \boldsymbol{\mu})^T \mathbf{C}(\mathbf{x} - \boldsymbol{\mu}) &= \left\{ \frac{N^2}{\sigma^2} \left[\mu^2 - 2\mu \frac{1}{N} \sum_{n=0}^{N-1} x[n] + \left(\frac{1}{N} \sum_{n=0}^{N-1} x[n] \right)^2 \right] + \sum_{n=1}^{N-1} x^2[n] \right\} \\ &= \left[\frac{N^2}{\sigma^2} (\mu^2 - 2\mu \bar{x} + \bar{x}^2)^2 + \sum_{n=1}^{N-1} x^2[n] \right] \\ &= \left[\frac{N^2}{\sigma^2} (\mu - \bar{x})^2 + \sum_{n=1}^{N-1} x^2[n] \right]. \end{aligned} \tag{5.25}$$

Finalmente, sustituyendo las ecuaciones 5.24 y 5.25 en la ecuación 5.23, se obtiene que la PDF de los datos se puede expresar como

$$p(\mathbf{x}, \boldsymbol{\theta}) = \frac{1}{(2\pi)^{\frac{N}{2}} \sigma} \exp \left\{ -\frac{1}{2} \left[\frac{N^2}{\sigma^2} (\mu - \bar{x})^2 + \sum_{n=1}^{N-1} x^2[n] \right] \right\},$$

y observando que se puede factorizar como

$$p(\mathbf{x}, \boldsymbol{\theta}) = \underbrace{\frac{1}{(2\pi)^{\frac{N}{2}}\sigma} \exp\left\{-\frac{N^2}{2\sigma^2}(\mu - \bar{x})^2\right\}}_{g(T(\mathbf{x}), \boldsymbol{\theta})} \underbrace{\exp\left\{-\frac{1}{2} \sum_{n=1}^{N-1} x^2[n]\right\}}_{h(\mathbf{x})},$$

se concluye que $T(\mathbf{x}) = \bar{x}$ es un estadístico suficiente de $\boldsymbol{\theta} = [\mu \sigma^2]^T$.

5.4.17. Problema 17

Se considera una senoide de frecuencia conocida en WGN, o

$$x[n] = A \cos 2\pi f_0 n + w[n], \quad n = 0, \dots, N-1,$$

donde $w[n]$ es WGN de varianza σ^2 . Encontrar el estimador MVU de

- la amplitud A asumiendo que σ^2 es conocido
- la amplitud A y la varianza σ^2 del ruido.

Se puede asumir que el estadístico suficiente es completo. Sugerencia: los resultados de los ejemplos de las secciones 5.3.2 y 5.3.5 pueden ser útiles.

Solución

- Se considera el caso en que σ^2 es conocido y se quiere calcular A . Como se determinó en el ejemplo de la sección 5.3.2, la PDF de los datos es

$$p(\mathbf{x}, A) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left[-\frac{1}{2\sigma^2} \left(\sum_{n=0}^{N-1} x^2[n] - 2A \sum_{n=0}^{N-1} x[n] \cos 2\pi f_0 n + A^2 \sum_{n=0}^{N-1} \cos^2 2\pi f_0 n \right) \right]$$

y como en este caso σ^2 es conocido, se puede factorizar como

$$p(\mathbf{x}, A) = \underbrace{\frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left[-\frac{1}{2\sigma^2} \left(-2A \sum_{n=0}^{N-1} x[n] \cos 2\pi f_0 n + A^2 \sum_{n=0}^{N-1} \cos^2 2\pi f_0 n \right) \right]}_{g(T(\mathbf{x}), A)} \times \underbrace{\exp \left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} x^2[n] \right]}_{h(\mathbf{x})},$$

concluyendo que el estadístico suficiente es

$$T(\mathbf{x}) = \sum_{n=0}^{N-1} x[n] \cos 2\pi f_0 n.$$

El estimador MVU es una función del estadístico suficiente que lo haga insesgado. Para intentar encontrarla, se calcula la media del estadístico,

$$\begin{aligned} E(T(\mathbf{x})) &= \sum_{n=0}^{N-1} E(x[n]) \cos 2\pi f_0 n \\ &\stackrel{(a)}{=} A \sum_{n=0}^{N-1} \cos^2 2\pi f_0 n, \end{aligned}$$

donde en (a) se empleó que $E(x[n]) = A \cos 2\pi f_0 n$. Se observa que con $g(x) = x / \sum_{n=0}^{N-1} \cos^2 2\pi f_0 n$ se tiene que

$$\hat{A} = g(T(\mathbf{x})) = \frac{\sum_{n=0}^{N-1} x[n] \cos 2\pi f_0 n}{\sum_{n=0}^{N-1} \cos^2 2\pi f_0 n}$$

es insesgado y por lo tanto, MVU.

- b. En el caso en que A y σ^2 son desconocidos, se encontró en el ejemplo de la sección 5.3.2 que el estadístico suficiente es

$$\mathbf{T}(\mathbf{x}) = \begin{bmatrix} T_1(\mathbf{x}) \\ T_2(\mathbf{x}) \end{bmatrix} = \begin{bmatrix} \sum_{n=0}^{N-1} x[n] \cos^2 2\pi f_0 n \\ \sum_{n=0}^{N-1} x^2[n] \end{bmatrix}.$$

Para encontrar una función que lo haga insesgado,

$$\mathbf{g}(\mathbf{T}(\mathbf{x})) = \begin{bmatrix} g_1(T_1(\mathbf{x}), T_2(\mathbf{x})) \\ g_2(T_1(\mathbf{x}), T_2(\mathbf{x})) \end{bmatrix}$$

se seguirá el procedimiento análogo al ejemplo de la sección 5.3.5. Al igual que en el caso anterior, la función $g_1(T_1(\mathbf{x})) = T_1(\mathbf{x}) / \sum_{n=0}^{N-1} \cos^2 2\pi f_0 n$ hace a $T_1(\mathbf{x})$ un estimador insesgado de A , resultando en que el estimador MVU de A es

$$\hat{A} = g_1(T_1(\mathbf{x})) = \frac{\sum_{n=0}^{N-1} x[n] \cos 2\pi f_0 n}{\sum_{n=0}^{N-1} \cos^2 2\pi f_0 n}.$$

Para buscar una función g_2 de $T_1(\mathbf{x})$ y $T_2(\mathbf{x})$ que provea un estimador insesgado de σ^2 , se parte calculando la esperanza de $T_2(\mathbf{x})$,

$$\begin{aligned} E(T_2(\mathbf{x})) &= \sum_{n=0}^{N-1} E(x^2[n]) \\ &\stackrel{(a)}{=} \sum_{n=0}^{N-1} (\sigma^2 + A^2 \cos^2 2\pi f_0 n) \\ &= N\sigma^2 + A^2 \sum_{n=0}^{N-1} \cos^2 2\pi f_0 n, \end{aligned}$$

donde en (a) se tuvo en cuenta que $x[n] \sim \mathcal{N}(A \cos 2\pi f_0 n, \sigma^2)$ y $\sigma^2 = E(x^2[n]) - E^2(x[n])$ y por lo tanto, $E(x^2[n]) = \sigma^2 + A^2 \cos^2 2\pi f_0 n$. Considerando que $E(T_1(\mathbf{x})) = A \sum_{n=0}^{N-1} \cos^2 2\pi f_0 n$, se calculará

$E(T_1^2(\mathbf{x}))$ para ver si sirve para eliminar el término $A^2 \sum_{n=0}^{N-1} \cos^2 2\pi f_0 n$ en la esperanza de $T_2(\mathbf{x})$,

$$\begin{aligned}
E(T_1^2(\mathbf{x})) &= E \left[\left(\sum_{n=0}^{N-1} x[n] \cos 2\pi f_0 n \right)^2 \right] \\
&= E \left(\sum_{m=0}^{N-1} x[m] \cos 2\pi f_0 m \sum_{n=0}^{N-1} x[n] \cos 2\pi f_0 n \right) \\
&= E \left(\sum_{m=0}^{N-1} \sum_{n=0}^{N-1} x[m] x[n] \cos 2\pi f_0 m \cos 2\pi f_0 n \right) \\
&= \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} E(x[m] x[n]) \cos 2\pi f_0 m \cos 2\pi f_0 n.
\end{aligned}$$

Teniendo en cuenta que

$$E(x[m] x[n]) = \begin{cases} E(x[m]) E(x[n]) = A^2 \cos 2\pi f_0 m \cos 2\pi f_0 n, & m \neq n \\ E(x^2[n]) = \sigma^2 + A^2 \cos^2 2\pi f_0 n & m = n \end{cases}$$

y sustituyendo en el resultado anterior, se obtiene que

$$\begin{aligned}
E(T_1^2(\mathbf{x})) &= \sum_{n=0}^{N-1} \sigma^2 \cos^2 2\pi f_0 n + \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} A^2 \cos^2 2\pi f_0 m \cos^2 2\pi f_0 n \\
&= \sigma^2 \sum_{n=0}^{N-1} \cos^2 2\pi f_0 n + \left(A \sum_{n=0}^{N-1} \cos^2 2\pi f_0 n \right)^2.
\end{aligned}$$

Este resultado y la inspección del resultado de la esperanza de $T_2(\mathbf{x})$, indica que la función

$$g_2'(T_1(\mathbf{x}), T_2(\mathbf{x})) = T_2(\mathbf{x}) - \frac{T_1^2(\mathbf{x})}{\sum_{n=0}^{N-1} \cos^2 2\pi f_0 n}$$

elimina el término deseado de la esperanza de $T_2(\mathbf{x})$. Efectivamente,

$$\begin{aligned}
E[g_2'(T_1(\mathbf{x}), T_2(\mathbf{x}))] &= E(T_2(\mathbf{x})) - \frac{E(T_1^2(\mathbf{x}))}{\sum_{n=0}^{N-1} \cos^2 2\pi f_0 n} \\
&= N\sigma^2 + A^2 \sum_{n=0}^{N-1} \cos^2 2\pi f_0 n \\
&\quad - \frac{\sigma^2 \sum_{n=0}^{N-1} \cos^2 2\pi f_0 n + \left(A \sum_{n=0}^{N-1} \cos^2 2\pi f_0 n \right)^2}{\sum_{n=0}^{N-1} \cos^2 2\pi f_0 n} \\
&= N\sigma^2 + A^2 \sum_{n=0}^{N-1} \cos^2 2\pi f_0 n - \sigma^2 - A^2 \sum_{n=0}^{N-1} \cos^2 2\pi f_0 n \\
&= (N-1)\sigma^2.
\end{aligned}$$

Este último resultado muestra además, que la función

$$g_2(T_1(\mathbf{x}), T_2(\mathbf{x})) = \frac{1}{N-1} g_2'(T_1(\mathbf{x}), T_2(\mathbf{x})),$$

hace al estadístico suficiente insesgado. Se concluye que

$$\hat{\sigma}^2 = \frac{1}{N-1} \left[\sum_{n=0}^{N-1} x^2[n] - \frac{\left(\sum_{n=0}^{N-1} x[n] \cos^2 2\pi f_0 n \right)^2}{\sum_{n=0}^{N-1} \cos^2 2\pi f_0 n} \right]$$

es el estimador MVU de σ^2 . Resumiendo, se encontró que

$$\begin{bmatrix} \hat{A} \\ \hat{\sigma}^2 \end{bmatrix} = \begin{bmatrix} \frac{T_1(\mathbf{x})}{\sum_{n=0}^{N-1} \cos^2 2\pi f_0 n} \\ \frac{1}{N-1} \left(T_2(\mathbf{x}) - \frac{T_1^2(\mathbf{x})}{\sum_{n=0}^{N-1} \cos^2 2\pi f_0 n} \right) \end{bmatrix} = \begin{bmatrix} \frac{\sum_{n=0}^{N-1} x[n] \cos 2\pi f_0 n}{\sum_{n=0}^{N-1} \cos^2 2\pi f_0 n} \\ \frac{1}{N-1} \left\{ \sum_{n=0}^{N-1} x^2[n] - \frac{\left(\sum_{n=0}^{N-1} x[n] \cos^2 2\pi f_0 n \right)^2}{\sum_{n=0}^{N-1} \cos^2 2\pi f_0 n} \right\} \end{bmatrix}.$$

Notar que el estimador de σ^2 se puede escribir como

$$\hat{\sigma}^2 = \frac{1}{N-1} \left[\sum_{n=0}^{N-1} x^2[n] - \hat{A}^2 \sum_{n=0}^{N-1} \cos^2 2\pi f_0 n \right].$$

5.4.18. Problema 18

Si se observa $x[n]$ para $n = 0, \dots, N-1$ donde las muestras son IID con distribución $\mathcal{U}[\theta_1, \theta_2]$, encontrar un estadístico suficiente para $\boldsymbol{\theta} = [\theta_1 \theta_2]$.

Solución La PDF de una muestra en términos de la función escalón es

$$p(x[n]; \boldsymbol{\theta}) = \frac{1}{\theta_2 - \theta_1} u(x[n] - \theta_1) u(\theta_2 - x[n]), \quad (5.26)$$

y por lo tanto, la PDF de los datos es

$$\begin{aligned} p(\mathbf{x}; \boldsymbol{\theta}) &= \frac{1}{(\theta_2 - \theta_1)^N} \prod_{n=0}^{N-1} u(x[n] - \theta_1) u(\theta_2 - x[n]) \\ &= \frac{1}{(\theta_2 - \theta_1)^N} u(\min x[n] - \theta_1) u(\theta_2 - \max x[n]), \end{aligned}$$

que se puede factorizar como

$$p(\mathbf{x}; \boldsymbol{\theta}) = \underbrace{\frac{1}{(\theta_2 - \theta_1)^N} u(\min x[n] - \theta_1) u(\theta_2 - \max x[n])}_{g(T(\mathbf{x}), \boldsymbol{\theta})} \times \underbrace{1}_{h(\mathbf{x})},$$

lo que indica que un estadístico suficiente es

$$\mathbf{T}(\mathbf{x}) = \begin{bmatrix} \min x[n] \\ \max x[n] \end{bmatrix}.$$

Asumiendo que el estadístico es completo, se intentará encontrar el estimador MVU de $\boldsymbol{\theta}$. Para hacerlo, se parte calculando $E(\mathbf{T}(\mathbf{x}))$. Se comenzará calculando la media de $T_1 = \min x[n]$, para lo cual hay que calcular su PDF. Del problema de la sección 5.4.13 se sabe que la PDF de T_1 es (ecuación 5.22)

$$p_{T_1}(\xi) = N(1 - \Pr\{x[n] \leq \xi\})^{N-1} p_{x[n]}(\xi),$$

donde $p_{x[n]}(\xi)$ está dada por la ecuación 5.26. La distribución de probabilidad $\Pr\{x[n] \leq \xi\}$ de $x[n]$ se obtiene integrando $p_{x[n]}(\xi)$, y para $\theta_1 < \xi < \theta_2$ es

$$\begin{aligned}\Pr\{x[n] \leq \xi\} &= \int_{-\infty}^{\xi} p_{x[n]}(u) du \\ &= \int_{\theta_1}^{\xi} \frac{1}{\theta_2 - \theta_1} du \\ &= \frac{1}{\theta_2 - \theta_1} \Big|_{\theta_1}^{\xi} \\ &= \frac{\xi - \theta_1}{\theta_2 - \theta_1},\end{aligned}$$

resultando en

$$\Pr\{x[n] \leq \xi\} = \begin{cases} 0, & \xi < \theta_1 \\ \frac{\xi - \theta_1}{\theta_2 - \theta_1}, & \theta_1 < \xi < \theta_2 \\ 1, & \xi > \theta_2. \end{cases} \quad (5.27)$$

Sustituyendo las ecuaciones 5.26 y 5.27 en la PDF de T_1 , se obtiene que para $\theta_1 < \xi < \theta_2$,

$$\begin{aligned}p_{T_1}(\xi) &= N \left(1 - \frac{\xi - \theta_1}{\theta_2 - \theta_1}\right)^{N-1} \frac{1}{\theta_2 - \theta_1} \\ &= N \frac{(\theta_2 - \xi)^{N-1}}{(\theta_2 - \theta_1)^N},\end{aligned}$$

resultando en

$$p_{T_1}(\xi) = \begin{cases} 0, & \xi < \theta_1 \\ N \frac{(\theta_2 - \xi)^{N-1}}{(\theta_2 - \theta_1)^N}, & \theta_1 < \xi < \theta_2 \\ 0, & \xi > \theta_2. \end{cases}$$

La esperanza de $T_1 = \min x[n]$ es

$$\begin{aligned}E(T_1) &= \int_{-\infty}^{\infty} \xi p_{T_1}(\xi) d\xi \\ &= \frac{N}{(\theta_2 - \theta_1)^N} \int_{\theta_1}^{\theta_2} \xi (\theta_2 - \xi)^{N-1} d\xi \\ &\stackrel{(a)}{=} \frac{1}{(\theta_2 - \theta_1)^N} \left[-\xi (\theta_2 - \xi)^N \Big|_{\theta_1}^{\theta_2} + \int_{\theta_1}^{\theta_2} (\theta_2 - \xi)^N d\xi \right] \\ &= \frac{1}{(\theta_2 - \theta_1)^N} \left[\theta_1 (\theta_2 - \theta_1)^N - \frac{(\theta_2 - \xi)^{N+1}}{N+1} \Big|_{\theta_1}^{\theta_2} \right] \\ &= \frac{1}{(\theta_2 - \theta_1)^N} \left[\theta_1 (\theta_2 - \theta_1)^N + \frac{(\theta_2 - \theta_1)^{N+1}}{N+1} \right] \\ &= \theta_1 + \frac{\theta_2 - \theta_1}{N+1}.\end{aligned}$$

donde en (a) se aplicó integración por partes $\int u dv = uv - \int v du$ con $u = \xi$ y $dv = N(\theta_2 - \xi)^{N-1}$ y por lo tanto, $du = 1$ y $v = -(\theta_2 - \xi)^N$.

Para calcular la esperanza de $T_2 = \max x[n]$ se realiza un razonamiento análogo. La PDF de T_2 se calculó en el ejemplo de la sección 5.2.3 (ecuación 5.13) y es

$$p_{T_2}(\xi) = N \Pr\{x[n] \leq \xi\}^{N-1} p_{x[n]}(\xi),$$

donde la PDF $p_{x[n]}(\xi)$ y la distribución $\Pr\{x[n] \leq \xi\}$ de $x[n]$ están dadas por las ecuaciones 5.26 y 5.27 respectivamente, resultando en

$$p_{T_2}(\xi) = \begin{cases} 0, & \xi < \theta_1 \\ N \frac{(\xi - \theta_1)^{N-1}}{(\theta_2 - \theta_1)^N}, & \theta_1 < \xi < \theta_2 \\ 0, & \xi > \theta_2 \end{cases}.$$

La esperanza de $T_2 = \max x[n]$ es

$$\begin{aligned} E(T_2) &= \int_{-\infty}^{\infty} \xi p_{T_2}(\xi) d\xi \\ &= \frac{N}{(\theta_2 - \theta_1)^N} \int_{\theta_1}^{\theta_2} \xi (\xi - \theta_1)^{N-1} d\xi \\ &\stackrel{(a)}{=} \frac{1}{(\theta_2 - \theta_1)^N} \left[\xi (\xi - \theta_1)^N \Big|_{\theta_1}^{\theta_2} - \int_{\theta_1}^{\theta_2} (\xi - \theta_1)^N d\xi \right] \\ &= \frac{1}{(\theta_2 - \theta_1)^N} \left[\theta_2 (\theta_2 - \theta_1)^N - \frac{(\xi - \theta_1)^{N+1}}{N+1} \Big|_{\theta_1}^{\theta_2} \right] \\ &= \frac{1}{(\theta_2 - \theta_1)^N} \left[\theta_2 (\theta_2 - \theta_1)^N - \frac{(\theta_2 - \theta_1)^{N+1}}{N+1} \right] \\ &= \theta_2 - \frac{\theta_2 - \theta_1}{N+1}. \end{aligned}$$

donde en (a) se aplicó integración por partes $\int u dv = uv - \int v du$ con $u = \xi$ y $dv = N(\xi - \theta_1)^{N-1}$ y por lo tanto, $du = 1$ y $v = (\xi - \theta_1)^N$. Resumiendo, se obtuvo que

$$E(\mathbf{T}(\mathbf{x})) = \begin{bmatrix} E(\min x[n]) \\ E(\max x[n]) \end{bmatrix} = \begin{bmatrix} \theta_1 + \frac{\theta_2 - \theta_1}{N+1} \\ \theta_2 - \frac{\theta_2 - \theta_1}{N+1} \end{bmatrix}.$$

Como las esperanzas de los estadísticos suficientes son lineales con los parámetros, los estimadores MVU se obtienen resolviendo el sistema de ecuaciones

$$\begin{cases} T_1 &= \hat{\theta}_1 + \frac{\hat{\theta}_2 - \hat{\theta}_1}{N+1} \\ T_2 &= \hat{\theta}_2 - \frac{\hat{\theta}_2 - \hat{\theta}_1}{N+1} \end{cases}.$$

Al hacerlo, se obtiene que

$$\hat{\boldsymbol{\theta}} = \begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix} = \begin{bmatrix} \frac{(N-2)T_1 - T_2}{N-3} \\ \frac{-T_1 + (N-2)T_2}{N-3} \end{bmatrix},$$

que es el estimador MVU.

5.4.19. Problema 19

En el modelo lineal visto en el capítulo 4, los datos se modelan como

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w},$$

donde $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$ con σ^2 desconocido. Encontrar el estimador MVU de $\boldsymbol{\theta}$ usando los teoremas de Neyman-Fisher y RLBS. Asumir que el estadístico suficiente es completo. Comparar los resultados con los obtenidos en el capítulo 4. Sugerencia: probar primero que se cumple la siguiente identidad:

$$(\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T(\mathbf{x} - \mathbf{H}\boldsymbol{\theta}) = (\mathbf{x} - \mathbf{H}\hat{\boldsymbol{\theta}})^T(\mathbf{x} - \mathbf{H}\hat{\boldsymbol{\theta}}) + (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \mathbf{H}^T \mathbf{H} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}),$$

donde

$$\hat{\boldsymbol{\theta}} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x}.$$

Solución La PDF de los datos es (ver la ecuación 4.2)

$$p(\mathbf{x}, \boldsymbol{\theta}) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left[-\frac{1}{2\sigma^2} (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T (\mathbf{x} - \mathbf{H}\boldsymbol{\theta}) \right].$$

Para factorizar la PDF como indica el teorema de Neyman-Fisher (ecuación 5.17) se demostrará que

$$(\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T(\mathbf{x} - \mathbf{H}\boldsymbol{\theta}) = (\mathbf{x} - \mathbf{H}\hat{\boldsymbol{\theta}})^T(\mathbf{x} - \mathbf{H}\hat{\boldsymbol{\theta}}) + (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \mathbf{H}^T \mathbf{H} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}),$$

donde

$$\hat{\boldsymbol{\theta}} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x}.$$

Partiendo del lado derecho de la igualdad y desarrollando, se tiene que

$$\begin{aligned} & (\mathbf{x} - \mathbf{H}\hat{\boldsymbol{\theta}})^T(\mathbf{x} - \mathbf{H}\hat{\boldsymbol{\theta}}) \\ & + (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \mathbf{H}^T \mathbf{H} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) = [\mathbf{x}\mathbf{x}^T - 2\hat{\boldsymbol{\theta}}^T \mathbf{H}^T \mathbf{x} + \hat{\boldsymbol{\theta}}^T \mathbf{H}^T \mathbf{H} \hat{\boldsymbol{\theta}}] + [\boldsymbol{\theta}^T \mathbf{H}^T \mathbf{H} \boldsymbol{\theta} - 2\boldsymbol{\theta}^T \mathbf{H}^T \mathbf{H} \hat{\boldsymbol{\theta}} + \hat{\boldsymbol{\theta}}^T \mathbf{H}^T \mathbf{H} \hat{\boldsymbol{\theta}}] \\ & = \mathbf{x}\mathbf{x}^T - 2\hat{\boldsymbol{\theta}}^T \mathbf{H}^T \mathbf{x} + 2\hat{\boldsymbol{\theta}}^T \mathbf{H}^T \mathbf{H} \hat{\boldsymbol{\theta}} + \boldsymbol{\theta}^T \mathbf{H}^T \mathbf{H} \boldsymbol{\theta} - 2\boldsymbol{\theta}^T \mathbf{H}^T \mathbf{H} \hat{\boldsymbol{\theta}} \\ & \stackrel{(a)}{=} \mathbf{x}\mathbf{x}^T - 2[\mathbf{x}^T \mathbf{H} (\mathbf{H}^T \mathbf{H})^{-1}] \mathbf{H}^T \mathbf{x} + 2[\mathbf{x}^T \mathbf{H} (\mathbf{H}^T \mathbf{H})^{-1}] [\mathbf{H}^T \mathbf{x}] \\ & \quad + \boldsymbol{\theta}^T \mathbf{H}^T \mathbf{H} \boldsymbol{\theta} - 2\boldsymbol{\theta}^T [\mathbf{H}^T \mathbf{x}] \\ & = \mathbf{x}\mathbf{x}^T + \boldsymbol{\theta}^T \mathbf{H}^T \mathbf{H} \boldsymbol{\theta} - 2\boldsymbol{\theta}^T \mathbf{H}^T \mathbf{x} \\ & = (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T (\mathbf{x} - \mathbf{H}\boldsymbol{\theta}), \end{aligned}$$

donde teniendo en cuenta que $\hat{\boldsymbol{\theta}} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x}$ y $\mathbf{H}^T \mathbf{H}$ es simétrica, en (a) se substituyó $\hat{\boldsymbol{\theta}}^T = \mathbf{x}^T \mathbf{H} (\mathbf{H}^T \mathbf{H})^{-1}$ y $\mathbf{H}^T \mathbf{H} \hat{\boldsymbol{\theta}} = \mathbf{H}^T \mathbf{x}$, resultando en la igualdad que se quería demostrar. Sustituyendo este resultado en el exponente de la PDF, se observa que la PDF se puede factorizar como

$$p(\mathbf{x}, \boldsymbol{\theta}) = \underbrace{\frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left[-\frac{1}{2\sigma^2} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \mathbf{H}^T \mathbf{H} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \right]}_{g(\mathbf{T}(\mathbf{x}), \boldsymbol{\theta})} \underbrace{\exp \left[-\frac{1}{2\sigma^2} (\mathbf{x} - \mathbf{H}\hat{\boldsymbol{\theta}})^T (\mathbf{x} - \mathbf{H}\hat{\boldsymbol{\theta}}) \right]}_{h(\mathbf{x})},$$

concluyendo que un estadístico suficiente es

$$\mathbf{T}(\mathbf{x}) = \hat{\boldsymbol{\theta}} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x}.$$

Para buscar una función que lo haga inesgado, se calcula la esperanza del estadístico suficiente,

$$E(\mathbf{T}(\mathbf{x})) = E(\hat{\boldsymbol{\theta}}) = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T E(\mathbf{x}) = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{H} \boldsymbol{\theta} = \boldsymbol{\theta}.$$

Como el estadístico suficiente ya es inesgado, $\hat{\boldsymbol{\theta}}$ es el estimador MVU. Esto es acorde a que $\hat{\boldsymbol{\theta}}$ es eficiente, resultado obtenido la sección 4.3.

Capítulo 6

Mejor estimador lineal insesgado

6.1. Introducción

En los casos en los cuales no es posible encontrar el estimador MVU, un enfoque habitual es restringir al estimador a ser lineal en los datos, y encontrar aquel estimador lineal que es insesgado y tiene menor varianza. Este es el mejor estimador lineal insesgado (*best linear unbiased estimator*, BLUE). El estimador BLUE siempre puede encontrarse, y solo se requiere el conocimiento del primer y segundo momento de los datos en lugar de la PDF.

6.2. Estimador BLUE

6.2.1. Parámetro escalar

El estimador BLUE es el estimador de la forma

$$\hat{\theta} = \sum_{n=0}^{N-1} a_n x[n],$$

donde los valores a_n son constantes a determinar imponiendo las condiciones de que el estimador sea insesgado y tenga varianza mínima. La condición de insesgado requiere que el modelo de los datos sea de la forma

$$x[n] = \theta s[n] + w[n], \quad (6.1)$$

donde $s[n]$ es una señal conocida y $\mathbf{w} = [w[0] \dots, w[N-1]]^T$ es un vector de ruido de media nula y matriz de covarianza \mathbf{C} . Por lo tanto, el estimador BLUE es aplicable en el caso de la estimación de la amplitud de una señal conocida en ruido. Como se explica en [1], el estimador BLUE es

$$\hat{\theta} = \frac{\mathbf{s}^T \mathbf{C}^{-1} \mathbf{x}}{\mathbf{s}^T \mathbf{C}^{-1} \mathbf{s}}, \quad (6.2)$$

donde $\mathbf{s} = [s[0] \dots, s[N-1]]^T$, y tiene varianza mínima entre todos los estimadores lineales de

$$\text{var}(\hat{\theta}) = \frac{1}{\mathbf{s}^T \mathbf{C}^{-1} \mathbf{s}}. \quad (6.3)$$

6.2.2. Parámetro vectorial

El estimador BLUE en el caso de un parámetro vectorial en el modelo lineal general está establecido por el *teorema de Gauss-Markov*. En este contexto, el modelo lineal general no asume ruido gaussiano.

Teorema de Gauss-Markov Si los datos son de la forma del modelo lineal general

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}, \quad (6.4)$$

donde \mathbf{H} es una matriz $N \times p$ conocida, $\boldsymbol{\theta}$ es un vector $p \times 1$ de parámetros a estimar y \mathbf{w} es un vector de ruido $N \times 1$ de media nula y matriz de covarianza \mathbf{C} con PDF arbitraria, el BLUE de $\boldsymbol{\theta}$ es

$$\hat{\boldsymbol{\theta}} = (\mathbf{H}^T \mathbf{C}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C}^{-1} \mathbf{x} \quad (6.5)$$

y la varianza mínima de $\hat{\theta}_i$ es

$$\text{var}(\hat{\boldsymbol{\theta}}) = [(\mathbf{H}^T \mathbf{C}^{-1} \mathbf{H})^{-1}]_{ii}. \quad (6.6)$$

Además, la matriz de covarianza de $\boldsymbol{\theta}$ es

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}} = (\mathbf{H}^T \mathbf{C}^{-1} \mathbf{H})^{-1}. \quad (6.7)$$

6.3. Ejemplo de procesamiento de señales

6.3.1. Ejemplo: localización de una fuente

En el problema de la determinación de la posición de una fuente, como por ejemplo una aeronave, un enfoque es emplear varias antenas para detectar una señal emitida por la fuente. La localización se realiza en base a la medida de la diferencia de tiempo de llegada (*time difference of arrival*, TDOA) de la señal emitida a las distintas antenas. Por ejemplo, si se emplean tres antenas, que es la cantidad mínima necesaria para localizar una fuente, en el caso particular en que la señal llega simultáneamente a las tres antenas, significa que la fuente está equidistante a las antenas, y su posición es la intersección de las mediatrices de los segmentos determinados por las antenas. En el caso general en donde los tiempos de llegada no son iguales, el lugar geométrico de la fuente considerando dos antenas es una hipérbola, como se muestra en la figura 6.1 y con tres antenas, la posición queda determinada por la intersección de dos hipérbolas.

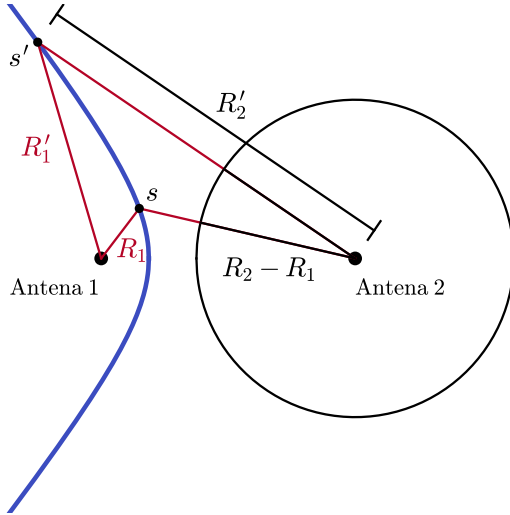


Figura 6.1: Lugar geométrico de la posición de la fuente para una diferencia de distancia fija a las antenas. Si los tiempos de llegada de la señal emitida a las antenas 1 y 2 son t_1 y t_2 respectivamente, la diferencia de distancia de la fuente a las antenas es $R_2 - R_1 = c(t_2 - t_1)$, donde c es la velocidad de propagación de la señal emitida y R_1 y R_2 es la distancia de la fuente a las antenas 1 y 2. El lugar geométrico de los puntos del plano que distan una diferencia de distancia constante a dos puntos fijos (focos) es una hipérbola. Esto implica que la fuente se encuentra en algún punto de la hipérbola.

Se examina a continuación el problema de localización de una fuente usando la teoría de la estimación. Se asume que hay N antenas en posiciones conocidas y que las medidas de los tiempos de llegada de la señal emitida a cada antena son t_i , con $i = 1, \dots, N - 1$, y el problema consiste en estimar la posición (x_s, y_s) de la fuente. Se asume que los tiempos de llegada están contaminados con ruido de media nula y matriz de covarianza conocida, pero de PDF desconocida. Para una señal emitida por la fuente en el instante $t = T_0$, las medidas se modelan como

$$t_i = T_0 + R_i/c + \epsilon_i, \quad i = 0, \dots, N - 1, \quad (6.8)$$

donde los ϵ_i son los ruidos de medición y c es la velocidad de propagación. Se asume que las medidas de ruido son de media nula y varianza σ^2 y no correlacionadas entre si. Para continuar, hay que relacionar la distancia R_i de cada antena a la fuente con la posición desconocida $\theta = [x_s \ y_s]^T$ de la fuente. Siendo (x_i, y_i) la posición conocida de la antena i -ésima, la distancia a la fuente es

$$R_i = \sqrt{(x_s - x_i)^2 + (y_s - y_i)^2}. \quad (6.9)$$

Sustituyendo la ecuación 6.9 en la ecuación 6.8 se observa que el modelo es no lineal con los parámetros desconocidos x_s y y_s . Para linealizar el modelo y poder aplicar el teorema de Markov-Gauss, se asume que se conoce la posición nominal (x_n, y_n) de la fuente. Esta posición nominal, que es cercana a la posición verdadera, se podría haber obtenido de medidas anteriores, situación que es típica en el caso del seguimiento de una fuente. Conociendo la posición nominal, se necesita estimar $\theta = [(x_s - x_n) \ (y_s -$

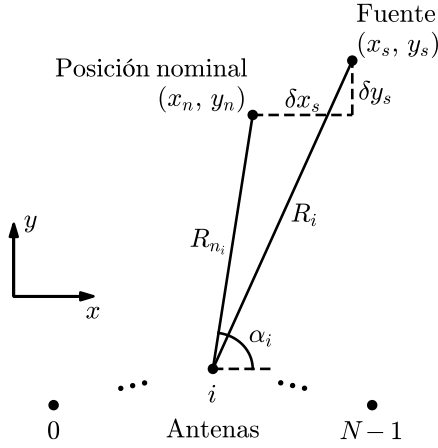


Figura 6.2: Geometría del problema de localización de fuentes.

$y_n)]^T = [\delta x_s \ \delta y_s]^T$, indicado en la figura 6.2. Considérese a R_i como una función de la posición de la fuente,

$$R_i(x, y) = \sqrt{(x - x_i)^2 + (y - y_i)^2}.$$

Realizando un desarrollo de Taylor de primer orden¹ en torno a la posición nominal, se tiene que

$$R_i(x, y) \approx R_i(x_n, y_n) + (x - x_n) \frac{\partial R_i(x_n, y_n)}{\partial x} + (y - y_n) \frac{\partial R_i(x_n, y_n)}{\partial y},$$

y como

$$\frac{\partial R_i(x, y)}{\partial x} = \frac{2(x - x_i)}{2\sqrt{(x - x_i)^2 + (y - y_i)^2}} = \frac{x - x_i}{R_i(x, y)}$$

y análogamente con la derivada respecto a y se tiene que

$$R_i(x, y) \approx R_i(x_n, y_n) + (x - x_n) \frac{x_n - x_i}{R_i(x_n, y_n)} + (y - y_n) \frac{y_n - y_i}{R_i(x_n, y_n)}.$$

Por lo tanto, evaluando en la posición de la fuente, se obtiene que

$$R_i(x_s, y_s) \approx R_i(x_n, y_n) + (x_s - x_n) \frac{x_n - x_i}{R_i(x_n, y_n)} + (y_s - y_n) \frac{y_n - y_i}{R_i(x_n, y_n)}.$$

Sean R_{n_i} las distancias de las antenas a la posición nominal (ver figura 6.2), la ecuación se puede escribir como

$$R_i \approx R_{n_i} + \frac{x_n - x_i}{R_{n_i}} \delta x_s + \frac{y_n - y_i}{R_{n_i}} \delta y_s,$$

¹El desarrollo de Taylor de primer orden de una función $f(x, y)$ de dos variables en torno a (a, b) es $f(x, y) \approx f(a, b) + (x - a)f_x(a, b) + (y - b)f_y(a, b)$.

y sustituyendo este resultado en la ecuación 6.8 se llega a que

$$t_i = T_0 + \frac{R_{n_i}}{c} + \frac{x_n - x_i}{R_{n_i}c} \delta x_s + \frac{y_n - y_i}{R_{n_i}c} \delta y_s + \epsilon_i,$$

que ahora es lineal con los parámetros desconocidos δx_s y δy_s . Teniendo en cuenta que

$$\frac{x_n - x_i}{R_{n_i}} = \cos \alpha_i \quad \frac{y_n - y_i}{R_{n_i}} = \sin \alpha_i,$$

donde α_i se definió en la figura 6.2 el modelo se simplifica como

$$t_i = T_0 + \frac{R_{n_i}}{c} + \frac{\cos \alpha_i}{c} \delta x_s + \frac{\sin \alpha_i}{c} \delta y_s + \epsilon_i.$$

El término R_{n_i}/c es conocido y puede incorporarse a la medida definiendo

$$\tau_i = t_i - \frac{R_{n_i}}{c},$$

por lo que el modelo lineal resulta en

$$\tau_i = T_0 + \frac{\cos \alpha_i}{c} \delta x_s + \frac{\sin \alpha_i}{c} \delta y_s + \epsilon_i,$$

donde los parámetros desconocidos son T_0 , δx_s y δy_s . Si bien el problema ya puede resolverse mediante el teorema de Gauss-Markov, conviene considerar la diferencia de los tiempos de llegada o medidas TDOA para eliminar T_0 ,

$$\begin{aligned} \xi_1 &= \tau_1 - \tau_0 \\ \xi_2 &= \tau_2 - \tau_1 \\ &\vdots \\ \xi_{N-1} &= \tau_{N-1} - \tau_{N-2}, \end{aligned}$$

y el modelo lineal final queda

$$\xi_i = \frac{1}{c} (\cos \alpha_i - \cos \alpha_{i-1}) \delta x_s + \frac{1}{c} (\sin \alpha_i - \sin \alpha_{i-1}) \delta y_s + \epsilon_i - \epsilon_{i-1}, \quad i = 1, \dots, N-1.$$

El modelo es de la forma del teorema de Gauss-Markov con

$$\boldsymbol{\theta} = \begin{bmatrix} \delta x_s \\ \delta y_s \end{bmatrix} \quad \mathbf{H} = \frac{1}{c} \begin{bmatrix} \cos \alpha_1 - \cos \alpha_0 & \sin \alpha_1 - \sin \alpha_0 \\ \cos \alpha_2 - \cos \alpha_1 & \sin \alpha_2 - \sin \alpha_1 \\ \vdots & \vdots \\ \cos \alpha_{N-1} - \cos \alpha_{N-2} & \sin \alpha_{N-1} - \sin \alpha_{N-2} \end{bmatrix} \quad \mathbf{w} = \begin{bmatrix} \epsilon_1 - \epsilon_0 \\ \epsilon_2 - \epsilon_1 \\ \vdots \\ \epsilon_{N-1} - \epsilon_{N-2} \end{bmatrix}.$$

El vector de ruido es de media nula, pero ahora es correlacionado. Para encontrar la matriz de covarianza, se observa que

$$\mathbf{w} = \underbrace{\begin{bmatrix} -1 & 1 & 0 & 0 & \dots & 0 \\ 0 & -1 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots & & \vdots \\ 0 & 0 & \dots & 0 & -1 & 1 \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} \epsilon_0 \\ \epsilon_1 \\ \vdots \\ \epsilon_{N-1} \end{bmatrix}}_{\boldsymbol{\epsilon}},$$

donde la matriz \mathbf{A} tiene dimensiones $(N-1) \times N$. Por lo tanto, la matriz de covarianza de \mathbf{w} es

$$\mathbf{C} = E(\mathbf{w}\mathbf{w}^T) = E(\mathbf{A}\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T\mathbf{A}^T) = \mathbf{A}E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T)\mathbf{A}^T = \mathbf{A}(\sigma^2\mathbf{I})\mathbf{A}^T = \sigma^2\mathbf{A}\mathbf{A}^T.$$

De la ecuación 6.5 se obtiene que el BLUE de la posición de la fuente es

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= (\mathbf{H}^T\mathbf{C}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}^{-1}\boldsymbol{\xi} \\ &= [\mathbf{H}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{H}]^{-1}\mathbf{H}^T(\mathbf{A}\mathbf{A}^T)^{-1}\boldsymbol{\xi} \end{aligned}$$

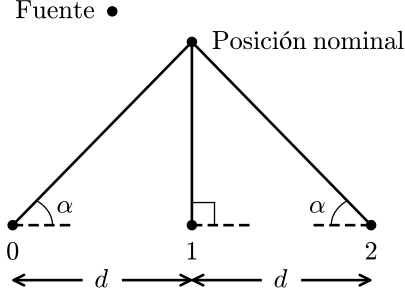


Figura 6.3: Ejemplo de geometría del problema de localización de fuentes para el conjunto mínimo de antenas.

y de la ecuación 6.7, la matriz de covarianza es

$$\mathbf{C}_{\hat{\theta}} = \sigma^2 [\mathbf{H}^T (\mathbf{A}\mathbf{A}^T)^{-1} \mathbf{H}]^{-1}. \quad (6.10)$$

Como ejemplo, considérese el conjunto de tres antenas mostrado en la figura 6.3. Para este conjunto mínimo de antenas y la posición nominal indicada, los parámetros del modelo son

$$\begin{aligned} \mathbf{H} &= \frac{1}{c} \begin{bmatrix} \cos \alpha_1 - \cos \alpha_0 & \sin \alpha_1 - \sin \alpha_0 \\ \cos \alpha_2 - \cos \alpha_1 & \sin \alpha_2 - \sin \alpha_1 \end{bmatrix} \\ &\stackrel{(a)}{=} \frac{1}{c} \begin{bmatrix} \cos \pi/2 - \cos \alpha & \sin \pi/2 - \sin \alpha \\ \cos(\pi - \alpha) - \cos \pi/2 & \sin(\pi - \alpha) - \sin \pi/2 \end{bmatrix} \\ &= \frac{1}{c} \begin{bmatrix} -\cos \alpha & 1 - \sin \alpha \\ -\cos \alpha & \sin \alpha - 1 \end{bmatrix}, \end{aligned}$$

donde en (a) se consideró que en este caso, $\alpha_0 = \alpha$, $\alpha_1 = \pi/2$ y $\alpha_2 = \pi - \alpha$, y

$$\mathbf{A} = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}.$$

Se calculará la matriz de covarianza, dada por la ecuación 6.10. Para hacerlo, se ve que

$$\mathbf{A}\mathbf{A}^T = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 1 & -1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \quad \text{y} \quad (\mathbf{A}\mathbf{A}^T)^{-1} = \frac{1}{3} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}.$$

Luego

$$\begin{aligned} (\mathbf{A}\mathbf{A}^T)^{-1} \mathbf{H} &= \frac{1}{3c} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} -\cos \alpha & 1 - \sin \alpha \\ -\cos \alpha & \sin \alpha - 1 \end{bmatrix} \\ &= \frac{1}{3c} \begin{bmatrix} -2\cos \alpha - \cos \alpha & 2(1 - \sin \alpha) - (1 - \sin \alpha) \\ \cos \alpha - 2\cos \alpha & (1 - \sin \alpha) - 2(1 - \sin \alpha) \end{bmatrix} \\ &= \frac{1}{3c} \begin{bmatrix} -3\cos \alpha & 1 - \sin \alpha \\ -3\cos \alpha & -(1 - \sin \alpha) \end{bmatrix} \end{aligned}$$

y

$$\begin{aligned} \mathbf{H}^T (\mathbf{A}\mathbf{A}^T)^{-1} \mathbf{H} &= \frac{1}{3c^2} \begin{bmatrix} -\cos \alpha & -\cos \alpha \\ 1 - \sin \alpha & -(1 - \sin \alpha) \end{bmatrix} \begin{bmatrix} -3\cos \alpha & 1 - \sin \alpha \\ -3\cos \alpha & -(1 - \sin \alpha) \end{bmatrix} \\ &= \frac{1}{3c^2} \begin{bmatrix} 3\cos^2 \alpha + 3\cos^2 \alpha & -(1 - \sin \alpha)\cos \alpha + (1 - \sin \alpha)\cos \alpha \\ -3(1 - \sin \alpha)\cos \alpha + 3(1 - \sin \alpha)\cos \alpha & (1 - \sin \alpha)^2 + (1 - \sin \alpha)^2 \end{bmatrix} \\ &= \frac{1}{3c^2} \begin{bmatrix} 6\cos^2 \alpha & 0 \\ 0 & 2(1 - \sin \alpha)^2 \end{bmatrix} \\ &= \frac{1}{c^2} \begin{bmatrix} 2\cos^2 \alpha & 0 \\ 0 & \frac{2}{3}(1 - \sin \alpha)^2 \end{bmatrix}. \end{aligned}$$

Como se trata de una matriz diagonal, la matriz inversa es diagonal y sus elementos son los inversos de los elementos de la diagonal, resultando en

$$\mathbf{C}_{\hat{\theta}} = \sigma^2 [\mathbf{H}^T (\mathbf{A}\mathbf{A}^T)^{-1} \mathbf{H}]^{-1} = \sigma^2 c^2 \begin{bmatrix} \frac{1}{2 \cos^2 \alpha} & 0 \\ 0 & \frac{3}{2(1 - \sin \alpha)^2} \end{bmatrix}.$$

El resultado muestra que cuando el ángulo α es mas pequeño se obtiene una localización mas precisa. Esto se logra haciendo que la distancia entre las antenas sea la mayor posible. Además, la localización mejora cuanto menor es la distancia de la fuente a las antenas, ya que esto también reduce el ángulo α .

6.4. Problemas

6.4.1. Problema 1

Si $x[n] = Ar^n + w[n]$ para $n = 0, \dots, N-1$, donde A es un parámetro desconocido, r es una constante conocida y $w[n]$ es ruido de media nula y varianza σ^2 , encontrar el BLUE de A y la varianza mínima. ¿La varianza mínima tiende a cero cuando $N \rightarrow \infty$?

Solución Los datos tienen la forma del modelo de la ecuación 6.1 con $s[n] = r^n$ y el estimador BLUE y la varianza mínima están dados por las ecuaciones 6.2 y 6.3 con $\mathbf{C} = \sigma^2 \mathbf{I}$. Con estos valores, el BLUE es

$$\hat{\theta} = \frac{\mathbf{s}^T \mathbf{C}^{-1} \mathbf{x}}{\mathbf{s}^T \mathbf{C}^{-1} \mathbf{s}} = \frac{\frac{1}{\sigma^2} \sum_{n=0}^{N-1} r^n x[n]}{\frac{1}{\sigma^2} \sum_{n=0}^{N-1} r^{2n}} = \frac{\sum_{n=0}^{N-1} r^n x[n]}{\sum_{n=0}^{N-1} r^{2n}}$$

y la varianza mínima es

$$\text{var}(\hat{\theta}) = \frac{1}{\mathbf{s}^T \mathbf{C}^{-1} \mathbf{s}} = \frac{\sigma^2}{\sum_{n=0}^{N-1} r^{2n}} = \begin{cases} \frac{\sigma^2}{N}, & |r| = 1 \\ \frac{1-r}{1-r^{2N}} \sigma^2, & |r| \neq 1 \end{cases}$$

Cuando $N \rightarrow \infty$, se cumple que

$$\text{var}(\hat{\theta}) \rightarrow \begin{cases} 0, & |r| \geq 1 \\ (1-r)\sigma^2, & |r| < 1 \end{cases},$$

es decir, la varianza no tiende a cero si $|r| < 1$.

6.4.2. Problema 2

Se observa

$$x[n] = A + w[n], \quad n = 0, \dots, N-1$$

donde $w[n]$ es ruido de media nula, no correlacionado y con varianza $\text{var}(w[n]) = \sigma_n^2$. En el caso en que $\sigma_n^2 = n+1$, examinar que ocurre con la varianza del BLUE cuando $N \rightarrow \infty$. Repetir para el caso en que $\sigma_n^2 = (n+1)^2$ y explicar los resultados.

Solución La varianza del BLUE está dada por la ecuación 6.3 con $\mathbf{s} = \mathbf{1}$ y matriz de covarianza

$$\mathbf{C} = \begin{bmatrix} \sigma_0^2 & 0 & \cdots & 0 \\ 0 & \sigma_1^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{N-1}^2 \end{bmatrix} \Rightarrow \mathbf{C}^{-1} = \begin{bmatrix} \frac{1}{\sigma_0^2} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma_1^2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sigma_{N-1}^2} \end{bmatrix}.$$

Por lo tanto,

$$\text{var}(\hat{A}) = \frac{1}{\sum_{n=0}^{N-1} \frac{1}{\sigma_n^2}}.$$

En el caso en que $\sigma_n^2 = n + 1$,

$$\text{var}(\hat{A}) = \frac{1}{\sum_{n=0}^{N-1} \frac{1}{n+1}} \xrightarrow{N \rightarrow \infty} 0,$$

ya que el denominador es la serie armónica, la cual diverge. En el caso en que $\sigma_n^2 = (n+1)^2$, se tiene que

$$\text{var}(\hat{A}) = \frac{1}{\sum_{n=0}^{N-1} \frac{1}{(n+1)^2}} \xrightarrow{N \rightarrow \infty} \frac{6}{\pi^2},$$

donde se usó que²

$$\sum_{n=0}^{N-1} \frac{1}{(n+1)^2} \xrightarrow{N \rightarrow \infty} \frac{\pi^2}{6}.$$

En este caso, la varianza de las muestras de ruido es tan grande que la varianza del estimador no tiende a cero.

6.4.3. Problema 3

Se considera el problema de la estimación de DC (ver el problema anterior) pero ahora las muestras de ruido están correlacionadas con matriz de covarianza

$$\mathbf{C} = \sigma^2 \begin{bmatrix} 1 & \rho & \mathbf{0} & \cdots & \mathbf{0} \\ \rho & 1 & & & \\ & \mathbf{0} & 1 & \rho & \cdots & \mathbf{0} \\ & \vdots & \vdots & \ddots & \vdots & \\ \mathbf{0} & \mathbf{0} & \cdots & 1 & \rho & \\ & & & \rho & 1 \end{bmatrix},$$

donde $|\rho| < 1$ y la dimensión N de la matriz es un número par. \mathbf{C} es una matriz diagonal a bloques y por lo tanto, fácil de invertir. Encontrar el BLUE y su varianza e interpretar los resultados.

Solución La matriz de covarianza indica que las muestras están correlacionadas de a pares, es decir, la muestra $w[n]$ está correlacionada con la muestra $w[n+1]$, con n par, pero no con el resto de las muestras. El BLUE y su varianza están dados por las ecuaciones 6.2 y 6.3 con $\mathbf{s} = \mathbf{1}$. Para encontrarlos, es necesario invertir la matriz \mathbf{C} . La matriz \mathbf{C} es diagonal a bloques, y cada bloque es

$$\mathbf{C}_{ii} = \sigma^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \Rightarrow \mathbf{C}_{ii}^{-1} = \frac{1}{\sigma^2(1-\rho^2)} \begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix}.$$

²Ver https://en.wikipedia.org/wiki/Basel_problem

Por lo tanto, la inversa de la matriz de covarianza es (ver el apéndice G)

$$\mathbf{C}^{-1} = \frac{1}{\sigma^2(1-\rho^2)} \begin{bmatrix} 1 & -\rho & \mathbf{0} & \cdots & \mathbf{0} \\ -\rho & 1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 1 & -\rho & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & 1 & -\rho \\ & & & & -\rho & 1 \end{bmatrix},$$

Luego,

$$\mathbf{C}^{-1}\mathbf{x} = \frac{1}{\sigma^2(1-\rho^2)} \begin{bmatrix} x[0] - \rho x[1] \\ -\rho x[0] + x[1] \\ x[2] - \rho x[3] \\ -\rho x[2] + x[3] \\ \vdots \\ x[N-2] - \rho x[N-1] \\ -\rho x[N-2] + x[N-1] \end{bmatrix}$$

y

$$\begin{aligned} \mathbf{1}^T \mathbf{C}^{-1} \mathbf{x} &= \frac{1}{\sigma^2(1-\rho^2)} \left(\sum_{n=0}^{N-1} x[n] - \sum_{n=0}^{N-1} \rho x[n] \right) \\ &= \frac{1}{\sigma^2(1-\rho^2)} (1-\rho) \sum_{n=0}^{N-1} x[n] \\ &\stackrel{(a)}{=} \frac{1}{\sigma^2(1+\rho)} \sum_{n=0}^{N-1} x[n], \end{aligned} \tag{6.11}$$

donde en (a) se observó que $(1-\rho^2) = (1+\rho)(1-\rho)$. Además,

$$\mathbf{C}^{-1} \mathbf{1} = \frac{1}{\sigma^2(1-\rho^2)} (1-\rho) \mathbf{1} = \frac{1}{\sigma^2(1+\rho)} \mathbf{1}$$

y

$$\mathbf{1}^T \mathbf{C}^{-1} \mathbf{1} = \frac{N}{\sigma^2(1+\rho)}. \tag{6.12}$$

De las ecuaciones 6.11 y 6.12 se obtiene que el BLUE es

$$\hat{A} = \frac{\mathbf{1}^T \mathbf{C}^{-1} \mathbf{x}}{\mathbf{1}^T \mathbf{C}^{-1} \mathbf{1}} = \frac{1}{N} \sum_{n=0}^{N-1} x[n] = \bar{x},$$

con varianza

$$\text{var}(\hat{A}) = \frac{1}{\mathbf{1}^T \mathbf{C}^{-1} \mathbf{1}} = \frac{\sigma^2(1+\rho)}{N}.$$

Se observa que en el caso en que $\rho = 0$, que implica que las muestras no están correlacionadas, $\text{var}(\hat{A}) = \sigma^2/N$, que es el caso usual. Si $\rho \rightarrow 1$, los pares de muestras están totalmente correlacionados y $\text{var}(\hat{A}) \rightarrow 2\sigma^2/N$ ya que la correlación total de a pares implica que $x[n] = x[n+1]$, lo que es equivalente a observar la mitad de las muestras. Finalmente, si $\rho \rightarrow -1$, $\text{var}(\hat{A}) \rightarrow 0$, ya que esto implica que $w[n] = -w[n+1]$ y de esta forma

$$\frac{x[n] + x[n+1]}{2} = \frac{(A + w[n]) + (A - w[n])}{2} = A,$$

es decir, el promediado de dos muestras sucesivas produce el valor exacto de A (ver el problema de la sección 3.6.9).

6.4.4. Problema 4

Se observan las muestras $\{x[0], x[1], \dots, x[N-1]\}$ IID con las siguientes PDFs:

a. Laplaciana

$$p(x[n]; \mu) = \frac{1}{2} \exp[-|x[n] - \mu|]$$

b. Gaussiana

$$p(x[n]; \mu) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(x[n] - \mu)^2\right]$$

Encontrar el BLUE de la media μ en ambos casos. ¿Qué se puede decir sobre el estimador MVU de μ ?

Solución Los datos se modelan como

$$\mathbf{x} = \mu \mathbf{1} + \mathbf{w}, \quad \text{con} \quad E(\mathbf{w}) = \mathbf{0} \quad \text{y} \quad \mathbf{C} = E(\mathbf{w}\mathbf{w}^T) = \text{var}(w[n])\mathbf{I}.$$

El BLUE está dado por la ecuación 6.2 con $\mathbf{s} = \mathbf{1}$, y es

$$\hat{\mu} = \frac{\mathbf{1}^T \mathbf{C}^{-1} \mathbf{x}}{\mathbf{1}^T \mathbf{C}^{-1} \mathbf{1}} = \frac{\mathbf{1}^T \mathbf{x}}{\mathbf{1}^T \mathbf{1}} = \frac{1}{N} \sum_{n=0}^{N-1} x[n] = \bar{x}$$

En el caso gaussiano, la media muestral es el estimador MVU, y por lo tanto, el BLUE produce el estimador MVU. Esto se debe a que el estimador MVU es lineal con los datos. En el caso de datos con distribución de Laplace, para afirmar que la media muestral es el estimador MVU, habría que calcular la CRLB y ver si la varianza de la media muestral alcanza la cota, o encontrar un estimador mediante el método de los estadísticos suficientes y ver que el estimador es la media muestral. Si se intentara alguno de estos dos métodos para encontrar el estimador MVU, se parte observando que la PDF de los datos es

$$p(\mathbf{x}; \mu) = \frac{1}{2^N} \exp\left(-\sum_{n=0}^{N-1} |x[n] - \mu|\right).$$

Para encontrar un estadístico suficiente, la PDF debe factorizarse como la ecuación 5.1 del teorema de Neyman-Fisher. Esto no parece ser posible, ya que el valor absoluto impide separar los datos del parámetro en una función $g(T(\mathbf{x}), \mu)$. Para calcular la CRLB, tomando el logaritmo, se tiene que

$$\ln p(\mathbf{x}; \mu) = -N \ln 2 - \sum_{n=0}^{N-1} |x[n] - \mu|$$

y derivando respecto al parámetro, llega a que

$$\frac{\partial \ln p(\mathbf{x}; \mu)}{\partial \mu} = -\sum_{n=0}^{N-1} \frac{\partial |x[n] - \mu|}{\partial \mu} = \sum_{n=0}^{N-1} \text{sig}(x[n] - \mu),$$

función que parece no poder ser factorizada como la ecuación 3.3. Se concluye que no es posible afirmar si la media muestral es el estimador MVU a partir de estadísticos suficientes o mediante la CRLB.

6.4.5. Problema 5

Se observan las muestras $\{x[0], \dots, x[N-1]\}$ IID con PDF log-normal

$$p(x[n], \theta) = \begin{cases} \frac{1}{\sqrt{2\pi}x[n]} \exp\left[-\frac{1}{2}(\ln x[n] - \theta)^2\right], & x[n] > 0 \\ 0, & x[n] < 0 \end{cases}$$

Probar que la media es $\exp(\theta + 1/2)$ y por lo tanto, no puede satisfacerse la restricción de insesgado del estimador lineal. Usando el enfoque de la transformación de variables aleatorias con $y[n] = \ln x[n]$, encontrar el BLUE para θ .

Solución Partiendo de la definición de la media, se ve que

$$\begin{aligned}
E(x[n]) &= \int_{-\infty}^{\infty} xp(x; \theta) dx \\
&= \int_0^{\infty} x \frac{1}{\sqrt{2\pi}x} e^{-\frac{1}{2}(\ln x - \theta)^2} dx \\
&= \int_0^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\ln x - \theta)^2} dx \\
&\stackrel{(a)}{=} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{u+\theta} e^{-\frac{1}{2}u^2} du \\
&= e^{\theta} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(u^2-2u)} du, \\
&\stackrel{(b)}{=} e^{\theta+\frac{1}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(u-1)^2} du \\
&\stackrel{(c)}{=} e^{\theta+\frac{1}{2}},
\end{aligned}$$

donde en (a) se realizó el cambio de variable $u = \ln x - \theta$, $du = dx/x$ y como $x = \exp(u + \theta)$, $dx = \exp(u + \theta)du$. Notar además que el límite inferior de integración es $u = -\infty$, ya que $u = \ln x - \theta \rightarrow -\infty$ cuando $x \rightarrow 0$. En (b) se empleó la técnica de “completar el cuadrado” observando que

$$e^{-\frac{1}{2}(u-1)^2} = e^{-\frac{1}{2}(u^2-2u-1)} = e^{-\frac{1}{2}(u^2-2u)} e^{-\frac{1}{2}}$$

y por lo tanto,

$$e^{-\frac{1}{2}(u^2-2u)} = e^{-\frac{1}{2}(u-1)^2} e^{\frac{1}{2}},$$

y en (c) que el resultado de la integral es 1 por tratarse de la integral en todo el dominio de una PDF $\mathcal{N}(1, 1)$. Se concluye que $E(x[n]) = \exp(\theta + 1/2)$, que es lo que se quería demostrar.

La restricción para que el BLUE sea insesgado es

$$E(\hat{\theta}) = \sum_{n=0}^{N-1} a_n E(x[n]) = \theta,$$

y en este caso la condición queda

$$\sum_{n=0}^{N-1} a_n e^{\theta+\frac{1}{2}} + \theta \quad \Rightarrow \quad \sum_{n=0}^{N-1} a_n = \frac{\theta}{e^{\theta+\frac{1}{2}}},$$

igualdad que no puede cumplirse para todo valor de θ .

Se considera ahora la transformación de la variable aleatoria $y = g(x) = \ln x$. La PDF de la variable transformada está dada por la ecuación 5.3 y como en este caso, a cada valor de y corresponde un único valor de x se tiene que

$$p_y(y) = \frac{p_x(x)}{|g'(x)|} = \frac{p_x(e^y)}{\frac{1}{e^y}},$$

donde en la última igualdad se consideró que $g'(x) = 1/x$ y que $x = e^y$. Por lo tanto, la PDF de $y[n] = \ln x[n]$ es

$$\begin{aligned}
p(y[n]; \theta) &= \frac{\frac{1}{\sqrt{2\pi}e^y[n]} \exp\left[-\frac{1}{2}(y[n] - \theta)^2\right]}{\frac{1}{e^{y[n]}}} \\
&= \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(y[n] - \theta)^2\right],
\end{aligned}$$

es decir, $y[n] \sim \mathcal{N}(\theta, 1)$, y por lo tanto, el BLUE para θ es la media muestral

$$\hat{\theta} = \frac{1}{N} \sum_{n=0}^{N-1} y[n] = \frac{1}{N} \sum_{n=0}^{N-1} \ln x[n].$$

6.4.6. Problema 6

En este problema se extiende el resultado del BLUE escalar. Asíumase que $E(x[n]) = \theta s[n] + \beta$, donde θ es el parámetro desconocido a estimar y β es una constante conocida. El vector \mathbf{x} de datos tiene matriz de covarianza \mathbf{C} . Se define un estimador lineal (en realidad, afín) para este problema como

$$\hat{\theta} = \sum_{n=0}^{N-1} a_n x[n] + b.$$

Probar que el BLUE está dado por

$$\hat{\theta} = \frac{\mathbf{s}^T \mathbf{C}^{-1} (\mathbf{x} - \beta \mathbf{1})}{\mathbf{s}^T \mathbf{C}^{-1} \mathbf{s}}.$$

Además, encontrar la varianza mínima.

Solución La condición de que el estimador sea insesgado es

$$E(\hat{\theta}) = \sum_{n=0}^{N-1} a_n E(x[n]) + b = \theta,$$

y sustituyendo $E(x[n])$ se obtiene que

$$\sum_{n=0}^{N-1} a_n (\theta s[n] + \beta) + b = \theta \quad \text{o} \quad \theta \sum_{n=0}^{N-1} a_n s[n] + \sum_{n=0}^{N-1} a_n \beta + b = \theta.$$

de lo que se deduce que las condiciones para que el estimador sea insesgado son

$$\sum_{n=0}^{N-1} a_n s[n] = 1 \quad \text{y} \quad \sum_{n=0}^{N-1} a_n \beta = -b.$$

Sustituyendo b en el estimador, queda

$$\hat{\theta} = \sum_{n=0}^{N-1} a_n x[n] - \sum_{n=0}^{N-1} a_n \beta = \sum_{n=0}^{N-1} a_n (x[n] - \beta)$$

y finalmente, definiendo $x'[n] = x[n] - \beta$

$$\hat{\theta} = \sum_{n=0}^{N-1} a_n x'[n]$$

que es el estimador lineal usual, pero ahora el vector de datos es $\mathbf{x}' = \mathbf{x} - \beta \mathbf{1}$, que tiene la misma matriz de covarianza que \mathbf{x} . Por lo tanto, el BLUE es

$$\hat{\theta} = \frac{\mathbf{s}^T \mathbf{C}^{-1} \mathbf{x}'}{\mathbf{s}^T \mathbf{C}^{-1} \mathbf{s}} = \frac{\mathbf{s}^T \mathbf{C}^{-1} (\mathbf{x} - \beta \mathbf{1})}{\mathbf{s}^T \mathbf{C}^{-1} \mathbf{s}}$$

y la varianza mínima es

$$\text{var}(\theta) = \frac{1}{\mathbf{s}^T \mathbf{C}^{-1} \mathbf{s}}.$$

6.4.7. Problema 7

Se asume que se observa $x[n] = As[n] + w[n]$ para $n = 0, \dots, N-1$, donde $w[n]$ es ruido de media nula con matriz de covarianza \mathbf{C} y $s[n]$ es una señal conocida. Se desea encontrar el estimador BLUE de la amplitud A . Calcular el estimador BLUE y discutir que ocurre si $\mathbf{s} = [s[0] \dots s[N-1]]^T$ es un vector propio de \mathbf{C} . Además, calcular la variannza mínima.

Solución El estimador BLUE de A es

$$\hat{A} = \frac{\mathbf{s}^T \mathbf{C}^{-1} \mathbf{x}}{\mathbf{s}^T \mathbf{C}^{-1} \mathbf{s}}.$$

Si \mathbf{s} es un vector propio de \mathbf{C} con valor propio λ , \mathbf{s} también es vector propio de \mathbf{C}^{-1} con valor propio $1/\lambda$, es decir, se cumple que

$$\mathbf{C}^{-1} \mathbf{s} = \frac{1}{\lambda} \mathbf{s}$$

y por lo tanto, también que

$$\mathbf{s}^T \mathbf{C}^{-1} = (\mathbf{C}^{-1} \mathbf{s})^T = \frac{1}{\lambda} \mathbf{s}^T,$$

donde se consideró que además \mathbf{C}^{-1} es una matriz simétrica. Sustituyendo estos resultados, el estimador queda

$$\hat{A} = \frac{\mathbf{s}^T \mathbf{x}}{\mathbf{s}^T \mathbf{s}}.$$

Notar que el resultado es el mismo que en el caso en que $\mathbf{C} = \sigma^2 \mathbf{I}$. La varianza mínima es

$$\text{var}(\hat{A}) = \frac{\lambda}{\mathbf{s}^T \mathbf{s}}.$$

6.4.8. Problema 8

Continuando con el problema anterior, siempre es posible representar al vector señal \mathbf{s} como una combinación lineal de vectores propios de \mathbf{C} . Esto se debe a que \mathbf{C} tiene N vectores propios linealmente independientes. Además, puede mostrarse que debido a la naturaleza simétrica de \mathbf{C} , esos vectores propios son ortonormales. Por lo tanto, una representación ortogonal de la señal es

$$\mathbf{s} = \sum_{i=0}^{N-1} \alpha_i \mathbf{v}_i,$$

donde $\{\mathbf{v}_0, \dots, \mathbf{v}_{N-1}\}$ son los vectores propios ortonormales de \mathbf{C} . Probar que la varianza mínima del BLUE es

$$\text{var}(\hat{A}) = \frac{1}{\sum_{i=0}^{N-1} \frac{\alpha_i^2}{\lambda_i}},$$

donde λ_i es el valor propio de \mathbf{C} asociado al vector propio \mathbf{v}_i . Luego, mostrar que la energía de la señal $\mathcal{E} = \mathbf{s}^T \mathbf{s}$ está dada por $\sum_{i=0}^{N-1} \alpha_i^2$. Finalmente, probar que si la energía se restringe a ser cierto valor \mathcal{E}_0 , la menor varianza posible del BLUE se obtiene eligiendo la señal

$$\mathbf{s} = c \mathbf{v}_{\min},$$

donde \mathbf{v}_{\min} es el vector propio de \mathbf{C} asociado al menor valor propio y c se elige de forma de satisfacer la restricción de la energía $\mathcal{E}_0 = \mathbf{s}^T \mathbf{s} = c^2$. Asumir que los valores propios de \mathbf{C} son todos distintos. Explicar porque este resultado tiene sentido.

Solución Se mostrará primero informalmente, que los vectores propios de una matriz simétrica son ortogonales. Este resultado está dado por el teorema espectral. Supóngase que \mathbf{x} y \mathbf{y} son vectores propios de la matriz simétrica \mathbf{A} con valores propios asociados λ y μ distintos. En este caso, se cumple que

$$\lambda \mathbf{x}^T \mathbf{y} = (\lambda \mathbf{x})^T \mathbf{y} = (\mathbf{A} \mathbf{x})^T \mathbf{y} = \mathbf{x}^T \mathbf{A}^T \mathbf{y} = \mathbf{x}^T \mathbf{A} \mathbf{y} = \mathbf{x}^T \mu \mathbf{y}$$

y por lo tanto,

$$(\lambda - \mu) \mathbf{x}^T \mathbf{y} = 0 \quad \Rightarrow \quad \mathbf{x}^T \mathbf{y} = 0,$$

es decir, \mathbf{x} y \mathbf{y} son ortonormales. Faltaría considerar el caso en que los valores propios son iguales, pero no se hará aquí.

La varianza mínima del estimador BLUE es

$$\text{var}(\hat{A}) = \frac{1}{\mathbf{s}^T \mathbf{C}^{-1} \mathbf{s}}.$$

Por lo tanto, considerando la representación de \mathbf{s} como combinación lineal de los vectores propios, se tiene que

$$\begin{aligned} \mathbf{C}^{-1} \mathbf{s} &= \mathbf{C}^{-1} \sum_{j=0}^{N-1} \alpha_j \mathbf{v}_j \\ &= \sum_{j=0}^{N-1} \alpha_j \mathbf{C}^{-1} \mathbf{v}_j \\ &= \sum_{j=0}^{N-1} \frac{\alpha_j}{\lambda_j} \mathbf{v}_j, \end{aligned}$$

donde en la última igualdad se consideró que si λ_j es el valor propio de \mathbf{C} asociado al vector propio \mathbf{v}_j , $1/\lambda_j$ es el valor propio de \mathbf{C}^{-1} asociado al vector propio \mathbf{v}_j . Luego,

$$\begin{aligned} \mathbf{s}^T \mathbf{C}^{-1} \mathbf{s} &= \left(\sum_{i=0}^{N-1} \alpha_i \mathbf{v}_i \right)^T \sum_{j=0}^{N-1} \frac{\alpha_j}{\lambda_j} \mathbf{v}_j \\ &= \sum_{i=0}^{N-1} \alpha_i \mathbf{v}_i^T \sum_{j=0}^{N-1} \frac{\alpha_j}{\lambda_j} \mathbf{v}_j \\ &= \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \frac{\alpha_i \alpha_j}{\lambda_j} \mathbf{v}_i^T \mathbf{v}_j \\ &= \sum_{i=0}^{N-1} \frac{\alpha_i^2}{\lambda_i}, \end{aligned}$$

donde en la última igualdad se consideró que los vectores propios son ortogonales, es decir, cumplen que

$$\mathbf{v}_i^T \mathbf{v}_j = \begin{cases} 1, & \text{si } i = j \\ 0, & \text{si } i \neq j. \end{cases}$$

Se concluye que

$$\text{var}(\hat{A}) = \frac{1}{\sum_{i=0}^{N-1} \frac{\alpha_i^2}{\lambda_i}},$$

que es lo que se quería demostrar.

La energía de la señal es

$$\begin{aligned} \mathcal{E} &= \mathbf{s}^T \mathbf{s} \\ &= \sum_{i=0}^{N-1} \alpha_i \mathbf{v}_i^T \sum_{j=0}^{N-1} \alpha_j \mathbf{v}_j \\ &= \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \alpha_i \alpha_j \mathbf{v}_i^T \mathbf{v}_j \\ &= \sum_{i=0}^{N-1} \alpha_i^2, \end{aligned}$$

donde nuevamente se consideró que los vectores propios son ortonormales.

Por último, se quiere encontrar la señal de energía \mathcal{E}_0 que minimiza la varianza. El problema consiste en buscar los valores de α_i que maximicen el denominador $\sum_{i=0}^{N-1} \alpha_i^2 / \lambda_i$ de la varianza sujeto a la restricción $\sum_{i=0}^{N-1} \alpha_i^2 = \mathcal{E}_0$ de la energía de la señal. Considerando multiplicadores de Lagrange, esto implica encontrar los puntos estacionarios de la función

$$\mathcal{L}(\boldsymbol{\alpha}, \lambda) = \sum_{i=0}^{N-1} \frac{\alpha_i^2}{\lambda_i} + \lambda \left(\sum_{i=0}^{N-1} \alpha_i^2 - \mathcal{E}_0 \right).$$

Derivando respecto a α_k , se tiene que

$$\frac{\partial \mathcal{L}(\boldsymbol{\alpha}, \lambda)}{\partial \alpha_k} = \frac{2\alpha_k}{\lambda_k} + 2\lambda\alpha_k$$

e igualando a cero, se obtienen las condiciones

$$\alpha_k \left(\frac{1}{\lambda_k} + \lambda \right) = 0, \quad k = 0, \dots, N-1.$$

Esto significa que $\alpha_k = 0$ o $\lambda = -1/\lambda_k$ para todo k . Pero por la hipótesis de que los valores propios de \mathbf{C} son todos distintos, $\lambda = -1/\lambda_k$ a lo sumo solo para un valor de k . Adicionalmente, no puede ocurrir que $\alpha_k = 0$ para todo k ya que no se cumpliría la restricción de la energía de la señal. Por lo tanto, la única posibilidad es que $\alpha_k = 0$ para todo k excepto para uno solo. Sea este $k = j$, donde j es el índice asociado al menor valor propio. De esta forma, $\alpha_k = 0$ para todo k excepto para $k = j$ y $\lambda = 1/\lambda_j$ por lo que

$$\mathbf{s} = \alpha_j \mathbf{v}_j \quad \text{y} \quad \mathcal{E}_0 = \alpha_j^2$$

y la varianza del estimador es

$$\text{var}(\hat{A}) = \frac{1}{\frac{\alpha_j^2}{\lambda_j}} = \frac{\lambda_j}{\alpha_j^2} = \frac{\lambda_j}{\mathcal{E}_0},$$

es decir

$$\text{var}(\hat{A}) = \frac{\lambda_{\min}}{\mathcal{E}_0} \quad \text{con} \quad \mathbf{s} = c\mathbf{v}_{\min} = \sqrt{\mathcal{E}_0} \mathbf{v}_{\min}.$$

6.4.9. Problema 9

En un sistema de comunicación de encendido y apagado (*on-off keyed*, OOK) se transmite una de dos señales

$$s_0(t) = 0, \quad 0 \leq t \leq T$$

para representar un 0 binario o

$$s_1(t) = A \cos 2\pi f_1 t, \quad 0 \leq t \leq T$$

para representar un 1 binario. Se asume que la amplitud A es positiva. Para determinar que bit fue transmitido, se muestrea la señal recibida obteniendo

$$x[n] = s_i[n] + w[n], \quad n = 0, \dots, N-1.$$

Si se transmitió un 0, $s_i[n] = 0$, y si se transmitió un 1, $s_i[n] = A \cos 2\pi f_1 n$ (asumiendo una tasa de muestreo de 1 muestra/s). Las muestras de ruido $w[n]$ modelan el ruido del canal. Un receptor sencillo podría estimar la amplitud A de la senoide y decidir que se envió un 1 si $\hat{A} > \gamma$ y un 0 si $\hat{A} < \gamma$. De esta forma, un problema equivalente es estimar la amplitud A de la señal recibida

$$x[n] = A \cos 2\pi f_1 n + w[n], \quad n = 0, \dots, N-1,$$

donde $w[n]$ es ruido de media nula con matriz de covarianza $\mathbf{C} = \sigma^2 \mathbf{I}$. Encontrar el BLUE para este problema e interpretar el detector resultante. Encontrar la mejor frecuencia en el rango $0 \leq f_1 < 1/2$ para usar en el transmisor.

Solución El modelo de los datos es en este caso

$$x[n] = As[n] + w[n], \quad \text{con} \quad s[n] = \cos 2\pi f_1 n.$$

Con $\mathbf{s} = [1 \cos 2\pi f_1, \dots, \cos 2\pi f_1(N-1)]^T$, el estimador BLUE es

$$\hat{A} = \frac{\mathbf{s}^T \mathbf{C}^{-1} \mathbf{x}}{\mathbf{s}^T \mathbf{C}^{-1} \mathbf{s}} = \frac{\mathbf{s}^T \mathbf{x}}{\mathbf{s}^T \mathbf{s}} = \frac{\sum_{n=0}^{N-1} s[n]x[n]}{\sum_{n=0}^{N-1} s^2[n]} = \frac{\sum_{n=0}^{N-1} x[n] \cos 2\pi f_1 n}{\sum_{n=0}^{N-1} \cos^2 2\pi f_1 n}.$$

Observando que el estimador puede escribirse como

$$\hat{A} = \frac{\frac{N}{2}}{\sum_{n=0}^{N-1} \cos^2 2\pi f_1 n} \frac{2}{N} \sum_{n=0}^{N-1} x[n] \cos 2\pi f_1 n$$

y que

$$\frac{\frac{N}{2}}{\sum_{n=0}^{N-1} \cos^2 2\pi f_1 n} \stackrel{(a)}{=} \frac{\frac{N}{2}}{\sum_{n=0}^{N-1} \frac{1 + \cos 4\pi f_1 n}{2}} = \frac{N}{N + \sum_{n=0}^{N-1} \cos 4\pi f_1 n} = \frac{1}{1 + \frac{1}{N} \sum_{n=0}^{N-1} \cos 4\pi f_1 n} \stackrel{(b)}{\approx} 1,$$

donde en (a) se empleó la identidad trigonométrica $\cos(2\theta) = 2\cos^2 \theta - 1$ y en (b) se tuvo en cuenta que $\frac{1}{N} \sum_{n=0}^{N-1} \cos 4\pi f_1 n \approx 0$ si f_1 no es cercano a 0 ni a $1/2$, como se mostró en el problema de la sección 3.6.7, se obtiene que

$$\hat{A} \approx \frac{2}{N} \sum_{n=0}^{N-1} x[n] \cos 2\pi f_1 n$$

si f_1 no es cercano a 0 ni a $1/2$. Se concluye que el estimador de la amplitud es el coeficiente de Fourier correspondiente al coseno de frecuencia f_1 (ver la ecuación E.2 del apéndice E). Además, la varianza es

$$\text{var}(\hat{A}) = \frac{1}{\mathbf{s}^T \mathbf{C}^{-1} \mathbf{s}} = \frac{\sigma^2}{\mathbf{s}^T \mathbf{s}} = \frac{\sigma^2}{\sum_{n=0}^{N-1} \cos^2 2\pi f_1 n} \geq \frac{\sigma^2}{N},$$

ya que $\sum_{n=0}^{N-1} \cos^2 2\pi f_1 n \leq N$ y la igualdad se da si $f_1 = 0$. Por lo tanto, la varianza se minimiza si se emplea una señal constante. Con esta elección, se maximiza la energía de la señal. En la figura 6.4 se muestra la varianza del estimador en función de la frecuencia de la sinusoide enviada.

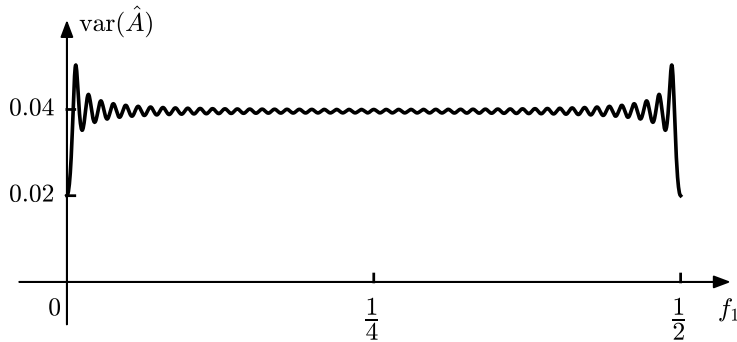


Figura 6.4: Varianza del estimador \hat{A} en función de la frecuencia f_1 de la sinusoide enviada para $N = 50$ y $\sigma^2 = 1$.

6.4.10. Problema 10

En este problema se continúa el problema anterior examinando la selección de la señal para un sistema OOK en la presencia de ruido coloreado. Sea el ruido $w[n]$ un proceso aleatorio WSS de media nula con ACF

$$r_{ww}[k] = \begin{cases} 1.81 & k = 0 \\ 0 & k = 1 \\ 0.9 & k = 2 \\ 0 & k \geq 3 \end{cases}$$

Encontrar y graficar la PSD para las frecuencias $0 \leq f \leq 1/2$. Como en el problema anterior, encontrar la frecuencia que produce la menor varianza del BLUE para $N = 50$. Explicar los resultados. Sugerencia: se necesita una computadora para hacerlo.

Solución La PSD $P_{ww}(f)$ del ruido es la transformada de Fourier de la autocorrelación,

$$\begin{aligned} P_{ww}(f) &= \mathcal{F}\{r_{ww}[k]\} \\ &= \sum_{k=-\infty}^{\infty} r_{ww}[k] e^{-j2\pi f k} \\ &= r_{ww}[-2] e^{j4\pi f} + r_{ww}[0] + r_{ww}[2] e^{-j4\pi f} \\ &= r_{ww}[0] + r_{ww}[2] (e^{j4\pi f} + e^{-j4\pi f}) \\ &= r_{ww}[0] + 2r_{ww}[2] \cos 4\pi f, \end{aligned}$$

resultando en

$$P_{ww}(f) = 1.81 + 1.8 \cos 4\pi f,$$

y se muestra en la figura 6.5. Para ver en que frecuencias el estimador tiene menor varianza, hay que

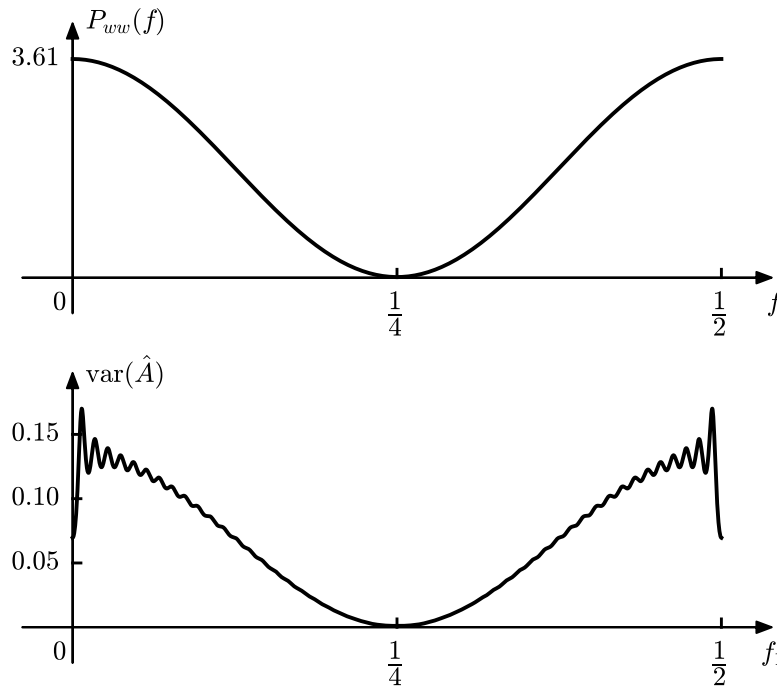


Figura 6.5: PSD del ruido coloreado y varianza del estimador BLUE. La varianza del estimador es menor en las frecuencias donde el ruido tiene menor potencia.

calcular la varianza en función de la frecuencia f_1 de la señal transmitida. La varianza del estimador es

$$\text{var}(\hat{A}) = \frac{1}{\mathbf{s}^T \mathbf{C}^{-1} \mathbf{s}},$$

con $\mathbf{s} = [1 \cos 2\pi f_1, \dots, \cos 2\pi f_1(N-1)]^T$, y la matriz de covarianza de los datos es

$$\mathbf{C} = \begin{bmatrix} r_{ww}[0] & r_{ww}[1] & \dots & r_{ww}[N-1] \\ r_{ww}[1] & r_{ww}[0] & \dots & r_{ww}[N-2] \\ \vdots & \vdots & \ddots & \vdots \\ r_{ww}[N-1] & r_{ww}[N-2] & \dots & r_{ww}[0] \end{bmatrix} = \begin{bmatrix} 1.81 & 0 & 0.9 & 0 & \dots & 0 \\ 0 & 1.81 & 0 & 0.9 & \dots & 0 \\ 0.9 & 0 & 1.81 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0.9 & 0 & 1.81 & 0 \\ 0 & \dots & 0 & 0.9 & 0 & 1.81 \end{bmatrix},$$

de dimensiones $N \times N$ con $N = 50$, toeplitz y simétrica. Se empleó una computadora para invertir \mathbf{C} y calcular $\text{var}(\hat{A})$ para distintos valores de f_1 . Como se observa en la figura 6.5, el estimador tiene menor varianza en $f_1 = 1/4$, coincidiendo con la frecuencia donde el ruido tiene menor PSD.

6.4.11. Problema 11

Se considera el problema de la estimación del nivel de DC en ruido WSS. Dado

$$x[n] = A + w[n], \quad n = 0, \dots, N-1,$$

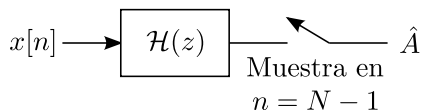
donde $w[n]$ es ruido WSS de media nula con ACF $r_{ww}[k]$, se quiere estimar A . Para estimar A se propone usar la muestra $n = N-1$ de la salida del filtro FIR mostrado en la figura 6.6. Notar que el estimador está dado por

$$\hat{A} = \sum_{k=0}^{N-1} h[k]x[N-1-k].$$

Se asume que la entrada $x[n]$ es nula en $n < 0$. Para obtener un buen estimador se necesita que el filtro permita pasar la continua y bloquee el ruido $w[n]$. Por lo tanto, se eligen los coeficientes $h[k]$ del filtro tal que $H(e^{j0}) = 1$ como restricción y minimicen la potencia del ruido en la muestra $n = N-1$. Encontrar los coeficientes del filtro óptimo y la potencia del ruido mínima a la salida del filtro óptimo. Explicar los resultados.

$$\mathcal{H}(z) = \sum_{k=0}^{N-1} h[k]z^{-k}$$

Figura 6.6: Filtro FIR para estimar A .



Solución La restricción para que el filtro permita pasar el componente de continua incambiado es $H(e^{j0}) = 1$, donde

$$H(e^{j2\pi f}) = \sum_{k=-\infty}^{\infty} h[k]e^{-j2\pi f k} = \sum_{k=0}^{N-1} h[k]e^{-j2\pi f k}.$$

Por lo tanto, evaluando en $f = 0$, la restricción es

$$H(e^{j0}) = \sum_{k=0}^{N-1} h[k] = 1,$$

que en notación matricial puede escribirse como

$$\mathbf{h}^T \mathbf{1} = 1, \quad \text{con} \quad \mathbf{h} = [h[0] \ h[1] \ \dots \ h[N-1]]^T \quad \text{y} \quad \mathbf{1} = [1 \ 1 \ \dots \ 1]^T.$$

Por otro lado, sea $y[n]$ la salida del filtro. De esta forma,

$$y[n] = \sum_{k=0}^{N-1} h[k]x[n-k],$$

y como el estimador de A es la muestra $n = N - 1$ de la salida del filtro,

$$\hat{A} = y[N - 1] = \sum_{k=0}^{N-1} h[k]x[N - 1 - k].$$

Como el filtro es lineal, la componente de la salida correspondiente al ruido en el instante $n = N - 1$ es

$$y_w[N - 1] = \sum_{k=0}^{N-1} h[k]w[N - 1 - k],$$

por lo que la potencia del ruido en ese instante es

$$\begin{aligned} E(y_w^2[N - 1]) &= E \left[\left(\sum_{k=0}^{N-1} h[k]w[N - 1 - k] \right)^2 \right] \\ &= E \left(\sum_{k=0}^{N-1} h[k]w[N - 1 - k] \sum_{l=0}^{N-1} h[l]w[N - 1 - l] \right) \\ &= \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} h[k]h[l]E(w[N - 1 - k]w[N - 1 - l]) \\ &\stackrel{(a)}{=} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} h[k]h[l]r_{ww}[k - l] \end{aligned}$$

donde en (a) se tuvo en cuenta que como el ruido es WSS, $E(w[N - 1 - k]w[N - 1 - l]) = E(w[k]w[l]) = r_{ww}[k - l]$. La potencia del ruido puede por lo tanto expresarse en forma matricial como

$$E(y_w^2[N - 1]) = \mathbf{h}^T \mathbf{C} \mathbf{h},$$

donde \mathbf{C} es la matriz de covarianza de tamaño $N \times N$ del ruido, con $\mathbf{C}_{k,l} = r_{ww}[k - l]$, teniendo en cuenta que el ruido es de media nula. Finalmente, el problema de encontrar los coeficientes óptimos del filtro se puede formular como

$$\text{minimizar } \mathbf{h}^T \mathbf{C} \mathbf{h} \quad \text{sujeto a } \mathbf{h}^T \mathbf{1} = 1.$$

Se concluye que el problema es equivalente a encontrar el BLUE con $\mathbf{s} = \mathbf{1}$ (ver el apéndice 6A de [1]). Por lo tanto, el filtro óptimo es

$$\mathbf{h}_{\text{opt}} = \frac{\mathbf{C}^{-1} \mathbf{1}}{\mathbf{1}^T \mathbf{C}^{-1} \mathbf{1}}.$$

Además, la potencia del ruido a la salida del filtro es

$$E(y_w^2[N - 1]) = \mathbf{h}_{\text{opt}}^T \mathbf{C} \mathbf{h}_{\text{opt}} = \frac{\mathbf{1}^T \mathbf{C}^{-1} \mathbf{C} \mathbf{C}^{-1} \mathbf{1}}{(\mathbf{1}^T \mathbf{C}^{-1} \mathbf{1})^2} = \frac{\mathbf{1}^T \mathbf{C}^{-1} \mathbf{1}}{(\mathbf{1}^T \mathbf{C}^{-1} \mathbf{1})^2} = \frac{1}{\mathbf{1}^T \mathbf{C}^{-1} \mathbf{1}}.$$

En la estimación del nivel de DC en ruido WSS, el estimador BLUE puede verse como la salida de un filtro lineal que permite pasar incambiado el nivel de continua y cuyos coeficientes se eligen de forma de minimizar la potencia del ruido a la salida del filtro.

Como ejemplo, se considera el caso en que el ruido es un proceso WSS de media nula con ACF

$$r_{ww}[k] = \begin{cases} 1.81 & k = 0 \\ 0 & k = 1 \\ r_{ww}[2] & k = 2 \\ 0 & k \geq 3 \end{cases}, \quad (6.13)$$

y el ruido contaminante es $w_1[n]$ cuando $r_{ww}[2] = 0.9$ y $w_2[n]$ cuando $r_{ww}[2] = -0.9$. En ambos casos, la potencia del ruido es $r_{ww}[0] = 1.81$. La PSD se calculó en el problema de la sección 6.4.10 y es

$$P_{ww}(f) = 1.81 + 2r_{ww}[2] \cos 4\pi f.$$

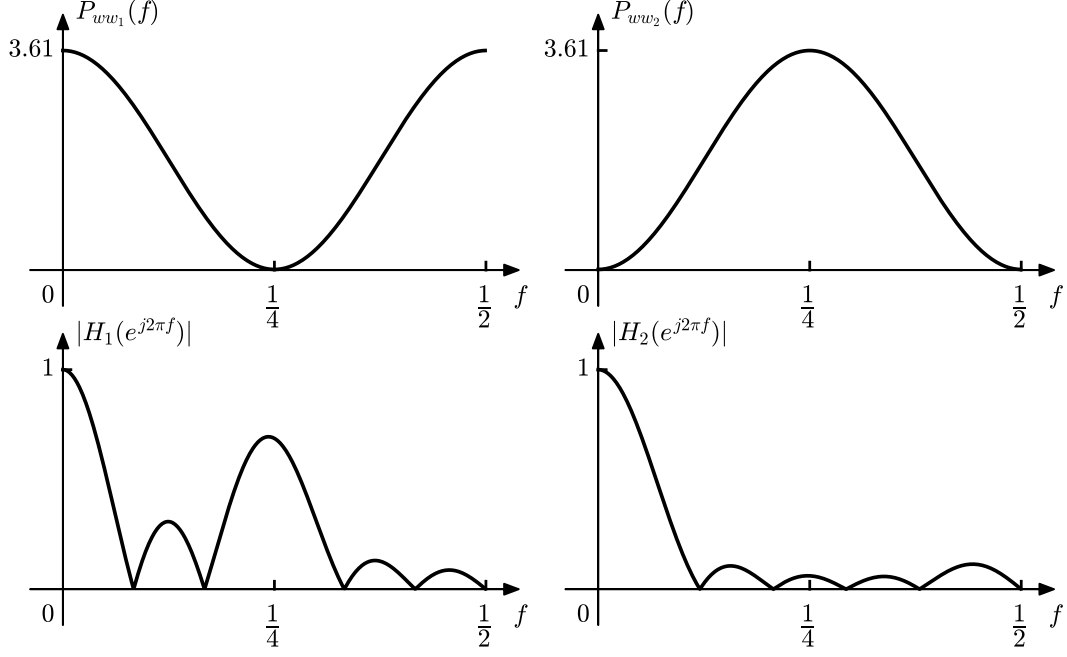


Figura 6.7: PSD del ruido y función de transferencia del filtro óptimo. $P_{ww_1}(f)$ es la PSD del ruido cuando $r_{ww}[2] = 0.9$ y $P_{ww_2}(f)$ es la PSD del ruido cuando $r_{ww}[2] = -0.9$. $H_1(e^{j2\pi f})$ y $H_2(e^{j2\pi f})$ son las funciones de transferencia de los filtros óptimos de orden $N - 1 = 9$ en cada caso.

Usando una computadora, se calcularon los coeficientes de filtro óptimo para los dos tipos de ruido usando $N = 10$ muestras. La PSD y la función de transferencia de los filtros óptimos, que son filtros FIR de orden $N - 1 = 9$, se muestran en la figura 6.7. Como se observa en la figura, la potencia del ruido $w_1[n]$ es máxima en torno a la continua y a la frecuencia máxima, mientras que la potencia del ruido $w_2[n]$ es máxima en torno a la frecuencia $f = 1/2$ y nula en continua y en la frecuencia máxima. También se observa que el filtro óptimo, que debe tener ganancia unidad en continua, tiene menor ganancia en las frecuencias donde el ruido tiene mayor potencia. La potencia del ruido a la salida del filtro es en cada caso

$$E(y_{w_1}^2[N - 1]) \approx 0.302 \quad \text{y} \quad E(y_{w_2}^2[N - 1]) \approx 0.028.$$

Esto indica que el ruido $w_2[n]$ es eliminado mas eficazmente, y se debe a que dicho ruido no tiene componentes espectrales en continua. El componente de continua no pueden ser eliminado debido a que es justamente es lo que se desea estimar.

Para observar el desempeño del filtro óptimo en el dominio del tiempo, se generaron señales $x[n] = A + w_{1,2}[n]$ con $A = 4$ y se filtraron con el filtro óptimo con $N = 50$. Los resultados se muestran en la figura 6.8. Se observa como el ruido $w_2[n]$ es filtrado de forma mas eficaz que el ruido $w_1[n]$. Se realiza a continuación algunas observaciones sobre la generación de los ruidos $w_{1,2}[n]$, que son procesos WSS dados por la ACF de la ecuación 6.13. Por tener autocorrelación de largo finito, se trata de procesos de media móvil (*moving average*, MA), y pueden obtenerse mediante el filtrado de ruido blanco con un filtro FIR (ver el capítulo de [5]), denominado filtro generador. Para encontrar los coeficientes de este filtro, se parte considerando que si se filtra un proceso $v[n]$ con PSD $P_{vv}(z)$ mediante un filtro con función de transferencia $H_G(z)$, la PSD del proceso $w[n]$ de salida es

$$P_{ww}(z) = P_{vv}(z)H_G(z)H_G(1/z) \quad (6.14)$$

si la respuesta al impulso $h_G[n]$ del filtro es real. Si el proceso de entrada al filtro es ruido blanco de potencia σ_v^2 , su ACF es $r_{vv}[k] = \sigma_v^2\delta[k]$ y su PSD es $P_{vv}(z) = \sigma_v^2$. Además, el proceso $w[n]$ que se quiere generar tiene en este caso ACF

$$r_{ww}[k] = r_{ww}[0]\delta[k] + r_{ww}[-2]\delta[k + 2] + r_{ww}[2]\delta[k - 2]$$

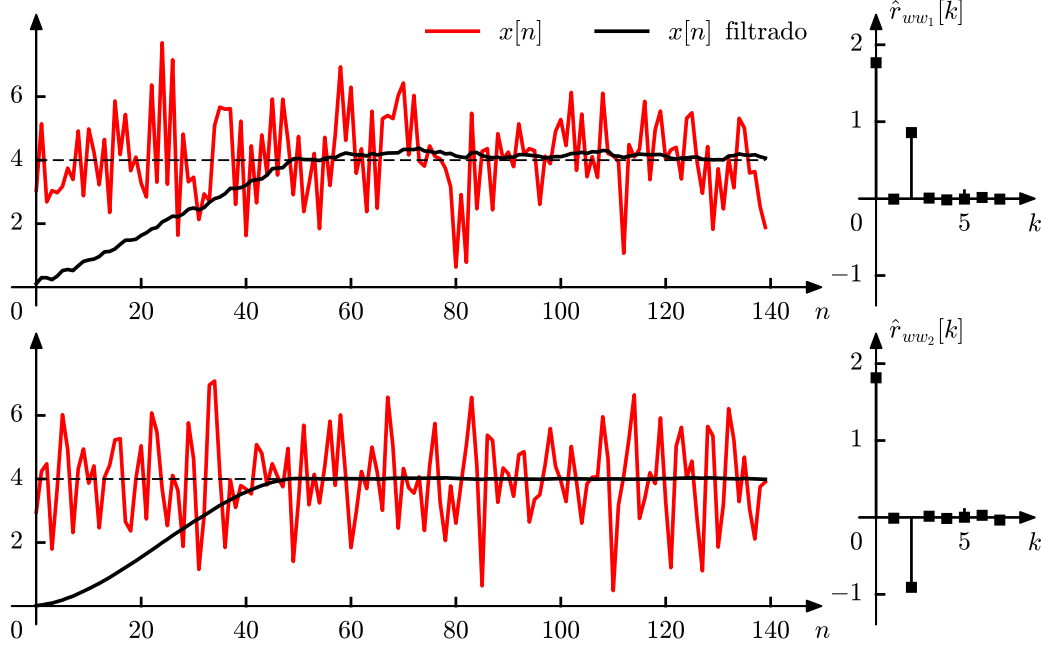


Figura 6.8: Nivel de DC $x[n] = A + w_{1,2}[n]$ contaminado con ruido WSS y salida del filtro óptimo calculado con $N = 50$ muestras. El nivel de DC es $A = 4$. También se grafican las primeras muestras de la autocorrelación muestral $\hat{r}_{ww_1}[k]$ y $\hat{r}_{ww_2}[k]$ calculadas usando 10000 muestras del ruido $w_1[n]$ y $w_2[n]$ respectivamente.

y como $r_{ww}[-2] = r_{ww}[2]$, su PSD es

$$P_{ww}(z) = r_{ww}[0] + r_{ww}[2]z^{-2} + r_{ww}[2]z^2.$$

Para calcular $H_G(z)$, se sustituye las PSD de $v[n]$ y $w[n]$ en la ecuación 6.14, obteniendo que

$$r_{ww}[0] + r_{ww}[2]z^{-2} + r_{ww}[2]z^2 = \sigma_v^2 H(z)H(1/z),$$

y por lo tanto,

$$H_G(z)H_G(1/z) = \frac{r_{ww}[0] + r_{ww}[2]z^{-2} + r_{ww}[2]z^2}{\sigma_v^2}.$$

Se observa que esta ecuación se cumple con un filtro con transferencia de la forma $H_G(z) = a + bz^{-2}$, ya que así

$$H_G(z)H_G(1/z) = (a + bz^{-2})(a + bz^2) = a^2 + abz^2 + abz^{-2} + b^2$$

e igualando con el lado derecho de la igualdad anterior, se tiene que

$$\begin{cases} a^2 + b^2 &= \frac{r_{ww}[0]}{\sigma_v^2} \\ ab &= \frac{r_{ww}[2]}{\sigma_v^2} \end{cases}.$$

Despejando b en la primera ecuación y sustituyendo en la primera, se obtiene que el coeficiente a cumple que

$$a^4 \sigma_v^4 - a^2 \sigma_v^2 r_{ww}[0] + r_{ww}^2[2] = 0 \quad \text{o} \quad x^2 - x r_{ww}[0] + r_{ww}^2[2] = 0, \quad \text{con } a = \pm \frac{\sqrt{x}}{\sigma_v}.$$

Por ejemplo, con los valores $r_{ww}[0] = 1.81$ y $r_{ww}[2] = 0.9$, los posibles valores a y b de los coeficientes del filtro generador son

a	1	-1	0.9	-0.9
b	0.9	-0.9	1	-1

Teniendo en cuenta que el filtro generador es de la forma $H_G(z) = a + bz^{-2}$, los ruidos $w_{1,2}[n]$ se generan con la ecuación en diferencias $w[n] = av[n] + bv[n-2]$, donde $v[n]$ es ruido blanco de potencia σ_v^2 . Para confirmar que los procesos generados efectivamente tienen la autocorrelación deseada, puede calcularse la autocorrelación muestral,

$$\hat{r}_{ww}[k] = \frac{1}{N-k} \sum_{n=0}^{N-1-k} x[n]x[n+k].$$

Esto se hizo usando $N = 10000$ muestras de los ruidos generados y el resultado se muestra en la figura 6.8.

6.4.12. Problema 12

Probar que el BLUE conmuta sobre transformaciones lineales (en realidad, afines) de $\boldsymbol{\theta}$. Es decir, si se quiere estimar

$$\boldsymbol{\alpha} = \mathbf{B}\boldsymbol{\theta} + \mathbf{b}$$

donde \mathbf{B} es una matriz conocida invertible $p \times p$ y \mathbf{b} es un vector conocido $p \times 1$, probar que el BLUE está dado por

$$\hat{\boldsymbol{\alpha}} = \mathbf{B}\hat{\boldsymbol{\theta}} + \mathbf{b}.$$

donde $\hat{\boldsymbol{\theta}}$ es el BLUE para $\boldsymbol{\theta}$. Asumir que $\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$, donde $E(\mathbf{w}) = \mathbf{0}$ y $E(\mathbf{w}\mathbf{w}^T) = \mathbf{C}$. Sugerencia: reemplazar $\boldsymbol{\theta}$ por $\boldsymbol{\alpha}$ en el modelo de los datos.

Solución Si el modelo de los datos es $\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$, el estimador BLUE de $\boldsymbol{\theta}$ es (ver la ecuación 6.5)

$$\hat{\boldsymbol{\theta}} = (\mathbf{H}^T \mathbf{C}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C}^{-1} \mathbf{x}.$$

Se quiere demostrar que el estimador del parámetro $\boldsymbol{\alpha}$, que es la transformación afín de $\boldsymbol{\theta}$ dada por $\boldsymbol{\alpha} = \mathbf{B}\boldsymbol{\theta} + \mathbf{b}$, es $\hat{\boldsymbol{\alpha}} = \mathbf{B}\hat{\boldsymbol{\theta}} + \mathbf{b}$. Para hacerlo, se despeja $\boldsymbol{\theta}$ de la transformación,

$$\boldsymbol{\theta} = \mathbf{B}^{-1}(\boldsymbol{\alpha} - \mathbf{b}),$$

y se sustituye en el modelo de los datos,

$$\begin{aligned} \mathbf{x} &= \mathbf{H}\boldsymbol{\theta} + \mathbf{w} \\ &= \mathbf{H}\mathbf{B}^{-1}(\boldsymbol{\alpha} - \mathbf{b}) + \mathbf{w} \\ &= \mathbf{H}\mathbf{B}^{-1}\boldsymbol{\alpha} - \mathbf{H}\mathbf{B}^{-1}\mathbf{b} + \mathbf{w}, \end{aligned}$$

resultando en

$$\mathbf{x} + \mathbf{H}\mathbf{B}^{-1}\mathbf{b} = \mathbf{H}\mathbf{B}^{-1}\boldsymbol{\alpha} + \mathbf{w},$$

que es el modelo lineal $\mathbf{x}' = \mathbf{H}'\boldsymbol{\alpha} + \mathbf{w}$ de parámetro $\boldsymbol{\alpha}$ con $\mathbf{x}' = \mathbf{x} + \mathbf{H}\mathbf{B}^{-1}\mathbf{b}$ y $\mathbf{H}' = \mathbf{H}\mathbf{B}^{-1}$. El estimador BLUE de $\boldsymbol{\alpha}$ es por lo tanto

$$\begin{aligned} \hat{\boldsymbol{\alpha}} &= (\mathbf{H}'^T \mathbf{C}^{-1} \mathbf{H}')^{-1} \mathbf{H}'^T \mathbf{C}^{-1} \mathbf{x}' \\ &= \left(\mathbf{B}^{-1T} \mathbf{H}^T \mathbf{C}^{-1} \mathbf{H} \mathbf{B}^{-1} \right)^{-1} \mathbf{B}^{-1T} \mathbf{H}^T \mathbf{C}^{-1} (\mathbf{x} + \mathbf{H}\mathbf{B}^{-1}\mathbf{b}) \\ &= \mathbf{B} (\mathbf{H}^T \mathbf{C}^{-1} \mathbf{H})^{-1} \mathbf{B}^T \mathbf{B}^{T-1} \mathbf{H}^T \mathbf{C}^{-1} (\mathbf{x} + \mathbf{H}\mathbf{B}^{-1}\mathbf{b}) \\ &= \mathbf{B} (\mathbf{H}^T \mathbf{C}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C}^{-1} \mathbf{x} + \mathbf{B} (\mathbf{H}^T \mathbf{C}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C}^{-1} \mathbf{H} \mathbf{B}^{-1} \mathbf{b} \\ &= \mathbf{B} (\mathbf{H}^T \mathbf{C}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C}^{-1} \mathbf{x} + \mathbf{B} \mathbf{B}^{-1} \mathbf{b} \\ &= \mathbf{B} (\mathbf{H}^T \mathbf{C}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C}^{-1} \mathbf{x} + \mathbf{b}, \end{aligned}$$

donde se consideró que $\mathbf{B}^{-1T} = \mathbf{B}^{T-1}$. Se concluye que el BLUE de $\boldsymbol{\alpha}$ es

$$\hat{\boldsymbol{\alpha}} = \mathbf{B}\hat{\boldsymbol{\theta}} + \mathbf{b},$$

que es lo que se quería demostrar.

6.4.13. Problema 13

En este problema se muestra que el BLUE es idéntico al *estimador por mínimos cuadrados ponderados* que se discute en el capítulo 8. En particular, el estimador por mínimos cuadrados ponderados se obtiene minimizando la función

$$J = (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T \mathbf{C}^{-1} (\mathbf{x} - \mathbf{H}\boldsymbol{\theta}).$$

Probar que el valor de $\boldsymbol{\theta}$ que minimiza J es el BLUE.

Solución Desarrollando la función J se tiene que

$$\begin{aligned} J &= (\mathbf{x}^T - \boldsymbol{\theta}^T \mathbf{H}^T) \mathbf{C}^{-1} (\mathbf{x} - \mathbf{H}\boldsymbol{\theta}) \\ &= \mathbf{x}^T \mathbf{C}^{-1} \mathbf{x} - 2\mathbf{x}^T \mathbf{C}^{-1} \mathbf{H}\boldsymbol{\theta} + \boldsymbol{\theta}^T \mathbf{H}^T \mathbf{C}^{-1} \mathbf{H}\boldsymbol{\theta}, \end{aligned}$$

y la derivada respecto al parámetro es

$$\frac{\partial J}{\partial \boldsymbol{\theta}} = -2\mathbf{H}^T \mathbf{C}^{-1} \mathbf{x} + 2\mathbf{H}^T \mathbf{C}^{-1} \mathbf{H}\boldsymbol{\theta},$$

donde se emplearon los resultados deducidos en el apéndice D que indican que

$$\frac{\partial \mathbf{b}^T \boldsymbol{\theta}}{\partial \boldsymbol{\theta}} = \mathbf{b}, \quad \frac{\partial \boldsymbol{\theta}^T \mathbf{A} \boldsymbol{\theta}}{\partial \boldsymbol{\theta}} = 2\mathbf{A}\boldsymbol{\theta}.$$

Finalmente, igualando la derivada a cero,

$$\mathbf{H}^T \mathbf{C}^{-1} \mathbf{H}\hat{\boldsymbol{\theta}} = \mathbf{H}^T \mathbf{C}^{-1} \mathbf{x}$$

y despejando el parámetro, se obtiene que el valor que minimiza J es

$$\hat{\boldsymbol{\theta}} = (\mathbf{H}^T \mathbf{C}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C}^{-1} \mathbf{x},$$

que es el BLUE.

6.4.14. Problema 14

Un proceso ruidoso se compone de variables aleatorias IID de media nula con PDF

$$p(w[n]) = \frac{1-\epsilon}{\sqrt{2\pi\sigma_B^2}} \exp \left[-\frac{1}{2} \left(\frac{w^2[n]}{\sigma_B^2} \right) \right] + \frac{\epsilon}{\sqrt{2\pi\sigma_I^2}} \exp \left[-\frac{1}{2} \left(\frac{w^2[n]}{\sigma_I^2} \right) \right],$$

donde $0 < \epsilon < 1$. Dicha PDF se denomina *mezcla gaussiana*. Se emplea para modelar ruido que es gaussiano de varianza σ_B^2 el $100(1-\epsilon)\%$ del tiempo y gaussiano de varianza σ_I^2 el resto del tiempo. Típicamente, $\sigma_I^2 \gg \sigma_B^2$ y $\epsilon \ll 1$, de forma que el ruido predominante es ruido de fondo con varianza σ_B^2 pequeña pero también contiene ocasionalmente ruido de alta potencia o interferencias modeladas como ruido gaussiano de varianza σ_I^2 grande. Mostrar que la varianza de esta PDF es

$$\sigma^2 = (1-\epsilon)\sigma_B^2 + \epsilon\sigma_I^2.$$

Si los datos son $\{w^2[0], \dots, w^2[N-1]\}$ y σ_B^2 y ϵ se asumen conocidos, encontrar el BLUE de σ_I^2 . Sugerencia: emplear los resultados del problema de la sección 6.4.12.

Solución Para calcular la varianza de $w[n]$ se parte observando que $E(w[n]) = 0$, ya que $w[n] \sim (1 - \epsilon)\mathcal{N}(0, \sigma_B^2) + \epsilon\mathcal{N}(0, \sigma_I^2)$. Por lo tanto, la varianza es

$$\begin{aligned}
\text{var}(w[n]) &= E(w^2[n]) \\
&= \int_{-\infty}^{\infty} w^2[n] p(w[n]) dw[n] \\
&= \int_{-\infty}^{\infty} w^2[n] \left\{ \frac{1 - \epsilon}{\sqrt{2\pi\sigma_B^2}} \exp\left[-\frac{1}{2} \left(\frac{w^2[n]}{\sigma_B^2}\right)\right] + \frac{\epsilon}{\sqrt{2\pi\sigma_I^2}} \exp\left[-\frac{1}{2} \left(\frac{w^2[n]}{\sigma_I^2}\right)\right] \right\} dw[n] \\
&= (1 - \epsilon) \int_{-\infty}^{\infty} w^2[n] \frac{1}{\sqrt{2\pi\sigma_B^2}} \exp\left[-\frac{1}{2} \left(\frac{w^2[n]}{\sigma_B^2}\right)\right] dw[n] \\
&\quad + \epsilon \int_{-\infty}^{\infty} w^2[n] \frac{1}{\sqrt{2\pi\sigma_I^2}} \exp\left[-\frac{1}{2} \left(\frac{w^2[n]}{\sigma_I^2}\right)\right] dw[n] \\
&\stackrel{(a)}{=} (1 - \epsilon)\sigma_B^2 + \epsilon\sigma_I^2,
\end{aligned}$$

donde en (a) se consideró que la primer integral es la varianza de una variable aleatoria $\mathcal{N}(0, \sigma_B^2)$ y la segunda integral es la varianza de una variable aleatoria $\mathcal{N}(0, \sigma_I^2)$.

Se calculará a continuación el BLUE de σ_I^2 . Teniendo en cuenta que los datos son $\{w^2[0], \dots, w^2[N-1]\}$, se considera la transformación $y[n] = w^2[n]$. De esta forma, se cumple que $E(y[n]) = \sigma^2$. Además, las muestras $y[n]$ son IID por ser las muestras $w[n]$ IID y por lo tanto, la matriz de covarianza del vector de las muestras $y[n]$ es $\mathbf{C} = \text{var}(y[n])\mathbf{I}$. Los datos transformados se pueden expresar como

$$\begin{aligned}
y[n] &= E(y[n]) + (y[n] - E(y[n])) \\
&= \sigma^2 + v[n]
\end{aligned}$$

o en notación matricial como

$$\mathbf{y} = \mathbf{1}\sigma^2 + \mathbf{v},$$

donde el ruido \mathbf{v} cumple que $E(\mathbf{v}) = \mathbf{0}$ y tiene matriz de covarianza la matriz de covarianza de \mathbf{y} . Por lo tanto, el BLUE de σ^2 esta dado por la ecuación 6.2 con $\mathbf{s} = \mathbf{1}$ y es

$$\hat{\sigma}^2 = \frac{\mathbf{1}^T \mathbf{C}^{-1} \mathbf{y}}{\mathbf{1}^T \mathbf{C}^{-1} \mathbf{1}} = \frac{\mathbf{1}^T \left(\frac{1}{\text{var}(y[n])} \mathbf{I} \right) \mathbf{y}}{\mathbf{1}^T \left(\frac{1}{\text{var}(y[n])} \mathbf{I} \right) \mathbf{1}} = \frac{\mathbf{1}^T \mathbf{y}}{\mathbf{1}^T \mathbf{1}} = \frac{1}{N} \sum_{n=0}^{N-1} y[n],$$

es decir,

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{n=0}^{N-1} w^2[n].$$

Además, como

$$\sigma^2 = (1 - \epsilon)\sigma_B^2 + \epsilon\sigma_I^2$$

se cumple que

$$\sigma_I^2 = \frac{\sigma^2 - (1 - \epsilon)\sigma_B^2}{\epsilon}.$$

En el problema de la sección 6.4.12 se dedujo que el BLUE se mantiene en trasformaciones afines del parámetro y por lo tanto, el BLUE de σ_I^2 es

$$\hat{\sigma}_I^2 = \frac{\hat{\sigma}^2 - (1 - \epsilon)\sigma_B^2}{\epsilon}$$

resultando en

$$\hat{\sigma}_I^2 = \frac{\frac{1}{N} \sum_{n=0}^{N-1} w^2[n] - (1 - \epsilon)\sigma_B^2}{\epsilon}.$$

6.4.15. Problema 15

Para el modelo lineal general

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{s} + \mathbf{w}$$

donde \mathbf{s} es un vector $N \times 1$ conocido y $E(\mathbf{w}) = 0$, $E(\mathbf{w}\mathbf{w}^T) = \mathbf{C}$, encontrar el BLUE.

Solución Definiendo $\mathbf{x}' = \mathbf{x} - \mathbf{s}$ el modelo de los datos se puede expresar como

$$\mathbf{x}' = \mathbf{H}\boldsymbol{\theta} + \mathbf{w},$$

que es el modelo lineal habitual. De la ecuación 6.5, el BLUE es

$$\hat{\boldsymbol{\theta}} = (\mathbf{H}^T \mathbf{C}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C}^{-1} \mathbf{x}',$$

resultando en

$$\hat{\boldsymbol{\theta}} = (\mathbf{H}^T \mathbf{C}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C}^{-1} (\mathbf{x} - \mathbf{s}).$$

6.4.16. Problema 16

En la implementación del BLUE para el modelo lineal descrito en el teorema de Gauss-Markov (sección 6.2.2), es posible usar sin saberlo una matriz de covarianza errónea para formar el estimador

$$\hat{\boldsymbol{\theta}} = (\mathbf{H}^T \hat{\mathbf{C}}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \hat{\mathbf{C}}^{-1} \mathbf{x}.$$

Para examinar el efecto de este error de modelado, asúmase que se quiere estimar el nivel de continua A en WGN en ruido no correlacionado. Calcular la varianza de este estimador y la varianza del estimador BLUE para el caso en que $N = 2$ y

$$\mathbf{C} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{y} \quad \hat{\mathbf{C}} = \begin{bmatrix} 1 & 0 \\ 0 & \alpha \end{bmatrix}.$$

Comparar las varianzas cuando $\alpha \rightarrow 0$, $\alpha = 1$ y $\alpha \rightarrow \infty$,

Solución El modelo de los datos es

$$\mathbf{x} = \mathbf{1}A + \mathbf{w},$$

y \mathbf{w} tiene matriz de covarianza \mathbf{C} . La varianza del estimador BLUE

$$\text{var}_{\text{BLUE}}(\hat{A}) = \frac{1}{\mathbf{1}^T \mathbf{C}^{-1} \mathbf{1}} = \frac{1}{2},$$

ya que $\mathbf{1}^T \mathbf{C}^{-1} \mathbf{1}$ es la suma de todos los elementos de \mathbf{C} . Para calcular la varianza del estimador empleando un modelo erróneo, se parte calculando la media del estimador. El estimador es

$$\hat{A} = \frac{\mathbf{1}^T \hat{\mathbf{C}}^{-1} \mathbf{x}}{\mathbf{1}^T \hat{\mathbf{C}}^{-1} \mathbf{1}},$$

por lo que su media es

$$\begin{aligned} E(\hat{A}) &= \frac{\mathbf{1}^T \hat{\mathbf{C}}^{-1} E(\mathbf{x})}{\mathbf{1}^T \hat{\mathbf{C}}^{-1} \mathbf{1}} \\ &= \frac{\mathbf{1}^T \hat{\mathbf{C}}^{-1} \mathbf{1} A}{\mathbf{1}^T \hat{\mathbf{C}}^{-1} \mathbf{1}} \\ &= A. \end{aligned}$$

Por lo tanto, el estimador es insesgado independientemente de la matriz $\hat{\mathbf{C}}$. La varianza del estimador es

$$\begin{aligned}
\text{var}(\hat{A}) &= E \left[(\hat{A} - A)^2 \right] \\
&= E \left[\left(\frac{\mathbf{1}^T \hat{\mathbf{C}}^{-1} E(\mathbf{x})}{\mathbf{1}^T \hat{\mathbf{C}}^{-1} \mathbf{1}} - A \right)^2 \right] \\
&= E \left[\left(\frac{\mathbf{1}^T \hat{\mathbf{C}}^{-1} E(\mathbf{x}) - \mathbf{1}^T \hat{\mathbf{C}}^{-1} \mathbf{1} A}{\mathbf{1}^T \hat{\mathbf{C}}^{-1} \mathbf{1}} \right)^2 \right] \\
&= E \left\{ \left[\frac{\mathbf{1}^T \hat{\mathbf{C}}^{-1} (E(\mathbf{x}) - A)}{\mathbf{1}^T \hat{\mathbf{C}}^{-1} \mathbf{1}} \right]^2 \right\} \\
&= E \left[\left(\frac{\mathbf{1}^T \hat{\mathbf{C}}^{-1} \mathbf{w}}{\mathbf{1}^T \hat{\mathbf{C}}^{-1} \mathbf{1}} \right)^2 \right] \\
&= E \left[\frac{\mathbf{1}^T \hat{\mathbf{C}}^{-1} \mathbf{w} \mathbf{w}^T \hat{\mathbf{C}}^{-1} \mathbf{1}}{(\mathbf{1}^T \hat{\mathbf{C}}^{-1} \mathbf{1})^2} \right] \\
&= \frac{\mathbf{1}^T \hat{\mathbf{C}}^{-1} E(\mathbf{w} \mathbf{w}^T) \hat{\mathbf{C}}^{-1} \mathbf{1}}{(\mathbf{1}^T \hat{\mathbf{C}}^{-1} \mathbf{1})^2},
\end{aligned}$$

concluyendo que

$$\text{var}(\hat{A}) = \frac{\mathbf{1}^T \hat{\mathbf{C}}^{-1} \mathbf{C} \hat{\mathbf{C}}^{-1} \mathbf{1}}{(\mathbf{1}^T \hat{\mathbf{C}}^{-1} \mathbf{1})^2}.$$

Sustituyendo los valores de las matrices \mathbf{C} y $\hat{\mathbf{C}}$, teniendo en cuenta que

$$\hat{\mathbf{C}}^{-1} \mathbf{C} \hat{\mathbf{C}}^{-1} = \hat{\mathbf{C}}^{-1} \mathbf{I} \hat{\mathbf{C}}^{-1} = \hat{\mathbf{C}}^{-1} \hat{\mathbf{C}}^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{\alpha} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{\alpha} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{\alpha^2} \end{bmatrix},$$

se obtiene que

$$\text{var}(\hat{A}) = \frac{1 + \frac{1}{\alpha^2}}{\left(1 + \frac{1}{\alpha}\right)^2} = \frac{\alpha^2 + 1}{(\alpha + 1)^2}.$$

Por lo tanto, se cumple que

$$\text{var}(\hat{A}) = \frac{\alpha^2 + 1}{(\alpha + 1)^2} = \begin{cases} 1 & \text{si } \alpha \rightarrow 0 \\ \frac{1}{2} & \text{si } \alpha = 1 \\ 1 & \text{si } \alpha \rightarrow \infty \end{cases}.$$

Cuando $\alpha = 1$, se obtiene el BLUE, que es el estimador lineal insesgado con menor varianza. En la figura 6.9 se muestra la varianza del estimador en función de α .

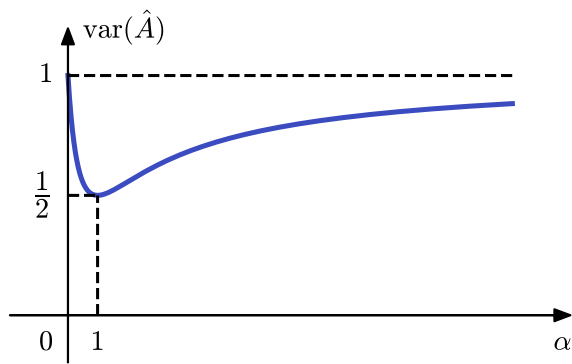


Figura 6.9: Varianza del estimador con modelado erróneo.

Capítulo 7

Estimación de máxima verosimilitud

7.1. Estimador de máxima verosimilitud

7.1.1. Definición

El estimador de máxima verosimilitud (*maximum likelihood estimator*, MLE) se define como el valor del parámetro θ que maximiza $p(\mathbf{x}; \theta)$ para \mathbf{x} fijo, es decir, el valor que maximiza la función de verosimilitud.

7.1.2. Propiedades asintóticas del MLE

Teorema Si la PDF $p(\mathbf{x}; \theta)$ de los datos \mathbf{x} satisface ciertas condiciones de regularidad, el estimador MLE del parámetro desconocido θ es asintóticamente distribuido como

$$\hat{\theta} \stackrel{a}{\sim} \mathcal{N}(\theta, I^{-1}(\theta))$$

donde $I(\theta)$ es la información de Fisher evaluada en el valor verdadero del parámetro desconocido.

En la demostración, para la cual se emplea la ley de los grandes números y el teorema central del límite, se especifican las condiciones de regularidad y los modos de convergencia del estimador MLE. Una introducción a los tipos de convergencia y teoremas sobre límites de secuencias de variables aleatorias puede encontrarse por ejemplo en el capítulo 7 de [2].

Demostración Se incluye a continuación una demostración poco rigurosa cuyo objetivo es ilustrar las ideas principales. Además, para simplificar la discusión, se asume que las observaciones son IID. Las condiciones de regularidad que tiene que cumplir la función de verosimilitud son las siguientes:

1. Las derivadas primera y segunda de la función de verosimilitud logarítmica están bien definidas.
- 2.

$$E \left[\frac{\partial \ln p(x[n]; \theta)}{\partial \theta} \right] = 0.$$

Esta es la misma condición de regularidad que la del teorema de CRLB.

Primero se demostrará que el MLE $\hat{\theta}$ es consistente. Esto significa que si θ_0 es el valor verdadero del parámetro, $\hat{\theta} \rightarrow \theta_0$ en probabilidad cuando $N \rightarrow \infty$. Para hacerlo, se necesitará la divergencia de Kullback-Leiber¹, que indica que

$$\int \ln \left[\frac{p(x)}{q(x)} \right] p(x) dx \geq 0, \quad (7.1)$$

donde $p(x)$ y $q(x)$ son densidades de probabilidad, y la igualdad se da si y solo si $q(x) = p(x)$. Para probar esta desigualdad, se observa que la formulación

$$\int \ln \left[\frac{q(x)}{p(x)} \right] p(x) dx \leq 0$$

¹Ver por ejemplo https://en.wikipedia.org/wiki/Kullback-Leibler_divergence.

es equivalente, y partiendo del lado izquierdo de esta última expresión, se ve que

$$\begin{aligned} \int \ln \left[\frac{q(x)}{p(x)} \right] p(x) dx &\stackrel{(a)}{\leq} \int \left[\frac{q(x)}{p(x)} - 1 \right] p(x) dx \\ &= \int q(x) dx - \int p(x) dx \\ &= 1 - 1 \\ &= 0, \end{aligned}$$

donde en (a) se empleó que $\ln x \leq x - 1$ para todo $x > 0$. La igualdad se da cuando $\ln x = x - 1$, es decir, en $x = 1$, que implica que $q(x) = p(x)$. Maximizar la función de verosimilitud logarítmica es equivalente a maximizar

$$\begin{aligned} \frac{1}{N} \ln p(\mathbf{x}; \theta) &= \frac{1}{N} \ln \prod_{n=0}^{N-1} p(x[n]; \theta) \\ &= \frac{1}{N} \sum_{n=0}^{N-1} \ln p(x[n]; \theta). \end{aligned}$$

Cuando $N \rightarrow \infty$, por la ley de los grandes números, esto converge en probabilidad al valor esperado. Por lo tanto,

$$\frac{1}{N} \sum_{n=0}^{N-1} \ln p(x[n]; \theta) \xrightarrow{P} E[\ln p(x[n]; \theta)] = \int \ln p(x[n]; \theta) p(x[n]; \theta_0) dx[n], \quad (7.2)$$

donde θ_0 es el valor verdadero del parámetro. Considerando la divergencia de Kullback-Leiber de la ecuación 7.1, se cumple que

$$\int \ln \left[\frac{p(x[n]; \theta_1)}{p(x[n]; \theta_2)} \right] p(x[n]; \theta_1) dx[n] \geq 0,$$

es decir,

$$\int \ln p(x[n]; \theta_1) p(x[n]; \theta_1) dx[n] \geq \int \ln p(x[n]; \theta_2) p(x[n]; \theta_1) dx[n].$$

Esto indica que el lado derecho de la ecuación 7.2 es maximizado para $\theta = \theta_0$. Por lo tanto, asumiendo condiciones de continuidad apropiadas, el lado izquierdo de la ecuación 7.2 también es maximizado para $\theta = \theta_0$. Como por definición el MLE $\hat{\theta}$ maximiza el lado izquierdo 7.2, se cumple que cuando $N \rightarrow \infty$, el MLE es $\hat{\theta} = \theta_0$. Se concluye que el MLE es consistente.

Se deriva a continuación la PDF asintótica del MLE, que cumple que

$$\hat{\theta} \xrightarrow{d} \mathcal{N}(\theta_0, I^{-1}(\theta_0)),$$

es decir, el estimador converge a una distribución normal de media θ_0 y varianza $I^{-1}(\theta_0)$. Para demostrarlo, se parte considerando el teorema del valor medio², que indica que bajo las hipótesis adecuadas se cumple que

$$f'(c) = \frac{f(b) - f(a)}{b - a}, \quad \text{o} \quad f(b) = f(a) + f'(c)(b - a) \quad \text{para } a < c < b.$$

Aplicando el teorema del valor medio a

$$f(\theta) = \frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \theta} \quad \text{con} \quad a = \theta_0, \quad b = \hat{\theta},$$

se cumple que

$$\left. \frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \theta} \right|_{\theta=\hat{\theta}} = \left. \frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \theta} \right|_{\theta=\theta_0} + \left. \frac{\partial^2 \ln p(\mathbf{x}; \theta)}{\partial \theta^2} \right|_{\theta=\hat{\theta}} (\hat{\theta} - \theta_0)$$

²Ver https://en.wikipedia.org/wiki/Mean_value_theorem.

para algún $\tilde{\theta}$ tal que $\theta_0 < \tilde{\theta} < \hat{\theta}$. Pero por definición del MLE,

$$\left. \frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \theta} \right|_{\theta=\hat{\theta}} = 0$$

y por lo tanto,

$$0 = \left. \frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \theta} \right|_{\theta=\theta_0} + \left. \frac{\partial^2 \ln p(\mathbf{x}; \theta)}{\partial \theta^2} \right|_{\theta=\tilde{\theta}} (\hat{\theta} - \theta_0).$$

Despejando $\hat{\theta} - \theta_0$ y multiplicando ambos lados de la igualdad por \sqrt{N} , se tiene que

$$\sqrt{N}(\hat{\theta} - \theta_0) = - \frac{\left. \sqrt{N} \frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \theta} \right|_{\theta=\theta_0}}{\left. \frac{\partial^2 \ln p(\mathbf{x}; \theta)}{\partial \theta^2} \right|_{\theta=\tilde{\theta}}}$$

y dividiendo el numerador y el denominador entre N se llega a que

$$\sqrt{N}(\hat{\theta} - \theta_0) = \frac{\frac{1}{\sqrt{N}} \left. \frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \theta} \right|_{\theta=\theta_0}}{- \frac{1}{N} \left. \frac{\partial^2 \ln p(\mathbf{x}; \theta)}{\partial \theta^2} \right|_{\theta=\tilde{\theta}}}. \quad (7.3)$$

Se considera ahora el denominador de la ecuación 7.3. Como los datos son IID, se cumple que

$$\frac{1}{N} \left. \frac{\partial^2 \ln p(\mathbf{x}; \theta)}{\partial \theta^2} \right|_{\theta=\tilde{\theta}} = \frac{1}{N} \sum_{n=0}^{N-1} \left. \frac{\partial^2 \ln p(x[n]; \theta)}{\partial \theta^2} \right|_{\theta=\tilde{\theta}}.$$

Como $\theta_0 < \tilde{\theta} < \hat{\theta}$, se cumple que $\tilde{\theta} \rightarrow \theta_0$ con $N \rightarrow \infty$ debido a la consistencia del MLE. Por lo tanto,

$$\begin{aligned} \frac{1}{N} \left. \frac{\partial^2 \ln p(\mathbf{x}; \theta)}{\partial \theta^2} \right|_{\theta=\tilde{\theta}} &\rightarrow \frac{1}{N} \sum_{n=0}^{N-1} \left. \frac{\partial^2 \ln p(x[n]; \theta)}{\partial \theta^2} \right|_{\theta=\theta_0} \\ &\stackrel{(a)}{\rightarrow} E \left[\left. \frac{\partial^2 \ln p(x[n]; \theta)}{\partial \theta^2} \right|_{\theta=\theta_0} \right] \\ &= -i(\theta_0). \end{aligned}$$

donde la convergencia en (a) es probabilidad por la ley de los grandes números. $i(\theta_0)$ es la información de Fisher de una muestra. Empleando nuevamente la hipótesis de IID, el numerador de la ecuación 7.3 es

$$\frac{1}{\sqrt{N}} \left. \frac{\partial \ln p(\mathbf{x}; \theta)}{\partial \theta} \right|_{\theta=\theta_0} = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} \left. \frac{\partial \ln p(x[n]; \theta)}{\partial \theta} \right|_{\theta=\theta_0}.$$

Se observa que

$$\xi_n = \left. \frac{\partial \ln p(x[n]; \theta)}{\partial \theta} \right|_{\theta=\theta_0}$$

es una variable aleatoria por ser una función de $x[n]$. Además, como los $x[n]$ son IID, también lo son los ξ_n . Por lo tanto, por el teorema central del límite, el numerador de la ecuación 7.3 tiene una PDF que converge a una PDF gaussiana. La media de esta PDF gaussiana es

$$\begin{aligned} E \left[\frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} \left. \frac{\partial \ln p(x[n]; \theta)}{\partial \theta} \right|_{\theta=\theta_0} \right] &= \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} E \left[\left. \frac{\partial \ln p(x[n]; \theta)}{\partial \theta} \right|_{\theta=\theta_0} \right] \\ &\stackrel{(a)}{=} \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} \left. \frac{\partial E[\ln p(x[n]; \theta)]}{\partial \theta} \right|_{\theta=\theta_0} \\ &\stackrel{(b)}{=} 0, \end{aligned}$$

donde en (a) se intercambi6 el orden de diferenciación y esperanza, lo cual es posible gracias a la condición de regularidad del punto 2., como se mostr6 en la sección 3.1.1, y en (b) se tuvo en cuenta que el término $E[\ln p(x[n]; \theta)]$ está dado por la integral de la ecuación 7.2, la cual es una función de θ que es máxima cuando $\theta = \theta_0$, por lo que su derivada respecto a θ evaluada en el máximo θ_0 debe ser nula. Además, la varianza es

$$\begin{aligned} E \left[\left(\frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} \frac{\partial \ln p(x[n]; \theta)}{\partial \theta} \bigg|_{\theta=\theta_0} \right)^2 \right] &\stackrel{(a)}{=} \frac{1}{N} \sum_{n=0}^{N-1} E \left[\left(\frac{\partial \ln p(x[n]; \theta)}{\partial \theta} \right)^2 \bigg|_{\theta=\theta_0} \right] \\ &\stackrel{(b)}{=} \frac{1}{N} N i(\theta_0) \\ &= i(\theta_0), \end{aligned}$$

donde en (a) se consider6 que las variables aleatorias ξ_n son IID y de media nula y en (b) se emple6 la expresi6n alternativa de la informaci6n de Fisher dada por la ecuaci6n 3.6. Para continuar la deducci6n, hay que referirse al teorema de Slutsky³, que indica que si la secuencia de variables aleatorias x_n converge en distribuci6n a la variable aleatoria x y la secuencia de variables aleatorias y_n converge en probabilidad a una constante c , se cumple que x_n/y_n converge en distribuci6n a la variable aleatoria x/c . En el caso de la ecuaci6n 7.3, se vi6 que el numerador y el denominador convergen respectivamente a

$$x_n \xrightarrow{d} x \sim \mathcal{N}(0, i(\theta_0)), \quad y_n \xrightarrow{p} c = i(\theta_0),$$

y por lo tanto, como

$$\text{var} \left(\frac{x}{c} \right) = \frac{\text{var}(x)}{c^2} = \frac{i(\theta_0)}{i^2(\theta_0)} = i^{-1}(\theta_0),$$

la ecuaci6n 7.3 converge asint6ticamente a

$$\sqrt{N}(\hat{\theta} - \theta_0) \stackrel{a}{\sim} \mathcal{N}(0, i^{-1}(\theta_0)),$$

o equivalentemente

$$\hat{\theta} \stackrel{a}{\sim} \mathcal{N} \left(\theta_0, \frac{1}{N i(\theta_0)} \right),$$

y finalmente

$$\hat{\theta} \stackrel{a}{\sim} \mathcal{N}(\theta_0, I^{-1}(\theta_0)),$$

donde en esta última ecuaci6n se tuvo en cuenta que como los datos son IID, la informaci6n de Fisher $I(\theta_0)$ de los datos es la suma de la informaci6n de Fisher $i(\theta_0)$ de cada muestra.

7.1.3. Propiedad de invarianza del MLE

Teorema: propiedad de invarianza del MLE. El MLE del parámetro $\alpha = g(\theta)$, donde la PDF $p(\mathbf{x}, \theta)$ es parametrizada por θ , está dada por

$$\hat{\alpha} = g(\hat{\theta})$$

donde $\hat{\theta}$ es el MLE de θ . El MLE de $\hat{\theta}$ se obtiene maximizando $p(\mathbf{x}, \theta)$. Si g no es una funci6n uno a uno (inyectiva), $\hat{\alpha}$ maximiza la funci6n de verosimilitud modificada $\bar{p}_T(\mathbf{x}, \alpha)$, definida como

$$\bar{p}_T(\mathbf{x}, \alpha) = \max_{\{\theta: \alpha=g(\theta)\}} p(\mathbf{x}; \theta).$$

7.1.4. MLE para el modelo lineal

Teorema: optimalidad del MLE para el modelo lineal. Si los datos observados \mathbf{x} son descriptos por el modelo lineal general (ecuaci6n 4.1)

$$\mathbf{x} = \mathbf{H}\theta + \mathbf{w}$$

³Ver por ejemplo https://en.wikipedia.org/wiki/Slutsky's_theorem.

donde \mathbf{H} es una matriz $N \times p$ conocida con $N > p$ y rango p , $\boldsymbol{\theta}$ es un vector $p \times 1$ de parámetros a estimar y \mathbf{w} es un vector de ruido $N \times 1$ con PDF $\mathcal{N}(\mathbf{0}, \mathbf{C})$, el MLE de $\boldsymbol{\theta}$ es

$$\hat{\boldsymbol{\theta}} = (\mathbf{H}^T \mathbf{C}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C}^{-1} \mathbf{x}. \quad (7.4)$$

$\hat{\boldsymbol{\theta}}$ es además un estimador eficiente ya que alcanza la CRLB y por lo tanto, es el estimador MVU. La PDF de $\hat{\boldsymbol{\theta}}$ es

$$\hat{\boldsymbol{\theta}} \sim \mathcal{N}(\boldsymbol{\theta}, (\mathbf{H}^T \mathbf{C}^{-1} \mathbf{H})^{-1}).$$

Este resultado puede generalizarse mediante la afirmación de que si existe un estimador eficiente, está dado por el MLE, como se muestra en el problema de la sección 7.6.12.

Demostración. Para el modelo lineal general, la PDF de los datos es

$$p(\mathbf{x}; \boldsymbol{\theta}) = \frac{1}{(2\pi)^{\frac{N}{2}} \det^{\frac{1}{2}}(\mathbf{C})} \exp \left[-\frac{1}{2} (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T \mathbf{C}^{-1} (\mathbf{x} - \mathbf{H}\boldsymbol{\theta}) \right]$$

por lo que el MLE se obtiene minimizando

$$J(\boldsymbol{\theta}) = (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T \mathbf{C}^{-1} (\mathbf{x} - \mathbf{H}\boldsymbol{\theta}), \quad (7.5)$$

y como es una función cuadrática en los elementos de $\boldsymbol{\theta}$ y \mathbf{C}^{-1} es una matriz definida positiva, la diferenciación produce el mínimo global. En la sección 3.4 sobre la CRLB para el caso gaussiano general, se demostró que si $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}(\boldsymbol{\theta}), \mathbf{C}(\boldsymbol{\theta}))$,

$$\frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \theta_k} = -\frac{1}{2} \text{tr} \left(\mathbf{C}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \theta_k} \right) + \frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})^T}{\partial \theta_k} \mathbf{C}^{-1}(\boldsymbol{\theta}) (\mathbf{x} - \boldsymbol{\mu}(\boldsymbol{\theta})) - \frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}(\boldsymbol{\theta}))^T \frac{\partial \mathbf{C}^{-1}(\boldsymbol{\theta})}{\partial \theta_k} (\mathbf{x} - \boldsymbol{\mu}(\boldsymbol{\theta}))$$

(ver la ecuación 3.33). En este caso, $\boldsymbol{\mu}(\boldsymbol{\theta}) = \mathbf{H}\boldsymbol{\theta}$ y la matriz de covarianza no depende de $\boldsymbol{\theta}$, por lo que la derivada primera se reduce a

$$\frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \theta_k} = \frac{\partial (\mathbf{H}\boldsymbol{\theta})^T}{\partial \theta_k} \mathbf{C}^{-1} (\mathbf{x} - \mathbf{H}\boldsymbol{\theta}),$$

y combinando las derivadas parciales para formar el gradiente se obtiene que

$$\frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{\partial (\mathbf{H}\boldsymbol{\theta})^T}{\partial \boldsymbol{\theta}} \mathbf{C}^{-1} (\mathbf{x} - \mathbf{H}\boldsymbol{\theta}).$$

Para calcular $\partial (\mathbf{H}\boldsymbol{\theta})^T / \partial \boldsymbol{\theta}$ se define

$$f(\boldsymbol{\theta}) = (\mathbf{H}\boldsymbol{\theta})^T \quad \Rightarrow \quad f(\boldsymbol{\theta} + d\boldsymbol{\theta}) = [\mathbf{H}(\boldsymbol{\theta} + d\boldsymbol{\theta})]^T = (\mathbf{H}\boldsymbol{\theta})^T + d\boldsymbol{\theta}^T \mathbf{H}^T$$

y comparando con la ecuación D.1, se llega a que $\partial (\mathbf{H}\boldsymbol{\theta})^T / \partial \boldsymbol{\theta} = \mathbf{H}^T$. Igualando el gradiente a cero, se obtiene que el MLE cumple que

$$\mathbf{H}^T \mathbf{C}^{-1} (\mathbf{x} - \mathbf{H}\hat{\boldsymbol{\theta}}) = \mathbf{0} \quad \Rightarrow \quad \mathbf{H}^T \mathbf{C}^{-1} \mathbf{H} \hat{\boldsymbol{\theta}} = \mathbf{H}^T \mathbf{C}^{-1} \mathbf{x},$$

resultando en

$$\hat{\boldsymbol{\theta}} = (\mathbf{H}^T \mathbf{C}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C}^{-1} \mathbf{x}.$$

En la sección 4.5 se mostró que este estimador es el MVU (ver la ecuación 4.20) y además es eficiente. Por lo tanto, $\hat{\boldsymbol{\theta}}$ es insesgado y tiene matriz de covarianza dada por la ecuación 4.21,

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}} = (\mathbf{H}^T \mathbf{C}^{-1} \mathbf{H})^{-1}$$

Finalmente, la PDF del MLE es gaussiana por ser una función lineal de \mathbf{x} . Se concluye que para el modelo lineal, el MLE es óptimo.

7.2. Ejemplos básicos

7.2.1. Ejemplo: MLE de los parámetros de ruido de Laplace

Se consideran las observaciones $\{x[0], x[1], \dots, x[N-1]\}$ IID con distribución de Laplace,

$$p(x; \mu, \sigma) = \frac{1}{\sqrt{2}\sigma} \exp\left(-\frac{\sqrt{2}|x - \mu|}{\sigma}\right).$$

En el ejemplo de la sección 3.1.1 se encontró el estimador MVU y la CRLB de σ cuando μ es conocido, pero no se pudo encontrar el MVU de μ con la técnica de Cramer-Rao. A continuación, se calcularán los estimadores empleando la técnica de máxima verosimilitud. La función de verosimilitud logarítmica es

$$\ln p(\mathbf{x}; \mu, \sigma) = -N \ln \sqrt{2} - N \ln \sigma - \frac{\sqrt{2}}{\sigma} \sum_{n=0}^{N-1} |x[n] - \mu|.$$

Supóngase que la media μ es conocida y se quiere encontrar el estimador MVU de la desviación estándar σ . La derivada respecto al parámetro es

$$\frac{\partial \ln p(\mathbf{x}; \sigma)}{\partial \sigma} = -\frac{N}{\sigma} + \frac{\sqrt{2}}{\sigma^2} \sum_{n=0}^{N-1} |x[n] - \mu|,$$

e igualando a cero, se obtiene que el MLE es

$$\hat{\sigma} = \frac{\sqrt{2}}{N} \sum_{n=0}^{N-1} |x[n] - \mu|.$$

Comparando con el resultado de la sección 3.1.1, se concluye que en este caso el MLE es el MVU eficiente.

Se asume ahora que σ es conocido y se quiere estimar el MLE de μ . El cálculo está basado en [11]. La función de verosimilitud logarítmica es

$$\ln p(\mathbf{x}; \mu) = -N \ln \sqrt{2}\sigma - \frac{\sqrt{2}}{\sigma} \sum_{n=0}^{N-1} |x[n] - \mu|.$$

Esta función es continua en todo μ y diferenciable excepto en los valores $\mu = x[n]$ con $n = 0, \dots, N-1$. Cuando la derivada existe, vale

$$\frac{\ln p(\mathbf{x}; \mu)}{\partial \mu} = \frac{\sqrt{2}}{\sigma} \sum_{n=0}^{N-1} \text{sig}(x[n] - \mu).$$

Sean

$$x_{(1)} < x_{(2)} < \dots < x_{(N)}$$

los estadísticos de orden de las observaciones, es decir, las observaciones ordenadas de menor a mayor (ver el capítulo 6 de [2]). La derivada

$$\sum_{n=0}^{N-1} \text{sig}(x[n] - \mu)$$

vale N para $\mu < x_{(1)}$, $N-2$ si $x_{(1)} < \mu < x_{(2)}$, $N-4$ si $x_{(2)} < \mu < x_{(3)}$, y así sucesivamente, como se muestra en la figura 7.1. Además, la gráfica de $\ln p(\mathbf{x}; \mu)$ es una función continua lineal a trozos, como se muestra en la figura. Se observa que si N es impar, $\ln p(\mathbf{x}; \mu)$ es estrictamente creciente en el intervalo $(-\infty, x_{(N+1)/2}]$ y estrictamente decreciente en el intervalo $[x_{(N+1)/2}, \infty)$ y por lo tanto, la verosimilitud se maximiza en $\mu = x_{(N+1)/2}$. Observar que si bien la función tiene un máximo absoluto, la derivada nunca se anula. Concretamente, la derivada no existe en el punto donde se da el máximo. En el caso en que N es par, el máximo de la función se da en el segmento horizontal $[x_{(N/2)}, x_{(N/2+1)}]$, como se aprecia en la figura 7.1. Teniendo en cuenta que la mediana se define como $x_{(N+1)/2}$ si N es impar y $(x_{(N/2)} + x_{(N/2+1)})/2$, es decir, el punto medio del segmento $[x_{(N/2)}, x_{(N/2+1)}]$ si N es par, se concluye que el MLE es

$$\hat{\mu} = \text{mediana}(\mathbf{x}).$$

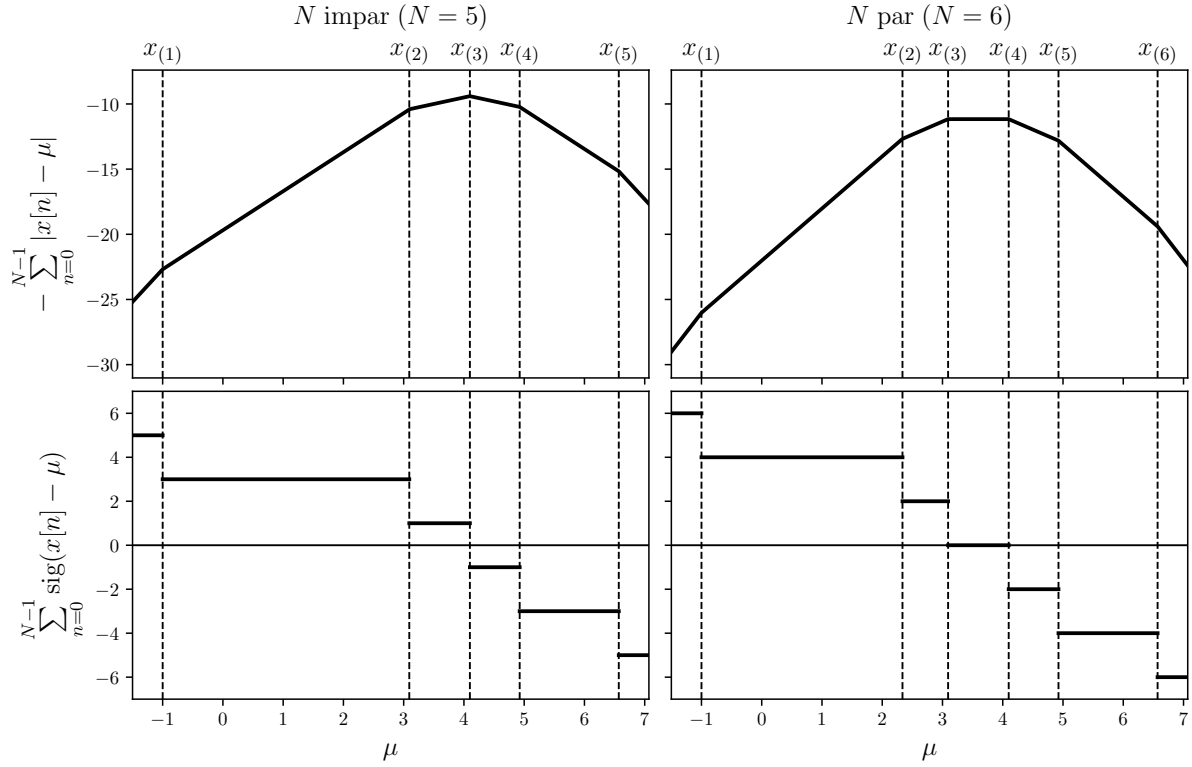


Figura 7.1: Función de verosimilitud logarítmica y su derivada cuando las observaciones son muestras IID con PDF de Laplace con $\mu = 4$ y $\sigma = 1$. Las líneas punteadas verticales indican los valores de las observaciones. Se observa que la verosimilitud es una función continua. En el caso en que N es impar, la verosimilitud es máxima en $x_{(N+1)/2}$, si bien la derivada no existe en ese valor. Si N es par, la verosimilitud es máxima en el segmento horizontal $[x_{N/2}, x_{N/2+1}]$. En ambos casos, la verosimilitud se maximiza en $\mu = \text{mediana}(\mathbf{x})$.

7.2.2. Ejemplo: MLE de la fase de una senoide

Se reconsidera el problema de la sección 3.3.1 de la estimación de la fase ϕ de una senoide en WGN, o

$$x[n] = A \cos(2\pi f_0 n + \phi) + w[n], \quad n = 0, \dots, N-1,$$

donde $w[n]$ es WGN de varianza σ^2 y la amplitud A y la frecuencia f_0 de la senoide son conocidas. El MLE se obtiene maximizando la PDF respecto al parámetro. La PDF es

$$p(\mathbf{x}, \phi) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} [x[n] - A \cos(2\pi f_0 n + \phi)]^2 \right\},$$

y su maximización es equivalente a la minimización de la función

$$J(\phi) = \sum_{n=0}^{N-1} [x[n] - A \cos(2\pi f_0 n + \phi)]^2.$$

Diferenciando respecto a ϕ se obtiene que

$$\frac{\partial J(\phi)}{\partial \phi} = 2 \sum_{n=0}^{N-1} [x[n] - A \cos(2\pi f_0 n + \phi)] A \sin(2\pi f_0 n + \phi),$$

e igualando a cero, se obtiene que el MLE cumple que

$$\sum_{n=0}^{N-1} x[n] \sin(2\pi f_0 n + \hat{\phi}) = A \sum_{n=0}^{N-1} \cos(2\pi f_0 n + \hat{\phi}) \sin(2\pi f_0 n + \hat{\phi}).$$

Empleando la identidad trigonométrica $\sin(2\theta) = 2 \sin \theta \cos \theta$ y la aproximación demostrada en el problema 3.6.7, el lado derecho de la igualdad puede ser aproximado como

$$\frac{1}{N} \sum_{n=0}^{N-1} \cos(2\pi f_0 n + \hat{\phi}) \sin(2\pi f_0 n + \hat{\phi}) = \frac{1}{2N} \sum_{n=0}^{N-1} \sin(4\pi f_0 n + 2\hat{\phi}) \approx 0$$

si f_0 no es cercano a 0 o a $1/2$. Por lo tanto, al dividir el lado derecho entre N e igualar a cero, se obtiene que el MLE aproximado cumple que

$$\sum_{n=0}^{N-1} x[n] \sin(2\pi f_0 n + \hat{\phi}) = 0$$

Considerando que $\sin(\alpha \pm \beta) = \sin \alpha \cos \beta \pm \cos \alpha \sin \beta$, se obtiene que

$$\sum_{n=0}^{N-1} x[n] \sin 2\pi f_0 n \cos \hat{\phi} = - \sum_{n=0}^{N-1} x[n] \cos 2\pi f_0 n \sin \hat{\phi},$$

y finalmente, el MLE de la fase es aproximadamente

$$\hat{\phi} = - \arctan \frac{\sum_{n=0}^{N-1} x[n] \sin 2\pi f_0 n}{\sum_{n=0}^{N-1} x[n] \cos 2\pi f_0 n}.$$

En el ejemplo de la sección 5.1.3 se encontró que no existe un estadístico suficiente individual en este problema. Los estadísticos suficientes están dados en la ecuación 5.9. Se observa que el estimador MLE es una función de los estadísticos suficientes. Como la factorización de Neyman-Fisher resultó en

$$p(\mathbf{x}; \phi) = g(T_1(\mathbf{x}), T_2(\mathbf{x}), \phi) h(\mathbf{x}),$$

maximizar $p(\mathbf{x}; \phi)$ respecto a ϕ es equivalente a maximizar g , ya que $h(\mathbf{x})$ siempre puede elegirse de forma que $h(\mathbf{x}) > 0$, y por lo tanto, $\hat{\phi}$ debe ser una función de $T_1(\mathbf{x})$ y $T_2(\mathbf{x})$.

Las propiedades asintóticas del MLE pueden alcanzarse para un largo de datos N pequeño si la SNR es suficientemente grande. En efecto, sustituyendo el modelo de los datos en el MLE de la fase, el estimador puede escribirse como

$$\hat{\phi} = - \arctan \frac{\sum_{n=0}^{N-1} [A \cos(2\pi f_0 n + \phi) + w[n]] \sin 2\pi f_0 n}{\sum_{n=0}^{N-1} [A \cos(2\pi f_0 n + \phi) + w[n]] \cos 2\pi f_0 n}.$$

Considerando las identidades trigonométricas $2 \cos \theta \sin \varphi = \sin(\theta + \varphi) - \sin(\theta - \varphi)$ y $2 \cos \theta \cos \varphi = \cos(\theta - \varphi) + \cos(\theta + \varphi)$, se tiene que

$$\begin{aligned} \cos(2\pi f_0 n + \phi) \sin 2\pi f_0 n &= \frac{1}{2} [\sin(4\pi f_0 n + \phi) - \sin \phi] \\ \cos(2\pi f_0 n + \phi) \cos 2\pi f_0 n &= \frac{1}{2} [\cos \phi + \cos(4\pi f_0 n + \phi)], \end{aligned}$$

y por lo tanto

$$\begin{aligned}
\hat{\phi} &= -\arctan \frac{\frac{A}{2} \sum_{n=0}^{N-1} [\sin(4\pi f_0 n + \phi) - \sin \phi] + \sum_{n=0}^{N-1} w[n] \sin 2\pi f_0 n}{\frac{A}{2} \sum_{n=0}^{N-1} [\cos \phi + \cos(4\pi f_0 n + \phi)] + \sum_{n=0}^{N-1} w[n] \cos 2\pi f_0 n} \\
&= -\arctan \frac{-\frac{NA}{2} \sin \phi + \frac{A}{2} \sum_{n=0}^{N-1} \sin(4\pi f_0 n + \phi) + \sum_{n=0}^{N-1} w[n] \sin 2\pi f_0 n}{\frac{NA}{2} \cos \phi + \frac{A}{2} \sum_{n=0}^{N-1} \cos(4\pi f_0 n + \phi) + \sum_{n=0}^{N-1} w[n] \cos 2\pi f_0 n} \\
&= \arctan \frac{\sin \phi - \frac{1}{N} \sum_{n=0}^{N-1} \sin(4\pi f_0 n + \phi) - \frac{2}{NA} \sum_{n=0}^{N-1} w[n] \sin 2\pi f_0 n}{\cos \phi + \frac{1}{N} \sum_{n=0}^{N-1} \cos(4\pi f_0 n + \phi) + \frac{2}{NA} \sum_{n=0}^{N-1} w[n] \cos 2\pi f_0 n}
\end{aligned}$$

y finalmente, considerando nuevamente la aproximación demostrada en el problema 3.6.7, se obtiene que

$$\hat{\phi} \approx \arctan \frac{\sin \phi - \frac{2}{NA} \sum_{n=0}^{N-1} w[n] \cos 2\pi f_0 n}{\cos \phi + \frac{2}{NA} \sum_{n=0}^{N-1} w[n] \sin 2\pi f_0 n}. \quad (7.6)$$

Si la cantidad de datos N es grande y/o si la potencia de la senoide ($A^2/2$) es grande, los términos de ruido son pequeños. La condición de que el error de estimación sea pequeño es lo que le permite al MLE alcanzar la distribución asintótica. Ver el problema de la sección 7.6.15 por una discusión adicional sobre esto.

7.2.3. Ejemplo: nivel de DC en ruido no independiente no gaussiano

Se consideran las observaciones

$$x[n] = A + w[n] \quad n = 0, \dots, N-1$$

donde cada muestra de $w[n]$ tiene PDF $p(w[n])$ simétrica respecto a $w[n] = 0$ y con valor máximo en $w[n] = 0$ como se muestra en la figura 7.2. Además, se asume que todas las muestras de ruido son

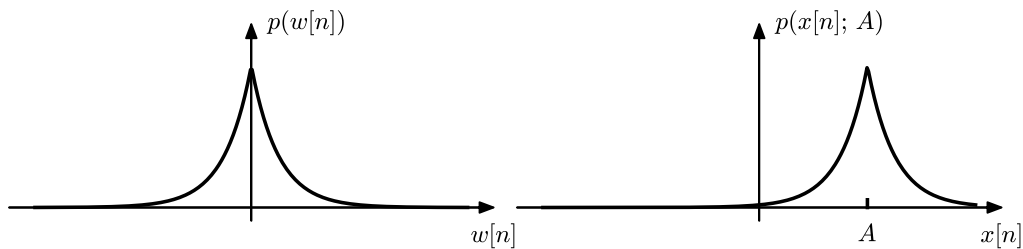


Figura 7.2: PDF no gaussiana del ruido $w[n]$ y de las observaciones $x[n] = A + w[n]$.

iguales, o $w[0] = w[1] = \dots = w[N-1]$. Como todas las observaciones son idénticas, alcanza con considerar una única observación en la estimación de A . Empleando solo $x[0]$, se observa que la PDF de $x[0]$ es una versión desplazada de $p(w[n])$ una cantidad A , como se muestra en la figura 7.2. Esto es

porque $p_{x[0]}(x[0]; A) = p_{w[0]}(x[0] - A)$ (ver el problema 1.1.3). El MLE de A es el valor que maximiza $p_{w[0]}(x[0] - A)$, y es

$$\hat{A} = x[0].$$

Este estimador tiene media

$$E(\hat{A}) = E(x[0]) = A$$

ya que la PDF del ruido es simétrica respecto a $w[0] = 0$. La varianza de \hat{A} es la misma que la varianza de $x[0]$ o $w[0]$. Por lo tanto,

$$\text{var}(\hat{A}) = \int_{-\infty}^{\infty} u^2 p_{w[0]}(u) du,$$

mientras que la CRLB, como se dedujo en el problema 3.6.2, es

$$\text{var}(\hat{A}) \geq \left[\int_{-\infty}^{\infty} \frac{\left(\frac{dp_{w[0]}(u)}{du} \right)^2}{p_{w[0]}(u)} du \right]^{-1},$$

y en general los dos valores son distintos (ver por ejemplo el problema de la sección 7.6.16). En este ejemplo, el error de estimación no decrece con el incremento del largo de los datos, sino que permanece constante. Además, la PDF de $\hat{A} = x[0]$, es claramente no gaussiana por ser la PDF de $w[0]$ desplazada. Finalmente, el estimador \hat{A} ni siquiera es consistente.

7.3. Algoritmo esperanza-maximización

7.3.1. Introducción

Un método para determinar de forma numérica el estimador MLE es el algoritmo esperanza-maximización (*expectation-maximization*, EM). Es un algoritmo iterativo en el que en cada paso se obtiene una estimación del parámetro que incrementa la función de verosimilitud. Bajo ciertas condiciones no muy exigentes, el algoritmo converge al menos a un máximo local. En esta sección se describe brevemente el algoritmo. Para hacerlo, se comienza considerando el ejemplo de estimación de las frecuencias de una mezcla de sinusoides, donde el modelo de los datos es

$$x[n] = \sum_{p=1}^p \cos 2\pi f_i n + w[n], \quad n = 0, \dots, N-1,$$

donde $w[n]$ es WGN de varianza σ^2 y las frecuencias $\mathbf{f} = [f_1 \dots f_p]$ son los parámetros desconocidos a estimar. Como la PDF de los datos es

$$p(\mathbf{x}, \mathbf{f}) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} \left(x[n] - \sum_{p=1}^p \cos 2\pi f_i n \right)^2 \right],$$

el algoritmo MLE requiere la minimización de la función de dimensión p

$$J(\mathbf{f}) = \left(x[n] - \sum_{p=1}^p \cos 2\pi f_i n \right)^2.$$

Supóngase que el conjunto original de datos pudiera ser remplazado por el conjunto de datos independientes

$$\begin{aligned} y_i[n] &= \cos 2\pi f_i n + w_i[n], & i &= 1, \dots, p \\ n &= 0, \dots, N-1 \end{aligned} \quad (7.7)$$

donde $w_i[n]$ es WGN con varianza σ_i^2 . Asumiendo que los ruidos $w_i[n]$ son procesos independientes, la PDF de los datos es

$$\begin{aligned} p(\mathbf{y}; \mathbf{f}) &= \prod_{i=1}^p p(\mathbf{y}_i; \mathbf{f}) \\ &= \prod_{i=1}^p \frac{1}{(2\pi\sigma_i^2)^{\frac{N}{2}}} \exp \left[-\frac{1}{2\sigma_i^2} \sum_{n=0}^{N-1} (y_i[n] - \cos 2\pi f_i n)^2 \right] \\ &= \left[\prod_{i=1}^p \frac{1}{(2\pi\sigma_i^2)^{\frac{N}{2}}} \right] \exp \left[-\sum_{i=1}^p \frac{1}{2\sigma_i^2} \sum_{n=0}^{N-1} (y_i[n] - \cos 2\pi f_i n)^2 \right], \end{aligned} \quad (7.8)$$

por lo que el problema quedó desacoplado y para maximizar la función de verosimilitud hay que minimizar

$$J(f_i) = \sum_{n=0}^{N-1} (y_i[n] - \cos 2\pi f_i n)^2, \quad i = 1, \dots, p.$$

El problema original de la minimización de una función de dimensión p se redujo a la minimización separada de p funciones de una dimensión. El nuevo conjunto de datos $\{y_1[n], \dots, y_p[n]\}$ se denomina *datos completos* y se relaciona con el conjunto de datos original como

$$x[n] = \sum_{i=1}^p y_i[n] \quad (7.9)$$

si el ruido se descompone como

$$w[n] = \sum_{i=1}^p w_i[n],$$

para lo cual se asume que los procesos de ruido $w_i[n]$ son independientes y

$$\sigma^2 = \sum_{i=1}^p \sigma_i^2[n].$$

El problema es como obtener los datos completos a partir de los datos originales, o los datos *incompletos*, teniendo que la descomposición no es única.

7.3.2. El algoritmo EM

Supóngase que se tiene la transformación de los datos incompletos a los datos completos

$$\mathbf{x} = \mathbf{g}(\mathbf{y}_1, \dots, \mathbf{y}_M) = \mathbf{g}(\mathbf{y})$$

donde en el ejemplo indicado antes $M = p$ y los elementos de \mathbf{y}_i están dados por la ecuación 7.7. El objetivo es encontrar el MLE de $\boldsymbol{\theta}$ mediante la maximización de $p_x(\mathbf{x}; \boldsymbol{\theta})$. Debido a que este problema es difícil por tratarse de la maximización de una función de muchas variables, en su lugar se maximiza $\ln p_y(\mathbf{y}; \boldsymbol{\theta})$. Como no se dispone de \mathbf{y} , se reemplaza la función de verosimilitud logarítmica por su esperanza condicional

$$E_{y|x}[\ln p_y(\mathbf{y}; \boldsymbol{\theta})] = \int \ln p_y(\mathbf{y}; \boldsymbol{\theta}) p(\mathbf{y}|\mathbf{x}; \boldsymbol{\theta}) d\mathbf{y}.$$

Finalmente, como se necesita conocer $\boldsymbol{\theta}$ para determinar $p(\mathbf{y}|\mathbf{x}; \boldsymbol{\theta})$ y poder establecer la esperanza condicional, se emplea la estimación actual. Si $\boldsymbol{\theta}_k$ es la estimación k -ésima del MLE de $\boldsymbol{\theta}$, se emplea el siguiente algoritmo iterativo:

Esperanza (E): Determinar la función de verosimilitud logarítmica promedio de los datos completos

$$U(\boldsymbol{\theta}, \boldsymbol{\theta}_k) = \int \ln p_y(\mathbf{y}; \boldsymbol{\theta}) p(\mathbf{y}|\mathbf{x}; \boldsymbol{\theta}_k) d\mathbf{y} \quad (7.10)$$

Maximización (M): Maximizar la función de verosimilitud logarítmica promedio de los datos completos

$$\boldsymbol{\theta}_{k+1} = \arg \max_{\boldsymbol{\theta}} U(\boldsymbol{\theta}, \boldsymbol{\theta}_k). \quad (7.11)$$

En la convergencia, podría obtenerse el estimador MLE. Este enfoque se denomina *algoritmo EM*. A continuación, se deduce el algoritmo para el problema de la estimación de las frecuencias de una mezcla de sinusoides empleado en la introducción.

7.3.3. Ejemplo: estimación de las frecuencias de mezcla de sinusoides

Continuando con el ejemplo de la sección 7.3.1, el objetivo es la deducción de la esperanza condicional de la función de verosimilitud logarítmica para aplicar el algoritmo EM. Se dedujo que la PDF de los datos completos está dada por la ecuación 7.8. Tomando el logaritmo se obtiene la función de verosimilitud logarítmica,

$$\begin{aligned} \ln p_y(\mathbf{y}; \boldsymbol{\theta}) &= \ln \left[\prod_{i=1}^p \frac{1}{(2\pi\sigma_i^2)^{\frac{N}{2}}} \right] - \sum_{i=1}^p \frac{1}{2\sigma_i^2} \sum_{n=0}^{N-1} (y_i[n] - \cos 2\pi f_i n)^2 \\ &= -\frac{N}{2} \sum_{i=1}^p \ln 2\pi\sigma_i^2 - \sum_{i=1}^p \frac{1}{2\sigma_i^2} \sum_{n=0}^{N-1} (y_i^2[n] - 2y_i[n] \cos 2\pi f_i n + \cos^2 2\pi f_i n) \\ &= -\frac{N}{2} \sum_{i=1}^p \ln 2\pi\sigma_i^2 - \sum_{i=1}^p \frac{1}{2\sigma_i^2} \sum_{n=0}^{N-1} y_i^2[n] + \sum_{i=1}^p \frac{1}{\sigma_i^2} \sum_{n=0}^{N-1} \left(y_i[n] \cos 2\pi f_i n - \frac{1}{2} \cos^2 2\pi f_i n \right). \end{aligned}$$

Considerando la aproximación

$$\sum_{n=0}^{N-1} \cos^2 2\pi f_i n = \sum_{n=0}^{N-1} \left(\frac{1}{2} + \frac{1}{2} \cos 4\pi f_i n \right) \approx \frac{N}{2},$$

donde se usó la ecuación B.1 demostrada en el apéndice B, se obtiene que

$$\begin{aligned} \ln p_y(\mathbf{y}; \boldsymbol{\theta}) &= -\frac{N}{2} \sum_{i=1}^p \ln 2\pi\sigma_i^2 - \sum_{i=1}^p \frac{1}{2\sigma_i^2} \sum_{n=0}^{N-1} y_i^2[n] - \frac{N}{4} \sum_{i=1}^p \frac{1}{\sigma_i^2} + \sum_{i=1}^p \frac{1}{\sigma_i^2} \sum_{n=0}^{N-1} y_i[n] \cos 2\pi f_i n \\ &= h(\mathbf{y}) + \sum_{i=1}^p \frac{1}{\sigma_i^2} \sum_{n=0}^{N-1} y_i[n] \cos 2\pi f_i n, \end{aligned}$$

donde $h(\mathbf{y})$ es la suma de los tres primeros sumandos y como no depende del parámetro, no afecta en la maximización de la función. Definiendo $\mathbf{c}_i = [1 \cos 2\pi f_i \dots \cos 2\pi f_i (N-1)]^T$, se llega a que

$$\ln p_y(\mathbf{y}; \boldsymbol{\theta}) = h(\mathbf{y}) + \sum_{i=1}^p \frac{1}{\sigma_i^2} \mathbf{c}_i^T \mathbf{y}_i.$$

Concatenando los \mathbf{c}_i y los \mathbf{y}_i como

$$\mathbf{c} = \begin{bmatrix} \frac{1}{\sigma_1^2} \mathbf{c}_1 \\ \vdots \\ \frac{1}{\sigma_p^2} \mathbf{c}_p \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_p \end{bmatrix}$$

se tiene que

$$\ln p_y(\mathbf{y}; \boldsymbol{\theta}) = h(\mathbf{y}) + \mathbf{c}^T \mathbf{y}.$$

Notar que los vectores \mathbf{c} y \mathbf{y} tienen dimensiones $Np \times 1$. De la ecuación 7.10 del paso E, la esperanza condicional es

$$\begin{aligned} U(\boldsymbol{\theta}, \boldsymbol{\theta}_k) &= E[\ln p_y(\mathbf{y}; \boldsymbol{\theta}) | \mathbf{x}; \boldsymbol{\theta}_k] \\ &= E[h(\mathbf{y}) | \mathbf{x}; \boldsymbol{\theta}_k] + \mathbf{c}^T E[\mathbf{y} | \mathbf{x}; \boldsymbol{\theta}_k]. \end{aligned} \quad (7.12)$$

Como se quiere maximizar $U(\boldsymbol{\theta}, \boldsymbol{\theta}_k)$ con respecto a $\boldsymbol{\theta}$, se puede omitir el valor esperado de $h(\mathbf{y})$ ya que no depende de $\boldsymbol{\theta}$. Ahora, los vectores \mathbf{y} y \mathbf{x} son conjuntamente gaussianos, ya que, del la ecuación 7.9

$$\mathbf{x} = \sum_{i=1}^p \mathbf{y}_i = [\mathbf{I} \mathbf{I} \dots \mathbf{I}] \mathbf{y},$$

donde \mathbf{I} es la matriz identidad $N \times N$ y la matriz de transformación está compuesta de p matrices identidad, por lo que tiene dimensiones $N \times Np$. Empleando el resultado de la esperanza condicional de vectores conjuntamente gaussianos deducido en el apéndice F dado por la ecuación 10.14, se tiene que

$$E(\mathbf{y}|\mathbf{x}; \boldsymbol{\theta}_k) = E(\mathbf{y}) + \mathbf{C}_{yx} \mathbf{C}_{xx}^{-1} (\mathbf{x} - E(\mathbf{x})).$$

Teniendo en cuenta que $\mathbf{y}_i = \mathbf{c}_i + \mathbf{w}_i$, las medias son

$$E(\mathbf{y}) = \begin{bmatrix} \mathbf{c}_1 \\ \vdots \\ \mathbf{c}_p \end{bmatrix} \quad E(\mathbf{x}) = \sum_{i=1}^p \mathbf{c}_i$$

y las matrices de covarianza son

$$\mathbf{C}_{xx} = \sigma^2 \mathbf{I}$$

de dimensiones $N \times N$ y

$$\begin{aligned} C_{yx} &= E \left(\begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \\ \vdots \\ \mathbf{w}_p \end{bmatrix} \mathbf{w}^T \right) \\ &= E \left(\begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \\ \vdots \\ \mathbf{w}_p \end{bmatrix} \left\{ \begin{bmatrix} \mathbf{I} & \mathbf{I} & \dots & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \\ \vdots \\ \mathbf{w}_p \end{bmatrix} \right\}^T \right) \\ &= \begin{bmatrix} \sigma_1^2 \mathbf{I} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \sigma_2^2 \mathbf{I} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \sigma_p^2 \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{I} \\ \mathbf{I} \\ \vdots \\ \mathbf{I} \end{bmatrix} \\ &= \begin{bmatrix} \sigma_1^2 \mathbf{I} \\ \sigma_2^2 \mathbf{I} \\ \vdots \\ \sigma_p^2 \mathbf{I} \end{bmatrix} \end{aligned}$$

Sustituyendo estos resultados en la esperanza condicional, se tiene que

$$\begin{aligned} E(\mathbf{y}|\mathbf{x}; \boldsymbol{\theta}_k) &= \begin{bmatrix} \mathbf{c}_1 \\ \vdots \\ \mathbf{c}_p \end{bmatrix} + \begin{bmatrix} \sigma_1^2 \mathbf{I} \\ \vdots \\ \sigma_p^2 \mathbf{I} \end{bmatrix} \frac{1}{\sigma^2} \mathbf{I} \left(\mathbf{x} - \sum_{i=1}^p \mathbf{c}_i \right) \\ &= \begin{bmatrix} \mathbf{c}_1 \\ \vdots \\ \mathbf{c}_p \end{bmatrix} + \begin{bmatrix} \frac{\sigma_1^2}{\sigma^2} \left(\mathbf{x} - \sum_{i=1}^p \mathbf{c}_i \right) \\ \vdots \\ \frac{\sigma_p^2}{\sigma^2} \left(\mathbf{x} - \sum_{i=1}^p \mathbf{c}_i \right) \end{bmatrix}, \end{aligned}$$

donde se observa que el problema quedó desacoplado en p problemas separados,

$$E(\mathbf{y}_i|\mathbf{x}; \boldsymbol{\theta}_k) = \mathbf{c}_i + \frac{\sigma_i^2}{\sigma^2} \left(\mathbf{x} - \sum_{i=1}^p \mathbf{c}_i \right), \quad i = 1, \dots, p$$

donde \mathbf{c}_i se calcular empleando $\boldsymbol{\theta}_k$. Notar que $E(\mathbf{y}_i|\mathbf{x}; \boldsymbol{\theta}_k)$ puede ser pensado como un estimador de los datos $y_i[n]$, ya que, llamando $\hat{\mathbf{y}}_i = E(\mathbf{y}_i|\mathbf{x}; \boldsymbol{\theta}_k)$,

$$\hat{y}_i[n] = \cos 2\pi f_{i_k} n + \frac{\sigma_i^2}{\sigma^2} \left(x[n] - \sum_{i=1}^p \cos 2\pi f_{i_k} n \right)$$

donde el término entre paréntesis es un estimador de $w[n]$ y por lo tanto, el segundo sumando es un estimador de $w_i[n]$ en el paso k -ésimo. Descartando el término independiente de $\boldsymbol{\theta}$ en la ecuación 7.12, se quiere maximizar sobre $\boldsymbol{\theta}$

$$U'(\boldsymbol{\theta}, \boldsymbol{\theta}_k) = \mathbf{c}^T \mathbf{y} = \sum_{i=1}^p \mathbf{c}_i^T \hat{\mathbf{y}}_i,$$

lo cual se hace maximizando cada termino de la suma separadamente,

$$f_{i_{k+1}} = \arg \max_{f_i} \mathbf{c}_i^T \hat{\mathbf{y}}_i.$$

Por último, como los σ_i^2 no son únicos, pueden elegirse arbitrariamente siempre que se cumpla que

$$\sum_{i=1}^p \sigma_i^2 = \sigma^2,$$

o equivalentemente

$$\sum_{i=1}^p \beta_i = \sum_{i=1}^p \frac{\sigma_i^2}{\sigma^2} = 1.$$

Resumiendo, el algoritmo EM para la estimación de las frecuencias es

Paso E: Para $i = 1, \dots, p$

$$\hat{y}_i[n] = \cos 2\pi f_{i_k} n + \beta_i \left(x[n] - \sum_{i=1}^p \cos 2\pi f_{i_k} n \right).$$

Paso M: Para $i = 1, \dots, p$

$$f_{i_{k+1}} = \arg \max_{f_i} \sum_{n=0}^{N-1} \hat{y}_i[n] \cos 2\pi f_i n,$$

donde los β_i pueden ser elegidos de forma arbitraria siempre que $\sum_{i=1}^p \beta_i = 1$.

El algoritmo desacopla iterativamente los datos originales en p conjunto de datos separados, donde cada uno consiste en una única senoide en WGN. La maximización obtenida corresponde al MLE de una única senoide en ruido con el conjunto de datos dado por los datos completos estimados, como se muestra en el problema de la sección 7.6.19. Mas detalles pueden encontrarse en [12].

7.4. MLE asintótico

En ocasiones es difícil evaluar el MLE de un parámetro con PDF gaussiana debido a la necesidad de invertir una matriz de covarianza de dimensiones grandes. Por ejemplo, si $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}(\boldsymbol{\theta}))$, el MLE se obtiene maximizando

$$p(\mathbf{x}, \boldsymbol{\theta}) = \frac{1}{(2\pi)^{\frac{N}{2}} \det^{\frac{N}{2}}(\mathbf{C})} \exp \left[-\frac{1}{2} \mathbf{x}^T \mathbf{C}^{-1} \mathbf{x} \right].$$

Si no es posible obtener una fórmula cerrada de la matriz inversa, la búsqueda del estimador requiere la inversión numérica de una matriz $N \times N$ en cada iteración. En el caso en que los datos \mathbf{x} son un proceso WSS de media nula de forma que la matriz de covarianza es Toeplitz, puede emplearse un método alternativo aproximado. Como se mostró en la sección 3.5, la función de verosimilitud logarítmica asintótica, es decir, cuando la cantidad de datos es grande, es (ecuación 3.50)

$$\ln p(\mathbf{x}; \boldsymbol{\theta}) = -\frac{N}{2} \ln 2\pi - \frac{N}{2} \int_{-\frac{1}{2}}^{\frac{1}{2}} \left[\ln P_{xx}(f) + \frac{I(f)}{P_{xx}(f)} \right] df. \quad (7.13)$$

donde

$$I(f) = \frac{1}{N} \left| \sum_{n=0}^{N-1} x[n] e^{-j2\pi f n} \right|^2$$

es el periodograma de los datos y $P_{xx}(f)$ es la PSD. La dependencia de la función de verosimilitud logarítmica con el parámetro $\boldsymbol{\theta}$ es a través de la PSD. Diferenciando 7.13 se obtiene que ((ecuación 3.51))

$$\frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \theta_i} = -\frac{N}{2} \int_{-\frac{1}{2}}^{\frac{1}{2}} \left[\frac{1}{P_{xx}(f)} - \frac{I(f)}{P_{xx}^2(f)} \right] \frac{\partial P_{xx}(f)}{\partial \theta_i} df,$$

por lo que las condiciones necesarias del MLE son

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \left[\frac{1}{P_{xx}(f)} - \frac{I(f)}{P_{xx}^2(f)} \right] \frac{\partial P_{xx}(f)}{\partial \theta_i} df = 0. \quad (7.14)$$

La derivada segunda está dada por la ecuación 3.52 por lo que podría emplearse algún método iterativo como el de Newton-Raphson para encontrar el máximo.

7.4.1. Ejemplo: proceso de media móvil gaussiano

Un proceso WSS tiene ACF

$$r_{xx}[k] = \begin{cases} 1 + b^2[1] + b^2[2] & k = 0 \\ b[1] + b[1]b[2] & k = 1 \\ b[2] & k = 2 \\ 0 & k \geq 3. \end{cases} \quad (7.15)$$

y se quiere calcular el MLE de los coeficientes $b[1]$ y $b[2]$. Para hacerlo, se comenzará mostrando que la PSD es

$$P_{xx}(f) = |1 + b[1]e^{-j2\pi f} + b[2]e^{-j4\pi f}|^2.$$

La transformada z de la ACF es

$$P_{xx}(z) = b[2]z^2 + (b[1] + b[1]b[2])z + (1 + b^2[1] + b^2[2]) + (b[1] + b[1]b[2])z^{-1} + b[2]z^{-2}. \quad (7.16)$$

Además, el proceso $x[n]$ se obtiene filtrando WGN de varianza $\sigma_{ww}^2 = 1$ mediante un filtro de función de transferencia $\mathcal{B}(z) = a_0 + a_1z^{-1} + a_2z^{-2}$, por lo que la PSD se puede expresar como (ver el capítulo 3 de [5])

$$\begin{aligned} P_{xx}(z) &= P_{ww}(z)\mathcal{B}(z)\mathcal{B}(1/z) \\ &= (a_0 + a_1z^{-1} + a_2z^{-2})(a_0 + a_1z^1 + a_2z^2) \\ &= a_0a_2z^2 + (a_0a_1 + a_1a_2)z + (a_0^2 + a_1^2 + a_2^2) + (a_0a_1 + a_1a_2)z^{-1} + a_0a_2z^{-2}. \end{aligned}$$

Igualando esta ecuación con la ecuación 7.16, se obtiene que los coeficientes del filtro generador son

$$a_0 = 1, \quad a_1 = b[1], \quad a_2 = b[2],$$

resultando en

$$\mathcal{B}(z) = 1 + b[1]z^{-1} + b[2]z^{-2}.$$

Se concluye que la PSD de los datos es

$$P_{xx}(f) = P_{ww}(f)|\mathcal{B}(f)|^2 = |1 + b[1]e^{-j2\pi f} + b[2]e^{-j4\pi f}|^2,$$

y se trata de un proceso MA de orden 2. Para calcular el MLE de los parámetros de filtro, es necesario invertir la matriz de covarianza. En su lugar, se puede maximizar la ecuación 7.13. Típicamente, se asume que el filtro $\mathcal{B}(z)$ es de fase mínima, es decir, todos los ceros se encuentran dentro del círculo unidad. En ese caso, como se muestra en el problema de la sección 7.6.22, se cumple que

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \ln P_{xx}(f) df = 0,$$

por lo que maximizar la ecuación 7.13 es equivalente a minimizar

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{I(f)}{|1 + b[1]e^{-j2\pi f} + b[2]e^{-j4\pi f}|^2} df.$$

Además, la función de transferencia se puede factorizar como

$$\mathcal{B}(z) = (1 - z_1 z^{-1})(1 - z_2 z^{-1})$$

donde z_1 y z_2 son los ceros del sistema y cumplen que son reales o complejos conjugados con $|z_1| < 1$ y $|z_2| < 1$. Por lo tanto, el MLE se encuentra minimizando

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{I(f)}{|1 - z_1 e^{-j2\pi f}|^2 |1 - z_2 e^{-j2\pi f}|^2} df \quad (7.17)$$

sobre los valores permitidos de z_1 y z_2 , y luego, considerando que

$$(1 - z_1 e^{-j2\pi f})(1 - z_2 e^{-j2\pi f}) = 1 - (z_1 + z_2)e^{-j2\pi f} + z_1 z_2 e^{-j4\pi f},$$

los coeficientes se obtienen como

$$b[1] = -(z_1 + z_2), \quad b[2] = z_1 z_2.$$

Para la minimización puede emplearse la búsqueda sobre una grilla en el plano complejo o algún método iterativo. La restricción de que z_1 y z_2 son reales o complejos conjugados limita la búsqueda, y es necesaria para que los coeficientes del filtro generador, y por lo tanto el proceso $x[n]$, sean reales.

Se realizó una simulación por computadora del cálculo del MLE asintótico de los parámetros $b[1]$ y $b[2]$ de un proceso MA(2), cuya ACF está dada por la ecuación 7.15, mediante la búsqueda del mínimo de la ecuación 7.17. Para esto se generaron $N = 150$ muestras del proceso MA(2) con parámetros $b[1] = -0.9$ y $b[2] = 0.81$, que corresponde a un filtro generador $\mathcal{B}(z)$ con ceros en $z_{1,2} = 0.9e^{\pm j\pi/6}$, se construyó una grilla en el semicírculo unidad del plano complejo y se evaluó la ecuación 7.17 en cada punto de la grilla. El estimador MLE de z_1 es el valor de la grilla en donde la función toma el menor valor, y $z_2 = z_1^*$. El resultado se muestra en la figura 7.3. Para la realización mostrada en la figura, el mínimo de la superficie está en \hat{z}_1 de módulo $|\hat{z}_1| \approx 0.89$ y argumento $\arg(\hat{z}_1) \approx 1.03$ rad, lo que corresponde a los estimadores

$$\hat{b}[1] = -2|\hat{z}_1| \cos[\arg(\hat{z}_1)] \approx -0.92 \quad \text{y} \quad \hat{b}[2] = |\hat{z}_1|^2 \approx 0.79.$$

El resultado mejora al incrementar la cantidad N de muestras del proceso. La ecuación 7.17 debe evaluarse para cada punto de la grilla y por lo tanto, puede ser computacionalmente costoso. Para hacerlo de forma eficiente, notar que $I(f)$ solo depende de los datos $x[n]$ y no de los valores z_1 de la grilla, por lo que se calcula una sola vez mediante la transformada discreta de Fourier (*discrete Fourier transform*, DFT) con relleno de ceros para obtener una cantidad de muestras apropiada en la frecuencia y lograr un cálculo preciso de la integral de forma numérica, empleando por ejemplo, la regla del trapecio. El denominador en la integral es distinto para cada complejo z_1 en la grilla, y para calcularlo de forma eficiente, se aplica la DTF a la secuencia $[1, -2|z_1| \cos[\arg(z_1)], |z_1|^2]$ con el relleno de ceros apropiado para tener la misma cantidad de muestras en frecuencia que el numerador. Se tuvo en cuenta además que el integrando es simétrico, por lo que alcanza con integrar en el intervalo entre 0 y 1/2.

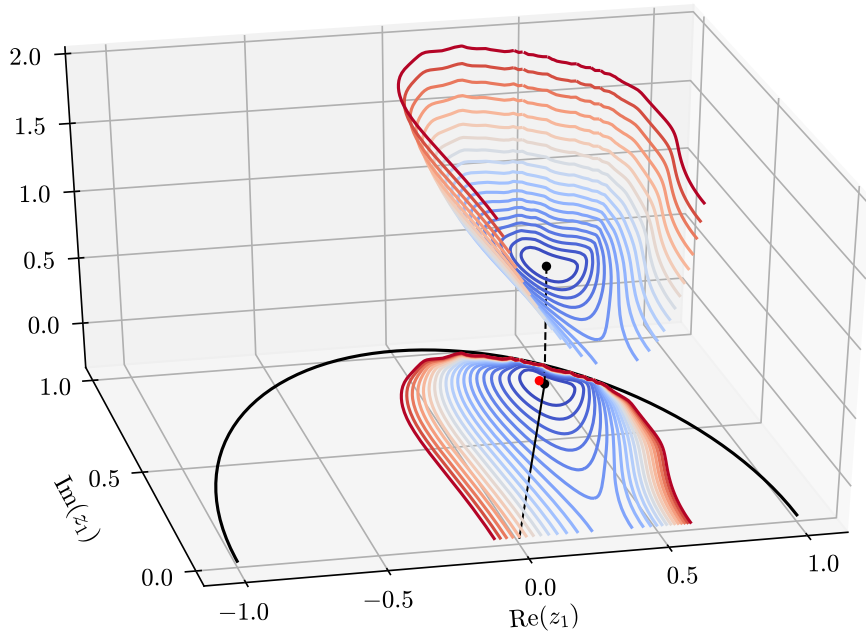


Figura 7.3: MLE asintótico de los parámetros de un proceso MA(2). Se muestra la superficie definida por la ecuación 7.17 en el plano complejo mediante curvas de nivel. El mínimo de la superficie corresponde al MLE del cero z_1 del proceso MA(2), y se marca con un punto negro. El valor verdadero del parámetro está marcado con el punto en rojo.

7.5. Ejemplos de procesamiento de señales

7.5.1. Ejemplo: estimación de los parámetros de una senoide

En la sección 3.4.4 se calculó la CRLB de los parámetros de una senoide en WGN. Se determinará a continuación el MLE de la amplitud A , la frecuencia f_0 y la fase ϕ . Los datos observados son

$$x[n] = A \cos(2\pi f_0 n + \phi) + w[n] \quad n = 0, 1, \dots, N-1,$$

con $A > 0$ y $0 < f_0 < 1/2$. La PDF de los datos es

$$p(\mathbf{x}; \boldsymbol{\theta}) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - A \cos(2\pi f_0 n + \phi))^2 \right],$$

por lo que el MLE de la amplitud A , la frecuencia f_0 y la fase ϕ se obtienen minimizando

$$J(A, f_0, \phi) = \sum_{n=0}^{N-1} (x[n] - A \cos(2\pi f_0 n + \phi))^2.$$

Considerando que $\cos(\alpha \pm \beta) = \cos \alpha \cos \beta \mp \sin \alpha \sin \beta$ se tiene que

$$J(A, f_0, \phi) = \sum_{n=0}^{N-1} (x[n] - A \cos \phi \cos 2\pi f_0 n + A \sin \phi \sin 2\pi f_0 n)^2.$$

Aunque J no es una forma cuadrática en A y ϕ , se puede transformar en una forma cuadrática definiendo

$$\alpha_1 = A \cos \phi, \quad \alpha_2 = A \sin \phi,$$

que se trata de una transformación uno a uno. La transformación inversa es

$$A = \sqrt{\alpha_1^2 + \alpha_2^2}, \quad \phi = \arctan\left(\frac{-\alpha_2}{\alpha_1}\right). \quad (7.18)$$

Sean

$$\mathbf{c} = [1 \cos 2\pi f_0 \dots \cos 2\pi f_0(N-1)]^T \quad \text{y} \quad \mathbf{s} = [0 \sin 2\pi f_0 \dots \sin 2\pi f_0(N-1)]^T.$$

La función a minimizar queda

$$\begin{aligned} J'(\alpha_1, \alpha_2, f_0) &= (\mathbf{x} - \alpha_1 \mathbf{c} - \alpha_2 \mathbf{s})(\mathbf{x} - \alpha_1 \mathbf{c} - \alpha_2 \mathbf{s})^T \\ &= (\mathbf{x} - \mathbf{H}\boldsymbol{\alpha})^T(\mathbf{x} - \mathbf{H}\boldsymbol{\alpha}), \end{aligned}$$

donde $\boldsymbol{\alpha} = [\alpha_1 \alpha_2]^T$ y $\mathbf{H} = [\mathbf{c} \mathbf{s}]$ es una matriz $N \times 2$. Se observa que la función a minimizar sobre $\boldsymbol{\alpha}$ es la misma que la encontrada en el modelo lineal dada por la ecuación 7.5 con $\mathbf{C} = \mathbf{I}$. Por lo tanto, de la ecuación 7.4, J' se minimiza con

$$\hat{\boldsymbol{\alpha}} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x}. \quad (7.19)$$

Luego,

$$\begin{aligned} J'(\hat{\alpha}_1, \hat{\alpha}_2, f_0) &= (\mathbf{x} - \mathbf{H}\hat{\boldsymbol{\alpha}})^T(\mathbf{x} - \mathbf{H}\hat{\boldsymbol{\alpha}}) \\ &= (\mathbf{x} - \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x})^T(\mathbf{x} - \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x}) \\ &= \mathbf{x}^T (\mathbf{I} - \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T)^T (\mathbf{I} - \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T) \mathbf{x} \\ &\stackrel{(a)}{=} \mathbf{x}^T (\mathbf{I} - \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T) \mathbf{x} \end{aligned}$$

donde en (a) se consideró que $\mathbf{I} - \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T$ es una matriz idempotente (una matriz idempotente cumple que $\mathbf{A}^2 = \mathbf{A}$), ya que

$$\begin{aligned} (\mathbf{I} - \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T)^T (\mathbf{I} - \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T) &= (\mathbf{I} - \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T)(\mathbf{I} - \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T) \\ &= \mathbf{I} - \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T - \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \\ &\quad + \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \\ &= \mathbf{I} - \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T - \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \\ &\quad + \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \\ &= \mathbf{I} - \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T. \end{aligned}$$

De esta forma, para encontrar \hat{f}_0 se necesita minimizar J' sobre f_0 o equivalentemente, maximizar

$$\mathbf{x}^T \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x}.$$

Como $\mathbf{H} = [\mathbf{c} \mathbf{s}]$, se tiene que

$$\mathbf{H}^T \mathbf{x} = \begin{bmatrix} \mathbf{c}^T \\ \mathbf{s}^T \end{bmatrix} \mathbf{x} = \begin{bmatrix} \mathbf{c}^T \mathbf{x} \\ \mathbf{s}^T \mathbf{x} \end{bmatrix}, \quad \mathbf{H}^T \mathbf{H} = \begin{bmatrix} \mathbf{c}^T \\ \mathbf{s}^T \end{bmatrix} \begin{bmatrix} \mathbf{c} & \mathbf{s} \end{bmatrix} = \begin{bmatrix} \mathbf{c}^T \mathbf{c} & \mathbf{c}^T \mathbf{s} \\ \mathbf{s}^T \mathbf{c} & \mathbf{s}^T \mathbf{s} \end{bmatrix}$$

y la función a maximizar es

$$\begin{bmatrix} \mathbf{c}^T \mathbf{x} \\ \mathbf{s}^T \mathbf{x} \end{bmatrix}^T \begin{bmatrix} \mathbf{c}^T \mathbf{c} & \mathbf{c}^T \mathbf{s} \\ \mathbf{s}^T \mathbf{c} & \mathbf{s}^T \mathbf{s} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{c}^T \mathbf{x} \\ \mathbf{s}^T \mathbf{x} \end{bmatrix}. \quad (7.20)$$

Una vez que se encuentra \hat{f}_0 empleando la expresión anterior, $\hat{\boldsymbol{\alpha}}$ se encuentra empleando la ecuación 7.19 y luego, \hat{A} y $\hat{\phi}$ mediante la ecuación 7.18. Puede obtenerse una aproximación del MLE considerando que

$$\begin{aligned} \mathbf{c}^T \mathbf{c} &= \sum_{n=0}^{N-1} \cos^2 2\pi f_0 n = \sum_{n=0}^{N-1} \frac{1 + \cos 4\pi f_0 n}{2} = \frac{N}{2} \left(1 + \frac{1}{N} \sum_{n=0}^{N-1} \cos 4\pi f_0 n \right) \approx \frac{N}{2} \\ \mathbf{s}^T \mathbf{s} &= \sum_{n=0}^{N-1} \sin^2 2\pi f_0 n = \sum_{n=0}^{N-1} \frac{1 - \cos 4\pi f_0 n}{2} = \frac{N}{2} \left(1 - \frac{1}{N} \sum_{n=0}^{N-1} \cos 4\pi f_0 n \right) \approx \frac{N}{2} \\ \mathbf{c}^T \mathbf{s} &= \mathbf{s}^T \mathbf{c} = \sum_{n=0}^{N-1} \sin 2\pi f_0 n \cos 2\pi f_0 n = \frac{1}{2} \sum_{n=0}^{N-1} \sin 4\pi f_0 n, \end{aligned}$$

donde en las aproximaciones se empleó el resultado demostrado en el apéndice B, que es válido si f_0 no es cercano a 0 ni a $1/2$. Por lo tanto,

$$\begin{aligned} \begin{bmatrix} \mathbf{c}^T \mathbf{c} & \mathbf{c}^T \mathbf{s} \\ \mathbf{s}^T \mathbf{c} & \mathbf{s}^T \mathbf{s} \end{bmatrix}^{-1} &\approx \begin{bmatrix} \frac{N}{2} & \frac{1}{2} \sum_{n=0}^{N-1} \sin 4\pi f_0 n \\ \frac{1}{2} \sum_{n=0}^{N-1} \sin 4\pi f_0 n & \frac{N}{2} \end{bmatrix}^{-1} \\ &= \frac{1}{\frac{N^2}{4} - \frac{1}{4} \left(\sum_{n=0}^{N-1} \sin 4\pi f_0 n \right)^2} \begin{bmatrix} \frac{N}{2} & -\frac{1}{2} \sum_{n=0}^{N-1} \sin 4\pi f_0 n \\ -\frac{1}{2} \sum_{n=0}^{N-1} \sin 4\pi f_0 n & \frac{N}{2} \end{bmatrix}. \end{aligned}$$

Teniendo en cuenta que

$$\frac{N^2}{4} - \frac{1}{4} \left(\sum_{n=0}^{N-1} \sin 4\pi f_0 n \right)^2 = \frac{N^2}{4} \left[1 - \left(\frac{1}{N} \sum_{n=0}^{N-1} \sin 4\pi f_0 n \right)^2 \right] \approx \frac{N^2}{4},$$

se tiene que

$$\begin{bmatrix} \mathbf{c}^T \mathbf{c} & \mathbf{c}^T \mathbf{s} \\ \mathbf{s}^T \mathbf{c} & \mathbf{s}^T \mathbf{s} \end{bmatrix}^{-1} = \begin{bmatrix} \frac{2}{N} & -\frac{2}{N^2} \sum_{n=0}^{N-1} \sin 4\pi f_0 n \\ -\frac{2}{N^2} \sum_{n=0}^{N-1} \sin 4\pi f_0 n & \frac{2}{N} \end{bmatrix} \approx \begin{bmatrix} \frac{2}{N} & 0 \\ 0 & \frac{2}{N} \end{bmatrix} = \frac{2}{N} \mathbf{I}.$$

Con estas consideraciones, la ecuación 7.20 a maximizar para encontrar \hat{f}_0 queda

$$\begin{aligned} \begin{bmatrix} \mathbf{c}^T \mathbf{x} & \mathbf{s}^T \mathbf{x} \end{bmatrix} \frac{2}{N} \mathbf{I} \begin{bmatrix} \mathbf{c}^T \mathbf{x} \\ \mathbf{s}^T \mathbf{x} \end{bmatrix} &= \frac{2}{N} [(\mathbf{c}^T \mathbf{x})^2 + (\mathbf{s}^T \mathbf{x})^2] \\ &= \frac{2}{N} \left[\left(\sum_{n=0}^{N-1} x[n] \cos 2\pi f_0 n \right)^2 + \left(\sum_{n=0}^{N-1} x[n] \sin 2\pi f_0 n \right)^2 \right] \\ &= \frac{2}{N} \left[\text{Re}^2 \left(\sum_{n=0}^{N-1} x[n] e^{-2j\pi f_0 n} \right) + \text{Im}^2 \left(\sum_{n=0}^{N-1} x[n] e^{-2j\pi f_0 n} \right) \right] \\ &= \frac{2}{N} \left| \sum_{n=0}^{N-1} x[n] e^{-2j\pi f_0 n} \right|^2, \end{aligned}$$

es decir, el MLE de la frecuencia se obtiene maximizando el periodograma

$$I(f) = \frac{1}{N} \left| \sum_{n=0}^{N-1} x[n] e^{-2j\pi f n} \right|^2$$

en la frecuencia f . Luego, de la ecuación 7.19 se tiene que

$$\begin{aligned} \hat{\alpha} &\approx \frac{2}{N} \begin{bmatrix} \hat{\mathbf{c}}^T \mathbf{x} \\ \hat{\mathbf{s}}^T \mathbf{x} \end{bmatrix} \\ &= \begin{bmatrix} \frac{2}{N} \sum_{n=0}^{N-1} x[n] \cos 2\pi \hat{f}_0 n \\ \frac{2}{N} \sum_{n=0}^{N-1} x[n] \sin 2\pi \hat{f}_0 n \end{bmatrix} \end{aligned}$$

y sustituyendo en ecuación 7.18 de la transformación se obtiene que

$$\hat{A} = \sqrt{\hat{\alpha}_1^2 + \hat{\alpha}_2^2} = \frac{2}{N} \left| \sum_{n=0}^{N-1} x[n] e^{-2j\pi \hat{f}_0 n} \right| \quad (7.21)$$

$$\hat{\phi} = \arctan \frac{-\sum_{n=0}^{N-1} x[n] \cos 2\pi \hat{f}_0 n}{\sum_{n=0}^{N-1} x[n] \sin 2\pi \hat{f}_0 n}. \quad (7.22)$$

7.5.2. Ejemplo: estimación de los parámetros de un modelo autorregresivo

En el ejemplo de la sección 3.5.2 se calculó la CRLB de los parámetros de un modelo AR. En esta sección se buscará el MLE. Para hacerlo, se empleará la función de verosimilitud logarítmica dada por la ecuación 7.13. Teniendo en cuenta que la PSD de un modelo AR es (ver la ecuación 3.53)

$$P_{xx}(f) = \frac{\sigma_u^2}{\left| 1 + \sum_{k=1}^p a[k] e^{-j2\pi f k} \right|^2} = \frac{\sigma_u^2}{|A(f)|^2},$$

la función de verosimilitud logarítmica es

$$\ln p(\mathbf{x}; \mathbf{a}, \sigma_u^2) = -\frac{N}{2} \ln 2\pi - \frac{N}{2} \int_{-\frac{1}{2}}^{\frac{1}{2}} \left[\ln \frac{\sigma_u^2}{|A(f)|^2} + \frac{I(f)}{\frac{\sigma_u^2}{|A(f)|^2}} \right] df.$$

Para que $1/A(z)$ sea estable, se requiere que $A(z)$ sea de fase mínima, y por lo tanto, como se muestra en el problema de la sección 7.6.22, se cumple que

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \ln |A(f)|^2 df = 0.$$

De esta forma

$$\ln p(\mathbf{x}; \mathbf{a}, \sigma_u^2) = -\frac{N}{2} \ln 2\pi - \frac{N}{2} \ln \sigma_u^2 - \frac{N}{2\sigma_u^2} \int_{-\frac{1}{2}}^{\frac{1}{2}} |A(f)|^2 I(f) df. \quad (7.23)$$

Diferenciando respecto a σ_u^2 e igualando a cero, se obtiene que el estimador de σ_u^2 cumple que

$$-\frac{N}{2\sigma_u^2} + \frac{N}{2\sigma_u^4} \int_{-\frac{1}{2}}^{\frac{1}{2}} |A(f)|^2 I(f) df = 0,$$

es decir,

$$\hat{\sigma}_u^2 = \int_{-\frac{1}{2}}^{\frac{1}{2}} |A(f)|^2 I(f) df.$$

Sustituyendo este resultado en la ecuación 7.23, resulta en

$$\ln p(\mathbf{x}; \mathbf{a}, \hat{\sigma}_u^2) = -\frac{N}{2} \ln 2\pi - \frac{N}{2} \ln \hat{\sigma}_u^2 - \frac{N}{2}.$$

El estimador $\hat{\mathbf{a}}$ es el valor de \mathbf{a} que maximiza esta ecuación, o equivalentemente, que minimiza $\hat{\sigma}_u^2$,

$$J(\mathbf{a}) = \int_{-\frac{1}{2}}^{\frac{1}{2}} |A(f)|^2 I(f) df.$$

Esta función es cuadrática en \mathbf{a} , y por lo tanto, la diferenciación conduce al mínimo global. Diferenciando respecto a $a[k]$ para $k = 1, \dots, p$ se tiene que

$$\begin{aligned}
\frac{\partial J(\mathbf{a})}{\partial a[k]} &= \int_{-\frac{1}{2}}^{\frac{1}{2}} \left[A(f) \frac{\partial A^*(f)}{\partial a[k]} + \frac{\partial A(f)}{\partial a[k]} A^*(f) \right] I(f) df \\
&= \int_{-\frac{1}{2}}^{\frac{1}{2}} [A(f)e^{j2\pi fk} + A^*(f)e^{-j2\pi fk}] I(f) df \\
&= \int_{-\frac{1}{2}}^{\frac{1}{2}} [A(f)I(f)e^{j2\pi fk} + A^*(f)I(f)e^{-j2\pi fk}] df \\
&\stackrel{(a)}{=} \int_{-\frac{1}{2}}^{\frac{1}{2}} A(f)I(f)e^{j2\pi fk} df + \int_{-\frac{1}{2}}^{\frac{1}{2}} A(-f)I(-f)e^{-j2\pi fk} df \\
&\stackrel{(b)}{=} \int_{-\frac{1}{2}}^{\frac{1}{2}} A(f)I(f)e^{j2\pi fk} df - \int_{\frac{1}{2}}^{-\frac{1}{2}} A(f')I(f')e^{j2\pi f'k} df' \\
&= 2 \int_{-\frac{1}{2}}^{\frac{1}{2}} A(f)I(f)e^{j2\pi fk} df
\end{aligned}$$

donde en (a) se tuvo en cuenta que $A(-f) = A^*(f)$ y $I(-f) = I(f)$ y en (b) se realizó el cambio de variable $f' = -f$ en la segunda integral. Igualando a cero, se obtiene que

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \left(1 + \sum_{l=1}^p a[l]e^{-j2\pi fl} \right) I(f)e^{j2\pi fk} df = 0 \quad k = 1, \dots, p \quad (7.24)$$

o

$$\sum_{l=1}^p a[l] \int_{-\frac{1}{2}}^{\frac{1}{2}} I(f)e^{j2\pi f(k-l)} df = - \int_{-\frac{1}{2}}^{\frac{1}{2}} I(f)e^{j2\pi fk} df \quad k = 1, \dots, p.$$

Observando que $\int_{-\frac{1}{2}}^{\frac{1}{2}} I(f)e^{j2\pi fk} df$ es la transformada inversa de Fourier del periodograma evaluada en k , y como se hace en el problema 7.6.25 puede demostrarse que es la ACF estimada, dada por

$$\hat{r}_{xx}[k] = \begin{cases} \frac{1}{N} \sum_{n=0}^{N-1-|k|} x[n]x[n+|k|] & |k| \leq N-1 \\ 0 & |k| \geq N, \end{cases} \quad (7.25)$$

se obtiene que el sistema de ecuaciones a resolver para el MLE aproximado de los parámetros del filtro AR es

$$\sum_{l=1}^p a[l] \hat{r}_{xx}[k-l] = -\hat{r}_{xx}[k] \quad k = 1, \dots, p,$$

o en forma matricial

$$\begin{bmatrix} \hat{r}_{xx}[0] & \hat{r}_{xx}[1] & \dots & \hat{r}_{xx}[p-1] \\ \hat{r}_{xx}[1] & \hat{r}_{xx}[0] & \dots & \hat{r}_{xx}[p-2] \\ \vdots & \vdots & \ddots & \vdots \\ \hat{r}_{xx}[p-1] & \hat{r}_{xx}[p-2] & \dots & \hat{r}_{xx}[0] \end{bmatrix} \begin{bmatrix} \hat{a}[1] \\ \hat{a}[2] \\ \vdots \\ \hat{a}[p] \end{bmatrix} = - \begin{bmatrix} \hat{r}_{xx}[1] \\ \hat{r}_{xx}[2] \\ \vdots \\ \hat{r}_{xx}[p] \end{bmatrix}. \quad (7.26)$$

Este sistema de ecuaciones son las *ecuaciones de Yule-Walker* (ver el capítulo 3 de [5]) y se trata del *método de la autocorrelación de predicción lineal* [7]. Para completar la discusión, se determinará de

forma explícita el MLE de σ_u^2 . Como se obtuvo previamente,

$$\begin{aligned}
\hat{\sigma}_u^2 &= \int_{-\frac{1}{2}}^{\frac{1}{2}} |\hat{A}(f)|^2 I(f) df \\
&= \int_{-\frac{1}{2}}^{\frac{1}{2}} \hat{A}(f) I(f) \hat{A}^*(f) df \\
&= \int_{-\frac{1}{2}}^{\frac{1}{2}} \hat{A}(f) I(f) \left(\sum_{k=0}^p \hat{a}[k] e^{j2\pi f k} \right) df \\
&= \sum_{k=0}^p \hat{a}[k] \int_{-\frac{1}{2}}^{\frac{1}{2}} \hat{A}(f) I(f) e^{j2\pi f k} df.
\end{aligned}$$

Pero todos los términos de la sumatoria son nulos excepto para $k = 0$, ya que la ecuación 7.24 indica que

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \hat{A}(f) I(f) e^{j2\pi f k} df = 0 \quad k = 1, \dots, p.$$

Por lo tanto, teniendo en cuenta que $a[0] = 1$,

$$\begin{aligned}
\hat{\sigma}_u^2 &= \int_{-\frac{1}{2}}^{\frac{1}{2}} \hat{A}(f) I(f) df \\
&= \sum_{k=0}^p \hat{a}[k] \int_{-\frac{1}{2}}^{\frac{1}{2}} I(f) e^{-j2\pi f k} df \\
&\stackrel{(a)}{=} \sum_{k=0}^p \hat{a}[k] \hat{r}_{xx}[-k] \\
&\stackrel{(b)}{=} \sum_{k=0}^p \hat{a}[k] \hat{r}_{xx}[k]
\end{aligned}$$

donde en (a) se observó que la integral es la transformada inversa de Fourier de $I(f)$ evaluada en $-k$ y en (b) que $\hat{r}_{xx}[-k] = \hat{r}_{xx}[k]$. Finalmente, el MLE es

$$\hat{\sigma}_u^2 = \hat{r}_{xx}[0] + \sum_{k=1}^p \hat{a}[k] \hat{r}_{xx}[k]. \quad (7.27)$$

7.6. Problemas

7.6.1. Problema 1

Se observa el conjunto de datos

$$x[n] = A + w[n], \quad n = 0, \dots, N-1,$$

donde A es un nivel de DC desconocido y $w[n]$ es WGN de varianza A desconocida. Este problema difiere del habitual en el hecho de que el parámetro desconocido A se refleja en la media y en la varianza. Mostrar que la CRLB es

$$\text{var}(\hat{A}) \geq \frac{A^2}{N(A + \frac{1}{2})}.$$

Solución La CRLB está dada por las ecuaciones 3.1 y 3.2. La PDF de los datos es

$$p(\mathbf{x}; A) = \frac{1}{(2\pi A)^{\frac{N}{2}}} \exp \left[-\frac{1}{2A} \sum_{n=0}^{N-1} (x[n] - A)^2 \right]$$

y al tomar el logaritmo queda

$$\ln p(\mathbf{x}; A) = -\frac{N}{2} \ln 2\pi - \frac{N}{2} \ln A - \frac{1}{2A} \sum_{n=0}^{N-1} (x[n] - A)^2.$$

La derivada primera respecto al parámetro es

$$\frac{\partial \ln p(\mathbf{x}; A)}{\partial A} = -\frac{N}{2A} + \frac{1}{A} \sum_{n=0}^{N-1} (x[n] - A) + \frac{1}{2A^2} \sum_{n=0}^{N-1} (x[n] - A)^2$$

y derivando nuevamente, se obtiene que

$$\frac{\partial \ln p(\mathbf{x}; A)}{\partial A} = \frac{N}{2A^2} - \frac{N}{A} - \frac{1}{A^2} \sum_{n=0}^{N-1} (x[n] - A) - \frac{1}{A^2} \sum_{n=0}^{N-1} (x[n] - A) - \frac{1}{A^3} \sum_{n=0}^{N-1} (x[n] - A)^2.$$

Tomando esperanza, se tiene que

$$\begin{aligned} E \left[\frac{\partial \ln p(\mathbf{x}; A)}{\partial A} \right] &= \frac{N}{2A^2} - \frac{N}{A} - \frac{1}{A^2} \sum_{n=0}^{N-1} [E(x[n]) - A] - \frac{1}{A^2} \sum_{n=0}^{N-1} [E(x[n]) - A] - \frac{1}{A^3} \sum_{n=0}^{N-1} E[(x[n] - A)^2] \\ &\stackrel{(a)}{=} \frac{N}{2A^2} - \frac{N}{A} - \frac{1}{A^3} NA \\ &= -\frac{N}{2A^2} - \frac{N}{A}, \end{aligned}$$

donde en (a) se consideró que $E(x[n]) = A$ y $E[(x[n] - A)^2] = A$. Se concluye que la información de Fisher es

$$I(A) = -E \left[\frac{\partial \ln p(\mathbf{x}; A)}{\partial A} \right] = \frac{N}{2A^2} + \frac{N}{A}$$

por lo que la CRLB es

$$\text{CRLB}(A) = \frac{1}{\frac{N}{2A^2} + \frac{N}{A}} = \frac{1}{\frac{N}{A^2} \left(\frac{1}{2} + A \right)},$$

es decir,

$$\text{var}(\hat{A}) \geq \frac{A^2}{N(A + \frac{1}{2})}.$$

7.6.2. Problema 2

En el problema anterior, considérese la media muestral como estimador. Calcular la varianza y compararla con la CRLB. ¿La media muestral alcanza la CRLB para N finito? ¿Qué ocurre cuando $N \rightarrow \infty$? ¿Es mejor la media muestral o el MLE como estimador?

Solución Se considera la media muestral

$$\bar{x} = \frac{1}{N} \sum_{n=0}^{N-1} x[n]$$

como estimador de A . Claramente, es un estimador insesgado, y su varianza es

$$\text{var}(\bar{x}) = \frac{1}{N^2} \sum_{n=0}^{N-1} \text{var}(x[n]) = \frac{1}{N^2} NA = \frac{A}{N}.$$

Del problema anterior, se sabe que

$$\text{CRLB}(A) = \frac{A^2}{N(A + \frac{1}{2})} < \frac{A^2}{NA} = \frac{A}{N} = \text{var}(\bar{x}).$$

Por lo tanto, como $\text{var}(\bar{x}) > \text{CRLB}(A)$, \bar{x} no alcanza la CRLB incluso cuando $N \rightarrow \infty$. Como el MLE alcanza la CRLB de forma asintótica (ver el ejemplo 7.2 de [1]), es mejor estimador.

7.6.3. Problema 3

Se observan N muestras IID con PDF

a. Gaussiana

$$p(x; \mu) = \frac{1}{\sqrt{2\pi}} \exp \left[-\frac{1}{2}(x - \mu)^2 \right].$$

b. Exponencial

$$p(x; \lambda) = \begin{cases} \lambda \exp(-\lambda x) & x > 0 \\ 0 & x < 0. \end{cases}$$

En cada caso, encontrar el MLE del parámetro desconocido, asegurando que efectivamente maximice la función de verosimilitud. ¿Los estimadores tienen sentido?

Solución

a. Gaussiana. La PDF de los datos es

$$p(\mathbf{x}; \mu) = \frac{1}{(2\pi)^{\frac{N}{2}}} \exp \left[-\frac{1}{2} \sum_{n=0}^{N-1} (x[n] - \mu)^2 \right].$$

Maximizar $p(\mathbf{x}; \mu)$ es equivalente a minimizar

$$J(\mu) = \sum_{n=0}^{N-1} (x[n] - \mu)^2,$$

y por tratarse de una función cuadrática en μ con concavidad positiva, tiene solo un mínimo global. Derivando respecto al parámetro, se tiene que

$$\frac{\partial J(\mu)}{\partial \mu} = -2 \sum_{n=0}^{N-1} (x[n] - \mu)$$

e igualando a cero, se tiene que el estimador cumple que

$$N\hat{\mu} = \sum_{n=0}^{N-1} x[n] \quad \Rightarrow \quad \hat{\mu} = \frac{1}{N} \sum_{n=0}^{N-1} x[n] = \bar{x}.$$

b. Exponencial. La PDF de los datos es

$$p(\mathbf{x}; \lambda) = \begin{cases} \lambda^N \exp \left(-\lambda \sum_{n=0}^{N-1} x[n] \right) & x[0] > 0, \dots, x[N-1] > 0 \\ 0 & \text{en otro caso.} \end{cases}$$

Para encontrar el máximo de $p(\mathbf{x}; \lambda)$, se observa que el mismo coincide con el máximo de $\ln p(\mathbf{x}; \lambda)$ por ser $\ln x$ una función monótonamente creciente con x . Por lo tanto, se buscará el máximo de

$$\ln p(\mathbf{x}; \lambda) = N \ln \lambda - \lambda \sum_{n=0}^{N-1} x[n].$$

Derivando respecto al parámetro, se tiene que

$$\frac{\partial \ln p(\mathbf{x}; \lambda)}{\partial \lambda} = \frac{N}{\lambda} - \sum_{n=0}^{N-1} x[n],$$

e igualando a cero, se llega a que

$$\frac{N}{\hat{\lambda}} = \sum_{n=0}^{N-1} x[n] \quad \Rightarrow \quad \hat{\lambda} = \frac{1}{\frac{1}{N} \sum_{n=0}^{N-1} x[n]} = \frac{1}{\bar{x}}.$$

Además, se cumple que

$$\frac{\partial^2 \ln p(\mathbf{x}; \lambda)}{\partial \lambda^2} = -\frac{N}{\lambda^2} < 0 \quad \forall \lambda,$$

indicando que $\ln p(\mathbf{x}; \lambda)$ es una función de concavidad negativa por lo que el máximo es el máximo global. Si x tiene PDF exponencial, $E(x) = 1/\lambda$, por lo que el estimador tiene sentido.

7.6.4. Problema 4

En ocasiones, los resultados asintóticos pueden ser confusos y deben aplicarse con cuidado. Como ejemplo, dos estimadores insesgados de θ tienen varianzas

$$\begin{aligned} \text{var}(\hat{\theta}_1) &= \frac{2}{N} \\ \text{var}(\hat{\theta}_2) &= \frac{1}{N} + \frac{100}{N^2}. \end{aligned}$$

Graficar las varianzas respecto a N para determinar el mejor estimador.

Solución La gráfica de las varianzas se muestra en la figura 7.4. Se observa que $\text{var}(\hat{\theta}_1) < \text{var}(\hat{\theta}_2)$ si $N < 100$ y $\text{var}(\hat{\theta}_2) < \text{var}(\hat{\theta}_1)$ si $N > 100$. Por lo tanto, $\hat{\theta}_1$ es mejor estimador si $N < 100$ y $\hat{\theta}_2$ es mejor estimador si $N > 100$.

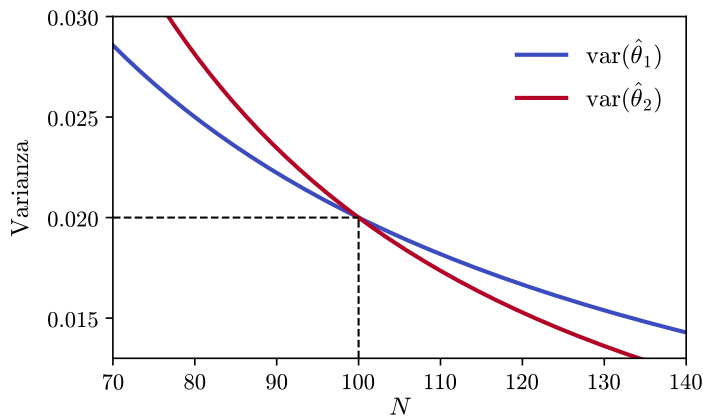


Figura 7.4: Varianzas de los estimadores $\hat{\theta}_1$ y $\hat{\theta}_2$ en función de la cantidad de muestras N . $\text{var}(\hat{\theta}_1) < \text{var}(\hat{\theta}_2)$ si $N < 100$ y $\text{var}(\hat{\theta}_2) < \text{var}(\hat{\theta}_1)$ si $N > 100$.

7.6.5. Problema 5

Una definición formal de consistencia de un estimador es la siguiente. Un estimador $\hat{\theta}$ es consistente si, dado $\epsilon > 0$,

$$\lim_{N \rightarrow \infty} \Pr\{|\hat{\theta} - \theta| > \epsilon\} = 0. \quad (7.28)$$

Probar que la media muestral es un estimador consistente para el problema de la estimación del nivel A de DC en WGN con varianza conocida. Sugerencia: emplear la desigualdad de Chebychev.

Solución Una secuencia de variables aleatorias $\hat{\theta}$ que converge a θ con N cumpliendo la ecuación 7.28, se dice que *converge en probabilidad* (ver el capítulo 7 de [2]). Por lo tanto, la definición de consistencia indica que el estimador $\hat{\theta}$ es consistente si converge en probabilidad al valor verdadero θ del parámetro. La desigualdad de Chebychev indica que si una variable aleatoria x tiene media μ y varianza σ^2 , para todo $\epsilon > 0$ se cumple que (ver el capítulo 5 de [2])

$$\Pr\{|x - \mu| \geq \epsilon\} \leq \frac{\sigma^2}{\epsilon^2}.$$

En el problema de la sección 2.1.3 se mostró que la media muestral cumple que $\bar{x} \sim \mathcal{N}(A, \sigma^2/N)$. Aplicando la desigualdad de Chebychev a \bar{x} , se tiene que

$$\Pr\{|\bar{x} - A| \geq \epsilon\} \leq \frac{\sigma^2}{N\epsilon^2}.$$

Por lo tanto, para todo $\epsilon > 0$ se cumple que

$$\lim_{N \rightarrow \infty} \Pr\{|\bar{x} - A| \geq \epsilon\} \leq \lim_{N \rightarrow \infty} \frac{\sigma^2}{N\epsilon^2} = 0$$

y por lo tanto, \bar{x} es consistente.

7.6.6. Problema 6

Otro resultado sobre consistencia establece que si $\alpha = g(\theta)$ para una función g continua y $\hat{\theta}$ es un estimador consistente de θ , se cumple que $\hat{\alpha} = g(\hat{\theta})$ es un estimador consistente de α . Empleando el argumento de linealización estadística y la definición formal de consistencia, mostrar que la afirmación es cierta. Sugerencia: linealizar g en torno al valor verdadero de θ .

Solución Hay que demostrar que

$$\lim_{N \rightarrow \infty} \Pr\{|\hat{\theta} - \theta| \geq \epsilon\} = 0 \quad \Rightarrow \quad \lim_{N \rightarrow \infty} \Pr\{|\hat{\alpha} - \alpha| \geq \epsilon\} = 0.$$

Sea θ_0 el valor verdadero del parámetro. Linealizando g en torno a θ_0 se tiene que

$$\alpha = g(\theta) \approx g(\theta_0) + \left. \frac{dg}{d\theta} \right|_{\theta=\theta_0} (\theta - \theta_0).$$

Por lo tanto,

$$\begin{aligned} \hat{\alpha} - \alpha &\approx \left[g(\theta_0) + \left. \frac{dg}{d\theta} \right|_{\theta=\theta_0} (\hat{\theta} - \theta_0) \right] - \left[g(\theta_0) + \left. \frac{dg}{d\theta} \right|_{\theta=\theta_0} (\theta - \theta_0) \right] \\ &= \left. \frac{dg}{d\theta} \right|_{\theta=\theta_0} (\hat{\theta} - \theta). \end{aligned}$$

De esta forma

$$\begin{aligned} \Pr\{|\hat{\alpha} - \alpha| \geq \epsilon\} &\approx \Pr\left\{ \left| \left. \frac{dg}{d\theta} \right|_{\theta=\theta_0} \right| |\hat{\theta} - \theta| \geq \epsilon \right\} \\ &= \Pr\left\{ |\hat{\theta} - \theta| \geq \frac{\epsilon}{\left| \left. \frac{dg}{d\theta} \right|_{\theta=\theta_0} \right|} \right\} \\ &= \Pr\{|\hat{\theta} - \theta| \geq \epsilon'\}. \end{aligned}$$

Finalmente

$$\lim_{N \rightarrow \infty} \Pr\{|\hat{\alpha} - \alpha| \geq \epsilon\} = \lim_{N \rightarrow \infty} \Pr\{|\hat{\theta} - \theta| \geq \epsilon'\} = 0,$$

donde la segunda igualdad se cumple para todo $\epsilon' > 0$ debido a que $\hat{\theta}$ es consistente y $|dg/d\theta|_{\theta=\theta_0}|$ es finito por ser la función g continua. Se concluye que $\hat{\alpha}$ es consistente.

7.6.7. Problema 7

Se consideran N observaciones IID de la familia de PDF exponenciales

$$p(x; \theta) = \exp[A(\theta)B(x) + C(x) + D(\theta)] \quad (7.29)$$

donde A , B , C y D son funciones de sus respectivos argumentos. Encontrar la ecuación a resolver para encontrar el MLE. Aplicar los resultados a las PDF del problema de la sección 7.6.3.

Solución La PDF de los datos es

$$\begin{aligned} p(\mathbf{x}; \theta) &= \prod_{n=0}^{N-1} \exp[A(\theta)B(x[n]) + C(x[n]) + D(\theta)] \\ &= \exp \left[A(\theta) \sum_{n=0}^{N-1} B(x[n]) + \sum_{n=0}^{N-1} C(x[n]) + ND(\theta) \right], \end{aligned}$$

y su máximo se obtiene maximizando

$$J(\theta) = A(\theta) \sum_{n=0}^{N-1} B(x[n]) + \sum_{n=0}^{N-1} C(x[n]) + ND(\theta).$$

La derivada respecto al parámetro es

$$\frac{\partial J(\theta)}{\partial \theta} = A'(\theta) \sum_{n=0}^{N-1} B(x[n]) + ND'(\theta),$$

e igualando a cero, se obtiene que el MLE $\hat{\theta}$ cumple que

$$A'(\hat{\theta}) \sum_{n=0}^{N-1} B(x[n]) + ND'(\hat{\theta}) = 0. \quad (7.30)$$

Se aplicará el resultado para encontrar el MLE de las PDF de la familia exponencial del problema de la sección 7.6.3.

a. Gaussiana

$$\begin{aligned} p(x; \mu) &= \frac{1}{\sqrt{2\pi}} \exp \left[-\frac{1}{2}(x - \mu)^2 \right] \\ &= \exp \left[-\frac{1}{2} \ln 2\pi - \frac{x^2}{2} + x\mu - \frac{\mu^2}{2} \right]. \end{aligned}$$

Por lo tanto, asociando términos con la ecuación 7.29, se tiene que

$$A(\mu) = \mu, \quad B(x) = x, \quad C(x) = -\frac{1}{2} \ln 2\pi - \frac{x^2}{2}, \quad D(\mu) = -\frac{\mu^2}{2}.$$

y aplicando la ecuación 7.30, el MLE cumple que

$$1 \times \sum_{n=0}^{N-1} x[n] - N\hat{\mu} = 0,$$

concluyendo que

$$\hat{\mu} = \frac{1}{N} \sum_{n=0}^{N-1} x[n] = \bar{x}.$$

b. Exponencial. Si $x > 0$,

$$\begin{aligned} p(x; \lambda) &= \lambda \exp(-\lambda x) \\ &= \exp(\ln \lambda - \lambda x), \end{aligned}$$

y asociando términos con la ecuación 7.29, se tiene que

$$A(\mu) = -\lambda, \quad B(x) = x, \quad C(x) = 0, \quad D(\lambda) = \ln \lambda.$$

Aplicando la ecuación 7.30, el MLE cumple que

$$-1 \times \sum_{n=0}^{N-1} x[n] + N \times \frac{1}{\lambda} = 0$$

concluyendo que

$$\hat{\lambda} = \frac{1}{\frac{1}{N} \sum_{n=0}^{N-1} x[n]} = \frac{1}{\bar{x}}.$$

7.6.8. Problema 8

Si se observan N muestras IID de un experimento Bernoulli (tirar una moneda) con probabilidades

$$\begin{aligned} \Pr\{x[n] = 1\} &= p \\ \Pr\{x[n] = 0\} &= 1 - p \end{aligned}$$

encontrar el MLE.

Solución Teniendo en cuenta que la PDF de una muestra es

$$p(x[n]; p) = p^{x[n]}(1 - p)^{1-x[n]},$$

la PDF de los datos es

$$\begin{aligned} p(\mathbf{x}, p) &= \prod_{n=0}^{N-1} p^{x[n]}(1 - p)^{1-x[n]} \\ &= p^{\sum_{n=0}^{N-1} x[n]}(1 - p)^{N - \sum_{n=0}^{N-1} x[n]}. \end{aligned}$$

La derivada respecto al parámetro es

$$\begin{aligned} \frac{\partial p(\mathbf{x}, p)}{\partial p} &= \left(\sum_{n=0}^{N-1} x[n] \right) p^{(\sum_{n=0}^{N-1} x[n])-1} (1 - p)^{N - \sum_{n=0}^{N-1} x[n]} \\ &\quad - p^{\sum_{n=0}^{N-1} x[n]} \left(N - \sum_{n=0}^{N-1} x[n] \right) (1 - p)^{N-1 - \sum_{n=0}^{N-1} x[n]} \\ &= p^{(\sum_{n=0}^{N-1} x[n])-1} (1 - p)^{N-1 - \sum_{n=0}^{N-1} x[n]} \left[\left(\sum_{n=0}^{N-1} x[n] \right) (1 - p) - p \left(N - \sum_{n=0}^{N-1} x[n] \right) \right], \end{aligned}$$

e igualando a cero, el MLE cumple que

$$\left(\sum_{n=0}^{N-1} x[n] \right) (1 - \hat{p}) - \hat{p} \left(N - \sum_{n=0}^{N-1} x[n] \right) = 0,$$

es decir

$$\hat{p} = \frac{1}{N} \sum_{n=0}^{N-1} x[n] = \bar{x}.$$

7.6.9. Problema 9

Para N observaciones IID de una PDF $\mathcal{U}[0, \theta]$, encontrar el MLE de θ .

Solución La PDF de una muestra es

$$p(x[n]; \theta) = \begin{cases} \frac{1}{\theta}, & x[n] \in [0, \theta] \\ 0, & \text{en otro caso} \end{cases}$$

y la PDF de los datos es

$$p(\mathbf{x}; \theta) = \begin{cases} \frac{1}{\theta^N}, & x[0] \in [0, \theta], \dots, x[N-1] \in [0, \theta] \\ 0, & \text{en otro caso.} \end{cases}$$

Se observa que la función de verosimilitud no es acotada, ya que

$$\lim_{\theta \rightarrow 0} p(\mathbf{x}; \theta) = \lim_{\theta \rightarrow 0} \frac{1}{\theta^N} = \infty.$$

Por lo tanto, la función se maximiza cuando θ es lo mas pequeño posible. Como se cumple que $\theta > x[n]$ para $n = 0, \dots, N-1$, el valor mas pequeño posible es $\theta_{\min} = \max x[n]$. De esta forma, el MLE es

$$\hat{\theta} = \max x[n].$$

7.6.10. Problema 10

Si se observa el conjunto de datos

$$x[n] = As[n] + w[n], \quad n = 0, \dots, N-1,$$

donde $s[n]$ es conocida y $w[n]$ es WGN de varianza σ^2 conocida, encontrar el MLE de A . Determinar la PDF del MLE y si se verifica la PDF asintótica del MLE.

Solución La PDF de los datos es

$$p(\mathbf{x}, A) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - As[n])^2 \right].$$

Para maximizar la función de verosimilitud, hay que minimizar el exponente

$$J(A) = \sum_{n=0}^{N-1} (x[n] - As[n])^2.$$

La derivada es

$$\frac{\partial J(A)}{\partial A} = \sum_{n=0}^{N-1} 2(x[n] - As[n])(-s[n]).$$

e igualando a cero, el MLE cumple que

$$\sum_{n=0}^{N-1} (x[n] - \hat{A}s[n])s[n] = 0$$

es decir,

$$\hat{A} = \frac{\sum_{n=0}^{N-1} x[n]s[n]}{\sum_{n=0}^{N-1} s^2[n]}.$$

Además, \hat{A} es gaussiana por ser la combinación lineal de variables aleatorias gaussianas. La media de \hat{A} es

$$E(\hat{A}) = \frac{\sum_{n=0}^{N-1} E(x[n])s[n]}{\sum_{n=0}^{N-1} s^2[n]} \stackrel{(a)}{=} \frac{A \sum_{n=0}^{N-1} s^2[n]}{\sum_{n=0}^{N-1} s^2[n]} = A,$$

donde en (a) se consideró que $E(x[n]) = As[n]$. La varianza de \hat{A} es

$$\text{var}(\hat{A}) = \frac{\sum_{n=0}^{N-1} \text{var}(x[n])s^2[n]}{\left(\sum_{n=0}^{N-1} s^2[n]\right)^2} = \frac{\sigma^2 \sum_{n=0}^{N-1} s^2[n]}{\left(\sum_{n=0}^{N-1} s^2[n]\right)^2} = \frac{\sigma^2}{\sum_{n=0}^{N-1} s^2[n]} = \frac{\sigma^2}{\mathbf{s}^T \mathbf{s}} \stackrel{(a)}{=} I^{-1}(A),$$

donde en (a) se notó que el modelo de los datos de este problema es un caso particular del modelo lineal (ecuación 4.1) con $\boldsymbol{\theta} = A$ y $\mathbf{H} = \mathbf{s}$, por lo que el lado izquierdo de la igualdad es el inverso de la información de Fisher del parámetro, como indica la ecuación 4.4. Se concluye que la PDF del MLE es

$$\hat{A} \sim \mathcal{N}(A, I^{-1}(A)),$$

que coincide con la PDF asintótica del MLE pero en este caso es válido para un número finito de muestras.

7.6.11. Problema 11

Encontrar la ecuación a resolver para el MLE del coeficiente de correlación ρ en el problema de la sección 3.6.15. Encontrar una solución aproximada cuando $N \rightarrow \infty$. Sugerencia: notar que $\sum_{n=0}^{N-1} x_0^2[n]/N \rightarrow 1$ y $\sum_{n=0}^{N-1} x_1^2[n]/N \rightarrow 1$ donde $\mathbf{x}[n] = [x_0[n] \ x_1[n]]$.

Solución Hay que maximizar la función de verosimilitud. Como el logaritmo es una función monotonamente creciente, es equivalente maximizar la función de verosimilitud logarítmica. La derivada de la función logarítmica se calculó en el problema de la sección 3.6.15 y está dada por la ecuación 3.71,

$$\frac{\partial \ln p(\mathbf{x}; \rho)}{\partial \rho} = \frac{N\rho}{(1-\rho^2)} + \frac{1}{(1-\rho^2)^2} \sum_{n=0}^{N-1} [(1+\rho^2)x_0[n]x_1[n] - \rho(x_0^2[n] + x_1^2[n])].$$

Igualando a cero, se obtiene que el MLE de ρ cumple que

$$\frac{N\hat{\rho}}{(1-\hat{\rho}^2)} + \frac{1}{(1-\hat{\rho}^2)^2} \sum_{n=0}^{N-1} [(1+\hat{\rho}^2)x_0[n]x_1[n] - \hat{\rho}(x_0^2[n] + x_1^2[n])] = 0,$$

o

$$\hat{\rho}(1-\hat{\rho}^2) + (1+\hat{\rho}^2) \frac{1}{N} \sum_{n=0}^{N-1} x_0[n]x_1[n] - \hat{\rho} \left(\frac{1}{N} \sum_{n=0}^{N-1} x_0^2[n] + \frac{1}{N} \sum_{n=0}^{N-1} x_1^2[n] \right) = 0.$$

Cuando $N \rightarrow \infty$, por la ley de los grandes números, se cumple que

$$\frac{1}{N} \sum_{n=0}^{N-1} x_0^2[n] \rightarrow E(x_0^2[n]) = \text{var}(x_0[n]) = 1, \quad \frac{1}{N} \sum_{n=0}^{N-1} x_1^2[n] \rightarrow E(x_1^2[n]) = \text{var}(x_1[n]) = 1,$$

por lo que la condición del MLE se puede aproximar como

$$\hat{\rho}(1-\hat{\rho}^2) + (1+\hat{\rho}^2) \frac{1}{N} \sum_{n=0}^{N-1} x_0[n]x_1[n] - 2\hat{\rho} = 0.$$

Definiendo $\hat{C}_{12} = \sum_{n=0}^{N-1} x_0[n]x_1[n]/N$ y operando, se tiene que

$$\hat{\rho} - \hat{\rho}^3 + \hat{C}_{12} + \hat{C}_{12}\hat{\rho}^2 - 2\hat{\rho} = 0 \quad \Rightarrow \quad \hat{\rho}^3 - \hat{C}_{12}\hat{\rho}^2 + \hat{\rho} - \hat{C}_{12} = 0.$$

Observando que el polinomio se puede escribir como

$$\hat{\rho}^2(\hat{\rho} - \hat{C}_{12}) + (\hat{\rho} - \hat{C}_{12}) = 0 \quad \Rightarrow \quad (\hat{\rho}^2 + 1)(\hat{\rho} - \hat{C}_{12}) = 0,$$

se concluye que

$$\hat{\rho} = \hat{C}_{12} = \frac{1}{N} \sum_{n=0}^{N-1} x_0[n]x_1[n].$$

7.6.12. Problema 12

En este problema se prueba que si existe un estimador eficiente, el método de máxima verosimilitud lo obtiene. Asumiendo un parámetro escalar, si existe un estimador eficiente, se cumple que

$$\frac{\partial \ln p(\mathbf{x}; \rho)}{\partial \rho} = I(\theta)(\hat{\theta} - \theta).$$

Usar esto para probar el teorema.

Solución Maximizar la función de verosimilitud es equivalente a maximizar la función de verosimilitud logarítmica. Asumiendo que $I(\theta) > 0$ para todo θ , ya que de lo contrario la PDF no dependería del parámetro, el MLE cumple que

$$I(\theta_{\text{MLE}})(\hat{\theta} - \theta_{\text{MLE}}) = 0$$

y por lo tanto,

$$\theta_{\text{MLE}} = \hat{\theta},$$

donde $\hat{\theta}$ es el estimador eficiente.

7.6.13. Problema 13

Se observan N muestras IID del nivel de continua en WGN de varianza conocida σ^2 . Mediante una simulación por computadora, verificar que la PDF del MLE o la media muestral es $\mathcal{N}(0, \sigma^2/N)$. Graficar las PDFs teórica y la obtenida mediante la simulación. Usar $A = 1$, $\sigma^2 = 0.1$, $N = 50$ y $M = 1000$ realizaciones. ¿Qué ocurre si M se incrementa a 5000?

Solución Con los valores indicados, la PDF teórica de la media muestral es $\mathcal{N}(1, 0.002)$. Se realizó una simulación por computadora repitiendo la estimación $M = 1000$ y $M = 10000$ veces. El resultado se muestra en la figura 7.5. Se observa que al incrementar la cantidad de realizaciones, la PDF experimental se ajusta mas a la PDF teórica.

7.6.14. Problema 14

Se consideran los datos $x[n]$ IID para $n = 0, \dots, N-1$, con $x[n] \sim \mathcal{N}(0, \sigma^2)$. De acuerdo al teorema de Slutsky, si \bar{x} es la media muestral y

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{n=0}^{N-1} (x[n] - \bar{x})^2$$

es la varianza muestral, se cumple que

$$\frac{\bar{x}}{\hat{\sigma}/\sqrt{N}} \stackrel{a}{\sim} \mathcal{N}(0, 1).$$

Si bien esto puede ser probado analíticamente, requiere familiaridad con el concepto de convergencia de variables aleatorias. En lugar de ello, implementar una simulación por computadora para encontrar la PDF de $\bar{x}/(\hat{\sigma}/\sqrt{N})$ para $\sigma^2 = 1$ y $N = 10$ y $N = 100$. Comparar el resultado con la PDF asintótica teórica.

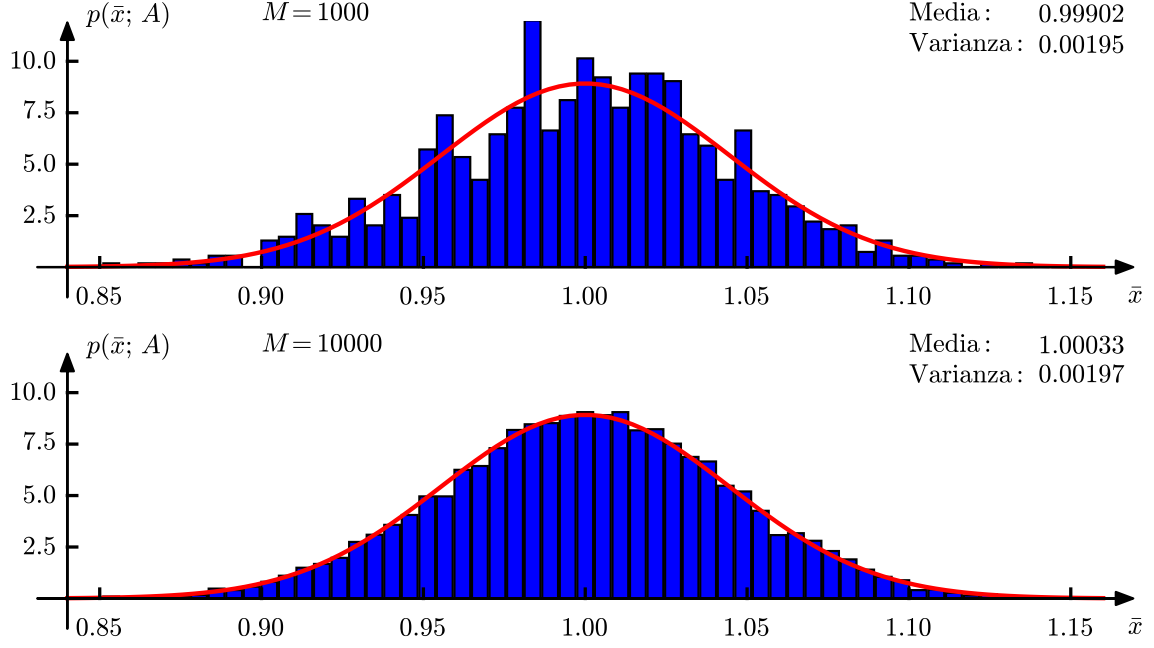


Figura 7.5: PDF teórica y experimental del MLE. La PDF teórica (en rojo) es $\mathcal{N}(1, 0.002)$.

Solución Se realizará primero un esbozo de la demostración analítica para el caso en que $x[n]$ son IID con $E(x[n]) = \mu$ y $\text{var}(x[n]) = \sigma^2$ pero con PDF arbitraria. En ese caso,

$$\frac{\bar{x} - \mu}{\hat{\sigma}/\sqrt{N}} \xrightarrow{d} \mathcal{N}(0, 1).$$

La notación significa convergencia en distribución. Para mostrarlo, se considera primero la variable aleatoria $\sqrt{N}(\bar{x} - \mu)$. Como $E[\sqrt{N}(\bar{x} - \mu)] = 0$ y $\text{var}[\sqrt{N}(\bar{x} - \mu)] = \sigma^2$, por el teorema central del límite clásico⁴, se cumple que con $N \rightarrow \infty$

$$\sqrt{N}(\bar{x} - \mu) \xrightarrow{d} x \sim \mathcal{N}(0, \sigma^2).$$

Por otro lado, operando con la varianza muestral, se ve que

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{N} \sum_{n=0}^{N-1} (x[n] - \bar{x})^2 \\ &= \frac{1}{N} \sum_{n=0}^{N-1} [(x[n] - \mu) - (\bar{x} - \mu)]^2 \\ &= \frac{1}{N} \sum_{n=0}^{N-1} (x[n] - \mu)^2 - \frac{1}{N} \sum_{n=0}^{N-1} 2(\bar{x} - \mu)(x[n] - \mu) + \frac{1}{N} \sum_{n=0}^{N-1} (\bar{x} - \mu)^2 \\ &= \frac{1}{N} \sum_{n=0}^{N-1} (x[n] - \mu)^2 - 2(\bar{x} - \mu) \frac{1}{N} \sum_{n=0}^{N-1} (x[n] - \mu) + (\bar{x} - \mu)^2 \\ &= \frac{1}{N} \sum_{n=0}^{N-1} (x[n] - \mu)^2 - 2(\bar{x} - \mu) \left(\frac{1}{N} \sum_{n=0}^{N-1} x[n] - \mu \right) + (\bar{x} - \mu)^2, \end{aligned}$$

⁴Ver https://en.wikipedia.org/wiki/Central_limit_theorem#Classical_CLT, por ejemplo.

resultando en

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{n=0}^{N-1} (x[n] - \mu)^2 - (\bar{x} - \mu)^2.$$

Por la ley de los grandes números, se cumple que

$$\bar{x} \xrightarrow{a.s.} E(x[n]) = \mu \quad \text{y} \quad \frac{1}{N} \sum_{n=0}^{N-1} (x[n] - \mu)^2 \xrightarrow{a.s.} E[(x[n] - \mu)^2] = \sigma^2,$$

y por lo tanto,

$$\hat{\sigma}^2 \xrightarrow{a.s.} \sigma^2.$$

La notación significa convergencia casi segura, como indica la ley fuerte de los grandes números. En la ley débil de los grandes números, la convergencia es en probabilidad. Pero como convergencia casi segura implica convergencia en probabilidad (ver el capítulo 7 de [2]), en cualquier caso se cumple que

$$\hat{\sigma}^2 \xrightarrow{p} \sigma^2.$$

Finalmente, empleando el teorema de Slutsky, referido previamente en este capítulo, se tiene que

$$\frac{\sqrt{N}(\bar{x} - \mu)}{\hat{\sigma}^2} \xrightarrow[p]{d} \frac{x}{\sigma^2} \sim \mathcal{N}(0, \sigma^2) \quad \Rightarrow \quad \frac{\sqrt{N}(\bar{x} - \mu)}{\hat{\sigma}} \xrightarrow[p]{d} \frac{x}{\sigma} \sim \mathcal{N}(0, 1).$$

Se realizó una simulación por computadora calculando $\sqrt{N}\bar{x}/\hat{\sigma}$ con valores de $N = 10$ y $N = 1000$ una cantidad $M = 1000$ veces. Los resultados de muestras en la figura 7.6, comparando el histograma de los valores prácticos con la PDF teórica $\mathcal{N}(0, 1)$. Como es de esperar, cuando mayor es N mejor es la aproximación al resultado asintótico teórico. Se observa que con $N = 10$ hay una diferencia considerable fundamentalmente en la varianza.

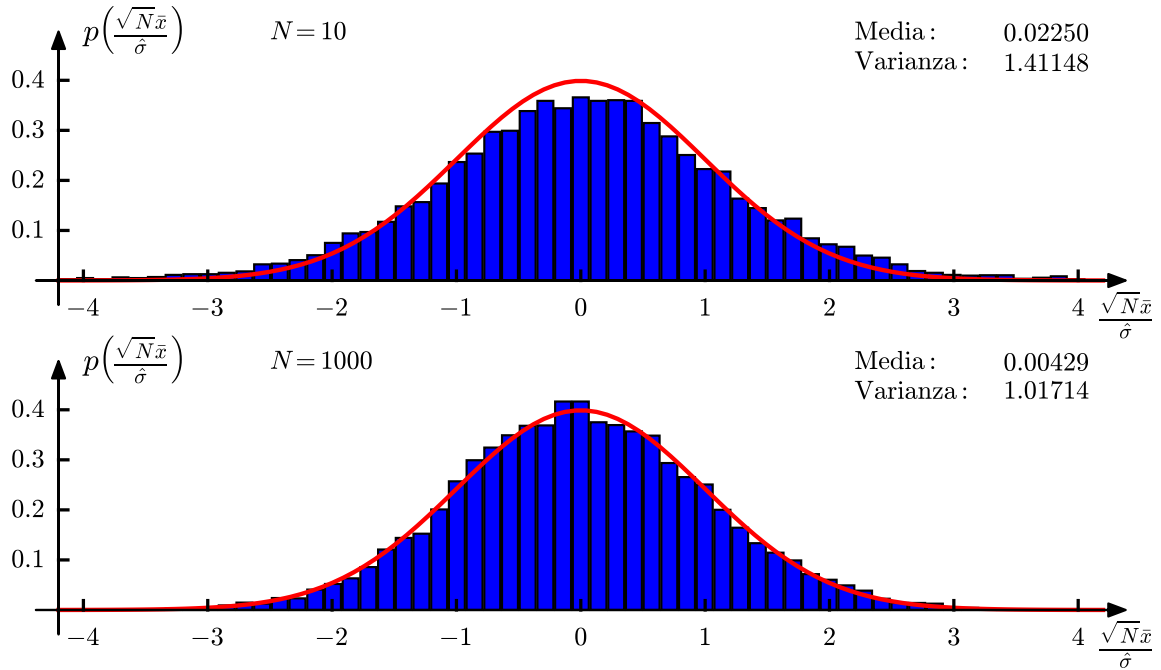


Figura 7.6: PDF teórica y experimental de $\sqrt{N}\bar{x}/\hat{\sigma}$. La PDF teórica (en rojo) es $\mathcal{N}(0, 1)$. Cuando N es pequeño, la varianza experimental es considerablemente mayor que la varianza asintótica teórica.

7.6.15. Problema 15

En este problema se muestra que el MLE alcanza la PDF asintótica cuando el error de estimación es pequeño. Se considera el ejemplo de la sección 7.2.2 del MLE de la fase de una senoide y sea

$$\begin{aligned}\epsilon_s &= -\frac{2}{NA} \sum_{n=0}^{N-1} w[n] \sin 2\pi f_0 n \\ \epsilon_c &= \frac{2}{NA} \sum_{n=0}^{N-1} w[n] \cos 2\pi f_0 n\end{aligned}$$

en la ecuación 7.6. Se asume que f_0 no es cercano a 0 ni a $1/2$ de forma que se pueden realizar las aproximaciones del problema 3.6.7. Probar primero que ϵ_s y ϵ_c son variables aleatorias gaussianas aproximadamente no correlacionadas y por lo tanto independientes. Además, determinar la PDF de ϵ_s y ϵ_c . Luego, asumiendo que ϵ_s y ϵ_c son pequeños, emplear una expansión en series de Taylor de $\hat{\phi}$ en torno al valor verdadero ϕ ,

$$\begin{aligned}\hat{\phi} &= g(\epsilon_s, \epsilon_c) \\ &\approx g(0, 0) + \left. \frac{\partial g(\epsilon_s, \epsilon_c)}{\partial \epsilon_s} \right|_{\epsilon_s=0, \epsilon_c=0} \epsilon_s + \left. \frac{\partial g(\epsilon_s, \epsilon_c)}{\partial \epsilon_c} \right|_{\epsilon_s=0, \epsilon_c=0} \epsilon_c,\end{aligned}$$

donde la función g esta dada por la ecuación 7.6. Usar esta expansión para encontrar la PDF de $\hat{\phi}$. Comparar la varianza con la CRLB calculada en el ejemplo 3.4. Notar que para que ϵ_s y ϵ_c sean pequeños, se debe cumplir que $N \rightarrow \infty$ y/o $A \rightarrow \infty$.

Solución La ecuación 7.6 expresada en función los términos de error ϵ_s y ϵ_c es

$$\hat{\phi} \approx \arctan \frac{\sin \phi + \epsilon_s}{\cos \phi + \epsilon_c}.$$

Se probará primero que ϵ_s y ϵ_c son aproximadamente no correlacionadas, es decir, que se cumple que

$$\text{cov}(\epsilon_s, \epsilon_c) = E(\epsilon_s \epsilon_c) - E(\epsilon_s)E(\epsilon_c) = 0.$$

Teniendo en cuenta que $E(\epsilon_s) = E(\epsilon_c) = 0$, se tiene que

$$\begin{aligned}\text{cov}(\epsilon_s, \epsilon_c) &= E(\epsilon_s \epsilon_c) \\ &= -\frac{4}{N^2 A^2} \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} E(w[m]w[n]) \sin(2\pi f_0 m) \cos(2\pi f_0 n) \\ &= -\frac{4\sigma^2}{N^2 A^2} \sum_{n=0}^{N-1} \sin(2\pi f_0 n) \cos(2\pi f_0 n),\end{aligned}$$

donde en la última igualdad se empleó que las muestras $w[n]$ de ruido son independientes y por lo tanto

$$E(w[m]w[n]) = \begin{cases} \sigma^2, & m = n \\ 0, & m \neq n \end{cases} = \sigma^2 \delta[m - n]. \quad (7.31)$$

Considerando la identidad trigonométrica $\sin 2\theta = 2 \sin \theta \cos \theta$ y empleando la aproximación demostrada en el problema de la sección 3.6.7, se obtiene que

$$\text{cov}(\epsilon_s, \epsilon_c) = -\frac{2\sigma^2}{N^2 A^2} \sum_{n=0}^{N-1} \sin 4\pi f_0 n \approx 0,$$

concluyendo que ϵ_s y ϵ_c son aproximadamente no correlacionadas. Además, son variables aleatorias gaussianas por ser una combinación lineal de variables aleatorias gaussianas independientes. Finalmente,

por ser aproximadamente no correlacionadas y gaussianas, ϵ_s y ϵ_c son aproximadamente independientes. Para calcular su PDF solo falta calcular las varianzas. Procediendo de forma análoga al cálculo de la covarianza, se tiene que

$$\begin{aligned}\text{var}(\epsilon_s) &= E(\epsilon_s^2) \\ &\stackrel{(a)}{=} \frac{4\sigma^2}{N^2 A^2} \sum_{n=0}^{N-1} \sin^2 2\pi f_0 n \\ &\stackrel{(b)}{=} \frac{2\sigma^2}{N^2 A^2} \left(N - \sum_{n=0}^{N-1} \cos 4\pi f_0 n \right) \\ &\stackrel{(c)}{\approx} \frac{2\sigma^2}{N A^2},\end{aligned}$$

donde en (a) se empleó la ecuación 7.31 en (b) se empleó la identidad trigonométrica $2\sin^2 \theta = 1 - \cos 2\theta$ y en (c) la aproximación demostrada en el problema 3.6.7. Análogamente,

$$\begin{aligned}\text{var}(\epsilon_c) &= E(\epsilon_c^2) \\ &= \frac{4\sigma^2}{N^2 A^2} \sum_{n=0}^{N-1} \cos^2 2\pi f_0 n \\ &\stackrel{(a)}{=} \frac{2\sigma^2}{N^2 A^2} \left(N + \sum_{n=0}^{N-1} \cos 4\pi f_0 n \right) \\ &\approx \frac{2\sigma^2}{N A^2},\end{aligned}$$

donde en (a) se empleó la identidad trigonométrica $2\cos^2 \theta = 1 + \cos 2\theta$. Se concluye que

$$\epsilon_s, \epsilon_c \sim \mathcal{N}\left(0, \frac{2\sigma^2}{N A^2}\right).$$

Se aproximará el estimador mediante la serie de Taylor indicada. Como

$$g(\epsilon_s, \epsilon_c) = \arctan \frac{\sin \phi + \epsilon_s}{\cos \phi + \epsilon_c},$$

y $d \arctan x / dx = 1/(1+x^2)$, se tiene que

$$\frac{\partial g(\epsilon_s, \epsilon_c)}{\partial \epsilon_s} = \frac{1}{1 + \left(\frac{\sin \phi + \epsilon_s}{\cos \phi + \epsilon_c}\right)^2} \frac{1}{\cos \phi + \epsilon_c} \quad \text{y} \quad \frac{\partial g(\epsilon_s, \epsilon_c)}{\partial \epsilon_c} = -\frac{1}{1 + \left(\frac{\sin \phi + \epsilon_s}{\cos \phi + \epsilon_c}\right)^2} \frac{\sin \phi + \epsilon_s}{(\cos \phi + \epsilon_c)^2}.$$

Además, $g(0, 0) = \phi$,

$$\left. \frac{\partial g(\epsilon_s, \epsilon_c)}{\partial \epsilon_s} \right|_{\epsilon_s=0, \epsilon_c=0} = \frac{1}{1 + \frac{\sin^2 \phi}{\cos^2 \phi}} \frac{1}{\cos \phi} = \cos \phi$$

y

$$\left. \frac{\partial g(\epsilon_s, \epsilon_c)}{\partial \epsilon_c} \right|_{\epsilon_s=0, \epsilon_c=0} = -\frac{1}{1 + \frac{\sin^2 \phi}{\cos^2 \phi}} \frac{\sin \phi}{\cos^2 \phi} = -\sin \phi.$$

Sustituyendo estos resultados en la serie de Taylor se concluye que

$$\hat{\phi} \approx \phi + \epsilon_s \cos \phi - \epsilon_c \sin \phi.$$

Para encontrar la PDF de $\hat{\phi}$, se observa que se trata de una variable aleatoria gaussiana por ser la combinación lineal de dos variables aleatorias gaussianas independientes. Además, $E(\hat{\phi}) = \phi$ y

$$\begin{aligned}\text{var}(\hat{\phi}) &\stackrel{(a)}{=} \text{var}(\epsilon_s) \cos^2 \phi + \text{var}(\epsilon_c) \sin^2 \phi \\ &= \frac{2\sigma^2}{NA^2} (\cos^2 + \sin^2) \\ &= \frac{2\sigma^2}{NA^2},\end{aligned}$$

donde en (a) se tuvo en cuenta que ϵ_s y ϵ_c son independientes. Se concluye que

$$\hat{\phi} \sim \mathcal{N}\left(\phi, \frac{2\sigma^2}{NA^2}\right).$$

En el ejemplo de la sección 3.3.1 se demuestra que la CRLB de ϕ es

$$\text{var}(\hat{\phi}) \geq \frac{2\sigma^2}{NA^2}.$$

Por lo tanto, el MLE alcanza aproximadamente la CRLB.

7.6.16. Problema 16

Para el ejemplo de la sección 7.2.3, determinar la $\text{var}(\hat{A})$ así como la CRLB para el ruido de PDF de Laplace

$$p(w[0]) = \frac{1}{2} e^{-|w[n]|}$$

¿El MLE alcanza la CRLB cuando $N \rightarrow \infty$?

Solución Como se indica en el ejemplo, la varianza del MLE es

$$\text{var}(\hat{A}) = \int_{-\infty}^{\infty} u^2 p(u) du,$$

y con $w[n]$ con PDF de Laplace queda

$$\begin{aligned}\text{var}(\hat{A}) &= \frac{1}{2} \int_{-\infty}^{\infty} u^2 e^{-|u|} du \\ &= \int_0^{\infty} u^2 e^{-u} du \\ &\stackrel{(a)}{=} -u^2 e^{-u} \Big|_0^{\infty} - 2 \int_0^{\infty} u e^{-u} du \\ &\stackrel{(b)}{=} 2 \left(u e^{-u} \Big|_0^{\infty} + \int_0^{\infty} e^{-u} du \right) \\ &= -2e^{-u} \Big|_0^{\infty} \\ &= 2,\end{aligned}$$

donde en (a) se aplicó integración por partes $\int x dy = xy - \int y dx$ con $x = u^2$ y $dy = e^{-u} du$ y por lo tanto, $dx = 2u du$ y $y = -e^{-u}$ y en (b) se tuvo en cuenta que el primer término es nulo y nuevamente se aplicó integración por partes en la integral con $x = u$ y $dy = -e^{-u} du$ y por lo tanto, $dx = du$ y $y = e^{-u}$. Por otro lado, la CRLB es (ver el problema 3.6.2)

$$\text{var}(\hat{A}) \geq \left[\int_{-\infty}^{\infty} \frac{\left(\frac{dp_{w[0]}(u)}{du} \right)^2}{p_{w[0]}(u)} du \right]^{-1}.$$

Teniendo en cuenta que en este caso

$$\frac{dp_{w[0]}(u)}{du} = \begin{cases} -\frac{1}{2}e^{-u}, & u > 0 \\ \frac{1}{2}e^u, & u < 0 \end{cases} \Rightarrow \left(\frac{dp_{w[0]}(u)}{du} \right)^2 = \begin{cases} \frac{1}{4}e^{-2u}, & u > 0 \\ \frac{1}{4}e^{2u}, & u < 0 \end{cases} = \frac{1}{4}e^{-2|u|},$$

la CRLB queda

$$\begin{aligned} \text{var}(\hat{A}) &\geq \left[\int_{-\infty}^{\infty} \frac{\frac{1}{4}e^{-2|u|}}{\frac{1}{2}e^{-|u|}} du \right]^{-1} \\ &\stackrel{(a)}{=} \left[2 \int_0^{\infty} \frac{\frac{1}{4}e^{-2u}}{\frac{1}{2}e^{-u}} du \right]^{-1} \\ &= \left[\int_0^{\infty} e^{-u} du \right]^{-1} \\ &= 1, \end{aligned}$$

donde en (a) se consideró que tanto el numerador como el denominador del integrando son funciones pares resultando en una función par. Se obtuvo que $\text{var}(\hat{A}) = 2$ mientras que $\text{CRLB}(A) = 1$. Además, como la varianza del MLE es constante con N , no se cumple que el MLE alcance la CRLB con $N \rightarrow \infty$.

7.6.17. Problema 17

Para N observaciones IID de una PDF $\mathcal{N}(0, 1/\theta)$ donde $\theta > 0$, encontrar el MLE de θ y su PDF asintótica.

Solución El MLE de θ puede obtenerse maximizando $p(\mathbf{x}; \theta)$, donde

$$p(\mathbf{x}; \theta) = \left(\frac{\theta}{2\pi} \right)^{\frac{N}{2}} \exp \left(-\frac{\theta}{2} \sum_{n=0}^{N-1} x^2[n] \right),$$

pero el problema se resolverá considerando que las observaciones tienen PDF $\mathcal{N}(0, \sigma^2)$, buscando el MLE de σ^2 y luego empleando la propiedad de invarianza ante transformaciones del MLE (sección 7.1.3) que establece que el MLE de θ es $\hat{\theta} = 1/\hat{\sigma}^2$. La PDF de los datos es

$$p(\mathbf{x}; \sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left(-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} x^2[n] \right).$$

Se maximizará la función de verosimilitud logarítmica,

$$\ln p(\mathbf{x}; \sigma^2) = -\frac{N}{2} \ln 2\pi - \frac{N}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{n=0}^{N-1} x^2[n].$$

La derivada respecto al parámetro es

$$\frac{\partial \ln p(\mathbf{x}; \sigma^2)}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{n=0}^{N-1} x^2[n],$$

e igualando a cero se obtiene que el MLE de σ^2 es

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{n=0}^{N-1} x^2[n].$$

Finalmente, el MLE de θ es

$$\hat{\theta} = \frac{1}{\hat{\sigma}^2} = \frac{1}{\frac{1}{N} \sum_{n=0}^{N-1} x^2[n]}.$$

De la sección 7.1.2, la PDF asintótica del MLE es

$$\hat{\theta} \stackrel{a}{\sim} \mathcal{N}(\theta, I^{-1}(\theta)).$$

Además, de la ecuación 3.9, se cumple que

$$I^{-1}(\theta) = I^{-1}(\sigma^2) \left(\frac{\partial g}{\partial \sigma^2} \right)^2,$$

donde $I(\sigma^2)$ fue calculado en el ejemplo 3.2.2 y es

$$I(\sigma^2) = \frac{N}{2\sigma^4}$$

y teniendo en cuenta que

$$\theta = g(\sigma^2) = \frac{1}{\sigma^2} \quad \Rightarrow \quad \frac{\partial g}{\partial \sigma^2} = -\frac{1}{\sigma^4},$$

la información del Fisher de θ es

$$I^{-1}(\theta) = \frac{2\sigma^4}{N} \left(-\frac{1}{\sigma^4} \right)^2 = \frac{2}{N\sigma^4} = \frac{2\theta^2}{N}.$$

Se concluye que

$$\hat{\theta} \stackrel{a}{\sim} \mathcal{N}\left(\theta, \frac{2\theta^2}{N}\right).$$

7.6.18. Problema 18

Graficar la función

$$g(x) = e^{-\frac{1}{2}x^2} + \frac{1}{10}e^{-\frac{1}{2}(x-10)^2}$$

sobre el dominio $-3 \leq x \leq 13$. Encontrar el máximo de la función a partir de la gráfica. Luego, emplear el método de Newton-Raphson para encontrar el máximo. Para eso, emplear los valores iniciales de $x_0 = 0.5$ y $x_0 = 9.5$. ¿Qué se puede decir sobre la importancia del valor inicial?

Solución En la figura 7.7 se muestra la gráfica de $g(x)$. Mediante su inspección se observa que hay dos máximos locales y el máximo global se encuentra en $x = 0$. También se deduce que hay un mínimo local entre los máximos. Para determinar los máximos y el mínimo analíticamente, hay que encontrar los ceros de la derivada, que es

$$g'(x) = -xe^{-\frac{1}{2}x^2} - \frac{x-10}{10}e^{-\frac{1}{2}(x-10)^2}.$$

Resolver $g'(x) = 0$ analíticamente no es posible, pero es posible hacerlo por computadora empleando el método iterativo de Newton-Raphson (ver la sección 7.7 de [1]). Para buscar un cero de $g'(x)$, o un extremo local de $g(x)$, se parte de un valor inicial x_0 , y en el paso k -ésimo la estimación del cero es

$$x_k = x_{k-1} - \frac{g'(x_{k-1})}{g''(x_{k-1})},$$

donde x_{k-1} es la estimación en el paso anterior. En este caso,

$$\begin{aligned} g''(x) &= -e^{-\frac{1}{2}x^2} - x \left[-xe^{-\frac{1}{2}x^2} \right] - \frac{x}{10}e^{-\frac{1}{2}(x-10)^2} - \frac{x-10}{10} \left[-\frac{x-10}{10}e^{-\frac{1}{2}(x-10)^2} \right] \\ &= (x^2 - 1)e^{-\frac{1}{2}x^2} + \frac{1}{10} [(x-10)^2 - 1] e^{-\frac{1}{2}(x-10)^2}. \end{aligned}$$

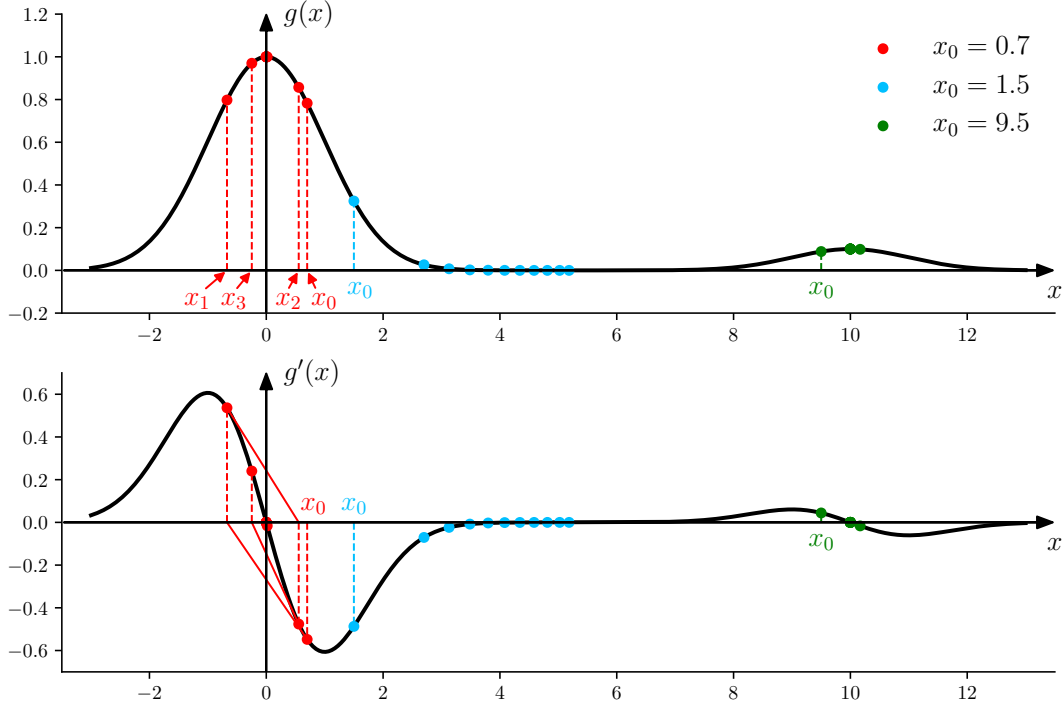


Figura 7.7: Método de Newton-Raphson para encontrar el máximo global de $g(x)$. Se muestran las primeras 10 iteraciones para encontrar los ceros de $g'(x)$ empleando distintos valores iniciales.

Se aplicó el método de Newton-Raphson para estimar el máximo global de $g(x)$ empleando como estimaciones iniciales los valores $x_0 = 0.7$, $x_0 = 1.5$ y $x_0 = 9.5$. En la figura 7.7 se muestran los resultados, específicamente, las 10 primeras iteraciones del método de Newton-Raphson. Se observa que en el caso en que $x_0 = 0.7$, el algoritmo converge al máximo global de $g(x)$ en $x = 0$, mientras que en el caso en que $x_0 = 9.5$, la convergencia es al máximo local en $x = 10$. Cuando el valor inicial es $x_0 = 1.5$, la convergencia es al mínimo en $x \approx 5.24$. Se concluye que para encontrar el máximo global, el valor inicial x_0 del método de Newton-Raphson debe ser cercano al valor del máximo $x_0 = 0$.

7.6.19. Problema 19

Si

$$x[n] = \cos 2\pi f_0 n + w[n], \quad n = 0, \dots, N-1$$

donde $w[n]$ es WGN de varianza σ^2 conocida, mostrar que el MLE de la frecuencia se obtiene aproximadamente maximizando (f_0 no cercano a 0 o $1/2$)

$$\sum_{n=0}^{N-1} x[n] \cos 2\pi f_0 n$$

sobre el intervalo $0 < f_0 < 1/2$. Luego, realizar una simulación Monte Carlo por computadora con $N = 10$, $f_0 = 0.25$, $\sigma^2 = 0.01$ y graficar la función a maximizar. Aplicar el algoritmo de Newton-Raphson para determinar el máximo y comparar los resultados con el obtenido mediante una búsqueda de grilla.

Solución La PDF de los datos es

$$p(\mathbf{x}; f_0) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - \cos 2\pi f_0 n)^2 \right],$$

por lo que la función de verosimilitud se maximiza al minimizar

$$\begin{aligned}\tilde{J}(f_0) &= \sum_{n=0}^{N-1} (x[n] - \cos 2\pi f_0 n)^2 \\ &= \sum_{n=0}^{N-1} x^2[n] - 2 \sum_{n=0}^{N-1} x[n] \cos 2\pi f_0 n + \sum_{n=0}^{N-1} \cos^2 2\pi f_0 n \\ &\stackrel{(a)}{\approx} \sum_{n=0}^{N-1} x^2[n] - 2 \sum_{n=0}^{N-1} x[n] \cos 2\pi f_0 n + \frac{N}{2},\end{aligned}$$

donde en (a) se tuvo en cuenta que la última sumatoria se puede aproximar como

$$\sum_{n=0}^{N-1} \cos^2 2\pi f_0 n = \sum_{n=0}^{N-1} \frac{1 + \cos 4\pi f_0 n}{2} \approx \frac{N}{2}$$

empleando la aproximación demostrada en el problema 3.6.7. Además, como $\tilde{J}(f_0)$ se minimiza maximizando el segundo sumando, el MLE se obtiene maximizando

$$J(f_0) = \sum_{n=0}^{N-1} x[n] \cos 2\pi f_0 n$$

en $0 < f_0 < 1/2$, que es lo que se quería demostrar. En la gráfica de la figura 7.8 se muestran 10 realizaciones de la función $J(f_0)$ obtenidas mediante una simulación por computadora empleando los valores $N = 10$, $f_0 = 0.25$ y $\sigma^2 = 0.01$. Se observa que efectivamente el máximo se encuentra en todos los casos cercano al valor verdadero $f_0 = 0.25$ del parámetro.

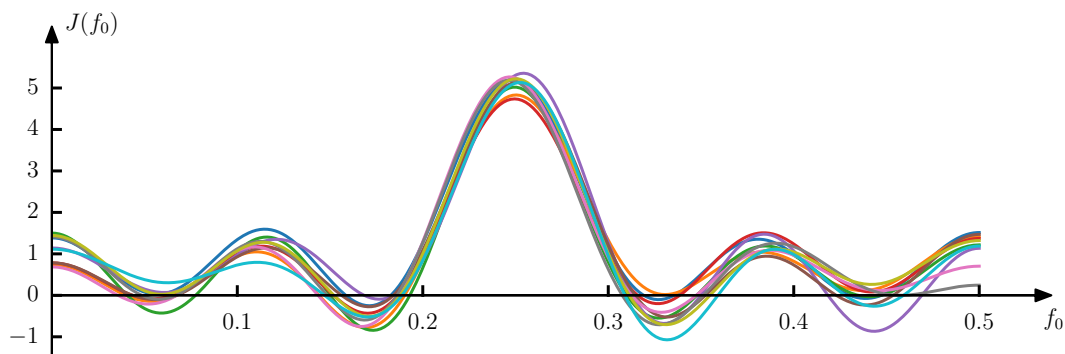


Figura 7.8: Función $J(f_0)$ a maximizar para obtener el MLE de f_0 . El valor verdadero del parámetro a estimar es $f_0 = 0.25$.

Se realizó la búsqueda del máximo en $M = 500$ realizaciones empleando una búsqueda de grilla y el método de Newton-Raphson con distintos valores iniciales. En este caso, la estimación en el paso k -ésimo del método de Newton-Raphson es

$$f_{0k+1} = f_{0k} - \frac{g(f_0)}{g'(f_0)},$$

con

$$g(f_0) = J'(f_0) = -2\pi \sum_{n=0}^{N-1} nx[n] \sin 2\pi f_0 n, \quad g'(f_0) = J''(f_0) = -(2\pi)^2 \sum_{n=0}^{N-1} n^2 x[n] \cos 2\pi f_0 n.$$

En la siguiente tabla se muestran los resultados.

	Búsqueda de grilla	Newton-Raphson			
		valor inicial (f_{0_0})			
		0.22	0.23	0.26	0.28
$E(\hat{f}_0)$	0.2500902	0.3233025	0.2500887	0.2500887	0.17483975
$\text{var}(\hat{f}_0)$	2.7×10^{-6}	1.3×10^{-1}	2.6×10^{-6}	2.6×10^{-6}	5.0×10^0

Se observa que con los valores iniciales ($f_{0_0} = 0.23$ y $f_{0_0} = 0.26$), el método de Newton-Raphson converge al máximo global en las 500 realizaciones. Con los valores iniciales ($f_{0_0} = 0.23$ y $f_{0_0} = 0.26$) el método no siempre converge al máximo global. En particular, con ($f_{0_0} = 0.23$) el método converge al mínimo ubicado en aproximadamente $f_0 = 0.32$ en la mayoría de la realizaciones y con ($f_{0_0} = 0.28$) converge al mínimo en $f_0 = 0.17$ en la mayoría de las realizaciones. En ambos casos, la varianza es grande debido a que en el resto de las realizaciones converge a un extremo arbitrario. En la figura 7.9 se muestra la estimación en las 500 realizaciones para todos los casos. Se observa que el método de Newton-Raphson converge al máximo global si el valor inicial es cercano al valor verdadero, que es el obtenido con la búsqueda de grilla.

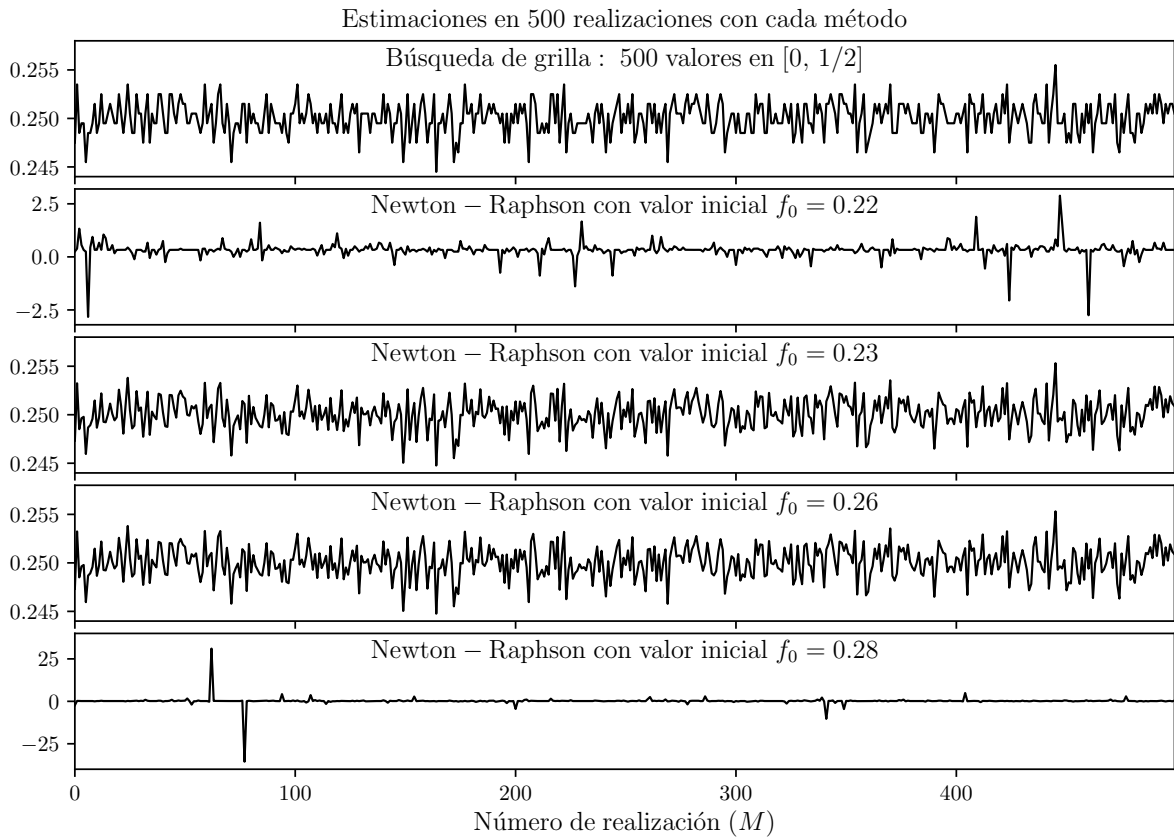


Figura 7.9: Estimaciones del MLE de f_0 en 500 realizaciones empleando búsqueda de grilla y el método de Newton-Raphson con distintos valores iniciales. El valor verdadero del parámetro a estimar es $f_0 = 0.25$. Para la búsqueda en grilla se emplearon 500 valores de f_0 en el intervalo $[0, 1/2]$.

7.6.20. Problema 20

Se considera el conjunto de datos

$$x[n] = s[n] + w[n], \quad n = 0, \dots, N-1,$$

donde $s[n]$ es desconocido para $n = 0, \dots, N-1$ y $w[n]$ es WGN de varianza conocida σ^2 . Determinar el MLE de $s[n]$ y su PDF. ¿Se cumplen las propiedades asintóticas del MLE? Es decir, ¿es el MLE

insesgado, eficiente, gaussiano y consistente?

Solución La PDF de los datos es

$$p(\mathbf{x}; \mathbf{s}) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - s[n])^2 \right],$$

por lo que el MLE se obtiene minimizando

$$J(\mathbf{s}) = \sum_{n=0}^{N-1} (x[n] - s[n])^2.$$

Derivando respecto al parámetro, se obtiene que

$$\frac{\partial J(\mathbf{s})}{\partial s[k]} = -2(x[k] - s[k]), \quad k = 0, \dots, N-1$$

e igualando a cero, se concluye que el MLE es

$$\hat{s}[k] = x[k], \quad k = 0, \dots, N-1,$$

es decir

$$\hat{\mathbf{s}} = \mathbf{x}.$$

Claramente, el MLE es insesgado, ya que

$$E(\hat{\mathbf{s}}) = E(\mathbf{x}) = \mathbf{s}.$$

Además, la matriz de covarianza del MLE es

$$\mathbf{C}_{\hat{\mathbf{s}}} = E[(\hat{\mathbf{s}} - \mathbf{s})(\hat{\mathbf{s}} - \mathbf{s})^T] = E(\mathbf{w}\mathbf{w}^T) = \sigma^2 \mathbf{I}.$$

Se concluye que el MLE es gaussiano con

$$\hat{\mathbf{s}} \sim \mathcal{N}(\mathbf{s}, \sigma^2 \mathbf{I}).$$

Par ver si es eficiente, hay que calcular la CRLB. La función de verosimilitud logarítmica es

$$\ln p(\mathbf{x}; \mathbf{s}) = -\frac{N}{2} \ln 2\pi\sigma^2 - \frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - s[n])^2.$$

Diferenciando,

$$\frac{\partial \ln p(\mathbf{x}; \mathbf{s})}{\partial s[k]} = \frac{1}{\sigma^2} (x[k] - s[k]),$$

y diferenciando nuevamente

$$\frac{\partial^2 \ln p(\mathbf{x}; \mathbf{s})}{\partial s[k] \partial s[l]} = \begin{cases} -\frac{1}{\sigma^2}, & k = l \\ 0 & k \neq l \end{cases} = -\frac{1}{\sigma^2} \delta[k - l],$$

resultando que la matriz de información de Fisher es

$$\mathbf{I}(\mathbf{s}) = \frac{1}{\sigma^2} \mathbf{I},$$

y la CRLB es

$$\text{var}(\hat{s}[k]) \geq [\mathbf{I}^{-1}(\mathbf{s})]_{kk} = \sigma^2.$$

Como el MLE alcanza la CRLB, es eficiente. Además, como

$$\lim_{N \rightarrow \infty} \text{var}(\hat{s}[k]) = \sigma^2 \neq 0,$$

el MLE no es consistente.

7.6.21. Problema 21

Para N observaciones IID con PDF $\mathcal{N}(A, \sigma^2)$, donde A y σ^2 son ambos desconocidos, encontrar el MLE de la SNR $\alpha = A^2/\sigma^2$.

Solución Por la propiedad de invarianza del MLE el MLE de α es $\hat{\alpha} = \hat{A}^2/\hat{\sigma}^2$, donde \hat{A} y $\hat{\sigma}^2$ son los MLE de A y σ^2 respectivamente. Para encontrar el MLE de $\boldsymbol{\theta} = [A \ \sigma^2]^T$ se parte de la PDF de los datos, que es

$$p(\mathbf{x}; \boldsymbol{\theta}) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - A)^2 \right].$$

Del ejemplo de la sección 3.2.2 se tiene que las derivadas de la función de verosimilitud logarítmica son

$$\frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial A} = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} (x[n] - A) \quad \frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{n=0}^{N-1} (x[n] - A)^2$$

Igualando a cero y despejando A en la primer ecuación se obtiene que

$$\hat{A} = \frac{1}{N} \sum_{n=0}^{N-1} x[n] = \bar{x}.$$

Resolviendo para σ^2 en la segunda ecuación y empleando el resultado anterior, se obtiene que

$$\frac{1}{\hat{\sigma}^2} \sum_{n=0}^{N-1} (x[n] - \bar{x})^2 = N \quad \Rightarrow \quad \hat{\sigma}^2 = \frac{1}{N} \sum_{n=0}^{N-1} (x[n] - \bar{x})^2.$$

Se concluye que el MLE de $\boldsymbol{\theta}$ es

$$\hat{\boldsymbol{\theta}} = \begin{bmatrix} \bar{x} \\ \frac{1}{N} \sum_{n=0}^{N-1} (x[n] - \bar{x})^2 \end{bmatrix}.$$

Finalmente el MLE de α es

$$\hat{\alpha} = \frac{\hat{A}^2}{\hat{\sigma}^2} = \frac{\bar{x}^2}{\frac{1}{N} \sum_{n=0}^{N-1} (x[n] - \bar{x})^2}.$$

7.6.22. Problema 22

Probar que si

$$\mathcal{A}(z) = 1 + \sum_{k=0}^p a[k]z^{-k}$$

es un polinomio de fase mínima (todas las raíces se encuentran dentro del círculo unidad), se cumple que

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \ln |A(f)|^2 df = 0.$$

Para hacerlo, probar primero que

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \ln |A(f)|^2 df = 2 \operatorname{Re} \left\{ \frac{1}{2\pi j} \oint_C \ln \mathcal{A}(z) \frac{dz}{z} \right\},$$

donde el contorno C es la circunferencia unidad en el plano z . Luego, notar que

$$\frac{1}{2\pi j} \oint_C \ln \mathcal{A}(z) \frac{dz}{z}$$

es la transformada z inversa de $\ln \mathcal{A}(z)$ evaluada en $n = 0$. Finalmente, emplear el hecho de que un polinomio $\mathcal{A}(z)$ de fase mínima conduce a una transformada z inversa de $\ln \mathcal{A}(z)$ que es una secuencia causal.

Solución Esta propiedad se probó en la deducción de la CRLB asintótica para procesos WSS gaussianos de la sección 3.5 y por lo tanto, se incluye a continuación solo un esbozo de la demostración. En la ecuación 3.48 se deduce que

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \ln |A(f)|^2 df = 2 \operatorname{Re} [\mathcal{Z}^{-1} \{\ln \mathcal{A}(z)\}|_{n=0}].$$

Teniendo en cuenta que $\mathcal{A}(z)$ es de fase mínima, todos los polos de $\ln \mathcal{A}(z)$, que son los ceros de $\mathcal{A}(z)$, se encuentran dentro del círculo unidad. Por lo tanto, la región de convergencia de $\ln \mathcal{A}(z)$ es causal y contiene a la circunferencia unidad, resultando en una transformada z inversa causal y estable. El teorema del valor inicial de la transformada z indica que para una secuencia $x[n]$ que es nula en $n < 0$, es decir, que es causal, se cumple que

$$\lim_{z \rightarrow \infty} X(z) = x[0].$$

Por lo tanto,

$$\begin{aligned} \mathcal{Z}^{-1} \{\ln \mathcal{A}(z)\}|_{n=0} &= \lim_{z \rightarrow \infty} \ln \mathcal{A}(z) \\ &= \ln \lim_{z \rightarrow \infty} \mathcal{A}(z) \\ &= \ln a[0] \\ &= \ln 1 \\ &= 0, \end{aligned}$$

concluyendo que

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \ln |\mathcal{A}(z)|^2 df = 0.$$

7.6.23. Problema 23

Encontrar el MLE asintótico de la potencia total P_0 de la PSD

$$P_{xx}(f) = P_0 Q(f)$$

donde

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} Q(f) df = 1.$$

Si $Q(f) = 1$ para todo f tratándose de un proceso WGN, simplificar los resultados. Sugerencia: emplear los resultados del problema de la sección 7.6.25 para la segunda parte.

Solución EL MLE asintótico para un proceso gaussiano fue deducido en la sección 7.4 y cumple la condición (ecuación 7.14)

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \left[\frac{1}{P_{xx}(f)} - \frac{I(f)}{P_{xx}^2(f)} \right] \frac{\partial P_{xx}(f)}{\partial \theta_i} df = 0,$$

donde

$$I(f) = \frac{1}{N} \left| \sum_{n=0}^{N-1} x[n] e^{-j2\pi f n} \right|^2$$

es el periodograma. Sustituyendo $P_{xx}(f)$ y considerando que el parámetro a estimar es P_0 , se tiene que

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \left[\frac{1}{\hat{P}_0 Q(f)} - \frac{I(f)}{\hat{P}_0^2 Q^2(f)} \right] Q(f) df = \int_{-\frac{1}{2}}^{\frac{1}{2}} \left[\frac{1}{\hat{P}_0} - \frac{I(f)}{\hat{P}_0^2 Q(f)} \right] df = \frac{1}{\hat{P}_0} - \frac{1}{\hat{P}_0^2} \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{I(f)}{Q(f)} df = 0,$$

concluyendo que el MLE es

$$\hat{P}_0 = \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{I(f)}{Q(f)} df.$$

En el caso en que $Q(f) = 1$ para todo f , el MLE queda

$$\begin{aligned}\hat{P}_0 &= \int_{\frac{1}{2}}^{\frac{1}{2}} I(f) df \\ &\stackrel{(a)}{=} \frac{1}{N} \int_{\frac{1}{2}}^{\frac{1}{2}} X'(f) X'^*(f) df \\ &= \frac{1}{N} \int_{\frac{1}{2}}^{\frac{1}{2}} |X'(f)|^2 df \\ &= \frac{1}{N} \sum_{n=-\infty}^{\infty} x'^2[n]\end{aligned}$$

donde en (a) se consideró que $I(f) = X'(f)X'^*(f)/N$ como se muestra en el problema de la sección 7.6.25 y en (b) se empleó la identidad de Parseval. Considerando la definición de $x'[n]$ en el problema 7.6.25, se concluye que

$$\hat{P}_0 = \frac{1}{N} \sum_{n=0}^{N-1} x^2[n].$$

7.6.24. Problema 24

En el ejemplo de la sección 7.5.1 se mostró que el pico del periodograma es el MLE de la frecuencia con la condición de que f_0 no esté cerca de 0 o $1/2$. Graficar el periodograma para $N = 10$ para las frecuencias $f_0 = 0.25$ y $f_0 = 0.05$ de los datos no ruidosos

$$x[n] = \cos 2\pi f_0 n, \quad n = 0, \dots, N-1.$$

¿Qué ocurre si la aproximación no es válida? Repetir el problema empleando la función exacta

$$J(f) = \mathbf{x}^T \mathbf{H} (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x}$$

donde \mathbf{H} está definido en el ejemplo de la sección 7.5.1.

Solución En el ejemplo de la sección 7.5.1 se mostró que el MLE de f_0 se obtiene maximizando la función

$$J(f) = \mathbf{x}^T \mathbf{H} (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x},$$

donde $\mathbf{H} = [\mathbf{c} \ \mathbf{s}]$ es una matriz $N \times 2$ con

$$\mathbf{c} = [1 \ \cos 2\pi f \ \dots \ \cos 2\pi f(N-1)]^T \quad \text{y} \quad \mathbf{s} = [0 \ \sin 2\pi f \ \dots \ \sin 2\pi f(N-1)]^T.$$

Además, empleando aproximaciones como la deducida en el ejercicio 3.6.7, se mostró que para f_0 no cercano a 0 o $1/2$, la función a maximizar se puede aproximar por el periodograma, es decir,

$$J(f) \approx I(f) = \frac{2}{N} \left| \sum_{n=0}^{N-1} x[n] e^{-j2\pi f n} \right|^2.$$

En la figura 7.10 se muestra la función $J(f)$ exacta y aproximada cuando $f_0 = 0.25$ y $f_0 = 0.05$. Se observa que para $f_0 = 0.25$ el periodograma es una buena aproximación de $J(f)$ y por lo tanto, el estimador MLE es bueno. Cuando $f_0 = 0.05$, que es cercano a 0, no se cumple la aproximación deducida en el ejercicio 3.6.7 y por lo tanto, no se cumple que $J(f) \approx I(f)$. En ese caso, el pico del periodograma está desplazado del valor verdadero. Esto se debe al solapamiento de los picos producidos por las sinusoides en $f_0 = 0.05$ y $f_0 = -0.05$, que no son bien resueltas en el periodograma debido a que el número de muestras $N = 10$ es pequeño.

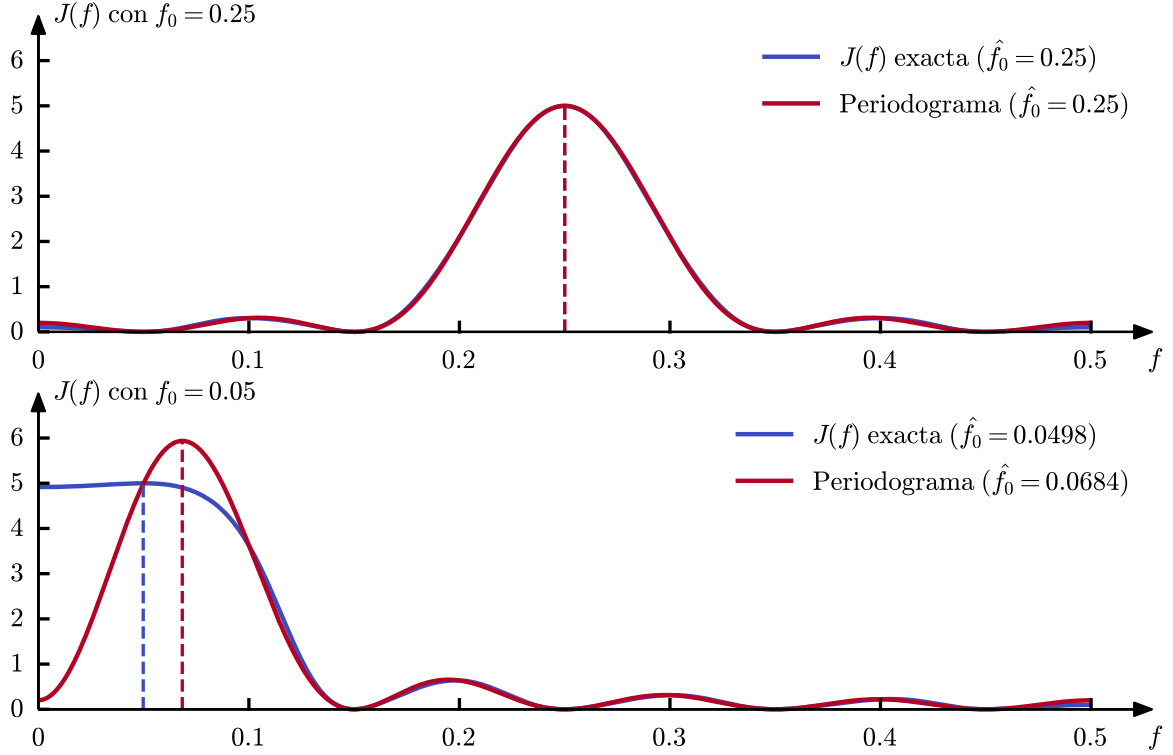


Figura 7.10: MLE de f_0 . Se muestra la función a maximizar $J(f)$ exacta y la aproximación por el periodograma $I(f)$ para los valores verdaderos $f_0 = 0.25$ y $f_0 = 0.05$. Se indica el máximo de la función en cada caso.

7.6.25. Problema 25

Probar que la transformada inversa de Fourier del periodograma es

$$\hat{r}_{xx}[k] = \begin{cases} \frac{1}{N} \sum_{n=0}^{N-1-|k|} x[n]x[n+|k|] & |k| \leq N-1 \\ 0 & |k| \geq N. \end{cases}$$

Sugerencia: notar que el periodograma puede escribirse como

$$I(f) = \frac{1}{N} X'(f) X'^*(f),$$

donde $X'(f)$ es la transformada de Fourier de la secuencia

$$x'[n] = \begin{cases} x[n] & 0 \leq n \leq N-1 \\ 0 & \text{en otro caso.} \end{cases}$$

Solución Por definición, el periodograma es

$$\begin{aligned} I(f) &= \frac{1}{N} \left| \sum_{n=0}^{N-1} x[n] e^{-j2\pi f n} \right|^2 \\ &= \frac{1}{N} \left| \sum_{n=-\infty}^{\infty} x'[n] e^{-j2\pi f n} \right|^2 \\ &= \frac{1}{N} X'(f) X'^*(f), \end{aligned}$$

donde $X'(f)$ es la transformada de Fourier de $x'[n]$, es decir,

$$x'[n] \xleftrightarrow{\mathcal{F}} X'(f) = \sum_{n=-\infty}^{\infty} x'[n] e^{-j2\pi f n}.$$

$I(f)$ puede escribirse como

$$\begin{aligned} I(f) &= \frac{1}{N} \left(\sum_{m=-\infty}^{\infty} x'[m] e^{-j2\pi f m} \right) \left(\sum_{n=-\infty}^{\infty} x'[n] e^{-j2\pi f n} \right)^* \\ &\stackrel{(a)}{=} \frac{1}{N} \left(\sum_{m=-\infty}^{\infty} x'[m] e^{-j2\pi f m} \right) \left(\sum_{n=-\infty}^{\infty} x'[n] e^{j2\pi f n} \right) \\ &= \frac{1}{N} \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} x'[m] x'[n] e^{-j2\pi f(m-n)} \\ &\stackrel{(b)}{=} \frac{1}{N} \sum_{l=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} x'[n+l] x'[n] e^{-j2\pi f l}, \end{aligned}$$

donde en (a) se consideró que $x'[n]$ es real y en (b) se realizó el cambio de variable $l = m - n$. Aplicando la transformada de Fourier inversa se tiene que

$$\begin{aligned} \mathcal{F}^{-1}\{I(f)\} &= \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{1}{N} \sum_{l=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} x'[n+l] x'[n] e^{-j2\pi f l} e^{j2\pi f k} df \\ &= \frac{1}{N} \sum_{l=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} x'[n+l] x'[n] \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{j2\pi f(k-l)} df \end{aligned}$$

Ahora,

$$\begin{aligned} \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{j2\pi f(k-l)} df &= \left. \frac{e^{j2\pi f(k-l)}}{j2\pi(k-l)} \right|_{-\frac{1}{2}}^{\frac{1}{2}} \\ &= \frac{e^{j\pi(k-l)} + e^{-j\pi(k-l)}}{j2\pi(k-l)} \\ &= \frac{\sin \pi(k-l)}{\pi(k-l)} \\ &\stackrel{(a)}{=} \delta[k-l]. \end{aligned}$$

donde en (a) se empleó la regla de L'Hôpital para evaluar $\sin \pi x / \pi x$ en $x = 0$,

$$\lim_{x \rightarrow 0} \frac{\sin \pi x}{\pi x} = \lim_{x \rightarrow 0} \frac{\pi \cos \pi x}{\pi} = 1.$$

Por lo tanto, la transformada inversa de Fourier del periodograma queda

$$\begin{aligned}\mathcal{F}^{-1}\{I(f)\} &= \frac{1}{N} \sum_{l=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} x'[n+l]x'[n]\delta[k-l] \\ &= \frac{1}{N} \sum_{n=-\infty}^{\infty} x'[n+k]x'[n].\end{aligned}$$

Considerando que

$$\begin{aligned}x'[n] &= x[n] \neq 0 & \text{en} & \quad n = [0, N-1] \\ x'[n+k] &= x[n+k] \neq 0 & \text{en} & \quad n = [-k, N-1-k]\end{aligned},$$

para $0 \leq k \leq N-1$ se tiene que

$$\mathcal{F}^{-1}\{I(f)\} = \frac{1}{N} \sum_{n=0}^{N-1-k} x[n+k]x[n]$$

y para $-(N-1) \leq k < 0$,

$$\begin{aligned}\mathcal{F}^{-1}\{I(f)\} &= \frac{1}{N} \sum_{n=-k}^{N-1} x[n+k]x[n] \\ &\stackrel{(a)}{=} \frac{1}{N} \sum_{m=0}^{N-1+k} x[m]x[m-k] \\ &\stackrel{(b)}{=} \frac{1}{N} \sum_{m=0}^{N-1-|k|} x[m]x[m+|k|]\end{aligned}$$

donde en (a) se realizó el cambio de variable $m = n+k$ y en (b) se tuvo en cuenta que como $k < 0$, $k = -|k|$. Combinando los resultados, se concluye que

$$\mathcal{F}^{-1}\{I(f)\} = \hat{r}_{xx}[k] = \begin{cases} \frac{1}{N} \sum_{n=0}^{N-1-|k|} x[n]x[n+|k|] & |k| \leq N-1 \\ 0 & |k| \geq N, \end{cases}$$

que es lo que se quería demostrar.

El resultado puede obtenerse de forma mas directa empleando propiedades de la transformada de Fourier. Como

$$I(f) = \frac{1}{N} X'(f)X'^*(f),$$

considerando que

$$x[n] \xleftrightarrow{\mathcal{F}} X(f) \quad \Rightarrow \quad x[-n] \xleftrightarrow{\mathcal{F}} X^*(f)$$

y que

$$\begin{aligned}x[n] &\xleftrightarrow{\mathcal{F}} X(f) \\ y[n] &\xleftrightarrow{\mathcal{F}} Y(f)\end{aligned} \quad \Rightarrow \quad x[n] * y[n] \xleftrightarrow{\mathcal{F}} X(f)Y(f)$$

se cumple que

$$x[n] * x[-n] \xleftrightarrow{\mathcal{F}} X(f)X^*(f).$$

Por lo tanto,

$$\mathcal{F}^{-1}\{I(f)\} = \frac{1}{N} x'[n] * x'[-n] = \frac{1}{N} \sum_{n=-\infty}^{\infty} x'[-n]x'[k-n] = \frac{1}{N} \sum_{m=-\infty}^{\infty} x'[m]x'[m+k].$$

7.6.26. Problema 26

La PSD de un proceso AR(1) está dada por

$$P_{xx}(f) = \frac{\sigma_u^2}{|1 + a[1]e^{-j2\pi f}|^2}.$$

Si $a[1]$ y σ^2 son desconocidos, encontrar el MLE de $P_{xx}(f)$ para $f = f_0$ empleando las ecuaciones 7.26 y 7.27. ¿Cual es la varianza asintótica del MLE? Sugerencia: emplear las ecuaciones 3.19 y 3.59.

Solución Para calcular el MLE de $P_{xx}(f_0)$, se calculará primero el MLE de $a[1]$ y σ^2 . Luego, el MLE de $P_{xx}(f_0)$ se obtiene inmediatamente gracias a la propiedad de invarianza ante transformaciones del MLE. De las ecuaciones 7.26 y 7.27 de Yule-Walker con $p = 1$ se tiene que

$$\begin{cases} \hat{r}_{xx}[0]\hat{a}[1] = -\hat{r}_{xx}[1] \\ \hat{\sigma}_u^2 = r_{xx}[0] + \hat{a}[1]\hat{r}_{xx}[1] \end{cases},$$

donde $\hat{r}_{xx}[k]$ es la ACF muestral de los datos $x[n]$, que está dada por la ecuación 7.25. Por lo tanto, de la primer ecuación se obtiene que

$$\hat{a}[1] = -\frac{\hat{r}_{xx}[1]}{\hat{r}_{xx}[0]}$$

y sustituyendo este resultado en la segunda ecuación, se obtiene que

$$\hat{\sigma}_u^2 = \hat{r}_{xx}[0] - \frac{\hat{r}_{xx}^2[1]}{\hat{r}_{xx}[0]} = \frac{\hat{r}_{xx}^2[0] - \hat{r}_{xx}^2[1]}{\hat{r}_{xx}[0]}.$$

Finalmente, por la propiedad de invarianza, el MLE de $P_{xx}(f_0)$ es

$$\hat{P}_{xx}(f_0) = \frac{\hat{\sigma}_u^2}{|1 + \hat{a}[1]e^{-j2\pi f_0}|^2}.$$

Para calcular la varianza asintótica del MLE, se considera que se cumple que el MLE es asintóticamente distribuido como

$$\hat{P}_{xx}(f_0) \stackrel{a}{\sim} \mathcal{N}(P_{xx}(f_0), \mathbf{I}^{-1}(P_{xx}(f_0))),$$

como se explicó en la sección 7.1.2. Además, como $P_{xx}(f_0)$ es una transformación del parámetro vectorial $\boldsymbol{\theta} = [a[1] \ \sigma_u^2]^T$, $\mathbf{I}^{-1}(P_{xx}(f_0))$ está dado por la ecuación 3.19,

$$\mathbf{I}^{-1}(P_{xx}(f_0)) = \frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathbf{I}^{-1}(\boldsymbol{\theta}) \frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T.$$

donde la matriz de información de Fisher $\mathbf{I}(\boldsymbol{\theta})$ fue calculada en el ejemplo del cálculo del MLE de los parámetros de un proceso AR de la sección 3.5.2 y está dada por la ecuación 3.59, que con $p = 1$ es

$$\mathbf{I}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{N}{\sigma_u^2} r_{xx}[0] & 0 \\ 0 & \frac{N}{2\sigma_u^4} \end{bmatrix} \Rightarrow \mathbf{I}^{-1}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\sigma_u^2}{Nr_{xx}[0]} & 0 \\ 0 & \frac{2\sigma_u^4}{N} \end{bmatrix}.$$

Además, la transformación g es en este caso

$$g(\boldsymbol{\theta}) = \frac{\sigma_u^2}{|1 + a[1]e^{-j2\pi f_0}|^2} = \frac{\sigma_u^2}{|A(f_0)|^2},$$

donde se definió $A(f_0) = 1 + a[1]e^{-j2\pi f_0}$. Considerando que

$$\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \frac{\partial g(\boldsymbol{\theta})}{\partial a[1]} & \frac{\partial g(\boldsymbol{\theta})}{\partial \sigma_u^2} \end{bmatrix},$$

se tiene que

$$\begin{aligned}
\frac{\partial g(\boldsymbol{\theta})}{\partial a[1]} &= -\frac{\sigma_u^2}{|A(f_0)|^4} \frac{\partial |A(f_0)|^2}{\partial a[1]} \\
&= -\frac{\sigma_u^2}{|A(f_0)|^4} \frac{\partial A(f_0)A^*(f_0)}{\partial a[1]} \\
&= -\frac{\sigma_u^2}{|A(f_0)|^4} \left[A(f_0) \frac{\partial A^*(f_0)}{\partial a[1]} + A^*(f_0) \frac{\partial A(f_0)}{\partial a[1]} \right] \\
&= -\frac{\sigma_u^2}{|A(f_0)|^4} [A(f_0)e^{j2\pi f_0} + A^*(f_0)e^{-j2\pi f_0}] \\
&= -\frac{2\sigma_u^2}{|A(f_0)|^4} \operatorname{Re} \{ A(f_0)e^{j2\pi f_0} \}
\end{aligned}$$

y como

$$\operatorname{Re} \{ A(f_0)e^{j2\pi f_0} \} = \operatorname{Re} \{ (1 + a[1]e^{-j2\pi f})e^{j2\pi f_0} \} = \operatorname{Re} \{ (e^{j2\pi f_0} + a[1]) \} = a[1] + \cos 2\pi f_0,$$

se obtiene que

$$\frac{\partial g(\boldsymbol{\theta})}{\partial a[1]} = -\frac{2\sigma_u^2(a[1] + \cos 2\pi f_0)}{|A(f)|^4}.$$

Por otro lado,

$$\frac{\partial g(\boldsymbol{\theta})}{\partial \sigma_u^2} = \frac{1}{|A(f_0)|^2},$$

concluyendo que

$$\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \begin{bmatrix} -\frac{2\sigma_u^2(a[1] + \cos 2\pi f_0)}{|A(f_0)|^4} & \frac{1}{|A(f_0)|^2} \end{bmatrix}.$$

Luego,

$$\mathbf{I}^{-1}(\boldsymbol{\theta}) \frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T = \begin{bmatrix} \frac{\sigma_u^2}{Nr_{xx}[0]} & 0 \\ 0 & \frac{2\sigma_u^4}{N} \end{bmatrix} \begin{bmatrix} -\frac{2\sigma_u^2(a[1] + \cos 2\pi f_0)}{|A(f_0)|^4} \\ \frac{1}{|A(f_0)|^2} \end{bmatrix} = \begin{bmatrix} -\frac{2\sigma_u^4(a[1] + \cos 2\pi f_0)}{Nr_{xx}[0]|A(f_0)|^4} \\ \frac{2\sigma_u^4}{N|A(f_0)|^2} \end{bmatrix},$$

y

$$\begin{aligned}
\frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathbf{I}^{-1}(\boldsymbol{\theta}) \frac{\partial g(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T &= \begin{bmatrix} -\frac{2\sigma_u^2(a[1] + \cos 2\pi f_0)}{|A(f_0)|^4} & \frac{1}{|A(f_0)|^2} \end{bmatrix} \begin{bmatrix} -\frac{2\sigma_u^4(a[1] + \cos 2\pi f_0)}{Nr_{xx}[0]|A(f_0)|^4} \\ \frac{2\sigma_u^4}{N|A(f_0)|^2} \end{bmatrix} \\
&= \frac{4\sigma_u^6(a[1] + \cos 2\pi f_0)^2}{Nr_{xx}[0]|A(f_0)|^8} + \frac{2\sigma_u^4}{N|A(f_0)|^4}.
\end{aligned}$$

Se concluye que

$$\mathbf{I}^{-1}(P_{xx}(f_0)) = \frac{4P_{xx}^4(f_0)(a[1] + \cos 2\pi f_0)^2}{N\sigma_u^2 r_{xx}[0]} + \frac{2P_{xx}^2(f_0)}{N} \quad (7.32)$$

es la varianza asintótica del MLE de $P_{xx}(f_0)$.

Se realizó una simulación por computadora que consistió en generar $N = 40$ y $N = 400$ muestras del proceso AR(1) con parámetros $a[1] = -0.9$ y $\sigma_u^2 = 1$, y se calculó el MLE $\hat{P}_{xx}(f)$ y su varianza $\operatorname{var}(\hat{P}_{xx}(f))$. En la figura 7.11 se muestran los resultados, en donde se grafica el promedio de $M = 10000$ realizaciones de $\hat{P}_{xx}(f)$ y $\operatorname{var}(\hat{P}_{xx}(f))$ junto con el valor verdadero de $P_{xx}(f)$ y la varianza asintótica $\operatorname{CRLB}(P_{xx}(f))$ dada por la ecuación 7.32, en donde $r_{xx}[0]$ se calcula como (ver el ejemplo de la sección 3.5.2)

$$r_{xx}[0] = \frac{\sigma_u^2}{1 - a^2[1]}.$$

Como es de esperar, se observa que cuando el largo N de los datos es pequeño, la varianza del estimador difiere considerablemente de la varianza asintótica teórica, pero se acerca cada vez mas al incrementar N .

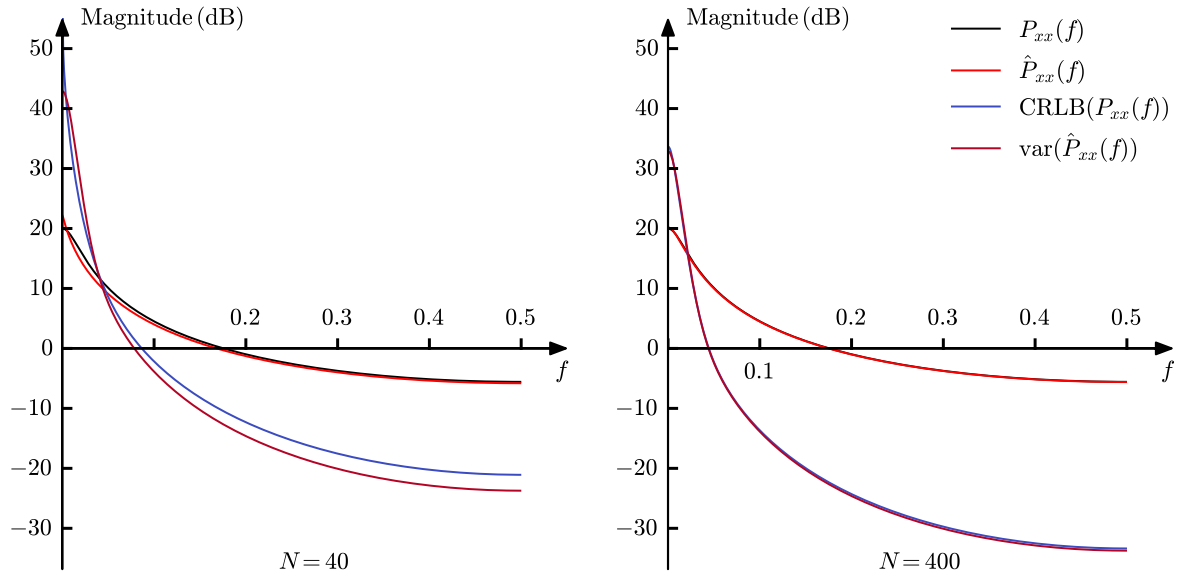


Figura 7.11: Comparación de la varianza asintótica teórica del MLE de $P_{xx}(f)$ con la varianza calculada en $M = 10000$ realizaciones empleando un largo de datos de $N = 40$ y $N = 400$ muestras.

Capítulo 8

Mínimos cuadrados

8.1. Enfoque de mínimos cuadrados

El enfoque empleado hasta ahora para considerar que un estimador es bueno, es que sea insesgado y tenga varianza mínima. Al elegir la varianza como medida de calidad, se busca minimizar la discrepancia entre el estimador y el valor verdadero del parámetro *en promedio*. En el enfoque de mínimos cuadrados (*least squares*, LS) se intenta minimizar la diferencia cuadrática entre los datos $x[n]$ *observados* y el modelo $s[n]$ asumido de los datos. La señal $s[n]$ es determinística y depende del parámetro θ desconocido que se quiere estimar. La diferencia entre los datos observados $x[n]$ y el modelo $s[n]$ se debe a ruido de observación o inexactitudes de modelado. El estimador de mínimos cuadrados (*least squares estimator*, LSE) de θ es aquel que minimiza el error cuadrático medio

$$J(\theta) = \sum_{n=0}^{N-1} (x[n] - s[n])^2.$$

En este enfoque no se asume ninguna hipótesis sobre las propiedades estadísticas de los datos $x[n]$. El LSE es generalmente empleado en situaciones donde no se dispone de una caracterización estadística precisa de los datos, pero también cuando es difícil o no es posible encontrar un estimador óptimo.

En el caso en que el modelo de señal es lineal con el parámetro a estimar, el error cuadrático es una función cuadrática con el parámetro, por lo que el mínimo se obtiene fácilmente y el problema de encontrar el LSE se denomina problema de mínimos cuadrados lineal. En el caso no lineal, en general no es posible obtener una expresión analítica del mínimo, y deben emplearse métodos iterativos o búsqueda de grilla, y el problema se denomina mínimos cuadrados no lineal.

8.2. Mínimos cuadrados lineales

8.2.1. Caso escalar

Para aplicar el enfoque de LS lineal, se asume que el modelo de señal es

$$s[n] = \theta h[n],$$

donde $h[n]$ es una señal conocida. La función del error LS es

$$J(\theta) = \sum_{n=0}^{N-1} (x[n] - \theta h[n])^2. \quad (8.1)$$

El LSE se obtiene diferenciando respecto al parámetro e igualando a cero. La derivada respecto al parámetro es

$$\frac{\partial J(\theta)}{\partial \theta} = -2 \sum_{n=0}^{N-1} (x[n] - \theta h[n]) h[n],$$

e igualando a cero y despejando el parámetro se obtiene que

$$\hat{\theta} = \frac{\sum_{n=0}^{N-1} x[n]h[n]}{\sum_{n=0}^{N-1} h^2[n]}. \quad (8.2)$$

El error mínimo se obtiene sustituyendo la ecuación 8.2 en la 8.1 y es

$$\begin{aligned} J_{\min} = J(\hat{\theta}) &= \sum_{n=0}^{N-1} (x[n] - \hat{\theta}h[n])(x[n] - \hat{\theta}h[n]) \\ &= \sum_{n=0}^{N-1} x[n](x[n] - \hat{\theta}h[n]) - \hat{\theta} \sum_{n=0}^{N-1} h[n](x[n] - \hat{\theta}h[n]) \\ &\stackrel{(a)}{=} \sum_{n=0}^{N-1} x^2[n] - \hat{\theta} \sum_{n=0}^{N-1} x[n]h[n], \end{aligned}$$

donde en (a) se consideró que la segunda sumatoria es nula. Efectivamente,

$$\begin{aligned} \sum_{n=0}^{N-1} h[n](x[n] - \hat{\theta}h[n]) &\stackrel{(a)}{=} \sum_{n=0}^{N-1} h[n] \left(x[n] - \frac{\sum_{n=0}^{N-1} x[n]h[n]}{\sum_{n=0}^{N-1} h^2[n]} h[n] \right) \\ &= \sum_{n=0}^{N-1} h[n]x[n] - \frac{\sum_{n=0}^{N-1} x[n]h[n]}{\sum_{n=0}^{N-1} h^2[n]} \sum_{n=0}^{N-1} h^2[n] \\ &= 0, \end{aligned}$$

donde en (a) se substituyó el LSE dado por la ecuación 8.2. Sustituyendo nuevamente la ecuación 8.2 el error mínimo se puede escribir como

$$J_{\min} = \sum_{n=0}^{N-1} x^2[n] - \frac{\left(\sum_{n=0}^{N-1} x[n]h[n] \right)^2}{\sum_{n=0}^{N-1} h^2[n]}. \quad (8.3)$$

En el problema de la sección 8.8.2 se muestra que se cumple que

$$0 \leq J_{\min} \leq \sum_{n=0}^{N-1} x^2[n]. \quad (8.4)$$

8.2.2. Caso vectorial

La extensión al caso en que el parámetro a estimar es un vector $\boldsymbol{\theta}$ de dimensiones $p \times 1$ es directa. El modelo de señal $\mathbf{s} = [s[0] \ s[1] \ \dots \ s[N-1]]^T$ lineal con el parámetro es

$$\mathbf{s} = \mathbf{H}\boldsymbol{\theta},$$

donde \mathbf{H} es una matriz conocida $N \times p$ con $N > p$, de rango completo p , denominada *matriz de observación*. El LSE se encuentra minimizando

$$J(\boldsymbol{\theta}) = \sum_{n=0}^{N-1} (x[n] - s[n])^2 = (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T (\mathbf{x} - \mathbf{H}\boldsymbol{\theta}).$$

Calculando el gradiente e igualando a cero se obtiene que el LSE es

$$\hat{\boldsymbol{\theta}} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x}. \quad (8.5)$$

Las ecuaciones $\mathbf{H}^T \mathbf{H} \hat{\boldsymbol{\theta}} = \mathbf{H}^T \mathbf{x}$ a resolver para calcular $\hat{\boldsymbol{\theta}}$ se llaman *ecuaciones normales*. La hipótesis del rango completo de \mathbf{H} asegura que $\mathbf{H}^T \mathbf{H}$ es invertible. En el problema de la sección 8.8.4 se deriva el resultado de una forma alternativa. Además, el error mínimo es

$$J_{\min} = J(\hat{\boldsymbol{\theta}}) = \mathbf{x}^T (\mathbf{I} - \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T) \mathbf{x}. \quad (8.6)$$

8.2.3. Interpretación geométrica

El modelo de la señal es

$$\mathbf{s} = \mathbf{H}\boldsymbol{\theta} = \begin{bmatrix} \mathbf{h}_1 & \mathbf{h}_2 & \dots & \mathbf{h}_p \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_p \end{bmatrix} = \sum_{i=1}^p \theta_i \mathbf{h}_i,$$

donde \mathbf{h}_i son las columnas de la matriz \mathbf{H} . Esto indica que el modelo de señal es una combinación lineal de las columnas de \mathbf{H} . Definiendo el largo de un vector $\boldsymbol{\xi} = [\xi_1 \ \xi_2 \ \dots \ \xi_N]^T$ como la norma euclídeana,

$$\|\boldsymbol{\xi}\| = \sqrt{\sum_{i=1}^N \xi_i^2} = \sqrt{\boldsymbol{\xi}^T \boldsymbol{\xi}}$$

el error LS puede expresarse como

$$J(\boldsymbol{\theta}) = (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T (\mathbf{x} - \mathbf{H}\boldsymbol{\theta}) = \|\mathbf{x} - \mathbf{H}\boldsymbol{\theta}\|^2 = \left\| \mathbf{x} - \sum_{i=1}^p \theta_i \mathbf{h}_i \right\|^2.$$

Se observa que el enfoque de LS lineal minimiza el cuadrado de la distancia desde el vector de datos \mathbf{x} al vector señal $\sum_{i=1}^p \theta_i \mathbf{h}_i$, que es una combinación lineal de las columnas de \mathbf{H} . El vector de datos pertenece a un espacio N -dimensional denominado R^N , mientras que todos los posibles vectores señal, por ser combinación lineal de $p < N$ vectores de dimensión N , pertenecen a un subespacio p -dimensional de R^N denominado S^p . La hipótesis de rango completo de \mathbf{H} asegura que las columnas de \mathbf{H} son linealmente independientes y por lo tanto generan un subespacio de dimensión exactamente p . El vector $\hat{\mathbf{s}} \in S^p$ con mínima distancia euclídeana a \mathbf{x} es la proyección ortogonal de \mathbf{x} en S^p . Esto significa que el vector de error $\boldsymbol{\epsilon} = \mathbf{x} - \hat{\mathbf{s}}$ debe ser ortogonal a S^p o equivalentemente, ortogonal a todas las columnas de \mathbf{H} ,

$$\boldsymbol{\epsilon}^T \mathbf{H} = \mathbf{0}^T. \quad (8.7)$$

Esta condición es conocida como *principio de ortogonalidad*. En la figura 8.1 se muestra un ejemplo en donde el vector de datos \mathbf{x} tiene dimensión $N = 3$ y por lo tanto pertenece a R^3 y el parámetro tiene dimensión $p = 2$ y de forma correspondiente, la matriz \mathbf{H} tiene dos columnas $[\mathbf{h}_1 \ \mathbf{h}_2]$ las cuales generan el subespacio S^2 .

La señal estimada, que es la proyección de \mathbf{x} en S^p , esta dada por

$$\hat{\mathbf{s}} = \mathbf{H}\hat{\boldsymbol{\theta}} = \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x} = \mathbf{P}\mathbf{x},$$

donde en la segunda identidad se empleó que el estimador está dado por la ecuación 8.5. La matriz $\mathbf{P} = \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T$ tiene dimensiones $N \times N$ y se llama *matriz de proyección ortogonal* o simplemente *matriz de proyección*. Tiene las siguientes características:

1. Es simétrica: $\mathbf{P}^T = \mathbf{P}$.
2. Es idempotente: $\mathbf{P}^2 = \mathbf{P}$.

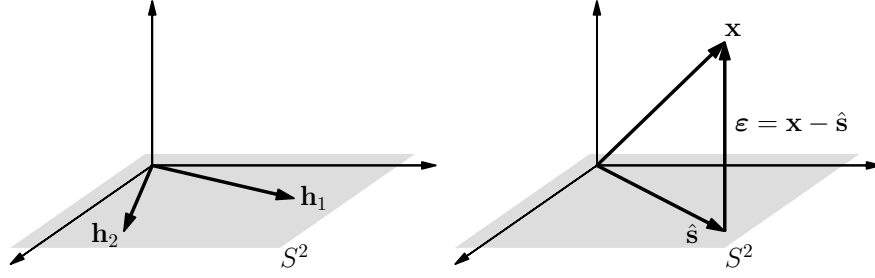


Figura 8.1: Interpretación geométrica del LSE. El vector de datos \mathbf{x} pertenece a R^3 y el modelo de señal tiene una matriz $\mathbf{H} = [\mathbf{h}_1 \ \mathbf{h}_2]$ con dos columnas de R^3 , las cuales generan el subespacio S^2 . El estimador $\hat{\mathbf{s}}$ es el vector de S^2 con distancia mínima al vector de datos \mathbf{x} y es la proyección ortogonal de \mathbf{x} en S^2 . Además, el error de estimación ϵ es el vector ortogonal a S^2 y es la proyección de \mathbf{x} en el subespacio ortogonal a S^2 .

La primer propiedad se muestra en el problema 8.8.11. El hecho de que debe ser idempotente se deduce de que \mathbf{P} aplicado a $\mathbf{P}\mathbf{x}$ conduce al mismo vector, ya que $\mathbf{P}\mathbf{x}$ ya pertenece a S^p . Además, la matriz de proyección es singular, ya que es de dimensión $N \times N$ y rango p , como se muestra en el problema 8.8.12. Si esto no fuera cierto, \mathbf{x} podría obtenerse a partir de $\hat{\mathbf{s}}$ mediante la inversión de \mathbf{P} , lo cual es imposible debido a que infinitos \mathbf{x} tienen la misma proyección $\hat{\mathbf{s}}$.

De la misma forma, el vector de error $\epsilon = \mathbf{x} - \hat{\mathbf{s}} = (\mathbf{I} - \mathbf{P})\mathbf{x}$ es la proyección de \mathbf{x} en el subespacio complemento o subespacio ortogonal al subespacio de la señal. Se puede verificar que $\mathbf{P}^\perp = \mathbf{I} - \mathbf{P}$ también es una matriz de proyección. De la ecuación 8.6 puede deducirse que el error mínimo es justamente

$$J_{\min} = \|\epsilon\|^2.$$

8.3. Mínimos cuadrados de orden recursivo

En muchos casos, el modelo de los datos es desconocido y debe asumirse. Por ejemplo, considérese un conjunto de datos $x[n]$ que varía suavemente en n , como se muestra en la figura 8.2. Si el modelo de los

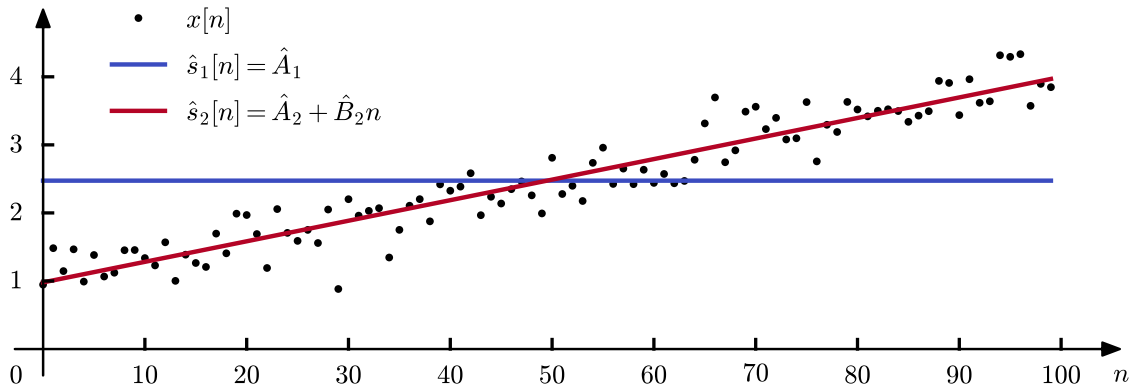


Figura 8.2: Datos y LSE del modelo con un parámetro (constante) y dos parámetros (lineal). Los datos se generaron como $s[n] = 1 + 0.03n + w[n]$, donde $w[n]$ es WGN con varianza $\sigma^2 = 0.1$.

datos es desconocido, se podría considerar un modelo constante, lineal, cuadrático o incluso de mayor

orden en n . Por ejemplo, los modelos constante y lineal son

$$\begin{aligned}s_1[n] &= A \\ s_2[n] &= A + Bn.\end{aligned}$$

Empleando el LSE con

$$\mathbf{H}_1 = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \quad \mathbf{H}_2 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & N-1 \end{bmatrix}$$

se obtienen los estimadores

$$\hat{A}_1 = \bar{x}$$

para el modelo constante y

$$\begin{aligned}\hat{A}_2 &= \frac{2(2N-1)}{N(N+1)} \sum_{n=0}^{N-1} x[n] - \frac{6}{N(N+1)} \sum_{n=0}^{N-1} nx[n] \\ \hat{B}_2 &= -\frac{6}{N(N+1)} \sum_{n=0}^{N-1} x[n] + \frac{12}{N(N^2-1)} \sum_{n=0}^{N-1} nx[n],\end{aligned} \tag{8.8}$$

para el modelo lineal, como se muestra en el problema de la sección 8.8.13. La figura 8.2 muestra que el modelo lineal con dos parámetros se ajusta mejor a los datos. Surge la duda de si debería agregarse un término cuadrático al modelo de los datos. Considerando que los datos están contaminados con ruido, puede ocurrir que al agregar mas parámetros el modelo se ajuste al ruido. Esta situación no es deseable, pero es inevitable si el modelo es desconocido. En la práctica, se elije el modelo mas simple que describa mejor a los datos. Considerando que el error LS mínimo se reduce al agregar mas parámetros, una estrategia podría ser incrementar la cantidad de parámetros hasta que el decrecimiento del error LS mínimo no sea significativo. En la figura 8.3 se muestra el error LS mínimo J_{\min} en función del

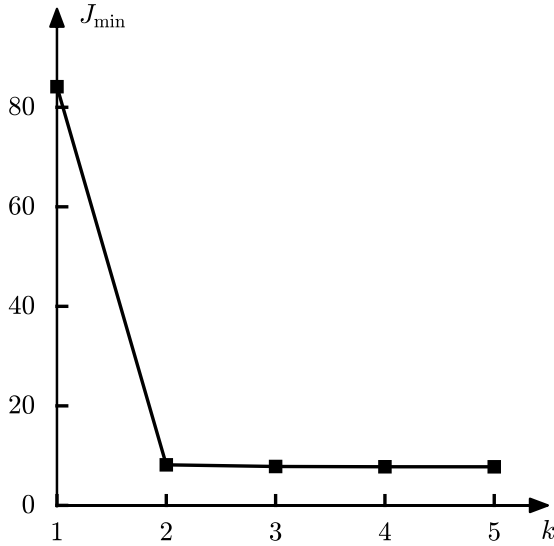


Figura 8.3: Efecto del número k de parámetros del modelo en el error LS mínimo J_{\min} . El error LS cae significativamente en $k = 2$. Al incrementar el número de parámetros mas allá de ese valor, la caída es pequeña, sugiriendo que el número de parámetros de la señal es efectivamente $k = 2$.

número k de parámetros del modelo, que es un polinomio en n de grado $k-1$. Los modelos $s_1[n]$ y $s_2[n]$ corresponden a $k=1$ y $k=2$. Se observa que un decrecimiento grande de J_{\min} ocurre con $k=2$. Para ordenes mayores, la caída del error LS mínimo no es significativa, indicando que se está modelando además el ruido. En este ejemplo, los datos efectivamente son lineales en n . En el caso en que A y B sean estimados de forma exacta, el error LS mínimo sería

$$J_{\min} = J(A, B) = J(\hat{A}, \hat{B}) = \sum_{n=0}^{N-1} (x[n] - s[n])^2 = \sum_{n=0}^{N-1} w^2[n] \approx N\sigma^2,$$

y teniendo en cuenta que en este ejemplo $N = 100$ y $\sigma^2 = 0.1$, resulta en $J_{\min} \approx 10$. Esto se verifica en la figura 8.3, y se confirma que el modelo elegido es el apropiado.

El ejemplo sugiere que es importante poder determinar el LSE para varios modelos de señal. La forma directa de hacerlo, es calcular el estimador para cada modelo empleando la ecuación 8.5. Sin embargo, la costo computacional puede reducirse empleando el enfoque LS de orden recursivo. Mediante este método, se actualiza el orden del modelo LSE en cada recursión. Específicamente, se calcula el LSE de un modelo con \mathbf{H} de dimensiones $N \times (k + 1)$ a partir del LSE del modelo con \mathbf{H} de dimensiones $N \times k$. A continuación se resume la técnica y se incluye una deducción algebraica.

8.3.1. Algoritmo LS de orden recursivo

Sea la matriz de observación \mathbf{H}_k de dimensiones $N \times k$ y el LSE $\boldsymbol{\theta}_k$ correspondiente,

$$\hat{\boldsymbol{\theta}}_k = (\mathbf{H}_k^T \mathbf{H}_k)^{-1} \mathbf{H}_k^T \mathbf{x}. \quad (8.9)$$

El error LS mínimo del modelo basado en \mathbf{H}_k es

$$J_{\min_k} = (\mathbf{x} - \mathbf{H}_k \hat{\boldsymbol{\theta}}_k)^T (\mathbf{x} - \mathbf{H}_k \hat{\boldsymbol{\theta}}_k). \quad (8.10)$$

El orden del modelo se incrementa a $k + 1$ agregando una columna \mathbf{h}_{k+1} a la matriz de observación. Esto genera la nueva matriz de observación

$$\mathbf{H}_{k+1} = [\mathbf{H}_k \quad \mathbf{h}_{k+1}] = [N \times k \quad N \times 1].$$

La actualizaciones de $\hat{\boldsymbol{\theta}}_{k+1}$ es

$$\hat{\boldsymbol{\theta}}_{k+1} = \begin{bmatrix} \hat{\boldsymbol{\theta}}_k - \frac{(\mathbf{H}_k^T \mathbf{H}_k)^{-1} \mathbf{H}_k^T \mathbf{h}_{k+1} \mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{x}}{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1}} \\ \frac{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{x}}{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1}} \end{bmatrix} = \begin{bmatrix} k \times 1 \\ 1 \times 1 \end{bmatrix} \quad (8.11)$$

donde

$$\mathbf{P}_k^\perp = \mathbf{I} - \mathbf{H}_k (\mathbf{H}_k^T \mathbf{H}_k)^{-1} \mathbf{H}_k^T \quad (8.12)$$

es la matriz de proyección sobre el subespacio ortogonal al generado por las columnas de \mathbf{H}_k . Para evitar invertir $\mathbf{H}_k^T \mathbf{H}_k$ se define

$$\mathbf{D}_k = (\mathbf{H}_k^T \mathbf{H}_k)^{-1} \quad (8.13)$$

y se emplea la fórmula recursiva

$$\mathbf{D}_{k+1} = \begin{bmatrix} \mathbf{D}_k + \frac{\mathbf{D}_k \mathbf{H}_k^T \mathbf{h}_{k+1} \mathbf{h}_{k+1}^T \mathbf{H}_k \mathbf{D}_k}{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1}} & -\frac{\mathbf{D}_k \mathbf{H}_k^T \mathbf{h}_{k+1}}{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1}} \\ -\frac{\mathbf{h}_{k+1}^T \mathbf{H}_k \mathbf{D}_k}{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1}} & \frac{1}{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1}} \end{bmatrix} = \begin{bmatrix} k \times k & k \times 1 \\ 1 \times k & 1 \times k \end{bmatrix}. \quad (8.14)$$

El error LS mínimo se actualiza como

$$J_{\min_{k+1}} = J_{\min_k} - \frac{(\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{x})^2}{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1}}. \quad (8.15)$$

El algoritmo no requiere la inversión de matrices. La recursión comienza calculando $\hat{\boldsymbol{\theta}}_1$, J_{\min_1} y \mathbf{D}_1 empleando las ecuaciones 8.9, 8.10 y 8.13 respectivamente. Las ecuaciones 8.11, 8.13 y 8.15 se denominan *método de mínimos cuadrados de orden recursivo*. En la figura 8.4 se muestra el código de una implementación del algoritmo. Ese código fue el empleado para obtener la gráfica de la figura 8.3.

```

import numpy as np
from numpy.linalg import inv, norm

# vector con el número de muestra: n = 0, ..., N-1
n = np.arange(N).reshape(-1, 1)
# número de paso
k = 1
# matriz de observación H en el paso k = 1
H = np.power(n, k - 1)
# matriz D en el paso k = 1
D = inv(H.T @ H)
# estimador en el paso k = 1
theta = D @ (H.T @ x)
# error LS mínimo en el paso k = 1
Jmin = norm(x - H @ theta, ord=2) ** 2
for k in np.arange(2, kmax + 1):
    # cálculo de P_perp en el paso k - 1
    P_perp = np.identity(N) - H @ D @ H.T
    # columna a agregar en H en el paso k:
    # vector con los elementos del vector n elevados a la k - 1
    h = np.power(n, k - 1)
    # denominador de los elementos del vector theta y la matriz D en el paso k
    den = h.T @ P_perp @ h
    ## cálculo del estimador en el paso k
    # elemento theta_12 del estimador: dimensiones (k - 1) x 1
    theta_11 = theta - (D @ H.T @ h @ h.T @ P_perp @ x) / den
    # elemento theta_22 del estimador: dimensiones 1 x 1
    theta_12 = (h.T @ P_perp @ x) / den
    # concatenación en filas para formar el estimador theta en el paso k
    theta = np.concatenate((theta_11, theta_12), axis=0)
    ## construcción de la matrix D en el paso k
    D_11 = D + (D @ H.T @ h @ h.T @ H @ D) / den
    D_21 = - (D @ H.T @ h) / den
    # concatenación para formar D
    D = np.concatenate((np.concatenate((D_11, D_21), axis=1),
                           np.concatenate((D_21.T, 1 / den), axis=1)), axis=0)

    # construcción de H en el paso k: [H h]
    H = np.concatenate((H, h), axis=1)
    ## cálculo del error mínimo en el paso k
    Jmin = Jmin - (h.T @ P_perp @ x) ** 2 / den

```

Figura 8.4: Implementación en el lenguaje Python del algoritmo LS de orden recursivo para el ejemplo del modelo de datos polinómico en n . En la iteración k -ésima se calcula el estimador y el error LS mínimo en función de las estimaciones en el paso $k - 1$. Para que la implementación sea genérica, basta cambiar la especificación de la matriz \mathbf{H}_1 en el paso inicial y la columna \mathbf{h}_k a agregar a la matriz \mathbf{H}_{k-1} en el paso k -ésimo.

Deducción del algoritmo Para deducir el algoritmo, se expresará el estimador, la matriz $\mathbf{D} = (\mathbf{H}^T \mathbf{H})^{-1}$ y el error LS mínimo en el paso $k + 1$ en función de sus valores en el paso k . El estimador en el paso $k + 1$ es

$$\hat{\boldsymbol{\theta}}_{k+1} = (\mathbf{H}_{k+1}^T \mathbf{H}_{k+1})^{-1} \mathbf{H}_{k+1}^T \mathbf{x}.$$

En el paso $k + 1$ se agrega la columna \mathbf{h}_{k+1} conocida a la matriz \mathbf{H}_k del paso k -ésimo,

$$\mathbf{H}_{k+1} = [\mathbf{H}_k \quad \mathbf{h}_{k+1}].$$

Por lo tanto,

$$\mathbf{H}_{k+1}^T \mathbf{H}_{k+1} = \begin{bmatrix} \mathbf{H}_k^T \\ \mathbf{h}_{k+1}^T \end{bmatrix} [\mathbf{H}_k \quad \mathbf{h}_{k+1}] = \begin{bmatrix} \mathbf{H}_k^T \mathbf{H}_k & \mathbf{H}_k^T \mathbf{h}_{k+1} \\ \mathbf{h}_{k+1}^T \mathbf{H}_k & \mathbf{h}_{k+1}^T \mathbf{h}_{k+1} \end{bmatrix} = \begin{bmatrix} k \times k & k \times 1 \\ 1 \times k & 1 \times 1 \end{bmatrix}.$$

Cómo se explica en el apéndice G, la matriz inversa de una matriz a bloques es

$$\begin{bmatrix} \mathbf{A} & \mathbf{b} \\ \mathbf{b}^T & c \end{bmatrix}^{-1} = \begin{bmatrix} \left(\mathbf{A} - \frac{\mathbf{b}\mathbf{b}^T}{c} \right)^{-1} & -\frac{1}{c} \left(\mathbf{A} - \frac{\mathbf{b}\mathbf{b}^T}{c} \right)^{-1} \mathbf{b} \\ -\frac{1}{c} \mathbf{b}^T \left(\mathbf{A} - \frac{\mathbf{b}\mathbf{b}^T}{c} \right)^{-1} & \frac{1}{c - \mathbf{b}^T \mathbf{A}^{-1} \mathbf{b}} \end{bmatrix},$$

donde se consideraron los elementos (1, 1), (1, 2) y (2, 1) de la matriz de la ecuación G.2 y el elemento (2, 2) de la matriz de la ecuación G.3, y por lo tanto,

$$(\mathbf{H}_{k+1}^T \mathbf{H}_{k+1})^{-1} = \begin{bmatrix} \mathbf{E}_k & -\frac{\mathbf{E}_k \mathbf{H}_{k+1}^T \mathbf{h}_{k+1}}{\mathbf{h}_{k+1}^T \mathbf{h}_{k+1}} \\ -\frac{\mathbf{h}_{k+1}^T \mathbf{H}_{k+1} \mathbf{E}_k}{\mathbf{h}_{k+1}^T \mathbf{h}_{k+1}} & \frac{1}{\mathbf{h}_{k+1}^T \mathbf{h}_{k+1} - \mathbf{h}_{k+1}^T \mathbf{H}_k^T (\mathbf{H}_k^T \mathbf{H}_k)^{-1} \mathbf{H}_k \mathbf{h}_{k+1}} \end{bmatrix},$$

donde se definió la matriz \mathbf{E}_k como la correspondiente al término $(\mathbf{A} - \mathbf{b}\mathbf{b}^T/c)^{-1}$, que es una matriz simétrica por ser \mathbf{A} simétrica. Empleando la identidad de Woodbury (ver la ecuación A.2 del apéndice A), que indica que

$$\left(\mathbf{A} - \frac{\mathbf{b}\mathbf{b}^T}{c} \right)^{-1} = \mathbf{A}^{-1} + \frac{\mathbf{A}^{-1} \mathbf{b} \mathbf{b}^T \mathbf{A}^{-1}}{c - \mathbf{b}^T \mathbf{A}^{-1} \mathbf{b}},$$

se tiene que

$$\begin{aligned} \mathbf{E}_k &= (\mathbf{H}_k^T \mathbf{H}_k)^{-1} + \frac{(\mathbf{H}_k^T \mathbf{H}_k)^{-1} \mathbf{H}_k^T \mathbf{h}_{k+1} \mathbf{h}_{k+1}^T \mathbf{H}_k (\mathbf{H}_k^T \mathbf{H}_k)^{-1}}{\mathbf{h}_{k+1}^T \mathbf{h}_{k+1} - \mathbf{h}_{k+1}^T \mathbf{H}_k (\mathbf{H}_k^T \mathbf{H}_k)^{-1} \mathbf{H}_k^T \mathbf{h}_{k+1}} \\ &= (\mathbf{H}_k^T \mathbf{H}_k)^{-1} + \frac{(\mathbf{H}_k^T \mathbf{H}_k)^{-1} \mathbf{H}_k^T \mathbf{h}_{k+1} \mathbf{h}_{k+1}^T \mathbf{H}_k (\mathbf{H}_k^T \mathbf{H}_k)^{-1}}{\mathbf{h}_{k+1}^T [\mathbf{I} - \mathbf{H}_k (\mathbf{H}_k^T \mathbf{H}_k)^{-1} \mathbf{H}_k^T] \mathbf{h}_{k+1}} \\ &= (\mathbf{H}_k^T \mathbf{H}_k)^{-1} + \frac{(\mathbf{H}_k^T \mathbf{H}_k)^{-1} \mathbf{H}_k^T \mathbf{h}_{k+1} \mathbf{h}_{k+1}^T \mathbf{H}_k (\mathbf{H}_k^T \mathbf{H}_k)^{-1}}{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1}}, \end{aligned}$$

en donde $\mathbf{P}_k^\perp = \mathbf{I} - \mathbf{P}_k = \mathbf{I} - \mathbf{H}_k (\mathbf{H}_k^T \mathbf{H}_k)^{-1} \mathbf{H}_k^T$. Notar que el denominador de \mathbf{E}_k es el mismo que el elemento (4, 4) de $(\mathbf{H}_{k+1}^T \mathbf{H}_{k+1})^{-1}$. Definiendo $\mathbf{D}_k = (\mathbf{H}_k^T \mathbf{H}_k)^{-1}$, se tiene que

$$\mathbf{D}_{k+1} = (\mathbf{H}_{k+1}^T \mathbf{H}_{k+1})^{-1} = \begin{bmatrix} \mathbf{E}_k & -\frac{\mathbf{E}_k \mathbf{H}_k^T \mathbf{h}_{k+1}}{\mathbf{h}_{k+1}^T \mathbf{h}_{k+1}} \\ -\frac{\mathbf{h}_{k+1}^T \mathbf{H}_k \mathbf{E}_k}{\mathbf{h}_{k+1}^T \mathbf{h}_{k+1}} & \frac{1}{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1}} \end{bmatrix}$$

con

$$\mathbf{E}_k = \mathbf{D}_k + \frac{\mathbf{D}_k \mathbf{H}_k^T \mathbf{h}_{k+1} \mathbf{h}_{k+1}^T \mathbf{H}_k \mathbf{D}_k}{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1}}.$$

Sustituyendo \mathbf{E}_k en los elementos de la anti-diagonal de \mathbf{D}_{k+1} , que son transpuestos, se ve que

$$\begin{aligned} \frac{\mathbf{E}_k \mathbf{H}_k^T \mathbf{h}_{k+1}}{\mathbf{h}_{k+1}^T \mathbf{h}_{k+1}} &= \frac{\mathbf{D}_k \mathbf{H}_k^T \mathbf{h}_{k+1}}{\mathbf{h}_{k+1}^T \mathbf{h}_{k+1}} + \frac{\mathbf{D}_k \mathbf{H}_k^T \mathbf{h}_{k+1} \mathbf{h}_{k+1}^T \mathbf{H}_k \mathbf{D}_k \mathbf{H}_k^T \mathbf{h}_{k+1}}{(\mathbf{h}_{k+1}^T \mathbf{h}_{k+1})(\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1})} \\ &= \frac{\mathbf{D}_k \mathbf{H}_k^T \mathbf{h}_{k+1} \mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1} + \mathbf{D}_k \mathbf{H}_k^T \mathbf{h}_{k+1} \mathbf{h}_{k+1}^T \mathbf{H}_k \mathbf{D}_k \mathbf{H}_k^T \mathbf{h}_{k+1}}{(\mathbf{h}_{k+1}^T \mathbf{h}_{k+1})(\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1})} \\ &= \frac{\mathbf{D}_k \mathbf{H}_k^T \mathbf{h}_{k+1} [\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1} + \mathbf{h}_{k+1}^T \mathbf{H}_k \mathbf{D}_k \mathbf{H}_k^T \mathbf{h}_{k+1}]}{(\mathbf{h}_{k+1}^T \mathbf{h}_{k+1})(\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1})} \\ &\stackrel{(a)}{=} \frac{\mathbf{D}_k \mathbf{H}_k^T \mathbf{h}_{k+1} [\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1} + \mathbf{h}_{k+1}^T \mathbf{P}_k \mathbf{h}_{k+1}]}{(\mathbf{h}_{k+1}^T \mathbf{h}_{k+1})(\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1})} \\ &= \frac{\mathbf{D}_k \mathbf{H}_k^T \mathbf{h}_{k+1} [\mathbf{h}_{k+1}^T (\mathbf{P}_k^\perp + \mathbf{P}_k) \mathbf{h}_{k+1}]}{(\mathbf{h}_{k+1}^T \mathbf{h}_{k+1})(\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1})} \\ &\stackrel{(b)}{=} \frac{\mathbf{D}_k \mathbf{H}_k^T \mathbf{h}_{k+1}}{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1}} \end{aligned}$$

donde en (a) se consideró que $\mathbf{P}_k = \mathbf{H}_k \mathbf{D}_k \mathbf{H}_k^T$ y en (b) que $\mathbf{P}_k^\perp + \mathbf{P}_k = \mathbf{I}$. Finalmente se tiene que

$$\mathbf{D}_{k+1} = \begin{bmatrix} \mathbf{D}_k + \frac{\mathbf{D}_k \mathbf{H}_k^T \mathbf{h}_{k+1} \mathbf{h}_{k+1}^T \mathbf{H}_k \mathbf{D}_k}{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1}} & -\frac{\mathbf{D}_k \mathbf{H}_k^T \mathbf{h}_{k+1}}{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1}} \\ -\frac{\mathbf{h}_{k+1}^T \mathbf{H}_k \mathbf{D}_k}{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1}} & \frac{1}{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1}} \end{bmatrix} = \begin{bmatrix} k \times k & k \times 1 \\ 1 \times k & 1 \times 1 \end{bmatrix}. \quad (8.16)$$

El estimador es

$$\begin{aligned} \hat{\boldsymbol{\theta}}_{k+1} &= \mathbf{D}_{k+1} \mathbf{H}_{k+1}^T \mathbf{x} \\ &= \mathbf{D}_{k+1} \begin{bmatrix} \mathbf{H}_k^T \mathbf{x} \\ \mathbf{h}_{k+1}^T \mathbf{x} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{D}_k \mathbf{H}_k^T \mathbf{x} + \frac{\mathbf{D}_k \mathbf{H}_k^T \mathbf{h}_{k+1} \mathbf{h}_{k+1}^T \mathbf{H}_k \mathbf{D}_k \mathbf{H}_k^T \mathbf{x}}{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1}} - \frac{\mathbf{D}_k \mathbf{H}_k^T \mathbf{h}_{k+1} \mathbf{h}_{k+1}^T \mathbf{x}}{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1}} \\ -\frac{\mathbf{h}_{k+1}^T \mathbf{H}_k \mathbf{D}_k \mathbf{H}_k^T \mathbf{x}}{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1}} + \frac{\mathbf{h}_{k+1}^T \mathbf{x}}{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1}} \end{bmatrix} \\ &= \begin{bmatrix} \hat{\boldsymbol{\theta}}_k - \frac{\mathbf{D}_k \mathbf{H}_k^T \mathbf{h}_{k+1} \mathbf{h}_{k+1}^T (\mathbf{I} - \mathbf{H}_k \mathbf{D}_k \mathbf{H}_k^T) \mathbf{x}}{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1}} \\ \frac{\mathbf{h}_{k+1}^T (\mathbf{I} - \mathbf{H}_k \mathbf{D}_k \mathbf{H}_k^T) \mathbf{x}}{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1}} \end{bmatrix} \\ &= \begin{bmatrix} \hat{\boldsymbol{\theta}}_k - \frac{\mathbf{D}_k \mathbf{H}_k^T \mathbf{h}_{k+1} \mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{x}}{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1}} \\ \frac{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{x}}{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1}} \end{bmatrix} \end{aligned}$$

Finalmente, la actualización del error LS mínimo es

$$\begin{aligned} J_{\min_{k+1}} &= (\mathbf{x} - \mathbf{H}_{k+1} \hat{\boldsymbol{\theta}}_{k-1})^T (\mathbf{x} - \mathbf{H}_{k+1} \hat{\boldsymbol{\theta}}_{k-1}) \\ &= \mathbf{x}^T (\mathbf{x} - \mathbf{H}_{k+1} \hat{\boldsymbol{\theta}}_{k-1}) - \hat{\boldsymbol{\theta}}_{k-1}^T \mathbf{H}_{k+1}^T (\mathbf{x} - \mathbf{H}_{k+1} \hat{\boldsymbol{\theta}}_{k-1}) \\ &\stackrel{(a)}{=} \mathbf{x}^T (\mathbf{x} - \mathbf{H}_{k+1} \hat{\boldsymbol{\theta}}_{k-1}) - \hat{\boldsymbol{\theta}}_{k-1}^T \mathbf{H}_{k+1}^T \boldsymbol{\epsilon}_{k+1} \\ &\stackrel{(b)}{=} \mathbf{x}^T \mathbf{x} - \mathbf{x}^T \mathbf{H}_{k+1} \hat{\boldsymbol{\theta}}_{k-1} \\ &= \mathbf{x}^T \mathbf{x} - \mathbf{x}^T \begin{bmatrix} \mathbf{H}_k & \mathbf{h}_{k+1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\theta}}_k - \frac{\mathbf{D}_k \mathbf{H}_k^T \mathbf{h}_{k+1} \mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{x}}{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1}} \\ \frac{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{x}}{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1}} \end{bmatrix} \\ &= \mathbf{x}^T \mathbf{x} - \mathbf{x}^T \mathbf{H}_k \hat{\boldsymbol{\theta}}_k + \frac{\mathbf{x}^T \mathbf{H}_k \mathbf{D}_k \mathbf{H}_k^T \mathbf{h}_{k+1} \mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{x}}{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1}} - \frac{\mathbf{x}^T \mathbf{h}_{k+1} \mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{x}}{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1}} \\ &\stackrel{(c)}{=} J_{\min_k} - \frac{\mathbf{x}^T (\mathbf{I} - \mathbf{H}_k \mathbf{D}_k \mathbf{H}_k^T) \mathbf{h}_{k+1} \mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{x}}{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1}} \\ &= J_{\min_k} - \frac{\mathbf{x}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1} \mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{x}}{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1}} \\ &= J_{\min_k} - \frac{(\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{x})^2}{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1}}, \end{aligned}$$

donde en (a), $\epsilon_{k+1} = \mathbf{x} - \mathbf{H}_{k+1}\hat{\boldsymbol{\theta}}_{k-1}$ es el error de estimación en el paso $k+1$, en (b) se consideró el principio de ortogonalidad dado por la ecuación 8.7, que indica que $\mathbf{H}_{k+1}^T \epsilon_{k+1} = \mathbf{0}$ y en (c) que $J_{\min_k} = \mathbf{x}^T \mathbf{x} - \mathbf{x}^T \mathbf{H}_k \hat{\boldsymbol{\theta}}_k$, como indica la identidad (b) para el paso $k+1$. En el problema 8.8.14 se deriva el resultado empleando un enfoque geométrico.

La matriz de proyección también puede ser actualizada de forma recursiva. Efectivamente, como

$$\mathbf{P}_{k+1} = \mathbf{H}_{k+1}(\mathbf{H}_{k+1}^T \mathbf{H}_{k+1})^{-1} \mathbf{H}_{k+1}^T = \mathbf{H}_{k+1} \mathbf{D}_{k+1} \mathbf{H}_{k+1}^T,$$

se puede emplear la ecuación de la actualización de \mathbf{D}_k dada por la ecuación 8.16. De esta forma,

$$\begin{aligned} \mathbf{P}_{k+1} &= [\mathbf{H}_k \quad \mathbf{h}_{k+1}] \begin{bmatrix} \mathbf{D}_k + \frac{\mathbf{D}_k \mathbf{H}_k^T \mathbf{h}_{k+1} \mathbf{h}_{k+1}^T \mathbf{H}_k \mathbf{D}_k}{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1}} & -\frac{\mathbf{D}_k \mathbf{H}_k^T \mathbf{h}_{k+1}}{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1}} \\ -\frac{\mathbf{h}_{k+1}^T \mathbf{H}_k \mathbf{D}_k}{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1}} & \frac{1}{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1}} \end{bmatrix} \begin{bmatrix} \mathbf{H}_k^T \\ \mathbf{h}_{k+1}^T \end{bmatrix} \\ &= [\mathbf{H}_k \quad \mathbf{h}_{k+1}] \begin{bmatrix} \mathbf{D}_k \mathbf{H}_k^T + \frac{\mathbf{D}_k \mathbf{H}_k^T \mathbf{h}_{k+1} \mathbf{h}_{k+1}^T \mathbf{H}_k \mathbf{D}_k \mathbf{H}_k^T}{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1}} - \frac{\mathbf{D}_k \mathbf{H}_k^T \mathbf{h}_{k+1} \mathbf{h}_{k+1}^T}{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1}} \\ -\frac{\mathbf{h}_{k+1}^T \mathbf{H}_k \mathbf{D}_k \mathbf{H}_k^T}{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1}} + \frac{\mathbf{h}_{k+1}^T}{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1}} \end{bmatrix} \\ &= \mathbf{H}_k \mathbf{D}_k \mathbf{H}_k^T + \frac{\mathbf{H}_k \mathbf{D}_k \mathbf{H}_k^T \mathbf{h}_{k+1} \mathbf{h}_{k+1}^T \mathbf{H}_k \mathbf{D}_k \mathbf{H}_k^T}{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1}} - \frac{\mathbf{H}_k \mathbf{D}_k \mathbf{H}_k^T \mathbf{h}_{k+1} \mathbf{h}_{k+1}^T}{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1}} \\ &\quad - \frac{\mathbf{h}_{k+1} \mathbf{h}_{k+1}^T \mathbf{H}_k \mathbf{D}_k \mathbf{H}_k^T}{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1}} + \frac{\mathbf{h}_{k+1} \mathbf{h}_{k+1}^T}{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1}} \\ &= \mathbf{P}_k + \frac{\mathbf{P}_k \mathbf{h}_{k+1} \mathbf{h}_{k+1}^T \mathbf{P}_k}{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1}} - \frac{\mathbf{P}_k \mathbf{h}_{k+1} \mathbf{h}_{k+1}^T}{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1}} - \frac{\mathbf{h}_{k+1} \mathbf{h}_{k+1}^T \mathbf{P}_k}{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1}} + \frac{\mathbf{h}_{k+1} \mathbf{h}_{k+1}^T}{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1}} \end{aligned}$$

resultando en

$$\mathbf{P}_{k+1} = \mathbf{P}_k + \frac{(\mathbf{I} - \mathbf{P}_k) \mathbf{h}_{k+1} \mathbf{h}_{k+1}^T (\mathbf{I} - \mathbf{P}_k)}{\mathbf{h}_{k+1}^T (\mathbf{I} - \mathbf{P}_k) \mathbf{h}_{k+1}}. \quad (8.17)$$

Esta ecuación se llama *matriz de proyección ortogonal recursiva*. En el problema de la sección 8.8.18 se emplea para deducir la ecuación recursiva del error LS mínimo.

8.4. Mínimos cuadrados secuencial

En las aplicaciones en donde hay que procesar muestras que se reciben una a una, es de interés poder actualizar el estimador cuando se observa una nueva muestra. Específicamente, supóngase que se determinó el LSE $\hat{\boldsymbol{\theta}}$ empleando las muestras $\{x[0], x[1], \dots, x[N-1]\}$. Si se observa la muestra $x[N]$, se desea actualizar $\hat{\boldsymbol{\theta}}$ en lugar de resolver la ecuación 8.5 empleando todas las muestras. Este procedimiento se denomina *mínimos cuadrados secuencial*.

8.4.1. Caso escalar

Ejemplo: nivel de DC en ruido no correlacionado

Se considera el problema de estimación del nivel de DC donde el ruido $w[n]$ es no correlacionado de varianza σ_n^2 . Aplicando el enfoque de LS ponderado (ver el problema de la sección 8.8.9), si la matriz de ponderación \mathbf{W} es diagonal con $[\mathbf{W}]_{ii} = 1/\sigma_i^2$, el LSE ponderado, como se calculó en el problema de la sección 8.8.8 es

$$\hat{A}[N-1] = \frac{\sum_{n=0}^{N-1} \frac{x[n]}{\sigma_n^2}}{\sum_{n=0}^{N-1} \frac{1}{\sigma_n^2}},$$

donde el argumento de \hat{A} indica el índice de la última muestra observada en el calculo del estimador. Sea $\hat{A}[N]$ el estimador incluyendo una nueva muestra $x[N]$. Para encontrar el LSE secuencial, se observa que

$$\begin{aligned}
\hat{A}[N] &= \frac{\sum_{n=0}^N \frac{x[n]}{\sigma_n^2}}{\sum_{n=0}^N \frac{1}{\sigma_n^2}} \\
&= \frac{\sum_{n=0}^{N-1} \frac{x[n]}{\sigma_n^2} + \frac{x[N]}{\sigma_N^2}}{\sum_{n=0}^{N-1} \frac{1}{\sigma_n^2}} \\
&= \frac{\left(\sum_{n=0}^{N-1} \frac{1}{\sigma_n^2}\right) \hat{A}[N-1] + \frac{x[N]}{\sigma_N^2}}{\sum_{n=0}^N \frac{1}{\sigma_n^2}} + \frac{\frac{x[N]}{\sigma_N^2}}{\sum_{n=0}^N \frac{1}{\sigma_n^2}} \\
&= \frac{\left(\sum_{n=0}^N \frac{1}{\sigma_n^2} - \frac{1}{\sigma_N^2}\right) \hat{A}[N-1] + \frac{x[N]}{\sigma_N^2}}{\sum_{n=0}^N \frac{1}{\sigma_n^2}} + \frac{\frac{x[N]}{\sigma_N^2}}{\sum_{n=0}^N \frac{1}{\sigma_n^2}} \\
&= \hat{A}[N-1] - \frac{\frac{1}{\sigma_N^2}}{\sum_{n=0}^N \frac{1}{\sigma_n^2}} + \frac{\frac{x[N]}{\sigma_N^2}}{\sum_{n=0}^N \frac{1}{\sigma_n^2}}
\end{aligned}$$

resultando en

$$\hat{A}[N] = \hat{A}[N-1] - \frac{\frac{1}{\sigma_N^2}}{\sum_{n=0}^N \frac{1}{\sigma_n^2}} (x[N] - \hat{A}[N-1])$$

El nuevo estimador es igual al estimador previo mas un *término de corrección*. El factor de ganancia que multiplica al término de corrección depende de la confianza en la nueva muestra. Si la nueva muestra es ruidosa o $\sigma_N^2 \rightarrow \infty$, el estimador previo no se corrige. En caso contrario, si la nueva muestra no tiene ruido o $\sigma_N^2 \rightarrow 0$, $\hat{A}[N] \rightarrow x[N]$, lo que significa que las muestras previas se descartan. Por lo tanto, el factor de ganancia representa la confianza en la nueva muestra relativa a la confianza de las muestras previas. Continuando con la interpretación del resultado, se sabe por el problema de la sección 8.8.8 que

$$\text{var}(\hat{A}[N-1]) = \frac{1}{\sum_{n=0}^{N-1} \frac{1}{\sigma_n^2}}.$$

y el factor de ganancia de la corrección N -ésima es

$$\begin{aligned}
K[N] &= \frac{\frac{1}{\sigma_N^2}}{\sum_{n=0}^N \frac{1}{\sigma_n^2}} \\
&= \frac{\frac{1}{\sigma_N^2}}{\frac{1}{\sigma_N^2} + \sum_{n=0}^{N-1} \frac{1}{\sigma_n^2}} \\
&= \frac{\frac{1}{\sigma_N^2}}{\frac{1}{\sigma_N^2} + \frac{1}{\text{var}(\hat{A}[N-1])}}
\end{aligned}$$

resultando en

$$K[N] = \frac{\text{var}(\hat{A}[N-1])}{\text{var}(\hat{A}[N-1]) + \sigma_N^2}.$$

Como $0 \leq K[N] \leq 1$, la corrección es grande si $K[N]$ es grande, cosa que ocurre cuando $\text{var}(\hat{A}[N-1])$ es grande. Asimismo, si la varianza del estimador previo es pequeña, también lo es la corrección. El hecho de que la ganancia dependa de $\text{var}(\hat{A}[N-1])$ puede emplearse para encontrar una expresión recursiva, ya que

$$\begin{aligned}
\text{var}(\hat{A}[N]) &= \frac{1}{\sum_{n=0}^N \frac{1}{\sigma_n^2}} \\
&= \frac{1}{\frac{1}{\sigma_N^2} + \sum_{n=0}^{N-1} \frac{1}{\sigma_n^2}} \\
&= \frac{1}{\frac{1}{\sigma_N^2} + \frac{1}{\text{var}(\hat{A}[N-1])}} \\
&= \frac{\text{var}(\hat{A}[N-1])\sigma_N^2}{\text{var}(\hat{A}[N-1]) + \sigma_N^2} \\
&= \frac{\text{var}(\hat{A}[N-1]) \left[\sigma_N^2 + \text{var}(\hat{A}[N-1]) - \text{var}(\hat{A}[N-1]) \right]}{\text{var}(\hat{A}[N-1]) + \sigma_N^2} \\
&= \left(1 - \frac{\text{var}(\hat{A}[N-1])}{\text{var}(\hat{A}[N-1]) + \sigma_N^2} \right) \text{var}(\hat{A}[N-1]) \\
&= (1 - K[N]) \text{var}(\hat{A}[N-1]).
\end{aligned}$$

Para calcular la ganancia de forma recursiva, se procede como sigue:

Actualización del estimador:

$$\hat{A}[N] = \hat{A}[N-1] + K[N](x[N] - \hat{A}[N-1]) \quad (8.18)$$

donde

$$K[N] = \frac{\text{var}(\hat{A}[N-1])}{\text{var}(\hat{A}[N-1]) + \sigma_N^2} \quad (8.19)$$

Actualización de la varianza:

$$\text{var}(\hat{A}[N]) = (1 - K[N]) \text{var}(\hat{A}[N-1]). \quad (8.20)$$

Para comenzar la recursión, se parte de

$$\begin{aligned} \hat{A}[0] &= x[0] \\ \text{var}(\hat{A}[0]) &= \sigma_0^2. \end{aligned}$$

Luego se calcula $K[1]$ con la ecuación 8.19 y $\hat{A}[1]$ con la ecuación 8.18. Después se calcula $\text{var}(\hat{A}[1])$ con la ecuación 8.20. Continuando de esta forma, se calcula $K[2]$, $\hat{A}[2]$ y $\text{var}(\hat{A}[2])$ y así sucesivamente. Observar que en el proceso de calcular la ganancia de forma recursiva, también se obtiene una recursión para la varianza del LSE. Se puede mostrar que el error LS mínimo también puede calcularse recursivamente como

$$J_{\min}[N] = J_{\min}[N-1] + \frac{\left(x[N] - \hat{A}[N-1]\right)^2}{\text{var}(\hat{A}[N-1]) + \sigma_N^2}, \quad (8.21)$$

como se hace en el problema de la sección 8.8.19.

8.4.2. Caso vectorial

A continuación se generaliza el cálculo del LSE secuencial para el caso en que el parámetro es vectorial. Considérese la minimización de la función J del error LS ponderado (ver el problema de la sección 8.8.9),

$$J = (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T \mathbf{C}^{-1} (\mathbf{x} - \mathbf{H}\boldsymbol{\theta}), \quad (8.22)$$

con $\mathbf{W} = \mathbf{C}^{-1}$, donde \mathbf{C} es la matriz de covarianza del ruido de media nula. Las hipótesis del problema son las mismas que el BLUE (ver el problema de la sección 6.4.13). Como se encontró en el problema de la sección 8.8.9 (ver también la ecuación 6.5), el LSE ponderado es

$$\hat{\boldsymbol{\theta}} = (\mathbf{H}^T \mathbf{C}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C}^{-1} \mathbf{x}$$

con matriz de covarianza

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}} = (\mathbf{H}^T \mathbf{C}^{-1} \mathbf{H})^{-1}.$$

dada por la ecuación 6.7). Solo si \mathbf{C} es diagonal, es decir, el ruido es no correlacionado, $\hat{\boldsymbol{\theta}}$ puede ser calculado secuencialmente en el tiempo. Asumiendo que esto es así, sea

$$\begin{aligned} \mathbf{C}[n] &= \text{diag}(\sigma_0^2, \sigma_1^2, \dots, \sigma_n^2) \\ \mathbf{H}[n] &= \begin{bmatrix} \mathbf{H}[n-1] \\ \mathbf{h}^T[n] \end{bmatrix} = \begin{bmatrix} n \times p \\ 1 \times p \end{bmatrix} \\ \mathbf{x}[n] &= [\mathbf{x}[0] \ \mathbf{x}[1] \ \dots \ \mathbf{x}[n]] \end{aligned}$$

y $\hat{\boldsymbol{\theta}}$ el LSE ponderado de $\boldsymbol{\theta}$ calculado a partir de $\mathbf{x}[n]$ o $(n+1)$ muestras. El estimador del lote de datos (*batch estimator*) es

$$\hat{\boldsymbol{\theta}}[n] = (\mathbf{H}^T[n] \mathbf{C}^{-1}[n] \mathbf{H}[n])^{-1} \mathbf{H}^T[n] \mathbf{C}^{-1}[n] \mathbf{x}[n] \quad (8.23)$$

con matriz de covarianza

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}} = \boldsymbol{\Sigma}[n] = (\mathbf{H}^T[n] \mathbf{C}^{-1}[n] \mathbf{H}[n])^{-1}. \quad (8.24)$$

Notar que $\mathbf{C}[n]$ es la matriz de covarianza del ruido mientras que $\boldsymbol{\Sigma}[n]$ es la matriz de covarianza del LSE. El estimador secuencial es

Actualización del estimador:

$$\hat{\boldsymbol{\theta}}[n] = \hat{\boldsymbol{\theta}}[n-1] + \mathbf{K}[n] \left(x[n] - \mathbf{h}^T[n] \hat{\boldsymbol{\theta}}[n-1] \right) \quad (8.25)$$

donde

$$\mathbf{K}[n] = \frac{\boldsymbol{\Sigma}[n-1] \mathbf{h}[n]}{\sigma_n^2 + \mathbf{h}^T[n] \boldsymbol{\Sigma}[n-1] \mathbf{h}[n]} \quad (8.26)$$

Actualización de la matriz de covarianza:

$$\Sigma[n] = (\mathbf{I} - \mathbf{K}[n]\mathbf{h}^T[n])\Sigma[n-1] \quad (8.27)$$

El factor de ganancia $\mathbf{K}[n]$ es un vector $p \times 1$ y la matriz de covarianza $\Sigma[n]$ tiene dimensiones $p \times p$. En la figura 8.5 se resume el algoritmo. Para comenzar la recursión es necesario especificar valores iniciales de $\hat{\boldsymbol{\theta}}[n-1]$ y $\Sigma[n-1]$ de forma de calcular $\mathbf{K}[n]$ con la ecuación 8.26 y luego $\hat{\boldsymbol{\theta}}[n]$ y $\Sigma[n]$ con las ecuaciones 8.25 y 8.27 respectivamente. Una forma de determinar los valores iniciales $\hat{\boldsymbol{\theta}}[n-1]$ y $\Sigma[n-1]$ es mediante las ecuaciones por lotes 8.23 y 8.24, lo que requiere que $\mathbf{H}^T[n-1]\mathbf{C}^{-1}[n-1]\mathbf{H}[n-1]$ sea invertible. Para que esto suceda, $\mathbf{H}[n-1]$ debe tener rango mayor o igual a p . Como $\mathbf{H}[n-1]$ tiene dimensiones $n \times p$, se debe cumplir que $n \geq p$, asumiendo que las p columnas son linealmente independientes. Por lo tanto, el procedimiento para el cálculo de LS secuencial típicamente comienza determinando $\hat{\boldsymbol{\theta}}[p-1]$ y $\Sigma[p-1]$ empleando el estimador de lotes de las ecuaciones 8.23 y 8.24 (ya que $\mathbf{H}[p-1]$ tiene dimensiones $p \times p$) y luego se emplean las ecuaciones secuenciales 8.25, 8.26 y 8.27 para $n \geq p$. Un segundo método de inicializar la recursión es asignar valores a $\hat{\boldsymbol{\theta}}[-1]$ y $\Sigma[-1]$ y calcular el LS secuencial para $n \geq 0$. Esto tiene el efecto de sesgar el estimador hacia $\hat{\boldsymbol{\theta}}[-1]$. La forma usual de minimizar el efecto del sesgo es elegir $\Sigma[-1]$ grande o $\Sigma[-1] = \alpha\mathbf{I}$ con α grande, los que implica dar poca confianza a $\Sigma[-1]$. El LSE para $n \geq p$ resulta en el mismo estimador que al emplear el estimador por lotes si $\alpha \rightarrow \infty$, como se muestra en el problema de la sección 8.8.23. Finalmente, el error LS mínimo también puede ser calculado secuencialmente como

$$J_{\min}[n] = J_{\min}[n-1] + \frac{(x[n] - \mathbf{h}^T[n]\hat{\boldsymbol{\theta}}[n-1])^2}{\sigma_n^2 + \mathbf{h}^T[n]\Sigma[n-1]\mathbf{h}[n]}. \quad (8.28)$$

El algoritmo de mínimos cuadrados secuencial es usualmente referido en la literatura como *algoritmo de mínimos cuadrados recursivo* (*recursive least squares*, RLS), como por ejemplo en el capítulo 10 de [4]. A continuación se incluye la deducción de las ecuaciones 8.25, 8.26, 8.27 y 8.28 del algoritmo LS secuencial.

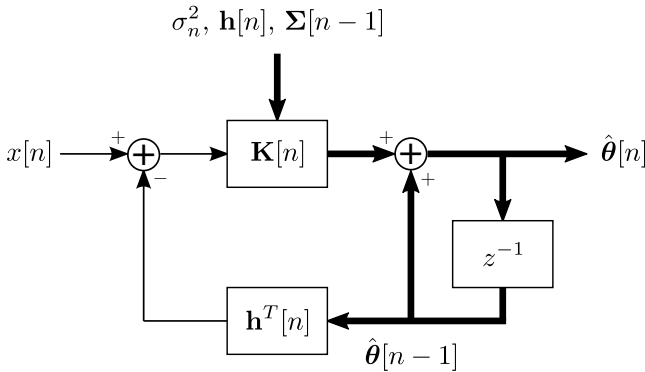


Figura 8.5: Estimador por mínimos cuadrados secuencial. Las flechas gruesas indican procesamiento de vectores.

Deducción del algoritmo LS secuencial El objetivo es expresar el estimador por lotes en el tiempo n en función del estimador en el tiempo $n-1$. Para hacerlo, se parte de la matriz de covarianza del estimador $\hat{\boldsymbol{\theta}}[n]$, dada por la ecuación 8.24, y se opera como sigue

$$\begin{aligned} \Sigma[n] &= (\mathbf{H}^T[n]\mathbf{C}^{-1}[n]\mathbf{H}[n])^{-1} \\ &= \left([\mathbf{H}^T[n-1] \quad \mathbf{h}^T[n]] \begin{bmatrix} \mathbf{C}^{-1}[n-1] & \mathbf{0} \\ \mathbf{0}^T & \sigma_n^2 \end{bmatrix} \begin{bmatrix} \mathbf{H}[n-1] \\ \mathbf{h}[n] \end{bmatrix} \right)^{-1} \\ &\stackrel{(a)}{=} \left([\mathbf{H}^T[n-1] \quad \mathbf{h}^T[n]] \begin{bmatrix} \mathbf{C}^{-1}[n-1] & \mathbf{0} \\ \mathbf{0}^T & \frac{1}{\sigma_n^2} \end{bmatrix} \begin{bmatrix} \mathbf{H}[n-1] \\ \mathbf{h}[n] \end{bmatrix} \right)^{-1} \\ &\stackrel{(b)}{=} \left(\mathbf{H}^T[n-1]\mathbf{C}^{-1}[n-1]\mathbf{H}[n-1] + \frac{1}{\sigma_n^2}\mathbf{h}[n]\mathbf{h}^T[n] \right)^{-1} \\ &= \left(\Sigma^{-1}[n-1] + \frac{1}{\sigma_n^2}\mathbf{h}[n]\mathbf{h}^T[n] \right)^{-1} \end{aligned}$$

donde en (a) se invirtió la matriz de covarianza del ruido teniendo en cuenta que es diagonal y en (b) se realizaron las operaciones entre matrices. Empleando la identidad de Woodbury, dada por la ecuación A.2 deducida en el apéndice A, se tiene que

$$\begin{aligned}\Sigma[n] &= \Sigma[n-1] - \frac{\Sigma[n-1]\mathbf{h}[n]\mathbf{h}^T[n]\Sigma[n-1]}{\sigma_n^2 + \mathbf{h}^T[n]\Sigma[n-1]\mathbf{h}[n]} \\ &= \left(\mathbf{I} - \frac{\Sigma[n-1]\mathbf{h}[n]\mathbf{h}^T[n]}{\sigma_n^2 + \mathbf{h}^T[n]\Sigma[n-1]\mathbf{h}[n]} \right) \Sigma[n-1].\end{aligned}$$

Definiendo

$$\mathbf{K}[n] = \frac{\Sigma[n-1]\mathbf{h}[n]}{\sigma_n^2 + \mathbf{h}^T[n]\Sigma[n-1]\mathbf{h}[n]}$$

se obtiene que

$$\Sigma[n] = (\mathbf{I} - \mathbf{K}[n]\mathbf{h}^T[n]) \Sigma[n-1],$$

que son respectivamente las ecuaciones 8.26 y 8.24 del algoritmo LS secuencial. Para encontrar la recursión del estimador, se parte del estimador por lotes, que está dado por la ecuación 8.23 y operando se ve que

$$\begin{aligned}\hat{\boldsymbol{\theta}}[n] &= (\mathbf{H}^T[n]\mathbf{C}^{-1}[n]\mathbf{H}[n])^{-1}\mathbf{H}^T[n]\mathbf{C}^{-1}[n]\mathbf{x}[n] \\ &= \Sigma[n] \left(\begin{bmatrix} \mathbf{H}^T[n-1] & \mathbf{h}[n] \end{bmatrix} \begin{bmatrix} \mathbf{C}^{-1}[n-1] & \mathbf{0} \\ \mathbf{0}^T & \sigma_n^2 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{x}[n-1] \\ x[n] \end{bmatrix} \right) \\ &\stackrel{(a)}{=} (\mathbf{I} - \mathbf{K}[n]\mathbf{h}^T[n]) \Sigma[n-1] \left(\begin{bmatrix} \mathbf{H}^T[n-1] & \mathbf{h}[n] \end{bmatrix} \begin{bmatrix} \mathbf{C}^{-1}[n-1] & \mathbf{0} \\ \mathbf{0}^T & \frac{1}{\sigma_n^2} \end{bmatrix} \begin{bmatrix} \mathbf{x}[n-1] \\ x[n] \end{bmatrix} \right) \\ &\stackrel{(b)}{=} (\mathbf{I} - \mathbf{K}[n]\mathbf{h}^T[n]) \Sigma[n-1] \left(\mathbf{H}^T[n-1]\mathbf{C}^{-1}[n-1]\mathbf{x}[n-1] + \frac{1}{\sigma_n^2}\mathbf{h}[n]x[n] \right),\end{aligned}$$

donde en (a) se invirtió la matriz de covarianza del ruido y se substituyó $\Sigma[n]$ por el resultado obtenido previamente y en (b) se realizaron las operaciones entre matrices. Se encontrará una expresión para el término $\mathbf{H}^T[n-1]\mathbf{C}^{-1}[n-1]\mathbf{x}[n-1]$ considerando al estimador por lotes en el tiempo $n-1$, que es

$$\begin{aligned}\hat{\boldsymbol{\theta}}[n-1] &= (\mathbf{H}^T[n-1]\mathbf{C}^{-1}[n-1]\mathbf{H}[n-1])^{-1}\mathbf{H}^T[n-1]\mathbf{C}^{-1}[n-1]\mathbf{x}[n-1] \\ &= \Sigma[n-1]\mathbf{H}^T[n-1]\mathbf{C}^{-1}[n-1]\mathbf{x}[n-1].\end{aligned}$$

Por lo tanto,

$$\mathbf{H}^T[n-1]\mathbf{C}^{-1}[n-1]\mathbf{x}[n-1] = \Sigma^{-1}[n-1]\hat{\boldsymbol{\theta}}[n-1], \quad (8.29)$$

obteniendo que

$$\hat{\boldsymbol{\theta}}[n] = (\mathbf{I} - \mathbf{K}[n]\mathbf{h}^T[n]) \Sigma[n-1] \left(\Sigma^{-1}[n-1]\hat{\boldsymbol{\theta}}[n-1] + \frac{1}{\sigma_n^2}\mathbf{h}[n]x[n] \right),$$

y desarrollando esta ecuación, se tiene que

$$\hat{\boldsymbol{\theta}}[n] = \hat{\boldsymbol{\theta}}[n-1] + \frac{1}{\sigma_n^2}\Sigma[n-1]\mathbf{h}[n]x[n] - \mathbf{K}[n]\mathbf{h}^T[n]\hat{\boldsymbol{\theta}}[n-1] - \frac{1}{\sigma_n^2}\mathbf{K}[n]\mathbf{h}^T[n]\Sigma[n-1]\mathbf{h}[n]x[n].$$

De la definición de $\mathbf{K}[n]$, se observa que el término $\Sigma[n-1]\mathbf{h}[n]$ en el segundo sumando puede expresarse como

$$\Sigma[n-1]\mathbf{h}[n] = \mathbf{K}[n] (\sigma_n^2 + \mathbf{h}^T[n]\Sigma[n-1]\mathbf{h}[n]).$$

Por lo tanto,

$$\begin{aligned}
\hat{\boldsymbol{\theta}}[n] &= \hat{\boldsymbol{\theta}}[n-1] + \frac{1}{\sigma_n^2} \mathbf{K}[n] (\sigma_n^2 + \mathbf{h}^T[n] \boldsymbol{\Sigma}[n-1] \mathbf{h}[n]) x[n] - \mathbf{K}[n] \mathbf{h}^T[n] \hat{\boldsymbol{\theta}}[n-1] \\
&\quad - \frac{1}{\sigma_n^2} \mathbf{K}[n] \mathbf{h}^T[n] \boldsymbol{\Sigma}[n-1] \mathbf{h}[n] x[n] \\
&= \hat{\boldsymbol{\theta}}[n-1] + \mathbf{K}[n] x[n] + \frac{1}{\sigma_n^2} \mathbf{K}[n] \mathbf{h}^T[n] \boldsymbol{\Sigma}[n-1] \mathbf{h}[n] x[n] - \mathbf{K}[n] \mathbf{h}^T[n] \hat{\boldsymbol{\theta}}[n-1] \\
&\quad - \frac{1}{\sigma_n^2} \mathbf{K}[n] \mathbf{h}^T[n] \boldsymbol{\Sigma}[n-1] \mathbf{h}[n] x[n] \\
&= \hat{\boldsymbol{\theta}}[n-1] + \mathbf{K}[n] x[n] - \mathbf{K}[n] \mathbf{h}^T[n] \hat{\boldsymbol{\theta}}[n-1],
\end{aligned}$$

resultando en

$$\hat{\boldsymbol{\theta}}[n] = \hat{\boldsymbol{\theta}}[n-1] + \mathbf{K}[n] \left(x[n] - \mathbf{h}^T[n] \hat{\boldsymbol{\theta}}[n-1] \right),$$

que es la ecuación 8.25 del algoritmo LS secuencial.

También puede obtenerse una expresión recursiva para el error LS mínimo. El error mínimo por lotes es (ver ecuación 8.22)

$$\begin{aligned}
J_{\min}[n] &= (\mathbf{x}[n] - \mathbf{H}[n] \hat{\boldsymbol{\theta}}[n])^T \mathbf{C}^{-1}[n] (\mathbf{x}[n] - \mathbf{H}[n] \hat{\boldsymbol{\theta}}[n]) \\
&= \mathbf{x}^T[n] \mathbf{C}^{-1}[n] (\mathbf{x}[n] - \mathbf{H}[n] \hat{\boldsymbol{\theta}}[n]) - \hat{\boldsymbol{\theta}}^T[n] \mathbf{H}^T[n] \mathbf{C}^{-1}[n] (\mathbf{x}[n] - \mathbf{H}[n] \hat{\boldsymbol{\theta}}[n]) \\
&\stackrel{(a)}{=} \mathbf{x}^T[n] \mathbf{C}^{-1}[n] (\mathbf{x}[n] - \mathbf{H}[n] \hat{\boldsymbol{\theta}}[n]),
\end{aligned}$$

donde en (a) se tuvo en cuenta el principio de ortogonalidad, dado por la ecuación 8.7, que indica que el error de estimación $\boldsymbol{\epsilon}[n] = \mathbf{x}[n] - \mathbf{H}[n] \hat{\boldsymbol{\theta}}[n]$ es ortogonal al subespacio generado por las columnas de $\mathbf{H}[n]$ y por lo tanto $\mathbf{H}^T[n] \boldsymbol{\epsilon}[n] = \mathbf{0}$. Continuando,

$$\begin{aligned}
J_{\min}[n] &= [\mathbf{x}^T[n-1] \quad x[n]] \begin{bmatrix} \mathbf{C}^{-1}[n-1] & \mathbf{0} \\ \mathbf{0}^T & \frac{1}{\sigma_n^2} \end{bmatrix} \begin{bmatrix} \mathbf{x}[n-1] - \mathbf{H}[n-1] \hat{\boldsymbol{\theta}}[n] \\ x[n] - \mathbf{h}^T[n] \hat{\boldsymbol{\theta}}[n] \end{bmatrix} \\
&= \mathbf{x}^T[n-1] \mathbf{C}^{-1}[n-1] \left(\mathbf{x}[n-1] - \mathbf{H}[n-1] \hat{\boldsymbol{\theta}}[n] \right) + \frac{1}{\sigma_n^2} x[n] \left(x[n] - \mathbf{h}^T[n] \hat{\boldsymbol{\theta}}[n] \right) \\
&\stackrel{(a)}{=} \mathbf{x}^T[n-1] \mathbf{C}^{-1}[n-1] \left[\mathbf{x}[n-1] - \mathbf{H}[n-1] \hat{\boldsymbol{\theta}}[n-1] - \mathbf{H}[n-1] \mathbf{K}[n] \left(x[n] - \mathbf{h}^T[n] \hat{\boldsymbol{\theta}}[n-1] \right) \right] \\
&\quad + \frac{1}{\sigma_n^2} x[n] \left[x[n] - \mathbf{h}^T[n] \hat{\boldsymbol{\theta}}[n-1] - \mathbf{h}^T[n] \mathbf{K}[n] \left(x[n] - \mathbf{h}^T[n] \hat{\boldsymbol{\theta}}[n-1] \right) \right] \\
&\stackrel{(b)}{=} \mathbf{x}^T[n-1] \mathbf{C}^{-1}[n-1] \left(\mathbf{x}[n-1] - \mathbf{H}[n-1] \hat{\boldsymbol{\theta}}[n-1] - \mathbf{H}[n-1] \mathbf{K}[n] e[n] \right) \\
&\quad + \frac{1}{\sigma_n^2} x[n] \left(e[n] - \mathbf{h}^T[n] \mathbf{K}[n] e[n] \right) \\
&= \mathbf{x}^T[n-1] \mathbf{C}^{-1}[n-1] \left(\mathbf{x}[n-1] - \mathbf{H}[n-1] \hat{\boldsymbol{\theta}}[n-1] \right) - \mathbf{x}^T[n-1] \mathbf{C}^{-1}[n-1] \mathbf{H}[n-1] \mathbf{K}[n] e[n] \\
&\quad + \frac{1}{\sigma_n^2} x[n] \left(1 - \mathbf{h}^T[n] \mathbf{K}[n] \right) e[n],
\end{aligned}$$

donde en (a) usó la actualización de $\hat{\boldsymbol{\theta}}[n]$ dada por la ecuación 8.25 obtenida previamente y en (b) se definió $e[n] = x[n] - \mathbf{h}^T[n] \hat{\boldsymbol{\theta}}[n-1]$. Luego, observando que el primer sumando es $J_{\min}[n-1]$, se obtiene

que

$$\begin{aligned}
J_{\min}[n] &= J_{\min}[n-1] - \mathbf{x}^T[n-1]\mathbf{C}^{-1}[n-1]\mathbf{H}[n-1]\mathbf{K}[n]e[n] + \frac{1}{\sigma_n^2}x[n](1 - \mathbf{h}^T[n]\mathbf{K}[n])e[n] \\
&\stackrel{(a)}{=} J_{\min}[n-1] - \hat{\boldsymbol{\theta}}^T[n-1]\boldsymbol{\Sigma}^{-1}[n-1]\mathbf{K}[n]e[n] + \frac{1}{\sigma_n^2}x[n](1 - \mathbf{h}^T[n]\mathbf{K}[n])e[n] \\
&\stackrel{(b)}{=} J_{\min}[n-1] - \frac{\hat{\boldsymbol{\theta}}^T[n-1]\boldsymbol{\Sigma}^{-1}[n-1]\boldsymbol{\Sigma}[n-1]\mathbf{h}[n]e[n]}{\sigma_n^2 + \mathbf{h}^T[n]\boldsymbol{\Sigma}[n-1]\mathbf{h}[n]} \\
&\quad + \frac{1}{\sigma_n^2}x[n]\left(1 - \frac{\mathbf{h}^T[n]\boldsymbol{\Sigma}[n-1]\mathbf{h}[n]}{\sigma_n^2 + \mathbf{h}^T[n]\boldsymbol{\Sigma}[n-1]\mathbf{h}[n]}\right)e[n] \\
&\stackrel{(c)}{=} J_{\min}[n-1] - \frac{\hat{\boldsymbol{\theta}}^T[n-1]\mathbf{h}[n]e[n]}{\sigma_n^2 + \mathbf{h}^T[n]\boldsymbol{\Sigma}[n-1]\mathbf{h}[n]} + \frac{x[n]e[n]}{\sigma_n^2 + \mathbf{h}^T[n]\boldsymbol{\Sigma}[n-1]\mathbf{h}[n]} \\
&= J_{\min}[n-1] + \frac{(x[n] - \hat{\boldsymbol{\theta}}^T[n-1]\mathbf{h}[n])e[n]}{\sigma_n^2 + \mathbf{h}^T[n]\boldsymbol{\Sigma}[n-1]\mathbf{h}[n]},
\end{aligned}$$

donde en (a) se substituyó $\mathbf{x}[n-1]^T\mathbf{C}^{-1}[n-1]\mathbf{H}[n-1]$ empleando la ecuación 8.29, en (b) se substituyó $\mathbf{K}[n]$ empleando la ecuación 8.26 deducida previamente y en (c) se eliminaron las matrices inversas en el segundo sumando y se sacó denominador común en el término entre paréntesis del tercer sumando. Finalmente, notando que el término entre paréntesis es $e[n]$, se concluye que

$$J_{\min}[n] = J_{\min}[n-1] + \frac{e^2[n]}{\sigma_n^2 + \mathbf{h}^T[n]\boldsymbol{\Sigma}[n-1]\mathbf{h}[n]}.$$

8.5. Mínimos cuadrados restringidos

El problema de LS restringidos se aplica cuando hay que imponer restricciones a los parámetros a estimar. Un ejemplo típico es cuando se sabe a priori que algunos de ellos son iguales. Asíumase que el parámetro $\boldsymbol{\theta}$ esta sujeto a $r < p$ restricciones lineales. El conjunto de restricciones puede expresarse como

$$\mathbf{A}\boldsymbol{\theta} = \mathbf{b}, \quad (8.30)$$

donde \mathbf{A} es una matriz conocida de tamaño $r \times p$ de rango completo r y \mathbf{b} es un vector conocido $r \times 1$. Para encontrar el LSE sujeto a las restricciones lineales se emplea la técnica de los multiplicadores de Lagrange. El estimador $\hat{\boldsymbol{\theta}}_c$ restringido se obtiene minimizando el lagrangeano

$$J_c = (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T(\mathbf{x} - \mathbf{H}\boldsymbol{\theta}) + \boldsymbol{\lambda}^T(\mathbf{A}\boldsymbol{\theta} - \mathbf{b}).$$

Puede verificarse que al hacerlo, la solución es

$$\hat{\boldsymbol{\theta}}_c = \hat{\boldsymbol{\theta}} - (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T[\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T]^{-1}(\mathbf{A}\hat{\boldsymbol{\theta}} - \mathbf{b}), \quad (8.31)$$

donde $\hat{\boldsymbol{\theta}} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x}$ es el estimador sin restricciones.

Ejemplo: señal restringida Supóngase que el modelo de la señal es

$$s[n] = \begin{cases} \theta_1 & n = 0 \\ \theta_2 & n = 1 \\ 0 & n = 2 \end{cases}$$

y se observa $\{x[0], x[1], x[2]\}$. En notación matricial el modelo se expresa como

$$\mathbf{s} = \mathbf{H}\boldsymbol{\theta}, \quad \text{con} \quad \mathbf{H} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \quad \text{y} \quad \boldsymbol{\theta} = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}.$$

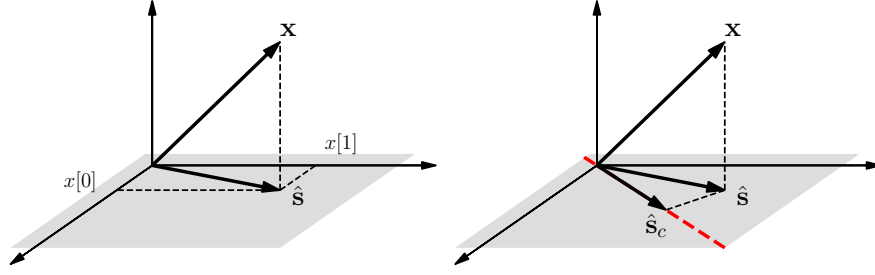


Figura 8.6: Mínimos cuadrados restringidos para el ejemplo de la señal restringida. En el caso sin restricciones (izquierda), la señal $\hat{\mathbf{s}}$ estimada pertenece al subespacio generador por \mathbf{H} , que es el plano en gris. En el caso con restricciones (derecha) la señal $\hat{\mathbf{s}}_c$ estimada pertenece al subespacio restringido, que está marcado en rojo, y es la proyección ortogonal de la señal estimada sin restricciones sobre éste.

Observar que el vector señal

$$\mathbf{s} = \begin{bmatrix} \theta_1 \\ \theta_2 \\ 0 \end{bmatrix}.$$

pertenece al plano mostrado en la figura 8.6. Considerando que en este caso $\mathbf{H}^T \mathbf{H} = \mathbf{I}$, el LSE sin restricciones es

$$\hat{\boldsymbol{\theta}} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x} = \begin{bmatrix} x[0] \\ x[1] \end{bmatrix}$$

y el estimador de la señal es

$$\hat{\mathbf{s}} = \begin{bmatrix} x[0] \\ x[1] \\ 0 \end{bmatrix}.$$

Asúmase ahora que se sabe a priori que $\theta_1 = \theta_2$. En términos de la ecuación 8.30 de la restricciones, esta restricción se escribe como

$$\begin{bmatrix} 1 & -1 \end{bmatrix} \boldsymbol{\theta} = 0, \quad \Rightarrow \quad \mathbf{A} = \begin{bmatrix} 1 & -1 \end{bmatrix} \quad \text{y} \quad \mathbf{b} = 0.$$

De la ecuación 8.31 se deduce que estimador con restricciones es

$$\begin{aligned} \hat{\boldsymbol{\theta}}_c &= \hat{\boldsymbol{\theta}} - (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{A}^T [\mathbf{A} (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{A}^T]^{-1} (\mathbf{A} \hat{\boldsymbol{\theta}} - \mathbf{b}) \\ &\stackrel{(a)}{=} \hat{\boldsymbol{\theta}} - \mathbf{A}^T (\mathbf{A} \mathbf{A}^T)^{-1} \mathbf{A} \hat{\boldsymbol{\theta}} \\ &= [\mathbf{I} - \mathbf{A}^T (\mathbf{A} \mathbf{A}^T)^{-1} \mathbf{A}] \hat{\boldsymbol{\theta}} \\ &= \left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \right) \begin{bmatrix} x[0] \\ x[1] \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} x[0] \\ x[1] \end{bmatrix}, \end{aligned}$$

donde en (a) se tuvo en cuenta que $\mathbf{H}^T \mathbf{H} = \mathbf{I}$ y $\mathbf{b} = 0$. Esto resulta en

$$\hat{\boldsymbol{\theta}}_c = \begin{bmatrix} \frac{1}{2}(x[0] + x[1]) \\ \frac{1}{2}(x[0] + x[1]) \end{bmatrix}$$

y la señal estimada restringida es

$$\hat{\mathbf{s}}_c = \mathbf{H}\hat{\boldsymbol{\theta}}_c = \begin{bmatrix} \frac{1}{2}(x[0] + x[1]) \\ \frac{1}{2}(x[0] + x[1]) \\ 0 \end{bmatrix}$$

como se muestra en la figura 8.6. En este problema, se podría haber incluido la restricción en el modelo de la señal como

$$s[n] = \begin{cases} \theta & n = 0 \\ \theta & n = 1 \\ 0 & n = 2 \end{cases}$$

que en forma matricial es

$$\mathbf{s} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \boldsymbol{\theta}$$

y \mathbf{s} debe pertenecer al subespacio restringido generado por el vector $[1 \ 1 \ 0]^T$ indicado en la figura 8.6. La señal estimada restringida $\hat{\mathbf{s}}_c$ puede obtenerse como la proyección de la señal estimada sin restricciones sobre el subespacio restringido. Ver el problema de la sección 8.8.24.

8.6. Mínimos cuadrados no lineales

Mediante el procedimiento de mínimos cuadrados, el parámetro $\boldsymbol{\theta}$ del modelo se determina minimizando el error LS

$$J = (\mathbf{x} - \mathbf{s}(\boldsymbol{\theta}))^T (\mathbf{x} - \mathbf{s}(\boldsymbol{\theta})),$$

donde $\mathbf{s}(\boldsymbol{\theta})$ es el modelo de señal de los datos \mathbf{x} , cuya dependencia del parámetro $\boldsymbol{\theta}$ se muestra explícitamente. Notar que si $\mathbf{x} - \mathbf{s}(\boldsymbol{\theta}) \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$, el LSE es también el MLE, ya que el error J es el exponente de la función de verosimilitud (ver el capítulo 7). En el problema LS lineal, la señal se expresa como $\mathbf{s}(\boldsymbol{\theta}) = \mathbf{H}\boldsymbol{\theta}$. En el caso general, $\mathbf{s}(\boldsymbol{\theta})$ es una función no lineal de dimensión N y no puede expresarse de esa forma. En esta situación, la minimización de J es mucho más difícil o incluso, imposible. En la práctica, el cálculo del LSE no lineal se basa en enfoques iterativos y sufre las mismas limitaciones de la determinación del MLE mediante métodos numéricos. El método más eficaz de búsqueda en una grilla solo es viable si la dimensión p de $\boldsymbol{\theta}$ es pequeña.

Hay algunos métodos para reducir la complejidad del problema LS no lineal que pueden emplearse en algunos pocos casos específicos. Uno es la *transformación de los parámetros*, que consiste en buscar una transformación uno a uno de $\boldsymbol{\theta}$ que produzca un modelo lineal de señal en un nuevo espacio. Específicamente, se busca una función \mathbf{g} de dimensión p de forma que el parámetro transformado $\boldsymbol{\alpha} = \mathbf{g}(\boldsymbol{\theta})$ produzca un modelo lineal de señal, es decir,

$$\mathbf{s}(\boldsymbol{\theta}(\boldsymbol{\alpha})) = \mathbf{s}(\mathbf{g}^{-1}(\boldsymbol{\alpha})) = \mathbf{H}\boldsymbol{\alpha}.$$

De esta forma, el LSE lineal de $\boldsymbol{\alpha}$ se obtiene mediante la ecuación 8.5 y luego, el LSE no lineal de $\boldsymbol{\theta}$ se obtiene como $\hat{\boldsymbol{\theta}} = \mathbf{g}^{-1}(\hat{\boldsymbol{\alpha}})$. Este método se basa en la propiedad de que la minimización puede realizarse en cualquier espacio transformado obtenido mediante una transformación uno a uno y luego aplicar la transformación inversa para volver al espacio original, como se muestra en el problema de la sección 8.8.26. Otro método para reducir la complejidad es el de la *separación de parámetros*. En este caso, se aprovecha el hecho de que si bien el modelo es no lineal, podría ser lineal en alguno de los parámetros. En el problema de la sección 8.8.1 se explica la técnica. Una situación en donde esto ocurre es en el ejemplo de la sección 7.5.1 y también en el problema de la sección 8.8.28.

A continuación se desarrolla el enfoque para resolver el problema LS no lineal general empleando el método numérico de Newton-Raphson. El error LS a minimizar es

$$J = (\mathbf{x} - \mathbf{s}(\boldsymbol{\theta}))^T (\mathbf{x} - \mathbf{s}(\boldsymbol{\theta})) = \sum_{n=0}^{N-1} (x[n] - s[n])^2.$$

Las condiciones necesarias pueden obtenerse diferenciando J , y son

$$\frac{\partial J}{\partial \theta_j} = -2 \sum_{i=0}^{N-1} (x[i] - s[i]) \frac{\partial s[i]}{\partial \theta_j} = 0, \quad j = 1, \dots, p.$$

Si se define la matriz jacobiana $N \times p$ como

$$\left[\frac{\partial \mathbf{s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right]_{ij} = \frac{\partial s[i]}{\partial \theta_j} \quad \begin{array}{l} i = 0, \dots, N-1 \\ j = 1, \dots, p \end{array}$$

las condiciones necesarias quedan

$$\sum_{i=0}^{N-1} (x[i] - s[i]) \left[\frac{\partial \mathbf{s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right]_{ij} = 0, \quad j = 1, \dots, p, \quad (8.32)$$

que en forma matricial puede escribirse como

$$\frac{\partial \mathbf{s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T (\mathbf{x} - \mathbf{s}(\boldsymbol{\theta})) = \mathbf{0},$$

y consiste en un sistema de p ecuaciones no lineales. Observar que en el caso en que el modelo de la señal es lineal con el parámetro, $\mathbf{s}(\boldsymbol{\theta}) = \mathbf{H}\boldsymbol{\theta}$, de forma que

$$s[i] = [\mathbf{H}\boldsymbol{\theta}]_i = \sum_{j=1}^p h_{ij} \theta_j \quad \Rightarrow \quad \frac{\partial s[i]}{\partial \theta_j} = h_{ij},$$

resultando en que $\partial \mathbf{s}(\boldsymbol{\theta}) / \partial \boldsymbol{\theta} = \mathbf{H}$, las condiciones necesarias quedan

$$\mathbf{H}^T (\mathbf{x} - \mathbf{H}\hat{\boldsymbol{\theta}}) = \mathbf{0} \quad \Rightarrow \quad \mathbf{H}^T \mathbf{H} \hat{\boldsymbol{\theta}} = \mathbf{H}^T \mathbf{x} \quad \Rightarrow \quad \hat{\boldsymbol{\theta}} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x},$$

que es el LSE usual lineal. El sistema no lineal de ecuaciones puede resolverse mediante el método iterativo de Newton-Raphson (ver la sección 7.7 de [1]). La función a resolver es

$$\mathbf{g}(\boldsymbol{\theta}) = \frac{\partial \mathbf{s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}^T (\mathbf{x} - \mathbf{s}(\boldsymbol{\theta})), \quad (8.33)$$

por lo que la iteración de Newton-Raphson es

$$\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k - \left(\frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)^{-1} \mathbf{g}(\boldsymbol{\theta}) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_k}. \quad (8.34)$$

Para encontrar la matriz jacobiana de \mathbf{g} , que es en realidad la matriz hessiana escalada de J , se tiene en cuenta que el elemento i -ésimo de \mathbf{g} es (ver la ecuación 8.32)

$$[\mathbf{g}(\boldsymbol{\theta})]_i = \sum_{n=0}^{N-1} (x[n] - s[n]) \frac{\partial s[n]}{\partial \theta_i}, \quad i = 1, \dots, p,$$

y por lo tanto

$$\begin{aligned} \frac{\partial [\mathbf{g}(\boldsymbol{\theta})]_i}{\partial \theta_j} &= \frac{\partial}{\partial \theta_j} \left[\sum_{n=0}^{N-1} (x[n] - s[n]) \frac{\partial s[n]}{\partial \theta_i} \right] \\ &= \sum_{n=0}^{N-1} \left[(x[n] - s[n]) \frac{\partial^2 s[n]}{\partial \theta_i \partial \theta_j} - \frac{\partial s[n]}{\partial \theta_i} \frac{\partial s[n]}{\partial \theta_j} \right]. \end{aligned}$$

Para expresar esto de forma mas sucinta, sea

$$[\mathbf{H}(\boldsymbol{\theta})]_{ij} = \left[\frac{\partial \mathbf{s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right]_{ij} = \frac{\partial s[i]}{\partial \theta_j} \quad \begin{array}{l} i = 0, \dots, N-1 \\ j = 1, \dots, p \end{array}, \quad (8.35)$$

y

$$[\mathbf{G}_n(\boldsymbol{\theta})]_{ij} = \frac{\partial^2 s[n]}{\partial \theta_i \partial \theta_j} \quad \begin{array}{ll} i &= 0, \dots, p \\ j &= 1, \dots, p \end{array} \quad (8.36)$$

De esta forma

$$\begin{aligned} \frac{\partial [\mathbf{g}(\boldsymbol{\theta})]_i}{\partial \theta_j} &= \sum_{n=0}^{N-1} [(x[n] - s[n])[\mathbf{G}_n(\boldsymbol{\theta})]_{ij} - [\mathbf{H}(\boldsymbol{\theta})]_{ni}[\mathbf{H}(\boldsymbol{\theta})]_{nj}] \\ &= \sum_{n=0}^{N-1} [\mathbf{G}_n(\boldsymbol{\theta})]_{ij}(x[n] - s[n]) - \sum_{n=0}^{N-1} [\mathbf{H}^T(\boldsymbol{\theta})]_{in}[\mathbf{H}(\boldsymbol{\theta})]_{nj} \\ &= \sum_{n=0}^{N-1} [\mathbf{G}_n(\boldsymbol{\theta})]_{ij}(x[n] - s[n]) - [\mathbf{H}^T(\boldsymbol{\theta})\mathbf{H}(\boldsymbol{\theta})]_{ij}, \end{aligned}$$

resultando en que

$$\frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{n=0}^{N-1} \mathbf{G}_n(\boldsymbol{\theta})(x[n] - s[n]) - \mathbf{H}^T(\boldsymbol{\theta})\mathbf{H}(\boldsymbol{\theta}) \quad (8.37)$$

Finalmente, sustituyendo las ecuaciones 8.33 y 8.37 en la ecuación 8.34 se obtiene que la iteración de Newton-Raphson para encontrar el estimador LS no lineal es

$$\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k + \left(\mathbf{H}^T(\boldsymbol{\theta}_k)\mathbf{H}(\boldsymbol{\theta}_k) - \sum_{n=0}^{N-1} \mathbf{G}_n(\boldsymbol{\theta}_k)(x[n] - s[n]) \right)^{-1} \mathbf{H}^T(\boldsymbol{\theta}_k)(\mathbf{x} - \mathbf{s}(\boldsymbol{\theta}_k)), \quad (8.38)$$

donde $\mathbf{H}(\boldsymbol{\theta})$ está dada por la ecuación 8.35 y $\mathbf{G}_n(\boldsymbol{\theta})$ por la ecuación 8.36, que son las derivadas parciales primera y segunda de la señal respecto a los parámetros. Es interesante notar que en caso de un modelo lineal, $\mathbf{s}(\boldsymbol{\theta}) = \mathbf{H}\boldsymbol{\theta}$, se cumple que $\mathbf{H}(\boldsymbol{\theta}) = \mathbf{H}$ y $\mathbf{G}_n(\boldsymbol{\theta}) = \mathbf{0}$, y por lo tanto

$$\begin{aligned} \boldsymbol{\theta}_{k+1} &= \boldsymbol{\theta}_k + (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T(\mathbf{x} - \mathbf{H}\boldsymbol{\theta}_k) \\ &= (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x}. \end{aligned}$$

Esto indica que la convergencia se alcanza en un solo paso, y se debe a que la naturaleza cuadrática del error J resulta en una función $\mathbf{g}(\boldsymbol{\theta})$ lineal. Es esperable que para modelos de señal aproximadamente lineales, la convergencia sea rápida.

8.7. Ejemplos de procesamiento de señales

Se incluyen a continuación algunos ejemplos en procesamiento de señales donde se emplea el LSE. En estas aplicaciones, el estimador MVU no puede obtenerse. Adicionalmente, la caracterización estadística del ruido puede no conocerse, e incluso si es conocida, el estimador MLE es generalmente muy complicado de calcular.

8.7.1. Ejemplo: diseño de filtros digitales

Un problema común en procesamiento de señales es el diseño de un filtro digital cuya respuesta en frecuencia cumpla ciertas especificaciones. Alternativamente, se puede intentar hacer coincidir la respuesta al impulso con la respuesta al impulso deseada, y es el enfoque examinado en el presente ejemplo. Un filtro de respuesta al impulso infinita (*infinite impulse response*, IIR) tiene función de transferencia

$$\begin{aligned} \mathcal{H}(z) &= \frac{\mathcal{B}(z)}{\mathcal{A}(z)} \\ &= \frac{b[0] + b[1]z^{-1} + \dots + b[q]z^{-q}}{1 + a[1]z^{-1} + \dots + a[p]z^{-p}}. \end{aligned}$$

Si la respuesta en frecuencia deseada es $H_d(f) = \mathcal{H}_d(e^{j2\pi f})$, su transformada de Fourier es la respuesta al impulso deseada,

$$h_d[n] = \mathcal{F}^{-1}\{H_d(f)\}.$$

La respuesta al impulso de filtro es una ecuación en diferencias recursiva que se obtiene aplicando la transformada z inversa a $\mathcal{H}(z)$. Para hacerlo, se multiplica por el denominador a ambos lados de la igualdad de la función de transferencia,

$$\mathcal{H}(z) + a[1]\mathcal{H}(z)z^{-1} + \cdots + a[p]\mathcal{H}(z)z^{-p} = b[0] + b[1]z^{-1} + \cdots + b[q]z^{-q}$$

y al aplicar la transformada z inversa, se obtiene que

$$h[n] + a[1]h[n-1] + \cdots + a[p]h[n-p] = b[0]\delta[n] + b[1]\delta[n-1] + \cdots + b[q]\delta[n-q],$$

donde se tuvo en cuenta que

$$\delta[n - n_0] \xleftrightarrow{\mathcal{Z}} z^{-n_0} \quad \text{y} \quad \delta[n - n_0] * h[n] = h[n - n_0] \xleftrightarrow{\mathcal{Z}} z^{-n_0} \mathcal{H}(z).$$

Despejando $h[n]$, la expresión recursiva de la respuesta al impulso queda

$$h[n] = \begin{cases} -\sum_{k=1}^p a[k]h[n-k] + \sum_{k=0}^q b[k]\delta[n-k] & n \geq 0 \\ 0 & n < 0. \end{cases}$$

La solución directa mediante LS sería elegir los coeficientes $\{a[k], b[k]\}$ de forma de minimizar

$$J = \sum_{n=0}^{N-1} (h_d[n] - h[n])^2,$$

donde N es un entero suficientemente grande de forma que $h[n]$ ya se haya extinguido y sea esencialmente nula. Notar que $h_d[n]$ son los datos y $h[n]$ es el modelo de la señal. El ruido en los datos se atribuye a error de modelado, por lo que su caracterización estadística es desconocida. Desafortunadamente, este enfoque produce un problema LS no lineal (ver el problema de la sección 8.8.28). Como ejemplo, si

$$\mathcal{H}(z) = \frac{b[0]}{1 + a[1]z^{-1}},$$

se tiene que

$$h[n] = \begin{cases} b[0](-a[1])^n & n \geq 0 \\ 0 & n < 0. \end{cases}$$

y

$$J = \sum_{n=0}^{N-1} (h_d[n] - b[0](-a[1])^n)^2,$$

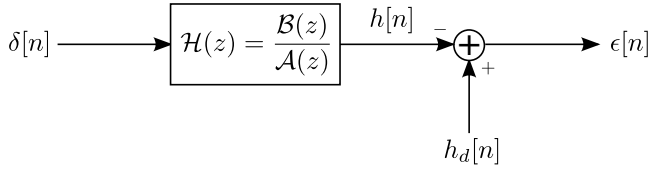
que es fuertemente no lineal en $a[1]$. De hecho, es la presencia del denominador $\mathcal{A}(z)$ que produce que el error LS no sea cuadrático. Para evitar este problema, se puede filtrar $h_d[n]$ y $h[n]$ con $\mathcal{A}(z)$ de forma de cancelar el denominador de $\mathcal{H}(z)$, como se muestra en la figura 8.7. Por lo tanto, se minimiza el error LS filtrado

$$J_f = \sum_{n=0}^{N-1} (h_d[n] - b[n])^2$$

donde $h_{d_f}[n]$ está dado por

$$h_{d_f}[n] = \sum_{k=0}^p a[k]h_d[n-k]$$

Error de mínimos cuadrados



Error de mínimos cuadrados filtrado

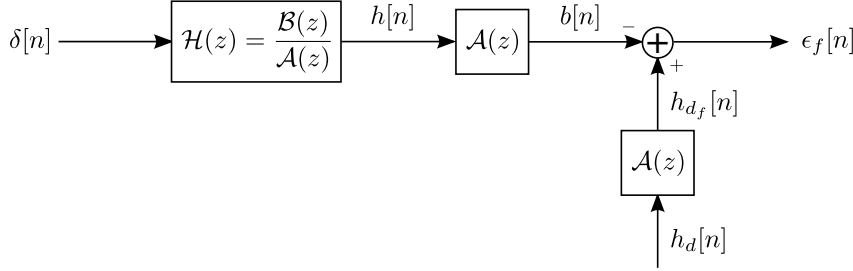


Figura 8.7: Conversión del problema LS no lineal a lineal para el diseño de un filtro.

con $a[0] = 1$. El error LS filtrado es por lo tanto

$$\begin{aligned} J_f &= \sum_{n=0}^{N-1} \left(\sum_{k=0}^p a[k] h_d[n-k] - b[n] \right)^2 \\ &\stackrel{(a)}{=} \sum_{n=0}^{N-1} \left(h_d[n] + \sum_{k=1}^p a[k] h_d[n-k] - b[n] \right)^2 \\ &= \sum_{n=0}^{N-1} \left[h_d[n] - \left(- \sum_{k=1}^p a[k] h_d[n-k] + b[n] \right) \right]^2, \end{aligned}$$

donde en (a) se separó el sumando correspondiente a $k = 0$ de la sumatoria en k . Para obtener los coeficientes del filtro que minimizan J_f , se observa que los coeficientes $b[n]$ solo aparecen en los primeros $(q + 1)$ términos, ya que $b[n] = 0$ en $n > q$. Como resultado, se tiene que

$$J_f(\mathbf{a}, \mathbf{b}) = \sum_{n=0}^q \left[h_d[n] - \left(- \sum_{k=1}^p a[k] h_d[n-k] + b[n] \right) \right]^2 + \sum_{n=q+1}^{N-1} \left[h_d[n] - \left(- \sum_{k=1}^p a[k] h_d[n-k] \right) \right]^2,$$

donde $\mathbf{a} = [a[1] \dots a[p]]^T$ y $\mathbf{b} = [b[0] \dots b[q]]^T$. La primera sumatoria puede minimizarse, en realidad, anularse, haciendo que

$$h_d[n] - \left(- \sum_{k=1}^p a[k] h_d[n-k] + b[n] \right) = 0, \quad n = 0, \dots, q,$$

ya que $h_d[n]$ puede elegirse de forma de anular la expresión para cada n . Despejando $b[n]$ se obtiene que

$$b[n] = h_d[n] + \sum_{k=1}^p a[k] h_d[n-k], \quad n = 0, \dots, q.$$

Observando que

$$\begin{aligned} b[0] &= h_d[0] \\ b[1] &= h_d[1] + a[1]h_d[0] \\ b[2] &= h_d[2] + a[1]h_d[1] + a[2]h_d[0] \\ &\vdots \\ b[q] &= h_d[q] + a[1]h_d[q-1] + \dots + a[p]h_d[q-p], \end{aligned}$$

en notación matricial, el LSE de los coeficientes de numerador es

$$\hat{\mathbf{b}} = \mathbf{h} + \mathbf{H}_0 \hat{\mathbf{a}}$$

donde

$$\mathbf{h} = [h_d[0] \quad h_d[1] \quad \dots \quad h_d[q]]^T \quad \text{y} \quad \mathbf{H}_0 = \begin{bmatrix} 0 & 0 & \dots & 0 \\ h_d[0] & 0 & \dots & 0 \\ h_d[1] & h_d[0] & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ h_d[p-1] & h_d[p-2] & \dots & h_d[0] \\ \vdots & \vdots & & \vdots \\ h_d[q-1] & h_d[q-2] & \dots & h_d[q-p] \end{bmatrix}.$$

El vector \mathbf{h} tiene dimensiones $(q+1) \times 1$, mientras que la matriz \mathbf{H}_0 tiene dimensiones $(q+1) \times p$ y es Toeplitz con la primera columna $[0, h_d[0], h_d[1], \dots, h_d[q-1]]^T$ y la primera fila $[0, \dots, 0]$ asumiendo que $q \geq p$. Para encontrar el LSE de los coeficientes \mathbf{a} del denominador hay que minimizar

$$\begin{aligned} J_f(\mathbf{a}, \hat{\mathbf{b}}) &= \sum_{n=q+1}^{N-1} \left[h_d[n] - \left(- \sum_{k=1}^p a[k] h_d[n-k] \right) \right]^2 \\ &= (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T (\mathbf{x} - \mathbf{H}\boldsymbol{\theta}), \end{aligned}$$

donde $\boldsymbol{\theta} = \mathbf{a}$ y

$$\mathbf{x} = [h_d[q+1] \quad h_d[q+2] \quad \dots \quad h_d[N-1]] \quad \text{y} \quad \mathbf{H} = \begin{bmatrix} h_d[q] & h_d[q-1] & \dots & h_d[q-p+1] \\ h_d[q+1] & h_d[q] & \dots & h_d[q-p+2] \\ \vdots & \vdots & & \vdots \\ h_d[N-2] & h_d[N-3] & \dots & h_d[N-1-p] \end{bmatrix}.$$

El vector \mathbf{x} tiene dimensiones $(N-1-q) \times 1$ y \mathbf{H} tiene dimensiones $(N-1-q) \times p$ y es Toeplitz con la primera columna $[h_d[q], h_d[q+1], \dots, h_d[N-2]]^T$ y la primera fila $[h_d[q], h_d[q-1], \dots, h_d[q-p+1]]$. Por lo tanto, el LSE de los coeficientes del denominador es

$$\hat{\mathbf{a}} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x}.$$

Este método para el diseño de filtros digitales se llama *método de mínimos cuadrados de Prony*. Como ejemplo, se considera el diseño de un filtro pasabajos cuya respuesta en frecuencia ideal es

$$H'_d(f) = \begin{cases} 1 & |f| < f_c \\ 0 & |f| > f_c, \end{cases}$$

donde f_c es la frecuencia de corte. La correspondiente respuesta al impulso es

$$h'_d[n] = \frac{\sin 2\pi f_c n}{\pi n}, \quad -\infty < n < \infty$$

la cual es infinita y no causal. Para asegurar causalidad, se trunca manteniendo solamente las muestras $|n| \leq n_0$ y se retarda n_0 muestras. Luego, para aproximar la respuesta al impulso deseada con el método LS de Prony, se asume que se dispone de N muestras de la respuesta al impulso

$$h_d[n] = \frac{\sin 2\pi f_c (n - n_0)}{\pi (n - n_0)}, \quad n = 0, \dots, N-1. \quad (8.39)$$

Se realizó una simulación por computadora empleando una señal deseada con frecuencia de corte $f_c = 0.1$, largo de $N = 51$ muestras y retardo de $n_0 = 25$ muestras. Usando el método LS de Prony con $p = q = 10$ se diseñó un filtro digital con el objetivo de igualar su respuesta al impulso con la respuesta al impulso deseada. El resultado se muestra en la figura 8.8. Respecto al filtro deseado, se observa que el efecto del

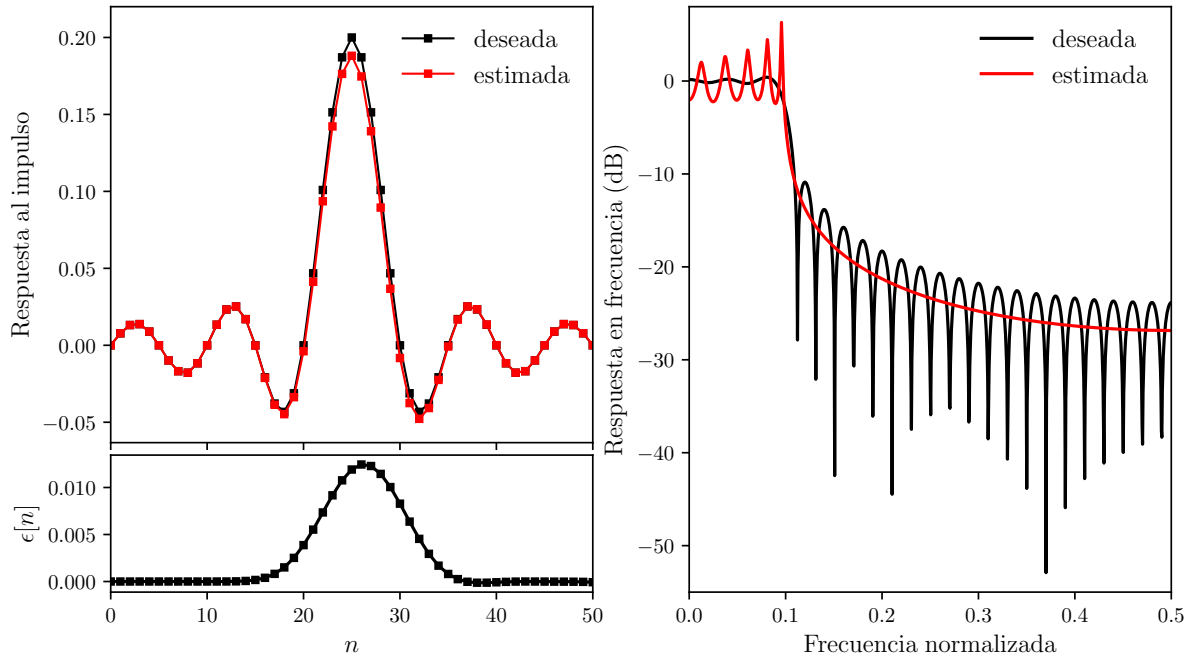


Figura 8.8: Diseño de un filtro digital empleando el método de LS de Prony. Se compara la respuesta al impulso deseada y la obtenida $h_d[n]$ y $h[n]$ respectivamente, y las correspondientes respuestas en frecuencia $H_d(f)$ y $H(f)$. También se grafica el error de estimación $\epsilon[n] = h_d[n] - h[n]$. Los parámetros empleados son $p = q = 10$, $N = 51$ muestras y $f_c = 0.1$.

```
import numpy as np
from scipy import linalg

# hd es la señal deseada dada, con hd.shape = (N, )
# el filtro de Prony a diseñar tiene función de transferencia con
# q coeficientes en el numerador y p coeficientes en el denominador
## estimación de los coeficientes del denominador, a[k], k = 1,..., p
x = hd[q + 1:]
H = linalg.toeplitz(hd[q: N - 1], hd[q: q - p: -1])
a = linalg.solve(H.T @ H, -H.T @ x)
## estimación de los coeficientes del numerador, b[k], k = 0,..., q
h = hd[: q + 1]
H0 = linalg.toeplitz(np.concatenate(([0], hd[: q])), np.zeros((p, )))
b = h + H0 @ a
```

Figura 8.9: Implementación en el lenguaje Python del método LS de Prony para el diseño de filtros digitales.

truncamiento de la respuesta al impulso ideal se manifiesta en la estructura de lóbulos laterales en la respuesta en frecuencia deseada. La respuesta en frecuencia del filtro digital obtenido con el método de Prony tiene picos en la banda pasante y roll-off mas suave sin lóbulos laterales en comparación con la respuesta en frecuencia deseada. La aproximación al filtro deseado mejora si se incrementan los valores de p y q . En la figura 8.9 se muestra el código en el lenguaje Python para el cálculo de los coeficientes $\hat{\mathbf{a}}$ y $\hat{\mathbf{b}}$. Para los parámetros elegidos en esta simulación, la matriz $\mathbf{H}^T \mathbf{H}$ en el cálculo de $\hat{\mathbf{a}}$ está mal condicionada, y la inversión numérica produce un solución con errores numéricos intolerables. Una forma de regularizar la matriz es sumarle una matriz diagonal $\epsilon \mathbf{I}$ con ϵ pequeño. Esta técnica se denomina *regularización de Tikhonov*¹.

¹Ver https://en.wikipedia.org/wiki/Tikhonov_regularization.

8.7.2. Ejemplo: estimación de los parámetros AR de un modelo ARMA

A continuación se describe un método para la estimación de los parámetros AR de un modelo autorregresivo de media móvil (*autoregressive-moving-average*, ARMA). Un modelo ARMA de un proceso WSS asume que la PSD es (por ejemplo el capítulo 3 de [5])

$$P_{xx}(f) = \frac{\sigma_u^2 |B(f)|^2}{|A(f)|^2}$$

donde

$$B(f) = 1 + \sum_{k=1}^q b[k] e^{-j2\pi f k} \quad \text{y} \quad A(f) = 1 + \sum_{k=1}^p a[k] e^{-j2\pi f k}.$$

Los coeficientes $b[k]$ se llaman parámetros MA del filtro, los coeficientes $a[k]$ se llaman parámetros AR del filtro y σ_u^2 es la varianza del ruido blanco conductor. El proceso se obtiene alimentando un filtro causal de respuesta de frecuencia $B(f)/A(f)$ con ruido blanco de varianza σ_u^2 . Si los $b[k]$ son nulos, se obtiene el proceso AR presentado en el ejemplo de la sección 3.5.2. La estimación de los parámetros AR de un modelo AR empleando el MLE asintótico fue discutido en el ejemplo de la sección 7.5.2. En el caso de un proceso ARMA, no es posible calcular el MLE asintótico. Un enfoque alternativo se basa en el *modelado de error de ecuación* de la ACF, en el cual se calculan los coeficientes AR, mientras que los coeficientes MA y la varianza del ruido se estiman con alguna otra técnica.

Para determinar la ACF, se parte de la transformada z de la PSD del modelo es (ver el capítulo 3 de [5])

$$\mathcal{P}_{xx}(z) = \frac{\sigma_u^2 \mathcal{B}(z) \mathcal{B}(z^{-1})}{\mathcal{A}(z) \mathcal{A}(z^{-1})}$$

donde $B(f) = \mathcal{B}(e^{j2\pi f})$, $A(f) = \mathcal{A}(e^{j2\pi f})$. Se deducirá una ecuación en diferencias para la ACF, que será la base del enfoque de modelado de error de ecuación. Tomando la transformada z inversa de $\mathcal{A}(z) \mathcal{P}_{xx}(z)$, se tiene que

$$\mathcal{Z}^{-1}\{\mathcal{A}(z) \mathcal{P}_{xx}(z)\} = \mathcal{Z}^{-1}\left\{\sigma_u^2 \mathcal{B}(z) \frac{\mathcal{B}(z^{-1})}{\mathcal{A}(z^{-1})}\right\}.$$

Considerando que

$$h[n] = \mathcal{Z}^{-1}\left\{\frac{\mathcal{B}(z^{-1})}{\mathcal{A}(z^{-1})}\right\},$$

por la propiedad de inversión temporal de la transformada z , que indica que

$$h[n] \xleftrightarrow{\mathcal{Z}} \mathcal{H}(z) \quad \Rightarrow \quad h[-n] \xleftrightarrow{\mathcal{Z}} \mathcal{H}(z^{-1}),$$

se obtiene que

$$\begin{aligned} \mathcal{Z}^{-1}\left\{\sigma_u^2 \mathcal{B}(z) \frac{\mathcal{B}(z^{-1})}{\mathcal{A}(z^{-1})}\right\} &= \sigma_u^2 b[n] * h[-n] \\ &= \sigma_u^2 \sum_{k=-\infty}^{\infty} b[k] h[k-n] \\ &= \sigma_u^2 \sum_{k=0}^q b[k] h[k-n] \\ &= \sigma_u^2 (b[0] h[-n] + \cdots + b[q] h[q-n]) \\ &= 0 \quad \text{si } q-n < 0, \end{aligned}$$

donde en la última igualdad se consideró que $h[n]$ es causal y por lo tanto, $h[n] = 0$ si $n < 0$. Por lo tanto

$$\mathcal{Z}^{-1}\{\mathcal{A}(z) \mathcal{P}_{xx}(z)\} = 0 \quad \text{si } n > q,$$

y teniendo en cuenta que la secuencia es la convolución entre $a[n]$ y $r_{xx}[n]$, se obtiene que

$$\sum_{k=0}^p a[k] r_{xx}[n-k] = 0 \quad \text{si } n > q. \quad (8.40)$$

con $a[0] = 1$. Estas ecuaciones se denominan *ecuaciones de Yule-Walker modificadas*. Notar del ejemplo de la sección 7.5.2 que las ecuaciones de Yule-Walker son idénticas excepto que se cumplen para $n > 0$, ya que $q = 0$ en un proceso AR.

En la práctica, la ACF se estima como

$$\hat{r}_{xx}[k] = \frac{1}{N} \sum_{n=0}^{N-1-|k|} x[n]x[n+|k|]$$

asumiendo que $x[n]$ se conoce en $n = 0, \dots, N-1$. Sustituyendo en la ecuación 8.40, se obtiene que

$$\sum_{k=0}^p a[k] \hat{r}_{xx}[n-k] = \epsilon[n] \quad n > q,$$

donde $\epsilon[n]$ denota el error en la estimación de la ACF con la ACF muestral. El modelo queda

$$\hat{r}_{xx}[n] = - \sum_{k=1}^p a[k] \hat{r}_{xx}[n-k] + \epsilon[n] \quad n > q,$$

el cual es lineal con los parámetros AR desconocidos. Si la ACF es estimada en los retardos $n = 0, \dots, M$, con $M \leq N-1$, el LSE de $a[k]$ minimiza

$$\begin{aligned} J &= \sum_{n=q+1}^M \left[\hat{r}_{xx}[n] - \left(- \sum_{k=1}^p a[k] \hat{r}_{xx}[n-k] \right) \right]^2 \\ &= (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T (\mathbf{x} - \mathbf{H}\boldsymbol{\theta}), \end{aligned}$$

con

$$\mathbf{x} = \begin{bmatrix} \hat{r}_{xx}[q+1] \\ \hat{r}_{xx}[q+2] \\ \vdots \\ \hat{r}_{xx}[M] \end{bmatrix} \quad \boldsymbol{\theta} = \begin{bmatrix} a[1] \\ a[2] \\ \vdots \\ a[p] \end{bmatrix} \quad \mathbf{H} = \begin{bmatrix} \hat{r}_{xx}[q] & \hat{r}_{xx}[q-1] & \dots & \hat{r}_{xx}[q-p+1] \\ \hat{r}_{xx}[q+1] & \hat{r}_{xx}[q] & \dots & \hat{r}_{xx}[q-p+2] \\ \vdots & \vdots & \ddots & \vdots \\ \hat{r}_{xx}[M-1] & \hat{r}_{xx}[M-2] & \dots & \hat{r}_{xx}[M-p] \end{bmatrix}.$$

El LSE de $\boldsymbol{\theta}$ es $(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x}$ y se llama *ecuaciones por mínimos cuadrados de Yule-Walker modificadas*.

8.7.3. Ejemplo: cancelador de ruido adaptativo

Un problema común en procesamiento de señales es la reducción de ruido, como por ejemplo, la supresión de la interferencia de 50 Hz de la red eléctrica en un circuito eléctrico. Una forma de hacerlo es mediante un cancelador de ruido adaptativo. Se asume que se dispone de una señal de ruido de referencia que luego de ser filtrada adecuadamente, puede igualarse al ruido y cancelarlo mediante la resta. Por ejemplo, para cancelar el ruido de 50 Hz se necesita conocer la amplitud y la fase, pero en general son desconocidas. En este caso, el objetivo del filtrado es igualar la amplitud y la fase del ruido de interferencia. En la figura 8.10 se muestra un esquema del cancelador. El canal primario contiene el

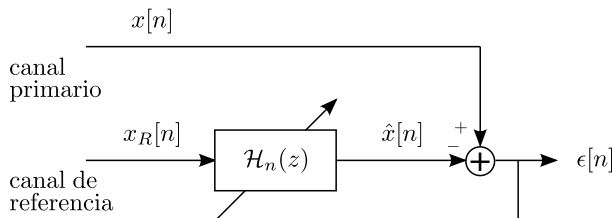


Figura 8.10: Esquema del cancelador de ruido adaptativo en tiempo discreto.

ruido $x[n]$ a ser cancelado. El canal de referencia contiene una secuencia conocida $x_R[n]$ que es similar pero no igual al ruido. Para realizar la cancelación, el ruido de referencia es filtrado con el objetivo de igualar el ruido a cancelar. Los coeficientes del filtro $\mathcal{H}_n(z)$ se determinan en cada instante n de forma

de que $\hat{x}[k] \approx x[k]$ para $k = 0, 1, \dots, n$. Como se busca que $\epsilon[n] \approx 0$, los coeficientes del filtro en el instante n se eligen de forma de minimizar

$$\begin{aligned} J &= \sum_{k=0}^n \epsilon[k] \\ &= \sum_{k=0}^n (x[k] - \hat{x}[k]) \\ &= \sum_{k=0}^n \left(x[k] - \sum_{l=0}^{p-1} h_n[l] x_R[k-l] \right)^2 \end{aligned}$$

donde $h_n[l]$ son los coeficientes del filtro en el tiempo n . Se reconoce que la estimación de los coeficientes puede resolverse como un problema de LS secuencial (ver la sección 8.4). Hay que observar que el problema de LSE asume que la interferencia es estacionaria. En caso contrario, el filtro debería cambiar sus coeficientes rápidamente para responder al cambio en la interferencia. La velocidad de respuesta al cambio depende de la cantidad de términos incluidos en $J[n]$. Si la interferencia sufre un cambio abrupto en el tiempo $n = n_0$, se espera que los coeficientes sean erróneos hasta $n \gg n_0$. Para permitir que el filtro se adapte más rápidamente a los cambios se puede ponderar más fuertemente los errores más recientes en $J[n]$ introduciendo un factor de ponderación o *factor de olvido* λ , con $0 < \lambda < 1$, como sigue

$$J[n] = \sum_{k=0}^n \lambda^{n-k} \left(x[k] - \sum_{l=0}^{p-1} h_n[l] x_R[k-l] \right)^2.$$

Esta modificación desestima exponencialmente los errores previos, permitiendo que el filtro reaccione más rápidamente a los cambios. En contrapartida, las estimaciones son más ruidosas por tomar en cuenta efectivamente menos errores LS. Notar que la solución no cambia si se minimiza

$$J[n] = \sum_{k=0}^n \frac{1}{\lambda^k} \left(x[k] - \sum_{l=0}^{p-1} h_n[l] x_R[k-l] \right)^2$$

para cada n , ya que λ^n es una constante multiplicativa de $J[n]$ para cada n . De esta forma, el problema es el LS secuencial ponderado descrito en la sección 8.4. En este caso, el modelo de la señal en el tiempo n es

$$\underbrace{\begin{bmatrix} x[0] \\ x[1] \\ \vdots \\ x[n] \end{bmatrix}}_{\mathbf{x}[n]} = \underbrace{\begin{bmatrix} x_R[0] & 0 & \dots & 0 \\ x_R[1] & x_R[0] & \dots & 0 \\ \vdots & \vdots & & \vdots \\ x_R[p-1] & x_R[p-2] & \dots & x_R[0] \\ \vdots & \vdots & & \vdots \\ x_R[n] & x_R[n-1] & \dots & x_R[n-p+1] \end{bmatrix}}_{\mathbf{H}[n]} \underbrace{\begin{bmatrix} h_n[0] \\ h_n[1] \\ \vdots \\ h_n[p-1] \end{bmatrix}}_{\boldsymbol{\theta}[n]}$$

y $\mathbf{C}[n] = \text{diag}(1, \lambda, \dots, \lambda^n)$, y se quiere minimizar

$$J[n] = (\mathbf{x}[n] - \mathbf{H}[n]\boldsymbol{\theta}[n])^T \mathbf{C}^{-1}[n] (\mathbf{x}[n] - \mathbf{H}[n]\boldsymbol{\theta}[n])$$

para cada n . De las ecuaciones 8.25-8.27 del LSE secuencial con

$$\begin{aligned} \hat{\boldsymbol{\theta}}[n] &= [h_n[0] \quad h_n[1] \quad \dots \quad h_n[p-1]]^T, \\ \mathbf{h}[n] &= [x_R[n] \quad x_R[n-1] \quad \dots \quad x_R[n-p+1]]^T \end{aligned}$$

y los pesos $\sigma_n^2 = \lambda^n$, las ecuaciones del algoritmo resultan en

$$\hat{\boldsymbol{\theta}}[n] = \hat{\boldsymbol{\theta}}[n-1] + \mathbf{K}[n]e[n]$$

donde

$$\begin{aligned} e[n] &= x[n] - \mathbf{h}^T[n] \hat{\boldsymbol{\theta}}[n-1] \\ \mathbf{h}[n] &= [x_R[n] \ x_R[n-1] \ \dots \ x_R[n-p+1]]^T \\ \mathbf{K}[n] &= \frac{\boldsymbol{\Sigma}[n-1] \mathbf{h}[n]}{\lambda^n + \mathbf{h}^T[n] \boldsymbol{\Sigma}[n-1] \mathbf{h}[n]} \\ \boldsymbol{\Sigma}[n] &= (\mathbf{I} - \mathbf{K}[n] \mathbf{h}^T[n]) \boldsymbol{\Sigma}[n-1]. \end{aligned}$$

Observar que $e[n] = x[n] - \mathbf{h}^T[n] \hat{\boldsymbol{\theta}}[n-1]$ en el algoritmo no es el error de estimación en el paso n , que está dado por $e[n] = x[n] - \hat{x}[n] = x[n] - \mathbf{h}^T[n] \hat{\boldsymbol{\theta}}[n]$, si no que es el error de estimación en el paso n considerando los coeficientes del filtro calculados en el paso $n-1$. El factor de olvido tiene que cumplir $0 < \lambda < 1$, y típicamente se elige cercano a 1. En la figura 8.12 se muestra la implementación en el lenguaje Python del algoritmo.

A continuación se muestra el resultado de una simulación por computadora. La interferencia es la señal sinusoidal

$$x[n] = A[n] \cos(2\pi(0.1)n + \pi/4), \quad \text{con} \quad A[n] = \begin{cases} 12, & 0 \leq n \leq N/2 \\ 4, & N/2 + 1 \leq n \leq N-1, \end{cases}$$

con $N = 100$. Notar que tiene un cambio abrupto de amplitud. La señal de referencia es

$$x_R[n] = \cos(2\pi(0.1)n),$$

y se implementó un cancelador de ruido con dos coeficientes ($p = 2$) ya que solo tiene que modificar a la señal de referencia en amplitud y fase para hacer que coincida con la interferencia. Para la inicialización del LSE secuencial se eligió

$$\hat{\boldsymbol{\theta}}[-1] = 0, \quad \boldsymbol{\Sigma}[-1] = 10^5 \mathbf{I}$$

y un factor de olvido $\lambda = 0.9$. La varianza del estimador en la matriz de covarianza inicial es grande para darle poca confianza al estimador inicial, que fue elegido de forma arbitraria. El factor de olvido es relativamente bajo (olvido rápido) para que el filtro pueda adaptarse rápidamente al cambio en la interferencia. En la figura 8.11 se muestra el resultado. Se observa que la interferencia es cancelada relativamente rápido. Como el factor de olvido es bajo, la señal estimada $\hat{x}[n]$ se adapta relativamente rápido al cambio en la interferencia. También se observa que los coeficientes del filtro adaptativo convergen a los valores ideales. Para calcular los valores ideales de los coeficientes del filtro del cancelador, se considera que si la entrada a un filtro es $e^{j\omega n}$, la salida es $\mathcal{H}(e^{j\omega})e^{j\omega n}$, y como en este caso se requiere que si la entrada es la referencia $x_R[n] = e^{j\pi n/5}$ la salida sea la interferencia $x[n] = a[n]e^{j\pi/4}e^{j\pi n/5}$, los coeficientes del filtro ideal cumplen que

$$\mathcal{H}(e^{j\pi/10}) = h[0] + h[1]e^{-j\pi/10} = a[n]e^{j\pi/4}.$$

Aplicando la fórmula de Euler, igualando la parte real e imaginaria a ambos lados de la igualdad y resolviendo el sistema de dos ecuaciones y dos incógnitas se llega a que

$$h[0] = A[n] \frac{\sin 9\pi/20}{\sin \pi/5} \quad \text{y} \quad h[1] = -A[n] \frac{\sin \pi/4}{\sin \pi/5}.$$

En la figura 8.12 se muestra la parte principal del código en Python del cancelador.

8.8. Problemas

8.8.1. Problema 1

Se quiere minimizar el error LS

$$J = \sum_{n=0}^{N-1} (x[n] - A \cos 2\pi f_0 n - B r^n)^2$$

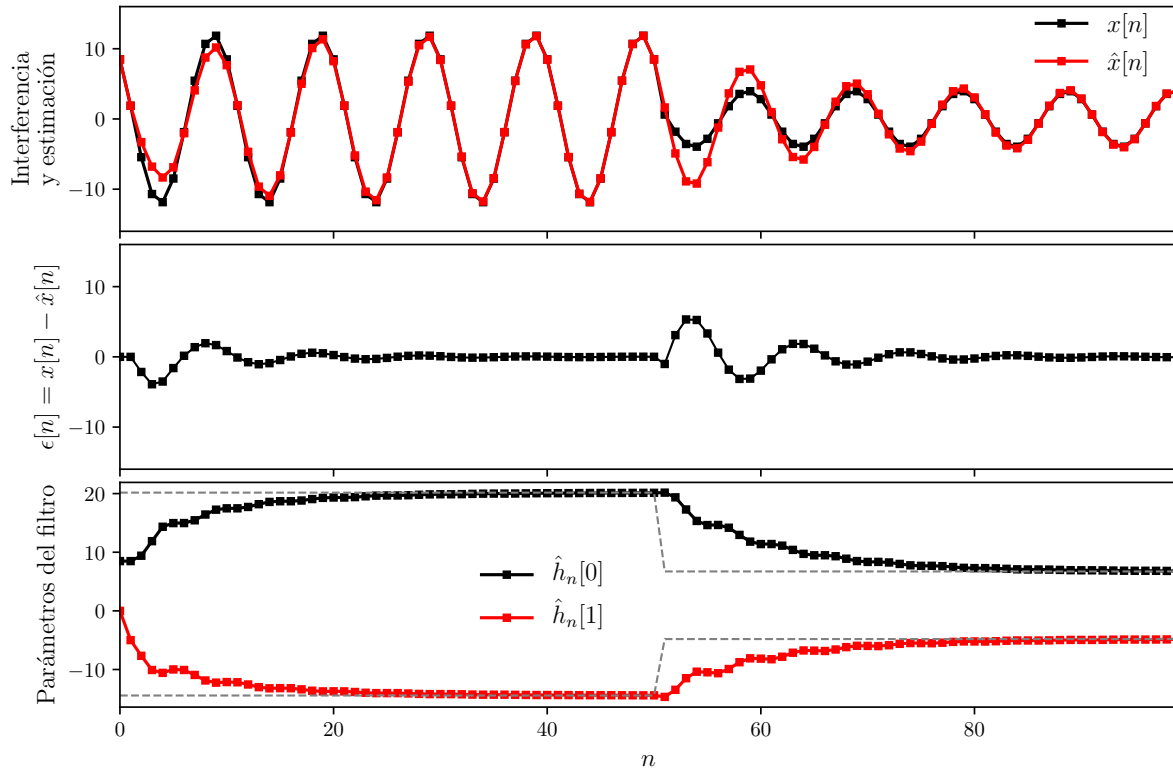


Figura 8.11: Cancelador de ruido. Se muestra la interferencia y la señal estimada a partir de la referencia, el error de estimación y los coeficientes del filtro del cancelador estimados. También se grafica los coeficientes del filtro ideal del cancelador.

```
import numpy as np

# x_r y x_i son la señal de interferencia y la señal de referencia
# respectivamente con x_r.shape = x_i.shape = (N, ).
# p es el orden del filtro adaptativo y lamb es el factor de olvido.
# theta_i es la estimador inicial (en n=-1), con theta_i.shape = (p, ),
# y sigma_i es la matriz de covarianza del estimador inicial con
# sigma_i.shape = (p, p).
theta = theta_i
sigma = sigma_i
# agregado de ceros a x_r, ya que h[n] = [x_r[n] x_r[n-1] ... x_r[n-p+1]]
x_r_pad = np.concatenate((np.zeros((p - 1,)), x_r))
for i in np.arange(N):
    h = x_r_pad[i + (p - 1): i - 1 if i - 1 >= 0 else None: -1]
    e = x_i[i] - h @ theta
    K = (sigma @ h) / (lamb ** i + h @ sigma @ h)
    theta = theta + K * e
    sigma = (np.eye(p) - K[:, None] @ h[None, :]) @ sigma
```

Figura 8.12: Implementación del cancelador en el lenguaje Python.

para encontrar el LSE de $\theta = [A f_0 B r]^T$, donde $0 < r < 1$. ¿Se trata de un problema LS lineal o no lineal? ¿Es el error LS cuadrático en alguno de los parámetros? En caso afirmativo, ¿en cuales? ¿Cómo se puede resolver este problema de minimización empleando una computadora?

Solución Claramente es un problema LS no lineal debido a los parámetros f_0 y r . Sin embargo, es un problema LS lineal en los parámetros A y B , y por lo tanto se puede reducir su complejidad empleando

la técnica de separación de parámetros como se explica a continuación. El modelo de la señal se puede expresar como

$$\mathbf{s} = \mathbf{H}(\boldsymbol{\alpha})\boldsymbol{\beta},$$

donde el vector $\boldsymbol{\alpha}$ contiene el conjunto de parámetros no lineales y $\boldsymbol{\beta}$ contiene el conjunto de parámetros lineales en el modelo. En este caso, el modelo de señal es

$$s[n] = A \cos 2\pi f_0 n - Br^n, \quad n = 0, \dots, N-1,$$

y separando los parámetros puede expresarse en notación matricial como

$$\mathbf{s} = \begin{bmatrix} 1 & 1 \\ \cos 2\pi f_0 & r \\ \cos 4\pi f_0 & r^2 \\ \vdots & \vdots \\ \cos 2\pi f_0(N-1) & r^{N-1} \end{bmatrix} \begin{bmatrix} A \\ B \end{bmatrix}.$$

Por lo tanto,

$$\boldsymbol{\alpha} = \begin{bmatrix} f_0 \\ r \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} A \\ B \end{bmatrix}, \quad \mathbf{H}(\boldsymbol{\alpha}) = \begin{bmatrix} 1 & 1 \\ \cos 2\pi f_0 & r \\ \cos 4\pi f_0 & r^2 \\ \vdots & \vdots \\ \cos 2\pi f_0(N-1) & r^{N-1} \end{bmatrix}.$$

Como

$$J(\boldsymbol{\alpha}, \boldsymbol{\beta}) = (\mathbf{x} - \mathbf{H}(\boldsymbol{\alpha})\boldsymbol{\beta})^T (\mathbf{x} - \mathbf{H}(\boldsymbol{\alpha})\boldsymbol{\beta}),$$

el valor de $\boldsymbol{\beta}$ que minimiza J para un valor de $\boldsymbol{\alpha}$ dado es el LSE lineal usual,

$$\hat{\boldsymbol{\beta}} = (\mathbf{H}^T(\boldsymbol{\alpha})\mathbf{H}(\boldsymbol{\alpha}))^{-1}\mathbf{H}(\boldsymbol{\alpha})\mathbf{x} \quad (8.41)$$

y de la ecuación 8.6, el error LS mínimo es

$$J(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \mathbf{x}^T [\mathbf{I} - \mathbf{H}(\boldsymbol{\alpha})(\mathbf{H}^T(\boldsymbol{\alpha})\mathbf{H}(\boldsymbol{\alpha}))^{-1}\mathbf{H}(\boldsymbol{\alpha})] \mathbf{x},$$

y depende únicamente de $\boldsymbol{\alpha}$. Por lo tanto, para calcular el LSE de $\boldsymbol{\alpha}$ hay que maximizar

$$\mathbf{x}^T \mathbf{H}(\boldsymbol{\alpha})(\mathbf{H}^T(\boldsymbol{\alpha})\mathbf{H}(\boldsymbol{\alpha}))^{-1}\mathbf{H}(\boldsymbol{\alpha})\mathbf{x}.$$

La maximización puede hacerse mediante una búsqueda de grilla en $0 \leq f_0 \leq 0.5$ y $0 < r < 1$. Resolviendo la ecuación 8.41 con $\boldsymbol{\alpha} = \hat{\boldsymbol{\alpha}} = [\hat{f}_0 \hat{r}]$ donde \hat{f}_0 y \hat{r} son los valores obtenidos en la maximización, se obtiene \hat{A} y \hat{B} .

8.8.2. Problema 2

Mostrar que se cumple la desigualdad indicada en la ecuación 8.4.

Solución De la ecuación 8.1 se deduce que $J(\theta) \geq 0$ por lo que $J_{\min} \geq 0$. Además, de la ecuación 8.3 se observa que $J_{\min} \leq \sum_{n=0}^{N-1} x^2[n]$, concluyendo que

$$0 \leq J_{\min} \leq \sum_{n=0}^{N-1} x^2[n].$$

8.8.3. Problema 3

Para el modelo de señal

$$s[n] = \begin{cases} A & 0 \leq n \leq M-1 \\ -A & M \leq n \leq N-1, \end{cases}$$

encontrar el LSE de A y el error LS mínimo. Asumir que se observa $x[n] = s[n] + w[n]$ para $n = 0, \dots, N-1$. Si ahora $w[n]$ es WGN con varianza σ^2 , encontrar la PDF del LSE.

Solución De la ecuación 8.2, el LSE es

$$\hat{A} = \frac{\sum_{n=0}^{N-1} x[n]h[n]}{\sum_{n=0}^{N-1} h^2[n]}.$$

El modelo de la señal es

$$s[n] = Ah[n], \quad \text{con} \quad h[n] = \begin{cases} 1 & 0 \leq n \leq M-1 \\ -1 & M \leq n \leq N-1, \end{cases}$$

por lo que el LSE es

$$\hat{A} = \frac{\sum_{n=0}^{M-1} x[n] - \sum_{n=M}^{N-1} x[n]}{N}.$$

De la ecuación 8.3

$$J_{\min} = \sum_{n=0}^{N-1} x^2[n] - \frac{\left(\sum_{n=0}^{N-1} x[n]h[n] \right)^2}{\sum_{n=0}^{N-1} h^2[n]},$$

resultando en

$$J_{\min} = \sum_{n=0}^{N-1} x^2[n] - N\hat{A}.$$

Si $w[n]$ es WGN con varianza σ^2 , $x[n] \sim \mathcal{N}(Ah[n], \sigma^2)$. Por lo tanto,

$$\begin{aligned} E(\hat{A}) &= \frac{\sum_{n=0}^{M-1} E(x[n]) - \sum_{n=M}^{N-1} E(x[n])}{N} \\ &= \frac{A \sum_{n=0}^{M-1} h[n] - A \sum_{n=M}^{N-1} h[n]}{N} \\ &= \frac{AM - A(N-M)(-1)}{N} \\ &= A, \end{aligned}$$

y

$$\begin{aligned} \text{var}(\hat{A}) &\stackrel{(a)}{=} \frac{\sum_{n=0}^{M-1} \text{var}(x[n]) + \sum_{n=M}^{N-1} \text{var}(x[n])}{N} \\ &= \frac{\sigma^2 M + \sigma^2 (N-M)}{N^2} \\ &= \frac{\sigma^2}{N}, \end{aligned}$$

donde en (a) se consideró que como las muestras $w[n]$ son independientes también lo son las muestras $x[n]$. Como \hat{A} es una combinación lineal de variables aleatorias gaussianas independientes, \hat{A} también es gaussiana, se concluye que

$$\hat{A} \sim \mathcal{N}\left(A, \frac{\sigma^2}{N}\right).$$

8.8.4. Problema 4

Deducir que el LSE de un parámetro vectorial está dado por la ecuación 8.5 verificando que se cumple la identidad

$$J = (\mathbf{x} - \mathbf{H}\hat{\boldsymbol{\theta}})^T(\mathbf{x} - \mathbf{H}\hat{\boldsymbol{\theta}}) + (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T \mathbf{H}^T \mathbf{H} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}),$$

donde

$$\hat{\boldsymbol{\theta}} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x}.$$

Para completar la prueba, mostrar que J se minimiza cuando $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}$, asumiendo que \mathbf{H} es de rango completo por lo que $\mathbf{H}^T \mathbf{H}$ es definida positiva.

Solución Que la identidad se cumple fue probado en el problema de la sección 5.4.19. Para deducir que el error se minimiza cuando $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}$, se observa que el primer sumando es una constante. Además, como $\mathbf{H}^T \mathbf{H}$ es definida positiva, el segundo sumando es no-negativo. Esto implica que J es mínimo cuando el segundo sumando es nulo, lo que se da solo si $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}$.

8.8.5. Problema 5

Para el modelo de señal

$$s[n] = \sum_{i=1}^p A_i \cos(2\pi f_i n)$$

donde las frecuencias f_i son conocidas y se quiere estimar las amplitudes A_i , encontrar las ecuaciones normales del LSE (no intentar resolverlas). Luego, si las frecuencias son $f_i = i/N$, encontrar explícitamente el LSE y el error LS mínimo. Finalmente, si $x[n] = s[n] + w[n]$ donde $w[n]$ es WGN de varianza σ^2 , determinar la PDF del LSE asumiendo las frecuencias indicadas. Sugerencia: las columnas de \mathbf{H} son ortogonales para las frecuencias dadas.

Solución El modelo de la señal en forma matricial es

$$\mathbf{s} = \mathbf{H}\boldsymbol{\theta}, \quad \text{con} \quad \mathbf{H} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ \cos 2\pi f_1 & \cos 2\pi f_2 & \dots & \cos 2\pi f_p \\ \cos 4\pi f_1 & \cos 4\pi f_2 & \dots & \cos 4\pi f_p \\ \vdots & \vdots & & \vdots \\ \cos 2\pi f_1(N-1) & \cos 2\pi f_2(N-1) & \dots & \cos 2\pi f_p(N-1) \end{bmatrix} \quad \text{y} \quad \boldsymbol{\theta} = \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_p \end{bmatrix}.$$

Las ecuaciones normales son

$$\mathbf{H}^T \mathbf{H} \boldsymbol{\theta} = \mathbf{H}^T \mathbf{x}.$$

Si las frecuencias son $f_i = i/N$ y \mathbf{h}_i es la columna i -ésima de \mathbf{H} , se cumple que

$$\mathbf{h}_i^T \mathbf{h}_j = \sum_{n=0}^{N-1} \cos\left(\frac{2\pi i n}{N}\right) \cos\left(\frac{2\pi j n}{N}\right) = \frac{N}{2} \delta_{ij},$$

como se mostró en el problema de la sección 4.6.5 (ver también el ejemplo de la sección 4.4.2). Por lo tanto,

$$\mathbf{H}^T \mathbf{H} = \frac{N}{2} \mathbf{I}$$

y el LSE es

$$\hat{\boldsymbol{\theta}} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x} = \frac{2}{N} \mathbf{H}^T \mathbf{x}.$$

Además, como elemento i -ésimo de $\mathbf{H}^T \mathbf{x}$ es

$$[\mathbf{H}^T \mathbf{x}]_i = \sum_{n=0}^{N-1} x[n] \cos\left(\frac{2\pi i n}{N}\right),$$

el elemento i -ésimo θ_i del LSE es

$$\hat{A}_i = \frac{2}{N} \sum_{n=0}^{N-1} x[n] \cos\left(\frac{2\pi i n}{N}\right).$$

De la ecuación 8.6, el error LS mínimo es

$$\begin{aligned} J_{\min} &= \mathbf{x}^T (\mathbf{I} - \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T) \mathbf{x} \\ &= \mathbf{x}^T \mathbf{x} - \frac{2}{N} \mathbf{x}^T \mathbf{H} \mathbf{H}^T \mathbf{x} \\ &\stackrel{(a)}{=} \mathbf{x}^T \mathbf{x} - \frac{N}{2} \hat{\boldsymbol{\theta}}^T \hat{\boldsymbol{\theta}}, \end{aligned}$$

donde en (a) se empleó que $\mathbf{H}^T \mathbf{x} = (N/2) \hat{\boldsymbol{\theta}}$. Finalmente,

$$J_{\min} = \sum_{n=0}^{N-1} x^2[n] - \frac{N}{2} \sum_{i=1}^p \hat{A}_i^2.$$

Si $x[n] = s[n] + w[n]$ con $w[n]$ WGN de varianza σ^2 , el LSE es el estimador MVU del modelo lineal y tiene las mismas propiedades, como la indicada en la sección 4.3. La esperanza del estimador es

$$E(\hat{\boldsymbol{\theta}}) = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T E(\mathbf{x}) = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{s} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{H} \boldsymbol{\theta} = \boldsymbol{\theta},$$

y a matriz de covarianza es

$$\begin{aligned} \mathbf{C}_{\hat{\boldsymbol{\theta}}} &= E \left\{ [\hat{\boldsymbol{\theta}} - E(\hat{\boldsymbol{\theta}})] [\hat{\boldsymbol{\theta}} - E(\hat{\boldsymbol{\theta}})]^T \right\} \\ &= E \left\{ [(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x} - \boldsymbol{\theta}] [(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x} - \boldsymbol{\theta}]^T \right\} \\ &= E \left\{ (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T (\mathbf{x} - \mathbf{H} \boldsymbol{\theta}) (\mathbf{x} - \mathbf{H} \boldsymbol{\theta})^T \mathbf{H} (\mathbf{H}^T \mathbf{H})^{-1} \right\} \\ &= (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T E(\mathbf{w} \mathbf{w}^T) \mathbf{H} (\mathbf{H}^T \mathbf{H})^{-1} \\ &= (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T (\sigma^2 \mathbf{I}) \mathbf{H} (\mathbf{H}^T \mathbf{H})^{-1} \\ &= \sigma^2 (\mathbf{H}^T \mathbf{H})^{-1} \end{aligned}$$

resultando en

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}} = \frac{2\sigma^2}{N} \mathbf{I}.$$

Finalmente, como $\hat{\boldsymbol{\theta}}$ es una combinación lineal de \mathbf{x} se concluye que

$$\hat{\boldsymbol{\theta}} = \mathcal{N} \left(\boldsymbol{\theta}, \frac{2\sigma^2}{N} \mathbf{I} \right).$$

8.8.6. Problema 6

Para el LSE general dado por la ecuación 8.5, encontrar la PDF del LSE en el caso en que $\mathbf{x} \sim \mathcal{N}(\mathbf{H} \boldsymbol{\theta}, \sigma^2 \mathbf{I})$. ¿El LSE es insesgado?

Solución Como indica la ecuación 8.5, el LSE es

$$\hat{\boldsymbol{\theta}} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x}.$$

Por tratarse de una combinación lineal de variables aleatorias gaussianas independientes, tiene PDF gaussiana, y por lo tanto, queda caracterizada por la media y la matriz de covarianza. La esperanza es

$$E(\hat{\boldsymbol{\theta}}) = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T E(\mathbf{x}) = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{H} \boldsymbol{\theta} = \boldsymbol{\theta},$$

indicando que el LSE es insesgado. La matriz de covarianza es

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}} = E \left\{ [\hat{\boldsymbol{\theta}} - E(\hat{\boldsymbol{\theta}})] [\hat{\boldsymbol{\theta}} - E(\hat{\boldsymbol{\theta}})]^T \right\}$$

y teniendo en cuenta que

$$\begin{aligned}\hat{\boldsymbol{\theta}} - E(\hat{\boldsymbol{\theta}}) &= (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x} - \boldsymbol{\theta} \\ &= (\mathbf{H}^T \mathbf{H})^{-1} (\mathbf{H}^T \mathbf{x} - \mathbf{H}^T \mathbf{H} \boldsymbol{\theta}) \\ &= (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T (\mathbf{x} - \mathbf{H} \boldsymbol{\theta}),\end{aligned}$$

la matriz de covarianza queda

$$\begin{aligned}\mathbf{C}_{\hat{\boldsymbol{\theta}}} &= (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T E[(\mathbf{x} - \mathbf{H} \boldsymbol{\theta})(\mathbf{x} - \mathbf{H} \boldsymbol{\theta})^T] \mathbf{H} (\mathbf{H}^T \mathbf{H})^{-1} \\ &= (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T (\sigma^2 \mathbf{I}) \mathbf{H} (\mathbf{H}^T \mathbf{H})^{-1} \\ &= \sigma^2 (\mathbf{H}^T \mathbf{H})^{-1}.\end{aligned}$$

Se concluye que

$$\hat{\boldsymbol{\theta}} \sim \mathcal{N}(\boldsymbol{\theta}, \sigma^2 (\mathbf{H}^T \mathbf{H})^{-1}).$$

8.8.7. Problema 7

En este problema se considera la estimación de la varianza σ^2 del ruido en el modelo $\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$, donde \mathbf{w} es ruido de media nula y varianza con matriz de covarianza $\sigma^2 \mathbf{I}$. Se propone el estimador

$$\hat{\sigma}^2 = \frac{1}{N} J_{\min} = \frac{1}{N} (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T (\mathbf{x} - \mathbf{H}\boldsymbol{\theta}),$$

donde $\hat{\boldsymbol{\theta}}$ es el LSE de $\boldsymbol{\theta}$ dado por la ecuación 8.5. ¿Es este estimador insesgado? En caso de que no lo sea, proponer uno que sea insesgado. Explicar los resultados. Sugerencia: serán útiles las identidades $E(\mathbf{x}^T \mathbf{y}) = E(\text{tr}(\mathbf{y} \mathbf{x}^T)) = \text{tr}(E(\mathbf{y} \mathbf{x}^T))$ y $\text{tr}(\mathbf{A} \mathbf{B}) = \text{tr}(\mathbf{B} \mathbf{A})$.

Solución Partiendo de la ecuación 8.6 para expresar J_{\min} , el estimador propuesto se puede escribir como

$$\begin{aligned}\hat{\sigma}^2 &= \frac{1}{N} \mathbf{x}^T (\mathbf{I} - \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T) \mathbf{x} \\ &= \frac{1}{N} (\mathbf{H}\boldsymbol{\theta} - \mathbf{w})^T (\mathbf{I} - \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T) (\mathbf{H}\boldsymbol{\theta} - \mathbf{w}) \\ &= \frac{1}{N} [(\mathbf{H}\boldsymbol{\theta} - \mathbf{w})^T (\mathbf{H}\boldsymbol{\theta} - \mathbf{w}) - (\mathbf{H}\boldsymbol{\theta} - \mathbf{w})^T \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T (\mathbf{H}\boldsymbol{\theta} - \mathbf{w})] \\ &= \frac{1}{N} [(\boldsymbol{\theta}^T \mathbf{H}^T \mathbf{H} \boldsymbol{\theta} + \mathbf{w}^T \mathbf{w} - 2\boldsymbol{\theta}^T \mathbf{H}^T \mathbf{w}) \\ &\quad - (\boldsymbol{\theta}^T \mathbf{H}^T \mathbf{H} (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{H} \boldsymbol{\theta} + \mathbf{w}^T \mathbf{H} (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{w} - 2\boldsymbol{\theta}^T \mathbf{H}^T \mathbf{H} (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{w})] \\ &= \frac{1}{N} [(\boldsymbol{\theta}^T \mathbf{H}^T \mathbf{H} \boldsymbol{\theta} + \mathbf{w}^T \mathbf{w} - 2\boldsymbol{\theta}^T \mathbf{H}^T \mathbf{w}) - (\boldsymbol{\theta}^T \mathbf{H}^T \mathbf{H} \boldsymbol{\theta} + \mathbf{w}^T \mathbf{H} (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{w} - 2\boldsymbol{\theta}^T \mathbf{H}^T \mathbf{w})],\end{aligned}$$

resultando en

$$\hat{\sigma}^2 = \frac{1}{N} (\mathbf{w}^T \mathbf{w} - \mathbf{w}^T \mathbf{H} (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{w}) = \frac{1}{N} \mathbf{w}^T (\mathbf{I} - \mathbf{H} (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T) \mathbf{w}.$$

Tomando esperanza, se tiene que

$$E(\hat{\sigma}^2) = \frac{1}{N} [E(\mathbf{w}^T \mathbf{w}) - E(\mathbf{w}^T \mathbf{H} (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{w})].$$

La primer esperanza es

$$E(\mathbf{w}^T \mathbf{w}) = E\left(\sum_{n=0}^{N-1} w^2[n]\right) = \sum_{n=0}^{N-1} E(w^2[n]) = N\sigma^2.$$

Para calcular la segunda esperanza, se considera la identidad $E(\mathbf{x}^T \mathbf{y}) = E(\text{tr}(\mathbf{y} \mathbf{x}^T)) = \text{tr}(E(\mathbf{y} \mathbf{x}^T))$, que es fácil de mostrar, con $\mathbf{x} = \mathbf{w}$ y $\mathbf{y} = \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{w}$. De esta forma

$$\begin{aligned} E(\mathbf{w}^T \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{w}) &= \text{tr}(E(\mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{w} \mathbf{w}^T)) \\ &= \text{tr}(\mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T E(\mathbf{w} \mathbf{w}^T)) \\ &\stackrel{(a)}{=} \text{tr}(\sigma^2 \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T) \\ &\stackrel{(b)}{=} \sigma^2 \text{tr}((\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{H}) \\ &= \sigma^2 \text{tr}(\mathbf{I}_{p \times p}) \\ &= p\sigma^2, \end{aligned}$$

donde en (a) se tuvo en cuenta que $E(\mathbf{w} \mathbf{w}^T) = \sigma^2 \mathbf{I}$ y en (b) se empleó la identidad $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$, que se demostró en el apéndice H. Combinando los resultados, se obtiene que

$$E(\hat{\sigma}^2) = \frac{1}{N}(N\sigma^2 - p\sigma^2) = \frac{N-p}{N}\sigma^2.$$

Se concluye que el estimador propuesto no es insesgado, pero solo por un factor constante. Por lo tanto, un estimador insesgado es el estimador propuesto multiplicado por el inverso del factor, es decir,

$$\check{\sigma}^2 = \frac{N}{N-p} \hat{\sigma}^2,$$

que resulta en el estimador insesgado

$$\check{\sigma}^2 = \frac{1}{N-p} J_{\min}.$$

8.8.8. Problema 8

Se considera el problema de la estimación del nivel de DC

$$x[n] = A + w[n], \quad n = 0, \dots, N-1,$$

donde $w[n]$ es ruido no correlacionado de media nula y varianza σ_n^2 , y se considera el error LS ponderado

$$J(A) = \sum_{n=0}^{N-1} \frac{1}{\sigma_n^2} (x[n] - A)^2.$$

Calcular el LSE de A y calcular la media y la varianza.

Solución La derivada del error LS es

$$\frac{\partial J(A)}{\partial A} = -2 \sum_{n=0}^{N-1} \frac{1}{\sigma_n^2} (x[n] - A),$$

e igualando a cero se obtiene que

$$\hat{A} \sum_{n=0}^{N-1} \frac{1}{\sigma_n^2} = \sum_{n=0}^{N-1} \frac{x[n]}{\sigma_n^2} \quad \Rightarrow \quad \hat{A} = \frac{\sum_{n=0}^{N-1} \frac{x[n]}{\sigma_n^2}}{\sum_{n=0}^{N-1} \frac{1}{\sigma_n^2}}.$$

La media del estimador es

$$E(\hat{A}) = \frac{\sum_{n=0}^{N-1} \frac{E(x[n])}{\sigma_n^2}}{\sum_{n=0}^{N-1} \frac{1}{\sigma_n^2}} = \frac{\sum_{n=0}^{N-1} \frac{A}{\sigma_n^2}}{\sum_{n=0}^{N-1} \frac{1}{\sigma_n^2}} = A,$$

y la varianza es

$$\text{var}(\hat{A}) \stackrel{(a)}{=} \frac{\sum_{n=0}^{N-1} \frac{\text{var}(x[n])}{\sigma_n^4}}{\left(\sum_{n=0}^{N-1} \frac{1}{\sigma_n^2}\right)^2} = \frac{\sum_{n=0}^{N-1} \frac{\sigma_n^2}{\sigma_n^4}}{\left(\sum_{n=0}^{N-1} \frac{1}{\sigma_n^2}\right)^2} = \frac{\sum_{n=0}^{N-1} \frac{1}{\sigma_n^2}}{\left(\sum_{n=0}^{N-1} \frac{1}{\sigma_n^2}\right)^2} = \frac{1}{\sum_{n=0}^{N-1} \frac{1}{\sigma_n^2}},$$

donde en (a) se tuvo en cuenta que el ruido $w[n]$ es no correlacionado.

8.8.9. Problema 9

Calcular el LSE y el error mínimo en el problema de LS ponderado observando que si \mathbf{W} es una matriz definida positiva, se puede expresar como $\mathbf{W} = \mathbf{D}^T \mathbf{D}$, donde \mathbf{D} es una matriz $N \times N$ invertible.

Solución La función de error en el problema de LS ponderado es

$$J(\boldsymbol{\theta}) = (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T \mathbf{W}(\mathbf{x} - \mathbf{H}\boldsymbol{\theta}).$$

Como la matriz \mathbf{W} es simétrica y definida positiva, admite una descomposición de Cholesky como $\mathbf{W} = \mathbf{D}^T \mathbf{D}$. Teniendo esto en cuenta, el error se puede escribir como

$$\begin{aligned} J(\boldsymbol{\theta}) &= (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T \mathbf{D}^T \mathbf{D}(\mathbf{x} - \mathbf{H}\boldsymbol{\theta}) \\ &= [\mathbf{D}(\mathbf{x} - \mathbf{H}\boldsymbol{\theta})]^T [\mathbf{D}(\mathbf{x} - \mathbf{H}\boldsymbol{\theta})] \\ &= (\mathbf{D}\mathbf{x} - \mathbf{D}\mathbf{H}\boldsymbol{\theta})^T (\mathbf{D}\mathbf{x} - \mathbf{D}\mathbf{H}\boldsymbol{\theta}) \\ &= (\mathbf{x}' - \mathbf{H}'\boldsymbol{\theta})^T (\mathbf{x}' - \mathbf{H}'\boldsymbol{\theta}), \end{aligned}$$

que es la función de error habitual con $\mathbf{x}' = \mathbf{D}\mathbf{x}$ y $\mathbf{H}' = \mathbf{D}\mathbf{H}$. Por lo tanto, de la ecuación 8.5 se obtiene que el LSE es

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= (\mathbf{H}'^T \mathbf{H}')^{-1} \mathbf{H}'^T \mathbf{x}' \\ &= (\mathbf{H}^T \mathbf{D}^T \mathbf{D} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{D}^T \mathbf{D} \mathbf{x} \end{aligned}$$

resultando en

$$\hat{\boldsymbol{\theta}} = (\mathbf{H}^T \mathbf{W} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{W} \mathbf{x}.$$

Además, de la ecuación 8.6, el error mínimo es

$$\begin{aligned} J_{\min} &= \mathbf{x}'^T (\mathbf{I} - \mathbf{H}'(\mathbf{H}'^T \mathbf{H}')^{-1} \mathbf{H}'^T) \mathbf{x}' \\ &= \mathbf{x}^T \mathbf{D}^T (\mathbf{I} - \mathbf{D}\mathbf{H}(\mathbf{H}^T \mathbf{D}^T \mathbf{D} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{D}^T) \mathbf{D} \mathbf{x} \\ &= \mathbf{x}^T (\mathbf{D}^T \mathbf{D} - \mathbf{D}^T \mathbf{D} \mathbf{H}(\mathbf{H}^T \mathbf{D}^T \mathbf{D} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{D}^T \mathbf{D}) \mathbf{x}, \end{aligned}$$

resultando en

$$J_{\min} = \mathbf{x}^T (\mathbf{W} - \mathbf{W} \mathbf{H}(\mathbf{H}^T \mathbf{W} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{W}) \mathbf{x}.$$

8.8.10. Problema 10

En referencia a la figura 8.1, mostrar que

$$\|\hat{\mathbf{s}}\|^2 + \|\mathbf{x} - \hat{\mathbf{s}}\|^2 = \|\mathbf{x}\|^2.$$

Esto puede ser pensado como el teorema de Pitágoras de mínimos cuadrados.

Solución Considerando que

$$\hat{\mathbf{s}} = \mathbf{H}\hat{\boldsymbol{\theta}} = \mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x} = \mathbf{P}\mathbf{x},$$

donde $\mathbf{P} = \mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T$ es la matriz de proyección sobre el subespacio generado por las columnas de \mathbf{H} . Además,

$$\mathbf{x} - \hat{\mathbf{s}} = \mathbf{x} - \mathbf{P}\mathbf{x} = (\mathbf{I} - \mathbf{P})\mathbf{x},$$

donde $\mathbf{I} - \mathbf{P}$ es la matriz de proyección sobre el subespacio ortogonal al generado por las columnas de \mathbf{H} . Por lo tanto,

$$\begin{aligned}\|\hat{\mathbf{s}}\|^2 + \|\mathbf{x} - \hat{\mathbf{s}}\|^2 &= \mathbf{x}^T\mathbf{P}^T\mathbf{P}\mathbf{x} + \mathbf{x}^T(\mathbf{I} - \mathbf{P})^T(\mathbf{I} - \mathbf{P})\mathbf{x} \\ &\stackrel{(a)}{=} \mathbf{x}^T\mathbf{P}\mathbf{P}\mathbf{x} + \mathbf{x}^T(\mathbf{I} - \mathbf{P})(\mathbf{I} - \mathbf{P})\mathbf{x} \\ &\stackrel{(b)}{=} \mathbf{x}^T\mathbf{P}\mathbf{x} + \mathbf{x}^T(\mathbf{I} - \mathbf{P})\mathbf{x} \\ &= \mathbf{x}^T\mathbf{x} \\ &= \|\mathbf{x}\|^2,\end{aligned}$$

donde en (a) se empleó que una matriz de proyección es simétrica, como se muestra en el problema siguiente, y en (b) que una matriz de proyección es idempotente, como se muestra en el problema de la sección 8.8.12.

8.8.11. Problema 11

En este problema se prueba que una matriz de proyección \mathbf{P} es simétrica. Sea $\mathbf{x} = \boldsymbol{\xi} + \boldsymbol{\xi}^\perp$, donde $\boldsymbol{\xi}$ pertenece al subespacio alcanzado por la matriz de proyección, o $\mathbf{P}\mathbf{x} = \boldsymbol{\xi}$, y $\boldsymbol{\xi}^\perp$ pertenece al subespacio ortogonal o $\mathbf{P}\boldsymbol{\xi}^\perp = \mathbf{0}$. Para vectores arbitrarios $\mathbf{x}_1, \mathbf{x}_2$ en R^N mostrar que

$$\mathbf{x}_1^T\mathbf{P}\mathbf{x}_2 - \mathbf{x}_2^T\mathbf{P}\mathbf{x}_1 = 0$$

descomponiendo \mathbf{x}_1 y \mathbf{x}_2 como se indicó arriba. Finalmente, probar el resultado deseado.

Solución La descomposición del vector \mathbf{x} se muestra en la figura 8.13. Considerando esta descomposi-

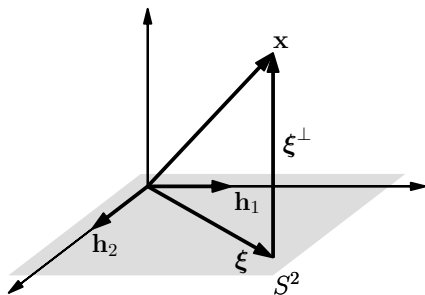


Figura 8.13: Descomposición del vector \mathbf{x} para el caso en que $\mathbf{x} \in R^3$ y la matriz de proyección \mathbf{P} conduce al subespacio S^2 .

ción, se observa que se cumple que

$$\begin{aligned}\mathbf{x}_1^T\mathbf{P}\mathbf{x}_2 - \mathbf{x}_2^T\mathbf{P}\mathbf{x}_1 &= (\boldsymbol{\xi}_1 + \boldsymbol{\xi}_1^\perp)^T\mathbf{P}(\boldsymbol{\xi}_2 + \boldsymbol{\xi}_2^\perp) - (\boldsymbol{\xi}_2 + \boldsymbol{\xi}_2^\perp)^T\mathbf{P}(\boldsymbol{\xi}_1 + \boldsymbol{\xi}_1^\perp) \\ &= (\boldsymbol{\xi}_1 + \boldsymbol{\xi}_1^\perp)^T\mathbf{P}\boldsymbol{\xi}_2 + (\boldsymbol{\xi}_1 + \boldsymbol{\xi}_1^\perp)^T\mathbf{P}\boldsymbol{\xi}_2^\perp - (\boldsymbol{\xi}_2 + \boldsymbol{\xi}_2^\perp)^T\mathbf{P}\boldsymbol{\xi}_1 - (\boldsymbol{\xi}_2 + \boldsymbol{\xi}_2^\perp)^T\mathbf{P}\boldsymbol{\xi}_1^\perp \\ &\stackrel{(a)}{=} (\boldsymbol{\xi}_1 + \boldsymbol{\xi}_1^\perp)^T\boldsymbol{\xi}_2 - (\boldsymbol{\xi}_2 + \boldsymbol{\xi}_2^\perp)^T\boldsymbol{\xi}_1 \\ &= \boldsymbol{\xi}_1^T\boldsymbol{\xi}_2 + \boldsymbol{\xi}_1^{\perp T}\boldsymbol{\xi}_2 - \boldsymbol{\xi}_2^T\boldsymbol{\xi}_1 - \boldsymbol{\xi}_2^{\perp T}\boldsymbol{\xi}_1 \\ &\stackrel{(b)}{=} \boldsymbol{\xi}_1^T\boldsymbol{\xi}_2 - \boldsymbol{\xi}_2^T\boldsymbol{\xi}_1 \\ &\stackrel{(c)}{=} 0,\end{aligned}$$

donde en (a) se consideró que como $\xi_i \in S^p$, se cumple que $\mathbf{P}\xi_i = \xi_i$, y como ξ_i^\perp es ortogonal a S^p , se cumple que $\mathbf{P}\xi_i^\perp = \mathbf{0}$, en (b) se tuvo en cuenta que ξ_i y ξ_j^\perp son ortogonales y por lo tanto $\xi_i^\perp{}^T \xi_j$ y en (c) que los sumandos son escalares y por lo tanto, invariantes a la transposición. Para ver que \mathbf{P} es simétrica se considera la identidad que se acaba de deducir,

$$\mathbf{x}_1^T \mathbf{P} \mathbf{x}_2 = \mathbf{x}_2^T \mathbf{P} \mathbf{x}_1.$$

El lado izquierdo de la igualdad se puede escribir como

$$\mathbf{x}_1^T \mathbf{P} \mathbf{x}_2 = (\mathbf{x}_1^T \mathbf{P} \mathbf{x}_2)^T = \mathbf{x}_2^T \mathbf{P}^T \mathbf{x}_1,$$

donde nuevamente se tuvo en cuenta que se trata de magnitudes escalares. Finalmente, como

$$\mathbf{x}_2^T \mathbf{P}^T \mathbf{x}_1 = \mathbf{x}_2^T \mathbf{P} \mathbf{x}_1,$$

se concluye que $\mathbf{P}^T = \mathbf{P}$ y por lo tanto, \mathbf{P} es una matriz simétrica.

8.8.12. Problema 12

Probar las siguientes propiedades de la matriz de proyección

$$\mathbf{P} = \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T.$$

- \mathbf{P} es idempotente.
- \mathbf{P} es semidefinida positiva.
- Los valores propios de \mathbf{P} son 0 o 1.
- El rango de \mathbf{P} es p . Emplear el hecho de que la traza de una matriz es igual a la suma de sus valores propios.

Solución Hay que tener en cuenta que si la matriz \mathbf{H} tiene dimensiones $N \times p$, la matriz \mathbf{P} tiene dimensiones $N \times N$.

- Efectivamente \mathbf{P} es idempotente, ya que

$$\mathbf{P}^2 = \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T = \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T = \mathbf{P}.$$

- Para ver que \mathbf{P} es semidefinida positiva, hay que demostrar que $\mathbf{x}^T \mathbf{P} \mathbf{x} \geq 0$ para todo \mathbf{x} . Se observa que

$$\mathbf{x}^T \mathbf{P} \mathbf{x} \stackrel{(a)}{=} \mathbf{x}^T \mathbf{P} \mathbf{P} \mathbf{x} \stackrel{(b)}{=} \mathbf{x}^T \mathbf{P}^T \mathbf{P} \mathbf{x} = (\mathbf{P} \mathbf{x})^T (\mathbf{P} \mathbf{x}) = \|\mathbf{P} \mathbf{x}\|^2 \geq 0,$$

donde en (a) se empleó que \mathbf{P} es idempotente, como se mostró en a., y en (b) que \mathbf{P} es simétrica, como se demostró en el problema 8.8.11.

- Los valores propios de una matriz cumplen que $\mathbf{P} \mathbf{x} = \lambda \mathbf{x}$, con $\mathbf{x} \neq \mathbf{0}$. Sea \mathbf{x} un valor propio de \mathbf{P} con valor propio λ asociado. De esta forma, se cumple que

$$\mathbf{P} \mathbf{x} \stackrel{(a)}{=} \mathbf{P} \mathbf{P} \mathbf{x} = \mathbf{P} \lambda \mathbf{x} = \lambda \mathbf{P} \mathbf{x} = \lambda^2 \mathbf{x},$$

donde en (a) se consideró que \mathbf{P} es idempotente, como se demostró en el primer punto. Por lo tanto, si λ es el valor propio asociado al vector propio \mathbf{x} , λ^2 también es valor propio asociado a \mathbf{x} . Como un vector propio no puede tener mas de un valor propio distinto asociado, se concluye que $\lambda = 0$ o $\lambda = 1$.

- Para ver que $\text{rank}(\mathbf{P}) = p$, se parte considerando que la traza de una matriz es invariable a permutaciones cíclicas, como se muestra en el apéndice H. Por lo tanto,

$$\text{tr}(\mathbf{P}) = \text{tr}[\mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T] = \text{tr}[(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{H}] \stackrel{(a)}{=} \text{tr}(\mathbf{I}_{p \times p}) = p,$$

donde en (a), $\mathbf{I}_{p \times p}$ es la matriz identidad de tamaño $p \times p$, cuya traza es p . Además, como la suma de los valores propios de una matriz coincide con su traza (ver el apéndice H), se cumple que

$$\sum_{i=1}^N \lambda_i(\mathbf{P}) = p,$$

y teniendo en cuenta que los valores propios de \mathbf{P} son 0 o 1, como se demostró en el punto anterior, se deduce que \mathbf{P} tiene p valores propios de valor 1 y $N - p$ valores propios de valor 0. Además, como una matriz real y simétrica tiene rango igual a la cantidad de valores propios distintos de cero por ser diagonalizable (ver el apéndice H), se concluye que $\text{rank}(\mathbf{P}) = p$.

8.8.13. Problema 13

Comprobar el resultado de la ecuación 8.8 del estimador para el problema del ajuste de una recta.

Solución El modelo de los datos es

$$s[n] = A + Bn, \quad n = 0, \dots, N-1,$$

y en forma matricial, se expresa como

$$\mathbf{s} = \mathbf{H}\boldsymbol{\theta}, \quad \text{con} \quad \mathbf{H} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & N-1 \end{bmatrix} \quad \text{y} \quad \boldsymbol{\theta} = \begin{bmatrix} A \\ B \end{bmatrix}.$$

Como indica la ecuación 8.5, el LSE es

$$\hat{\boldsymbol{\theta}} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x}.$$

En este caso,

$$\mathbf{H}^T \mathbf{H} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ 0 & 1 & \dots & N-1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & N-1 \end{bmatrix} = \begin{bmatrix} \sum_{n=0}^{N-1} 1 & \sum_{n=0}^{N-1} n \\ \sum_{n=0}^{N-1} n & \sum_{n=0}^{N-1} n^2 \end{bmatrix} \stackrel{(a)}{=} N \begin{bmatrix} 1 & \frac{N-1}{2} \\ \frac{N-1}{2} & \frac{(N-1)(2N-1)}{6} \end{bmatrix}.$$

donde en (a) se emplearon las identidades dadas por las ecuaciones 3.20. La matriz inversa es

$$(\mathbf{H}^T \mathbf{H})^{-1} = \frac{1}{N} \begin{bmatrix} \frac{2(2N-1)}{N+1} & -\frac{6}{N+1} \\ -\frac{6}{N+1} & \frac{12}{N^2-1} \end{bmatrix},$$

como se calculó en el ejemplo de la sección 3.2.3 (ver la ecuación 3.21). Además, se tiene que

$$\mathbf{H}^T \mathbf{x} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ 0 & 1 & \dots & N-1 \end{bmatrix} \begin{bmatrix} x[0] \\ \vdots \\ x[N-1] \end{bmatrix} = \begin{bmatrix} \sum_{n=0}^{N-1} x[n] \\ \sum_{n=0}^{N-1} nx[n] \end{bmatrix}.$$

Combinando los resultados, se obtiene que

$$\begin{bmatrix} \hat{A} \\ \hat{B} \end{bmatrix} = \frac{1}{N} \begin{bmatrix} \frac{2(2N-1)}{N+1} & -\frac{6}{N+1} \\ -\frac{6}{N+1} & \frac{12}{N^2-1} \end{bmatrix} \begin{bmatrix} \sum_{n=0}^{N-1} x[n] \\ \sum_{n=0}^{N-1} nx[n] \end{bmatrix},$$

resultando en

$$\begin{aligned}\hat{A} &= \frac{2(2N-1)}{N(N+1)} \sum_{n=0}^{N-1} x[n] - \frac{6}{N(N+1)} \sum_{n=0}^{N-1} nx[n] \\ \hat{B} &= -\frac{6}{N(N+1)} \sum_{n=0}^{N-1} x[n] + \frac{12}{N(N^2-1)} \sum_{n=0}^{N-1} nx[n],\end{aligned}$$

que es lo que se quería mostrar. Notar que el estimador coincide con el estimador eficiente cuando el ruido es WGN del ejemplo de la sección 3.2.3. Sin embargo, no se puede asegurar que el LSE obtenido sea eficiente debido a que no están especificadas las características estadísticas del ruido.

8.8.14. Problema 14

En este problema se derivan las ecuaciones de actualización del algoritmo LS de orden recursivo emplantando un enfoque geométrico. Asíase que se dispone de $\hat{\boldsymbol{\theta}}_k$ y por lo tanto se conoce

$$\hat{\mathbf{s}}_k = \mathbf{H}_k \hat{\boldsymbol{\theta}}_k. \quad (8.42)$$

Si se agrega la columna \mathbf{h}_{k+1} a la matriz de observación, el la señal LS estimada es

$$\hat{\mathbf{s}}_{k+1} = \hat{\mathbf{s}}_k + \alpha \mathbf{h}'_{k+1},$$

donde \mathbf{h}'_{k+1} es el componente de \mathbf{h}_{k+1} ortogonal al subespacio generado por $\{\mathbf{h}_1, \dots, \mathbf{h}_k\}$. En primer lugar, encontrar \mathbf{h}'_{k+1} y luego determinar α notando que $\alpha \mathbf{h}'_{k+1}$ es la proyección de \mathbf{x} en \mathbf{h}'_{k+1} . Finalmente, como

$$\hat{\mathbf{s}}_{k+1} = \mathbf{H}_{k+1} \hat{\boldsymbol{\theta}}_{k+1}$$

determinar $\hat{\boldsymbol{\theta}}_{k+1}$.

Solución Como se discute en la sección 8.2.3, la matriz de proyección del subespacio generado por las columnas de \mathbf{H}_k es $\mathbf{P}_k = \mathbf{H}(\mathbf{H}_k^T \mathbf{H}_k) \mathbf{H}_k^T$ y la matriz de proyección del subespacio ortogonal es $\mathbf{P}_k^\perp = \mathbf{I} - \mathbf{P}_k$. Por lo tanto, la proyección de \mathbf{h}_{k+1} en el subespacio ortogonal al generado por las columnas de \mathbf{H}_k es

$$\mathbf{h}'_{k+1} = \mathbf{P}_k^\perp \mathbf{h}_{k+1}. \quad (8.43)$$

Para determinar α , se parte notando que dados dos vectores \mathbf{u} y \mathbf{v} , el vector proyección $\text{proj}_{\mathbf{u}}(\mathbf{v})$ de \mathbf{v} sobre \mathbf{u} es²

$$\text{proj}_{\mathbf{u}}(\mathbf{v}) = \frac{\mathbf{u}^T \mathbf{v}}{\|\mathbf{u}\|^2} \mathbf{u}.$$

Efectivamente, la matriz de proyección sobre el subespacio generado por \mathbf{u} es

$$\mathbf{P} = \mathbf{u}(\mathbf{u}^T \mathbf{u})^{-1} \mathbf{u}^T = \frac{\mathbf{u} \mathbf{u}^T}{\mathbf{u}^T \mathbf{u}} = \frac{\mathbf{u} \mathbf{u}^T}{\|\mathbf{u}\|^2},$$

y por lo tanto, la proyección de \mathbf{v} sobre ese subespacio es

$$\text{proj}_{\mathbf{u}}(\mathbf{v}) = \mathbf{P} \mathbf{v} = \frac{\mathbf{u} \mathbf{u}^T}{\|\mathbf{u}\|^2} \mathbf{v} \stackrel{(a)}{=} \frac{\mathbf{u}^T \mathbf{v}}{\|\mathbf{u}\|^2} \mathbf{u}.$$

donde en (a) se tuvo en cuenta que $[\mathbf{u}^T \mathbf{u}]_{ij} = \mathbf{u}_i \mathbf{u}_j$ y por lo tanto

$$[(\mathbf{u}^T \mathbf{u}) \mathbf{v}]_i = \sum_j \mathbf{u}_i \mathbf{u}_j \mathbf{v}_j = \mathbf{u}_i (\mathbf{u}^T \mathbf{v}) = [(\mathbf{u}^T \mathbf{v}) \mathbf{u}]_i.$$

Con estas consideraciones, la proyección de \mathbf{x} sobre \mathbf{h}'_{k+1} es

$$\alpha \mathbf{h}'_{k+1} = \frac{\mathbf{h}_{k+1}^T \mathbf{x}}{\|\mathbf{h}'_{k+1}\|^2} \mathbf{h}'_{k+1},$$

²Ver https://en.wikipedia.org/wiki/Gram-Schmidt_process

y por lo tanto

$$\alpha = \frac{\mathbf{h}_{k+1}^T \mathbf{x}}{\|\mathbf{h}'_{k+1}\|^2} = \frac{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{x}}{\|\mathbf{P}_k^\perp \mathbf{h}_{k+1}\|^2} = \frac{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{x}}{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{P}_k^{\perp T} \mathbf{P}_k^\perp \mathbf{h}_{k+1}}.$$

Además, considerando que \mathbf{P}_k^\perp es una matriz simétrica e idempotente por ser una matriz de proyección, se obtiene que

$$\alpha = \frac{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{x}}{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1}}. \quad (8.44)$$

La señal LS estimada en el paso $k+1$ es

$$\begin{aligned} \hat{\mathbf{s}}_{k+1} &= \hat{\mathbf{s}}_k + \alpha \mathbf{h}'_{k+1} \\ &\stackrel{(a)}{=} \mathbf{H}_k \hat{\boldsymbol{\theta}}_k + \frac{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{x}}{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1}} \mathbf{P}_k^\perp \mathbf{h}_{k+1} \\ &\stackrel{(b)}{=} \mathbf{H}_k \hat{\boldsymbol{\theta}}_k + (\mathbf{I} - \mathbf{P}_k) \mathbf{h}_{k+1} \frac{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{x}}{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1}} \\ &= \mathbf{H}_k \hat{\boldsymbol{\theta}}_k + \frac{\mathbf{h}_{k+1} \mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{x}}{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1}} - \frac{\mathbf{P}_k \mathbf{h}_{k+1} \mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{x}}{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1}} \\ &\stackrel{(c)}{=} \mathbf{H}_k \hat{\boldsymbol{\theta}}_k + \frac{\mathbf{h}_{k+1} \mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{x}}{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1}} - \frac{\mathbf{H}_k (\mathbf{H}_k^T \mathbf{H}_k)^{-1} \mathbf{H}_k^T \mathbf{h}_{k+1} \mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{x}}{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1}} \\ &= [\mathbf{H}_k \quad \mathbf{h}_k] \begin{bmatrix} \hat{\boldsymbol{\theta}}_k - \frac{(\mathbf{H}_k^T \mathbf{H}_k)^{-1} \mathbf{H}_k^T \mathbf{h}_{k+1} \mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{x}}{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1}} \\ \frac{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{x}}{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1}} \end{bmatrix} \\ &= \mathbf{H}_{k+1} \hat{\boldsymbol{\theta}}_{k+1} \end{aligned}$$

donde en (a) se sustituyeron las ecuaciones 8.42, 8.43 y 8.44, en (b) se consideró que $\mathbf{P}_k^\perp = \mathbf{I} - \mathbf{P}_k$ y en (c) que $\mathbf{P}_k = \mathbf{H}_k (\mathbf{H}_k^T \mathbf{H}_k)^{-1} \mathbf{H}_k^T$. Se concluye que $\hat{\boldsymbol{\theta}}_{k+1}$ está dado por la ecuación 8.11.

8.8.15. Problema 15

Emplear el algoritmo LS de orden recursivo para encontrar los valores de A y B que minimizan

$$J = \sum_{n=0}^{N-1} (x[n] - A - Br^n)^2.$$

Se asume que el parámetro r es conocido.

Solución El modelo de señal de orden 1 es

$$s_1[n] = A_1,$$

que en forma matricial se expresa como

$$\mathbf{s}_1 = \mathbf{H}_1 A_1, \quad \text{con} \quad \mathbf{H}_1 = \mathbf{1},$$

por lo que el estimador de orden 1 es

$$\hat{A}_1 = (\mathbf{H}_1^T \mathbf{H}_1)^T \mathbf{H}_1^T \mathbf{x} = (\mathbf{1}^T \mathbf{1})^T \mathbf{1}^T \mathbf{x} = \frac{1}{N} \sum_{n=0}^{N-1} x[n],$$

es decir,

$$\hat{\boldsymbol{\theta}}_1 = \hat{A}_1 = \bar{x}.$$

El modelo de señal de orden 2 es

$$s_2[n] = A_2 + B_2 r^n$$

y en forma matricial se expresa como

$$\mathbf{s}_2 = \mathbf{H}_2 \boldsymbol{\theta}_2, \quad \text{con} \quad \mathbf{H}_2 = \begin{bmatrix} 1 & 1 \\ 1 & r \\ 1 & r^2 \\ \vdots & \vdots \\ 1 & r^{N-1} \end{bmatrix} = [\mathbf{H}_1 \quad \mathbf{h}_2] \quad \text{y} \quad \boldsymbol{\theta}_2 = \begin{bmatrix} A \\ B \end{bmatrix}.$$

La actualización al estimador de orden 2 en el LS recursivo está dado por la ecuación 8.11 y es

$$\hat{\boldsymbol{\theta}}_2 = \begin{bmatrix} \hat{\boldsymbol{\theta}}_1 - \frac{(\mathbf{H}_1^T \mathbf{H}_1)^{-1} \mathbf{H}_1^T \mathbf{h}_2 \mathbf{h}_2^T \mathbf{P}_1^\perp \mathbf{x}}{\mathbf{h}_2^T \mathbf{P}_1^\perp \mathbf{h}_2} \\ \frac{\mathbf{h}_2^T \mathbf{P}_1^\perp \mathbf{x}}{\mathbf{h}_2^T \mathbf{P}_1^\perp \mathbf{h}_2} \end{bmatrix}$$

donde

$$\mathbf{P}_1^\perp = \mathbf{I} - \mathbf{H}_1 (\mathbf{H}_1^T \mathbf{H}_1)^{-1} \mathbf{H}_1^T.$$

Empleando las definiciones, se tiene que

$$(\mathbf{H}_1^T \mathbf{H}_1)^{-1} = \frac{1}{N}, \quad \mathbf{P}_1^\perp = \mathbf{I} - \frac{1}{N} \mathbf{1} \mathbf{1}^T, \quad \mathbf{P}_1^\perp \mathbf{x} = \mathbf{x} - \frac{1}{N} \mathbf{1} \mathbf{1}^T \mathbf{x} = \mathbf{x} - \frac{1}{N} \mathbf{1} \sum_{n=0}^{N-1} x[n] = \mathbf{x} - \mathbf{1} \bar{x}$$

$$\mathbf{h}_2^T \mathbf{P}_1^\perp \mathbf{x} = \mathbf{h}_2^T \mathbf{x} - \mathbf{h}_2^T \mathbf{1} \bar{x} = \sum_{n=0}^{N-1} x[n] r^n - \bar{x} \sum_{n=0}^{N-1} r^n, \quad (\mathbf{H}_1^T \mathbf{H}_1)^{-1} \mathbf{H}_1^T \mathbf{h}_2 = \frac{1}{N} \sum_{n=0}^{N-1} r^n$$

$$\mathbf{h}_2^T \mathbf{P}_1^\perp \mathbf{h}_2 = \mathbf{h}_2^T \mathbf{h}_2 - \frac{1}{N} \mathbf{h}_2^T \mathbf{1} \mathbf{1}^T \mathbf{h}_2 = \mathbf{h}_2^T \mathbf{h}_2 - \frac{1}{N} (\mathbf{h}_2^T \mathbf{1})^2 = \sum_{n=0}^{N-1} r^{2n} - \frac{1}{N} \left(\sum_{n=0}^{N-1} r^n \right)^2,$$

y sustituyendo en la actualización del estimador, se obtiene que

$$\hat{\boldsymbol{\theta}}_2 = \begin{bmatrix} \bar{x} - \frac{\frac{1}{N} \sum_{n=0}^{N-1} r^n \left(\sum_{n=0}^{N-1} x[n] r^n - \bar{x} \sum_{n=0}^{N-1} r^n \right)}{\sum_{n=0}^{N-1} r^{2n} - \frac{1}{N} \left(\sum_{n=0}^{N-1} r^n \right)^2} \\ \frac{\sum_{n=0}^{N-1} x[n] r^n - \bar{x} \sum_{n=0}^{N-1} r^n}{\sum_{n=0}^{N-1} r^{2n} - \frac{1}{N} \left(\sum_{n=0}^{N-1} r^n \right)^2} \end{bmatrix},$$

es decir,

$$\hat{A}_2 = \bar{x} - \frac{1}{N} \sum_{n=0}^{N-1} r^n \hat{B}_2, \quad \hat{B}_2 = \frac{\sum_{n=0}^{N-1} x[n] r^n - \bar{x} \sum_{n=0}^{N-1} r^n}{\sum_{n=0}^{N-1} r^{2n} - \frac{1}{N} \left(\sum_{n=0}^{N-1} r^n \right)^2}.$$

8.8.16. Problema 16

Se consideran los modelos

$$\begin{aligned} s_1[n] &= A \quad 0 \leq n \leq N-1 \\ s_2[n] &= \begin{cases} A & 0 \leq n \leq M-1 \\ B & M \leq n \leq N-1. \end{cases} \end{aligned}$$

La segunda señal es útil para modelar un salto del nivel en $n = M$. Alternativamente, el segundo modelo se puede expresar como

$$s_2[n] = Au_1[n] + (B - A)u_2[n]$$

donde

$$\begin{aligned} u_1[n] &= 1 \quad 0 \leq n \leq N-1 \\ u_2[n] &= \begin{cases} 0 & 0 \leq n \leq M-1 \\ 1 & M \leq n \leq N-1. \end{cases} \end{aligned}$$

Sea $\theta_1 = A$ y $\theta_2 = [A \ (B - A)]^T$. Encontrar el LSE y el error LS mínimo para cada modelo empleando el LS de orden recursivo. Discutir como podría emplear los resultados para detectar un salto en el nivel.

Solución El modelo la primer señal es

$$\mathbf{s}_1 = \mathbf{H}_1 A_1, \quad \text{con} \quad \mathbf{H}_1 = \mathbf{1},$$

por lo que, al igual que el problema anterior, el estimador

$$\hat{\theta}_1 = \hat{A}_1 = \bar{x}.$$

El modelo de señal de orden 2 es

$$\mathbf{s}_2 = [\mathbf{H}_1 \quad \mathbf{h}_2] \theta_2, \quad \text{con} \quad \mathbf{h}_2 = \begin{bmatrix} 0 & \dots & 0 & 1 & \dots & 1 \end{bmatrix}^T \quad \text{y} \quad \theta_2 = \begin{bmatrix} A \\ B - A \end{bmatrix}.$$

$n=0$ $n=M$

La actualización al estimador de orden 2 en el LS recursivo está dado por la ecuación 8.11 con

$$(\mathbf{H}_1^T \mathbf{H}_1)^{-1} = \frac{1}{N}, \quad \mathbf{P}_1^\perp = \mathbf{I} - \frac{1}{N} \mathbf{1} \mathbf{1}^T, \quad \mathbf{P}_1^\perp \mathbf{x} = \mathbf{x} - \frac{1}{N} \mathbf{1} \mathbf{1}^T \mathbf{x} = \mathbf{x} - \frac{1}{N} \mathbf{1} \sum_{n=0}^{N-1} x[n] = \mathbf{x} - \mathbf{1} \bar{x}$$

ya que $\mathbf{H}_1 = \mathbf{1}$ al igual que el problema anterior. Además, considerando la definición de \mathbf{h}_2 ,

$$\mathbf{h}_2^T \mathbf{P}_1^\perp \mathbf{x} = \mathbf{h}_2^T \mathbf{x} - \mathbf{h}_2^T \mathbf{1} \bar{x} = \sum_{n=M}^{N-1} x[n] - (N - M) \bar{x}, \quad (\mathbf{H}_1^T \mathbf{H}_1)^{-1} \mathbf{H}_1^T \mathbf{h}_2 = \frac{N - M}{N},$$

$$\mathbf{h}_2^T \mathbf{P}_1^\perp \mathbf{h}_2 = \mathbf{h}_2^T \mathbf{h}_2 - \frac{1}{N} \mathbf{h}_2^T \mathbf{1} \mathbf{1}^T \mathbf{h}_2 = \mathbf{h}_2^T \mathbf{h}_2 - \frac{1}{N} (\mathbf{h}_2^T \mathbf{1})^2 = (N - M) - \frac{1}{N} (N - M)^2 = \frac{M}{N} (N - M).$$

Por lo tanto, el estimador es

$$\begin{aligned}
\hat{A}_2 &= \hat{\theta}_1 - \frac{(\mathbf{H}_1^T \mathbf{H}_1)^{-1} \mathbf{H}_1^T \mathbf{h}_2 \mathbf{h}_2^T \mathbf{P}_1^\perp \mathbf{x}}{\mathbf{h}_2^T \mathbf{P}_1^\perp \mathbf{h}_2} \\
&= \bar{x} - \frac{\frac{N-M}{N} \left[\sum_{n=M}^{N-1} x[n] - (N-M)\bar{x} \right]}{\frac{M}{N}(N-M)} \\
&= \bar{x} - \frac{1}{M} \sum_{n=M}^{N-1} x[n] + \frac{N-M}{M} \bar{x} \\
&= -\frac{1}{M} \sum_{n=M}^{N-1} x[n] + \frac{N}{M} \bar{x} \\
&= -\frac{1}{M} \sum_{n=M}^{N-1} x[n] + \frac{1}{M} \left(\sum_{n=0}^{M-1} x[n] + \sum_{n=M}^{N-1} x[n] \right) \\
&= \frac{1}{M} \sum_{n=0}^{M-1} x[n]
\end{aligned}$$

resultando en

$$\hat{A}_2 = \frac{1}{M} \sum_{n=0}^{M-1} x[n],$$

y

$$\begin{aligned}
\widehat{B-A} &= \frac{\mathbf{h}_2^T \mathbf{P}_1^\perp \mathbf{x}}{\mathbf{h}_2^T \mathbf{P}_1^\perp \mathbf{h}_2} \\
&= \frac{\sum_{n=M}^{N-1} x[n] - (N-M)\bar{x}}{\frac{M}{N}(N-M)} \\
&= \frac{M}{N} \left(\frac{1}{N-M} \sum_{n=M}^{N-1} x[n] - \bar{x} \right),
\end{aligned}$$

resultando en

$$\widehat{B-A} = \frac{M}{N}(\bar{x}_1 - \bar{x}), \quad \text{con} \quad \bar{x}_1 = \frac{1}{N-M} \sum_{n=M}^{N-1} x[n].$$

De la ecuación 8.6, el error mínimo en el estimador de primer orden es

$$\begin{aligned}
J_{\min_1} &= \mathbf{x}^T \mathbf{P}_1^\perp \mathbf{x} \\
&= \mathbf{x}^T (\mathbf{x} - \mathbf{1}\bar{x}) \\
&= \sum_{n=0}^{N-1} x^2[n] - \bar{x} \sum_{n=0}^{N-1} x[n],
\end{aligned}$$

resultando en

$$J_{\min_1} = \sum_{n=0}^{N-1} x^2[n] - N\bar{x}^2.$$

La actualización del error LS para el modelo de segundo orden está dado por la ecuación 8.15,

$$J_{\min_2} = J_{\min_1} - \frac{(\mathbf{h}_2^T \mathbf{P}_1^\perp \mathbf{x})^2}{\mathbf{h}_2^T \mathbf{P}_1^\perp \mathbf{h}_2}.$$

El decremento del error es

$$\begin{aligned}
\frac{(\mathbf{h}_2^T \mathbf{P}_1^\perp \mathbf{x})^2}{\mathbf{h}_2^T \mathbf{P}_1^\perp \mathbf{h}_2} &= (\mathbf{h}_2^T \mathbf{P}_1^\perp \mathbf{x}) \frac{(\mathbf{h}_2^T \mathbf{P}_1^\perp \mathbf{x})}{\mathbf{h}_2^T \mathbf{P}_1^\perp \mathbf{h}_2} \\
&\stackrel{(a)}{=} (N - M)(\bar{x}_1 - \bar{x}) \widehat{B - A} \\
&= (N - M)(\bar{x}_1 - \bar{x}) \frac{M}{N}(\bar{x}_1 - \bar{x}) \\
\\
\frac{(\mathbf{h}_2^T \mathbf{P}_1^\perp \mathbf{x})^2}{\mathbf{h}_2^T \mathbf{P}_1^\perp \mathbf{h}_2} &= (\mathbf{h}_2^T \mathbf{P}_1^\perp \mathbf{x}) \frac{(\mathbf{h}_2^T \mathbf{P}_1^\perp \mathbf{x})}{\mathbf{h}_2^T \mathbf{P}_1^\perp \mathbf{h}_2} \\
&= \left(\sum_{n=M}^{N-1} x[n] - (N - M)\bar{x} \right) \widehat{B - A} \\
&= (N - M)(\bar{x}_1 - \bar{x}) \frac{M}{N}(\bar{x}_1 - \bar{x})
\end{aligned}$$

resultando en

$$J_{\min_2} = J_{\min_1} - \frac{M}{N}(N - M)(\bar{x}_1 - \bar{x})^2.$$

Un salto en el nivel en $n = M$ se manifiesta en un decrecimiento grande del error J_{\min_2} del modelo de segundo orden. Efectivamente, si no hay un cambio de nivel $\bar{x}_1 \approx \bar{x}$ y por lo tanto $J_{\min_2} \approx J_{\min_1}$, mientras si hay un cambio de nivel, $(\bar{x}_1 - \bar{x})^2$ tendrá un valor significativo.

8.8.17. Problema 17

Probar que $\mathbf{P}_k^\perp \mathbf{x}$ es ortogonal al subespacio generado por las columnas de \mathbf{H}_k .

Solución Hay que probar que

$$\mathbf{H}_k^T \mathbf{P}_k^\perp \mathbf{x} = \mathbf{0}.$$

Considerando la definición de la matriz de proyección \mathbf{P}_k^\perp , dada por la ecuación 8.12, se tiene que

$$\begin{aligned}
\mathbf{H}_k^T \mathbf{P}_k^\perp \mathbf{x} &= \mathbf{H}_k(\mathbf{I} - \mathbf{H}_k(\mathbf{H}_k^T \mathbf{H}_k)^{-1} \mathbf{H}_k^T) \mathbf{x} \\
&= \mathbf{H}_k^T \mathbf{x} - \mathbf{H}_k^T \mathbf{H}_k(\mathbf{H}_k^T \mathbf{H}_k)^{-1} \mathbf{H}_k^T \mathbf{x} \\
&= \mathbf{H}_k^T \mathbf{x} - \mathbf{H}_k^T \mathbf{x} \\
&= \mathbf{0}.
\end{aligned}$$

8.8.18. Problema 18

Empleando la matriz de proyección ortogonal recursiva (ecuación 8.17) derivar la actualización del error LS mínimo (ecuación 8.15).

Solución Partiendo de la ecuación 8.6 del error LS mínimo, se tiene que

$$\begin{aligned}
J_{\min_{k+1}} &= \mathbf{x}^T (\mathbf{I} - \mathbf{P}_{k+1}) \mathbf{x} \\
&= \mathbf{x}^T \mathbf{x} - \mathbf{x}^T \mathbf{P}_{k+1} \mathbf{x} \\
&= \mathbf{x}^T \mathbf{x} - \mathbf{x}^T \left[\mathbf{P}_k + \frac{(\mathbf{I} - \mathbf{P}_k) \mathbf{h}_{k+1} \mathbf{h}_{k+1}^T (\mathbf{I} - \mathbf{P}_k)}{\mathbf{h}_{k+1}^T (\mathbf{I} - \mathbf{P}_k) \mathbf{h}_{k+1}} \right] \mathbf{x} \\
&= \mathbf{x}^T \mathbf{x} - \mathbf{x}^T \mathbf{P}_k \mathbf{x} - \frac{\mathbf{x}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1} \mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{x}}{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1}} \\
&= J_{\min_k} - \frac{(\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{x})^2}{\mathbf{h}_{k+1}^T \mathbf{P}_k^\perp \mathbf{h}_{k+1}}.
\end{aligned}$$

8.8.19. Problema 19

Verificar la ecuación 8.21 del cálculo secuencial del error LS mínimo empleando la ecuación 8.18 de la actualización del LSE.

Solución En el problema de la estimación del nivel de DC en ruido no correlacionado el LSE se obtiene minimizando el error LS ponderado (ver el problema 8.8.8)

$$J(A) = \sum_{n=0}^{N-1} \frac{1}{\sigma_n^2} (x[n] - A)^2.$$

Para obtener la expresión recursiva del error LS mínimo se observa que

$$\begin{aligned} J_{\min}[N] &= \sum_{n=0}^N \frac{1}{\sigma_n^2} (x[n] - \hat{A}[N])^2 \\ &= \sum_{n=0}^{N-1} \frac{1}{\sigma_n^2} (x[n] - \hat{A}[N])^2 + \frac{1}{\sigma_N^2} (x[N] - \hat{A}[N])^2 \\ &\stackrel{(a)}{=} \sum_{n=0}^{N-1} \frac{1}{\sigma_n^2} \left[x[n] - \hat{A}[N-1] - K[N](x[N] - \hat{A}[N-1]) \right]^2 \\ &\quad + \frac{1}{\sigma_N^2} \left[x[N] - \hat{A}[N-1] - K[N](x[N] - \hat{A}[N-1]) \right]^2 \\ &= \sum_{n=0}^{N-1} \frac{1}{\sigma_n^2} (x[n] - \hat{A}[N-1])^2 + 2K[N](x[N] - \hat{A}[N-1]) \sum_{n=0}^{N-1} \frac{1}{\sigma_n^2} (x[n] - \hat{A}[N-1]) \\ &\quad + K^2[N](x[N] - \hat{A}[N-1])^2 \sum_{n=0}^{N-1} \frac{1}{\sigma_n^2} + \frac{1}{\sigma_N^2} \left[x[N] - \hat{A}[N-1] - K[N](x[N] - \hat{A}[N-1]) \right]^2, \end{aligned}$$

donde en (a) se substituyó $\hat{A}[N]$ empleando la ecuación 8.18. El primer sumando es $J_{\min}[N-1]$ y la sumatoria del segundo sumando es nula (ver el problema 8.8.8). Operando con los dos últimos sumandos se tiene que

$$\begin{aligned} &K^2[N](x[N] - \hat{A}[N-1])^2 \sum_{n=0}^{N-1} \frac{1}{\sigma_n^2} + \frac{1}{\sigma_N^2} \left[x[N] - \hat{A}[N-1] - K[N](x[N] - \hat{A}[N-1]) \right]^2 \\ &\stackrel{(a)}{=} \frac{K^2[N]}{\text{var}(\hat{A}[N-1])} (x[N] - \hat{A}[N-1])^2 + \frac{1}{\sigma_N^2} (x[N] - \hat{A}[N-1])^2 (1 - K[N])^2 \\ &= (x[N] - \hat{A}[N-1])^2 \left[\frac{K^2[N]}{\text{var}(\hat{A}[N-1])} + \frac{1}{\sigma_N^2} (1 - K[N])^2 \right] \\ &\stackrel{(b)}{=} (x[N] - \hat{A}[N-1])^2 \left[\frac{\text{var}(\hat{A}[N-1])}{(\text{var}(\hat{A}[N-1]) + \sigma_N^2)^2} + \frac{\sigma_N^2}{(\text{var}(\hat{A}[N-1]) + \sigma_N^2)^2} \right] \\ &= \frac{(x[N] - \hat{A}[N-1])^2}{\text{var}(\hat{A}[N-1]) + \sigma_N^2}, \end{aligned}$$

donde en (a) se empleó que

$$\text{var}(\hat{A}[N-1]) = \frac{1}{\sum_{n=0}^{N-1} \frac{1}{\sigma_n^2}},$$

como se encontró en el problema de la sección 8.8.8 y en (b) se empleó la ecuación 8.19, que indica que

$$K[N] = \frac{\text{var}(\hat{A}[N-1])}{\text{var}(\hat{A}[N-1]) + \sigma_N^2} \quad \Rightarrow \quad K[N] - 1 = \frac{\sigma_N^2}{\text{var}(\hat{A}[N-1]) + \sigma_N^2}.$$

Se concluye que

$$J_{\min}[N] = J_{\min}[N-1] + \frac{(x[N] - \hat{A}[N-1])^2}{\text{var}(\hat{A}[N-1]) + \sigma_N^2},$$

que es lo que se quería probar.

8.8.20. Problema 20

Sea $x[n] = Ar^n + w[n]$, donde $w[n]$ es WGN de varianza $\sigma^2=1$. Encontrar el LSE secuencial de A asumiendo que r es conocido. Además, determinar la varianza del LSE de forma secuencial y luego resolver para obtener una función explícita de n para la varianza. Tómese $\hat{A}[0] = x[0]$ y $\text{var}(\hat{A}[0]) = \sigma^2 = 1$.

Solución El LSE secuencial está dado por las ecuaciones 8.25-8.27. En este caso, el modelo de la señal es

$$\mathbf{s}[n] = \mathbf{H}[n]A, \quad \text{con} \quad \mathbf{H}[n] = [1 \quad r \quad r^2 \quad \dots \quad r^n]^T,$$

y por lo tanto,

$$\mathbf{h}[n] = r^n$$

escalar. Además, como el estimador es escalar, la matriz de covarianza es la varianza del estimador,

$$\Sigma[n] = \text{var}(\hat{A}[n]).$$

Por lo tanto, el LSE secuencial es

$$\hat{A}[n] = \hat{A}[n-1] + K[n] \left(x[n] - \hat{A}[n-1]r^n \right)$$

donde

$$K[n] = \frac{\text{var}(\hat{A}[n-1])r^n}{1 + \text{var}(\hat{A}[n-1])r^{2n}}$$

y

$$\text{var}(\hat{A}[n]) = (1 - K[n]r^n) \text{var}(\hat{A}[n-1]).$$

Para resolver la recursión de la varianza, defínase $v_n = \text{var}(\hat{A}[n])$. De esta forma, sustituyendo la ganancia en la ecuación de la recursión de la varianza, se tiene que

$$\begin{aligned} v_n &= \left(1 - \frac{v_{n-1}r^{2n}}{1 + v_{n-1}r^{2n}} \right) v_{n-1} \\ &= \frac{v_{n-1}}{1 + v_{n-1}r^{2n}}. \end{aligned}$$

Para simplificar las cuentas, considérese

$$\frac{1}{v_n} = \frac{1 + v_{n-1}r^{2n}}{v_{n-1}} = \frac{1}{v_{n-1}} + r^{2n}.$$

Los primeros términos de la recursión son

$$\frac{1}{v_0} = 1, \quad \frac{1}{v_1} = \frac{1}{v_0} + r^2 = 1 + r^2, \quad \frac{1}{v_2} = \frac{1}{v_1} + r^4 = 1 + r^2 + r^4, \quad \frac{1}{v_3} = \frac{1}{v_2} + r^6 = 1 + r^2 + r^4 + r^6$$

y en general se cumple que

$$\frac{1}{v_n} = \sum_{k=0}^n r^{2k},$$

concluyendo que

$$v_n = \text{var}(\hat{A}[n]) = \frac{1}{\sum_{k=0}^n r^{2k}}.$$

8.8.21. Problema 21

Empleando las ecuaciones 8.19 y 8.20 de la actualización de la ganancia y la varianza respectivamente del LSE secuencial, resolver para las secuencias ganancia y varianza si $\sigma_n^2 = r^n$. Emplear $\text{var}(\hat{A}[0]) = \text{var}(x[0]) = \sigma_0^2 = 1$ para la inicialización. Luego, examinar que ocurre con la ganancia y la varianza cuando $N \rightarrow \infty$ si $r = 1$, $0 < r < 1$ y $r > 1$. Sugerencia: resolver para $1/\text{var}(\hat{A}[N])$.

Solución Las ecuaciones de la ganancia y la varianza son respectivamente

$$K[N] = \frac{\text{var}(\hat{A}[N-1])}{\text{var}(\hat{A}[N-1]) + \sigma_N^2} \quad \text{y} \quad \text{var}(\hat{A}[N]) = (1 - K[N]) \text{var}(\hat{A}[N-1]).$$

Sea $v_N = \text{var}(\hat{A}[N])$. Sustituyendo la ecuación de la ganancia en la ecuación de la varianza y considerando que $\sigma_N^2 = r^N$ se tiene que

$$v_N = \left(1 - \frac{v_{N-1}}{v_{N-1} + r^N}\right) v_{N-1} = \frac{v_{N-1} r^N}{v_{N-1} + r^N},$$

y por lo tanto,

$$\frac{1}{v_N} = \frac{v_{N-1} + r^N}{v_{N-1} r^N} = \frac{1}{r^N} + \frac{1}{v_{N-1}}.$$

Los primeros términos de la recursión son

$$\frac{1}{v_0} = 1, \quad \frac{1}{v_1} = \frac{1}{r} + \frac{1}{v_0} = \frac{1}{r} + 1, \quad \frac{1}{v_2} = \frac{1}{r^2} + \frac{1}{v_1} = \frac{1}{r^2} + \frac{1}{r} + 1, \quad \frac{1}{v_3} = \frac{1}{r^3} + \frac{1}{v_2} = \frac{1}{r^3} + \frac{1}{r^2} + \frac{1}{r} + 1,$$

y generalizando, se obtiene que

$$\frac{1}{v_N} = \sum_{n=0}^N r^{-n}.$$

Multiplicando y dividiendo el lado derecho de la igualdad entre r^N , el inverso de la varianza se puede expresar como

$$\frac{1}{v_N} = \frac{\sum_{n=0}^N r^{N-n}}{r^N} = \frac{\sum_{k=0}^N r^k}{r^N},$$

donde en la segunda igualdad se realizó el cambio de variable $k = N - n$. Se concluye que la secuencia de la varianza es

$$v_N = \text{var}(\hat{A}[N]) = \frac{r^N}{\sum_{n=0}^N r^n}.$$

Sustituyendo este resultado en la ecuación de la ganancia, se tiene que

$$K[N] = \frac{\frac{r^{N-1}}{\sum_{n=0}^{N-1} r^n}}{\frac{r^{N-1}}{\sum_{n=0}^{N-1} r^n} + r^N} = \frac{r^{N-1}}{r^{N-1} + r^N \sum_{n=0}^{N-1} r^n} = \frac{1}{1 + r \sum_{n=0}^{N-1} r^n} = \frac{1}{1 + \sum_{n=0}^{N-1} r^{n+1}} = \frac{1}{1 + \sum_{k=1}^N r^k},$$

donde en la última igualdad se realizó el cambio de variable $k = n + 1$. Se concluye que

$$K[N] = \frac{1}{\sum_{n=0}^N r^n}.$$

A continuación se realiza el estudio de la convergencia.

- En el caso en que $r = 1$,

$$\text{var}(\hat{A}[N]) = \frac{1}{N+1} \xrightarrow{N \rightarrow \infty} 0 \quad \text{y} \quad K[N] = \frac{1}{N+1} \xrightarrow{N \rightarrow \infty} 0.$$

- Si $0 < r < 1$, considerando que

$$\sum_{n=0}^{\infty} r^n = \frac{1}{1-r}, \quad \text{si } |r| < 1$$

se tiene que

$$\text{var}(\hat{A}[N]) = \frac{r^N}{\sum_{n=0}^N r^n} \xrightarrow{N \rightarrow \infty} \frac{0}{\sum_{n=0}^{\infty} r^n} = 0(1-r) = 0 \quad \text{y} \quad K[N] = \frac{1}{\sum_{n=0}^N r^n} \xrightarrow{N \rightarrow \infty} \frac{1}{\sum_{n=0}^{\infty} r^n} = 1-r.$$

- Si $r > 1$,

$$\text{var}(\hat{A}[N]) = \frac{1}{\sum_{n=0}^N \left(\frac{1}{r}\right)^n} \xrightarrow{N \rightarrow \infty} \frac{1}{\sum_{n=0}^{\infty} \left(\frac{1}{r}\right)^n} = \frac{1}{\frac{1}{1-\frac{1}{r}}} = \frac{r-1}{r}$$

y

$$K[N] = \frac{1}{\sum_{n=0}^N r^n} \xrightarrow{N \rightarrow \infty} \frac{1}{\sum_{n=0}^{\infty} r^n} = 0.$$

En este caso, como $\sigma_N^2 = r^N$ con $r > 1$, los datos son cada vez mas ruidosos al crecer N . De forma acorde, $K[N] \rightarrow 0$, que implica que cada vez se da menos peso a los datos nuevos para calcular el estimador. Asimismo, la varianza no se anula.

8.8.22. Problema 22

Realizar una simulación por computadora para graficar $\hat{A}[N]$ dado por la ecuación 8.18. Asumir que el modelo de los datos es

$$x[n] = A + w[n]$$

donde $w[n]$ es WGN con $\sigma_n^2 = r^n$. Inicializar el estimador empleando $\hat{A}[0]$ y $\text{var}(\hat{A}[0]) = \text{var}(x[0]) = \sigma_0^2 = 1$. Además, graficar las secuencias de la ganancia y la varianza.

Solución El análisis teórico de este problema se realizó en el problema anterior. En la figura 8.14 se muestra el resultado de la simulación. Se empleó la misma secuencia de ruido con el escalado variable apropiado para generar $w[n]$ con varianza r^n para los distintos valores de $r[n]$. Concretamente, se generó WGN $v[n] \sim \mathcal{N}(0, 1)$, y luego $w[n] = a[n]v[n]$, donde $a[n]$ es una secuencia determinística. Como

$$\text{var}(w[n]) = \text{var}(a[n]v[n]) = a^2[n] \text{var}(v[n]) = a^2[n]$$

y se quiere que $\text{var}(w[n]) = r^n$, $a^2[n] = r^n$, por lo que $a[n] = r^{n/2}$. Como se dedujo en el problema anterior, cuando $0 < r < 1$, la varianza del ruido decrece y la varianza del estimador tiende a cero. Esto implica que la estimación es perfecta con $N \rightarrow \infty$. Además, $K[N] \rightarrow 1-r$ con $N \rightarrow \infty$. En el caso en que $r > 1$, la potencia del ruido crece. La ganancia tiende a cero rápidamente, lo que significa que las muestras se toman cada vez menos en cuenta cuando N crece. Esto puede verse como emplear un número finito de muestras para calcular el estimador. Por esta razón, la varianza del estimador no tiende a cero.

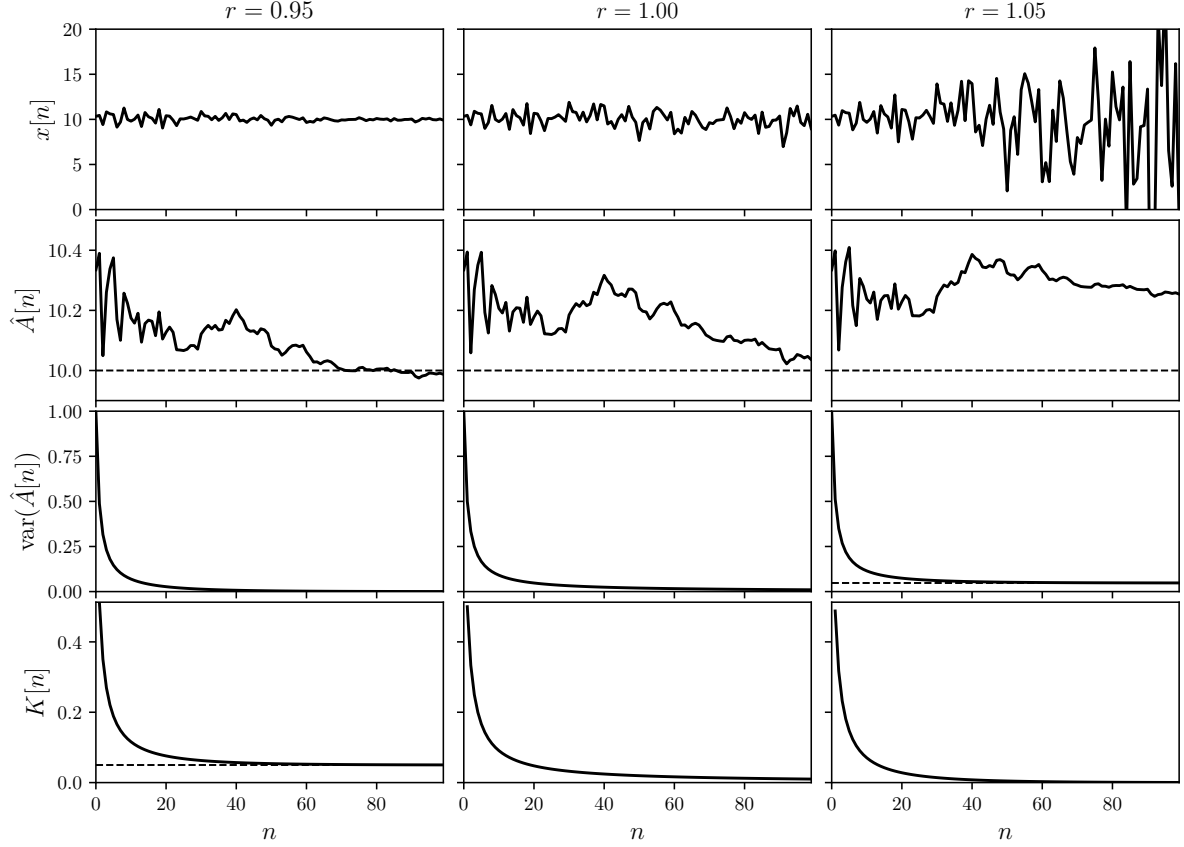


Figura 8.14: Estimador LSE secuencial de A , con $x[n] = A + w[n]$. La varianza del ruido $w[n]$ es r^n . Se muestran los resultados para $r = 0.95, 1, 1.05$. Con $r > 1$, la varianza del estimador no tiende a cero, por lo que la estimación no alcanza nunca el valor correcto.

8.8.23. Problema 23

En este problema se examina la inicialización del LSE secuencial. Asíumase que se elige $\hat{\boldsymbol{\theta}}[-1]$ y $\boldsymbol{\Sigma}[-1] = \alpha \mathbf{I}$ para inicializar el LSE. Se mostrará que si $\alpha \rightarrow \infty$, el estimador LSE por lotes

$$\begin{aligned} \hat{\boldsymbol{\theta}}_B[n] &= (\mathbf{H}^T[n] \mathbf{C}^{-1}[n] \mathbf{H}[n])^{-1} \mathbf{H}^T[n] \mathbf{C}^{-1}[n] \mathbf{x}[n] \\ &= \left(\sum_{k=0}^n \frac{1}{\sigma_k^2} \mathbf{h}[k] \mathbf{h}^T[k] \right)^{-1} \left(\sum_{k=0}^n \frac{1}{\sigma_k^2} x[k] \mathbf{h}[k] \right) \end{aligned}$$

es idéntico al estimador secuencial para $n \geq p$. Primero, asíumase que los vectores de observación $\{\mathbf{h}[-p], \mathbf{h}[-(p-1)], \dots, \mathbf{h}[-1]\}$ y las varianzas del ruido $\{\sigma_{-p}^2, \sigma_{-(p-1)}^2, \dots, \sigma_{-1}^2\}$ existen, de forma que para el LSE y matriz de covarianza inicial elegidos es posible calcular el LSE por lotes. Por lo tanto,

$$\hat{\boldsymbol{\theta}}[-1] = \hat{\boldsymbol{\theta}}_B[-1] = (\mathbf{H}^T[-1] \mathbf{C}^{-1}[-1] \mathbf{H}[-1])^{-1} \mathbf{H}^T[-1] \mathbf{C}^{-1}[-1] \mathbf{x}[-1]$$

y

$$\boldsymbol{\Sigma}[-1] = (\mathbf{H}^T[-1] \mathbf{C}^{-1}[-1] \mathbf{H}[-1])^{-1}$$

donde

$$\mathbf{H}[-1] = \begin{bmatrix} \mathbf{h}^T[-p] \\ \mathbf{h}^T[-(p-1)] \\ \vdots \\ \mathbf{h}^T[-1] \end{bmatrix}$$

$$\mathbf{C}[-1] = \text{diag}(\sigma_{-p}^2, \sigma_{-(p-1)}^2, \dots, \sigma_{-1}^2).$$

De esta forma, se puede ver al estimador inicial del LSE secuencial como el resultado de aplicar un estimador por lotes con los vectores de observación iniciales. Debido a que el LSE secuencial inicializado mediante el estimador por lotes es idéntico al estimador por lotes empleando todos los vectores de observación, se cumple que el LSE secuencial con las condiciones iniciales asumidas es

$$\hat{\boldsymbol{\theta}}_S[n] = \left(\sum_{k=-p}^n \frac{1}{\sigma_k^2} \mathbf{h}[k] \mathbf{h}^T[k] \right)^{-1} \left(\sum_{k=-p}^n \frac{1}{\sigma_k^2} x[k] \mathbf{h}[k] \right). \quad (8.45)$$

Mostrar que esto puede reescribirse como

$$\hat{\boldsymbol{\theta}}_S[n] = \left(\boldsymbol{\Sigma}^{-1}[-1] + \sum_{k=0}^n \frac{1}{\sigma_k^2} \mathbf{h}[k] \mathbf{h}^T[k] \right)^{-1} \left(\boldsymbol{\Sigma}^{-1}[-1] + \sum_{k=0}^n \frac{1}{\sigma_k^2} x[k] \mathbf{h}[k] \right).$$

Luego, examinar que ocurre cuando $\alpha \rightarrow \infty$ para $0 \leq n \leq p-1$ y para $n \geq p$,

Solución Las matrices de observación y de covarianza del ruido del estimador por lotes con las condiciones iniciales indicadas son

$$\mathbf{H}[n] = \begin{bmatrix} \mathbf{h}^T[-p] \\ \vdots \\ \mathbf{h}^T[-1] \\ \mathbf{h}^T[0] \\ \vdots \\ \mathbf{h}^T[n] \end{bmatrix}$$

$$\mathbf{C}[n] = \text{diag}(\sigma_{-p}^2, \dots, \sigma_{-1}^2, \sigma_0^2, \dots, \sigma_n^2).$$

y por lo tanto, al aplicar la ecuación 8.23 se observa que efectivamente el estimador secuencial está dado por la ecuación 8.45. Esta ecuación se puede escribir como

$$\hat{\boldsymbol{\theta}}_S[n] = \left(\sum_{k=-p}^{-1} \frac{1}{\sigma_k^2} \mathbf{h}[k] \mathbf{h}^T[k] + \sum_{k=0}^n \frac{1}{\sigma_k^2} \mathbf{h}[k] \mathbf{h}^T[k] \right)^{-1} \left(\sum_{k=-p}^{-1} \frac{1}{\sigma_k^2} x[k] \mathbf{h}[k] + \sum_{k=0}^n \frac{1}{\sigma_k^2} x[k] \mathbf{h}[k] \right).$$

Por otro lado, se cumple que

$$\begin{aligned} \hat{\boldsymbol{\theta}}[-1] &= (\mathbf{H}^T[-1] \mathbf{C}^{-1}[-1] \mathbf{H}[-1])^{-1} \mathbf{H}^T[-1] \mathbf{C}^{-1}[-1] \mathbf{x}[-1] \\ &= \boldsymbol{\Sigma}^{-1}[-1] \mathbf{H}^T[-1] \mathbf{C}^{-1}[-1] \mathbf{x}[-1], \end{aligned}$$

y por lo tanto,

$$\begin{aligned} \boldsymbol{\Sigma}^{-1}[-1] \hat{\boldsymbol{\theta}}[-1] &= \mathbf{H}^T[-1] \mathbf{C}^{-1}[-1] \mathbf{x}[-1] \\ &= \sum_{k=-p}^{-1} \frac{1}{\sigma_k^2} x[k] \mathbf{h}[k]. \end{aligned}$$

Además,

$$\begin{aligned}\boldsymbol{\Sigma}^{-1}[-1] &= \mathbf{H}^T[-1]\mathbf{C}^{-1}[-1]\mathbf{H}[-1] \\ &= \sum_{k=-p}^{-1} \frac{1}{\sigma_k^2} \mathbf{h}[h]\mathbf{h}^T[k].\end{aligned}$$

Combinando estos resultados, se obtiene que

$$\hat{\boldsymbol{\theta}}_S[n] = \left(\boldsymbol{\Sigma}^{-1}[-1] + \sum_{k=0}^n \frac{1}{\sigma_k^2} \mathbf{h}[h]\mathbf{h}^T[k] \right)^{-1} \left(\boldsymbol{\Sigma}^{-1}[-1]\hat{\boldsymbol{\theta}}[-1] + \sum_{k=0}^n \frac{1}{\sigma_k^2} x[h]\mathbf{h}[k] \right),$$

que es lo que se quería mostrar. Asíumase ahora que se elige $\boldsymbol{\Sigma}[-1] = \alpha \mathbf{I}$ para inicializar el LSE. Si $\alpha \rightarrow \infty$,

$$\boldsymbol{\Sigma}^{-1}[-1] = \frac{\mathbf{I}}{\alpha} \rightarrow \mathbf{0}$$

y por lo tanto,

$$\left(\boldsymbol{\Sigma}^{-1}[-1] + \sum_{k=0}^n \frac{1}{\sigma_k^2} \mathbf{h}[h]\mathbf{h}^T[k] \right)^{-1} \xrightarrow{\alpha \rightarrow \infty} \left(\sum_{k=0}^n \frac{1}{\sigma_k^2} \mathbf{h}[h]\mathbf{h}^T[k] \right)^{-1} = (\mathbf{H}^T[n]\mathbf{C}^{-1}[n]\mathbf{H}[n])^{-1}.$$

$\mathbf{H}[n]$ es una matriz $(n+1) \times p$. En el caso en que $n+1 < p$ se cumple que

$$\text{rank}(\mathbf{H}[n]) \leq n+1 < p.$$

Por lo tanto,

$$\text{rank}(\mathbf{H}^T[n]\mathbf{C}^{-1}[n]\mathbf{H}[n]) \stackrel{(a)}{=} \text{rank}(\mathbf{H}^T[n]\mathbf{H}[n]) \stackrel{(b)}{=} \text{rank}(\mathbf{H}[n]) \leq n+1 < p$$

donde en (a) se consideró que $\mathbf{C}^{-1}[n]$ es una matriz diagonal y por lo tanto, de rango completo, y en (b) que $\text{rank}(\mathbf{A}^T\mathbf{A}) = \text{rank}(\mathbf{A})$, como se explica en el apéndice H. Teniendo en cuenta que $\mathbf{H}^T[n]\mathbf{C}^{-1}[n]\mathbf{H}[n]$ es una matriz $p \times p$ con $\text{rank}(\mathbf{H}^T[n]\mathbf{C}^{-1}[n]\mathbf{H}[n]) < p$, se concluye que no es invertible y por lo tanto, no es posible tomar $\alpha \rightarrow \infty$ si $n+1 < p$. En el caso en que $n+1 \geq p$, $\mathbf{H}^T[n]\mathbf{C}^{-1}[n]\mathbf{H}[n]$ es invertible (asumiendo que las p columnas son linealmente independientes) y se cumple que

$$\hat{\boldsymbol{\theta}}_S[n] \xrightarrow{\alpha \rightarrow \infty} \left(\sum_{k=0}^n \frac{1}{\sigma_k^2} \mathbf{h}[h]\mathbf{h}^T[k] \right)^{-1} \left(\sum_{k=0}^n \frac{1}{\sigma_k^2} x[h]\mathbf{h}[k] \right) = \hat{\boldsymbol{\theta}}_B[n],$$

que es lo que se quería mostrar.

8.8.24. Problema 24

En el ejemplo de la sección 8.5 de la estimación de una señal restringida, determinar $\hat{\mathbf{s}}_c$ primero proyectando \mathbf{x} en el subespacio generado por \mathbf{h}_1 y \mathbf{h}_2 para producir $\hat{\mathbf{s}}$ y luego proyectando $\hat{\mathbf{s}}$ en el subespacio restringido.

Solución La proyección ortogonal de \mathbf{x} sobre el subespacio generado por las columnas de \mathbf{H} es

$$\hat{\mathbf{s}} = \mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x}, \quad \text{con} \quad \mathbf{H} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

Considerando que

$$\mathbf{H}^T\mathbf{H} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathbf{I} \quad \text{y} \quad \mathbf{H}\mathbf{H}^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

se obtiene que

$$\hat{\mathbf{s}} = \begin{bmatrix} x[0] \\ x[1] \\ 0 \end{bmatrix}.$$

Como se discutió en el ejemplo, el subespacio restringido es el generado por el vector

$$\mathbf{h}_c = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$$

y la proyección ortogonal de $\hat{\mathbf{s}}$ sobre este subespacio es

$$\hat{\mathbf{s}}_c = \mathbf{h}_c(\mathbf{h}_c^T \mathbf{h}_c)^{-1} \mathbf{h}_c^T \hat{\mathbf{s}}$$

Considerando que

$$\mathbf{h}_c^T \mathbf{h}_c = 2 \quad \text{y} \quad \mathbf{h}_c \mathbf{h}_c^T = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

se obtiene que

$$\hat{\mathbf{s}}_c = \frac{1}{2} \begin{bmatrix} x[0] + x[1] \\ x[0] + x[1] \\ 0 \end{bmatrix},$$

que coincide con el resultado de la sección 8.5.

8.8.25. Problema 25

Si el modelo de la señal es

$$s[n] = A + B(-1)^n, \quad n = 0, \dots, N-1,$$

con N par, encontrar el LSE de $\boldsymbol{\theta} = [A \ B]^T$. Asumiendo que $A = B$, repetir el problema empleando el enfoque de LS restringidos. Comparar los resultados.

Solución En notación matricial, el modelo de la señal es

$$\mathbf{s} = \mathbf{H}\boldsymbol{\theta}, \quad \text{con} \quad \mathbf{H} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & -1 \end{bmatrix}.$$

El LSE es

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x} \\ &= \frac{1}{N} \mathbf{H}^T \mathbf{x} \end{aligned}$$

donde se tuvo en cuenta que

$$\mathbf{H}^T \mathbf{H} = \begin{bmatrix} N & 0 \\ 0 & N \end{bmatrix} = N\mathbf{I},$$

resultando en

$$\hat{\boldsymbol{\theta}} = \frac{1}{N} \begin{bmatrix} \sum_{n=0}^{N-1} x[n] \\ \sum_{n=0}^{N-1} (-1)^n x[n] \end{bmatrix}.$$

Asumiendo que $A = B$, la restricción en la forma de la ecuación 8.30 es

$$\begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} A \\ B \end{bmatrix} = 0,$$

por lo que $\mathbf{A} = \begin{bmatrix} 1 & -1 \end{bmatrix}$ y $\mathbf{b} = 0$. El estimador restringido, que está dado por la ecuación 8.31 es

$$\begin{aligned} \hat{\theta}_c &= \hat{\theta} - (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{A}^T [\mathbf{A} (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{A}^T]^{-1} (\mathbf{A} \hat{\theta} - \mathbf{b}) \\ &\stackrel{(a)}{=} \hat{\theta} - \frac{1}{N} \mathbf{A}^T \left(\frac{\mathbf{A} \mathbf{A}^T}{N} \right)^{-1} \mathbf{A} \hat{\theta} \\ &= [\mathbf{I} - \mathbf{A}^T (\mathbf{A} \mathbf{A}^T)^{-1} \mathbf{A}] \hat{\theta}, \end{aligned}$$

donde en (a) se consideró que $\mathbf{H}^T \mathbf{H} = N\mathbf{I}$. Teniendo en cuenta que

$$\mathbf{I} - \mathbf{A}^T (\mathbf{A} \mathbf{A}^T)^{-1} \mathbf{A} = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix},$$

como se calculó en la sección 8.5, y sustituyendo $\hat{\theta}$ se obtiene que

$$\hat{\theta}_c = \frac{1}{2N} \begin{bmatrix} \sum_{n=0}^{N-1} x[n] + \sum_{n=0}^{N-1} (-1)^n x[n] \\ \sum_{n=0}^{N-1} x[n] + \sum_{n=0}^{N-1} (-1)^n x[n] \end{bmatrix}.$$

De esta forma,

$$\begin{aligned} \hat{A}_c &= \hat{B}_c = \frac{1}{2N} \sum_{n=0}^{N-1} [x[n] + (-1)^n x[n]] \\ &= \frac{1}{2N} \sum_{n=0}^{N-1} x[n] [1 + (-1)^n] \\ &= \frac{1}{2N} \sum_{\substack{n=0 \\ n \text{ par}}}^{N-1} 2x[n] \\ &= \frac{1}{N} \sum_{n=0}^{\frac{N}{2}-1} x[2n] \\ &= \frac{1}{2} \times \frac{2}{N} \sum_{n=0}^{\frac{N}{2}-1} x[2n], \end{aligned}$$

es decir

$$\hat{A}_c = \frac{\bar{x}'}{2}, \quad \text{con} \quad \mathbf{x}' = [x[0] \ x[2] \ \dots \ x[N-2]].$$

Esto tiene sentido, ya que el modelo de señal con $A = B$ es

$$s[n] = A + (-1)^n A = 2A \begin{cases} 2A, & n \text{ par} \\ 0, & n \text{ impar}, \end{cases}$$

y el LSE restringido \hat{A}_c es el promedio de las muestras pares de $x[n]$ dividido 2.

8.8.26. Problema 26

Considérese la minimización de $h(\theta)$ respecto a θ y asúmase que

$$\hat{\theta} = \arg \min_{\theta} h(\theta).$$

Sea $\alpha = g(\theta)$, donde g es una función uno a uno. Probar que si $\hat{\alpha}$ minimiza $h(g^{-1}(\hat{\alpha}))$, se cumple que $\hat{\theta} = g^{-1}(\hat{\alpha})$.

Solución Por hipótesis, se cumple que $\hat{\alpha}$ minimiza $h(g^{-1}(\hat{\alpha}))$, lo que implica que

$$\left. \frac{dh(g^{-1}(\alpha))}{d\alpha} \right|_{\alpha=\hat{\alpha}} = 0.$$

Aplicando la regla de la cadena de la diferenciación, se tiene que

$$\frac{dh(g^{-1}(\alpha))}{d\alpha} = h'(g^{-1}(\alpha)) \frac{dg^{-1}(\alpha)}{d\alpha} \stackrel{(a)}{=} h'(g^{-1}(\alpha)) \frac{1}{g'(g^{-1}(\alpha))}$$

donde en (a) se consideró que la derivada de una función inversa está dada por³ $[f^{-1}]'(a) = 1/f'(f^{-1}(a))$. Por lo tanto, se cumple que

$$h'(g^{-1}(\alpha)) \frac{1}{g'(g^{-1}(\alpha))} \Big|_{\alpha=\hat{\alpha}} = 0$$

Considerando que g es una función inyectiva, se cumple que $g'(\theta)$ es estrictamente positiva o negativa para todo θ , por lo que la condición resulta en

$$h'(g^{-1}(\hat{\alpha})) = 0.$$

Como

$$\hat{\theta} = \arg \min_{\theta} h(\theta) \quad \Leftrightarrow \quad h'(\hat{\theta}) = 0,$$

se concluye que

$$\hat{\theta} = g^{-1}(\hat{\alpha}).$$

Una forma simple de ver lo mismo, es partiendo de que por hipótesis se cumple que

$$h(g^{-1}(\hat{\alpha})) \leq h(g^{-1}(\alpha)) \quad \forall \alpha,$$

y por lo tanto

$$h(\theta_0) \leq h(\theta) \quad \forall \theta, \quad \text{con } \theta_0 = g^{-1}(\hat{\alpha}).$$

Pero como θ_0 minimiza $h(\theta)$, $\hat{\theta} = \theta_0$, resultando en que

$$\hat{\theta} = g^{-1}(\hat{\alpha}).$$

8.8.27. Problema 27

Sean los datos observados

$$x[n] = e^{\theta} + w[n], \quad n = 0, \dots, N-1.$$

Determinar una iteración de Newton-Raphson para encontrar el LSE de θ . ¿Es posible evitar la optimización no lineal y encontrar el LSE analíticamente?

Solución En este caso, el modelo de señal es $s[n] = e^{\theta}$, por lo que la función de error a minimizar es

$$J(\theta) = \sum_{n=0}^{N-1} (x[n] - e^{\theta})^2.$$

Diferenciando respecto al parámetro, se obtiene que la condición necesaria es

$$\frac{\partial J}{\partial \theta} = -2 \sum_{n=0}^{N-1} (x[n] - e^{\theta}) e^{\theta} = 0 \quad \Rightarrow \quad \sum_{n=0}^{N-1} (x[n] - e^{\hat{\theta}}) = 0 \quad \Rightarrow \quad N e^{\hat{\theta}} = \sum_{n=0}^{N-1} x[n] \quad \Rightarrow \quad e^{\hat{\theta}} = \bar{x}$$

resultando en que el LSE es

$$\hat{\theta} = \ln \bar{x}.$$

³Ver https://en.wikipedia.org/wiki/Inverse_functions_and_differentiation.

Notar que si se realiza la transformación del parámetro $\alpha = g(\theta) = e^\theta$, el problema es lineal y el estimador LSE de α es $\hat{\alpha} = \bar{x}$. Como la transformación g es biyectiva, el estimador LSE de θ es $\hat{\theta} = g^{-1}(\hat{\alpha}) = \ln \hat{\alpha}$, como se mostró en el problema de la sección 8.8.26. Si bien el problema no es lineal, se encontró una expresión analítica para el estimador.

Para encontrar una solución iterativa empleando el algoritmo de Newton-Raphson, que está dado por la ecuación 8.34, se parte de que la función a encontrar la raíz es

$$g(\theta) = e^\theta \sum_{n=0}^{N-1} (x[n] - e^\theta) = e^\theta \left(\sum_{n=0}^{N-1} x[n] - N e^\theta \right).$$

Teniendo en cuenta que la derivada es

$$g'(\theta) = e^\theta \left(\sum_{n=0}^{N-1} x[n] - N e^\theta \right) - e^\theta N e^\theta = e^\theta \left(\sum_{n=0}^{N-1} x[n] - 2N e^\theta \right).$$

la iteración de Newton-Raphson queda

$$\begin{aligned} \theta_{k+1} &= \theta_k - \frac{g(\theta_k)}{g'(\theta_k)} \\ &= \theta_k - \frac{e^{\theta_k} \left(\sum_{n=0}^{N-1} x[n] - N e^{\theta_k} \right)}{e^{\theta_k} \left(\sum_{n=0}^{N-1} x[n] - 2N e^{\theta_k} \right)} \end{aligned}$$

resultando en

$$\theta_{k+1} = \theta_k + \frac{\bar{x} - e^{\theta_k}}{2e^{\theta_k} - \bar{x}}.$$

También podría haberse considerado como función a resolver a

$$g(\theta) = \sum_{n=0}^{N-1} (x[n] - e^\theta),$$

resultando en una iteración distinta. El mismo resultado se obtiene empleando directamente la ecuación 8.38 de la iteración de Newton-Raphson para el LSE no lineal genérico. Como en este caso $s[n] = e^\theta$, o $\mathbf{s} = e^\theta \mathbf{1}$, se tiene que

$$\frac{\partial s[i]}{\partial \theta} = e^\theta, \quad i = 0, \dots, N-1 \quad \Rightarrow \quad \mathbf{H}(\theta) = e^\theta \mathbf{1}$$

y

$$\frac{\partial^2 s[n]}{\partial \theta^2} = e^\theta, \quad n = 0, \dots, N-1 \quad \Rightarrow \quad \mathbf{G}_n(\theta) = e^\theta,$$

la iteración de Newton-Raphson es

$$\begin{aligned}
\theta_{k+1} &= \theta_k + \left(\mathbf{H}^T(\theta_k) \mathbf{H}(\theta_k) - \sum_{n=0}^{N-1} \mathbf{G}_n(\theta_k)(x[n] - s[n]) \right)^{-1} \mathbf{H}^T(\theta_k)(\mathbf{x} - \mathbf{s}(\theta_k)) \\
&= \theta_k + \left(e^{2\theta_k} \mathbf{1}^T \mathbf{1} - e^{\theta_k} \sum_{n=0}^{N-1} (x[n] - e^{\theta_k}) \right)^{-1} e^{\theta_k} \mathbf{1}^T (\mathbf{x} - e^{\theta_k} \mathbf{1}) \\
&= \theta_k + \frac{e^{\theta_k} \sum_{n=0}^{N-1} (x[n] - e^{\theta_k})}{e^{\theta_k} \left[N e^{\theta_k} - \sum_{n=0}^{N-1} (x[n] - e^{\theta_k}) \right]} \\
&= \theta_k + \frac{\sum_{n=0}^{N-1} x[n] - N e^{\theta_k}}{2 N e^{\theta_k} - \sum_{n=0}^{N-1} x[n]} \\
&= \theta_k + \frac{\bar{x} - e^{\theta_k}}{2 e^{\theta_k} - \bar{x}}.
\end{aligned}$$

8.8.28. Problema 28

En el ejemplo de la sección 8.7.1 el LSE verdadero debe minimizar

$$J = \sum_{n=0}^{N-1} (h_d[n] - h[n])^2.$$

En este problema se deducen las ecuaciones que hay que optimizar para encontrar el LSE verdadero. Se asume que $p = q + 1$. Primero, mostrar que el modelo de señal puede escribirse como

$$\mathbf{s} = \begin{bmatrix} h[0] \\ h[1] \\ \vdots \\ h[N-1] \end{bmatrix} = \underbrace{\begin{bmatrix} g[0] & 0 & \dots & 0 \\ g[1] & g[0] & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ g[q] & g[q-1] & \dots & g[0] \\ g[q+1] & g[q] & \dots & g[1] \\ \vdots & \vdots & & \vdots \\ g[N-1] & g[N-2] & \dots & g[N-1-q] \end{bmatrix}}_{\mathbf{G}} \underbrace{\begin{bmatrix} b[0] \\ b[1] \\ \vdots \\ b[q] \end{bmatrix}}_{\mathbf{b}}$$

donde

$$g[n] = \mathcal{Z}^{-1} \left\{ \frac{1}{\mathcal{A}(z)} \right\}$$

y la matriz \mathbf{G} tiene dimensiones $N \times (q+1)$. Si $\mathbf{x} = [h_d[0] \ h_d[1] \ \dots \ h_d[N-1]]^T$ mostrar que

$$J(\mathbf{a}, \hat{\mathbf{b}}) = \mathbf{x}^T (\mathbf{I} - \mathbf{G}(\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T) \mathbf{x}$$

donde $\hat{\mathbf{b}}$ es el LSE de \mathbf{b} para un \mathbf{a} dado. Finalmente, probar que

$$\mathbf{I} - \mathbf{G}(\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T = \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \quad (8.46)$$

donde \mathbf{A}^T es la matriz de dimensiones $(N - q - 1) \times N = (N - p) \times N$

$$\mathbf{A}^T = \begin{bmatrix} a[p] & a[p-1] & \dots & 1 & 0 & 0 & \dots & 0 \\ 0 & a[p] & \dots & a[1] & 1 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 0 & a[p] & a[p-1] & \dots & 1 \end{bmatrix}. \quad (8.47)$$

Por lo tanto, el LSE de \mathbf{a} se obtiene minimizando

$$\mathbf{x}^T \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{x},$$

y luego que es encontrado empleando una técnica de optimización no lineal, se tiene que

$$\begin{aligned}\hat{\mathbf{b}} &= (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T \mathbf{x} \\ \hat{\mathbf{s}} &= (\mathbf{I} - \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T) \mathbf{x},\end{aligned}$$

donde los elementos de \mathbf{A} son reemplazados por el LSE de \mathbf{a} . Notar que este problema es un ejemplo de un problema LS no lineal separable. Sugerencia: para probar la identidad 8.46 considerar $\mathbf{L} = [\mathbf{A} \ \mathbf{G}]$, la cual es invertible debido a que es de rango completo y calcular

$$\mathbf{L}(\mathbf{L}^T \mathbf{L})^{-1} \mathbf{L}^T = \mathbf{I}.$$

También se necesita observar que $\mathbf{A}^T \mathbf{G} = \mathbf{0}$, que surge de que $a[n] * g[n] = \delta[n]$.

Solución Este problema se analiza en la sección 9.18 de [13]. Se comenzará deduciendo el modelo de la señal. El sistema es un modelo ARMA con función de transferencia

$$\mathcal{H}(z) = \frac{\mathcal{B}(z)}{\mathcal{A}(z)} = \frac{b[0] + b[1]z^{-1} + \dots + b[q]z^{-q}}{1 + a[1]z^{-1} + \dots + a[p]z^{-p}}$$

con $p = q + 1$. Aplicando la transformada z inversa se obtiene la respuesta al impulso, y es

$$\begin{aligned}h[n] &= \mathcal{Z}^{-1}\{\mathcal{H}(z)\} \\ &= \mathcal{Z}^{-1}\left\{\frac{b[0] + b[1]z^{-1} + \dots + b[q]z^{-q}}{\mathcal{A}(z)}\right\} \\ &= b[0]\mathcal{Z}^{-1}\left\{\frac{1}{\mathcal{A}(z)}\right\} + b[1]\mathcal{Z}^{-1}\left\{\frac{z^{-1}}{\mathcal{A}(z)}\right\} + \dots + b[q]\mathcal{Z}^{-1}\left\{\frac{z^{-q}}{\mathcal{A}(z)}\right\} \\ &= b[0]g[n] + b[1]g[n-1] + \dots + b[q]g[n-q] \\ &= \sum_{k=0}^q b[k]g[n-k],\end{aligned}$$

donde se definió

$$g[n] = \mathcal{Z}^{-1}\left\{\frac{1}{\mathcal{A}(z)}\right\}$$

y se empleó que $\mathcal{Z}^{-1}\{z^{-k}\mathcal{G}(z)\} = g[n-k]$. Por lo tanto, se tiene que

$$\begin{aligned}h[0] &= b[0]h[0] \\ h[1] &= b[0]g[1] + g[1]g[0] \\ &\vdots \\ h[q] &= b[0]g[q] + b[1]g[q-1] + \dots + b[q]g[0] \\ &\vdots \\ h[N-1] &= b[0]g[N-1] + b[1]g[N-2] + \dots + b[q]g[N-1-q]\end{aligned}$$

que en notación matricial se expresa como

$$\mathbf{s} = \begin{bmatrix} h[0] \\ h[1] \\ \vdots \\ h[N-1] \end{bmatrix} = \underbrace{\begin{bmatrix} g[0] & 0 & \dots & 0 \\ g[1] & g[0] & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ g[q] & g[q-1] & \dots & g[0] \\ g[q+1] & g[q] & \dots & g[1] \\ \vdots & \vdots & & \vdots \\ g[N-1] & g[N-2] & \dots & g[N-1-q] \end{bmatrix}}_{\mathbf{G}} \underbrace{\begin{bmatrix} b[0] \\ b[1] \\ \vdots \\ b[q] \end{bmatrix}}_{\mathbf{b}},$$

donde la matriz \mathbf{G} tiene dimensiones $N \times (q+1)$ o $N \times p$ y depende del parámetro \mathbf{a} . Se observa que es un modelo lineal en el parámetro \mathbf{b} . Si las observaciones son $\mathbf{x} = [h_d[0] \ h_d[1] \ \dots \ h_d[N-1]]^T$, el error LS es

$$\begin{aligned} J(\mathbf{a}, \mathbf{b}) &= \sum_{n=0}^{N-1} (h_d[n] - h[n])^2 \\ &= (\mathbf{x} - \mathbf{s})^T (\mathbf{x} - \mathbf{s}) \\ &= (\mathbf{x} - \mathbf{G}(\mathbf{a})\mathbf{b})^T (\mathbf{x} - \mathbf{G}(\mathbf{a})\mathbf{b}), \end{aligned}$$

donde por única vez se hizo explícita la dependencia de la matriz \mathbf{G} con el parámetro \mathbf{a} . Por tratarse de un problema LS lineal en \mathbf{b} , el LSE $\hat{\mathbf{b}}$ que minimiza el error para un \mathbf{a} dado es

$$\hat{\mathbf{b}} = (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T \mathbf{x}$$

y el error mínimo es

$$\begin{aligned} J(\mathbf{a}, \hat{\mathbf{b}}) &= (\mathbf{x} - \mathbf{G}\hat{\mathbf{b}})^T (\mathbf{x} - \mathbf{G}\hat{\mathbf{b}}) \\ &= (\mathbf{x} - \mathbf{G}(\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T \mathbf{x})^T (\mathbf{x} - \mathbf{G}(\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T \mathbf{x}) \\ &= \mathbf{x}^T (\mathbf{I} - \mathbf{G}(\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T) (\mathbf{I} - \mathbf{G}(\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T) \mathbf{x}, \end{aligned}$$

donde se consideró que la matriz de proyección $\mathbf{I} - \mathbf{G}(\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T$ es simétrica y teniendo en cuenta que además es idempotente (ver los problemas de las secciones 8.8.11 y 8.8.12), resulta en

$$J(\mathbf{a}, \hat{\mathbf{b}}) = \mathbf{x}^T (\mathbf{I} - \mathbf{G}(\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T) \mathbf{x}.$$

Para probar la identidad de la ecuación 8.46, se necesita primero observar que por definición

$$\mathcal{G}(z) = \frac{1}{\mathcal{A}(z)} = \frac{1}{1 + a[1]z^{-1} + \dots + a[p]z^{-p}} \quad \Rightarrow \quad \mathcal{G}(z)(1 + a[1]z^{-1} + \dots + a[p]z^{-p}) = 1,$$

y aplicando la transformada z inversa, se obtiene que

$$g[n] + a[1]g[n-1] + \dots + a[p]g[n-p] = \delta[n] \quad \text{o} \quad \sum_{k=0}^p a[k]g[n-k] = \delta[n]. \quad (8.48)$$

Definiendo $a[n] = 0$ en $n < 0$ y $n > p$ y $g[n] = 0$ en $n < 0$, la expresión se puede escribir como

$$\sum_{k=-\infty}^{\infty} a[k]g[n-k] = \delta[n] \quad \text{o} \quad (a * g)[n] = \delta[n]. \quad (8.49)$$

Este resultado se puede obtener directamente empleando la propiedad de convolución de la transformada z ,

$$\mathcal{A}(z)\mathcal{G}(z) = 1 \quad \xrightarrow{\mathcal{Z}^{-1}} \quad (a * g)[n] = \delta[n].$$

Continuando con la prueba de 8.46, se considera la matriz $\mathbf{L} = [\mathbf{A} \ \mathbf{G}]$, donde \mathbf{A} es la matriz de la ecuación 8.47. Como la matriz \mathbf{A} tiene dimensiones $N \times (N-p)$ y la matriz \mathbf{G} tiene dimensiones $N \times p$, la matriz \mathbf{L} es cuadrada de dimensiones $N \times N$. Se verá que además que es de rango completo. Para hacerlo, se parte observando que debido a su estructura, las matrices \mathbf{A} y \mathbf{G} son de rango completo, con rangos $N-p$ y p respectivamente. Si se prueba que las columnas de \mathbf{A} son linealmente independientes a las columnas de \mathbf{G} se concluye que \mathbf{L} tiene N columnas linealmente independientes y por lo tanto, es de rango completo N . Se mostrará que las matrices \mathbf{A} y \mathbf{G} son ortogonales entre si, o $\mathbf{A}^T \mathbf{G} = \mathbf{0}$, lo que implica que sus columnas son linealmente independientes. La columna i -ésima de \mathbf{A} , que es la fila i -ésima de \mathbf{A}^T , con $0 \leq i \leq N-p-1$, y la columna j -ésima de \mathbf{G} , con $0 \leq j \leq p-1$, son respectivamente

$$\begin{aligned} [\mathbf{A}]_i &= [\quad 0 \quad \dots \quad 0 \quad 0 \quad a[p] \quad a[p-1] \quad \dots \quad a[0] \quad 0 \quad \dots \quad 0 \quad]^T \\ &\quad \quad \quad k=0 \quad \quad \quad k=i \quad \quad \quad k=i+p \quad \quad \quad k=N-1 \\ [\mathbf{G}]_j &= [\quad 0 \quad \dots \quad 0 \quad g[0] \quad g[1] \quad \dots \quad \dots \quad \dots \quad \dots \quad g[N-1-j] \quad]^T. \\ &\quad \quad \quad k=0 \quad \quad \quad k=j \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad k=N-1 \end{aligned}$$

Por lo tanto, el elemento (k, i) de \mathbf{A} y el elemento (k, j) de \mathbf{G} son respectivamente

$$\begin{aligned} [\mathbf{A}]_{ki} &= a[p+i-k] & \text{con } a[n] = 0 \text{ en } n < 0 \text{ y } n > p \\ [\mathbf{G}]_{kj} &= g[k-j] & \text{con } g[n] = 0 \text{ en } n < 0. \end{aligned}$$

En la figura 8.15 se ilustra gráficamente los elementos de una columna de la matriz \mathbf{A} , los cuales son una versión invertida temporalmente y desplazada de $a[n]$. Teniendo en cuenta estas consideraciones, el

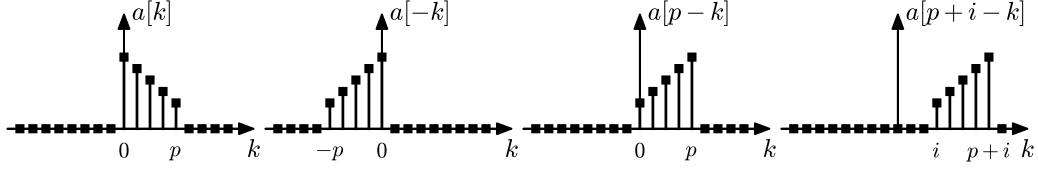


Figura 8.15: Columna i -ésima de la matriz \mathbf{A} . Se trata de la secuencia $a[p+i-k]$.

elemento (i, j) de la matriz $\mathbf{A}^T \mathbf{G}$ es

$$\begin{aligned} [\mathbf{A}^T \mathbf{G}]_{ij} &= \sum_{k=0}^{N-1} [\mathbf{A}^T]_{ik} [\mathbf{G}]_{kj} \\ &= \sum_{k=0}^{N-1} [\mathbf{A}]_{ki} [\mathbf{G}]_{kj} \\ &= \sum_{k=0}^{N-1} a[p+i-k] g[k-j] \\ &\stackrel{(a)}{=} \sum_{k=-\infty}^{\infty} a[p+i-k] g[k-j] \\ &\stackrel{(b)}{=} \sum_{l=-\infty}^{\infty} a[p+i-j-l] g[l] \\ &\stackrel{(c)}{=} (a * g)[p+i-j] \\ &\stackrel{(d)}{=} \delta[p+i-j] \\ &\stackrel{(e)}{=} 0, \end{aligned}$$

donde en (a) se extendieron los límites de la sumatoria teniendo en cuenta que $a[n] = 0$ en $n < 0$ y $n > p$ y $g[n] = 0$ en $n < 0$, en (b) se realizó el cambio de variable $l = k - j$, en (c) se notó que la sumatoria es la convolución $(a * g)[n]$ evaluada en $n = p + i - j$, en (d) se empleó el resultado de la ecuación 8.49 y en (e) se tuvo en cuenta que como $0 \leq i \leq N - p - 1$ y $0 \leq j \leq p - 1$

$$p + i - j \geq p + i_{\min} - j_{\max} = p + 0 - (p - 1) = 1$$

y por lo tanto, $\delta[p+i-j] = 0$ para todo par (i, j) . Se concluye que efectivamente $\mathbf{A}^T \mathbf{G} = \mathbf{0}$, que implica que las columnas de \mathbf{A} y \mathbf{G} son linealmente independientes y por lo tanto \mathbf{L} es una matriz cuadrada $N \times N$ de rango completo, que es lo que se quería probar. Como consecuencia, el subespacio generado por \mathbf{L} es el espacio \mathbb{R}^N por lo que la matriz de proyección asociada es la identidad, es decir, se cumple que

$$\mathbf{L}(\mathbf{L}^T \mathbf{L})^{-1} \mathbf{L}^T = \mathbf{I}.$$

El hecho de que \mathbf{L} sea cuadrada y de rango completo asegura que \mathbf{L} es invertible, por lo que también lo es $\mathbf{L}^T \mathbf{L}$. Empleando la definición de \mathbf{L} se tiene que

$$\mathbf{L}^T \mathbf{L} = \begin{bmatrix} \mathbf{A}^T \\ \mathbf{G}^T \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{G} \end{bmatrix} = \begin{bmatrix} \mathbf{A}^T \mathbf{A} & \mathbf{A}^T \mathbf{G} \\ \mathbf{G}^T \mathbf{A} & \mathbf{G}^T \mathbf{G} \end{bmatrix} \stackrel{(a)}{=} \begin{bmatrix} \mathbf{A}^T \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^T \mathbf{G} \end{bmatrix},$$

donde en (a) se empleó que $\mathbf{A}^T \mathbf{G} = \mathbf{0}$, resultado obtenido previamente, que implica que también $\mathbf{G}^T \mathbf{A} = \mathbf{0}$. Luego (ver el apéndice G),

$$(\mathbf{L}^T \mathbf{L})^{-1} = \begin{bmatrix} (\mathbf{A}^T \mathbf{A})^{-1} & \mathbf{0} \\ \mathbf{0} & (\mathbf{G}^T \mathbf{G})^{-1} \end{bmatrix}.$$

Finalmente

$$\begin{aligned} \mathbf{L}(\mathbf{L}^T \mathbf{L})^{-1} \mathbf{L}^T &= [\mathbf{A} \quad \mathbf{G}] \begin{bmatrix} (\mathbf{A}^T \mathbf{A})^{-1} & \mathbf{0} \\ \mathbf{0} & (\mathbf{G}^T \mathbf{G})^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{A}^T \\ \mathbf{G}^T \end{bmatrix} \\ &= [\mathbf{A} \quad \mathbf{G}] \begin{bmatrix} (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \\ (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T \end{bmatrix} \\ &= \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T + \mathbf{G}(\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T. \end{aligned}$$

Se concluye que

$$\mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T + \mathbf{G}(\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T = \mathbf{I} \quad \Rightarrow \quad \mathbf{I} - \mathbf{G}(\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T = \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T.$$

Como se explica en la letra del problema, el error LS mínimo con el LSE de \mathbf{b} para un \mathbf{a} dado es

$$J(\mathbf{a}, \hat{\mathbf{b}}) = \mathbf{x}^T (\mathbf{I} - \mathbf{G}(\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T) \mathbf{x} = \mathbf{x}^T \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{x}.$$

El procedimiento para encontrar el LSE de \mathbf{a} y \mathbf{b} consiste en calcular primero el LSE de \mathbf{a} como

$$\hat{\mathbf{a}} = \arg \min_{\mathbf{a}} \mathbf{x}^T \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{x}$$

mediante alguna técnica de optimización no lineal, y luego calcular el LSE de \mathbf{b} como

$$\hat{\mathbf{b}} = (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T \mathbf{x},$$

donde los elementos de la matriz \mathbf{G} se obtienen mediante la ecuación 8.48 empleando los valores de $\hat{\mathbf{a}}$.

Se realizó una simulación por computadora para estimar los parámetros $\hat{\mathbf{a}}$ y $\hat{\mathbf{b}}$ mediante esta técnica igual a la realizada en la sección 8.7.1 para estimar los parámetros $\hat{\mathbf{a}}$ y $\hat{\mathbf{b}}$ mediante el método LS de Prony. La señal deseada es el filtro pasabajos cuya respuesta al impulso está dado por la ecuación 8.39, con frecuencia de corte $f_c = 0.1$, largo de $N = 51$ muestras y retardo de $n_0 = 25$ muestras. Con $p = q + 1 = 10$ se diseñó un filtro digital con el objetivo de igualar su respuesta al impulso con la respuesta al impulso deseada. El resultado se muestra en la figura 8.16. En la figura 8.17 se muestra la parte principal del código para el cálculo de los parámetros. Para la optimización no lineal necesaria para el cálculo de $\hat{\mathbf{a}}$ se empleó la función de Python `scipy.optimize.minimize` con el algoritmo BFGS⁴. Se pudo observar que el resultado es muy sensible a la condición inicial empleada en el algoritmo de optimización.

⁴Ver https://en.wikipedia.org/wiki/Broyden-Fletcher-Goldfarb-Shanno_algorithm

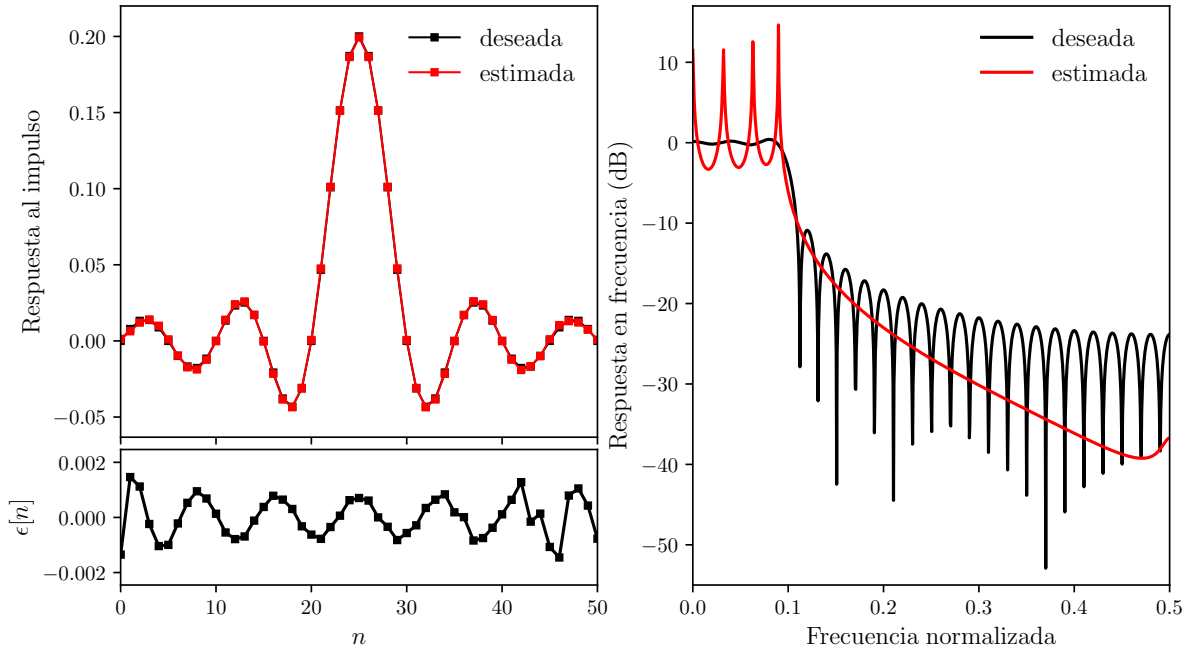


Figura 8.16: Diseño de un filtro digital empleando el LSE verdadero en lugar del método LS de Prony. Se compara la respuesta al impulso deseada y la obtenida, $h_d[n]$ y $h[n]$ respectivamente, y las correspondientes respuestas en frecuencia $H_d(f)$ y $H(f)$. También se grafica el error de estimación $\epsilon[n] = h_d[n] - h[n]$. Los parámetros empleados son $p = q + 1 = 10$, $N = 51$ muestras y $f_c = 0.1$.

```
import numpy as np
from scipy import signal, linalg, optimize

# hd es la señal deseada dada, con hd.shape = (N, )
# el filtro a diseñar tiene función de transferencia con q coeficientes
# en el numerador y p coeficientes en el denominador, con p = q + 1.

# función no lineal en a a optimizar:  $hd^T @ A @ (A^T @ A)^{-1} @ A^T @ hd$ 
def ja(a):
    c = np.concatenate(([a[0]], np.zeros((N - (p + 1), ))))
    r = np.concatenate((a, [1], np.zeros((N - (p + 1), ))))
    AT = linalg.toeplitz(c, r)
    return hd @ AT.T @ linalg.inv(AT @ AT.T) @ AT @ hd

## cálculo de a empleando scipy.optimize.minimize
# a0: valor inicial para el algoritmo de optimización
a0 = np.zeros((p, ))
a0[0] = 0.2
res = optimize.minimize(ja, a0)
a = res.x
a = np.concatenate(([1], a[:-1]))

## cálculo de b
# cálculo de la matriz G empleando el valor de a obtenido en la optimización
delta = np.zeros((N, ))
delta[0] = 1
g = signal.lfilter([1], a, delta)
G = linalg.toeplitz(g, np.concatenate(([g[0]], np.zeros((p - 1, )))))
b = linalg.inv(G.T @ G) @ G.T @ hd
```

Figura 8.17: Implementación en el lenguaje Python del diseño de filtros digitales empleando el LSE verdadero. El problema LS es lineal en el parámetro **b** y no lineal en el parámetro **a**. Para la optimización no lineal se empleó la función `scipy.optimize.minimize` para minimizar el resultado de la función `ja`.

Capítulo 9

Método de los momentos

9.1. Introducción

La estimación mediante el método de los momentos se basa en la solución de una ecuación teórica que involucra los momentos de una PDF. Como ejemplo, considérese que se observa $x[n]$ para $n = 0, \dots, N-1$, que son muestras IID de una PDF mezcla de gaussianas (ver el problema de la sección 6.4.14)

$$p(x[n]; \epsilon) = \frac{1-\epsilon}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{x^2[n]}{2\sigma_1^2}\right) + \frac{\epsilon}{\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{x^2[n]}{2\sigma_2^2}\right),$$

o de forma mas sucinta

$$p(x[n]; \epsilon) = (1-\epsilon)\phi_1(x[n]) + \epsilon\phi_2(x[n]), \quad \text{con} \quad \phi_i(x[n]) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{x^2[n]}{2\sigma_i^2}\right).$$

El parámetro ϵ se llama parámetro de mezcla, y cumple que $0 < \epsilon < 1$, y σ_1^2 y σ_2^2 son las varianzas de las PDF individuales. Si σ_1^2 y σ_2^2 son conocidos y se quiere estimar ϵ , todos los métodos de estimación MVU fracasan. El MLE requiere de la maximización de una función no lineal de ϵ , que podría implementarse mediante un búsqueda de grilla. El método de los momentos brinda una solución mas simple. Se parte observando que

$$\begin{aligned} E(x^2[n]) &= \int_{-\infty}^{\infty} x^2[n] [(1-\epsilon)\phi_1(x[n]) + \epsilon\phi_2(x[n])] dx[n] \\ &= (1-\epsilon)\sigma_1^2 + \epsilon\sigma_2^2, \end{aligned} \quad (9.1)$$

como se muestra en el problema 6.4.14. Esta ecuación teórica vincula el parámetro desconocido ϵ con el segundo momento. Si se reemplaza $E(x^2[n])$ por su estimador natural $\frac{1}{N} \sum_{n=0}^{N-1} x^2[n]$ se tiene que

$$\frac{1}{N} \sum_{n=0}^{N-1} x^2[n] = (1-\epsilon)\sigma_1^2 + \epsilon\sigma_2^2$$

y resolviendo para ϵ , se obtiene el estimador por el método de los momentos,

$$\hat{\epsilon} = \frac{\frac{1}{N} \sum_{n=0}^{N-1} x^2[n] - \sigma_1^2}{\sigma_2^2 - \sigma_1^2}. \quad (9.2)$$

En este caso, el estimador es insesgado, pero no siempre es así. La varianza de $\hat{\epsilon}$ es

$$\begin{aligned}\text{var}(\hat{\epsilon}) &= \frac{1}{\sigma_2^2 - \sigma_1^2} \text{var} \left(\frac{1}{N} \sum_{n=0}^{N-1} x^2[n] \right) \\ &= \frac{1}{N(\sigma_2^2 - \sigma_1^2)} \text{var}(x^2[n]) \\ &= \frac{1}{N(\sigma_2^2 - \sigma_1^2)} [E(x^4[n]) - E^2(x^2[n])],\end{aligned}$$

y como

$$\begin{aligned}E(x^4[n]) &= \int_{-\infty}^{\infty} x^4[n] [(1-\epsilon)\phi_1(x[n]) + \epsilon\phi_2(x[n])] dx[n] \\ &= (1-\epsilon)\mu_{41} + \epsilon\mu_{42} \\ &= (1-\epsilon)3\sigma_1^4 + \epsilon3\sigma_2^2\end{aligned}$$

donde μ_{4i} es el momento cuarto de $\phi_i(x[n])$, que por ser una PDF $\mathcal{N}(0, \sigma_i^2)$ es $\mu_{4i} = 3\sigma_i^4$. Empleando este resultado y el de la ecuación 9.1, se obtiene que

$$\text{var}(\hat{\epsilon}) = \frac{3(1-\epsilon)\sigma_1^4 + 3\epsilon\sigma_2^2 - [(1-\epsilon)\sigma_1^2 + \epsilon\sigma_2^2]}{N(\sigma_2^2 - \sigma_1^2)}. \quad (9.3)$$

Para determinar la pérdida de desempeño, se podría calcular la CRLB numéricamente y comparar con este resultado. Obsérvese que el estimador es consistente en el sentido de que $\hat{\epsilon} \rightarrow \epsilon$ en probabilidad con $N \rightarrow \infty$. Esto es porque de 9.2 se deduce que $E(\hat{\epsilon}) = \epsilon$, y de 9.3 que $\text{var}(\hat{\epsilon}) \rightarrow 0$ con $N \rightarrow \infty$. En general, como los estimadores de los momentos sustituidos en la ecuación teórica tienden a los valores verdaderos de los momentos con $N \rightarrow \infty$, la ecuación a resolver tiende a la ecuación teórica. Como resultado, el estimador obtenido mediante el método de los momentos es consistente (ver el problema de la sección 7.6.5).

9.2. Método de los momentos

9.2.1. Parámetro escalar

A continuación se resume el método de los momentos cuando el parámetro es escalar. Asíumase que el momento k -ésimo $\mu_k = E(x^k[n])$ depende del parámetro θ como

$$\mu_k = h(\theta).$$

Primero se resuelve para θ ,

$$\theta = h^{-1}(\mu_k)$$

asumiendo que existe h^{-1} . Luego, se reemplaza el momento teórico por su estimador natural

$$\hat{\mu}_k = \frac{1}{N} \sum_{n=0}^{N-1} x^k[n]$$

para obtener el estimador por el método de los momentos

$$\hat{\theta} = h^{-1} \left(\frac{1}{N} \sum_{n=0}^{N-1} x^k[n] \right).$$

Ejemplo: PDF exponencial

Se consideran N observaciones IID de la PDF exponencial

$$p(x[n], \lambda) = \begin{cases} \lambda \exp(-\lambda x[n]), & x[n] \geq 0 \\ 0, & x[n] < 0. \end{cases}$$

Se desea estimar el parámetro λ , donde $\lambda > 0$. El momento primero es

$$\mu_1 = E(x[n]) = \int_0^\infty x[n] \lambda \exp(-\lambda x[n]) dx[n] = \frac{1}{\lambda},$$

como se calculó en el problema de la sección 5.4.15. Resolviendo para λ , se tiene que

$$\lambda = \frac{1}{\mu_1}$$

y sustituyendo el momento primero por su estimador natural se obtiene que el estimador por el método de los momentos es

$$\hat{\lambda} = \frac{1}{\frac{1}{N} \sum_{n=0}^{N-1} x[n]}.$$

9.2.2. Parámetro vectorial

Se considera ahora el caso en que el parámetro es un vector de dimensiones $p \times 1$. Para estimar el parámetro se requieren p ecuaciones teóricas de los momentos. Por lo tanto, supóngase que

$$\begin{aligned} \mu_1 &= h_1(\theta_1, \dots, \theta_p) \\ \mu_2 &= h_2(\theta_1, \dots, \theta_p) \\ &\vdots \\ \mu_p &= h_p(\theta_1, \dots, \theta_p) \end{aligned}$$

o en notación matricial

$$\boldsymbol{\mu} = \mathbf{h}(\boldsymbol{\theta}). \quad (9.4)$$

Resolviendo para $\boldsymbol{\theta}$,

$$\boldsymbol{\theta} = \mathbf{h}^{-1}(\boldsymbol{\mu}). \quad (9.5)$$

el estimador por el método de los momentos es

$$\hat{\boldsymbol{\theta}} = \mathbf{h}^{-1}(\hat{\boldsymbol{\mu}}) \quad \text{con} \quad \hat{\boldsymbol{\mu}} = \begin{bmatrix} \frac{1}{N} \sum_{n=0}^{N-1} x[n] \\ \frac{1}{N} \sum_{n=0}^{N-1} x^2[n] \\ \vdots \\ \frac{1}{N} \sum_{n=0}^{N-1} x^p[n] \end{bmatrix}.$$

Puede ocurrir que los primeros p momentos sean insuficientes para determinar todos los parámetros a estimar. En ese caso, se necesita encontrar un conjunto de p ecuaciones teóricas de los momentos que permitan resolver la ecuación 9.4 para $\boldsymbol{\theta}$ y obtener la ecuación 9.5. Es preferible emplear las ecuaciones con los momentos de menor orden posible, ya que típicamente la varianza de los estimadores de los momentos crecen con el orden. Además, es deseable que el sistema de ecuaciones sea lineal para mantener la simplicidad del método. En el caso de un parámetro vectorial es posible que se requieran los momentos cruzados, como en el caso del ejemplo de la sección 9.4 y del problema de la sección 9.5.5.

9.3. Evaluación estadística de los estimadores

En el método de los momentos, no se sabe de antemano si el comportamiento del estimador es bueno. Una vez obtenido el estimador, podría ser posible determinar sus propiedades estadísticas, como la media y la varianza, al igual que en el ejemplo de la mezcla de gaussianas de la introducción. Como

$$\hat{\boldsymbol{\theta}} = \mathbf{h}^{-1}(\hat{\boldsymbol{\mu}}) = \mathbf{g}(\mathbf{x}), \quad (9.6)$$

en principio podría determinarse la PDF de $\hat{\theta}$ empleando las fórmulas convencionales de transformación de variables aleatorias. Sin embargo, en la práctica esto suele ser imposible por su intratabilidad matemática. Como la ecuación 9.6 es la forma general de todo estimador, los métodos descriptos a continuación se aplican a cualquier estimador, además de a aquellos obtenidos mediante el método de los momentos. Estos métodos permiten evaluar el desempeño de un estimador determinando expresiones aproximadas para la media y la varianza. En general, las aproximaciones mejoran cuando se incrementa la cantidad de datos, siendo un enfoque asintótico.

9.3.1. Evaluación cuando el registro de datos es grande

Considérese un parámetro escalar $\hat{\theta}$ que fue estimado empleando la ecuación 9.6. Para determinar la media y varianza aproximada de $\hat{\theta}$ se asumirá que depende de $r < N$ estadísticos $\{T_1(\mathbf{x}), \dots, T_r(\mathbf{x})\}$ cuyas varianzas son pequeñas lo que asegura que la PDF de $[T_1 \dots T_r]^T$ estará concentrada en torno a su media. En el ejemplo de la PDF exponencial de la sección 9.2.1, el estimador de λ se puede escribir como

$$\hat{\lambda} = g(T_1(\mathbf{x})), \quad \text{donde} \quad T_1(\mathbf{x}) = \frac{1}{N} \sum_{n=0}^{N-1} x[n] \quad \text{y} \quad g(T_1) = \frac{1}{T_1}.$$

Para N grande la PDF de T_1 estará fuertemente concentrar en torno a su media, ya que $\text{var}(T_1) = 1/(N\lambda^2)$, como se mostrará en el siguiente ejemplo. Por lo tanto, es posible emplear el argumento de linealización estadística (ver el capítulo 3 de [1]) para aproximar g en torno a la media de T_1 mediante una expansión de Taylor de primer orden. En general, se asume que

$$\hat{\theta} = g(\mathbf{T})$$

donde $\mathbf{T} = [T_1 \dots T_r]^T$. Luego se realiza un desarrollo de Taylor de primer orden de g en torno al punto $\mathbf{T} = E(\mathbf{T}) = \boldsymbol{\mu}$,

$$\begin{aligned} \hat{\theta} = g(\mathbf{T}) &\approx g(\boldsymbol{\mu}) + \left. \frac{\partial g}{\partial \mathbf{T}} \right|_{\mathbf{T}=\boldsymbol{\mu}}^T (\mathbf{T} - \boldsymbol{\mu}) \\ &= g(\boldsymbol{\mu}) + \sum_{k=1}^r \left. \frac{\partial g}{\partial T_k} \right|_{\mathbf{T}=\boldsymbol{\mu}} (T_k - \mu_k), \end{aligned} \quad (9.7)$$

y tomando esperanza se obtiene que

$$E(\hat{\theta}) = g(\boldsymbol{\mu}). \quad (9.8)$$

Con este grado de aproximación se cumple que $E(\hat{\theta}) = E(g(\mathbf{T})) = g(E(\mathbf{T}))$, es decir, que la esperanza conmuta bajo la función no lineal g . Notar que se requiere calcular el primer momento de \mathbf{T} para obtener $\boldsymbol{\mu}$. La aproximación de la varianza es

$$\begin{aligned} \text{var}(\hat{\theta}) &= E[(\hat{\theta} - E(\hat{\theta}))^2] \\ &\stackrel{(a)}{=} E \left\{ \left[g(\boldsymbol{\mu}) + \left. \frac{\partial g}{\partial \mathbf{T}} \right|_{\mathbf{T}=\boldsymbol{\mu}}^T (\mathbf{T} - \boldsymbol{\mu}) - g(\boldsymbol{\mu}) \right]^2 \right\} \\ &= E \left\{ \left[\left. \frac{\partial g}{\partial \mathbf{T}} \right|_{\mathbf{T}=\boldsymbol{\mu}}^T (\mathbf{T} - \boldsymbol{\mu}) \right]^2 \right\} \\ &= E \left[\left. \frac{\partial g}{\partial \mathbf{T}} \right|_{\mathbf{T}=\boldsymbol{\mu}}^T (\mathbf{T} - \boldsymbol{\mu})(\mathbf{T} - \boldsymbol{\mu})^T \left. \frac{\partial g}{\partial \mathbf{T}} \right|_{\mathbf{T}=\boldsymbol{\mu}} \right] \\ &= \left. \frac{\partial g}{\partial \mathbf{T}} \right|_{\mathbf{T}=\boldsymbol{\mu}}^T E[(\mathbf{T} - \boldsymbol{\mu})(\mathbf{T} - \boldsymbol{\mu})^T] \left. \frac{\partial g}{\partial \mathbf{T}} \right|_{\mathbf{T}=\boldsymbol{\mu}} \end{aligned}$$

donde en (a) se reemplazó $\hat{\theta}$ empleando la ecuación 9.7 y $E(\hat{\theta})$ empleando la ecuación 9.8. Finalmente, considerando que $E[(\mathbf{T} - \boldsymbol{\mu})(\mathbf{T} - \boldsymbol{\mu})^T] = \mathbf{C}_T$ es la matriz de covarianza de \mathbf{T} , se obtiene que

$$\text{var}(\hat{\theta}) = \frac{\partial g}{\partial \mathbf{T}} \bigg|_{\mathbf{T}=\boldsymbol{\mu}}^T \mathbf{C}_T \frac{\partial g}{\partial \mathbf{T}} \bigg|_{\mathbf{T}=\boldsymbol{\mu}}. \quad (9.9)$$

Notar que para la varianza se requiere la media y la matriz de covarianza de \mathbf{T} . El razonamiento análogo puede emplearse para calcular los desarrollos de Taylor de mayor orden, como se hace en los problemas de la secciones 9.5.8 y 9.5.9, pero el álgebra se hace cada vez mas tediosa.

Ejemplo: PDF exponencial (continuación)

En el ejemplo de la PDF exponencial de la sección 9.2.1, se obtuvo que el estimador por el método de los momentos es

$$\hat{\lambda} = \frac{1}{\frac{1}{N} \sum_{n=0}^{N-1} x[n]},$$

donde las muestras $x[n]$ son IID con PDF exponencial. Para calcular la media y la varianza, se puede emplear las ecuaciones 9.8 y 9.9. En este caso, se tiene que

$$\hat{\lambda} = g(T_1), \quad \text{donde} \quad T_1 = \frac{1}{N} \sum_{n=0}^{N-1} x[n] \quad \text{y} \quad g(T_1) = \frac{1}{T_1}.$$

La media de T_1 es

$$\mu_1 = E(T_1) = \frac{1}{N} \sum_{n=0}^{N-1} E(x[n]) = E(x[n]) = \frac{1}{\lambda},$$

como se calculó en la sección 9.2.1. De la ecuación 9.8, se obtiene que la media aproximada del estimador es

$$E(\hat{\lambda}) = g(\mu_1) = \frac{1}{\frac{1}{\lambda}} = \lambda.$$

Para calcular la varianza mediante 9.9, se parte observando que la varianza de T_1 es

$$\text{var}(T_1) = \text{var}\left(\frac{1}{N} \sum_{n=0}^{N-1} x[n]\right) = \frac{\text{var}(x[n])}{N}.$$

Además,

$$\text{var}(x[n]) = E(x^2[n]) - E^2(x[n]),$$

con

$$\begin{aligned} E(x^2[n]) &= \int_0^\infty u^2 \lambda e^{-\lambda u} du \\ &\stackrel{(a)}{=} -u^2 e^{-\lambda u} \bigg|_0^\infty + 2 \int_0^\infty u e^{-\lambda u} du \\ &\stackrel{(b)}{=} 2 \left(-u e^{-\lambda u} \bigg|_0^\infty + \frac{1}{\lambda} \int_0^\infty e^{-\lambda u} du \right) \\ &= \frac{2}{\lambda} \left(-\frac{1}{\lambda} e^{-\lambda u} \bigg|_0^\infty \right) \\ &= \frac{2}{\lambda^2}, \end{aligned}$$

donde en (a) se aplicó integración por partes $\int x dy = xy - \int y dx$ con $x = u^2$ y $dy = \lambda e^{-\lambda u} du$ y por lo tanto, $dx = 2u du$ y $y = -e^{-\lambda u}$ y en (b) se tuvo en cuenta que el primer término es nulo y nuevamente

se aplicó integración por partes en la integral con $x = u$ y $dy = e^{-\lambda u} du$ y por lo tanto, $dx = du$ y $y = -(1/\lambda)e^{-\lambda u}$. Combinando los resultados, se obtiene que

$$\text{var}(x[n]) = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2},$$

y por lo tanto,

$$\text{var}(T_1) = \frac{1}{N\lambda^2}.$$

Empleando la ecuación 9.9, la varianza aproximada del estimador es

$$\begin{aligned} \text{var}(\hat{\lambda}) &= \left. \frac{\partial g}{\partial T_1} \right|_{T_1=\mu_1} \text{var}(T_1) \left. \frac{\partial g}{\partial T_1} \right|_{T_1=\mu_1} \\ &\stackrel{(a)}{=} (-\lambda^2) \frac{1}{N\lambda^2} (-\lambda^2) \\ &= \frac{\lambda^2}{N}, \end{aligned}$$

donde en (a) se tuvo en cuenta que como $g(T_1) = 1/T_1$,

$$\left. \frac{\partial g}{\partial T_1} \right|_{T_1=\mu_1} = -\frac{1}{\mu_1^2} = -\lambda^2.$$

Se concluye que el estimador es aproximadamente insesgado y tiene una varianza aproximada que decrece con N . Para este problema, el estimador $\hat{\lambda}$ obtenido por el método de los momentos es el MLE, como se dedujo en el problema de la sección 7.6.3. De esta forma, su PDF asintótica es (ver la sección 7.1.2)

$$\hat{\lambda} \stackrel{a}{\sim} \mathcal{N}(\lambda, \lambda^2/N),$$

ya que la información de Fisher $I(\lambda)$ es λ^2/N (ver el problema 7.6.3). Por lo tanto, la aproximación es buena solo si $N \rightarrow \infty$. Ver el problema de la sección 9.5.10 por la deducción de las aproximaciones de segundo orden de la media y la varianza de $\hat{\lambda}$.

9.3.2. Evaluación cuando la SNR es grande

La premisa principal para la evaluación del desempeño del estimador empleando el enfoque de la aproximación mediante la serie de Taylor de la función g es que g sea aproximadamente lineal en el rango de \mathbf{T} en donde $p(\mathbf{T}; \theta)$ es esencialmente no nula. Esto ocurre naturalmente cuando el registro de datos es grande. Otra situación en donde ocurre es en la estimación de un parámetro de una señal en ruido cuando la SNR es grande. En este caso, expandiendo la función en torno a la señal es posible obtener aproximaciones de la media y la varianza. Si la SNR es grande, los resultados son precisos, ya que que el ruido produce solo una pequeña perturbación del valor del estimador. A continuación se desarrolla este segundo caso.

Se considera el modelo de datos

$$x[n] = s[n; \theta] + w[n], \quad n = 0, \dots, N-1$$

donde $w[n]$ es ruido de media nula y covarianza \mathbf{C} . Un estimador genérico del parámetro θ escalar es

$$\begin{aligned} \hat{\theta} &= g(\mathbf{x}) \\ &= g(\mathbf{s}(\theta) + \mathbf{w}) \\ &= h(\mathbf{w}). \end{aligned}$$

En este caso, se puede elegir el estadístico \mathbf{T} como los datos originales, ya que como la SNR es grande, la PDF de $\mathbf{T} = \mathbf{x}$ está concentrada en torno a su media, que es la señal. Luego, se realiza un desarrollo de Taylor de primer orden de g en torno a la media de \mathbf{x} , que es $\boldsymbol{\mu} = \mathbf{s}(\theta)$, o equivalentemente, de h en torno a $\mathbf{w} = \mathbf{0}$. Esto resulta en

$$\hat{\theta} \approx h(\mathbf{0}) + \sum_{n=0}^{N-1} \left. \frac{\partial h}{\partial w[n]} \right|_{\mathbf{w}=\mathbf{0}} w[n]. \quad (9.10)$$

Siguiendo el mismo razonamiento que en la sección anterior, se cumple de forma aproximada que

$$E(\hat{\theta}) = h(\mathbf{0}) = g(\mathbf{s}(\theta)) \quad (9.11)$$

$$\text{var}(\hat{\theta}) = \left. \frac{\partial h}{\partial \mathbf{w}} \right|_{\mathbf{w}=\mathbf{0}}^T \mathbf{C} \left. \frac{\partial h}{\partial \mathbf{w}} \right|_{\mathbf{w}=\mathbf{0}}. \quad (9.12)$$

En la mayoría de los casos, si no hay ruido, el estimador conduce al valor verdadero del parámetro, ya que $g(\mathbf{s}(\theta)) = \theta$. Por lo tanto, si la SNR es buena, el estimador es insesgado.

9.4. Ejemplo de procesamiento de señales

En este ejemplo, se emplea el método de los momentos y la evaluación del desempeño aproximado para el problema de estimación de la frecuencia. Se asume que se observa

$$x[n] = A \cos(2\pi f_0 n + \phi) + w[n], \quad n = 0, \dots, N-1.$$

donde $w[n]$ es ruido blanco de media nula y varianza σ^2 . El parámetro a estimar es la frecuencia f_0 . Este problema fue discutido en el ejemplo de la sección 7.5.1, en donde se encontró que el MLE de la frecuencia está dado por la ubicación del pico del periodograma. Para calcular el estimador por el método de los momentos, se considerará que la fase ϕ es una variable aleatoria independiente de $w[n]$ con PDF $\mathcal{U}[0, 2\pi]$. Con esta hipótesis, la señal $s[n] = A \cos(2\pi f_0 n + \phi)$ puede verse como la realización de un proceso aleatorio WSS. Efectivamente, la media y la ACF de $s[n]$ son independientes de n , ya que la media es

$$\begin{aligned} E(s[n]) &= E[A \cos(2\pi f_0 n + \phi)] \\ &\stackrel{(a)}{=} \int_0^{2\pi} A \cos(2\pi f_0 n + \phi) \frac{1}{2\pi} d\phi \\ &= \frac{A}{2\pi} \sin(2\pi f_0 n + \phi) \Big|_0^{2\pi} \\ &= \frac{A}{2\pi} [\sin(2\pi f_0 n + 2\pi) - \sin(2\pi f_0 n)] \\ &= 0, \end{aligned}$$

donde en (a) se consideró que la PDF de ϕ es $1/(2\pi)$ en $[0, 2\pi]$ y nula en otro caso. La ACF es

$$\begin{aligned} r_{ss}[n, k] &= E(s[n]s[n+k]) \\ &= E[A^2 \cos(2\pi f_0 n + \phi) \cos(2\pi f_0(n+k) + \phi)] \\ &\stackrel{(a)}{=} A^2 E \left[\frac{1}{2} \cos(4\pi f_0 n + 2\pi f_0 k + 2\phi) + \frac{1}{2} \cos 2\pi f_0 k \right] \\ &\stackrel{(b)}{=} \frac{A^2}{2} \cos 2\pi f_0 k, \end{aligned}$$

donde en (a) se empleó la identidad trigonométrica $2 \cos \theta \cos \varphi = \cos(\theta - \varphi) + \cos(\theta + \varphi)$ y en (b) que la esperanza del primer sumando, que se obtiene de forma análoga a como se realizó para la media, es nula, y el segundo sumando es una constante. La ACF del proceso observado es

$$\begin{aligned} r_{xx}[k] &= r_{ss}[k] + r_{ww}[k] \\ &= \frac{A^2}{2} \cos 2\pi f_0 k + \sigma^2 \delta[k], \end{aligned} \quad (9.13)$$

donde en la primera igualdad se tuvo en cuenta que los procesos $s[n]$ y $w[n]$ son independientes y de media nula. Para simplificar la discusión y el estimador obtenido, se asumirá que la amplitud de la señal es conocida (ver el problema de la sección 9.5.12 para el caso en que la amplitud es desconocida). Sea $A = \sqrt{2}$, de forma de que la ACF es

$$r_{xx}[k] = \cos 2\pi f_0 k + \sigma^2 \delta[k].$$

Para estimar la frecuencia por el método de los momentos, se observa que

$$r_{xx}[1] = \cos 2\pi f_0$$

y por lo tanto, sin asumir el conocimiento de σ^2 , el estimador puede implementarse como

$$\hat{f}_0 = \frac{1}{2\pi} \arccos \hat{r}_{xx}[1]$$

donde $\hat{r}_{xx}[1]$ es un estimador de $r_{xx}[1]$. El estimador natural de la ACF para $k = 1$ es

$$\hat{r}_{xx}[1] = \frac{1}{N-1} \sum_{n=0}^{N-2} x[n]x[n+1]$$

por lo que se concluye que el estimador de la frecuencia es

$$\hat{f}_0 = \frac{1}{2\pi} \arccos \left[\frac{1}{N-1} \sum_{n=0}^{N-2} x[n]x[n+1] \right]. \quad (9.14)$$

Este estimador, si bien fue motivado por el modelo sinusoidal con fase aleatoria, es válido si ϕ es determinístico y desconocido (ver el problema de la sección 9.5.11). En la determinación de la media y la varianza aproximadas se asume que ϕ es determinístico. Para evaluar el desempeño del estimador se usará una expansión de Taylor de primer orden en torno al punto $\mathbf{w} = \mathbf{0}$, y por lo tanto, los cálculos son válidos cuando la SNR es grande. La media se obtiene mediante la ecuación 9.11 con

$$h(\mathbf{w}) = \frac{1}{2\pi} \arccos \left[\frac{1}{N-1} \sum_{n=0}^{N-2} \left(\sqrt{2} \cos(2\pi f_0 n + \phi) + w[n] \right) \left(\sqrt{2} \cos(2\pi f_0 (n+1) + \phi) + w[n+1] \right) \right] \quad (9.15)$$

y evaluando en $\mathbf{w} = \mathbf{0}$, se obtiene que

$$\begin{aligned} E(\hat{f}_0) &= \frac{1}{2\pi} \arccos \left[\frac{1}{N-1} \sum_{n=0}^{N-2} \sqrt{2} \cos(2\pi f_0 n + \phi) \sqrt{2} \cos(2\pi f_0 (n+1) + \phi) \right] \\ &= \frac{1}{2\pi} \arccos \left[\frac{1}{N-1} \sum_{n=0}^{N-2} (\cos(4\pi f_0 n + 2\pi f_0 + 2\phi) + \cos 2\pi f_0) \right]. \end{aligned}$$

La sumatoria del término de doble frecuencia es nula (ver el problema de la sección 3.6.7), por lo que si la SNR es buena, se obtiene que

$$\begin{aligned} E(\hat{f}_0) &= \frac{1}{2\pi} \arccos [\cos 2\pi f_0] \\ &= f_0. \end{aligned}$$

La varianza se obtiene con la ecuación 9.12. De la ecuación 9.15,

$$h(\mathbf{w}) = \frac{1}{2\pi} \arccos[u(\mathbf{w})] \quad \text{con} \quad u(\mathbf{w}) = \frac{1}{N-1} \sum_{n=0}^{N-2} (s[n] + w[n])(s[n+1] + w[n+1]),$$

por lo que las derivadas primeras son

$$\begin{aligned} \left. \frac{\partial h}{\partial w[i]} \right|_{\mathbf{w}=\mathbf{0}} &= \left. \frac{\partial h}{\partial u} \frac{\partial u}{\partial w[i]} \right|_{\mathbf{w}=\mathbf{0}} \\ &= \frac{1}{2\pi} \frac{\partial \arccos u}{\partial u} \frac{\partial u}{\partial w[i]} \bigg|_{\mathbf{w}=\mathbf{0}} \\ &= -\frac{1}{2\pi} \frac{1}{\sqrt{1-u^2}} \frac{\partial u}{\partial w[i]} \bigg|_{\mathbf{w}=\mathbf{0}}. \end{aligned}$$

Además,

$$u(\mathbf{w})|_{\mathbf{w}=\mathbf{0}} = \frac{1}{N-1} \sum_{n=0}^{N-2} s[n]s[n+1] \\ \approx \cos 2\pi f_0,$$

como se mostró previamente y por lo tanto

$$\left. \frac{\partial h}{\partial w[i]} \right|_{\mathbf{w}=\mathbf{0}} = -\frac{1}{2\pi\sqrt{1-\cos^2 2\pi f_0}} \left. \frac{\partial u}{\partial w[i]} \right|_{\mathbf{w}=\mathbf{0}} \\ = -\frac{1}{2\pi \sin 2\pi f_0} \left. \frac{\partial u}{\partial w[i]} \right|_{\mathbf{w}=\mathbf{0}}$$

También, como

$$u(\mathbf{w}) = \frac{1}{N-1} [(s[0] + w[0])(s[1] + w[1]) + \dots + (s[i-1] + w[i-1])(s[i] + w[i]) \\ + (s[i] + w[i])(s[i+1] + w[i+1]) + \dots + (s[N-2] + w[N-2])(s[N-1] + w[N-1])],$$

se cumple que

$$\left. \frac{\partial u}{\partial w[i]} \right|_{\mathbf{w}=\mathbf{0}} = \begin{cases} \frac{1}{N-1}(s[1] + w[1]), & i = 0 \\ \frac{1}{N-1} [(s[i-1] + w[i-1]) + (s[i+1] + w[i+1])], & i = 1, \dots, N-2 \\ \frac{1}{N-1}(s[N-2] + w[N-2]), & i = N-1, \end{cases}$$

y por lo tanto,

$$\left. \frac{\partial u}{\partial w[i]} \right|_{\mathbf{w}=\mathbf{0}} = \begin{cases} \frac{1}{N-1} s[1], & i = 0 \\ \frac{1}{N-1} (s[i-1] + s[i+1]), & i = 1, \dots, N-2 \\ \frac{1}{N-1} s[N-2], & i = N-1. \end{cases}$$

Finalmente,

$$\left. \frac{\partial h}{\partial w[i]} \right|_{\mathbf{w}=\mathbf{0}} = \begin{cases} -\frac{1}{2\pi(N-1)\sin 2\pi f_0} s[1], & i = 0 \\ -\frac{1}{2\pi(N-1)\sin 2\pi f_0} (s[i-1] + s[i+1]), & i = 1, \dots, N-2 \\ -\frac{1}{2\pi(N-1)\sin 2\pi f_0} s[N-2], & i = N-1. \end{cases}$$

Considerando este resultado, la varianza (ecuación 9.12) es

$$\text{var}(\hat{f}_0) = \left. \frac{\partial h}{\partial \mathbf{w}} \right|_{\mathbf{w}=\mathbf{0}}^T \mathbf{C} \left. \frac{\partial h}{\partial \mathbf{w}} \right|_{\mathbf{w}=\mathbf{0}} \\ \stackrel{(a)}{=} \sigma^2 \sum_{n=0}^{N-1} \left(\left. \frac{\partial h}{\partial w[n]} \right|_{\mathbf{w}=\mathbf{0}} \right)^2 \\ = \frac{\sigma^2}{2\pi(N-1)\sin 2\pi f_0} \left[s^2[1] + \sum_{n=1}^{N-2} (s[n-1] + s[n+1])^2 + s^2[N-2] \right]$$

donde en (a) se tuvo en cuenta que $\mathbf{C} = \sigma^2 \mathbf{I}$. Observando que

$$\begin{aligned} s[n-1] + s[n+1] &= \sqrt{2} \cos(2\pi f_0(n-1) + \phi) + \sqrt{2} \cos(2\pi f_0(n+1) + \phi) \\ &\stackrel{(a)}{=} 2\sqrt{2} \cos\left(\frac{4\pi f_0 n + 2\phi}{2}\right) \cos\left(\frac{4\pi f_0}{2}\right) \\ &= 2\sqrt{2} \cos(2\pi f_0 n + \phi) \cos 2\pi f_0 \\ &= 2s[n] \cos 2\pi f_0, \end{aligned}$$

donde en (a) se empleó la identidad trigonométrica $\cos \theta + \cos \varphi = 2 \cos\left(\frac{\theta+\varphi}{2}\right) \cos\left(\frac{\theta-\varphi}{2}\right)$, se llega a que

$$\text{var}(\hat{f}_0) = \frac{\sigma^2}{2\pi(N-1) \sin 2\pi f_0} \left[s^2[1] + 4 \cos^2 2\pi f_0 \sum_{n=1}^{N-2} s^2[n] + s^2[N-2] \right] \quad (9.16)$$

Observar que la varianza decrece como $1/N^2$ mientras que la CRLB decrece como $1/N^3$, como indica la ecuación 3.41. Además, en cuanto a la dependencia con f_0 la varianza crece rápidamente cuando f_0 se acerca a 0 o a $1/2$.

9.5. Problemas

9.5.1. Problema 1

Si se realizan N observaciones $\{x[0], x[1], \dots, x[N-1]\}$ IID de una PDF de Rayleigh

$$p(x; \sigma^2) = \begin{cases} \frac{x}{\sigma^2} \exp\left(-\frac{1}{2} \frac{x^2}{\sigma^2}\right), & x > 0 \\ 0, & x < 0 \end{cases}$$

encontrar un estimador por el método de los momentos de σ^2 .

Solución El primer momento de una variable aleatoria de Rayleigh es

$$E(x) = \int_{-\infty}^{\infty} \frac{x^2}{\sigma^2} \exp\left(-\frac{1}{2} \frac{x^2}{\sigma^2}\right) dx.$$

Para calcular la integral, se considera una variable aleatoria $\mathcal{N}(0, \sigma^2)$, de la cual se sabe que el segundo momento es $E(x^2) = \sigma^2$, y por lo tanto

$$\begin{aligned} \sigma^2 &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} x^2 \exp\left(-\frac{1}{2} \frac{x^2}{\sigma^2}\right) dx \\ &\stackrel{(a)}{=} \frac{2}{\sqrt{2\pi\sigma^2}} \int_0^{\infty} x^2 \exp\left(-\frac{1}{2} \frac{x^2}{\sigma^2}\right) dx \\ &\stackrel{(b)}{=} \sqrt{\frac{2}{\pi}} \sigma \int_0^{\infty} \frac{x^2}{\sigma^2} \exp\left(-\frac{1}{2} \frac{x^2}{\sigma^2}\right) dx, \end{aligned}$$

donde en (a) se consideró que el integrando es una función par y en (b) se multiplicó y dividió entre σ . Observando que la integral es la integral que se quería calcular, despejando se obtiene que

$$E(x) = \int_{-\infty}^{\infty} \frac{x^2}{\sigma^2} \exp\left(-\frac{1}{2} \frac{x^2}{\sigma^2}\right) dx = \sqrt{\frac{\pi}{2}} \sigma,$$

y por lo tanto,

$$E^2(x) = \frac{\pi}{2} \sigma^2 \quad \Rightarrow \quad \sigma^2 = \frac{2}{\pi} E^2(x).$$

Se concluye que el estimador por el método de los momentos es

$$\hat{\sigma}^2 = \frac{2}{\pi} \left(\frac{1}{N} \sum_{n=0}^{N-1} x[n] \right)^2 = \frac{2}{\pi} \bar{x}^2.$$

Este estimador no es insesgado. Efectivamente,

$$\begin{aligned}
E(\hat{\sigma}^2) &= \frac{2}{\pi N^2} E \left[\left(\sum_{n=0}^{N-1} x[n] \right)^2 \right] \\
&= \frac{2}{\pi N^2} E \left(\sum_{m=0}^{N-1} \sum_{n=0}^{N-1} x[m]x[n] \right) \\
&= \frac{2}{\pi N^2} E \left(\sum_{n=0}^{N-1} x^2[n] + \sum_{\substack{m=0 \\ m \neq n}}^{N-1} \sum_{n=0}^{N-1} x[m]x[n] \right) \\
&= \frac{2}{\pi N^2} \left[\sum_{n=0}^{N-1} E(x^2[n]) + \sum_{\substack{m=0 \\ m \neq n}}^{N-1} \sum_{n=0}^{N-1} E(x[m])E(x[n]) \right] \\
&= \frac{2}{\pi N^2} [NE(x^2[n]) + N(N-1)E^2(x[n])] .
\end{aligned}$$

El momento segundo de una variable aleatoria de Rayleigh es $E(x^2[n]) = 2\sigma^2$, como se calculó en el problema de la sección 5.4.15. Se concluye que

$$E(\hat{\sigma}^2) = \frac{2}{\pi N^2} \left[2N\sigma^2 + \frac{\pi}{2}N(N-1)\sigma^2 \right] ,$$

lo que implica que el estimador no es insesgado. Sin embargo,

$$\lim_{n \rightarrow \infty} E(\hat{\sigma}^2) = \frac{2}{\pi N^2} \left(\frac{\pi}{2}N^2\sigma^2 \right) = \sigma^2 ,$$

lo que implica que es asintóticamente insesgado. Además, como se indicó previamente, es consistente.

En el problema de la sección 5.4.15 se dedujo que el segundo momento de una variable aleatoria de Rayleigh es

$$E(x^2) = 2\sigma^2 ,$$

por lo que otro estimador por el método de los momentos es

$$\check{\sigma}^2 = \frac{1}{2N} \sum_{n=0}^{N-1} x^2[n] .$$

Este estimador es insesgado,

$$E(\check{\sigma}^2) = \frac{1}{2N} \sum_{n=0}^{N-1} E(x^2[n]) = \frac{1}{2N} (2N\sigma^2) = \sigma^2 .$$

Además, como se encontró en el mismo problema, es el estimador MVU.

9.5.2. Problema 2

Si se realizan N observaciones $\{x[0], x[1], \dots, x[N-1]\}$ IID de una PDF de Laplace

$$p(x; \sigma^2) = \frac{1}{\sqrt{2}\sigma} \exp \left(-\frac{\sqrt{2}|x|}{\sigma} \right)$$

encontrar un estimador por el método de los momentos de σ .

Solución Como la PDF es simétrica, el momento primero es nulo por lo que no puede emplearse para el método de los momentos. El momento segundo es

$$\begin{aligned}
E(x^2) &= \int_{-\infty}^{\infty} \frac{x^2}{\sqrt{2}\sigma} \exp\left(-\frac{\sqrt{2}|x|}{\sigma}\right) dx \\
&= \frac{\sqrt{2}}{\sigma} \int_0^{\infty} x^2 \exp\left(-\frac{\sqrt{2}x}{\sigma}\right) dx \\
&\stackrel{(a)}{=} -x^2 \exp\left(-\frac{\sqrt{2}x}{\sigma}\right) \Big|_0^{\infty} + 2 \int_0^{\infty} x \exp\left(-\frac{\sqrt{2}x}{\sigma}\right) dx \\
&\stackrel{(b)}{=} -\sqrt{2}\sigma x \exp\left(-\frac{\sqrt{2}x}{\sigma}\right) \Big|_0^{\infty} + \sqrt{2}\sigma \int_0^{\infty} \exp\left(-\frac{\sqrt{2}x}{\sigma}\right) dx \\
&= \sqrt{2}\sigma \left[-\frac{\sigma}{\sqrt{2}} \exp\left(-\frac{\sqrt{2}x}{\sigma}\right) \Big|_0^{\infty} \right] \\
&= \sigma^2,
\end{aligned}$$

donde en (a) se aplicó integración por partes $\int u dv = uv - \int v du$ con $u = x^2$ y $dv = (\sqrt{2}/\sigma)e^{-\sqrt{2}x/\sigma} dx$ y por lo tanto, $du = 2x dx$ y $v = -e^{-\sqrt{2}x/\sigma}$ y en (b) se tuvo en cuenta que el primer término es nulo y nuevamente se aplicó integración por partes en la integral con $u = \sqrt{2}x$ y $dv = \sqrt{2}e^{-\sqrt{2}x/\sigma} dx$ y por lo tanto, $du = \sqrt{2} dx$ y $v = -\sigma e^{-\sqrt{2}x/\sigma}$. Como

$$\sigma = \sqrt{E(x^2)},$$

un estimador por el método de los momentos es

$$\hat{\sigma} = \sqrt{\frac{1}{N} \sum_{n=0}^{N-1} x^2[n]}.$$

En el ejemplo de la estimación de los parámetros de ruido de Laplace de la sección 3.1.1 se encontró que el estimador MVU eficiente de σ es

$$\hat{\sigma} = \frac{\sqrt{2}}{N} \sum_{n=0}^{N-1} |x[n]|.$$

9.5.3. Problema 3

Asúmase que se observan N muestras $\{\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{N-1}\}$ IID de una PDF gaussiana bivariada, donde cada \mathbf{x} es un vector aleatorio 2×1 con PDF $\mathbf{x} \sim \mathcal{N}(0, \mathbf{C})$. Si

$$\mathbf{C} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

encontrar un estimador de ρ por el método de los momentos. Además determinar una ecuación cúbica a resolver para el MLE de ρ . Comentar sobre la complejidad de implementación de los dos estimadores.

Solución Sea $\mathbf{x}_n = [u_n \ v_n]^T$. Teniendo en cuenta que

$$\rho = \text{cov}(u_n v_n) = E(u_n v_n) - E(u_n)E(v_n) = E(u_n v_n),$$

el estimador por el método de los momentos es

$$\hat{\rho} = \frac{1}{N} \sum_{n=0}^{N-1} u_n v_n.$$

En el problema de la sección 7.6.11 se encontró que el MLE de ρ cumple que

$$\hat{\rho}(1 - \hat{\rho}^2) + (1 + \hat{\rho}^2) \frac{1}{N} \sum_{n=0}^{N-1} u_n v_n - \hat{\rho} \left(\frac{1}{N} \sum_{n=0}^{N-1} u_n^2 + \frac{1}{N} \sum_{n=0}^{N-1} v_n^2 \right) = 0.$$

Claramente, el estimador por el método de los momentos es mas fácil de implementar.

9.5.4. Problema 4

Si se observan N muestras $\{x[0], x[1], \dots, x[N-1]\}$ IID de una PDF $\mathcal{N}(\mu, \sigma^2)$, encontrar un estimador por el método de los momentos de $\boldsymbol{\theta} = [\mu \ \sigma^2]^T$.

Solución Teniendo en cuenta que

$$\mu = E(x[n]), \quad \sigma^2 = E(x^2[n]) - E^2(x[n]),$$

los estimadores por el método de los momentos son

$$\hat{\mu} = \frac{1}{N} \sum_{n=0}^{N-1} x[n] = \bar{x}, \quad \hat{\sigma}^2 = \frac{1}{N} \sum_{n=0}^{N-1} x^2[n] - \bar{x}^2.$$

9.5.5. Problema 5

En el ejemplo de la sección 8.7.2 se determinó que la ACF de un proceso ARMA satisface la ecuación en diferencias recursiva (ecuación 8.40)

$$r_{xx}[n] = - \sum_{k=1}^p a[k] r_{xx}[n-k] \quad \text{si } n > q,$$

donde $\{a[1], \dots, a[p]\}$ son los coeficientes AR del filtro y q es el orden del filtro MA. Proponer un estimador por el método de los momentos para los coeficientes AR del filtro.

Solución La ACF se define como

$$r_{xx}[k] = E(x[n]x[n+k]),$$

por lo que su estimador natural es

$$\hat{r}_{xx}[k] = \frac{1}{N - |k|} \sum_{n=0}^{N-1-|k|} x[n]x[n+|k|].$$

De esta forma, las ecuaciones para el método de los momentos son

$$\sum_{k=1}^p \hat{a}[k] \hat{r}_{xx}[n-k] = -\hat{r}_{xx}[n] \quad \text{si } n > q.$$

Como hay p parámetros a determinar, eligiendo $n = q+1, \dots, q+p$, se obtiene el sistema lineal de ecuaciones

$$\begin{cases} \hat{a}[1]\hat{r}_{xx}[q] & + & \hat{a}[2]\hat{r}_{xx}[q-1] & + & \dots & + & \hat{a}[p]\hat{r}_{xx}[q-p+1] & = & -\hat{r}_{xx}[q+1] \\ \hat{a}[1]\hat{r}_{xx}[q+1] & + & \hat{a}[2]\hat{r}_{xx}[q] & + & \dots & + & \hat{a}[p]\hat{r}_{xx}[q-p+2] & = & -\hat{r}_{xx}[q+2] \\ \vdots & & \vdots & & & & \vdots & & \vdots \\ \hat{a}[1]\hat{r}_{xx}[q+p-1] & + & \hat{a}[2]\hat{r}_{xx}[q+p-2] & + & \dots & + & \hat{a}[p]\hat{r}_{xx}[q] & = & -\hat{r}_{xx}[q+p] \end{cases}$$

que en forma matricial se puede escribir como

$$\begin{bmatrix} \hat{r}_{xx}[q] & \hat{r}_{xx}[q-1] & \dots & \hat{r}_{xx}[q-p+1] \\ \hat{r}_{xx}[q+1] & \hat{r}_{xx}[q] & \dots & \hat{r}_{xx}[q-p+2] \\ \vdots & \vdots & \ddots & \vdots \\ \hat{r}_{xx}[q+p-1] & \hat{r}_{xx}[q+p-2] & \dots & \hat{r}_{xx}[q] \end{bmatrix} \begin{bmatrix} \hat{a}[1] \\ \hat{a}[2] \\ \vdots \\ \hat{a}[p] \end{bmatrix} = - \begin{bmatrix} \hat{r}_{xx}[q+1] \\ \hat{r}_{xx}[q+2] \\ \vdots \\ \hat{r}_{xx}[q+p] \end{bmatrix}.$$

Resolviendo el sistema de p ecuaciones con p incógnitas se obtiene el estimador de los parámetros AR por el método de los momentos. Notar que en este caso, los momentos involucrados son los momentos cruzados.

9.5.6. Problema 6

Para el nivel de DC en WGN o

$$x[n] = A + w[n], \quad n = 0, \dots, N-1$$

donde $w[n]$ es WGN con varianza σ^2 , se quiere estimar A^2 . Se propone usar

$$\widehat{A^2} = (\bar{x})^2.$$

Para este estimador, encontrar la media y varianza aproximadas empleando el enfoque de la expansión de Taylor de primer orden.

Solución La aproximación de Taylor de primer orden de la media y la varianza están dadas por las ecuaciones 9.8 y 9.9 con

$$T = \bar{x}, \quad \text{y} \quad g(T) = T^2.$$

Como

$$\mu = E(T) = E(\bar{x}) = A,$$

de la ecuación 9.8, la aproximación de la media es

$$E(\widehat{A^2}) = g(\mu) = A^2.$$

Además, teniendo en cuenta que

$$\text{var}(T) = \frac{\sigma^2}{N} \quad \text{y} \quad \left. \frac{\partial g}{\partial T} \right|_{T=\mu} = 2T \Big|_{T=\mu} = 2\mu,$$

de la ecuación 9.9, la aproximación de la varianza es

$$\text{var}(\widehat{A^2}) = \text{var}(T) \left(\left. \frac{\partial g}{\partial T} \right|_{T=\mu} \right)^2 = \frac{4\mu^2\sigma^2}{N}$$

resultando en

$$\text{var}(\widehat{A^2}) = \frac{4A^2\sigma^2}{N}.$$

9.5.7. Problema 7

Para los datos observados

$$x[n] = \cos \phi + w[n], \quad n = 0, \dots, N-1$$

donde $w[n]$ es WGN con varianza σ^2 , encontrar un estimador por el método de los momentos para ϕ . Asumiendo que la SNR es grande, determinar la media y la varianza aproximada para el estimador empleando el enfoque de la expansión de Taylor de primer orden.

Solución Como

$$E(x[n]) = \cos \phi$$

y por lo tanto

$$\phi = \arccos [E(x[n])]$$

el estimador por el método de los momentos es

$$\hat{\phi} = \arccos(\bar{x}).$$

Asumiendo que la SNR es grande, la aproximación de la media y la varianza están dadas por las ecuaciones 9.11 y 9.12 con

$$\hat{\phi} = h(\mathbf{w}) = \arccos \left[\frac{1}{N} \sum_{n=0}^{N-1} (\cos \phi + w[n]) \right].$$

De la ecuación 9.11, la aproximación de la media es

$$E(\hat{\phi}) = h(\mathbf{0}) = \arccos \left(\frac{1}{N} \sum_{n=0}^{N-1} \cos \phi \right) = \arccos(\cos \phi),$$

resultando en

$$E(\hat{\phi}) = \phi.$$

La aproximación de la varianza está dada por la ecuación 9.12. Como

$$h(\mathbf{w}) = \arccos[u(\mathbf{w})] \quad \text{con} \quad u(\mathbf{w}) = \frac{1}{N} \sum_{n=0}^{N-1} (\cos \phi + w[n]),$$

la derivada es

$$\begin{aligned} \left. \frac{\partial h}{\partial w[i]} \right|_{\mathbf{w}=\mathbf{0}} &= \left. \frac{\partial h}{\partial u} \frac{\partial u}{\partial w[i]} \right|_{\mathbf{w}=\mathbf{0}} \\ &= \frac{\partial \arccos u}{\partial u} \frac{\partial u}{\partial w[i]} \left[\frac{1}{N} \sum_{n=0}^{N-1} (\cos \phi + w[n]) \right] \\ &= \frac{1}{\sqrt{1-u^2}} \left(-\frac{1}{N} \right) \Big|_{\mathbf{w}=\mathbf{0}} \\ &\stackrel{(a)}{=} -\frac{1}{N} \frac{1}{\sqrt{1-\cos^2 \phi}} \\ &= -\frac{1}{N \sin \phi}, \end{aligned}$$

por lo que la varianza queda

$$\begin{aligned} \text{var}(\hat{\theta}) &= \left. \frac{\partial h}{\partial \mathbf{w}} \right|_{\mathbf{w}=\mathbf{0}}^T \mathbf{C} \left. \frac{\partial h}{\partial \mathbf{w}} \right|_{\mathbf{w}=\mathbf{0}} \\ &\stackrel{(a)}{=} \sigma^2 \sum_{n=0}^{N-1} \left(\left. \frac{\partial h}{\partial w[n]} \right|_{\mathbf{w}=\mathbf{0}} \right)^2 \\ &= \sigma^2 \sum_{n=0}^{N-1} \frac{1}{N^2 \sin^2 \phi}, \end{aligned}$$

donde en (a) se consideró que $\mathbf{C} = \sigma^2 \mathbf{I}$. Se concluye que

$$\text{var}(\hat{\theta}) = \frac{\sigma^2}{N \sin^2 \phi}$$

9.5.8. Problema 8

En este problema se determina la aproximación de segundo orden de la media de un estimador empleando un desarrollo de Taylor de segundo orden. Para hacerlo, se expande $\hat{\theta} = g(\mathbf{T})$ en torno a la media de \mathbf{T} o $\mathbf{T} = \boldsymbol{\mu}$ manteniendo hasta el término cuadrático inclusive. Mostrar que la media aproximada de $\hat{\theta}$ es

$$E(\hat{\theta}) = g(\boldsymbol{\mu}) + \frac{1}{2} \text{tr}[\mathbf{G}(\boldsymbol{\mu})\mathbf{C}_T] \quad (9.17)$$

donde

$$[\mathbf{G}(\boldsymbol{\mu})]_{ij} = \left. \frac{\partial^2 g}{\partial T_i \partial T_j} \right|_{\mathbf{T}=\boldsymbol{\mu}}.$$

Sugerencia: emplear el resultado

$$E(\mathbf{x}^T \mathbf{y}) = E(\text{tr}(\mathbf{y}\mathbf{x}^T)) = \text{tr}(E(\mathbf{y}\mathbf{x}^T)).$$

Solución El desarrollo de Taylor de segundo orden de $\hat{\theta} = g(\mathbf{T})$ en torno a su media $\boldsymbol{\mu}$ es

$$\hat{\theta} = g(\mathbf{T}) \approx g(\boldsymbol{\mu}) + \left. \frac{\partial g}{\partial \mathbf{T}} \right|_{\mathbf{T}=\boldsymbol{\mu}}^T (\mathbf{T} - \boldsymbol{\mu}) + \frac{1}{2} (\mathbf{T} - \boldsymbol{\mu})^T \mathbf{G}(\boldsymbol{\mu}) (\mathbf{T} - \boldsymbol{\mu}).$$

Tomando esperanza, se obtiene que

$$\begin{aligned} E(\hat{\theta}) &= g(\boldsymbol{\mu}) + \frac{1}{2} E[(\mathbf{T} - \boldsymbol{\mu})^T \mathbf{G}(\boldsymbol{\mu}) (\mathbf{T} - \boldsymbol{\mu})] \\ &\stackrel{(a)}{=} g(\boldsymbol{\mu}) + \frac{1}{2} \text{tr} \{ E[\mathbf{G}(\boldsymbol{\mu}) (\mathbf{T} - \boldsymbol{\mu}) (\mathbf{T} - \boldsymbol{\mu})^T] \} \\ &= g(\boldsymbol{\mu}) + \frac{1}{2} \text{tr} \{ \mathbf{G}(\boldsymbol{\mu}) E[(\mathbf{T} - \boldsymbol{\mu}) (\mathbf{T} - \boldsymbol{\mu})^T] \}, \end{aligned}$$

donde en (a) se empleó el resultado sugerido $E(\mathbf{x}^T \mathbf{y}) = \text{tr}(E(\mathbf{y}\mathbf{x}^T))$, el cual es fácil de mostrar, con $\mathbf{x} = (\mathbf{T} - \boldsymbol{\mu})$ y $\mathbf{y} = \mathbf{G}(\boldsymbol{\mu})(\mathbf{T} - \boldsymbol{\mu})$. Finalmente, teniendo en cuenta que $E[(\mathbf{T} - \boldsymbol{\mu})(\mathbf{T} - \boldsymbol{\mu})^T]$ es la matriz de covarianza \mathbf{C}_T de \mathbf{T} se concluye que

$$E(\hat{\theta}) = g(\boldsymbol{\mu}) + \frac{1}{2} \text{tr}[\mathbf{G}(\boldsymbol{\mu})\mathbf{C}_T],$$

que es lo que se quería mostrar.

9.5.9. Problema 9

En este problema se determina la aproximación de segundo orden de la varianza de un estimador empleando un desarrollo de Taylor de segundo orden. Para hacerlo, se expande $\hat{\theta} = g(\mathbf{T})$ en torno a la media de \mathbf{T} o $\mathbf{T} = \boldsymbol{\mu}$ manteniendo hasta el término cuadrático inclusive. Luego, se emplea el resultado del problema de la sección 9.5.8 y finalmente se realiza la hipótesis de que $\mathbf{T} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{C}_T)$. Verificar que se cumple el siguiente resultado:

$$\text{var}(\hat{\theta}) = \left. \frac{\partial g}{\partial \mathbf{T}} \right|_{\mathbf{T}=\boldsymbol{\mu}}^T \mathbf{C}_T \left. \frac{\partial g}{\partial \mathbf{T}} \right|_{\mathbf{T}=\boldsymbol{\mu}} + \frac{1}{2} \text{tr}[(\mathbf{G}(\boldsymbol{\mu})\mathbf{C}_T)^2]. \quad (9.18)$$

La matriz $\mathbf{G}(\boldsymbol{\mu})$ de dimensiones $r \times r$ es la definida en el problema 9.5.8. Sugerencia: emplear el siguiente resultado, que es válido para una matriz \mathbf{A} simétrica $N \times N$,

$$\text{var}(\mathbf{x}^T \mathbf{A} \mathbf{x}) = 2 \text{tr}[(\mathbf{A} \mathbf{C}_x)^2]$$

si $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_x)$.

Solución El desarrollo de Taylor de segundo orden de $\hat{\theta} = g(\mathbf{T})$ en torno a su media $\boldsymbol{\mu}$ es

$$\hat{\theta} = g(\mathbf{T}) \approx g(\boldsymbol{\mu}) + \left. \frac{\partial g}{\partial \mathbf{T}} \right|_{\mathbf{T}=\boldsymbol{\mu}}^T (\mathbf{T} - \boldsymbol{\mu}) + \frac{1}{2} (\mathbf{T} - \boldsymbol{\mu})^T \mathbf{G}(\boldsymbol{\mu}) (\mathbf{T} - \boldsymbol{\mu}).$$

La varianza de $\hat{\theta}$ es

$$\begin{aligned} \text{var}(\hat{\theta}) &= E \left\{ \left[\hat{\theta} - E(\hat{\theta}) \right]^2 \right\} \\ &\stackrel{(a)}{=} E \left\{ \left[g(\boldsymbol{\mu}) + \left. \frac{\partial g}{\partial \mathbf{T}} \right|_{\mathbf{T}=\boldsymbol{\mu}}^T (\mathbf{T} - \boldsymbol{\mu}) + \frac{1}{2} (\mathbf{T} - \boldsymbol{\mu})^T \mathbf{G}(\boldsymbol{\mu}) (\mathbf{T} - \boldsymbol{\mu}) - g(\boldsymbol{\mu}) - \frac{1}{2} \text{tr} [\mathbf{G}(\boldsymbol{\mu}) \mathbf{C}_T] \right]^2 \right\} \\ &= E \left\{ \left[\left. \frac{\partial g}{\partial \mathbf{T}} \right|_{\mathbf{T}=\boldsymbol{\mu}}^T (\mathbf{T} - \boldsymbol{\mu}) + \frac{1}{2} (\mathbf{T} - \boldsymbol{\mu})^T \mathbf{G}(\boldsymbol{\mu}) (\mathbf{T} - \boldsymbol{\mu}) - \frac{1}{2} \text{tr} [\mathbf{G}(\boldsymbol{\mu}) \mathbf{C}_T] \right]^2 \right\}, \end{aligned}$$

donde en (a) se substituyó $\hat{\theta}$ por su desarrollo de Taylor y $E(\hat{\theta})$ por el resultado obtenido en el problema 9.5.8. Definiendo

$$\mathbf{x} = (\mathbf{T} - \boldsymbol{\mu}), \quad \mathbf{b} = \left. \frac{\partial g}{\partial \mathbf{T}} \right|_{\mathbf{T}=\boldsymbol{\mu}}, \quad \mathbf{A} = \mathbf{G}(\boldsymbol{\mu}),$$

donde \mathbf{A} es una matriz simétrica por ser la matriz hessiana de g , se tiene que

$$\begin{aligned} \text{var}(\hat{\theta}) &= E \left\{ \left[\mathbf{b}^T \mathbf{x} + \frac{1}{2} \mathbf{x}^T \mathbf{A} \mathbf{x} - \frac{1}{2} \text{tr} (\mathbf{A} \mathbf{C}_T) \right]^2 \right\} \\ &= E \left[\mathbf{b}^T \mathbf{x} \mathbf{x}^T \mathbf{b} + \frac{1}{4} (\mathbf{x}^T \mathbf{A} \mathbf{x})^2 + \frac{1}{4} \text{tr}^2 (\mathbf{A} \mathbf{C}_T) \right. \\ &\quad \left. + (\mathbf{b}^T \mathbf{x}) (\mathbf{x}^T \mathbf{A} \mathbf{x}) - (\mathbf{b}^T \mathbf{x}) \text{tr} (\mathbf{A} \mathbf{C}_T) - \frac{1}{2} \mathbf{x}^T \mathbf{A} \mathbf{x} \text{tr} (\mathbf{A} \mathbf{C}_T) \right] \\ &= \mathbf{b}^T E(\mathbf{x} \mathbf{x}^T) \mathbf{b} + \frac{1}{4} E[(\mathbf{x}^T \mathbf{A} \mathbf{x})^2] + \frac{1}{4} \text{tr}^2 (\mathbf{A} \mathbf{C}_T) \\ &\quad + E[(\mathbf{b}^T \mathbf{x}) (\mathbf{x}^T \mathbf{A} \mathbf{x})] - [\mathbf{b}^T E(\mathbf{x})] \text{tr} (\mathbf{A} \mathbf{C}_T) - \frac{1}{2} E(\mathbf{x}^T \mathbf{A} \mathbf{x}) \text{tr} (\mathbf{A} \mathbf{C}_T) \\ &= \mathbf{b}^T \mathbf{C}_T \mathbf{b} + \frac{1}{4} E[(\mathbf{x}^T \mathbf{A} \mathbf{x})^2] + \frac{1}{4} \text{tr}^2 (\mathbf{A} \mathbf{C}_T) - \frac{1}{2} E(\mathbf{x}^T \mathbf{A} \mathbf{x}) \text{tr} (\mathbf{A} \mathbf{C}_T), \end{aligned}$$

donde en la última igualdad se consideró que $\mathbf{x} = (\mathbf{T} - \boldsymbol{\mu}) \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_T)$ por lo que los momentos impares son nulos. Teniendo en cuenta la identidad $E(\mathbf{x}^T \mathbf{y}) = \text{tr}(E(\mathbf{y} \mathbf{x}^T))$, se cumple que

$$E(\mathbf{x}^T \mathbf{A} \mathbf{x}) = \text{tr}[E(\mathbf{A} \mathbf{x} \mathbf{x}^T)] = \text{tr}[\mathbf{A} E(\mathbf{x} \mathbf{x}^T)] = \text{tr}(\mathbf{A} \mathbf{C}_T).$$

Además, considerando la identidad dada por la ecuación 3.35 que indica que

$$E[\mathbf{y}^T \mathbf{A} \mathbf{y} \mathbf{y}^T \mathbf{B} \mathbf{y}] = \text{tr}(\mathbf{A} \mathbf{C}) \text{tr}(\mathbf{B} \mathbf{C}) + 2 \text{tr}(\mathbf{A} \mathbf{C} \mathbf{B} \mathbf{C}),$$

donde $\mathbf{y} \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$ y \mathbf{A} y \mathbf{B} son matrices simétricas y fue demostrada en la sección 3.4, se tiene que

$$E[(\mathbf{x}^T \mathbf{A} \mathbf{x})^2] = \text{tr}^2(\mathbf{A} \mathbf{C}_T) + 2 \text{tr}[(\mathbf{A} \mathbf{C}_T)^2].$$

Sustituyendo estos resultados en la expresión de la varianza, se obtiene que

$$\begin{aligned} \text{var}(\hat{\theta}) &= \mathbf{b}^T \mathbf{C}_T \mathbf{b} + \frac{1}{4} \{ \text{tr}^2(\mathbf{A} \mathbf{C}_T) + 2 \text{tr}[(\mathbf{A} \mathbf{C}_T)^2] \} + \frac{1}{4} \text{tr}^2(\mathbf{A} \mathbf{C}_T) - \frac{1}{2} \text{tr}^2(\mathbf{A} \mathbf{C}_T) \\ &= \mathbf{b}^T \mathbf{C}_T \mathbf{b} + \frac{1}{2} \text{tr}[(\mathbf{A} \mathbf{C}_T)^2], \end{aligned}$$

es decir,

$$\text{var}(\hat{\theta}) = \left. \frac{\partial g}{\partial \mathbf{T}} \right|_{\mathbf{T}=\boldsymbol{\mu}}^T \mathbf{C}_T \left. \frac{\partial g}{\partial \mathbf{T}} \right|_{\mathbf{T}=\boldsymbol{\mu}} + \frac{1}{2} \text{tr}[(\mathbf{G}(\boldsymbol{\mu}) \mathbf{C}_T)^2],$$

que es lo que se quería mostrar.

9.5.10. Problema 10

Empleando los resultados de los problemas 9.5.8 y 9.5.9 encontrar la media y varianza aproximada hasta segundo orden del estimador de λ de la PDF exponencial discutido en la sección 9.3. ¿Cómo se compara el resultado con las aproximaciones de primer orden? Asegurarse de justificar la PDF aproximadamente gaussiana de $\hat{\lambda}$, requerida para aplicar la expresión de la varianza. Además, comparar los resultados con los predichos por la teoría del MLE asintótico (recordar que el estimador por el método de los momentos es el MLE).

Solución Las expresiones de la media y la varianza aproximadas hasta segundo orden están dadas por las ecuaciones 9.17 y 9.18. Como se calculó en la sección 9.2.1, en el caso de las muestras IID con PDF gaussiana,

$$\hat{\lambda} = g(T_1), \quad \text{donde} \quad T_1 = \frac{1}{N} \sum_{n=0}^{N-1} x[n] \quad \text{y} \quad g(T_1) = \frac{1}{T_1}.$$

Además, en la sección 9.3 se obtuvo que la media y la varianza del estadístico es

$$\mu_1 = E(T_1) = \frac{1}{\lambda} \quad \text{y} \quad \text{var}(T_1) = \frac{1}{N\lambda^2}.$$

De la ecuación 9.17, la aproximación de la media es

$$\begin{aligned} E(\hat{\lambda}) &= g(\mu_1) + \frac{1}{2} \text{tr}[\mathbf{G}(\mu_1) \mathbf{C}_T] \\ &= g(\mu_1) + \frac{1}{2} G(\mu_1) \text{var}(T_1), \end{aligned}$$

donde

$$g(\mu_1) = \frac{1}{\frac{1}{\lambda}} = \lambda$$

y

$$G(\mu_1) = \left. \frac{\partial^2 g}{\partial T_1^2} \right|_{T_1=\mu_1} = \left. \frac{\partial}{\partial T_1} \left(-\frac{1}{T_1^2} \right) \right|_{T_1=\mu_1} = \left. \frac{2}{T_1^3} \right|_{T_1=\mu_1} = \frac{2}{\mu_1^3} = 2\lambda^3.$$

Combinando los resultados,

$$E(\hat{\lambda}) = \lambda + \frac{1}{2} (2\lambda^3) \left(\frac{1}{N\lambda^2} \right) = \lambda + \frac{\lambda}{N},$$

resultando en

$$E(\hat{\lambda}) = \lambda \left(1 + \frac{1}{N} \right).$$

De la ecuación 9.18, la varianza aproximada es

$$\begin{aligned} \text{var}(\hat{\theta}) &= \left. \frac{\partial g}{\partial \mathbf{T}} \right|_{\mathbf{T}=\boldsymbol{\mu}}^T \mathbf{C}_T \left. \frac{\partial g}{\partial \mathbf{T}} \right|_{\mathbf{T}=\boldsymbol{\mu}} + \frac{1}{2} \text{tr}[(\mathbf{G}(\boldsymbol{\mu}) \mathbf{C}_T)^2] \\ &= \left(\left. \frac{\partial g}{\partial T_1} \right|_{T_1=\mu_1} \right)^2 \text{var}(T_1) + \frac{1}{2} (G(\mu_1) \text{var}(T_1))^2 \\ &= (-\lambda^2)^2 \frac{1}{N\lambda^2} + \frac{1}{2} \left(2\lambda^3 \frac{1}{N\lambda^2} \right)^2 \\ &= \frac{\lambda^2}{N} + \frac{2\lambda^2}{N^2}, \end{aligned}$$

resultando en

$$\text{var}(\hat{\theta}) = \frac{\lambda^2}{N} \left(1 + \frac{2}{N} \right).$$

Notar que la ecuación de la varianza es válida si $T_1 = \bar{x} \sim \mathcal{N}(\mu_1, \text{var}(T_1))$, lo cual puede asumirse por el teorema central del límite. También se observa que a diferencia de la aproximación de primer orden, el estimador no es insesgado. Como se observó en la sección 9.3, el estimador coincide con el MLE y la PDF asintótica del MLE es $\hat{\lambda} \stackrel{a}{\sim} \mathcal{N}(\lambda, \lambda^2/N)$. La aproximación de primer orden coincide con el MLE asintótico, pero en la aproximación de segundo orden, la PDF del MLE asintótico se cumple si $2/N \ll 1$.

9.5.11. Problema 11

Probar que para una señal sinusoidal

$$s[n] = A \cos(2\pi f_0 n + \phi)$$

donde la fase ϕ es determinística pero desconocida se cumple que

$$\frac{1}{N-1} \sum_{n=0}^{N-2} s[n]s[n+1] \rightarrow \frac{A^2}{2} \cos 2\pi f_0$$

cuando $N \rightarrow \infty$. Asumir que f_0 no es cercana a 0 o 1/2. Teniendo en cuenta este resultado, comentar sobre el uso del estimador del método de los momentos de la sección 9.4 en el caso en que la fase es determinística.

Solución Se ve que

$$\begin{aligned} \frac{1}{N-1} \sum_{n=0}^{N-2} s[n]s[n+1] &= \frac{1}{N-1} \sum_{n=0}^{N-2} A^2 \cos(2\pi f_0 n + \phi) \cos(2\pi f_0 (n+1) + \phi) \\ &\stackrel{(a)}{=} \frac{A^2}{2(N-1)} \sum_{n=0}^{N-2} [\cos(4\pi f_0 n + 2\pi f_0 + 2\phi) + \cos 2\pi f_0] \\ &\xrightarrow{N \rightarrow \infty} \frac{A^2}{2} \cos 2\pi f_0, \end{aligned}$$

donde en (a) se empleó la identidad trigonométrica $2 \cos \theta \cos \varphi = \cos(\theta - \varphi) + \cos(\theta + \varphi)$ y en el límite se tuvo en cuenta que la sumatoria del término del doble de la frecuencia tiende a 0 cuando $N \rightarrow \infty$ si f_0 no es cercana a 0 o 1/2 (ver el problema 3.6.7). El método de los momentos es válido, ya que en este caso, la media entre realizaciones $r_{ss}[1] = E(s[n]s[n-1])$ es igual a la media en el tiempo $\hat{r}_{ss}[1]$ en el caso en que la fase ϕ se considera una variable aleatoria. Un proceso aleatorio que cumple esta condición se llama *proceso ergódico*. Si la fase es determinística, el promedio temporal $\hat{r}_{ss}[1]$ iguala a $r_{ss}[1]$ cuando $N \rightarrow \infty$.

Notar además que se cumple el resultado mas general

$$\hat{r}_{ss}[k] = \frac{1}{N-|k|} \sum_{n=0}^{N-1-|k|} s[n]s[n+k] \xrightarrow{N \rightarrow \infty} r_{ss}[k] = \frac{A^2}{2} \cos 2\pi f_0 k,$$

que se deduce de forma analoga al caso con $k = 1$.

9.5.12. Problema 12

Para extender la aplicabilidad del estimador de la frecuencia del ejemplo de la sección 9.4, se asume que la amplitud es desconocida. ¿Porque no puede emplearse en este caso el estimador dado por la ecuación 9.14? Se considera el estimador

$$\hat{f}_0 = \frac{1}{2\pi} \arccos \left[\frac{\frac{1}{N-1} \sum_{n=0}^{N-2} x[n]x[n+1]}{\frac{1}{N} \sum_{n=0}^{N-1} x^2[n]} \right].$$

Mostrar que a alta SNR y para N grande, $E(\hat{f}_0) = f_0$. Justificar este estimador como un estimador por el método de los momentos basado en la ACF. ¿Se espera que este estimador sea bueno si la SNR es baja?

Solución El estimador dado por la ecuación 9.14 solo es válido si $A = \sqrt{2}$. En el caso en que la amplitud tenga un valor arbitrario A conocido, el estimador sería

$$\hat{f}_0 = \frac{1}{2\pi} \arccos \left[\frac{2}{A^2} \hat{r}_{xx}[1] \right] = \frac{1}{2\pi} \arccos \left[\frac{2}{A^2} \frac{1}{N-1} \sum_{n=0}^{N-2} x[n]x[n+1] \right],$$

el cual se obtiene de la ACF $r_{xx}[k]$ (ecuación 9.13) evaluada en $k = 1$.

Para mostrar el estimador propuesto

$$\hat{f}_0 = g(\mathbf{x}) = g(\mathbf{s}(f_0) + \mathbf{w}) = \frac{1}{2\pi} \arccos \left[\frac{\frac{1}{N-1} \sum_{n=0}^{N-2} x[n]x[n+1]}{\frac{1}{N} \sum_{n=0}^{N-1} x^2[n]} \right]$$

cumple que $E(\hat{f}_0) = f_0$, de la ecuación 9.11, que es válida cuando la SNR es alta y por lo tanto, $\mathbf{x} = \mathbf{s}(f_0) + \mathbf{w} \approx \mathbf{s}(f_0)$, se tiene que

$$\begin{aligned} E(\hat{f}_0) &= g(\mathbf{s}(f_0)) \\ &= \frac{1}{2\pi} \arccos \left[\frac{\frac{1}{N-1} \sum_{n=0}^{N-2} s[n]s[n+1]}{\frac{1}{N} \sum_{n=0}^{N-1} s^2[n]} \right] \\ &\xrightarrow{N \rightarrow \infty} \frac{1}{2\pi} \arccos \left[\frac{\frac{A^2}{2} \cos 2\pi f_0}{\frac{A^2}{2}} \right] \\ &= f_0, \end{aligned}$$

donde en el límite con $N \rightarrow \infty$ se emplearon los resultados del problema de la sección 9.5.11.

El estimador es un estimador por el método de los momentos basado en la ACF. Efectivamente, de la ecuación 9.13 se tiene que

$$r_{xx}[0] = \frac{A^2}{2} + \sigma^2 \approx \frac{A^2}{2}$$

donde en la aproximación se consideró que la SNR es grande y por lo tanto, $A^2/2 \gg \sigma^2$, y

$$r_{xx}[1] = \frac{A^2}{2} \cos 2\pi f_0.$$

En consecuencia,

$$\frac{r_{xx}[1]}{r_{xx}[0]} = \cos 2\pi f_0,$$

resultando en

$$f_0 = \frac{1}{2\pi} \arccos \left[\frac{r_{xx}[1]}{r_{xx}[0]} \right].$$

Empleando los estimadores naturales de $r_{xx}[0]$ y $r_{xx}[1]$, se obtiene que el estimador por el método de los momentos es

$$\hat{f}_0 = \frac{1}{2\pi} \arccos \left[\frac{\hat{r}_{xx}[1]}{\hat{r}_{xx}[0]} \right] = \frac{1}{2\pi} \arccos \left[\frac{\frac{1}{N-1} \sum_{n=0}^{N-2} x[n]x[n+1]}{\frac{1}{N} \sum_{n=0}^{N-1} x^2[n]} \right].$$

En el caso en que la SNR no sea alta, se cumple que

$$\hat{f}_0 = \frac{1}{2\pi} \arccos \left[\frac{\hat{r}_{xx}[1]}{\hat{r}_{xx}[0]} \right] \xrightarrow{N \rightarrow \infty} \frac{1}{2\pi} \arccos \left[\frac{r_{xx}[1]}{r_{xx}[0]} \right] = \frac{1}{2\pi} \arccos \left[\frac{\frac{A^2}{2} \cos 2\pi f_0}{\frac{A^2}{2} + \sigma^2} \right] \neq f_0,$$

concluyendo que el estimador no es válido.

9.5.13. Problema 13

Para el ejemplo de la estimación de la frecuencia de la sección 9.4, asúmase que los parámetros son $A = \sqrt{2}$, $f_0 = 0.25$, $\phi = 0$ y N impar. Mostrar que de la ecuación 9.16 se obtiene que $\text{var}(\hat{f}_0) = 0$. Para entender este resultado, notar que

$$\left. \frac{\partial h}{\partial w[i]} \right|_{\mathbf{w}=\mathbf{0}} = 0$$

para $i = 0, \dots, N-1$. ¿Qué dice esto sobre la aproximación de la expansión de Taylor de primer orden? ¿Cómo se puede mejorar la situación?

Solución Partiendo de la ecuación 9.16 de la varianza considerando que $s[n] = A \cos 2\pi f_0 n$ y evaluando en $A = \sqrt{2}$, $f_0 = 0.25$ y $\phi = 0$ se tiene que

$$\begin{aligned} \text{var}(\hat{f}_0) &= \frac{\sigma^2}{2\pi(N-1) \sin 2\pi f_0} \left[s^2[1] + 4 \cos^2 2\pi f_0 \sum_{n=1}^{N-2} s^2[n] + s^2[N-2] \right] \\ &= \frac{\sigma^2}{2\pi(N-1) \sin \frac{\pi}{2}} \left[2 \cos^2 \frac{\pi}{2} + 4 \cos^2 \frac{\pi}{2} \sum_{n=1}^{N-2} 2 \cos^2 \frac{\pi}{2} n + 2 \cos^2 \frac{\pi}{2} (N-2) \right] \\ &= 0, \end{aligned}$$

ya que N es impar y $\cos \pi n/2 = 0$ si n es impar. Este resultado proviene de que en este caso

$$\left. \frac{\partial h}{\partial w[i]} \right|_{\mathbf{w}=\mathbf{0}} = 0.$$

Efectivamente, como se dedujo en la sección 9.4,

$$\left. \frac{\partial h}{\partial w[i]} \right|_{\mathbf{w}=\mathbf{0}} = \begin{cases} -\frac{1}{2\pi(N-1) \sin 2\pi f_0} s[1], & i = 0 \\ -\frac{2 \cos 2\pi f_0}{2\pi(N-1) \sin 2\pi f_0} s[i], & i = 1, \dots, N-2 \\ -\frac{1}{2\pi(N-1) \sin 2\pi f_0} s[N-2], & i = N-1, \end{cases}$$

y es fácil ver que esta expresión evaluada en los valores de los parámetros indicados es nula para todo i . Por lo tanto, la aproximación del estimador mediante una expansión de Taylor de primer orden de la ecuación 9.10 es

$$\hat{f}_0 \approx h(\mathbf{0}) + \left. \frac{\partial h}{\partial \mathbf{w}} \right|_{\mathbf{w}=\mathbf{0}}^T \mathbf{w} = h(\mathbf{0}),$$

lo que indica que se está aproximando al estimador mediante un constante independiente de los datos observados. Esto conduce a un resultado teórico inválido de la varianza dada por la ecuación 9.12. Para obtener un resultado válido, se podría emplear una aproximación de Taylor de segundo orden como en el problema de la sección 9.5.9.

Capítulo 10

La filosofía Bayesiana

10.1. Conocimiento a priori y estimación

Un regla fundamental en la teoría de la estimación es que el conocimiento a priori conduce a estimadores mas exactos. Por ejemplo, si se sabe que un parámetro está restringido a un intervalo conocido, cualquier buen estimador debería producir estimadores solo en ese intervalo. El estimador MVU del nivel A de DC en WGN es la media muestral \bar{x} . Sin embargo, como $\bar{x} \sim \mathcal{N}(A, \sigma^2/N)$, esto supone que A puede tomar cualquier valor en $-\infty < A < \infty$. Debido a limitaciones físicas del problema, podría saberse que A solo puede tomar valores en el intervalo $-A_0 \leq A \leq A_0$, y por lo tanto, la media muestral no es un buen estimador debido a que conduce a valores fuera de ese intervalo. Un mejor estimador sería la media muestral truncada,

$$\check{A} = \begin{cases} -A_0 & \bar{x} < -A_0 \\ \bar{x} & -A_0 \leq \bar{x} \leq A_0 \\ A_0 & \bar{x} > A_0, \end{cases}$$

que es consistente con las restricciones del parámetro. Dicho estimador tiene PDF

$$p_{\check{A}}(\xi; A) = \Pr\{\bar{x} \leq -A_0\}\delta(\xi + A_0) + p_{\hat{A}}(\xi; A)[u(\xi + A_0) - u(\xi - A_0)] + \Pr\{\bar{x} \geq A_0\}\delta(\xi - A_0),$$

donde $u(x)$ es la función escalón. Puede mostrarse que \check{A} es un estimador sesgado (ver el problema de la sección 10.5.7). Sin embargo, si se compara el error cuadrático medio (*mean square error*, MSE) de los dos estimadores, se observa que para cualquier A en el intervalo $-A_0 \leq A \leq A_0$

$$\begin{aligned} \text{mse}(\hat{A}) &= E[(\hat{A} - A)^2] \\ &= \int_{-\infty}^{\infty} (\xi - A)^2 p_{\hat{A}}(\xi; A) d\xi \\ &= \int_{-\infty}^{-A_0} (\xi - A)^2 p_{\hat{A}}(\xi; A) d\xi + \int_{-A_0}^{A_0} (\xi - A)^2 p_{\hat{A}}(\xi; A) d\xi + \int_{A_0}^{\infty} (\xi - A)^2 p_{\hat{A}}(\xi; A) d\xi \\ &> \int_{-\infty}^{-A_0} (-A_0 - A)^2 p_{\hat{A}}(\xi; A) d\xi + \int_{-A_0}^{A_0} (\xi - A)^2 p_{\hat{A}}(\xi; A) d\xi + \int_{A_0}^{\infty} (A_0 - A)^2 p_{\hat{A}}(\xi; A) d\xi \\ &= (-A_0 - A)^2 \Pr\{\bar{x} \leq -A_0\} + \int_{-A_0}^{A_0} (\xi - A)^2 p_{\hat{A}}(\xi; A) d\xi + (A_0 - A)^2 \Pr\{\bar{x} \geq A_0\} \\ &= E[(\check{A} - A)^2] \\ &= \text{mse}(\check{A}), \end{aligned}$$

donde en la desigualdad se consideró que $(\xi - A)^2 > (-A_0 - A)^2$ para todo $\xi < -A_0$ y $(\xi - A)^2 > (A_0 - A)^2$ para todo $\xi > A_0$. Se concluye que la media muestral truncada es mejor estimador en términos del MSE. Si bien \hat{A} es el MVU, se logró reducir el MSE al permitir que el estimador sea sesgado. En la estimación clásica, la minimización del MSE conduce a estimadores irrealizables que dependen del valor del parámetro a estimar. Al haber encontrado un mejor estimador, surge la pregunta de si existe un estimador óptimo para este problema. La respuesta es afirmativa, pero únicamente luego de reformular

el modelo de datos. Sabiendo que A debe pertenecer a cierto intervalo conocido, se supondrá que A es elegido de ese intervalo. El proceso de la elección de A se modela como un evento aleatorio con cierta PDF. Conociendo únicamente el intervalo y al no tener inclinación por un valor específico de A , tiene sentido asignar una PDF $\mathcal{U}[-A_0, A_0]$ a la variable aleatoria A . El hecho de considerar al parámetro A como una variable aleatoria es la desviación del enfoque clásico hacia el enfoque bayesiano. Como siempre, el problema es la estimación del valor de A , o de la realización de la variable aleatoria. Sin embargo, ahora es posible incorporar conocimiento de como A es escogido. Por ejemplo, se podría intentar encontrar el estimador \hat{A} que minimice el *MSE bayesiano*, definido como

$$\text{Bmse}(\hat{A}) = E[(A - \hat{A})^2] = \iint (A - \hat{A})^2 p(\mathbf{x}, A) d\mathbf{x} dA. \quad (10.1)$$

Notar que como A es una variable aleatoria, la esperanza se toma respecto a la PDF conjunta $p(\mathbf{x}, A)$. Esta definición del MSE es distinta al caso clásico, y se distingue empleando la notación Bmse . Notar que a diferencia del caso clásico, el MSE bayesiano no depende del parámetro A debido a la integración respecto a A .

Continuando con el ejemplo, se deducirá el estimador que minimiza el MSE bayesiano. Para hacerlo, se parte del teorema de Bayes para escribir

$$p(\mathbf{x}, A) = p(A|\mathbf{x})p(\mathbf{x})$$

de forma que

$$\text{Bmse}(\hat{A}) = \int \left[\int (A - \hat{A})^2 p(A|\mathbf{x}) dA \right] p(\mathbf{x}) d\mathbf{x}.$$

Como $p(\mathbf{x}) > 0$ para todo \mathbf{x} , si la integral entre paréntesis puede ser minimizada para cada \mathbf{x} , el MSE bayesiano es minimizado. Por lo tanto, fijando \mathbf{x} de forma de que \hat{A} es una variable escalar en lugar de una función de \mathbf{x} , se tiene que

$$\begin{aligned} \frac{\partial}{\partial \hat{A}} \int (A - \hat{A})^2 p(A|\mathbf{x}) dA &= \int \frac{\partial}{\partial \hat{A}} (A - \hat{A})^2 p(A|\mathbf{x}) dA \\ &= \int -2(A - \hat{A}) p(A|\mathbf{x}) dA \\ &= -2 \int A p(A|\mathbf{x}) dA + 2\hat{A} \int p(A|\mathbf{x}) dA \\ &= -2 \int A p(A|\mathbf{x}) dA + 2\hat{A}, \end{aligned}$$

en donde en la última igualdad se consideró que la segunda integral es 1. Igualando la derivada a cero y despejando resulta en

$$\hat{A} = \int A p(A|\mathbf{x}) dA$$

o finalmente

$$\hat{A} = E(A|\mathbf{x}).$$

Se obtuvo que el estimador óptimo que minimiza el MSE bayesiano es la media de la PDF *a posteriori* $p(A|\mathbf{x})$. La PDF *a posteriori* es la PDF de A luego de observar a los datos. En contraposición, $p(A)$ o

$$p(A) = \int p(\mathbf{x}, A) d\mathbf{x}$$

puede considerarse como la PDF *a priori*, indicando la PDF antes de observar los datos. El estimador que minimiza el MSE bayesiano se llama estimador de error cuadrático medio mínimo (*minimum mean square error*, MMSE). El efecto de observar los datos es concentrar la PDF de A debido a que los datos reducen la incertidumbre de A (ver el problema de la sección 10.5.15). En el problema de la sección 10.5.5 se deduce el estimador MMSE de forma alternativa.

Para determinar el estimador MMSE se requiere la PDF a posteriori, la cual puede obtenerse mediante la regla de Bayes como

$$\begin{aligned} p(A|\mathbf{x}) &= \frac{p(\mathbf{x}|A)p(A)}{p(\mathbf{x})} \\ &= \frac{p(\mathbf{x}|A)p(A)}{\int p(\mathbf{x}|A)p(A) dA}. \end{aligned}$$

Observar que el denominador es solo un factor de normalización independiente de A , necesario para asegurar que $p(A|\mathbf{x})$ integre a 1. Continuando con el ejemplo, la PDF a priori $p(A)$ es $\mathcal{U}[-A_0, A_0]$. Para determinar la PDF condicional $p(\mathbf{x}|A)$ se necesita asumir que la elección de A a partir de $p(A)$ no afecta a la PDF de las muestras de ruido $w[n]$, es decir, las muestras de ruido $w[n]$ son independientes de A . Por lo tanto, para $n = 0, \dots, N-1$

$$\begin{aligned} p_x(x[n]|A) &\stackrel{(a)}{=} p_w(x[n] - A|A) \\ &\stackrel{(b)}{=} p_w(x[n] - A) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{1}{2\sigma^2} (x[n] - A)^2 \right], \end{aligned}$$

donde en (a) se tuvo en cuenta que $x[n] = w[n] + A$ y por lo tanto, $p_x(\xi|A) = p_w(\xi - A|A)$ y en (b) se consideró que $w[n]$ y A son independientes. Finalmente,

$$p(\mathbf{x}|A) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - A)^2 \right]. \quad (10.2)$$

Si bien la PDF tiene una forma idéntica a la PDF $p(\mathbf{x}; A)$ del caso clásico usual, en este caso se trata de una PDF condicional (ver el problema de la sección 10.5.6). Combinando los resultados se obtiene que la PDF a posteriori es

$$p(A|\mathbf{x}) = \begin{cases} \frac{\frac{1}{2A_0(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - A)^2 \right]}{\int_{-A_0}^{A_0} \frac{1}{2A_0(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - A)^2 \right] dA} & |A| \leq A_0 \\ 0 & |A| > A_0, \end{cases}$$

donde se tuvo en cuenta que $p(A) = 1/(2A_0)$ en $|A| \leq A_0$ y 0 en otro caso. Como

$$\begin{aligned} \sum_{n=0}^{N-1} (x[n] - A)^2 &= \sum_{n=0}^{N-1} (x^2[n] - 2Ax[n] + A^2) \\ &= \sum_{n=0}^{N-1} x^2[n] - 2NA\bar{x} + NA^2 \\ &\stackrel{(a)}{=} NA^2 - 2NA\bar{x} + N\bar{x}^2 + \sum_{n=0}^{N-1} x^2[n] - N\bar{x}^2 \\ &= N(A - \bar{x})^2 + \sum_{n=0}^{N-1} x^2[n] - N\bar{x}^2, \end{aligned} \quad (10.3)$$

donde en (a) se sumó y restó $N\bar{x}^2$. El segundo y tercer sumando son independientes de A por lo que pueden factorizarse afuera de la integral y el factor se cancela con el numerador, es decir, en $|A| \leq A_0$ se

tiene que

$$\begin{aligned}
 p(A|\mathbf{x}) &= \frac{\frac{1}{2A_0(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2} \left(\sum_{n=0}^{N-1} x^2[n] - N\bar{x}^2\right)\right] \exp\left[-\frac{1}{2\sigma^2} N(A - \bar{x})^2\right]}{\frac{1}{2A_0(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2} \left(\sum_{n=0}^{N-1} x^2[n] - N\bar{x}^2\right)\right] \int_{-A_0}^{A_0} \exp\left[-\frac{1}{2\sigma^2} N(A - \bar{x})^2\right] dA} \\
 &= \frac{\exp\left[-\frac{1}{2\frac{\sigma^2}{N}} (A - \bar{x})^2\right]}{\int_{-A_0}^{A_0} \exp\left[-\frac{1}{2\frac{\sigma^2}{N}} (A - \bar{x})^2\right] dA}.
 \end{aligned}$$

Observando que la PDF tiene la forma de una gaussiana de media \bar{x} y varianza σ^2/N truncada, se puede escribir como

$$p(A|\mathbf{x}) = \begin{cases} \frac{1}{c\sqrt{2\pi\frac{\sigma^2}{N}}} \exp\left[-\frac{1}{2\frac{\sigma^2}{N}} (A - \bar{x})^2\right] & |A| \leq A_0 \\ 0 & |A| > A_0, \end{cases}$$

donde el factor c se determina con el requerimiento de que $p(A|\mathbf{x})$ integre a 1, resultando en

$$c = \int_{-A_0}^{A_0} \frac{1}{\sqrt{2\pi\frac{\sigma^2}{N}}} \exp\left[-\frac{1}{2\frac{\sigma^2}{N}} (A - \bar{x})^2\right] dA.$$

El estimador MMSE, que es la media de $p(A|\mathbf{x})$, es

$$\begin{aligned}
 \hat{A} &= E(A|\mathbf{x}) \\
 &= \int_{-\infty}^{\infty} A p(A|\mathbf{x}) dA \\
 &= \frac{\int_{-A_0}^{A_0} A \frac{1}{\sqrt{2\pi\frac{\sigma^2}{N}}} \exp\left[-\frac{1}{2\frac{\sigma^2}{N}} (A - \bar{x})^2\right] dA}{\int_{-A_0}^{A_0} \frac{1}{\sqrt{2\pi\frac{\sigma^2}{N}}} \exp\left[-\frac{1}{2\frac{\sigma^2}{N}} (A - \bar{x})^2\right] dA}. \tag{10.4}
 \end{aligned}$$

Si bien la expresión no puede ser evaluada de forma analítica, se observa que \hat{A} es una función de \bar{x} así como de A_0 y σ^2 (ver el problema de la sección 10.5.7). Debido al truncamiento, el estimador no es \bar{x} , si no que estará sesgado hacia cero. Esto se debe al conocimiento a priori incorporado sobre A mediante $p(A)$. Como se muestra en el problema de la sección 10.5.8, en ausencia de datos, el estimador MMSE es

$$\hat{A} = E(A) = 0.$$

El efecto de observar los datos es posicionar la media a posteriori entre $A = 0$ y $A = \bar{x}$ en un compromiso entre el conocimiento a priori y el contribuido por los datos.

Los resultados de este ejemplo son generales y se resumen a continuación. En el enfoque bayesiano se asume que el parámetro a estimar es una realización de una variable aleatoria θ . De esta forma, se le asigna una PDF a priori $p(\theta)$. Luego de que observar los datos, el conocimiento sobre el parámetro se resume en la PDF a posteriori $p(\theta|\mathbf{x})$. El estimador óptimo se define como aquel que minimiza el error cuadrático medio entre todas las realizaciones de θ y \mathbf{x} , o el error cuadrático medio bayesiano. Este estimador es la media de la PDF a posteriori,

$$\hat{\theta} = E(\theta|\mathbf{x}) = \int \theta p(\theta|\mathbf{x}) d\theta, \tag{10.5}$$

y la PDF a posteriori se puede calcular como

$$p(\theta|\mathbf{x}) = \frac{p(\mathbf{x}|\theta)p(\theta)}{\int p(\mathbf{x}|\theta)p(\theta) d\theta}. \quad (10.6)$$

10.2. Elección de la PDF a priori

Como se vio en la sección anterior, una vez elegida la PDF a priori, el estimador MMSE se obtiene mediante la ecuación 10.5. No hay dudas sobre su existencia como la hay con el estimador MVU. El único problema práctico es si es posible o no obtener una expresión analítica para $E(\theta|\mathbf{x})$. En el ejemplo anterior, donde la PDF a priori es uniforme, no se encontró una expresión analítica para la PDF a posteriori. En estos casos es necesario evaluar la ecuación 10.5 mediante una integración numérica, lo cual puede ser inviable si el parámetro es un vector con dimensiones grandes. El siguiente ejemplo ilustra un caso importante en donde puede obtenerse una expresión analítica del estimador MMSE.

10.2.1. Ejemplo: nivel de DC en WGN con PDF a priori gaussiana

Se modifica ahora el conocimiento a priori del ejemplo de la estimación del nivel de DC en WGN de la sección anterior, que condujo a un estimador que no se puede expresar de forma analítica. Se asume la PDF a priori gaussiana

$$p(A) = \frac{1}{\sqrt{2\pi\sigma_A^2}} \exp \left[-\frac{1}{2\sigma_A^2} (A - \mu_A)^2 \right].$$

Elijiendo $\mu_A = 0$ y $3\sigma_A = A_0$ puede considerarse como que se está incorporando el conocimiento a priori de que $|A| \leq A_0$ de forma similar a la PDF a priori uniforme, pero con la PDF gaussiana se le da preferencia a los valores de A cercanos a cero. Para calcular la PDF a posteriori, se parte calculando la PDF condicional $p(\mathbf{x}|A)$, que igual que antes es (ver las ecuaciones 10.2 y 10.3)

$$\begin{aligned} p(\mathbf{x}|A) &= \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - A)^2 \right] \\ &= \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} x^2[n] \right] \exp \left[-\frac{1}{2\sigma^2} (NA^2 - 2NA\bar{x}) \right], \end{aligned}$$

y la PDF a posteriori es

$$\begin{aligned} p(A|\mathbf{x}) &= \frac{p(\mathbf{x}|A)p(A)}{\int p(\mathbf{x}|A)p(A) dA} \\ &= \frac{\frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \frac{1}{\sqrt{2\pi\sigma_A^2}} \exp \left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} x^2[n] \right] \exp \left[-\frac{1}{2\sigma^2} (NA^2 - 2NA\bar{x}) \right] \exp \left[-\frac{1}{2\sigma_A^2} (A - \mu_A)^2 \right]}{\int_{-\infty}^{\infty} \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \frac{1}{\sqrt{2\pi\sigma_A^2}} \exp \left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} x^2[n] \right] \exp \left[-\frac{1}{2\sigma^2} (NA^2 - 2NA\bar{x}) \right] \exp \left[-\frac{1}{2\sigma_A^2} (A - \mu_A)^2 \right] dA} \\ &= \frac{\exp \left[-\frac{1}{2} \left(\frac{1}{\sigma^2} (NA^2 - 2NA\bar{x}) + \frac{1}{\sigma_A^2} (A - \mu_A)^2 \right) \right]}{\int_{-\infty}^{\infty} \exp \left[-\frac{1}{2} \left(\frac{1}{\sigma^2} (NA^2 - 2NA\bar{x}) + \frac{1}{\sigma_A^2} (A - \mu_A)^2 \right) \right] dA} \\ &= \frac{\exp \left[-\frac{1}{2} Q(A) \right]}{\int_{-\infty}^{\infty} \exp \left[-\frac{1}{2} Q(A) \right] dA}. \end{aligned}$$

El denominador no depende de A debido a la integración, siendo únicamente un factor de normalización, y además el argumento de la exponencial es cuadrático en A . Por lo tanto, $p(A|\mathbf{x})$ debe ser gaussiana,

cuya media y varianza dependen de los datos \mathbf{x} . Continuando, se tiene que $Q(A)$ es

$$\begin{aligned} Q(A) &= \frac{N}{\sigma^2} A^2 - \frac{2NA\bar{x}}{\sigma^2} + \frac{A^2}{\sigma_A^2} - \frac{2\mu_A A}{\sigma_A^2} + \frac{\mu_A^2}{\sigma_A^2} \\ &= \left(\frac{N}{\sigma^2} + \frac{1}{\sigma_A^2} \right) A^2 - 2 \left(\frac{N}{\sigma^2} \bar{x} + \frac{\mu_A}{\sigma_A^2} \right) A + \frac{\mu_A^2}{\sigma_A^2}, \end{aligned}$$

y definiendo

$$\sigma_{A|x}^2 = \frac{1}{\frac{N}{\sigma^2} + \frac{1}{\sigma_A^2}} \quad \text{y} \quad \mu_{A|x} = \left(\frac{N}{\sigma^2} \bar{x} + \frac{\mu_A}{\sigma_A^2} \right) \sigma_{A|x}^2$$

se puede escribir como

$$\begin{aligned} Q(A) &= \frac{A^2}{\sigma_{A|x}^2} - \frac{2\mu_{A|x} A}{\sigma_{A|x}^2} + \frac{\mu_{A|x}^2}{\sigma_{A|x}^2} \\ &= \frac{1}{\sigma_{A|x}^2} \left(A^2 - 2\mu_{A|x} A + \mu_{A|x}^2 \right) - \frac{\mu_{A|x}^2}{\sigma_{A|x}^2} + \frac{\mu_{A|x}^2}{\sigma_{A|x}^2} \\ &= \frac{1}{\sigma_{A|x}^2} (A - \mu_{A|x})^2 - \frac{\mu_{A|x}^2}{\sigma_{A|x}^2} + \frac{\mu_{A|x}^2}{\sigma_{A|x}^2}. \end{aligned}$$

Sustituyendo este resultado en la PDF a posteriori, se obtiene que

$$\begin{aligned} p(A|\mathbf{x}) &= \frac{\exp \left[-\frac{1}{2\sigma_{A|x}^2} (A - \mu_{A|x})^2 \right] \exp \left[-\frac{1}{2} \left(\frac{\mu_{A|x}^2}{\sigma_{A|x}^2} - \frac{\mu_A^2}{\sigma_A^2} \right) \right]}{\int_{-\infty}^{\infty} \exp \left[-\frac{1}{2\sigma_{A|x}^2} (A - \mu_{A|x})^2 \right] \exp \left[-\frac{1}{2} \left(\frac{\mu_{A|x}^2}{\sigma_{A|x}^2} - \frac{\mu_A^2}{\sigma_A^2} \right) \right] dA} \\ &= \frac{1}{\sqrt{2\pi\sigma_{A|x}^2}} \exp \left[-\frac{1}{2\sigma_{A|x}^2} (A - \mu_{A|x})^2 \right] \end{aligned}$$

donde en la última igualdad se consideró que $p(A|\mathbf{x})$ tiene que integrar a 1. Se observa que la PDF a posteriori es $\mathcal{N}(\mu_{A|x}, \sigma_{A|x}^2)$. De esto surge inmediatamente que el estimador MMSE es

$$\begin{aligned} \hat{A} &= E(A|\mathbf{x}) \\ &= \mu_{A|x} \\ &= \frac{\frac{N}{\sigma^2} \bar{x} + \frac{\mu_A}{\sigma_A^2}}{\frac{N}{\sigma^2} + \frac{1}{\sigma_A^2}} \\ &= \frac{\frac{N}{\sigma^2}}{\frac{N}{\sigma^2} + \frac{1}{\sigma_A^2}} \bar{x} + \frac{\frac{1}{\sigma_A^2}}{\frac{N}{\sigma^2} + \frac{1}{\sigma_A^2}} \mu_A, \end{aligned}$$

que puede escribirse como

$$\hat{A} = \frac{\sigma_A^2}{\sigma_A^2 + \frac{\sigma^2}{N}} \bar{x} + \frac{\frac{\sigma^2}{N}}{\sigma_A^2 + \frac{\sigma^2}{N}} \mu_A = \alpha \bar{x} + (1 - \alpha) \mu_A, \quad (10.7)$$

con

$$\alpha = \frac{\sigma_A^2}{\sigma_A^2 + \frac{\sigma^2}{N}},$$

que cumple que $0 < \alpha < 1$. α es un factor de peso entre el conocimiento a priori y los datos. Cuando se observan pocos datos, $\sigma_A^2 \ll \sigma^2/N$, α es pequeño y $\hat{A} \approx \mu_A$, y viceversa, cuando se observa una gran cantidad de datos $\sigma_A^2 \gg \sigma^2/N$, $\alpha \approx 1$ y $\hat{A} \approx \bar{x}$. Alternativamente, puede verse este proceso examinando la PDF cuando crece N . Cuando la cantidad de datos crece, la PDF a posteriori se concentra cada vez mas en torno a la media. Esto es porque la varianza a posteriori

$$\text{var}(A|\mathbf{x}) = \sigma_{A|x}^2 = \frac{1}{\frac{N}{\sigma^2} + \frac{1}{\sigma_A^2}}$$

decrece con N . Además, la media a posteriori también cambia con N . Para N pequeño será aproximadamente μ_A y se aproxima a \bar{x} a medida que crece N . De hecho, cuando $N \rightarrow \infty$, se cumple que $\hat{A} \rightarrow \bar{x}$, que a su vez tiende al valor verdadero A del parámetro. Se observa además que si no hay conocimiento a priori, hecho que puede modelarse dejando que $\sigma_A^2 \rightarrow \infty$, se cumple que $\hat{A} \rightarrow \bar{x}$ para cualquier largo de registro de datos, y se obtiene el estimador clásico. Finalmente, se mostrará como la incorporación de conocimiento a priori mejora la precisión de la estimación. Como

$$\begin{aligned} \text{Bmse}(\hat{A}) &= E[(A - \hat{A})^2] \\ &= \iint (A - \hat{A})^2 p(\mathbf{x}, A) d\mathbf{x} dA \\ &= \iint (A - \hat{A})^2 p(A|\mathbf{x}) p(\mathbf{x}) d\mathbf{x} dA, \end{aligned}$$

y $\hat{A} = E(A|\mathbf{x})$, se tiene que

$$\begin{aligned} \text{Bmse}(\hat{A}) &= \int \left[\int (A - E(A|\mathbf{x}))^2 p(A|\mathbf{x}) dA \right] p(\mathbf{x}) d\mathbf{x} \\ &= \int \text{var}(A|\mathbf{x}) p(\mathbf{x}) d\mathbf{x}. \end{aligned} \tag{10.8}$$

Esto indica que el MSE bayesiano es la varianza de la PDF a posteriori promediada de forma ponderada con la PDF de los datos \mathbf{x} . Por lo tanto,

$$\begin{aligned} \text{Bmse}(\hat{A}) &= \int \sigma_{A|x}^2 p(\mathbf{x}) d\mathbf{x} \\ &\stackrel{(a)}{=} \sigma_{A|x}^2 \\ &= \frac{1}{\frac{N}{\sigma^2} + \frac{1}{\sigma_A^2}} \end{aligned} \tag{10.9}$$

donde en (a) se consideró que en este caso $\sigma_{A|x}^2$ no depende de \mathbf{x} . Esto se puede escribir como

$$\text{Bmse}(\hat{A}) = \frac{\sigma^2}{N} \left(\frac{\sigma_A^2}{\sigma_A^2 + \frac{\sigma^2}{N}} \right),$$

y teniendo en cuenta que el factor entre paréntesis es menor que 1, se concluye que

$$\text{Bmse}(\hat{A}) < \frac{\sigma^2}{N},$$

donde σ^2/N es el MSE mínimo cuando no se incorpora conocimiento a priori, o cuando $\sigma_A^2 \rightarrow \infty$.

Las PDF a priori gaussianas son útiles en la práctica debido a su docilidad matemática. La principal propiedad que produce esto es la *propiedad de reproducción*. Si $p(A, \mathbf{x})$ es gaussiana, el marginal $p(A)$ también es gaussiano, así como la PDF a posteriori $p(A|\mathbf{x})$. Por lo tanto, la forma de la PDF se mantiene cuando se condiciona en \mathbf{x} . Otro ejemplo de una PDF con esta propiedad es dada en el problema de la sección 10.5.10.

10.3. Propiedades de la PDF gaussiana

En esta sección se analizan las propiedades de la PDF gaussiana bivariada y se describen las propiedades correspondientes en el caso multivariado. La principal propiedad que se analiza es que la PDF condicional también es gaussiana.

Sea $[x y]^T$ un vector conjuntamente gaussiano cuya PDF conjunta es

$$p(x, y) = \frac{1}{2\pi \det^{\frac{1}{2}}(\mathbf{C})} \exp \left[-\frac{1}{2} \begin{bmatrix} x - E(x) \\ y - E(y) \end{bmatrix}^T \mathbf{C}^{-1} \begin{bmatrix} x - E(x) \\ y - E(y) \end{bmatrix} \right].$$

Esta PDF se llama PDF gaussiana *bivariada*. El vector de media y la matriz de covarianza son respectivamente

$$E \left(\begin{bmatrix} x \\ y \end{bmatrix} \right) = \begin{bmatrix} E(x) \\ E(y) \end{bmatrix}$$

$$\mathbf{C} = \begin{bmatrix} \text{var}(x) & \text{cov}(x, y) \\ \text{cov}(y, x) & \text{var}(y) \end{bmatrix}.$$

La PDF gaussiana bivariada también puede escribirse como

$$p(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\frac{(x-\mu_x)^2}{\sigma_x^2} - 2\rho \frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \frac{(y-\mu_y)^2}{\sigma_y^2} \right] \right\} \quad (10.10)$$

con $\mu_x = E(x)$, $\mu_y = E(y)$, $\sigma_x^2 = \text{var}(x)$, $\sigma_y^2 = \text{var}(y)$ y $\rho = \text{cov}(x, y)/\sqrt{\text{var}(x)\text{var}(y)}$. Las PDF marginales $p(x)$ y $p(y)$ también son gaussianas, como puede verificarse realizando las integraciones (ver el capítulo 6 de [2])

$$p(x) = \int_{-\infty}^{\infty} p(x, y) dy = \frac{1}{\sqrt{2\pi \text{var}(x)}} \exp \left[-\frac{1}{2 \text{var}(x)} (x - E(x))^2 \right]$$

$$p(y) = \int_{-\infty}^{\infty} p(x, y) dx = \frac{1}{\sqrt{2\pi \text{var}(y)}} \exp \left[-\frac{1}{2 \text{var}(y)} (y - E(y))^2 \right].$$

Los contornos sobre los cuales la PDF $p(x, y)$ es constante son los valores de x e y para los cuales el exponente de la exponencial es constante. Estos se muestran en la figura 10.1 donde se observa que son elipses. Si $\text{var}(x) = \text{var}(y)$ y el coeficiente de correlación $\rho \neq 0$, el eje mayor de las elipses se encuentra en una recta de pendiente ± 1 . Si $\rho = 0$, los contornos de $p(x, y)$ constante son circunferencias, y con $\rho \neq 0$ el eje mayor de las elipses crece y el eje menor decrece con $|\rho| \rightarrow 1$.

Al observar $x = x_0$, la PDF condicional de y es

$$p(y|x_0) = \frac{p(x_0, y)}{p(x_0)} = \frac{p(x_0, y)}{\int_{-\infty}^{\infty} p(x_0, y) dy},$$

que indica que la PDF condicional de y es la sección $x = x_0$ de la superficie $p(x, y)$ (ver la figura 10.1) con la normalización apropiada para que integre a 1. Puede verse que al fijar $x = x_0$, $p(x_0, y)$ es cuadrática en y , por lo que la PDF condicional debe ser gaussiana. Como $p(y)$ también es gaussiana, esta propiedad puede formularse diciendo que si x e y son conjuntamente gaussianas, la PDF a priori $p(y)$ y la PDF a posteriori $p(y|x)$ son también gaussianas. En el siguiente teorema se especifica la forma exacta de la PDF condicional, que se deduce en el apéndice F.

Teorema (PDF condicional de gaussiana bivariada) Si x e y son conjuntamente gaussianas con vector de media $[E(x) E(y)]^T$ y matriz de covarianza

$$\mathbf{C} = \begin{bmatrix} \text{var}(x) & \text{cov}(x, y) \\ \text{cov}(y, x) & \text{var}(y) \end{bmatrix}.$$

de forma que

$$p(x, y) = \frac{1}{2\pi \det^{\frac{1}{2}}(\mathbf{C})} \exp \left[-\frac{1}{2} \begin{bmatrix} x - E(x) \\ y - E(y) \end{bmatrix}^T \mathbf{C}^{-1} \begin{bmatrix} x - E(x) \\ y - E(y) \end{bmatrix} \right].$$

La PDF condicional $p(y|x)$ también es gaussiana con

$$E(y|x) = E(y) + \frac{\text{cov}(x, y)}{\text{var}(x)}(x - E(x)) \quad (10.11)$$

$$\text{var}(y|x) = \text{var}(y) - \frac{\text{cov}^2(x, y)}{\text{var}(x)}. \quad (10.12)$$

Este resultado puede interpretarse de la siguiente forma. Antes de observar x , la variable aleatoria y está distribuida de acuerdo a la PDF a priori $p(y)$, o $y \sim \mathcal{N}(E(y), \text{var}(y))$. Luego de observar x , la variable aleatoria y está distribuida de acuerdo a la PDF a posteriori $p(y|x)$ dada por el teorema. Asumiendo que x e y no son independientes y por lo tanto, $\text{cov}(x, y) \neq 0$, la PDF a posterior se hace mas concentrada dado que hay menos incertidumbre sobre y . Para verificar esto, de la ecuación 10.12 se ve que

$$\begin{aligned} \text{var}(y|x) &= \text{var}(y) \left[1 - \frac{\text{cov}^2(x, y)}{\text{var}(x) \text{var}(y)} \right] \\ &= \text{var}(y)(1 - \rho^2) \end{aligned} \quad (10.13)$$

y por lo tanto $\text{var}(y|x) \leq \text{var}(y)$, ya que el coeficiente de correlación ρ cumple que $|\rho| < 1$. Como se vio previamente, el estimador MMSE de y al observar x es $E(y|x)$ por lo que, de la ecuación 10.11,

$$\hat{y} = E(y) + \frac{\text{cov}(x, y)}{\text{var}(x)}(x - E(x)).$$

En forma normalizada (una variable aleatoria de media nula y varianza unidad), esto se puede expresar como

$$\frac{\hat{y} - E(y)}{\sqrt{\text{var}(y)}} = \frac{\text{cov}(x, y)}{\sqrt{\text{var}(x) \text{var}(y)}} \frac{(x - E(x))}{\sqrt{\text{var}(x)}}.$$

o

$$\hat{y}_n = \rho x_n.$$

El coeficiente de correlación es un factor de escala sobre la variable aleatoria normalizada x_n para obtener el estimador MMSE de la variable aleatoria normalizada y_n . Si las variables aleatorias ya están normalizadas, los contornos de PDF constante son como en la figura 10.1. La ubicación de los picos de $p(x, y)$ cuando se considera como una función de y para cada x es la recta $y = \rho x$, y el estimador MMSE es $\hat{y} = E(y|x) = \rho x$ (ver el problema de la sección 10.5.12). El estimador MMSE emplea la correlación entre las variables aleatorias para estimar la realización de una basada en una realización de la otra. De las ecuaciones 10.8 y 10.13, el MSE mínimo es,

$$\begin{aligned} \text{Bmse}(\hat{y}) &= \int \text{var}(y|x) p(x) dx \\ &= \text{var}(y|x) \\ &= \text{var}(y)(1 - \rho^2) \end{aligned}$$

donde se tuvo en cuenta que $\text{var}(y|x)$ no depende de x , ya que $\text{var}(y)$ y ρ dependen únicamente de la matriz de covarianza. Esto indica que la calidad del estimador depende del coeficiente de correlación, el cual es una medida de la dependencia estadística entre x e y .

Para generalizar este resultado, se considera a continuación el caso multivariado. La demostración puede encontrarse en el apéndice F.

Teorema (PDF condicional de gaussiana multivariada) Si \mathbf{x} y \mathbf{y} son vectores conjuntamente gaussianos, donde \mathbf{x} es $k \times 1$ y \mathbf{y} es $l \times 1$, con vector de media $[E(\mathbf{x})^T \ E(\mathbf{y})^T]^T$ y matriz de covarianza particionada

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_{xx} & \mathbf{C}_{xy} \\ \mathbf{C}_{yx} & \mathbf{C}_{yy} \end{bmatrix}$$

tal que

$$p(\mathbf{x}, \mathbf{y}) = \frac{1}{(2\pi)^{\frac{k+l}{2}} \det^{\frac{1}{2}}(\mathbf{C})} \exp \left[-\frac{1}{2} \begin{bmatrix} \mathbf{x} - E(\mathbf{x}) \\ \mathbf{y} - E(\mathbf{y}) \end{bmatrix}^T \mathbf{C}^{-1} \begin{bmatrix} \mathbf{x} - E(\mathbf{x}) \\ \mathbf{y} - E(\mathbf{y}) \end{bmatrix} \right],$$

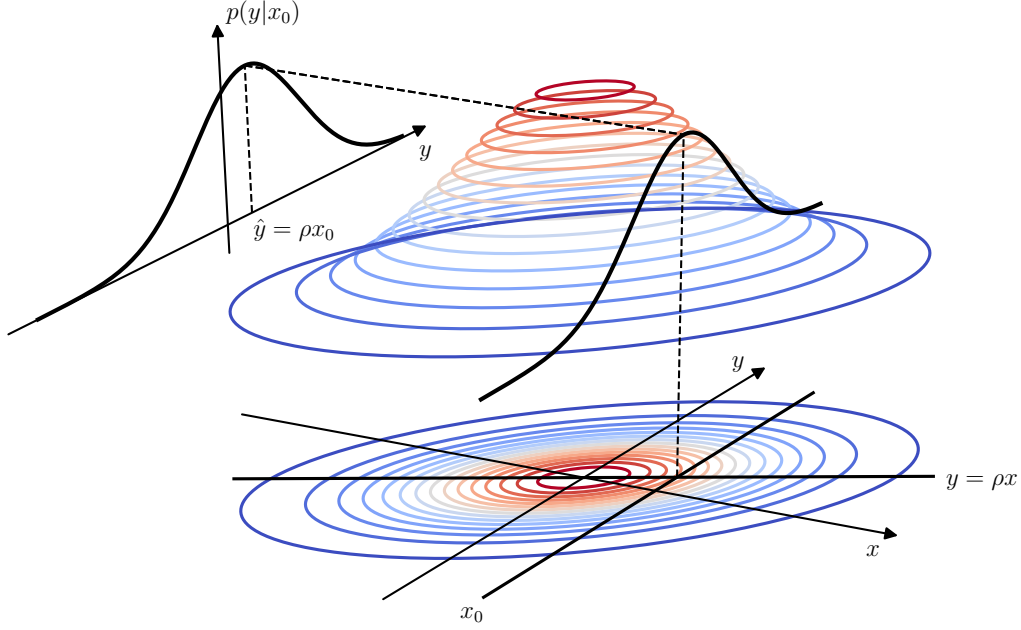


Figura 10.1: PDF conjunta gaussiana bivariada $p(x, y)$ y PDF condicional $p(y|x)$. Las curvas de nivel corresponden a los pares (x, y) donde la PDF conjunta $p(x, y)$ es constante. Al observar $x = x_0$, la PDF condicional $p(y|x_0)$ es la sección correspondiente a la recta $x = x_0$ de la superficie $p(x, y)$, o $p(x_0, y)$ normalizada para que integre a 1. La PDF condicional también es gaussiana. Para cada valor de x , el máximo de la PDF condicional, que a su vez es $E(y|x)$, se da en la recta $y = \rho x$.

la PDF condicional $p(\mathbf{y}|\mathbf{x})$ es también gaussiana con

$$E(\mathbf{y}|\mathbf{x}) = E(\mathbf{y}) + \mathbf{C}_{yx}\mathbf{C}_{xx}^{-1}(\mathbf{x} - E(\mathbf{x})) \quad (10.14)$$

$$\mathbf{C}_{y|x} = \mathbf{C}_{yy} - \mathbf{C}_{yx}\mathbf{C}_{xx}^{-1}\mathbf{C}_{xy}. \quad (10.15)$$

Notar que la matriz de covarianza de la PDF condicional no depende de \mathbf{x} , aunque esto no es cierto en general. Al igual que en el caso bivariado, la PDF a priori $p(\mathbf{y})$ así como la PDF a posteriori $p(\mathbf{y}|\mathbf{x})$ son gaussianas.

10.4. Modelo lineal bayesiano

De forma análoga al modelo lineal descrito en el capítulo 4, puede definirse el modelo lineal bayesiano. En particular, sea el modelo

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}, \quad (10.16)$$

donde \mathbf{x} es el vector $N \times 1$ de datos, \mathbf{H} es una matriz $N \times p$ conocida, $\boldsymbol{\theta}$ es un vector aleatorio $p \times 1$ con PDF a priori $\mathcal{N}(\boldsymbol{\mu}_\theta, \mathbf{C}_\theta)$ y \mathbf{w} es un vector $N \times 1$ de ruido con PDF $\mathcal{N}(\mathbf{0}, \mathbf{C}_w)$ independiente de $\boldsymbol{\theta}$. Este modelo de datos se denomina *modelo lineal bayesiano general*, y difiere del modelo lineal clásico en que $\boldsymbol{\theta}$ se modela como una variable aleatoria con PDF gaussiana. Se derivará el estimador MMSE, para lo cual es necesario calcular la PDF a posteriori $p(\boldsymbol{\theta}|\mathbf{x})$. Del teorema de la sección 10.3 para el caso multivariado, se sabe que si \mathbf{x} y $\boldsymbol{\theta}$ son conjuntamente gaussianos, la PDF a posteriori también es gaussiana. Por lo tanto, alcanza verificar que efectivamente es así, y aplicar el teorema para calcular la media y la varianza. Sea $\mathbf{z} = [\mathbf{x}^T \boldsymbol{\theta}^T]^T$, de forma de que de la ecuación 10.16 se cumple que

$$\mathbf{z} = \begin{bmatrix} \mathbf{H}\boldsymbol{\theta} + \mathbf{w} \\ \boldsymbol{\theta} \end{bmatrix} = \begin{bmatrix} \mathbf{H} & \mathbf{I} \\ \mathbf{I} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\theta} \\ \mathbf{w} \end{bmatrix},$$

donde \mathbf{H} es de dimensiones $N \times p$, la matriz identidad de arriba a la derecha es de dimensiones $N \times N$ y la de abajo a la izquierda de dimensiones $p \times p$, y la matriz $\mathbf{0}$ es de dimensiones $p \times N$. Como $\boldsymbol{\theta}$ y \mathbf{w} son independientes con PDF gaussiana, son conjuntamente gaussianos. Además, como \mathbf{z} es una transformación lineal de un vector gaussiano, también es gaussiano (ver el capítulo 6 de [2]). Se concluye puede aplicarse el teorema de la sección anterior con $\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$ y $\mathbf{y} = \boldsymbol{\theta}$. Los vectores de media son

$$\begin{aligned} E(\mathbf{x}) &= E(\mathbf{H}\boldsymbol{\theta} + \mathbf{w}) = \mathbf{H}E(\boldsymbol{\theta}) = \mathbf{H}\boldsymbol{\mu}_\theta \\ E(\mathbf{y}) &= E(\boldsymbol{\theta}) = \boldsymbol{\mu}_\theta. \end{aligned}$$

La matriz de covarianza de \mathbf{x} es

$$\begin{aligned} \mathbf{C}_{xx} &= E[(\mathbf{x} - E(\mathbf{x}))(\mathbf{x} - E(\mathbf{x}))^T] \\ &= E[(\mathbf{H}\boldsymbol{\theta} + \mathbf{w} - \mathbf{H}\boldsymbol{\mu}_\theta)(\mathbf{H}\boldsymbol{\theta} + \mathbf{w} - \mathbf{H}\boldsymbol{\mu}_\theta)^T] \\ &= E[(\mathbf{H}(\boldsymbol{\theta} - \boldsymbol{\mu}_\theta) + \mathbf{w})(\mathbf{H}(\boldsymbol{\theta} - \boldsymbol{\mu}_\theta) + \mathbf{w})^T] \\ &= E[(\mathbf{H}(\boldsymbol{\theta} - \boldsymbol{\mu}_\theta) + \mathbf{w})(\boldsymbol{\theta} - \boldsymbol{\mu}_\theta)^T \mathbf{H}^T + \mathbf{w}\mathbf{w}^T] \\ &\stackrel{(a)}{=} \mathbf{H}E[(\boldsymbol{\theta} - \boldsymbol{\mu}_\theta)(\boldsymbol{\theta} - \boldsymbol{\mu}_\theta)^T] \mathbf{H}^T + \mathbf{H}E(\boldsymbol{\theta} - \boldsymbol{\mu}_\theta)E(\mathbf{w}^T) + E(\mathbf{w})\mathbf{H}E[(\boldsymbol{\theta} - \boldsymbol{\mu}_\theta)^T] + E(\mathbf{w}\mathbf{w}^T) \\ &\stackrel{(b)}{=} \mathbf{H}\mathbf{C}_\theta \mathbf{H}^T + \mathbf{C}_w, \end{aligned}$$

donde en (a) se consideró que $\boldsymbol{\theta}$ y \mathbf{w} son independientes y en (b) que \mathbf{w} tiene media nula. La matriz de covarianza cruzada es

$$\begin{aligned} \mathbf{C}_{yx} &= E[(\mathbf{y} - E(\mathbf{y}))(\mathbf{x} - E(\mathbf{x}))^T] \\ &= E[(\boldsymbol{\theta} - \boldsymbol{\mu}_\theta)(\mathbf{H}(\boldsymbol{\theta} - \boldsymbol{\mu}_\theta) + \mathbf{w})^T] \\ &\stackrel{(a)}{=} E[(\boldsymbol{\theta} - \boldsymbol{\mu}_\theta)(\boldsymbol{\theta} - \boldsymbol{\mu}_\theta)^T \mathbf{H}^T] \\ &= \mathbf{C}_\theta \mathbf{H}^T, \end{aligned}$$

donde en (a) nuevamente se consideró que $\boldsymbol{\theta}$ y \mathbf{w} son independientes y que \mathbf{w} tiene media nula. Además, $\mathbf{C}_{yy} = \mathbf{C}_\theta$. Finalmente, aplicando la ecuación 10.14, se tiene que la media de la PDF a posteriori es

$$E(\boldsymbol{\theta}|\mathbf{x}) = \boldsymbol{\mu}_\theta + \mathbf{C}_\theta \mathbf{H}^T (\mathbf{H}\mathbf{C}_\theta \mathbf{H}^T + \mathbf{C}_w)^{-1} (\mathbf{x} - \mathbf{H}\boldsymbol{\mu}_\theta),$$

y a partir de la ecuación 10.15, la covarianza es

$$\mathbf{C}_{\theta|x} = \mathbf{C}_\theta - \mathbf{C}_\theta \mathbf{H}^T (\mathbf{H}\mathbf{C}_\theta \mathbf{H}^T + \mathbf{C}_w)^{-1} \mathbf{H}\mathbf{C}_\theta.$$

El siguiente teorema resume los resultados del modelo lineal bayesiano.

Teorema (PDF a posteriori del modelo lineal bayesiano general) Si los datos \mathbf{x} observados puede modelarse como

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$$

donde \mathbf{x} es un vector de datos $N \times 1$, \mathbf{H} es una matriz $N \times p$ conocida, $\boldsymbol{\theta}$ es un vector aleatorio $p \times 1$ con PDF a priori $\mathcal{N}(\boldsymbol{\mu}_\theta, \mathbf{C}_\theta)$, y \mathbf{w} es una vector $N \times 1$ de ruido con PDF $\mathcal{N}(\mathbf{0}, \mathbf{C}_w)$ e independiente de $\boldsymbol{\theta}$, la PDF a posteriori $p(\boldsymbol{\theta}|\mathbf{x})$ es gaussiana con media

$$E(\boldsymbol{\theta}|\mathbf{x}) = \boldsymbol{\mu}_\theta + \mathbf{C}_\theta \mathbf{H}^T (\mathbf{H}\mathbf{C}_\theta \mathbf{H}^T + \mathbf{C}_w)^{-1} (\mathbf{x} - \mathbf{H}\boldsymbol{\mu}_\theta) \quad (10.17)$$

y covarianza

$$\mathbf{C}_{\theta|x} = \mathbf{C}_\theta - \mathbf{C}_\theta \mathbf{H}^T (\mathbf{H}\mathbf{C}_\theta \mathbf{H}^T + \mathbf{C}_w)^{-1} \mathbf{H}\mathbf{C}_\theta. \quad (10.18)$$

Es útil tener en cuenta que la media y la covarianza de la PDF a posteriori pueden expresarse de forma alternativa como

$$E(\boldsymbol{\theta}|\mathbf{x}) = \boldsymbol{\mu}_\theta + (\mathbf{C}_\theta^{-1} + \mathbf{H}^T \mathbf{C}_w^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C}_w^{-1} (\mathbf{x} - \mathbf{H}\boldsymbol{\mu}_\theta) \quad (10.19)$$

y

$$\mathbf{C}_{\theta|x} = (\mathbf{C}_\theta^{-1} + \mathbf{H}^T \mathbf{C}_w^{-1} \mathbf{H})^{-1}, \quad (10.20)$$

como se muestra en el problema de la sección 10.5.13.

10.5. Problemas

10.5.1. Problema 1

En este problema se intenta aplicar el enfoque bayesiano en la estimación de un parámetro determinístico. Como el parámetro es determinístico, se le asigna la PDF a priori $p(\theta) = \delta(\theta - \theta_0)$, donde θ_0 es el valor verdadero del parámetro. Encontrar el estimador MMSE para esta PDF a priori y explicar los resultados.

Solución Partiendo de la definición del estimador MMSE (ecuación 10.5) se tiene que

$$\begin{aligned}\hat{\theta} &= E(\theta|\mathbf{x}) \\ &= \int \theta p(\theta|\mathbf{x}) d\theta \\ &\stackrel{(a)}{=} \frac{\int \theta p(\mathbf{x}|\theta)p(\theta) d\theta}{\int p(\mathbf{x}|\theta)p(\theta) d\theta} \\ &= \frac{\int \theta p(\mathbf{x}|\theta)\delta(\theta - \theta_0) d\theta}{\int p(\mathbf{x}|\theta)\delta(\theta - \theta_0) d\theta} \\ &= \frac{\theta_0 p(\mathbf{x}|\theta_0)}{p(\mathbf{x}|\theta_0)} \\ &= \theta_0,\end{aligned}$$

donde en (a) se empleó la ecuación 10.6 de la PDF a posteriori. El MMSE es el valor verdadero del parámetro ya que el conocimiento a priori es perfecto.

10.5.2. Problema 2

Dos variables aleatorias x e y se dicen condicionalmente independientes entre si si la PDF conjunta se factoriza como

$$p(x, y|z) = p(x|z)p(y|z)$$

donde z es la variable aleatoria condicionante. Se observa $x[n] = A + w[n]$ para $n = 0, 1$, donde A , $w[0]$ y $w[1]$ son variables aleatorias. Si A , $w[0]$ y $w[1]$ son todas independientes, probar que $x[0]$ y $x[1]$ son condicionalmente independientes. La variable aleatoria condicionante es A . ¿Son $x[0]$ y $x[1]$ independientes al no estar condicionadas, o es cierto que se cumple

$$p(x[0], x[1]) = p(x[0])p(x[1])?$$

Para responder esto considerar el caso en que A , $w[0]$ y $w[1]$ son independientes y tienen PDF $\mathcal{N}(0, 1)$.

Solución Hay que probar que

$$p(x[0], x[1]|A) = p(x[0]|A)p(x[1]|A).$$

Se cumple que

$$\begin{aligned}p_{x[0], x[1]}(x[0], x[1]|A) &= p_{w[0], w[1]}(x[0] - A, x[1] - A|A) \\ &\stackrel{(a)}{=} p_{w[0], w[1]}(x[0] - A, x[1] - A) \\ &\stackrel{(b)}{=} p_{w[0]}(x[0] - A)p_{w[1]}(x[1] - A) \\ &= p_{x[0]}(x[0]|A)p_{x[1]}(x[1]|A),\end{aligned}$$

donde en (a) se consideró que los $w[n]$ son independientes de A y en (b) que $w[0]$ y $w[1]$ son independientes entre si.

Para probar que no son independientes de forma incondicionada, se probará que no se cumple que

$$p(x[0], x[1]) = p(x[0])p(x[1])$$

en el caso en que A , $w[0]$ y $w[1]$ son independientes y tienen PDF $\mathcal{N}(0, 1)$. Se comenzará calculando la PDF conjunta

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{k}{2}} \det^{\frac{1}{2}}(\mathbf{C})} \exp \left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{C}^{-1}(\mathbf{x} - \boldsymbol{\mu}) \right].$$

En este caso, $k = 2$ con

$$\mathbf{x} = \begin{bmatrix} x[0] \\ x[1] \end{bmatrix} = \begin{bmatrix} A + w[0] \\ A + w[1] \end{bmatrix}.$$

La media es

$$\boldsymbol{\mu} = E(\mathbf{x}) = E \left(\begin{bmatrix} A + w[0] \\ A + w[1] \end{bmatrix} \right) = \begin{bmatrix} E(A) + E(w[0]) \\ E(A) + E(w[1]) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Para calcular la matriz de covarianza, se observa que

$$\begin{aligned} \mathbf{x}\mathbf{x}^T &= \begin{bmatrix} (A + w[0])^2 & (A + w[0])(A + w[1]) \\ (A + w[0])(A + w[1]) & (A + w[1])^2 \end{bmatrix} \\ &= \begin{bmatrix} A^2 + w^2[0] + 2Aw[0] & A^2 + Aw[0] + Aw[1] + w[0]w[1] \\ A^2 + Aw[0] + Aw[1] + w[0]w[1] & A^2 + w^2[1] + 2Aw[1] \end{bmatrix}, \end{aligned}$$

y la matriz de covarianza queda

$$\begin{aligned} \mathbf{C} &\stackrel{(a)}{=} E(\mathbf{x}\mathbf{x}^T) \\ &= E \left(\begin{bmatrix} A^2 + w^2[0] + 2Aw[0] & A^2 + Aw[0] + Aw[1] + w[0]w[1] \\ A^2 + Aw[0] + Aw[1] + w[0]w[1] & A^2 + w^2[1] + 2Aw[1] \end{bmatrix} \right) \\ &\stackrel{(b)}{=} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, \end{aligned}$$

donde en (a) se consideró que \mathbf{x} tiene media nula y en (b) que $E(A^2) = E(w^2[0]) = E(w^2[1]) = 1$ y $E(Aw[0]) = E(Aw[1]) = E(w[0]w[1]) = 0$ por ser variables aleatorias independientes con PDF $\mathcal{N}(0, 1)$. Como

$$\det(\mathbf{C}) = 3, \quad \mathbf{C}^{-1} = \frac{1}{3} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix},$$

y además

$$\begin{aligned} \mathbf{x}^T \mathbf{C}^{-1} \mathbf{x} &= \begin{bmatrix} x[0] & x[1] \end{bmatrix} \frac{1}{3} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} x[0] \\ x[1] \end{bmatrix} \\ &= \frac{1}{3} \begin{bmatrix} x[0] & x[1] \end{bmatrix} \begin{bmatrix} 2x[0] - x[1] \\ -x[0] + 2x[1] \end{bmatrix} \\ &= \frac{2}{3} (x^2[0] - x[0]x[1] + x^2[1]), \end{aligned}$$

combinando los resultados, se obtiene que la PDF conjunta queda

$$p(\mathbf{x}) = \frac{1}{2\pi\sqrt{3}} \exp \left[-\frac{1}{3} (x^2[0] - x[0]x[1] + x^2[1]) \right].$$

Se calculará a continuación $p(x[n])$ para calcular $p(x[0])p(x[1])$ y comparar el resultado con la PDF conjunta. Recordar que

$$x[n] = A + w[n],$$

donde A y $w[n]$ son independientes con PDF $\mathcal{N}(0, 1)$. Puede demostrarse que si $x \sim \mathcal{N}(\mu_x, \sigma_x^2)$ y $y \sim \mathcal{N}(\mu_y, \sigma_y^2)$ son independientes, la variable aleatoria $z = x + y$ tiene PDF $\mathcal{N}(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2)$ (ver el ejemplo 6-34 de [2]). Por lo tanto, $x[n] \sim \mathcal{N}(0, 2)$ y su PDF es

$$p(x[n]) = \frac{1}{\sqrt{4\pi}} \exp\left(-\frac{1}{4}x^2[n]\right)$$

y por lo tanto

$$p(x[0])p(x[1]) = \frac{1}{4\pi} \exp\left[-\frac{1}{4}(x^2[0] + x^2[1])\right].$$

Se concluye que

$$p(\mathbf{x}) \neq p(x[0])p(x[1]),$$

que implica que las variables aleatorias $x[0]$ y $x[1]$ no son independientes si no se condicionan en A .

10.5.3. Problema 3

Se observan los datos $x[n]$ para $n = 0, \dots, N-1$. Cada muestra tiene una PDF condicional

$$p(x[n]|\theta) = \begin{cases} \exp[-(x[n] - \theta)] & x[n] > \theta \\ 0 & x[n] < \theta, \end{cases}$$

y condicionadas en θ las observaciones son independientes. La PDF a priori es

$$p(\theta) = \begin{cases} \exp(-\theta) & \theta > 0 \\ 0 & \theta < 0. \end{cases}$$

Encontrar el estimador MMSE de θ . Ver el problema de la sección 10.5.2 para la definición de independencia condicional.

Solución El estimador MMSE es $\hat{\theta} = E(\theta|\mathbf{x})$, por lo que hay que calcular $p(\theta|\mathbf{x})$, que puede hacerse empleando la ecuación 10.6. La PDF condicional de los datos es

$$\begin{aligned} p(\mathbf{x}|\theta) &\stackrel{(a)}{=} \prod_{n=0}^{N-1} p(x[n]|\theta) \\ &= \begin{cases} \exp\left[-\sum_{n=0}^{N-1} (x[n] - \theta)\right] & x[0] > \theta, \dots, x[N-1] > \theta \quad \text{o} \quad \min x[n] > \theta \\ 0 & \text{en otro caso,} \end{cases} \\ &= \begin{cases} \exp[-N(\bar{x} - \theta)] & \min x[n] > \theta \\ 0 & \text{en otro caso,} \end{cases} \end{aligned}$$

donde en (a) se empleó la hipótesis de independencia condicional. La PDF conjunta es

$$\begin{aligned} p(\mathbf{x}|\theta)p(\theta) &= \begin{cases} \exp[-N(\bar{x} - \theta)] \exp(-\theta) & \min x[n] > \theta \quad \text{y} \quad \theta > 0 \\ 0 & \text{en otro caso} \end{cases} \\ &= \begin{cases} \exp(-N\bar{x}) \exp[(N-1)\theta] & 0 < \theta < \min x[n] \\ 0 & \text{en otro caso,} \end{cases} \end{aligned}$$

por lo que la PDF a posteriori es

$$p(\theta|\mathbf{x}) = \frac{p(\mathbf{x}|\theta)p(\theta)}{\int_{-\infty}^{\infty} p(\mathbf{x}|\theta)p(\theta) d\theta} = \frac{e^{-N\bar{x}} e^{(N-1)\theta}}{e^{-N\bar{x}} \int_0^{\xi} e^{(N-1)\theta} d\theta} = \frac{e^{(N-1)\theta}}{\int_0^{\xi} e^{(N-1)\theta} d\theta},$$

donde se definió $\xi = \min x[n]$. Como

$$\int_0^{\xi} e^{(N-1)\theta} d\theta = \frac{e^{(N-1)\theta}}{N-1} \Big|_0^{\xi} = \frac{e^{(N-1)\xi} - 1}{N-1},$$

la PDF a posteriori queda

$$p(\theta|\mathbf{x}) = \begin{cases} \frac{(N-1)e^{(N-1)\theta}}{e^{(N-1)\xi} - 1} & 0 < \theta < \xi \\ 0 & \text{en otro caso.} \end{cases}$$

El estimador MMSE es

$$\begin{aligned} \hat{\theta} &= E(\theta|\mathbf{x}) \\ &= \int_{-\infty}^{\infty} \theta p(\theta|\mathbf{x}) d\theta \\ &= \int_0^{\xi} \theta \frac{(N-1)e^{(N-1)\theta}}{e^{(N-1)\xi} - 1} d\theta \\ &\stackrel{(a)}{=} \frac{1}{(N-1)(e^{(N-1)\xi} - 1)} \int_0^{(N-1)\xi} x e^x dx \\ &\stackrel{(b)}{=} \frac{1}{(N-1)(e^{(N-1)\xi} - 1)} \left[x e^x \Big|_0^{(N-1)\xi} - \int_0^{(N-1)\xi} e^x dx \right] \\ &= \frac{1}{(N-1)(e^{(N-1)\xi} - 1)} \left[(N-1)\xi e^{(N-1)\xi} - e^x \Big|_0^{(N-1)\xi} \right] \\ &= \frac{1}{(N-1)(e^{(N-1)\xi} - 1)} \left[(N-1)\xi e^{(N-1)\xi} - (e^{(N-1)\xi} - 1) \right] \\ &= \frac{\xi e^{(N-1)\xi}}{e^{(N-1)\xi} - 1} - \frac{1}{N-1}, \end{aligned}$$

concluyendo que

$$\hat{\theta} = \frac{(\min x[n])e^{(N-1)\min x[n]}}{e^{(N-1)\min x[n]} - 1} - \frac{1}{N-1}.$$

10.5.4. Problema 4

Repetir el problema de la sección 10.5.3 pero con PDF condicional

$$p(x[n]|\theta) = \begin{cases} \frac{1}{\theta} & 0 \leq x[n] \leq \theta \\ 0 & \text{en otro caso} \end{cases}$$

y con θ con PDF a prior $\mathcal{U}[0, \beta]$. ¿Qué ocurre si β es grande de forma de que el conocimiento a priori es pequeño?

Solución Siguiendo el mismo razonamiento que en el problema de la sección 10.5.3, como las muestras son condicionalmente independientes, la PDF condicional de los datos es

$$p(\mathbf{x}|\theta) = \begin{cases} \frac{1}{\theta^N} & 0 \leq x[0] \leq \theta, \dots, 0 \leq x[N-1] \leq \theta \quad \text{o} \quad 0 \leq \min x[n] \leq \max x[n] \leq \theta \\ 0 & \text{en otro caso.} \end{cases}$$

Considerando que $\theta \sim \mathcal{U}[0, \beta]$,

$$\begin{aligned} p(\mathbf{x}|\theta)p(\theta) &= \begin{cases} \frac{1}{\theta^N} \frac{1}{\beta} & 0 \leq \min x[n] \leq \max x[n] \leq \theta \quad \text{y} \quad 0 \leq \theta \leq \beta \\ 0 & \text{en otro caso} \end{cases} \\ &= \begin{cases} \frac{1}{\theta^N} \frac{1}{\beta} & \max x[n] \leq \theta \leq \beta \\ 0 & \text{en otro caso.} \end{cases} \end{aligned}$$

y la PDF a poseteriori es

$$p(\theta|\mathbf{x}) = \frac{p(\mathbf{x}|\theta)p(\theta)}{\int_{-\infty}^{\infty} p(\mathbf{x}|\theta)p(\theta) d\theta} = \frac{\frac{1}{\theta^N} \frac{1}{\beta}}{\frac{1}{\beta} \int_{\xi}^{\beta} \frac{1}{\theta^N} d\theta} = \frac{\frac{1}{\theta^N}}{\int_{\xi}^{\beta} \frac{1}{\theta^N} d\theta}$$

donde se definió $\xi = \min(x[n])$. El denominador es

$$\int_{\xi}^{\beta} \frac{1}{\theta^N} d\theta = -\frac{\theta^{-(N-1)}}{N-1} \Big|_{\xi}^{\beta} = -\frac{1}{N-1} \left[\beta^{-(N-1)} - \xi^{-(N-1)} \right],$$

resultando en

$$p(\theta|\mathbf{x}) = \frac{(N-1)\theta^{-N}}{\xi^{-(N-1)} - \beta^{-(N-1)}}.$$

El estimador MMSE es

$$\begin{aligned} \hat{\theta} &= E(\theta|\mathbf{x}) \\ &= \int_{-\infty}^{\infty} \theta p(\theta|\mathbf{x}) d\theta \\ &= \frac{N-1}{\xi^{-(N-1)} - \beta^{-(N-1)}} \int_{\xi}^{\beta} \theta \theta^{-N} d\theta \\ &= \frac{N-1}{\xi^{-(N-1)} - \beta^{-(N-1)}} \int_{\xi}^{\beta} \theta^{-(N-1)} d\theta \\ &= \frac{N-1}{\xi^{-(N-1)} - \beta^{-(N-1)}} \left(-\frac{\theta^{-(N-2)}}{N-2} \Big|_{\xi}^{\beta} \right) \\ &= \frac{N-1}{\xi^{-(N-1)} - \beta^{-(N-1)}} \left(\frac{\xi^{-(N-2)} - \beta^{-(N-2)}}{N-2} \right) \\ &= \frac{N-1}{N-2} \frac{\xi^{-(N-2)} - \beta^{-(N-2)}}{\xi^{-(N-1)} - \beta^{-(N-1)}}, \end{aligned}$$

concluyendo que

$$\hat{\theta} = \frac{N-1}{N-2} \frac{(\max x[n])^{-(N-2)} - \beta^{-(N-2)}}{(\max x[n])^{-(N-1)} - \beta^{-(N-1)}}.$$

10.5.5. Problema 5

Derivar el estimador MMSE evaluando

$$\begin{aligned} \text{Bmse}(\hat{\theta}) &= E_{x, \theta} \left[(\theta - \hat{\theta})^2 \right] \\ &= E_{x, \theta} \left\{ \left[(\theta - E(\theta|\mathbf{x})) - (E(\theta|\mathbf{x}) - \hat{\theta}) \right]^2 \right\}. \end{aligned}$$

Sugerencia: emplear el resultado $E_{x, \theta}(\cdot) = E_x[E_{\theta|x}(\cdot)]$.

Solución Se mostrará primero que $E_{x, \theta}(\cdot) = E_x[E_{\theta|x}(\cdot)]$. Para hacerlo, se considera la variable aleatoria $g(\mathbf{x}, \theta)$ y se observa que

$$\begin{aligned} E_{x, \theta}[g(\mathbf{x}, \theta)] &= \iint g(\mathbf{x}, \theta)p(\mathbf{x}, \theta) d\mathbf{x} d\theta \\ &= \iint g(\mathbf{x}, \theta)p(\theta|\mathbf{x})p(\mathbf{x}) d\mathbf{x} d\theta \\ &= \int \left[\int g(\mathbf{x}, \theta)p(\theta|\mathbf{x}) d\theta \right] p(\mathbf{x}) d\mathbf{x} \\ &= \int E_{\theta|x}[g(\mathbf{x}, \theta)]p(\mathbf{x}) d\mathbf{x} \\ &= E_x\{E_{\theta|x}[g(\mathbf{x}, \theta)]\}. \end{aligned}$$

Considerando este resultado, el Bmse puede escribirse como

$$\begin{aligned} \text{Bmse}(\hat{\theta}) &= E_x \left[E_{\theta|x} \left\{ \left[(\theta - E(\theta|\mathbf{x})) - (E(\theta|\mathbf{x}) - \hat{\theta}) \right]^2 \right\} \right] \\ &= E_x \left\{ E_{\theta|x} [(\theta - E(\theta|\mathbf{x}))^2] + 2E_{\theta|x} [(\theta - E(\theta|\mathbf{x}))(E(\theta|\mathbf{x}) - \hat{\theta})] + E_{\theta|x} [(E(\theta|\mathbf{x}) - \hat{\theta})^2] \right\}. \end{aligned}$$

El segundo sumando es nulo, ya que

$$\begin{aligned} E_{\theta|x} [(\theta - E(\theta|\mathbf{x}))(E(\theta|\mathbf{x}) - \hat{\theta})] &\stackrel{(a)}{=} [E(\theta|\mathbf{x}) - \hat{\theta}] E_{\theta|x} [(\theta - E(\theta|\mathbf{x}))] \\ &= [E(\theta|\mathbf{x}) - \hat{\theta}] [E_{\theta|x}(\theta) - E(\theta|\mathbf{x})] \\ &= [E(\theta|\mathbf{x}) - \hat{\theta}] [E(\theta|\mathbf{x}) - E(\theta|\mathbf{x})] \\ &= 0, \end{aligned}$$

donde en (a) se consideró que $\hat{\theta}$ dado \mathbf{x} es una constante por ser una función de \mathbf{x} . Por lo tanto,

$$\text{Bmse}(\hat{\theta}) = E_x \left\{ E_{\theta|x} [(\theta - E(\theta|\mathbf{x}))^2] + E_{\theta|x} [(E(\theta|\mathbf{x}) - \hat{\theta})^2] \right\}.$$

Observando que el segundo sumando es no negativo para todo $\hat{\theta}$, para minimizar el Bmse hay que elegir $\hat{\theta}$ de forma de anularlo, resultando en

$$\hat{\theta} = E(\theta|\mathbf{x}).$$

10.5.6. Problema 6

En el ejemplo de la sección 10.2.1 modificar el modelo de datos como

$$x[n] = A + w[n], \quad n = 0, \dots, N-1,$$

donde $w[n]$ es WGN con varianza σ_+^2 si $A \geq 0$ y σ_-^2 si $A < 0$. Encontrar la PDF $p(\mathbf{x}|A)$. Compararla con $p(\mathbf{x}; A)$ del caso clásico en el que A es determinístico y $w[n]$ es WGN con varianza σ^2 si

- $\sigma_+^2 = \sigma_-^2$
- $\sigma_+^2 \neq \sigma_-^2$

Solución En este caso se cumple que

$$\text{var}(w[n]) = \begin{cases} \sigma_+^2 & \text{si } A \geq 0 \\ \sigma_-^2 & \text{si } A < 0, \end{cases}$$

por lo que el ruido no es independiente de A . La probabilidad condicional de una muestra es

$$p_{x[n]}(x[n]|A) = p_{w[n]}(x[n] - A|A) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma_+^2}} \exp\left[-\frac{1}{2\sigma_+^2}(x[n] - A)^2\right] & \text{si } A \geq 0 \\ \frac{1}{\sqrt{2\pi\sigma_-^2}} \exp\left[-\frac{1}{2\sigma_-^2}(x[n] - A)^2\right] & \text{si } A < 0, \end{cases}$$

Al condicionar en A , los $w[n]$ son independientes, y también lo son los $x[n]$. Por lo tanto,

$$p(\mathbf{x}|A) = p(x[0]|A)p(x[1]|A)\dots p(x[N-1]|A) = \begin{cases} \frac{1}{(2\pi\sigma_+^2)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma_+^2} \sum_{n=0}^{N-1} (x[n] - A)^2\right] & \text{si } A \geq 0 \\ \frac{1}{(2\pi\sigma_-^2)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma_-^2} \sum_{n=0}^{N-1} (x[n] - A)^2\right] & \text{si } A < 0. \end{cases}$$

Por otro lado, en el caso clásico con A determinístico,

$$p(\mathbf{x}; A) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - A)^2\right].$$

Se concluye que si $\sigma_+^2 \neq \sigma_-^2$, $p(\mathbf{x}|A) \neq p(\mathbf{x}; A)$. En el caso en que $\sigma_+^2 = \sigma_-^2 = \sigma^2$, $p(\mathbf{x}|A)$ y $p(\mathbf{x}; A)$ son idénticos.

10.5.7. Problema 7

Graficar \hat{A} dado por la ecuación 10.4 como una función de \bar{x} si $\sqrt{\sigma^2/N} = 1$ para $A_0 = 3$ y $A_0 = 10$. Comparar los resultados con $\hat{A} = \bar{x}$. Sugerencia: notar que el numerador puede evaluarse de forma analítica y el denominador está relacionado con la función de distribución de probabilidad de una variable aleatoria gaussiana.

Solución El estimador \hat{A} dado por la ecuación 10.4 es la esperanza de una variable aleatoria con PDF gaussiana truncada. Considérese el caso general de una variable aleatoria gaussiana con media μ y varianza σ^2 truncada entre A y B . La PDF está dada por¹

$$p(x) = \begin{cases} \frac{\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x - \mu)^2\right]}{\int_A^B \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x - \mu)^2\right] dx} & \text{si } A \leq x \leq B \\ 0 & \text{en otro caso.} \end{cases}$$

¹Ver por ejemplo: https://en.wikipedia.org/wiki/Truncated_normal_distribution

La media es

$$\begin{aligned}
E(x) &= \frac{\int_A^B x \frac{1}{\sqrt{2\pi}\sigma^2} \exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right] dx}{\int_A^B \frac{1}{\sqrt{2\pi}\sigma^2} \exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right] dx} \\
&\stackrel{(a)}{=} \frac{\int_{(A-\mu)/\sigma}^{(B-\mu)/\sigma} (\sigma u + \mu) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du}{\int_{(A-\mu)/\sigma}^{(B-\mu)/\sigma} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du} \\
&= \mu \frac{\int_{(A-\mu)/\sigma}^{(B-\mu)/\sigma} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du}{\int_{(A-\mu)/\sigma}^{(B-\mu)/\sigma} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du} + \sigma \frac{\int_{(A-\mu)/\sigma}^{(B-\mu)/\sigma} \frac{u}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du}{\int_{(A-\mu)/\sigma}^{(B-\mu)/\sigma} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du} \\
&= \mu + \sigma \frac{\left. -\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) \right|_{(A-\mu)/\sigma}^{(B-\mu)/\sigma}}{\int_{(A-\mu)/\sigma}^{(B-\mu)/\sigma} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du} \\
&= \mu + \sigma \frac{\frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}(A-\mu)^2\right] - \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}(B-\mu)^2\right]}{\int_{-\infty}^{(B-\mu)/\sigma} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du - \int_{-\infty}^{(A-\mu)/\sigma} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du},
\end{aligned}$$

donde en (a) se realizó el cambio de variable $u = (x - \mu)/\sigma$. Esto se puede expresar como

$$E(x) = \mu + \sigma \frac{\phi\left(\frac{A-\mu}{\sigma}\right) - \phi\left(\frac{B-\mu}{\sigma}\right)}{\Phi\left(\frac{B-\mu}{\sigma}\right) - \Phi\left(\frac{A-\mu}{\sigma}\right)},$$

donde

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \quad \text{y} \quad \Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du$$

son respectivamente las funciones de densidad y distribución de probabilidad de una variable aleatoria gaussiana normalizada. Como el estimador \hat{A} es la media de una variable aleatoria con PDF $\mathcal{N}(\bar{x}, \sigma^2/N)$ truncada entre $-A_0$ y A_0 , se puede calcular como

$$\hat{A} = \bar{x} + \frac{\sigma}{\sqrt{N}} \frac{\phi\left(\frac{-A_0 - \bar{x}}{\frac{\sigma}{\sqrt{N}}}\right) - \phi\left(\frac{A_0 - \bar{x}}{\frac{\sigma}{\sqrt{N}}}\right)}{\Phi\left(\frac{A_0 - \bar{x}}{\frac{\sigma}{\sqrt{N}}}\right) - \Phi\left(\frac{-A_0 - \bar{x}}{\frac{\sigma}{\sqrt{N}}}\right)}.$$

En la figura 10.2 se muestra una implementación en el lenguaje Python del estimador \hat{A} cuando $\sqrt{\sigma^2/N} = 1$. Recordar que la PDF a priori es $\mathcal{U}[-A_0, A_0]$. Si A_0 es pequeño, la PDF a posteriori es una gaussiana de media \bar{x} fuertemente truncada y la media difiere de \bar{x} . En la figura 10.3 se muestra el valor del estimador \hat{A} como una función de \bar{x} para $A_0 = 3$ y $A_0 = 10$ cuando $\sqrt{\sigma^2/N} = 1$. Se observa que para $A_0 = 10$ se cumple que $\hat{A} \approx \bar{x}$ si $\bar{x} < A_0$.

10.5.8. Problema 8

Una variable aleatoria θ tiene PDF $p(\theta)$. Se desea estimar una realización de θ sin la disponibilidad de datos. Para hacerlo, se propone un estimador MMSE que minimice $E[(\theta - \hat{\theta})^2]$, donde la esperanza

```
from scipy.stats import norm
hat_A = x + (norm.pdf(-A0 - x) - norm.pdf(A0 - x)) / (norm.cdf(A0 - x) - norm.cdf(-A0 - x))
```

Figura 10.2: Implementación en el lenguaje Python del estimador \hat{A} .

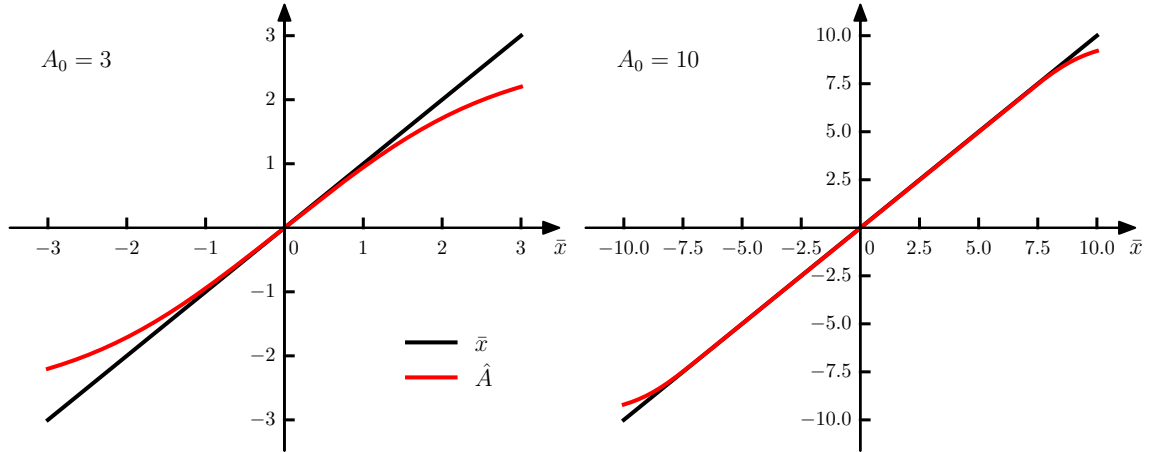


Figura 10.3: Comparación del estimador \hat{A} y la media muestral \bar{x} .

es respecto a $p(\theta)$ únicamente. Probar que el estimador MMSE es $\hat{\theta} = E(\theta)$. Aplicar los resultados al ejemplo de la sección 10.2.1 para mostrar que el MSE bayesiano mínimo se reduce cuando se incorporan los datos al estimador.

Solución Se quiere encontrar $\hat{\theta}$ que minimice

$$E[(\theta - \hat{\theta})^2] = \int (\theta - \hat{\theta})^2 p(\theta) d\theta.$$

Diferenciando respecto a $\hat{\theta}$ se tiene que

$$\begin{aligned} \frac{\partial}{\partial \hat{\theta}} E[(\theta - \hat{\theta})^2] &= \frac{\partial}{\partial \hat{\theta}} \int (\theta - \hat{\theta})^2 p(\theta) d\theta \\ &= \int \frac{\partial}{\partial \hat{\theta}} (\theta - \hat{\theta})^2 p(\theta) d\theta \\ &= -2 \int (\theta - \hat{\theta}) p(\theta) d\theta, \end{aligned}$$

e igualando a cero, se tiene que el MMSE mínimo es

$$\int (\theta - \hat{\theta}) p(\theta) d\theta = 0 \quad \Rightarrow \quad \hat{\theta} \int p(\theta) d\theta = \int \theta p(\theta) d\theta \quad \Rightarrow \quad \hat{\theta} = E(\theta),$$

que es lo que se quería mostrar. En el ejemplo de la sección 10.2.1 de la estimación del nivel de DC A en WGN con PDF a priori gaussiana $\mathcal{N}(\mu_A, \sigma_A^2)$, sin observar datos, el estimador MMSE sería

$$\hat{A} = E(A)$$

y el error mínimo es

$$\text{Bmse}(\hat{A}) = E[(A - E(A))^2] = \sigma_A^2.$$

Al incorporar datos, el error mínimo es (ver la ecuación 10.9)

$$\text{Bmse}(\hat{A}) = \sigma_{A|x}^2 = \frac{1}{\frac{N}{\sigma^2} + \frac{1}{\sigma_A^2}} < \frac{1}{\frac{1}{\sigma_A^2}} = \sigma_A^2,$$

concluyendo que el MSE bayesiano mínimo se reduce al incorporar los datos al estimador.

10.5.9. Problema 9

Un inspector de control de calidad tiene el trabajo de monitorear los valores de resistencia de resistencias manufacturadas. Lo hace seleccionando una resistencia de un lote y midiendo la resistencia con un óhmetro. Sabe que el óhmetro es de mala calidad y produce un error de medida, el que modela como una variable aleatoria $\mathcal{N}(0, 1)$. Por consiguiente, toma N medidas independientes. Además, sabe que las resistencias deberían ser de 100 ohms. Sin embargo, debido a las tolerancias de fabricación, generalmente tienen un error ϵ , con $\epsilon \sim \mathcal{N}(0, 0.011)$. Si el inspector elige una resistencia, ¿cuántas medidas con el óhmetro son necesarias para asegurar que un estimador MMSE de la resistencia R produzca el valor correcto a 0.1 ohm en promedio? ¿Cuántas medidas necesitaría si no tuviera conocimiento a priori de las tolerancias de fabricación?

Solución Sea la medida n -ésima

$$x[n] = R + w[n],$$

donde $w[n]$ es el error de medición con PDF $\mathcal{N}(0, \sigma^2)$, y R es el valor verdadero de la resistencia, que se modela como una variable aleatoria con PDF a priori $\mathcal{N}(\mu_R, \sigma_R^2)$. El error de medición está dado por las especificaciones del óhmetro, y en este caso es $w[n] \sim \mathcal{N}(0, 1)$. Además, como se sabe que las resistencias son de 100 ohms ($E(R) = 100$) con una tolerancia $\epsilon \sim \mathcal{N}(0, 0.011)$, el conocimiento a priori se modela como $R \sim \mathcal{N}(100, 0.011)$. Por lo tanto, $\sigma^2 = 1$, $\mu_R = 100$ y $\sigma_R^2 = 0.011$. Este problema es equivalente al problema de la estimación del nivel de DC en WGN con PDF a priori gaussiana discutido en la sección 10.2.1, donde se dedujo que el estimador MMSE está dado por la ecuación 10.7 y el MSE bayesiano mínimo está dado por la ecuación 10.9 y es

$$\text{Bmse}(\hat{R}) = \sigma_{R|x}^2 = \frac{1}{\frac{N}{\sigma^2} + \frac{1}{\sigma_R^2}}.$$

Se quiere saber el número de mediciones necesarias para que el estimador MMSE produzca el valor correcto en promedio con una diferencia menor a 0.1 ohm, o $|R - \hat{R}| \leq 0.1$ en promedio. Como por definición (ecuación 10.1)

$$\text{Bmse}(\hat{R}) = E[(R - \hat{R})^2],$$

se requiere que

$$\text{Bmse}(\hat{R}) = E[(R - \hat{R})^2] \leq (0.1)^2 = 0.01.$$

Por lo tanto,

$$\frac{1}{\frac{N}{\sigma^2} + \frac{1}{\sigma_R^2}} \leq 0.01,$$

y con los valores dados, la condición es

$$\frac{1}{N + \frac{1}{0.011}} \leq 0.01 \quad \Rightarrow \quad N + \frac{1}{0.011} \geq 100 \quad \Rightarrow \quad N \geq 100 - 90.9 \quad \Rightarrow \quad N \geq 9.1,$$

concluyendo que

$$N \geq 10.$$

Si no hay conocimiento a priori de R , hecho que se modela tomando $\sigma_R^2 \rightarrow \infty$, se tiene que

$$\text{Bmse}(\hat{R}) = \frac{\sigma^2}{N}.$$

La condición en este caso es

$$\frac{\sigma^2}{N} \leq 0.01 \quad \Rightarrow \quad \frac{1}{N} \leq 0.01 \quad \Rightarrow \quad N \geq 100.$$

10.5.10. Problema 10

En este problema se discute la propiedad de reproducción de algunas PDF. La PDF a posteriori es

$$p(\theta|\mathbf{x}) = \frac{p(\mathbf{x}|\theta)p(\theta)}{\int p(\mathbf{x}|\theta)p(\theta) d\theta},$$

donde el denominador no depende de θ . Si $p(\theta)$ se elige tal que al ser multiplicada por $p(\mathbf{x}|\theta)$ se obtenga una expresión de la misma forma que la PDF de θ , la PDF a posteriori $p(\theta|\mathbf{x})$ tendrá la misma forma de $p(\theta)$. Ese fue el caso del ejemplo de la sección 10.2.1 con la PDF gaussiana. Asíumase ahora que la PDF de $x[n]$ condicionada con θ es la PDF exponencial

$$p(x[n]|\theta) = \begin{cases} \theta \exp(-\theta x[n]) & x[n] > 0 \\ 0 & x[n] < 0 \end{cases}$$

donde los $x[n]$ son condicionalmente independientes (ver el problema de la sección 10.5.2). Luego, asíumase la PDF a priori gamma

$$p(\theta) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} \exp(-\lambda\theta) & \theta > 0 \\ 0 & \theta < 0 \end{cases}$$

donde $\lambda > 0$ y $\alpha > 0$. Encontrar la PDF a posteriori y compararla con la PDF a priori. Dicha PDF, en este caso la gamma, se denomina PDF *a priori conjugada*.

Solución Teniendo en cuenta que los $x[n]$ son condicionalmente independientes, la PDF condicional de los datos es

$$p(\mathbf{x}|\theta) = \begin{cases} \theta^N \exp\left(-\theta \sum_{n=0}^{N-1} x[n]\right) & x[0] > 0, x[1] > 0, \dots, x[N-1] > 0 \\ 0 & \text{en otro caso,} \end{cases}$$

que para $x[0] > 0, x[1] > 0, \dots, x[N-1] > 0$ puede escribirse como

$$p(\mathbf{x}|\theta) = \theta^N \exp(-N\theta\bar{x}).$$

Luego, para $x[0] > 0, x[1] > 0, \dots, x[N-1] > 0$ y $\theta > 0$ se tiene que

$$\begin{aligned} p(\mathbf{x}|\theta)p(\theta) &= \theta^N \exp(-N\theta\bar{x}) \frac{\lambda^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} \exp(-\lambda\theta) \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \theta^{N+\alpha-1} \exp[-(N\bar{x} + \lambda)\theta] \end{aligned}$$

y la PDF a posteriori queda

$$\begin{aligned} p(\theta|\mathbf{x}) &= \frac{p(\mathbf{x}|\theta)p(\theta)}{\int p(\mathbf{x}|\theta)p(\theta) d\theta} \\ &= \frac{\frac{\lambda^\alpha}{\Gamma(\alpha)} \theta^{N+\alpha-1} \exp[-(N\bar{x} + \lambda)\theta]}{\int \frac{\lambda^\alpha}{\Gamma(\alpha)} \theta^{N+\alpha-1} \exp[-(N\bar{x} + \lambda)\theta] d\theta} \\ &= \frac{\theta^{N+\alpha-1} \exp[-(N\bar{x} + \lambda)\theta]}{\int \theta^{N+\alpha-1} \exp[-(N\bar{x} + \lambda)\theta] d\theta} \\ &= \frac{\lambda'^{\alpha'}}{\Gamma(\alpha')} \theta^{\alpha'-1} \exp(-\lambda'\theta) \end{aligned}$$

donde en la última igualdad se notó que el numerador tiene la forma de una PDF gamma con $\alpha' = \alpha + N$ y $\lambda' = N\bar{x} + \lambda$ y el denominador es una constante de normalización para que $p(\theta|\mathbf{x})$ integre a 1, y por lo tanto debe corresponder a los factores acordes de la PDF gamma. Se concluye que

$$p(\theta|\mathbf{x}) = \begin{cases} \frac{(N\bar{x} + \lambda)^{\alpha+N}}{\Gamma(\alpha + N)} \theta^{\alpha+N-1} \exp[-(N\bar{x} + \lambda)\theta] & \theta > 0 \\ 0 & \theta < 0 \end{cases}$$

es una PDF gamma, al igual que la PDF a priori.

10.5.11. Problema 11

Se desea estimar el peso de una persona en base a su altura. Para ver si esto es factible, se toman los datos de $N = 200$ personas para generar los pares ordenados (h, P) , donde h denota la altura y P el peso. Los datos obtenidos se muestran en la figura 10.4. Explicar como deducir el peso de una persona basado en su altura empleando un estimador MMSE. ¿Qué hipótesis de modelado hay que realizar sobre los datos? Luego, el mismo experimento se realiza para personas de un planeta lejano. Los datos se muestran en la figura 10.4. ¿Cómo sería ahora el estimador MMSE del peso?

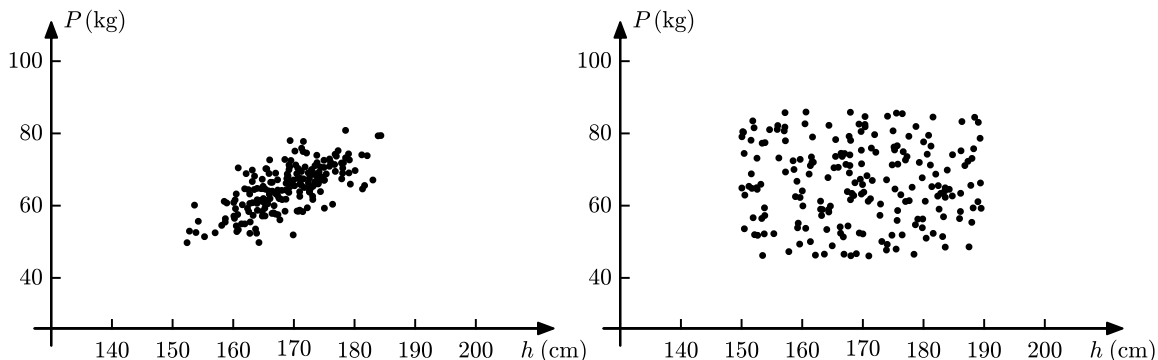


Figura 10.4: Pares (h, P) de la altura y peso respectivamente de $N = 200$ personas. A la izquierda se muestran los datos del planeta Tierra y a la derecha los datos de un planeta lejano.

Solución En la figura 10.4, se observa que los puntos del diagrama de dispersión correspondiente al planeta Tierra se encuentran concentrados en un elipse. Esto sugiere que el vector aleatorio $[h \ P]^T$ tiene PDF gaussiana bivariada. Por lo tanto, el estimador MMSE del peso en base de la altura está dado por la ecuación 10.11 y es

$$\hat{P} = E(P) + \frac{\text{cov}(h, P)}{\text{var}(h)}(h - E(h)).$$

Los puntos del diagrama de dispersión correspondiente al planeta lejano se encuentran concentrados en un rectángulo, y no hay un peso preferido para cada altura. Esto sugiere que la altura y el peso no están correlacionados y por lo tanto, $\text{cov}(h, P) = 0$. El estimador MMSE es

$$\hat{P} = E(P)$$

independiente de la altura.

10.5.12. Problema 12

Si $[x \ y]^T \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$, donde

$$\mathbf{C} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

y sea $g(y) = p(x_0, y)$ para algún $x = x_0$. Probar que $g(y)$ es maximizada para $y = \rho x_0$. Además, mostrar que $E(y|x_0) = \rho x_0$. ¿Porqué son los mismos valores? Si $\rho = 0$, ¿cual es el estimador MMSE de y basado en x ?

Solución De la ecuación 10.10 con $\mu_x = \mu_y = 0$ y $\sigma_x^2 = \sigma_y^2 = 1$, la PDF conjunta es

$$p(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp \left[-\frac{1}{2(1-\rho^2)} (x^2 - 2\rho xy + y^2) \right].$$

Fijando $x = x_0$, la función a maximizar en y es

$$g(y) = p(x_0, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp \left[-\frac{1}{2(1-\rho^2)} (x_0^2 - 2\rho x_0 y + y^2) \right].$$

Esta función es máxima cuando se minimiza el exponente

$$h(y) = \frac{1}{2(1-\rho^2)} (x_0^2 - 2\rho x_0 y + y^2),$$

es decir, cuando

$$\frac{\partial h(y)}{\partial y} = \frac{2y - 2\rho x_0}{2(1-\rho^2)} = 0,$$

concluyendo que el máximo se da en

$$y = \rho x_0.$$

Por otro lado, de la ecuación 10.11 con $E(x) = E(y) = 0$ y $\text{var}(x) = 1$, se tiene que

$$E(y|x_0) = \rho x_0.$$

Ambos valores coinciden debido a que

$$p(y|x_0) = \frac{p(x_0, y)}{p(x_0)},$$

y como el denominador es una constante, maximizar $p(x_0, y)$ equivale a maximizar $p(y|x_0)$. Como $p(y|x_0)$ es una gaussiana, el máximo se da en la media $E(y|x_0)$. Si $\rho = 0$, de la ecuación 10.11, se tiene que

$$E(y|x_0) = E(y) = 0.$$

En este caso, como x e y son no correlacionadas, x no aporta información sobre y .

10.5.13. Problema 13

Verificar las ecuaciones 10.19 y 10.20 empleando el lema de inversión de matrices. Sugerencia: verificar primero la ecuación 10.20 y emplear el resultado para verificar la ecuación 10.19.

Solución Se comenzará verificando la ecuación 10.20, que indica que

$$\mathbf{C}_{\theta|x} = (\mathbf{C}_\theta^{-1} + \mathbf{H}^T \mathbf{C}_w^{-1} \mathbf{H})^{-1}.$$

De la ecuación 10.18, la covarianza es

$$\mathbf{C}_{\theta|x} = \mathbf{C}_\theta - \mathbf{C}_\theta \mathbf{H}^T (\mathbf{H} \mathbf{C}_\theta \mathbf{H}^T + \mathbf{C}_w)^{-1} \mathbf{H} \mathbf{C}_\theta.$$

Como se demuestra en el apéndice A, el lema de inversión de matrices (ecuación A.1) indica que

$$\mathbf{A}^{-1} - \mathbf{A}^{-1} \mathbf{B} (\mathbf{D} \mathbf{A}^{-1} \mathbf{B} + \mathbf{C}^{-1})^{-1} \mathbf{D} \mathbf{A}^{-1} = (\mathbf{A} + \mathbf{B} \mathbf{C} \mathbf{D})^{-1}.$$

Por lo tanto, con $\mathbf{A}^{-1} = \mathbf{C}_\theta$, $\mathbf{B} = \mathbf{H}^T$, $\mathbf{D} = \mathbf{H}$ y $\mathbf{C}^{-1} = \mathbf{C}_w$ se tiene que

$$\begin{aligned}\mathbf{C}_{\theta|x} &= \mathbf{C}_\theta - \mathbf{C}_\theta \mathbf{H}^T (\mathbf{H} \mathbf{C}_\theta \mathbf{H}^T + \mathbf{C}_w)^{-1} \mathbf{H} \mathbf{C}_\theta \\ &= (\mathbf{C}_\theta^{-1} + \mathbf{H}^T \mathbf{C}_w^{-1} \mathbf{H})^{-1},\end{aligned}$$

que es lo que se quería mostrar. Se considerará esta última igualdad,

$$(\mathbf{C}_\theta^{-1} + \mathbf{H}^T \mathbf{C}_w^{-1} \mathbf{H})^{-1} = \mathbf{C}_\theta - \mathbf{C}_\theta \mathbf{H}^T (\mathbf{H} \mathbf{C}_\theta \mathbf{H}^T + \mathbf{C}_w)^{-1} \mathbf{H} \mathbf{C}_\theta, \quad (10.21)$$

para mostrar que se verifica la ecuación 10.19 de la media,

$$E(\boldsymbol{\theta}|\mathbf{x}) = \boldsymbol{\mu}_\theta + (\mathbf{C}_\theta^{-1} + \mathbf{H}^T \mathbf{C}_w^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C}_w^{-1} (\mathbf{x} - \mathbf{H} \boldsymbol{\mu}_\theta).$$

Notando que el término de la izquierda de la ecuación 10.21 aparece multiplicado por $\mathbf{H}^T \mathbf{C}_w^{-1}$ en la ecuación de la media, se parte multiplicando ambos lados de la igualdad 10.21 por este factor. De esta forma, se tiene que

$$\begin{aligned}(\mathbf{C}_\theta^{-1} + \mathbf{H}^T \mathbf{C}_w^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C}_w^{-1} &= \mathbf{C}_\theta \mathbf{H}^T \mathbf{C}_w^{-1} - \mathbf{C}_\theta \mathbf{H}^T (\mathbf{H} \mathbf{C}_\theta \mathbf{H}^T + \mathbf{C}_w)^{-1} \mathbf{H} \mathbf{C}_\theta \mathbf{H}^T \mathbf{C}_w^{-1} \\ &= \mathbf{C}_\theta \mathbf{H}^T [\mathbf{C}_w^{-1} - (\mathbf{H} \mathbf{C}_\theta \mathbf{H}^T + \mathbf{C}_w)^{-1} \mathbf{H} \mathbf{C}_\theta \mathbf{H}^T \mathbf{C}_w^{-1}].\end{aligned} \quad (10.22)$$

A continuación se opera con el término entre paréntesis rectos. Definiendo $\mathbf{A}^{-1} = \mathbf{C}_w^{-1}$ y $\mathbf{B} = \mathbf{H} \mathbf{C}_\theta \mathbf{H}^T$, el término entre paréntesis rectos se puede expresar como

$$\begin{aligned}\mathbf{A}^{-1} - (\mathbf{B} + \mathbf{A})^{-1} \mathbf{B} \mathbf{A}^{-1} &\stackrel{(a)}{=} \mathbf{A}^{-1} - (\mathbf{A}^{-1} \mathbf{B})(\mathbf{B}^{-1} \mathbf{A})(\mathbf{B} + \mathbf{A})^{-1} \mathbf{B} \mathbf{A}^{-1} \\ &= \mathbf{A}^{-1} - \mathbf{A}^{-1} \mathbf{B} [\mathbf{B}^{-1} \mathbf{A}(\mathbf{B} + \mathbf{A})^{-1} \mathbf{B}] \mathbf{A}^{-1} \\ &\stackrel{(b)}{=} \mathbf{A}^{-1} - \mathbf{A}^{-1} \mathbf{B} [\mathbf{B}^{-1}(\mathbf{B} + \mathbf{A}) \mathbf{A}^{-1} \mathbf{B}]^{-1} \mathbf{A}^{-1} \\ &= \mathbf{A}^{-1} - \mathbf{A}^{-1} \mathbf{B} (\mathbf{B}^{-1} \mathbf{B} \mathbf{A}^{-1} \mathbf{B} + \mathbf{B}^{-1} \mathbf{A} \mathbf{A}^{-1} \mathbf{B})^{-1} \mathbf{A}^{-1} \\ &= \mathbf{A}^{-1} - \mathbf{A}^{-1} \mathbf{B} (\mathbf{A}^{-1} \mathbf{B} + \mathbf{I})^{-1} \mathbf{A}^{-1} \\ &\stackrel{(c)}{=} (\mathbf{A} + \mathbf{B})^{-1},\end{aligned}$$

donde en (a) se multiplicó el segundo sumando por $(\mathbf{A}^{-1} \mathbf{B})$ y por $(\mathbf{A}^{-1} \mathbf{B})^{-1} = \mathbf{B}^{-1} \mathbf{A}$, en (b) se invirtió dos veces el término entre paréntesis rectos y en (c) se aplicó el lema de inversión de matrices (ecuación A.1) con $\mathbf{C} = \mathbf{I}$ y $\mathbf{D} = \mathbf{I}$. Se concluye que la ecuación 10.22 puede escribirse como

$$(\mathbf{C}_\theta^{-1} + \mathbf{H}^T \mathbf{C}_w^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C}_w^{-1} = \mathbf{C}_\theta \mathbf{H}^T (\mathbf{H} \mathbf{C}_\theta \mathbf{H}^T + \mathbf{C}_w)^{-1}.$$

Sustituyendo este resultado en la ecuación 10.19 de la media, se obtiene que

$$E(\boldsymbol{\theta}|\mathbf{x}) = \boldsymbol{\mu}_\theta + \mathbf{C}_\theta \mathbf{H}^T (\mathbf{H} \mathbf{C}_\theta \mathbf{H}^T + \mathbf{C}_w)^{-1} (\mathbf{x} - \mathbf{H} \boldsymbol{\mu}_\theta),$$

que es la ecuación 10.17.

10.5.14. Problema 14

Se observan los datos

$$x[n] = Ar^n + w[n], \quad n = 0, \dots, N-1$$

donde r es conocido, $w[n]$ es WGN con varianza σ^2 y $A \sim \mathcal{N}(0, \sigma^2)$ independiente de $w[n]$. Encontrar el estimador MMSE de A así como el MSE bayesiano mínimo.

Solución Se resolverá el problema de dos formas. Una es mediante el cálculo explícito de la PDF a posteriori y la otra considerando que el modelo de los datos es el modelo lineal bayesiano discutido en la sección 10.4.

La PDF a posteriori puede calcularse empleando la ecuación 10.6. La PDF condicional de los datos es

$$\begin{aligned} p(\mathbf{x}|A) &= \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - Ar^n)^2 \right] \\ &= \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left[-\frac{1}{2\sigma^2} \left(\sum_{n=0}^{N-1} x^2[n] - 2A \sum_{n=0}^{N-1} r^n x[n] + A^2 \sum_{n=0}^{N-1} r^{2n} \right) \right]. \end{aligned}$$

Luego,

$$\begin{aligned} p(\mathbf{x}|A)p(A) &= \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left[-\frac{1}{2\sigma^2} \left(\sum_{n=0}^{N-1} x^2[n] - 2A \sum_{n=0}^{N-1} r^n x[n] + A^2 \sum_{n=0}^{N-1} r^{2n} \right) \right] \frac{1}{\sqrt{2\pi\sigma_A^2}} \exp \left(-\frac{A^2}{2\sigma_A^2} \right) \\ &= \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \frac{1}{\sqrt{2\pi\sigma_A^2}} \exp \left(-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} x^2[n] \right) \\ &\quad \times \exp \left[-\frac{1}{2\sigma^2} \left(-2A \sum_{n=0}^{N-1} r^n x[n] + A^2 \sum_{n=0}^{N-1} r^{2n} \right) - \frac{A^2}{2\sigma_A^2} \right], \end{aligned}$$

y la PDF a posteriori queda

$$\begin{aligned} p(A|x) &= \frac{p(\mathbf{x}|A)p(A)}{\int_{-\infty}^{\infty} p(\mathbf{x}|A)p(A) dA} \\ &\stackrel{(a)}{=} \frac{\exp \left[-\frac{1}{2\sigma^2} \left(-2A \sum_{n=0}^{N-1} r^n x[n] + A^2 \sum_{n=0}^{N-1} r^{2n} \right) - \frac{A^2}{2\sigma_A^2} \right]}{\int_{-\infty}^{\infty} \exp \left[-\frac{1}{2\sigma^2} \left(-2A \sum_{n=0}^{N-1} r^n x[n] + A^2 \sum_{n=0}^{N-1} r^{2n} \right) - \frac{A^2}{2\sigma_A^2} \right] dA} \\ &= \frac{\exp \left[-\frac{1}{2}Q(A) \right]}{\int_{-\infty}^{\infty} \exp \left[-\frac{1}{2}Q(A) \right] dA}, \end{aligned}$$

donde en (a) se consideró que los factores que no dependen de A se cancelan entre el numerador y el denominador. El denominador no depende de A debido a la integración, siendo únicamente un factor de normalización, y además el argumento de la exponencial es cuadrático en A . Por lo tanto, $p(A|\mathbf{x})$ debe ser gaussiana. Continuando, se tiene que $Q(A)$ es

$$\begin{aligned} Q(A) &= A^2 \left(\frac{1}{\sigma^2} \sum_{n=0}^{N-1} r^{2n} + \frac{1}{\sigma_A^2} \right) - 2A \frac{1}{\sigma^2} \sum_{n=0}^{N-1} r^n x[n] \\ &= \left(\frac{1}{\sigma^2} \sum_{n=0}^{N-1} r^{2n} + \frac{1}{\sigma_A^2} \right) \left(A^2 - 2A \frac{\frac{1}{\sigma^2} \sum_{n=0}^{N-1} r^n x[n]}{\frac{1}{\sigma^2} \sum_{n=0}^{N-1} r^{2n} + \frac{1}{\sigma_A^2}} \right) \\ &= \left(\frac{1}{\sigma^2} \sum_{n=0}^{N-1} r^{2n} + \frac{1}{\sigma_A^2} \right) \left[\left(A - \frac{\frac{1}{\sigma^2} \sum_{n=0}^{N-1} r^n x[n]}{\frac{1}{\sigma^2} \sum_{n=0}^{N-1} r^{2n} + \frac{1}{\sigma_A^2}} \right)^2 - \left(\frac{\frac{1}{\sigma^2} \sum_{n=0}^{N-1} r^n x[n]}{\frac{1}{\sigma^2} \sum_{n=0}^{N-1} r^{2n} + \frac{1}{\sigma_A^2}} \right)^2 \right] \end{aligned}$$

y definiendo

$$\frac{1}{\sigma_{A|x}^2} = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} r^{2n} + \frac{1}{\sigma_A^2} \quad \Rightarrow \quad \sigma_{A|x}^2 = \frac{1}{\frac{1}{\sigma^2} \sum_{n=0}^{N-1} r^{2n} + \frac{1}{\sigma_A^2}}$$

y

$$\mu_{A|x} = \frac{\frac{1}{\sigma^2} \sum_{n=0}^{N-1} r^n x[n]}{\frac{1}{\sigma^2} \sum_{n=0}^{N-1} r^{2n} + \frac{1}{\sigma_A^2}}$$

se puede escribir como

$$Q(A) = \frac{1}{\sigma_{A|x}^2} \left[(A - \mu_{A|x})^2 - \mu_{A|x}^2 \right] = \frac{1}{\sigma_{A|x}^2} (A - \mu_{A|x})^2 - \frac{\mu_{A|x}^2}{\sigma_{A|x}^2}.$$

Sustituyendo este resultado en la PDF a posteriori, se obtiene que

$$\begin{aligned} p(A|\mathbf{x}) &= \frac{\exp \left[-\frac{1}{2\sigma_{A|x}^2} (A - \mu_{A|x})^2 \right] \exp \left(\frac{1}{2} \frac{\mu_{A|x}^2}{\sigma_{A|x}^2} \right)}{\int_{-\infty}^{\infty} \exp \left[-\frac{1}{2\sigma_{A|x}^2} (A - \mu_{A|x})^2 \right] \exp \left(\frac{1}{2} \frac{\mu_{A|x}^2}{\sigma_{A|x}^2} \right) dA} \\ &= \frac{1}{\sqrt{2\pi\sigma_{A|x}^2}} \exp \left[-\frac{1}{2\sigma_{A|x}^2} (A - \mu_{A|x})^2 \right], \end{aligned}$$

donde nuevamente se consideró que los factores que no dependen de A se cancelan en el numerador y el denominador y el denominador es una constante de forma de que la PDF integre a 1. Por lo tanto, el estimador MMSE es

$$\hat{A} = E(A|\mathbf{x}) = \mu_{A|x} = \frac{\frac{1}{\sigma^2} \sum_{n=0}^{N-1} r^n x[n]}{\frac{1}{\sigma^2} \sum_{n=0}^{N-1} r^{2n} + \frac{1}{\sigma_A^2}}.$$

Además, como $\text{var}(A|\mathbf{x}) = \sigma_{A|x}^2$ no depende de \mathbf{x} , de la ecuación 10.8, el MSE bayesiano mínimo es

$$\text{Bmse}(\hat{A}) = \sigma_{A|x}^2 = \frac{1}{\frac{1}{\sigma^2} \sum_{n=0}^{N-1} r^{2n} + \frac{1}{\sigma_A^2}}.$$

Considerando que el modelo de los datos se trata del modelo lineal bayesiano, el problema puede resolverse empleando los resultados de la sección 10.4. En este caso, el modelo de los datos es

$$\mathbf{x} = \mathbf{H}\mathbf{A} + \mathbf{w}, \quad \text{con} \quad \mathbf{H} = [1 \ r \ r^2 \ \dots \ r^{N-1}]^T.$$

El estimador MMSE está dado por la ecuación 10.19,

$$\hat{A} = E(A|\mathbf{x}) = \boldsymbol{\mu}_A + (\mathbf{C}_A^{-1} + \mathbf{H}^T \mathbf{C}_w^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C}_w^{-1} (\mathbf{x} - \mathbf{H} \boldsymbol{\mu}_A),$$

y con

$$\boldsymbol{\mu}_A = 0, \quad \mathbf{C}_A = \sigma_A^2, \quad \mathbf{C}_w = \sigma^2 \mathbf{I},$$

se tiene que

$$\mathbf{H}^T \mathbf{C}_w^{-1} \mathbf{H} = \frac{1}{\sigma^2} \mathbf{H}^T \mathbf{H} = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} r^{2n} \quad \text{y} \quad \mathbf{H}^T \mathbf{C}_w^{-1} (\mathbf{x} - \mathbf{H} \boldsymbol{\mu}_A) = \frac{1}{\sigma^2} \mathbf{H}^T \mathbf{x} = \frac{1}{\sigma^2} \sum_{n=0}^{N-1} r^n x[n],$$

resultando en

$$\hat{A} = \frac{\frac{1}{\sigma^2} \sum_{n=0}^{N-1} r^n x[n]}{\frac{1}{\sigma_A^2} + \frac{1}{\sigma^2} \sum_{n=0}^{N-1} r^{2n}}.$$

Para calcular el MSE bayesiano mínimo, de la ecuación 10.20 se obtiene que

$$\sigma_{A|x}^2 = \frac{1}{\frac{1}{\sigma_A^2} + \frac{1}{\sigma^2} \sum_{n=0}^{N-1} r^{2n}},$$

y como es independiente de los datos, de la ecuación 10.8, el MSE bayesiano mínimo es

$$\text{Bmse}(\hat{A}) = \sigma_{A|x}^2 = \frac{1}{\frac{1}{\sigma_A^2} + \frac{1}{\sigma^2} \sum_{n=0}^{N-1} r^{2n}}.$$

10.5.15. Problema 15

Una medida de aleatoriedad de una variable aleatoria θ es la *entropía*, definida como

$$H(\theta) = E(-\ln p(\theta)) = - \int \ln p(\theta) d\theta.$$

Si $\theta \sim \mathcal{N}(\mu_\theta, \sigma_\theta^2)$, calcular la entropía y relacionarla con la concentración de la PDF. Luego de observar los datos, la entropía de la PDF a posteriori puede definirse como

$$H(\theta|\mathbf{x}) = E(-\ln p(\theta|\mathbf{x})) = - \iint \ln p(\theta|\mathbf{x}) p(\mathbf{x}, \theta) d\mathbf{x} d\theta$$

y debe ser menor que $H(\theta)$. De esta forma, una medida de la información ganada al observar los datos es

$$I = H(\theta) - H(\theta|\mathbf{x}).$$

Probar que $I \geq 0$. ¿Bajo que condiciones se cumple que $I = 0$? Sugerencia: expresar $H(\theta)$ como

$$H(\theta) = - \iint \ln p(\theta) p(\mathbf{x}, \theta) d\mathbf{x} d\theta.$$

También se necesitará la desigualdad

$$\int \ln \frac{p_1(\mathbf{u})}{p_2(\mathbf{u})} p_1(\mathbf{u}) d\mathbf{u} \geq 0$$

para las PDFs $p_1(\mathbf{u})$ y $p_2(\mathbf{u})$. La igualdad se da si y solo si $p_1(\mathbf{u}) = p_2(\mathbf{u})$.

Solución Como $\theta \sim \mathcal{N}(\mu_\theta, \sigma_\theta^2)$, su PDF es

$$p(\theta) = \frac{1}{\sqrt{2\pi\sigma_\theta^2}} \exp \left[-\frac{1}{2\sigma_\theta^2} (\theta - \mu_\theta)^2 \right],$$

y al tomar logaritmo, se tiene que

$$\ln p(\theta) = -\frac{1}{2} \ln 2\pi\sigma_\theta^2 - \frac{1}{2\sigma_\theta^2}(\theta - \mu_\theta)^2.$$

Por lo tanto, la entropía es

$$\begin{aligned} H(\theta) &= E(-\ln p(\theta)) \\ &= \frac{1}{2} \ln 2\pi\sigma_\theta^2 + \frac{1}{2\sigma_\theta^2} E[(\theta - \mu_\theta)^2] \\ &\stackrel{(a)}{=} \frac{1}{2} \ln 2\pi\sigma_\theta^2 + \frac{1}{2}, \end{aligned}$$

donde en (a) se empleó que $E[(\theta - \mu_\theta)^2] = \sigma_\theta^2$. Esto resulta en

$$H(\theta) = \frac{1}{2} (1 + \ln 2\pi\sigma_\theta^2).$$

Se observa que

$$\lim_{\sigma_\theta^2 \rightarrow 0} H(\theta) = -\infty \quad \text{y} \quad \lim_{\sigma_\theta^2 \rightarrow \infty} H(\theta) = \infty.$$

Se concluye que cuando la PDF de θ es mas concentrada (menor σ_θ^2) y la aleatoriedad es pequeña, la entropía es pequeña y viceversa. Para mostrar que $I \geq 0$, partiendo de la definición de la información, se tiene que

$$\begin{aligned} I &= H(\theta) - H(\theta|\mathbf{x}) \\ &\stackrel{(a)}{=} - \iint \ln p(\theta) p(\mathbf{x}, \theta) d\mathbf{x} d\theta + \iint \ln p(\theta|\mathbf{x}) p(\mathbf{x}, \theta) d\mathbf{x} d\theta \\ &= \iint [\ln p(\theta|\mathbf{x}) - \ln p(\theta)] p(\mathbf{x}, \theta) d\mathbf{x} d\theta \\ &\stackrel{(b)}{=} \iint \ln \frac{p(\theta|\mathbf{x})}{p(\theta)} p(\theta|\mathbf{x}) p(\mathbf{x}) d\mathbf{x} d\theta \\ &= \int \left[\int \ln \frac{p(\theta|\mathbf{x})}{p(\theta)} p(\theta|\mathbf{x}) d\theta \right] p(\mathbf{x}) d\mathbf{x} \\ &\stackrel{(c)}{\geq} 0, \end{aligned}$$

donde en (a) se expresó $H(\theta)$ como indica la sugerencia, que proviene del hecho que

$$p(\theta) = \int p(\mathbf{x}, \theta) d\mathbf{x},$$

en (b) se empleó el teorema de Bayes para expresar $p(\mathbf{x}, \theta) = p(\theta|\mathbf{x})p(\mathbf{x})$ y en (c) se consideró la desigualdad indicada en la sugerencia, que es la divergencia de Kullback-Leiber (ver la sección 7.1.2 por una demostración), lo cual indica que la integral entre paréntesis es no negativa para todo \mathbf{x} , y como además $p(\mathbf{x}) \geq 0$ para todo \mathbf{x} la integral exterior es no negativa. En la divergencia de Kullback-Leiber la igualdad a 0 se da si y solo si $p_1(\mathbf{u}) = p_2(\mathbf{u})$, que en este caso es

$$p(\theta|\mathbf{x}) = p(\theta).$$

que indica que el parámetro θ es independiente de los datos \mathbf{x} (ver el ejemplo 6-39 de [2]). De esta forma, observar los datos no aporta información sobre el parámetro y el incremento de la información es nula.

10.5.16. Problema 16

Para el ejemplo del nivel de DC en WGN con PDF a priori gaussiana de la sección 10.2.1 mostrar que la información ganada al observar los datos es

$$I = \frac{1}{2} \ln \left(1 + \frac{\sigma_A^2}{\sigma^2/N} \right).$$

Solución La información ganada al observar los datos es (ver el problema de la sección 10.5.15)

$$I = H(A) - H(A|\mathbf{x}) = E(-\ln p(A)) - E(-\ln p(A|\mathbf{x})).$$

Como la PDF a priori de A es $\mathcal{N}(\mu_A, \sigma_A^2)$,

$$\ln p(A) = -\frac{1}{2} \ln 2\pi - \frac{1}{2} \ln \sigma_A^2 - \frac{1}{2\sigma_A^2} (A - \mu_A)^2$$

por lo que la entropía de A es

$$\begin{aligned} H(A) &= E(-\ln p(A)) \\ &= \frac{1}{2} \ln 2\pi + \frac{1}{2} \ln \sigma_A^2 + \frac{1}{2\sigma_A^2} E[(A - \mu_A)^2] \\ &= \frac{1}{2} \ln 2\pi + \frac{1}{2} \ln \sigma_A^2 + \frac{1}{2} \end{aligned}$$

donde en la última igualdad se tuvo en cuenta que $E[(A - \mu_A)^2] = \sigma_A^2$. De forma análoga, considerando que la PDF a posteriori $p(A|\mathbf{x})$ es $\mathcal{N}(\mu_{A|x}, \sigma_{A|x}^2)$, donde $\mu_{A|x}$ y $\sigma_{A|x}^2$ están dados por las ecuaciones 10.7 y 10.9 respectivamente, se tiene que

$$H(A|\mathbf{x}) = \frac{1}{2} \ln 2\pi + \frac{1}{2} \ln \sigma_{A|x}^2 + \frac{1}{2}.$$

La ganancia de información es

$$\begin{aligned} I &= H(A) - H(A|\mathbf{x}) \\ &= \frac{1}{2} \ln \sigma_A^2 - \frac{1}{2} \ln \sigma_{A|x}^2 \\ &= \frac{1}{2} \ln \frac{\sigma_A^2}{\sigma_{A|x}^2} \\ &= \frac{1}{2} \ln \sigma_A^2 \left(\frac{N}{\sigma^2} + \frac{1}{\sigma_A^2} \right), \end{aligned}$$

donde en la última igualdad se empleó la ecuación 10.9 para sustituir $\sigma_{A|x}^2$. Se concluye que

$$I = \frac{1}{2} \ln \left(\frac{\sigma_A^2}{\sigma^2/N} + 1 \right),$$

que es lo que se quería mostrar.

10.5.17. Problema 17

En la elección de una PDF a priori que sea no informativa o no asuma ningún conocimiento a priori, un criterio empleado es hacerlo de forma de maximizar la información dada por los datos. De esta forma, los datos son el principal contribuyente del estado de conocimiento del parámetro desconocido. Empleando los resultados del problema de la sección 10.5.15, este enfoque puede ser implementado eligiendo $p(\theta)$ de forma que I sea máxima. Para la PDF a priori gaussiana del ejemplo de la sección 10.2.1, ¿cómo deben elegirse μ_A y σ_A^2 para que $p(A)$ sea no informativa?

Solución Del problema de la sección 10.5.16, se observa que I se maximiza cuando $\sigma_A^2 \rightarrow \infty$ independientemente del valor de μ_A . En ese caso, $\hat{A} \rightarrow \bar{x}$, y la elección de la PDF a priori no influye en el estimador. La elección de $\sigma_A^2 \rightarrow \infty$ modela el hecho de que no hay conocimiento a priori, como se señaló en el ejemplo de la sección 10.5.15.

Capítulo 11

Estimadores bayesianos generales

11.1. Funciones de costo

Previamente se definió el estimador MMSE como aquel que minimiza $E[(\theta - \hat{\theta})^2]$, donde la esperanza es respecto a $p(\mathbf{x}, \theta)$. Si $\epsilon = \theta - \hat{\theta}$ es el error para una realización particular de \mathbf{x} y θ , y $\mathcal{C}(\epsilon) = \epsilon^2$, el criterio del MSE minimiza $E[\mathcal{C}(\epsilon)]$. La función determinística $\mathcal{C}(\epsilon)$ se llama *función de costo*. En el criterio MSE, los errores grandes son particularmente costosos. El costo promedio $\mathcal{R} = E[\mathcal{C}(\epsilon)]$ se llama *costo de Bayes* y mide el desempeño del estimador. Otras posibles funciones de costo empleadas son

$$\mathcal{C}(\epsilon) = |\epsilon|,$$

con la cual los errores se penalizan de forma proporcional, y la función “acierto o fallo”,

$$\mathcal{C}(\epsilon) = \begin{cases} 0 & |\epsilon| < \delta \\ 1 & |\epsilon| > \delta, \end{cases}$$

con $\delta > 0$, la cual asigna costo nulo a errores pequeños menores a δ y costo 1 a los errores que superan el umbral δ .

Ya se vió que el estimador MMSE $\hat{\theta} = E(\theta|\mathbf{x})$ minimiza el costo de Bayes para una función de costo cuadrática. A continuación se determinará el estimador óptimo para las otras funciones de costo. El costo de Bayes es

$$\begin{aligned} \mathcal{R} &= E[\mathcal{C}(\epsilon)] \\ &= \iint \mathcal{C}(\theta - \hat{\theta}) p(\mathbf{x}, \theta) d\mathbf{x} d\theta \\ &= \iint \mathcal{C}(\theta - \hat{\theta}) p(\theta|\mathbf{x}) p(\mathbf{x}) d\mathbf{x} d\theta \\ &= \int \left[\int \mathcal{C}(\theta - \hat{\theta}) p(\theta|\mathbf{x}) d\theta \right] p(\mathbf{x}) d\mathbf{x}. \end{aligned} \tag{11.1}$$

Al igual que el razonamiento realizado en la sección 10.1, el costo se minimiza minimizando la integral interior para cada \mathbf{x} y considerando que con \mathbf{x} fijo, $\hat{\theta}$ es una variable escalar. Para la función de costo valor absoluto, la integral interior en 11.1 es

$$\begin{aligned} g(\hat{\theta}) &= \int |\theta - \hat{\theta}| p(\theta|\mathbf{x}) d\theta \\ &= \int_{-\infty}^{\hat{\theta}} (\hat{\theta} - \theta) p(\theta|\mathbf{x}) d\theta + \int_{\hat{\theta}}^{\infty} (\theta - \hat{\theta}) p(\theta|\mathbf{x}) d\theta \end{aligned}$$

donde en la última igualdad se consideró que

$$|\theta - \hat{\theta}| = \theta - \hat{\theta} \quad \text{si } \theta > \hat{\theta} \quad \text{y} \quad |\theta - \hat{\theta}| = \hat{\theta} - \theta \quad \text{si } \theta < \hat{\theta}.$$

Para encontrar el valor de $\hat{\theta}$ que la minimiza, hay que igualar la derivada respecto a $\hat{\theta}$ a cero. Para diferenciar respecto a $\hat{\theta}$ hay que emplear la regla de Leibnitz¹, que indica que

$$\frac{\partial}{\partial u} \int_{\phi_1(u)}^{\phi_2(u)} h(u, v) dv = \int_{\phi_1(u)}^{\phi_2(u)} \frac{\partial h(u, v)}{\partial u} dv + \frac{d\phi_2(u)}{du} h(u, \phi_2(u)) - \frac{d\phi_1(u)}{du} h(u, \phi_1(u)).$$

Aplicando la regla de Leibnitz en la primera integral se tiene que

$$\begin{aligned} \frac{\partial}{\partial \hat{\theta}} \int_{-\infty}^{\hat{\theta}} (\hat{\theta} - \theta) p(\theta|\mathbf{x}) d\theta &\stackrel{(a)}{=} \int_{-\infty}^{\hat{\theta}} \frac{\partial}{\partial \hat{\theta}} [(\hat{\theta} - \theta) p(\theta|\mathbf{x})] d\theta + \frac{d\hat{\theta}}{d\hat{\theta}} (\hat{\theta} - \hat{\theta}) p(\theta|\mathbf{x}) \\ &= \int_{-\infty}^{\hat{\theta}} p(\theta|\mathbf{x}) d\theta, \end{aligned}$$

donde en (a) se consideró que el límite inferior de integración no depende de $\hat{\theta}$, y realizando el razonamiento análogo para la segunda integral, se obtiene que

$$\frac{dg(\hat{\theta})}{d\hat{\theta}} = \int_{-\infty}^{\hat{\theta}} p(\theta|\mathbf{x}) d\theta - \int_{\hat{\theta}}^{\infty} p(\theta|\mathbf{x}) d\theta = 0.$$

por lo que la función se minimiza con el valor de $\hat{\theta}$ que cumple que

$$\int_{-\infty}^{\hat{\theta}} p(\theta|\mathbf{x}) d\theta = \int_{\hat{\theta}}^{\infty} p(\theta|\mathbf{x}) d\theta.$$

Por definición, $\hat{\theta}$ es la mediana de la PDF a posteriori, o el valor que cumple que $\Pr\{\theta \leq \hat{\theta}|\mathbf{x}\} = 1/2$.

Para la función de costo “acierto o fallo” se tiene que $\mathcal{C}(\epsilon) = 1$ para

$$\epsilon > \delta \quad \text{y} \quad \epsilon < -\delta \quad \Rightarrow \quad \theta - \hat{\theta} > \delta \quad \text{y} \quad \theta - \hat{\theta} < -\delta \quad \Rightarrow \quad \theta > \hat{\theta} + \delta \quad \text{y} \quad \theta < \hat{\theta} - \delta,$$

por lo que la integral interior en la ecuación 11.1 queda

$$g(\hat{\theta}) = \int_{-\infty}^{\hat{\theta}-\delta} p(\theta|\mathbf{x}) d\theta + \int_{\hat{\theta}+\delta}^{\infty} p(\theta|\mathbf{x}) d\theta.$$

Considerando que

$$\int_{-\infty}^{\infty} p(\theta|\mathbf{x}) d\theta = 1,$$

se obtiene que

$$g(\hat{\theta}) = 1 - \int_{\hat{\theta}-\delta}^{\hat{\theta}+\delta} p(\theta|\mathbf{x}) d\theta,$$

que se minimiza maximizando

$$\int_{\hat{\theta}-\delta}^{\hat{\theta}+\delta} p(\theta|\mathbf{x}) d\theta.$$

Para δ arbitrariamente pequeño la integral se maximiza eligiendo el valor $\hat{\theta}$ correspondiente a la ubicación del máximo de $p(\theta|\mathbf{x})$. El estimador que minimiza el costo de Bayes para la función de costo “acierto o fallo” es la *moda* (ubicación del máximo) de la PDF a posteriori. Este estimador se llama estimador *máximo a posteriori* (MAP).

En resumen, los estimadores que minimizan el costo de Bayes para las funciones de costo cuadrática, valor absoluto y “acierto o fallo” son respectivamente la media, la mediana y la moda de la PDF a posteriori. Para algunas PDF a posteriori, los tres estimadores son idénticos, como por ejemplo, para la PDF a posteriori gaussiana. Ver además el problema de la sección 11.6.2.

¹Ver https://en.wikipedia.org/wiki/Leibniz_integral_rule

11.2. Estimadores de error cuadrático medio mínimo

En el capítulo 10 se dedujo que el estimador MMSE es $E(\theta|\mathbf{x})$. A continuación se generaliza el resultado para el caso en que el parámetro es vectorial y se analizan algunas propiedades.

Si $\boldsymbol{\theta}$ es un parámetro vectorial de dimensión $p \times 1$, para estimar por ejemplo θ_1 se considera el resto de los parámetros como *parámetros molestos*. Si $p(\mathbf{x}|\boldsymbol{\theta})$ es la PDF condicional de los datos y $p(\boldsymbol{\theta})$ la PDF a priori del parámetro vectorial, la PDF a posteriori de θ_1 puede obtenerse como

$$p(\theta_1|\mathbf{x}) = \int \cdots \int p(\boldsymbol{\theta}|\mathbf{x}) d\theta_2 \cdots d\theta_p \quad (11.2)$$

donde

$$p(\boldsymbol{\theta}|\mathbf{x}) = \frac{p(\mathbf{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int p(\mathbf{x}|\boldsymbol{\theta})p(\boldsymbol{\theta}) d\boldsymbol{\theta}}.$$

Empleando el mismo razonamiento que en el capítulo 10, el estimador de θ_1 es

$$\begin{aligned} \hat{\theta}_1 &= E(\theta_1|\mathbf{x}) \\ &= \int \theta_1 p(\theta_1|\mathbf{x}) d\theta_1, \end{aligned}$$

o en general

$$\hat{\theta}_i = \int \theta_i p(\theta_i|\mathbf{x}) d\theta_i, \quad i = 1, \dots, p.$$

Este es el estimador MMSE que minimiza

$$E[(\theta_i - \hat{\theta}_i)^2] = \int (\theta_i - \hat{\theta}_i)^2 p(\theta_i, \mathbf{x}) d\mathbf{x} d\theta_i,$$

que es el error cuadrático promediado respecto a la PDF marginal $p(\mathbf{x}, \theta_i)$. Se observa que el estimador MMSE para el caso vectorial no implica nada nuevo excepto por la necesidad de calcular la PDF a posteriori de cada parámetro. El estimador MMSE se puede expresar de forma vectorial de la siguiente forma. El estimador del primer parámetro se puede expresar como

$$\begin{aligned} \hat{\theta}_1 &= \int \theta_1 p(\theta_1|\mathbf{x}) d\theta_1 \\ &\stackrel{(a)}{=} \int \theta_1 \left[\int \cdots \int p(\boldsymbol{\theta}|\mathbf{x}) d\theta_2 \cdots d\theta_p \right] d\theta_1 \\ &= \int \cdots \int \theta_1 p(\boldsymbol{\theta}|\mathbf{x}) d\theta_1 \cdots d\theta_p \\ &= \int \theta_1 p(\boldsymbol{\theta}|\mathbf{x}) d\boldsymbol{\theta}, \end{aligned}$$

donde en (a) se empleó la ecuación 11.2, o en general,

$$\hat{\theta}_i = \int \theta_i p(\boldsymbol{\theta}|\mathbf{x}) d\boldsymbol{\theta}, \quad i = 1, \dots, p.$$

En notación vectorial, resulta en

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= \begin{bmatrix} \int \theta_1 p(\boldsymbol{\theta}|\mathbf{x}) d\boldsymbol{\theta} \\ \vdots \\ \int \theta_p p(\boldsymbol{\theta}|\mathbf{x}) d\boldsymbol{\theta} \end{bmatrix} \\ &= \int \boldsymbol{\theta} p(\boldsymbol{\theta}|\mathbf{x}) d\boldsymbol{\theta} \\ &= E(\boldsymbol{\theta}|\mathbf{x}). \end{aligned}$$

Es importante notar que por la deducción, el estimador MMSE $E(\boldsymbol{\theta}|\mathbf{x})$ minimiza el MSE para cada componente del parámetro vectorial desconocido, o $[\hat{\boldsymbol{\theta}}]_i = [E(\boldsymbol{\theta}|\mathbf{x})]_i$ minimiza $E[(\theta_i - \hat{\theta}_i)^2]$.

En el caso escalar, se obtuvo que el MSE bayesiano mínimo es (ver la ecuación 10.8)

$$\begin{aligned} \text{Bmse}(\hat{\theta}_1) &= \int \left[\int (\theta_1 - E(\theta_1|\mathbf{x}))^2 p(\theta_1|\mathbf{x}) d\theta_1 \right] p(\mathbf{x}) d\mathbf{x} \\ &= \int \text{var}(\theta_1|\mathbf{x}) p(\mathbf{x}) d\mathbf{x}. \end{aligned}$$

Empleando la ecuación 11.2 de la PDF a posteriori de θ_1 , se tiene que

$$\begin{aligned} \text{Bmse}(\hat{\theta}_1) &= \int \left\{ \int (\theta_1 - E(\theta_1|\mathbf{x}))^2 \left[\int \cdots \int p(\boldsymbol{\theta}|\mathbf{x}) d\theta_2 \cdots d\theta_p \right] d\theta_1 \right\} p(\mathbf{x}) d\mathbf{x} \\ &= \int \left[\int \cdots \int (\theta_1 - E(\theta_1|\mathbf{x}))^2 p(\boldsymbol{\theta}|\mathbf{x}) d\theta_1 \cdots d\theta_p \right] p(\mathbf{x}) d\mathbf{x} \\ &= \int \left[\int (\theta_1 - E(\theta_1|\mathbf{x}))^2 p(\boldsymbol{\theta}|\mathbf{x}) d\boldsymbol{\theta} \right] p(\mathbf{x}) d\mathbf{x}. \end{aligned}$$

La integral interior es la varianza de θ_1 para la PDF a posteriori $p(\boldsymbol{\theta}|\mathbf{x})$, y es el elemento (1, 1) de la matriz de covarianza $\mathbf{C}_{\theta|x}$ de la PDF a posteriori. Por lo, tanto, en el caso general, el MSE bayesiano mínimo es

$$\text{Bmse}(\hat{\theta}_i) = \int [\mathbf{C}_{\theta|x}]_{ii} p(\mathbf{x}) d\mathbf{x}, \quad i = 1, \dots, p \quad (11.3)$$

donde

$$\mathbf{C}_{\theta|x} = E_{\theta|x}[(\boldsymbol{\theta} - E(\boldsymbol{\theta}|\mathbf{x}))(\boldsymbol{\theta} - E(\boldsymbol{\theta}|\mathbf{x}))^T].$$

11.2.1. Análisis de Fourier bayesiano

Se considera el ejemplo del análisis de Fourier de la sección 4.4.2, pero para simplificar los cálculos se asume que $M = 1$. De esta forma, el modelo de los datos es

$$x[n] = a \cos 2\pi f_0 n + b \sin 2\pi f_0 n + w[n], \quad n = 0, \dots, N-1,$$

donde f_0 es un múltiplo de $1/N$, excepto 0 o $1/2$, casos en que $\sin 2\pi f_0 n = 0$, y $w[n]$ es WGN con varianza σ^2 . Se desea estimar $\boldsymbol{\theta} = [a \ b]^T$. Como desviación del modelo clásico, se asume ahora que a y b son variables aleatorias con PDF a priori

$$\boldsymbol{\theta} \sim (\mathbf{0}, \sigma_{\theta}^2 \mathbf{I})$$

y $\boldsymbol{\theta}$ es independiente de $w[n]$. Este tipo de modelo es referido como *sinusoide con desvanecimiento Rayleigh*, y es usado frecuentemente para representar una sinusoide que se propaga por un medio con atenuación (ver además el problema de la sección 11.6.6). Expresando el modelo en forma vectorial como

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w},$$

donde

$$\mathbf{H} = \begin{bmatrix} 1 & 0 \\ \cos 2\pi f_0 & \sin 2\pi f_0 \\ \vdots & \vdots \\ \cos 2\pi f_0(N-1) & \sin 2\pi f_0(N-1) \end{bmatrix},$$

se reconoce como el modelo lineal bayesiano de la sección 10.4. Del teorema de dicha sección puede obtenerse la media y la covarianza de la PDF a posteriori, dadas por las ecuaciones 10.19 y 10.20 respectivamente, las cuales indican que

$$\hat{\boldsymbol{\theta}} = E(\boldsymbol{\theta}|\mathbf{x}) = \boldsymbol{\mu}_{\theta} + (\mathbf{C}_{\theta}^{-1} + \mathbf{H}^T \mathbf{C}_w^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C}_w^{-1} (\mathbf{x} - \mathbf{H}\boldsymbol{\mu}_{\theta}) \quad \text{y} \quad \mathbf{C}_{\theta|x} = (\mathbf{C}_{\theta}^{-1} + \mathbf{H}^T \mathbf{C}_w^{-1} \mathbf{H})^{-1},$$

y con

$$\boldsymbol{\mu}_{\theta} = \mathbf{0}, \quad \mathbf{C}_{\theta} = \sigma_{\theta}^2 \mathbf{I}, \quad \mathbf{C}_w = \sigma^2 \mathbf{I},$$

se obtiene que

$$\hat{\boldsymbol{\theta}} = \left(\frac{1}{\sigma_{\theta}^2} \mathbf{I} + \frac{1}{\sigma^2} \mathbf{H}^T \mathbf{H} \right)^{-1} \frac{1}{\sigma^2} \mathbf{H}^T \mathbf{x} \quad \text{y} \quad \mathbf{C}_{\theta|x} = \left(\frac{1}{\sigma_{\theta}^2} \mathbf{I} + \frac{1}{\sigma^2} \mathbf{H}^T \mathbf{H} \right)^{-1}.$$

Considerando que las columnas de \mathbf{H} son ortogonales, como se mostró en el ejemplo de la sección 4.4.2,

$$\mathbf{H}^T \mathbf{H} = \frac{N}{2} \mathbf{I}$$

el se obtiene que

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= \left(\frac{1}{\sigma_{\theta}^2} \mathbf{I} + \frac{N}{2\sigma^2} \mathbf{I} \right)^{-1} \mathbf{H}^T \mathbf{x} \\ &= \frac{\frac{1}{\sigma^2}}{\frac{1}{\sigma_{\theta}^2} + \frac{N}{2\sigma^2}} \mathbf{H}^T \mathbf{x}, \end{aligned}$$

y teniendo en cuenta que

$$\mathbf{H}^T \mathbf{x} = \begin{bmatrix} \sum_{n=0}^{N-1} x[n] \cos 2\pi f_0 n \\ \sum_{n=0}^{N-1} x[n] \sin 2\pi f_0 n \end{bmatrix}$$

se concluye que el estimador MMSE es

$$\begin{aligned} \hat{a} &= \frac{1}{1 + \frac{2\sigma^2/N}{\sigma_{\theta}^2}} \left[\frac{2}{N} \sum_{n=0}^{N-1} x[n] \cos 2\pi f_0 n \right] \\ \hat{b} &= \frac{1}{1 + \frac{2\sigma^2/N}{\sigma_{\theta}^2}} \left[\frac{2}{N} \sum_{n=0}^{N-1} x[n] \sin 2\pi f_0 n \right]. \end{aligned}$$

Comparando con los resultados del ejemplo 4.4.2, la única diferencia con el caso clásico es el factor de escala, y si $\sigma_{\theta}^2 \gg 2\sigma^2/N$, que corresponde a que el conocimiento a priori es escaso en comparación al conocimiento de los datos, los resultados son idénticos. La matriz de covarianza a posteriori es

$$\mathbf{C}_{\theta|x} = \frac{\frac{1}{\sigma^2}}{\frac{1}{\sigma_{\theta}^2} + \frac{N}{2\sigma^2}} \mathbf{I},$$

y como no depende de \mathbf{x} , de la ecuación 11.3 se obtiene que el MSE bayesiano mínimo es

$$\begin{aligned} \text{Bmse}(\hat{a}) &= \frac{1}{\frac{1}{\sigma_{\theta}^2} + \frac{1}{2\sigma^2/N}} \\ \text{Bmse}(\hat{b}) &= \frac{1}{\frac{1}{\sigma_{\theta}^2} + \frac{1}{2\sigma^2/N}}. \end{aligned}$$

Es interesante notar que en ausencia de conocimiento a priori en el modelo lineal bayesiano, el estimador MMSE conduce al estimador MVU del modelo lineal clásico. Para verificarlo, de la ecuación 10.19, el estimador MMSE es

$$\hat{\boldsymbol{\theta}} = E(\boldsymbol{\theta}|\mathbf{x}) = \boldsymbol{\mu}_{\theta} + (\mathbf{C}_{\theta}^{-1} + \mathbf{H}^T \mathbf{C}_w^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C}_w^{-1} (\mathbf{x} - \mathbf{H} \boldsymbol{\mu}_{\theta}).$$

Si no hay conocimiento a priori, $\mathbf{C}_\theta^{-1} \rightarrow \mathbf{0}$ y por lo tanto

$$\begin{aligned}\hat{\boldsymbol{\theta}} &\rightarrow \boldsymbol{\mu}_\theta + (\mathbf{H}^T \mathbf{C}_w^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C}_w^{-1} (\mathbf{x} - \mathbf{H} \boldsymbol{\mu}_\theta) \\ &= \boldsymbol{\mu}_\theta + (\mathbf{H}^T \mathbf{C}_w^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C}_w^{-1} \mathbf{x} - (\mathbf{H}^T \mathbf{C}_w^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C}_w^{-1} \mathbf{H} \boldsymbol{\mu}_\theta \\ &= \boldsymbol{\mu}_\theta + (\mathbf{H}^T \mathbf{C}_w^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C}_w^{-1} \mathbf{x} - \boldsymbol{\mu}_\theta \\ &= (\mathbf{H}^T \mathbf{C}_w^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C}_w^{-1} \mathbf{x},\end{aligned}$$

que es el estimador MVU del modelo lineal (ecuación 4.20). Sin embargo, esto ocurre por diversas circunstancias fortuitas, como se muestra en el problema de la sección 11.6.7.

El estimador MMSE tiene varias propiedades útiles que serán empleadas mas adelante. En primer lugar, conmuta en transformaciones lineales (en realidad, afines). Asúmase que se quiere estimar $\boldsymbol{\alpha}$ con

$$\boldsymbol{\alpha} = \mathbf{A} \boldsymbol{\theta} + \mathbf{b},$$

donde \mathbf{A} es una matriz $r \times p$ y \mathbf{b} es un vector $r \times 1$. De esta forma, $\boldsymbol{\alpha}$ es un vector aleatorio para el cual el estimador MMSE es

$$\hat{\boldsymbol{\alpha}} = E(\boldsymbol{\alpha}|\mathbf{x}).$$

Debido a la linealidad de la operación de esperanza, se cumple que

$$\begin{aligned}\hat{\boldsymbol{\alpha}} &= E(\mathbf{A} \boldsymbol{\theta} + \mathbf{b}|\mathbf{x}) \\ &= \mathbf{A} E(\boldsymbol{\theta}|\mathbf{x}) + \mathbf{b} \\ &= \mathbf{A} \hat{\boldsymbol{\theta}} + \mathbf{b}.\end{aligned}$$

Ver además los problemas de las secciones 11.6.8 y 11.6.9. Una segunda propiedad importante está relacionada al estimador MMSE basado en dos vectores de datos \mathbf{x}_1 y \mathbf{x}_2 . Asúmase que $\boldsymbol{\theta}$, \mathbf{x}_1 y \mathbf{x}_2 son conjuntamente gaussianos y los vectores de datos son independientes. El estimador MMSE es

$$\hat{\boldsymbol{\theta}} = E(\boldsymbol{\theta}|\mathbf{x}_1, \mathbf{x}_2).$$

Siendo $\mathbf{x} = [\mathbf{x}_1^T \mathbf{x}_2^T]^T$, de la ecuación 10.14 se tiene que

$$\hat{\boldsymbol{\theta}} = E(\boldsymbol{\theta}|\mathbf{x}) = E(\boldsymbol{\theta}) + \mathbf{C}_{\theta x} \mathbf{C}_{xx}^{-1} (\mathbf{x} - E(\mathbf{x})).$$

La matriz de covarianza de los datos es

$$\mathbf{C}_{xx}^{-1} = \begin{bmatrix} \mathbf{C}_{x_1 x_1} & \mathbf{C}_{x_1 x_2} \\ \mathbf{C}_{x_2 x_1} & \mathbf{C}_{x_2 x_2} \end{bmatrix}^{-1} \stackrel{(a)}{=} \begin{bmatrix} \mathbf{C}_{x_1 x_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_{x_2 x_2} \end{bmatrix}^{-1} \stackrel{(b)}{=} \begin{bmatrix} \mathbf{C}_{x_1 x_1}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_{x_2 x_2}^{-1} \end{bmatrix}$$

donde en (a) se consideró que \mathbf{x}_1 y \mathbf{x}_2 son independientes y en (b) se empleó el resultado del apéndice G de la inversa de una matriz diagonal a bloques. Además

$$\mathbf{C}_{\theta x} = E(\boldsymbol{\theta} [\mathbf{x}_1^T \mathbf{x}_2^T]) = [E(\boldsymbol{\theta} \mathbf{x}_1^T) \quad E(\boldsymbol{\theta} \mathbf{x}_2^T)] = [\mathbf{C}_{\theta x_1} \quad \mathbf{C}_{\theta x_2}].$$

Combinando los resultados, se obtiene que

$$\begin{aligned}\hat{\boldsymbol{\theta}} &= E(\boldsymbol{\theta}) + [\mathbf{C}_{\theta x_1} \quad \mathbf{C}_{\theta x_2}] \begin{bmatrix} \mathbf{C}_{x_1 x_1}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_{x_2 x_2}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 - E(\mathbf{x}_1) \\ \mathbf{x}_2 - E(\mathbf{x}_2) \end{bmatrix} \\ &= E(\boldsymbol{\theta}) + \mathbf{C}_{\theta x_1} \mathbf{C}_{x_1 x_1}^{-1} (\mathbf{x}_1 - E(\mathbf{x}_1)) + \mathbf{C}_{\theta x_2} \mathbf{C}_{x_2 x_2}^{-1} (\mathbf{x}_2 - E(\mathbf{x}_2)).\end{aligned}$$

El estimador está compuesto del término del estimador a priori $E(\boldsymbol{\theta})$ así como los términos correspondientes a los conjuntos de datos independientes. Observar que si $E(\boldsymbol{\theta}) = 0$, se cumple que

$$\begin{aligned}\hat{\boldsymbol{\theta}} &= \mathbf{C}_{\theta x_1} \mathbf{C}_{x_1 x_1}^{-1} (\mathbf{x}_1 - E(\mathbf{x}_1)) + \mathbf{C}_{\theta x_2} \mathbf{C}_{x_2 x_2}^{-1} (\mathbf{x}_2 - E(\mathbf{x}_2)) \\ &= E(\boldsymbol{\theta}|\mathbf{x}_1) + E(\boldsymbol{\theta}|\mathbf{x}_2),\end{aligned}$$

o

$$\begin{aligned}\hat{\boldsymbol{\theta}} &= E(\boldsymbol{\theta}|\mathbf{x}_1, \mathbf{x}_2) \\ &= E(\boldsymbol{\theta}|\mathbf{x}_1) + E(\boldsymbol{\theta}|\mathbf{x}_2).\end{aligned} \tag{11.4}$$

Esto indica que el estimador MMSE cumple una propiedad de aditividad para conjuntos de datos independientes.

11.3. Estimadores máximo a posteriori

En el enfoque de estimación MAP se elige como estimador $\hat{\theta}$ el valor que maximiza la PDF a posteriori, o

$$\hat{\theta} = \arg \max_{\theta} p(\theta|\mathbf{x}).$$

Como se vio previamente, este estimador minimiza la función de costo “acierto o fallo”. Para encontrar el máximo de $p(\theta|\mathbf{x})$ se observa que

$$p(\theta|\mathbf{x}) = \frac{p(\mathbf{x}|\theta)p(\theta)}{p(\mathbf{x})}$$

y como el denominador no depende de θ , es equivalente la maximización del numerador $p(\mathbf{x}|\theta)p(\theta)$. Esto es reminiscente al estimador MLE, que maximiza $p(\mathbf{x}; \theta)$, excepto por la presencia de la PDF a priori. Por lo tanto, el estimador MAP es

$$\hat{\theta} = \arg \max_{\theta} p(\mathbf{x}|\theta)p(\theta),$$

o equivalentemente,

$$\hat{\theta} = \arg \max_{\theta} [\ln p(\mathbf{x}|\theta) + \ln p(\theta)].$$

Observar que el estimador MAP es usualmente mas fácil de determinar debido a que solo involucra una maximización y no una integración. El estimador MMSE requiere una integración para calcular la esperanza.

11.3.1. Ejemplo: PDF exponencial

Asúmase que

$$p(x[n]|\theta) = \begin{cases} \theta \exp(-\theta x[n]) & x[n] > 0 \\ 0 & x[n] < 0 \end{cases}$$

donde los $x[n]$ son condicionalmente IID, o

$$p(\mathbf{x}|\theta) = \prod_{n=0}^{N-1} p(x[n]|\theta)$$

y la PDF a priori es

$$p(\theta) = \begin{cases} \lambda \exp(-\lambda \theta) & \theta > 0 \\ 0 & \theta < 0. \end{cases}$$

El estimador MAP se obtiene maximizando

$$\begin{aligned} g(\theta) &= \ln p(\mathbf{x}|\theta) + \ln p(\theta) \\ &= \ln \left[\theta^N \exp \left(-\theta \sum_{n=0}^{N-1} x[n] \right) \right] + \ln [\lambda \exp(-\lambda \theta)] \\ &= N \ln \theta - N \theta \bar{x} + \ln \lambda - \lambda \theta \end{aligned}$$

para $\theta > 0$. La diferenciación respecto a θ produce

$$\frac{dg(\theta)}{d\theta} = \frac{N}{\theta} - N \bar{x} - \lambda$$

e igualando a cero se obtiene que el estimador MAP es

$$\hat{\theta} = \frac{1}{\bar{x} + \frac{\lambda}{N}}.$$

Notar que con $N \rightarrow \infty$, $\hat{\theta} \rightarrow 1/\bar{x}$. Recordar que $E(x[n]|\theta) = 1/\theta$ (ver el ejemplo de la sección 9.2.1), por lo que

$$\theta = \frac{1}{E(x[n]|\theta)},$$

por lo que el estimador MAP es razonable. Por otro lado, si $\lambda \rightarrow 0$ de forma tal que el PDF a priori es casi uniforme, se obtiene el estimador $1/\bar{x}$. De hecho, ese es el estimador MLE bayesiano, que es el estimador que se obtiene al maximizar $p(\mathbf{x}|\theta)$, ya que cuando $\lambda \rightarrow 0$ se tiene la situación de la figura 11.1 en donde la PDF condicional domina a la PDF a priori, y el máximo de g no es afectado por la PDF a priori.

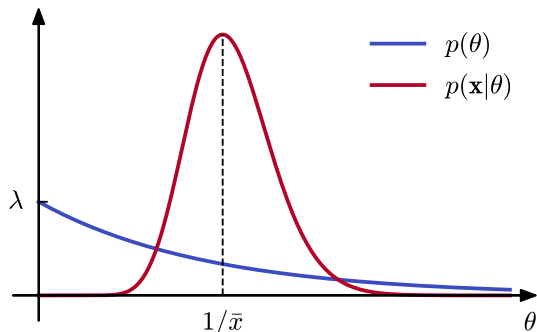


Figura 11.1: PDF a priori y PDF condicional de los datos. La PDF condicional domina a la PDF a priori si N es grande o si $\lambda \rightarrow 0$. En la figura, $N = 20$, $\lambda = 0.4$ y los datos se generaron con $\theta = 2$.

Para extender el estimador MAP al caso vectorial en el cual la PDF a posteriori es ahora $p(\theta|\mathbf{x})$, nuevamente se emplea que

$$p(\theta_1|\mathbf{x}) = \int \cdots \int p(\theta|\mathbf{x}) d\theta_2 \cdots d\theta_p,$$

y el MAP está dado por

$$\hat{\theta}_1 = \arg \max_{\theta_1} p(\theta_1|\mathbf{x}),$$

o en general

$$\hat{\theta}_i = \arg \max_{\theta_i} p(\theta_i|\mathbf{x}), \quad i = 1, \dots, p. \quad (11.5)$$

Este estimador minimiza la media de la función de costo “acierto o fallo”

$$\mathcal{R}_i = E[\mathcal{C}(\theta_i - \hat{\theta}_i)]$$

para cada i , donde la esperanza se toma sobre $p(\mathbf{x}, \theta_i)$. La ventaja de la no necesidad de integración del estimador MAP para el caso escalar se pierde en el caso vectorial debido al requerimiento del cálculo de las PDF marginales $p(\theta_i|\mathbf{x})$ mediante la ecuación 11.2. Sin embargo, puede proponerse el siguiente estimador MAP *vectorial*

$$\hat{\theta} = \arg \max_{\theta} p(\theta|\mathbf{x}),$$

en el cual la PDF a posteriori para el parámetro vectorial θ es maximizada para encontrar el estimador. De esta forma, se evita la integración para calcular las PDF marginales, ya que equivalentemente,

$$\hat{\theta} = \arg \max_{\theta} p(\mathbf{x}|\theta)p(\theta).$$

Este estimador en general no es el mismo que el dado por la ecuación 11.5. Para ver esto, considérese el ejemplo ilustrado en la figura 11.2, en el cual la PDF a posteriori del parámetro vectorial $p(\theta_1, \theta_2|\mathbf{x})$ es constante e igual a $1/3$ en la región $\{2 < \theta_1 < 3, 0 < \theta_2 < 1\}$ y constante e igual a $1/6$ en las regiones $\{0 < \theta_1 < 2, 1 < \theta_2 < 2\}$ y $\{3 < \theta_1 < 5, 1 < \theta_2 < 2\}$. Por lo tanto, el estimado MAP vectorial de θ_2 es cualquier valor $0 < \hat{\theta}_2 < 1$ donde la PDF es $1/3$. El estimador MAP de θ_2 se obtiene maximizando

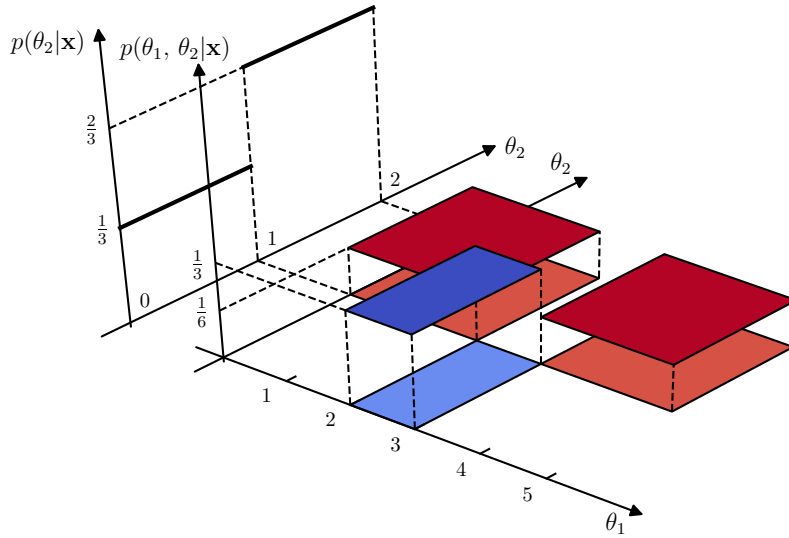


Figura 11.2: PDF a posteriori $p(\theta_1, \theta_2 | \mathbf{x})$ del parámetro vectorial $\boldsymbol{\theta} = [\theta_1 \theta_2]^T$ y PDF a posteriori marginal $p(\theta_2 | \mathbf{x})$. $p(\theta_1, \theta_2 | \mathbf{x})$ del parámetro vectorial es constante e igual a $1/3$ en la región en azul y constante e igual a $1/6$ en las regiones en rojo. El máximo de $p(\theta_1, \theta_2 | \mathbf{x})$ en θ_2 es cualquier valor $0 < \theta_2 < 1$, mientras que el máximo de $p(\theta_2 | \mathbf{x})$ en θ_2 es cualquier valor $1 < \theta_2 < 2$.

$p(\theta_2 | \mathbf{x})$. Como

$$\begin{aligned}
 p(\theta_2 | \mathbf{x}) &= \int p(\theta_1, \theta_2 | \mathbf{x}) d\theta_1 \\
 &= \begin{cases} \int_2^3 \frac{1}{3} d\theta_1, & 0 < \theta_2 < 1 \\ \int_0^2 \frac{1}{6} d\theta_1 + \int_3^5 \frac{1}{6} d\theta_1, & 1 < \theta_2 < 2 \end{cases} \\
 &= \begin{cases} \frac{1}{3}, & 0 < \theta_2 < 1 \\ \frac{2}{3}, & 1 < \theta_2 < 2, \end{cases}
 \end{aligned}$$

que también se muestra en la figura 11.2, el estimador MAP es cualquier valor en el intervalo $1 < \hat{\theta}_2 < 2$, y difiere del estimador MAP vectorial. Puede mostrarse que el estimador MAP vectorial efectivamente minimiza el costo de Bayes, pero uno distinto al correspondiente a la función de costo de “acierto o fallo” (ver el problema de la sección 11.6.11). El estimador MAP vectorial será referido solamente como estimador MAP.

Se llama estimador MLE bayesiano al valor de $\boldsymbol{\theta}$ que maximiza $p(\mathbf{x} | \boldsymbol{\theta})$. Es cierto en general que cuando $N \rightarrow \infty$, el estimador MAP tiende al estimador MLE Bayesiano. Para obtener el estimador MAP hay que maximizar $p(\mathbf{x} | \boldsymbol{\theta})p(\boldsymbol{\theta})$. Si la PDF a priori $p(\boldsymbol{\theta})$ es uniforme en el intervalo de $\boldsymbol{\theta}$ en el cual $p(\mathbf{x} | \boldsymbol{\theta})$ es esencialmente no nula, lo cual es cierto si $N \rightarrow \infty$ ya que la PDF de los datos domina sobre la PDF a priori, la maximización será equivalente a la maximización de $p(\mathbf{x} | \boldsymbol{\theta})$. Si $p(\mathbf{x} | \boldsymbol{\theta})$ tiene la misma

forma que la familia de PDFs $p(\mathbf{x}; \boldsymbol{\theta})$, el MLE bayesiano y el MLE clásico tendrán la misma forma (ver el problema de la sección 11.6.7).

Se mencionarán algunas propiedades del estimador MAP. Primero, si la PDF a posteriori es gaussiana, como por ejemplo en el caso del modelo lineal bayesiano, la moda o la ubicación del pico coincide con la media. Por lo tanto, el estimador MAP es idéntico al estimador MMSE si \mathbf{x} y $\boldsymbol{\theta}$ son conjuntamente gaussianas. Segundo, la propiedad de invarianza del estimador MLE (ver la sección 7.1.3) no se cumple para el estimador MAP. Sin embargo, si se cumple en el caso de que la transformación sea lineal, como se muestra en el problema de la sección 11.6.12.

11.4. Descripción del desempeño

En esta sección se describe la técnica para evaluar el desempeño de un estimador bayesiano, que debido a la naturaleza aleatoria del parámetro desconocido, difiere del caso clásico.

11.4.1. Caso escalar

En el problema clásico de estimación, interesa la media y la varianza de un estimador. De esta forma, si el estimador es gaussiano, se obtiene inmediatamente su PDF. Si la PDF está concentrada en torno al valor verdadero del parámetro, puede afirmarse que el estimador se comporta bien. En el caso de un parámetro aleatorio no puede aplicarse el mismo enfoque. La aleatoriedad del parámetro resulta en una PDF distinta para cada realización de θ . Esta PDF se denota $p(\hat{\theta}|\theta)$. Para buen desempeño, el estimador debe ser cercano a θ para cada valor posible de θ , o el error

$$\epsilon = \theta - \hat{\theta}$$

debe ser pequeño. Este razonamiento conduce al estimador MMSE, que minimiza $E_{x, \theta}[(\theta - \hat{\theta})^2]$. De esta forma, para un estimador bayesiano arbitrario, tiene sentido evaluar su desempeño mediante la PDF del error. Esta PDF, que ahora tiene en cuenta la aleatoriedad de θ , debe estar concentrada en torno a cero. Para el estimador MMSE $\hat{\theta} = E(\theta|\mathbf{x})$, el error es

$$\epsilon = \theta - E(\theta|\mathbf{x}),$$

y la media del error es

$$\begin{aligned} E_{x, \theta}(\epsilon) &= \iint [\theta - E(\theta|\mathbf{x})] p(\mathbf{x}, \theta) d\mathbf{x} d\theta \\ &= \iint [\theta - E(\theta|\mathbf{x})] p(\theta|\mathbf{x}) p(\mathbf{x}) d\mathbf{x} d\theta \\ &= \int \left[\int [\theta - E(\theta|\mathbf{x})] p(\theta|\mathbf{x}) d\theta \right] p(\mathbf{x}) d\mathbf{x} \\ &= \int \left[\int \theta p(\theta|\mathbf{x}) d\theta - E(\theta|\mathbf{x}) \right] p(\mathbf{x}) d\mathbf{x} \\ &= \int [E(\theta|\mathbf{x}) - E(\theta|\mathbf{x})] p(\mathbf{x}) d\mathbf{x} \\ &= 0, \end{aligned}$$

es decir, el error del estimador MMSE es cero en promedio. Teniendo en cuenta que $E(\epsilon) = 0$, la varianza del error del estimador MMSE es

$$\text{var}(\epsilon) = E_{x, \theta}(\epsilon^2),$$

y por lo tanto,

$$\text{var}(\epsilon) = E_{x, \theta}[(\theta - \hat{\theta})^2],$$

que es justamente el MSE bayesiano mínimo (ver la ecuación 10.1). Por último, si ϵ es gaussiano, se tiene que

$$\epsilon \sim \mathcal{N}(0, \text{Bmse}(\hat{\theta})). \quad (11.6)$$

Ejemplo: nivel de DC en WGN con PDF a priori gaussiana

En el ejemplo de la sección 10.2.1 de la estimación del nivel de DC en WGN con PDF a priori gaussiana se obtuvo que el estimador MMSE es (ecuación 10.7)

$$\hat{A} = \frac{\sigma_A^2}{\sigma_A^2 + \frac{\sigma^2}{N}} \bar{x} + \frac{\frac{\sigma^2}{N}}{\sigma_A^2 + \frac{\sigma^2}{N}} \mu_A$$

y el MSE bayesiano mínimo es (ecuación 10.9)

$$\text{Bmse}(\hat{A}) = \frac{1}{\frac{N}{\sigma^2} + \frac{1}{\sigma_A^2}}.$$

En este caso, el error es $\epsilon = A - \hat{A}$. Como \hat{A} depende linealmente de \mathbf{x} y \mathbf{x} y A son conjuntamente gaussianas, ϵ es gaussiana. Como resultado, de la ecuación 11.6 se tiene que

$$\epsilon \sim \mathcal{N}\left(0, \frac{1}{\frac{N}{\sigma^2} + \frac{1}{\sigma_A^2}}\right).$$

Con $N \rightarrow \infty$, la PDF se concentra cada vez mas en torno a cero y se dice que el estimador es *consistente en el sentido bayesiano* (ver el problema de la sección 2.1.8). Consistencia implica que para un registro de datos suficientemente grande, \hat{A} siempre será cercano a la realización de A sin importar el valor de la realización. Ver también el problema de la sección 11.6.10.

11.4.2. Caso vectorial

En el caso de un parámetro $\boldsymbol{\theta}$ vectorial, el error puede definirse como

$$\boldsymbol{\epsilon} = \boldsymbol{\theta} - \hat{\boldsymbol{\theta}}.$$

Al igual que en el caso escalar, si $\hat{\boldsymbol{\theta}}$ es el estimador MMSE, el error $\boldsymbol{\epsilon}$ tiene media nula. La matriz de covarianza es

$$E_{\mathbf{x}, \theta}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T) = E_{\mathbf{x}, \theta} [(\boldsymbol{\theta} - E(\boldsymbol{\theta}|\mathbf{x}))(\boldsymbol{\theta} - E(\boldsymbol{\theta}|\mathbf{x}))^T].$$

La esperanza se toma respecto a $p(\mathbf{x}, \boldsymbol{\theta})$. Notar que

$$\begin{aligned} [E(\boldsymbol{\theta}|\mathbf{x})]_i &= E(\theta_i|\mathbf{x}) = \int \theta_i p(\boldsymbol{\theta}|\mathbf{x}) d\boldsymbol{\theta} \\ &= \int \theta_i p(\theta_i|\mathbf{x}) d\theta_i, \end{aligned}$$

y depende únicamente de \mathbf{x} . Por lo tanto, $E(\boldsymbol{\theta}|\mathbf{x})$ es una función únicamente de \mathbf{x} , como debe ser para un estimador válido. La igualdad anterior surge de la ecuación 11.2, ya que

$$\begin{aligned} E(\theta_i|\mathbf{x}) &= \int \theta_i p(\boldsymbol{\theta}|\mathbf{x}) d\boldsymbol{\theta} \\ &= \int \cdots \int \theta_i p(\boldsymbol{\theta}|\mathbf{x}) d\theta_1 \cdots d\theta_p \\ &= \int \theta_i \left[\int \cdots \int p(\boldsymbol{\theta}|\mathbf{x}) d\theta_1 \cdots d\theta_{i-1} d\theta_{i+1} \cdots d\theta_p \right] d\theta_i \\ &= \int \theta_i p(\theta_i|\mathbf{x}) d\theta_i. \end{aligned}$$

Al igual que en el caso escalar, los elementos de la diagonal de la matriz de covarianza del error son el MSE bayesiano mínimo, ya que el elemento $[i, i]$ es

$$[E_{x,\theta}(\epsilon\epsilon^T)]_{ii} = \iint [(\theta_i - E(\theta_i|\mathbf{x}))^2 p(\mathbf{x}, \boldsymbol{\theta}) d\mathbf{x} d\boldsymbol{\theta}.$$

Integrando respecto a $\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_p$ de forma analoga a la realizada antes, se obtiene que

$$[E_{x,\theta}(\epsilon\epsilon^T)]_{ii} = \int [(\theta_i - E(\theta_i|\mathbf{x}))^2 p(\mathbf{x}, \theta_i) d\mathbf{x} d\theta_i,$$

que es el MSE bayesiano mínimo para $\hat{\theta}_i$ (ver la ecuación 10.1). De esta forma, los elementos de la diagonal de la matriz de covarianza del error $E_{x,\theta}(\epsilon\epsilon^T)$ son los errores MSE bayesianos mínimos. La matriz de covarianza del error se llama *matriz del error cuadrático medio bayesiano* y se denota $\mathbf{M}_{\hat{\theta}}$. Para determinar la PDF de ϵ se derivará la matriz de covarianza completa. La discusión se realizará para el caso particular del modelo lineal bayesiano únicamente (ver la sección 10.4). La matriz de covarianza es

$$\begin{aligned} \mathbf{M}_{\hat{\theta}} &= E_{x,\theta} [(\boldsymbol{\theta} - E(\boldsymbol{\theta}|\mathbf{x}))(\boldsymbol{\theta} - E(\boldsymbol{\theta}|\mathbf{x}))^T] \\ &= \iint (\boldsymbol{\theta} - E(\boldsymbol{\theta}|\mathbf{x}))(\boldsymbol{\theta} - E(\boldsymbol{\theta}|\mathbf{x}))^T p(\mathbf{x}, \boldsymbol{\theta}) d\mathbf{x} d\boldsymbol{\theta} \\ &= \int \left[\int (\boldsymbol{\theta} - E(\boldsymbol{\theta}|\mathbf{x}))(\boldsymbol{\theta} - E(\boldsymbol{\theta}|\mathbf{x}))^T p(\boldsymbol{\theta}|\mathbf{x}) d\boldsymbol{\theta} \right] p(\mathbf{x}) d\mathbf{x} \\ &= E_x E_{\theta|\mathbf{x}} [(\boldsymbol{\theta} - E(\boldsymbol{\theta}|\mathbf{x}))(\boldsymbol{\theta} - E(\boldsymbol{\theta}|\mathbf{x}))^T] \\ &= E_x (\mathbf{C}_{\theta|\mathbf{x}}) \end{aligned}$$

donde $\mathbf{C}_{\theta|\mathbf{x}}$ es la matriz de covarianza de la PDF a posteriori $p(\boldsymbol{\theta}|\mathbf{x})$. Si $\boldsymbol{\theta}$ y \mathbf{x} son conjuntamente gaussianos, $\mathbf{C}_{\theta|\mathbf{x}}$ no depende de \mathbf{x} (ver la sección 10.3) y está dada por la ecuación 10.15, resultando en

$$\mathbf{M}_{\hat{\theta}} = \mathbf{C}_{\theta|\mathbf{x}} = \mathbf{C}_{\theta\theta} - \mathbf{C}_{\theta\mathbf{x}} \mathbf{C}_{\mathbf{x}\mathbf{x}}^{-1} \mathbf{C}_{\mathbf{x}\theta}.$$

Restringiendo la deducción al modelo lineal bayesiano, de la ecuación 10.18,

$$\mathbf{M}_{\hat{\theta}} = \mathbf{C}_{\theta} - \mathbf{C}_{\theta} \mathbf{H}^T (\mathbf{H} \mathbf{C}_{\theta} \mathbf{H}^T + \mathbf{C}_w)^{-1} \mathbf{H} \mathbf{C}_{\theta}, \quad (11.7)$$

o de la ecuación 10.20

$$\mathbf{M}_{\hat{\theta}} = (\mathbf{C}_{\theta}^{-1} + \mathbf{H}^T \mathbf{C}_w^{-1} \mathbf{H})^{-1}, \quad (11.8)$$

donde se simplificó la notación $\mathbf{C}_{\theta\theta}$ a \mathbf{C}_{θ} . Finalmente, se observa que para el modelo lineal bayesiano, $\epsilon = \boldsymbol{\theta} - \hat{\boldsymbol{\theta}}$ es gaussiano, ya que de la ecuación 10.17

$$\begin{aligned} \epsilon &= \boldsymbol{\theta} - \hat{\boldsymbol{\theta}} \\ &= \boldsymbol{\theta} - \boldsymbol{\mu}_{\theta} - \mathbf{C}_{\theta} \mathbf{H}^T (\mathbf{H} \mathbf{C}_{\theta} \mathbf{H}^T + \mathbf{C}_w)^{-1} (\mathbf{x} - \mathbf{H} \boldsymbol{\mu}_{\theta}), \end{aligned}$$

y de esta forma, ϵ es una transformación lineal de $\mathbf{x}, \boldsymbol{\theta}$, que son conjuntamente gaussianos. Por lo tanto, el vector de error del estimador MMSE de los parámetros de un modelo lineal bayesiano se caracteriza como

$$\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{M}_{\hat{\theta}}),$$

donde $\mathbf{M}_{\hat{\theta}}$ está dado por 11.7 o por 11.8. A continuación se aplica el resultado en un ejemplo.

Análisis de Fourier bayesiano (continuación)

En el ejemplo de la sección 11.2.1, se tiene que

$$\mathbf{C}_{\hat{\theta}} = \sigma_{\theta}^2 \mathbf{I}, \quad \mathbf{C}_w = \sigma^2 \mathbf{I}, \quad \mathbf{H}^T \mathbf{H} = \frac{N}{2} \mathbf{I}.$$

De la ecuación 11.8

$$\begin{aligned}\mathbf{M}_{\hat{\theta}} &= (\mathbf{C}_{\theta}^{-1} + \mathbf{H}^T \mathbf{C}_w^{-1} \mathbf{H})^{-1} \\ &= \left(\frac{1}{\sigma_{\theta}^2} \mathbf{I} + \frac{N}{2\sigma^2} \mathbf{I} \right)^{-1} \\ &= \left(\frac{1}{\sigma_{\theta}^2} + \frac{N}{2\sigma^2} \right)^{-1} \mathbf{I}.\end{aligned}$$

Por lo tanto, el error $\boldsymbol{\epsilon} = [\epsilon_{\hat{a}} \ \epsilon_{\hat{b}}]$ tiene PDF

$$\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{M}_{\hat{\theta}})$$

con

$$\mathbf{M}_{\hat{\theta}} = \begin{bmatrix} \frac{1}{\frac{1}{\sigma_{\theta}^2} + \frac{1}{2\sigma^2/N}} & 0 \\ 0 & \frac{1}{\frac{1}{\sigma_{\theta}^2} + \frac{1}{2\sigma^2/N}} \end{bmatrix}.$$

Las componentes del error son independientes y además

$$\begin{aligned}\text{Bsme}(\hat{a}) &= [\mathbf{M}_{\hat{\theta}}]_{11} \\ \text{Bsme}(\hat{b}) &= [\mathbf{M}_{\hat{\theta}}]_{22},\end{aligned}$$

coincidiendo con los resultados del ejemplo de la sección 11.2.1. Para este ejemplo, los contornos de densidad de probabilidad constante del vector de error son circulares. Los contornos de densidad ocurren cuando el exponente $\boldsymbol{\epsilon}^T \mathbf{M}_{\hat{\theta}}^{-1} \boldsymbol{\epsilon}$ de la PDF es constante. Una forma de describir el desempeño del estimador es mediante la elipse de error o concentración, que en este caso es un círculo. Se trata de una elipse en la cual el vector de error pertenece a ella con probabilidad P . Sea

$$\boldsymbol{\epsilon}^T \mathbf{M}_{\hat{\theta}}^{-1} \boldsymbol{\epsilon} = c^2. \quad (11.9)$$

En el caso en que $p = 2$, si $[\mathbf{M}_{\hat{\theta}}^{-1}]_{11} = A$, $[\mathbf{M}_{\hat{\theta}}^{-1}]_{12} = [\mathbf{M}_{\hat{\theta}}^{-1}]_{21} = B$ y $[\mathbf{M}_{\hat{\theta}}^{-1}]_{22} = C$, la ecuación se puede expresar como

$$A\epsilon_{\hat{a}}^2 + 2B\epsilon_{\hat{a}}\epsilon_{\hat{b}} + C\epsilon_{\hat{b}}^2 = c^2,$$

y corresponde a un elipse centrada en el origen con ejes con inclinación arbitraria (ver el apéndice I.2). La condición $B^2 - 4AC < 0$ se cumple por ser $\mathbf{M}_{\hat{\theta}}$ una matriz de covarianza (ver el capítulo 6 de [2]). La probabilidad P de que $\boldsymbol{\epsilon}$ pertenezca a la elipse es

$$P = \Pr\{\boldsymbol{\epsilon}^T \mathbf{M}_{\hat{\theta}}^{-1} \boldsymbol{\epsilon} \leq c^2\}.$$

Pero $u = \boldsymbol{\epsilon}^T \mathbf{M}_{\hat{\theta}}^{-1} \boldsymbol{\epsilon}$ es una variable aleatoria $\chi^2(2)$ (ver el problema de la sección 11.6.13), por lo que su PDF es (ver la ecuación 2.1)

$$p(u) = \begin{cases} \frac{1}{2}e^{-u/2}, & u \geq 0 \\ 0, & u < 0 \end{cases},$$

de forma que

$$\begin{aligned}P &= \Pr\{\boldsymbol{\epsilon}^T \mathbf{M}_{\hat{\theta}}^{-1} \boldsymbol{\epsilon} \leq c^2\} \\ &= \int_0^{c^2} \frac{1}{2}e^{-u/2} du \\ &= -e^{-u/2} \Big|_0^{c^2} \\ &= 1 - e^{-c^2/2}\end{aligned}$$

y despejando c^2 , se tiene que

$$c^2 = 2 \ln \frac{1}{1-P}.$$

Se concluye que

$$\boldsymbol{\epsilon}^T \mathbf{M}_{\hat{\theta}}^{-1} \boldsymbol{\epsilon} = 2 \ln \frac{1}{1-P}$$

describe la elipse de error para la probabilidad P . El vector de error pertenece a esta elipse con probabilidad P . En general, los contornos son elípticos si la matriz de covarianza del error no es una matriz de identidad escalada (ver los problemas de las secciones 11.6.14 y 11.6.15).

A continuación se resumen los resultados en un teorema.

Teorema: Desempeño del estimador MMSE para el modelo lineal bayesiano Si los datos observados \mathbf{x} pueden ser modelados por el modelo lineal bayesiano del teorema de la sección 10.4, el estimador MMSE es (ecuaciones 10.17 y 10.19)

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= \boldsymbol{\mu}_{\theta} + \mathbf{C}_{\theta} \mathbf{H}^T (\mathbf{H} \mathbf{C}_{\theta} \mathbf{H}^T + \mathbf{C}_w)^{-1} (\mathbf{x} - \mathbf{H} \boldsymbol{\mu}_{\theta}) \\ &= \boldsymbol{\mu}_{\theta} + (\mathbf{C}_{\theta}^{-1} + \mathbf{H}^T \mathbf{C}_w^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C}_w^{-1} (\mathbf{x} - \mathbf{H} \boldsymbol{\mu}_{\theta}). \end{aligned}$$

El desempeño del estimador se mide con el error $\boldsymbol{\epsilon} = \boldsymbol{\theta} - \hat{\boldsymbol{\theta}}$, cuya PDF es gausiana con media nula y matriz de covarianza

$$\begin{aligned} \mathbf{C}_{\epsilon} &= E_{\mathbf{x}, \boldsymbol{\theta}} (\boldsymbol{\epsilon} \boldsymbol{\epsilon}^T) \\ &= \mathbf{C}_{\theta} - \mathbf{C}_{\theta} \mathbf{H}^T (\mathbf{H} \mathbf{C}_{\theta} \mathbf{H}^T + \mathbf{C}_w)^{-1} \mathbf{H} \mathbf{C}_{\theta} \\ &= (\mathbf{C}_{\theta}^{-1} + \mathbf{H}^T \mathbf{C}_w^{-1} \mathbf{H})^{-1}. \end{aligned}$$

La matriz de covarianza del error es también la matriz de MSE mínimo $\mathbf{M}_{\hat{\theta}}$, cuyos elementos de la diagonal son el MSE bayesiano mínimo, o

$$\begin{aligned} [\mathbf{M}_{\hat{\theta}}]_{ii} &= [\mathbf{C}_{\epsilon}]_{ii} \\ &= \text{Bsme}(\hat{\theta}_i). \end{aligned}$$

Ver los problemas de la sección 11.6.16 y 11.6.17 sobre una aplicación en el ajuste de una recta.

11.5. Ejemplo de procesamiento de señales

Hay diversos problemas de procesamiento de señales que se ajustan al modelo lineal bayesiano. Un ejemplo es el problema de *deconvolución*, que se describe a continuación. En un sistema de comunicación, se transmite una señal $s(t)$ a través de un canal con respuesta al impulso $h(t)$. A la salida del canal se observa la señal distorsionada por el canal $s_0(t)$. El objetivo es deconvolucionar $s(t)$ de una versión contaminada con ruido de $s_0(t) = s(t) * h(t)$. El modelo del sistema se muestra en la figura 11.3. Para simplificar el problema se asume que $h(t)$ es conocido. En caso contrario el problema es el de *deconvolución ciega*, considerablemente mas difícil de resolver. Se asume además que $s(t)$ es la realización de un proceso aleatorio. Este modelo es apropiado por ejemplo para señales de voz, en donde la señal cambia con el hablante y el contenido. Se observa la señal $s_0(t)$ corrompida con ruido $w(t)$. Para convertir el problema

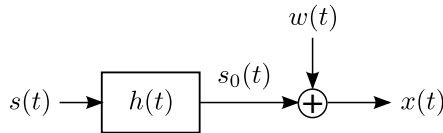


Figura 11.3: Modelo del sistema en el problema de deconvolución.

a tiempo discreto, se asume que $s(t)$ es esencialmente de banda limitada de B Hz de forma tal que la salida $s_0(t)$ también es de banda limitada de B Hz. Se asume además que el ruido de tiempo continuo es un proceso WSS gaussiano con PSD

$$P_{ww}(f) = \begin{cases} \frac{N_0}{2} & |f| < B \\ 0 & |f| > B. \end{cases}$$

De esta forma, la potencia del ruido es

$$\sigma^2 = \int_{-\infty}^{\infty} P_{ww}(f) df = \int_{-B}^B \frac{N_0}{2} df = N_0 B.$$

Luego del muestreo apropiado, el modelo de los datos en tiempo discreto es

$$x[n] = \sum_{m=0}^{n_s-1} h[n-m]s[m] + w[n], \quad n = 0, \dots, N-1.$$

La señal $s[n]$ es no nula en el intervalo $[0, n_s-1]$ y $h[n]$ es no nula en el intervalo $[0, n_h-1]$. Por lo tanto, $s_0[n]$ es no nula en el intervalo $[0, (n_s-1) + (n_h-1)] = [0, n_s+n_h-2]$ resultando en que $N = n_s+n_h-1$. Además, $w[n]$ es WGN con varianza $\sigma^2 = N B_0$. Si se asume que $s(t)$ es un proceso gaussiano, $s[n]$ es un proceso gaussiano en tiempo discreto. La muestra n -ésima de la señal observada es

$$x[n] = h[n]s[0] + h[n-1]s[1] + \dots + h[n-(n_s+1)]s[n_s-1], \quad \text{con } n = 0, \dots, N-1,$$

donde $h[n] = 0$ si $n < 0$ o $n > n_h-1$. En notación matricial, el modelo del sistema es

$$\underbrace{\begin{bmatrix} x[0] \\ x[1] \\ \vdots \\ x[N-1] \end{bmatrix}}_{\mathbf{x}} = \underbrace{\begin{bmatrix} h[0] & 0 & \dots & 0 \\ h[1] & h[0] & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ h[n_s-1] & h[n_s-2] & \dots & h[0] \\ \vdots & \vdots & & \vdots \\ h[N-1] & h[N-2] & \dots & h[N-p] \end{bmatrix}}_{\mathbf{H}} \underbrace{\begin{bmatrix} s[0] \\ s[1] \\ \vdots \\ s[n_s-1] \end{bmatrix}}_{\boldsymbol{\theta}} + \underbrace{\begin{bmatrix} w[0] \\ w[1] \\ \vdots \\ w[N-1] \end{bmatrix}}_{\mathbf{w}}.$$

Notar que en la matriz \mathbf{H} los elementos $h[n] = 0$ si $n > n_h-1$. Se observa que se trata del modelo lineal bayesiano con $p = n_s-1$. Para completar el modelo falta únicamente especificar la media y la covarianza de $\boldsymbol{\theta} = \mathbf{s}$. Si la señal es de media nula, como en el caso de una señal de voz, se puede asumir la PDF a priori

$$\mathbf{s} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_s).$$

Puede encontrarse una forma explícita de la matriz de covarianza asumiendo que la señal es WSS, al menos durante un intervalo de tiempo corto, como en la señal de voz. De esta forma,

$$[\mathbf{C}_s]_{ij} = r_{ss}[i-j]$$

donde $r_{ss}[k]$ es la ACF. Si se conoce la PSD, puede obtenerse la ACF y queda \mathbf{C}_s determinada. Como resultado, de la ecuación 10.17, el estimador MMSE de \mathbf{s} es

$$\hat{\mathbf{s}} = \mathbf{C}_s \mathbf{H}^T (\mathbf{H} \mathbf{C}_s \mathbf{H}^T + \sigma^2 \mathbf{I})^{-1} \mathbf{x}. \quad (11.10)$$

Un caso especial ocurre cuando $\mathbf{H} = \mathbf{I}$, donde \mathbf{I} es la matriz identidad $n_s \times n_s$, que corresponde cuando la respuesta al impulso del canal es $h[n] = \delta[n]$, y el modelo lineal es

$$\mathbf{x} = \boldsymbol{\theta} + \mathbf{w}.$$

Observar que en el modelo lineal clásico se asumió que \mathbf{H} tiene dimensiones $N \times p$ con $N > p$. Esto es importante desde el punto de vista práctico, ya que para $\mathbf{H} = \mathbf{I}$, el estimador MVU es

$$\hat{\boldsymbol{\theta}} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x} = \mathbf{x}.$$

En este caso no hay promediado debido a que se están estimado tantos parámetros como observaciones. En el caso bayesiano, se permite que $N = p$ o incluso que $N < p$ debido a que se tiene conocimiento a priori de $\boldsymbol{\theta}$. Si $N < p$, la ecuación 11.10 tiene solución, mientras que en la estimación clásica, el estimador MVU no existe ($\mathbf{H}^T \mathbf{H}$ no es invertible). Asumiendo que $\mathbf{H} = \mathbf{I}$, el estimador MMSE de \mathbf{s} es

$$\hat{\mathbf{s}} = \mathbf{C}_s (\mathbf{C}_s + \sigma^2 \mathbf{I})^{-1} \mathbf{x}. \quad (11.11)$$

Notar que el estimador puede escribirse como $\hat{s} = \mathbf{A}\mathbf{x}$, donde \mathbf{A} es una matriz $n_s \times n_s$. La matriz \mathbf{A} se llama *filtro de Wiener*. Notar que se trata de un filtro no causal, ya que para calcular cada muestra de la salida se emplean todas las muestras de la entrada. Como ejemplo, para el caso escalar en el que se quiere estimar $s[0]$ a partir de $x[0]$, el filtro de Wiener es

$$\begin{aligned}\hat{s}[0] &= \frac{r_{ss}[0]}{r_{ss}[0] + \sigma^2} x[0] \\ &= \frac{\eta}{\eta + 1} x[0]\end{aligned}$$

donde $\eta = r_{ss}[0]/\sigma^2$ es la SNR. Para una SNR alta, $\hat{s}[0] \approx x[0]$, mientras que para una SNR baja, $\hat{s}[0] \approx 0$. Un ejemplo mas interesante se obtiene si $s[n]$ es la realización de un proceso AR(1), o

$$s[n] = -a[1]s[n-1] + u[n]$$

donde $u[n]$ es WGN con varianza σ_u^2 y $a[1]$ es el coeficiente del filtro. La ACF de este proceso es

$$r_{ss}[k] = \frac{\sigma_u^2}{1 - a^2[1]} (-a[1])^{|k|}$$

y la PSD es

$$P_{ss}(f) = \frac{\sigma_u^2}{|1 + a[1] \exp(-j2\pi f)|^2}.$$

Estos resultados se obtienen mediante las ecuaciones de Yule-Walker (ver el capítulo 3 de [5]), que indican que para un proceso $x[n]$ AR(p),

$$\sum_{k=0}^p a[k]x[n-k] = u[n], \quad \text{con } a[0] = 1,$$

donde $u[n]$ es WGN con varianza σ_u^2 , la ACF cumple la ecuación en diferencias

$$\sum_{k=0}^p a[k]r_{xx}[n-k] = \begin{cases} \sigma_u^2 & n = 0 \\ 0 & n \geq 1. \end{cases}$$

En el caso en que $a[1] < 0$, la PSD es la de un proceso pasabajo, como se muestra en la figura 11.4 para $a[1] = -0.95$ y $\sigma_u^2 = 1$. También se muestra una realización del proceso $s[n]$ para los mismos parámetros AR. Cuando se adiciona WGN de forma de tener una SNR de 5 dB, se obtienen los datos $x[n]$ observados. Notar que para una SNR dada, la potencia σ^2 del ruido aditivo es

$$10 \log \frac{r_{ss}[0]}{\sigma^2} = \text{SNR} \quad \Rightarrow \quad \sigma^2 = \frac{r_{ss}[0]}{10^{\text{SNR}/10}}.$$

Mediante el filtro de Wiener dado por la ecuación 11.11, se obtiene la estimación $\hat{s}[n]$. Como se observa en la figura, el filtro de Wiener suaviza las fluctuaciones producidas por el ruido. Sin embargo, como contrapartida, la señal también sufre de suavizado. Puede mostrarse, como se hace en el problema de la sección 11.6.18, que el filtro de Wiener actúa como pasabajos en este caso.

11.6. Problemas

11.6.1. Problema 1

Se observan los datos $x[n]$, $n = 0, \dots, N-1$ que tienen PDF

$$p(x[n]|\mu) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{1}{2\sigma^2} (x[n] - \mu^2) \right].$$

Los $x[n]$ son independientes cuando se condicionan en μ . La media μ tiene PDF a priori

$$\mu \sim \mathcal{N}(\mu_0, \sigma_0^2).$$

Encontrar los estimadores MMSE y MAP de μ . ¿Qué ocurre cuando $\sigma_0^2 \rightarrow 0$ y $\sigma_0^2 \rightarrow \infty$?

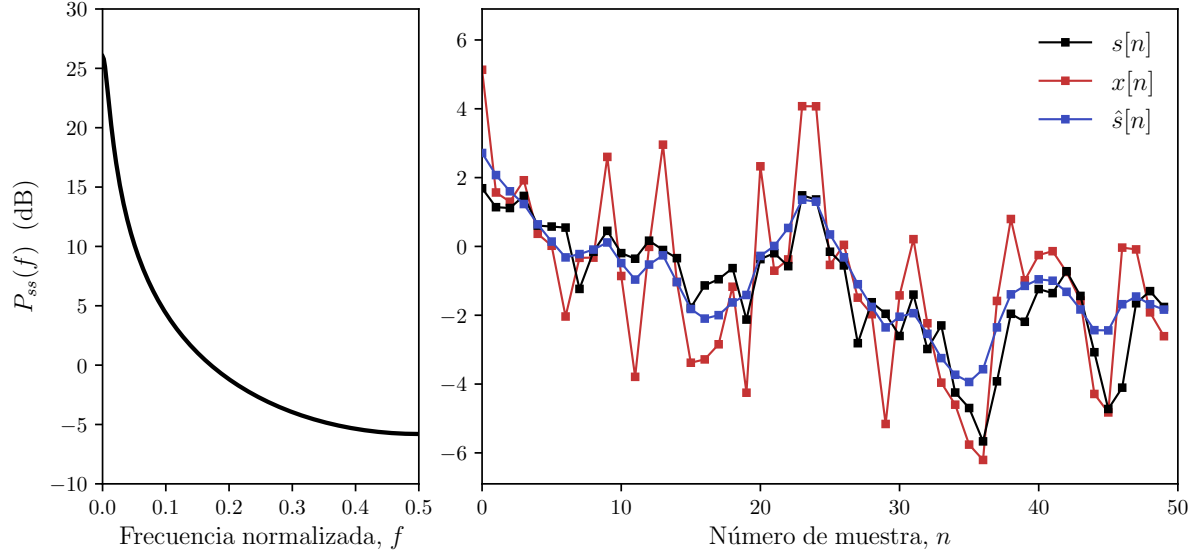


Figura 11.4: Filtro de Wiener. Se muestra la PSD de la señal $s[n]$ a estimar y una realización de la señal $s[n]$ a estimar, la señal $x[n] = s[n] + w[n]$, que es la señal a estimar contaminada con WGN y la señal $\hat{s}[n]$ estimada, que es a señal $x[n]$ filtrada con el filtro de Wiener.

Solución Este problema es el mismo que la estimación del nivel de DC en WGN con PDF a priori gaussiana del ejemplo de la sección 10.2.1 con $A = \mu$. En dicho ejemplo, se encontró que la PDF a posteriori es

$$p(\mu|\mathbf{x}) = \frac{1}{\sqrt{2\pi\sigma_{\mu|x}^2}} \exp \left[-\frac{1}{2\sigma_{\mu|x}^2} (A - \mu_{\mu|x})^2 \right]$$

con

$$\mu_{\mu|x} = \frac{\frac{N}{\sigma^2} \bar{x} + \frac{\mu_0}{\sigma_0^2}}{\frac{N}{\sigma^2} + \frac{1}{\sigma_0^2}} \quad \text{y} \quad \sigma_{\mu|x}^2 = \frac{1}{\frac{N}{\sigma^2} + \frac{1}{\sigma_0^2}}.$$

Como la PDF a posteriori es gaussiana, la media coincide con la moda, por lo que los estimadores MMSE y MAP coinciden,

$$\hat{\mu}_{\text{MMSE}} = E(\mathbf{x}|\mu) = \mu_{\mu|x}, \quad \text{y} \quad \hat{\mu}_{\text{MAP}} = \arg \max_{\mu} p(\mu|\mathbf{x}) = \mu_{\mu|x}.$$

Cuando $\sigma_0^2 \rightarrow 0$, se tiene que

$$\hat{\mu} \rightarrow \frac{\frac{\mu_0}{\sigma_0^2}}{\frac{1}{\sigma_0^2}} = \mu_0.$$

En este caso, no hay incertidumbre en el conocimiento a priori. Domina el conocimiento a priori y se descartan los datos. En el caso en que $\sigma_0^2 \rightarrow \infty$, se cumple que

$$\hat{\mu} \rightarrow \frac{\frac{N}{\sigma^2} \bar{x}}{\frac{N}{\sigma^2}} = \bar{x}.$$

Se descarta el conocimiento a priori y el estimador es la media muestral. Los datos dominan sobre el conocimiento a priori.

11.6.2. Problema 2

Sea la PDF a posteriori

$$p(\theta|x) = \frac{\epsilon}{\sqrt{2\pi}} \exp \left[-\frac{1}{2}(\theta - x)^2 \right] + \frac{1-\epsilon}{\sqrt{2\pi}} \exp \left[-\frac{1}{2}(\theta + x)^2 \right].$$

Graficar la PDF para $\epsilon = 1/2$ y $\epsilon = 3/4$. Luego, encontrar los estimadores MMSE y MAP para los mismos valores de ϵ .

Solución Las PDF son gaussianas bimodales y sus gráficas se muestran en la figura 11.5. Como se

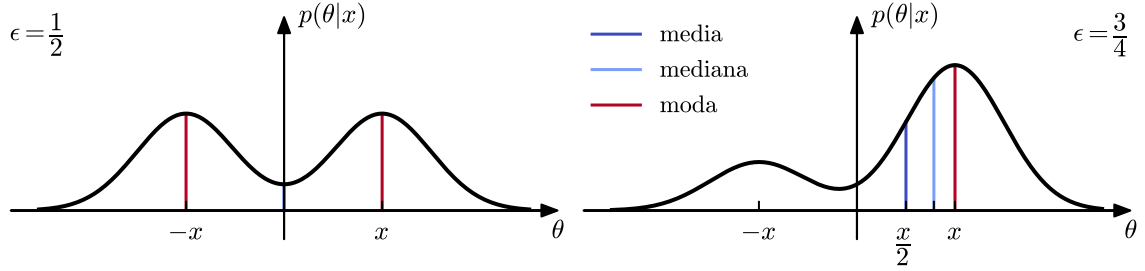


Figura 11.5: PDF a posteriori con $\epsilon = 1/2$ y $\epsilon = 3/4$.

discutió en la sección 11.1, los estimadores óptimos que minimizan el costo de Bayes son la media, la mediana y la moda si se emplean funciones de costo cuadrática, valor absoluto y “acierto o fallo” respectivamente. El estimador obtenido con la función de costo cuadrática es el estimador MMSE y el obtenido con la función de costo “acierto o fallo” es el estimador MAP. En el caso en que $\epsilon = 1/2$, la PDF es simétrica. Por lo tanto, la media de la PDF a posteriori es nula, y el estimador MMSE es $\hat{\theta}_{\text{MMSE}} = 0$. Por la propiedad de simetría de la PDF, la media y la mediana coinciden, por lo que el estimador con la función de costo valor absoluto coincide con el estimador MMSE. El estimador MAP, que es la moda de la PDF, es

$$\hat{\theta}_{\text{MAP}} = \arg \max_{\theta} p(\theta|x) = \pm x,$$

no único. Con $\epsilon = 3/4$, la PDF a posteriori no es simétrica. El estimador MMSE es

$$\begin{aligned} \hat{\theta}_{\text{MMSE}} &= E(\theta|x) \\ &= \int_{-\infty}^{\infty} \theta p(\theta|x) d\theta \\ &= \epsilon \int_{-\infty}^{\infty} \frac{\theta}{\sqrt{2\pi}} \exp \left[-\frac{1}{2}(\theta - x)^2 \right] d\theta + (1-\epsilon) \int_{-\infty}^{\infty} \frac{\theta}{\sqrt{2\pi}} \exp \left[-\frac{1}{2}(\theta + x)^2 \right] d\theta \\ &= \epsilon x + (1-\epsilon)(-x) \\ &= x(2\epsilon - 1), \end{aligned}$$

y evaluando en $\epsilon = 3/4$ se obtiene que

$$\hat{\theta}_{\text{MMSE}} = \frac{x}{2}.$$

El estimador MAP es

$$\hat{\theta}_{\text{MAP}} = x,$$

como se muestra en la figura 11.5. En la figura también se muestra la mediana, que es el estimador obtenido con la función de costo valor absoluto. Para encontrar el estimador, se considera que

$$\begin{aligned}\Pr\{\theta \leq z|x\} &= \int_{-\infty}^z p(\theta|x) d\theta \\ &= \epsilon \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(\theta - x)^2\right] d\theta + (1 - \epsilon) \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(\theta + x)^2\right] d\theta \\ &= \epsilon \Phi(z - x) + (1 - \epsilon) \Phi(z + x),\end{aligned}$$

donde

$$\Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{\theta^2/2} d\theta$$

es la distribución de probabilidad acumulada de una variable aleatoria gaussiana normalizada. El estimador, por ser la mediana, cumple que

$$\epsilon \Phi(\hat{\theta} - x) + (1 - \epsilon) \Phi(\hat{\theta} + x) = \frac{1}{2},$$

y no puede expresarse analíticamente. En la figura 11.6 se muestra el código para calcular el estimador empleando una computadora.

```
from scipy.stats import norm
from scipy import optimize

def j(z):
    return epsilon * norm.cdf(z-x) + (1 - epsilon) * norm.cdf(z+x) - 1/2

z0 = 0
j_root = optimize.root(j, z0)
```

Figura 11.6: Código en el lenguaje Python del cálculo de la mediana de una PDF gaussiana bimodal.

11.6.3. Problema 3

Para la PDF a posteriori

$$p(\theta|x) = \begin{cases} \exp[-(\theta - x)] & \theta > x \\ 0 & \theta < x \end{cases}$$

encontrar los estimadores MMSE y MAP.

Solución El estimador MMSE es

$$\begin{aligned}\hat{\theta}_{\text{MMSE}} &= E(\theta|x) = \int_{-\infty}^{\infty} \theta p(\theta|x) d\theta \\ &= \int_x^{\infty} \theta e^{-(\theta-x)} d\theta \\ &\stackrel{(a)}{=} -\theta e^{-(\theta-x)} \Big|_x^{\infty} + \int_x^{\infty} e^{-(\theta-x)} d\theta \\ &= x - e^{-(\theta-x)} \Big|_x^{\infty} \\ &= x + 1,\end{aligned}$$

donde en (a) se aplicó integración por partes $\int u dv = uv - \int v du$ con $u = \theta$ y $dv = e^{-(\theta-x)} d\theta$ y por lo tanto, $du = d\theta$ y $v = -e^{-(\theta-x)}$. Para calcular el estimador MAP se tiene en cuenta que la PDF a posteriori es decreciente en $[x, \infty)$ y por lo tanto

$$\hat{\theta}_{\text{MAP}} = \arg \max_{\theta} p(\theta|x) = x.$$

En la figura 11.7 se muestra la PDF a posteriori y los estimadores.

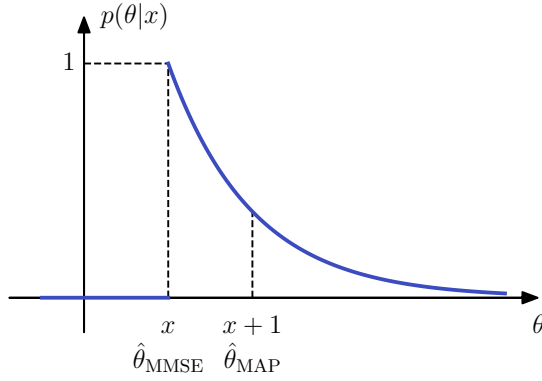


Figura 11.7: PDF a posteriori y estimadores MMSE y MAP.

11.6.4. Problema 4

Se observan los datos $x[n] = A + w[n]$ para $n = 0, \dots, N-1$. Se asume que el parámetro desconocido A tiene PDF a priori

$$p(A) = \begin{cases} \lambda \exp(-\lambda A) & A > 0 \\ 0 & A < 0, \end{cases}$$

donde $\lambda > 0$, y $w[n]$ es WGN con varianza σ^2 y es independiente de A . Encontrar el estimador MAP de A .

Solución La PDF a posteriori está dada por

$$p(A|\mathbf{x}) = \frac{p(\mathbf{x}|A)p(A)}{\int p(\mathbf{x}|A)p(A) dA}.$$

Como el denominador es una constante independiente de A , para encontrar el estimador MAP alcanza con maximizar el numerador. La PDF condicional de los datos es

$$p(\mathbf{x}|A) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - A)^2 \right],$$

por lo que el numerador de la PDF a posteriori para $A > 0$ es

$$\begin{aligned} p(\mathbf{x}|A)p(A) &= \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x[n] - A)^2 \right] \lambda \exp(-\lambda A) \\ &= \frac{\lambda}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} (x^2[n] - 2Ax[n] + A^2) - \lambda A \right] \\ &= \frac{\lambda}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left[-\frac{1}{2\sigma^2} \left(\sum_{n=0}^{N-1} x^2[n] - 2AN\bar{x} + NA^2 \right) - \lambda A \right] \\ &= \frac{\lambda}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left(-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} x^2[n] \right) \exp \left\{ -\frac{1}{2\sigma^2} [NA^2 - 2A(N\bar{x} - \sigma^2\lambda)] \right\} \\ &= \frac{\lambda}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left(-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} x^2[n] \right) \exp \left\{ -\frac{N}{2\sigma^2} \left[A^2 - 2A \left(\bar{x} - \frac{\sigma^2}{N}\lambda \right) \right] \right\}. \end{aligned}$$

Empleando la técnica de completar el cuadrado, $(A - B)^2 = A^2 - 2AB + B^2$, la segunda exponencial puede expresarse como

$$\begin{aligned} \exp \left\{ -\frac{N}{2\sigma^2} \left[A^2 - 2A \left(\bar{x} - \frac{\sigma^2}{N} \lambda \right) \right] \right\} &= \exp \left\{ -\frac{N}{2\sigma^2} \left[A^2 - 2A \left(\bar{x} - \frac{\sigma^2}{N} \lambda \right) + \left(\bar{x} - \frac{\sigma^2}{N} \lambda \right)^2 \right] \right\} \\ &\quad \times \exp \left[\frac{N}{2\sigma^2} \left(\bar{x} - \frac{\sigma^2}{N} \lambda \right)^2 \right] \\ &= \exp \left\{ -\frac{N}{2\sigma^2} \left[A - \left(\bar{x} - \frac{\sigma^2}{N} \lambda \right) \right]^2 \right\} \exp \left[\frac{N}{2\sigma^2} \left(\bar{x} - \frac{\sigma^2}{N} \lambda \right)^2 \right], \end{aligned}$$

por lo que el numerador de la PDF a posteriori queda

$$p(\mathbf{x}|A)p(A) = \frac{\lambda}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp \left(-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} x^2[n] \right) \exp \left[\frac{N}{2\sigma^2} \left(\bar{x} - \frac{\sigma^2}{N} \lambda \right)^2 \right] \exp \left\{ -\frac{N}{2\sigma^2} \left[A - \left(\bar{x} - \frac{\sigma^2}{N} \lambda \right) \right]^2 \right\}.$$

En la PDF a posteriori, los términos independientes de A del numerador se cancelan con los del denominador, por lo que se concluye que

$$p(A|\mathbf{x}) = \begin{cases} \frac{\exp \left\{ -\frac{N}{2\sigma^2} \left[A - \left(\bar{x} - \frac{\sigma^2}{N} \lambda \right) \right]^2 \right\}}{\int_0^\infty \exp \left\{ -\frac{N}{2\sigma^2} \left[A - \left(\bar{x} - \frac{\sigma^2}{N} \lambda \right) \right]^2 \right\} dA} & A > 0 \\ 0 & A < 0, \end{cases}$$

que es una PDF gaussiana truncada en 0. Teniendo en cuenta que el denominador es una constante, el estimador MAP es el valor de A que maximiza el numerador. En la figura 11.8 se muestra el numerador. Se pueden distinguir dos casos, cuando $\bar{x} > \lambda\sigma^2/N$ y $\bar{x} < \lambda\sigma^2/N$. Teniendo esto en cuenta, el estimador

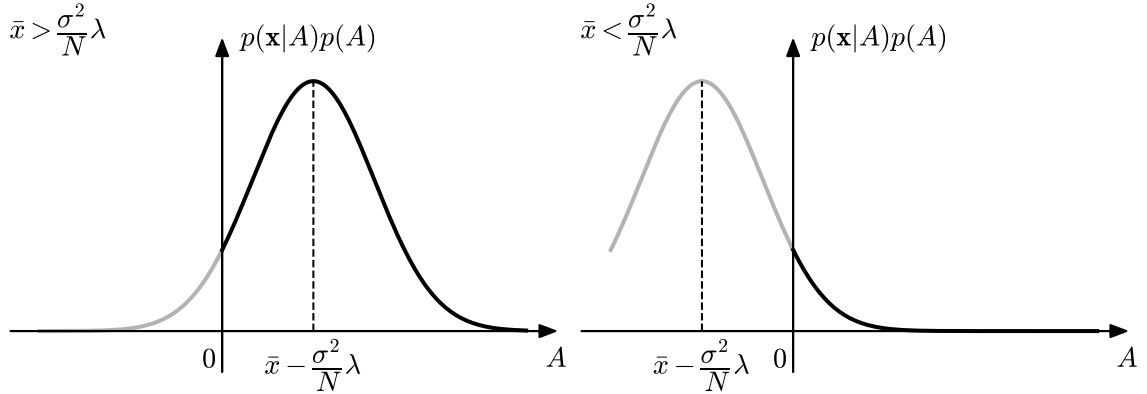


Figura 11.8: Numerador de la PDF a posteriori $p(\mathbf{x}|A)p(A)$ para el caso en que $\bar{x} > \lambda\sigma^2/N$ y $\bar{x} < \lambda\sigma^2/N$. El máximo de la PDF a posteriori se da en $\bar{x} - \lambda\sigma^2/N$ y 0 respectivamente.

MAP es

$$\hat{A} = \begin{cases} \bar{x} - \frac{\sigma^2}{N} \lambda & \bar{x} > \frac{\sigma^2}{N} \lambda \\ 0 & \bar{x} < \frac{\sigma^2}{N} \lambda \end{cases}$$

11.6.5. Problema 5

Si $\boldsymbol{\theta}$ y \mathbf{x} son conjuntamente gaussianos de forma que

$$\begin{bmatrix} \boldsymbol{\theta} \\ \mathbf{x} \end{bmatrix} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$$

donde

$$\boldsymbol{\mu} = \begin{bmatrix} E(\boldsymbol{\theta}) \\ E(\mathbf{x}) \end{bmatrix} \quad \text{y} \quad \begin{bmatrix} \mathbf{C}_{\theta\theta} & \mathbf{C}_{\theta x} \\ \mathbf{C}_{x\theta} & \mathbf{C}_{xx} \end{bmatrix},$$

encontrar el estimador MMSE de $\boldsymbol{\theta}$ en base a \mathbf{x} . Además, determinar el MSE bayesiano mínimo de cada componente de $\boldsymbol{\theta}$. Explicar que ocurre si $\mathbf{C}_{\theta x} = \mathbf{0}$.

Solución El estimador MMSE es

$$\hat{\boldsymbol{\theta}} = E(\boldsymbol{\theta}|\mathbf{x})$$

y de la ecuación 11.3, el MSE bayesiano mínimo de cada componente es

$$\text{Bmse}(\hat{\theta}_i) = \int [\mathbf{C}_{\theta|x}]_{ii} p(\mathbf{x}) d\mathbf{x}, \quad i = 1, \dots, p.$$

Además, de los resultados de la PDF condicional de vectores conjuntamente gaussianos obtenidos en la sección 10.3 (ecuaciones 10.14 y 10.15), el estimador MMSE es

$$\hat{\boldsymbol{\theta}} = E(\boldsymbol{\theta}|\mathbf{x}) = E(\boldsymbol{\theta}) + \mathbf{C}_{\theta x} \mathbf{C}_{xx}^{-1} \mathbf{C}_{x\theta},$$

y el MSE bayesiano mínimo es

$$\text{Bmse}(\hat{\theta}_i) = [\mathbf{C}_{\theta|x}]_{ii}, \quad i = 1, \dots, p$$

ya que la matriz de covarianza

$$\mathbf{C}_{\theta|x} = \mathbf{C}_{\theta\theta} + \mathbf{C}_{\theta x} \mathbf{C}_{xx}^{-1} \mathbf{C}_{x\theta}$$

es independiente de \mathbf{x} .

Si $\mathbf{C}_{\theta x} = \mathbf{0}$, el estimador MMSE y su la matriz de covarianza son respectivamente

$$\hat{\boldsymbol{\theta}} = E(\boldsymbol{\theta}) \quad \text{y} \quad \mathbf{C}_{\theta|x} = \mathbf{C}_{\theta\theta}.$$

Esto se debe a que si $\mathbf{C}_{\theta x} = \mathbf{0}$, $\boldsymbol{\theta}$ y \mathbf{x} no están correlacionados y como además son conjuntamente gaussianos, son independientes. Por lo tanto, los datos no aportan información sobre el parámetro. Se descartan los datos y se emplea únicamente el conocimiento a priori.

11.6.6. Problema 6

Se considera el modelo de datos del ejemplo de la sección 11.2.1 para una senoide en WGN. Se reescribe el modelo como

$$x[n] = A \cos(2\pi f_0 n + \phi) + w[n]$$

donde

$$\begin{aligned} A &= \sqrt{a^2 + b^2} \\ \phi &= \arctan\left(-\frac{b}{a}\right). \end{aligned}$$

Si $\boldsymbol{\theta} = [a \ b]^T \sim \mathcal{N}(\mathbf{0}, \sigma_\theta^2 \mathbf{I})$, mostrar que la PDF de A es Rayleigh, la PDF de ϕ es $\mathcal{U}[0, 2\pi]$, y A y ϕ son independientes.

Solución Este problema esta resuelto en el ejemplo 6-22 de [2]. El modelo es equivalente, ya que

$$\begin{aligned} A \cos(2\pi f_0 n + \phi) &\stackrel{(a)}{=} A (\cos 2\pi f_0 n \cos \phi - \sin 2\pi f_0 n \sin \phi) \\ &= (A \cos \phi) \cos 2\pi f_0 n + (-A \sin \phi) \sin 2\pi f_0 n, \end{aligned}$$

donde en (a) se empleó la identidad trigonométrica $\cos(\alpha \pm \beta) = \cos \alpha \cos \beta \mp \sin \alpha \sin \beta$. Por lo tanto

$$\begin{cases} a = A \cos \phi \\ b = -A \sin \phi \end{cases} \Rightarrow \begin{cases} A = \sqrt{a^2 + b^2} \\ \phi = \arctan\left(-\frac{b}{a}\right). \end{cases}$$

Para resolver el problema, se empleará el teorema fundamental de transformación de variables aleatorias (ver el capítulo 7 de [2]), que indica que dadas n funciones

$$g_1(\mathbf{x}), \dots, g_n(\mathbf{x}), \quad \text{con} \quad \mathbf{x} = [x_1, \dots, x_n]^T,$$

si se construye el vector aleatorio $\mathbf{y} = [y_1, \dots, y_n]^T$ con

$$y_1 = g_1(\mathbf{x}), \dots, y_n = g_n(\mathbf{x}),$$

la PDF de \mathbf{y} puede determinarse en términos de la PDF de \mathbf{x} . Para hacerlo, para un conjunto específico de valores y_1, \dots, y_n , se resuelve el sistema

$$g_1(\mathbf{x}) = y_1, \dots, g_n(\mathbf{x}) = y_n. \quad (11.12)$$

Si el sistema no tiene soluciones, $p(\mathbf{y}) = 0$. Si el sistema tiene una solución única $\mathbf{x} = [x_1, \dots, x_n]$,

$$p_y(y_1, \dots, y_n) = \frac{p_x(x_1, \dots, x_n)}{|\det[J(x_1, \dots, x_n)]|} \quad (11.13)$$

donde

$$J(x_1, \dots, x_n) = \begin{bmatrix} \frac{\partial g_1}{\partial x_1} & \dots & \frac{\partial g_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_n}{\partial x_1} & \dots & \frac{\partial g_n}{\partial x_n} \end{bmatrix}$$

es la matriz jacobiana de la transformación. Si el sistema tiene múltiples soluciones, se suman los términos correspondientes como en el caso escalar (ver la ecuación 2.3). En este caso, se tiene que

$$A = g_1(a, b) = \sqrt{a^2 + b^2}, \quad \phi = g_2(a, b) = \arctan\left(-\frac{b}{a}\right),$$

y la solución $a = A \cos \phi$, $b = -A \sin \phi$ es única, es decir, la transformación es inyectiva. La matriz Jacobiana de la transformación es

$$J(a, b) = \begin{bmatrix} \frac{\partial A}{\partial a} & \frac{\partial A}{\partial b} \\ \frac{\partial \phi}{\partial a} & \frac{\partial \phi}{\partial b} \end{bmatrix}.$$

Como

$$\frac{\partial A}{\partial a} = \frac{\partial (a^2 + b^2)^{\frac{1}{2}}}{\partial a} = \frac{1}{2}(a^2 + b^2)^{-\frac{1}{2}} 2a = \frac{a}{\sqrt{a^2 + b^2}} \quad \text{y análogamente} \quad \frac{\partial A}{\partial b} = \frac{b}{\sqrt{a^2 + b^2}},$$

y además,

$$\frac{\partial \phi}{\partial a} = \frac{\partial}{\partial a} \arctan\left(-\frac{b}{a}\right) = \frac{1}{1 + \left(\frac{b}{a}\right)^2} \frac{b}{a^2} = \frac{b}{a^2 + b^2}$$

y

$$\frac{\partial \phi}{\partial b} = \frac{\partial}{\partial b} \arctan \left(-\frac{b}{a} \right) = \frac{1}{1 + \left(\frac{b}{a} \right)^2} \left(-\frac{1}{a} \right) = -\frac{a}{a^2 + b^2},$$

resultando en

$$J(a, b) = \begin{bmatrix} \frac{a}{\sqrt{a^2 + b^2}} & \frac{b}{\sqrt{a^2 + b^2}} \\ \frac{b}{a^2 + b^2} & -\frac{a}{a^2 + b^2} \end{bmatrix} = \begin{bmatrix} \frac{a}{A} & \frac{b}{A} \\ \frac{b}{A^2} & -\frac{a}{A^2} \end{bmatrix}.$$

El determinante de la matriz Jacobiana es

$$\det[J(A \cos \phi, -A \sin \phi)] = -\frac{a^2}{A^3} - \frac{b^2}{A^3} = -\frac{A^2}{A^3} = -\frac{1}{A}.$$

Como $\boldsymbol{\theta} = [a \ b]^T \sim \mathcal{N}(\mathbf{0}, \sigma_\theta^2 \mathbf{I})$,

$$p_{a,b}(a, b) = \frac{1}{2\pi\sigma_\theta^2} \exp \left[-\frac{a^2 + b^2}{2\sigma_\theta^2} \right],$$

y por la ecuación 11.13 se tiene que

$$p_{A,\phi}(A, \phi) = \frac{p_{a,b}(A \cos \phi, -A \sin \phi)}{|\det[J(A \cos \phi, -A \sin \phi)]|}.$$

Combinando los resultados, se concluye que

$$p(A, \phi) = \frac{A}{2\pi\sigma_\theta^2} \exp \left(-\frac{A^2}{2\sigma_\theta^2} \right), \quad \text{con} \quad A > 0, \quad 0 \leq \phi < 2\pi.$$

Las PDF de A y ϕ se obtienen a partir de las PDF marginales. La PDF de A es

$$\begin{aligned} p(A) &= \int_0^{2\pi} p(A, \phi) d\phi \\ &= \frac{A}{2\pi\sigma_\theta^2} \exp \left(-\frac{A^2}{2\sigma_\theta^2} \right) \int_0^{2\pi} d\phi \\ &= \frac{A}{2\pi\sigma_\theta^2} \exp \left(-\frac{A^2}{2\sigma_\theta^2} \right) \left(\phi \Big|_0^{2\pi} \right) \end{aligned}$$

resultando en

$$p(A) = \frac{A}{\sigma_\theta^2} \exp \left(-\frac{A^2}{2\sigma_\theta^2} \right), \quad A > 0.$$

Se concluye a A es una variable aleatoria Rayleigh con parámetro de escala σ_θ^2 . La PDF de ϕ es

$$\begin{aligned} p(\phi) &= \int_0^\infty p(A, \phi) dA \\ &= \frac{1}{2\pi\sigma_\theta^2} \int_0^\infty A \exp \left(-\frac{A^2}{2\sigma_\theta^2} \right) dA \\ &= \frac{1}{2\pi\sigma_\theta^2} \left[-\sigma_\theta^2 \exp \left(-\frac{A^2}{2\sigma_\theta^2} \right) \Big|_0^\infty \right] \\ &= \frac{1}{2\pi} [-(0 - 1)], \end{aligned}$$

resultando en

$$p(\phi) = \frac{1}{2\pi}, \quad 0 \leq \phi < 2\pi,$$

es decir, $\phi \sim \mathcal{U}[0, 2\pi)$. Además, como

$$p(A, \phi) = \frac{A}{2\pi\sigma_\theta^2} \exp \left(-\frac{A^2}{2\sigma_\theta^2} \right) = \frac{A}{\sigma_\theta^2} \exp \left(-\frac{A^2}{2\sigma_\theta^2} \right) \times \frac{1}{2\pi} = p(A)p(\phi),$$

se concluye que A y ϕ son variables aleatorias independientes.

11.6.7. Problema 7

En el modelo lineal clásico, el estimador MVU es también eficiente, por lo que el estimador de máxima verosimilitud también lo produce, o equivalentemente, el estimador MVU se obtiene maximizando

$$p(\mathbf{x}; \boldsymbol{\theta}) = \frac{1}{(2\pi)^{\frac{N}{2}} \det^{\frac{1}{2}}(\mathbf{C})} \exp \left[-\frac{1}{2}(\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T \mathbf{C}^{-1}(\mathbf{x} - \mathbf{H}\boldsymbol{\theta}) \right].$$

Para el modelo lineal bayesiano, el estimador MMSE es idéntico al estimador MAP. Argumentar porque en ausencia de información a priori, el estimador MMSE para el modelo lineal bayesiano tiene forma idéntica al estimador MVU para el modelo lineal clásico.

Solución Como se muestra en la sección 7.1.4, el estimador MLE produce el estimador MVU para el modelo lineal. Por lo tanto, los distintos estimadores son

$$\hat{\boldsymbol{\theta}}_{\text{MLE}} = \hat{\boldsymbol{\theta}}_{\text{MVU}} = \arg \max_{\boldsymbol{\theta}} p(\mathbf{x}; \boldsymbol{\theta}), \quad \hat{\boldsymbol{\theta}}_{\text{MMSE}} = E(\boldsymbol{\theta}|\mathbf{x}), \quad \hat{\boldsymbol{\theta}}_{\text{MAP}} = \arg \max_{\boldsymbol{\theta}} p(\boldsymbol{\theta}|\mathbf{x}).$$

Los estimadores MMSE y MAP son iguales. Esto se debe a que la PDF a posteriori $p(\boldsymbol{\theta}|\mathbf{x})$ es gaussiana, como se dedujo en la sección 10.4 del modelo lineal bayesiano, y por lo tanto, la media coincide con el máximo de la PDF. Por otro lado, la PDF a posteriori se puede expresar como

$$p(\boldsymbol{\theta}|\mathbf{x}) = \frac{p(\mathbf{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int p(\mathbf{x}|\boldsymbol{\theta})p(\boldsymbol{\theta}) d\boldsymbol{\theta}}.$$

Teniendo en cuenta que el denominador es una constante de normalización, el MAP se obtiene maximizando $p(\mathbf{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})$ y sin conocimiento a priori, es equivalente a maximizar $p(\mathbf{x}|\boldsymbol{\theta})$. Esto se debe a que sin conocimiento a priori, $p(\boldsymbol{\theta})$ tiende a ser uniforme y el máximo de la PDF condicional $p(\mathbf{x}|\boldsymbol{\theta})$ no es afectado por la PDF a priori $p(\boldsymbol{\theta})$. Como en el modelo lineal bayesiano se cumple que

$$p(\mathbf{x}|\boldsymbol{\theta}) = p(\mathbf{x}; \boldsymbol{\theta}) = \frac{1}{(2\pi)^{\frac{N}{2}} \det^{\frac{1}{2}}(\mathbf{C})} \exp \left[-\frac{1}{2}(\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^T \mathbf{C}^{-1}(\mathbf{x} - \mathbf{H}\boldsymbol{\theta}) \right],$$

maximizar $p(\mathbf{x}; \boldsymbol{\theta})$, que conduce al estimador MLE o MVU, es equivalente a maximizar $p(\mathbf{x}|\boldsymbol{\theta})$, que conduce al estimador MAP. Por lo tanto, el MLE y el MAP coinciden. Como el MMSE coincide con el MAP, se concluye que el estimador MMSE coincide con el estimador MLE.

11.6.8. Problema 8

Sea una parámetro $\boldsymbol{\theta}$ que cambia con el tiempo de acuerdo a la relación determinística

$$\boldsymbol{\theta}[n] = \mathbf{A}\boldsymbol{\theta}[n-1], \quad n \geq 1,$$

donde \mathbf{A} es una matriz $p \times p$ invertible conocida y $\boldsymbol{\theta}[0]$ es un parámetro desconocido que se modela como un vector aleatorio. Notar que una vez que $\boldsymbol{\theta}[0]$ es especificado, también lo es $\boldsymbol{\theta}[n]$ para $n \geq 1$. Probar que el estimador MMSE de $\boldsymbol{\theta}[n]$ es

$$\hat{\boldsymbol{\theta}}[n] = \mathbf{A}^n \hat{\boldsymbol{\theta}}[0]$$

donde $\hat{\boldsymbol{\theta}}[0]$ es el estimador MMSE de $\boldsymbol{\theta}[0]$, o equivalentemente,

$$\hat{\boldsymbol{\theta}}[n] = \mathbf{A} \hat{\boldsymbol{\theta}}[n-1].$$

Solución El estimador MMSE de $\boldsymbol{\theta}[n]$ es

$$\begin{aligned} \hat{\boldsymbol{\theta}}[n] &= E(\boldsymbol{\theta}[n]|\mathbf{x}) \\ &\stackrel{(a)}{=} E(\mathbf{A}^n \boldsymbol{\theta}[0]|\mathbf{x}) \\ &\stackrel{(b)}{=} \mathbf{A}^n E(\boldsymbol{\theta}[0]|\mathbf{x}) \\ &\stackrel{(c)}{=} \mathbf{A}^n \hat{\boldsymbol{\theta}}[0], \end{aligned}$$

donde en (a) se tuvo en cuenta que el estimador cumple que

$$\boldsymbol{\theta}[n] = \mathbf{A}^n \boldsymbol{\theta}[0],$$

en (b) se aplicó la propiedad de linealidad de la esperanza y en (c) que el estimador MMSE de $\boldsymbol{\theta}[0]$ es

$$\hat{\boldsymbol{\theta}}[0] = E(\boldsymbol{\theta}[0]|\mathbf{x}),$$

concluyendo la demostración.

11.6.9. Problema 9

Un vehículo parte desde una posición desconocida $(x[0], y[0])$ y se mueve a velocidad constante como

$$x[n] = x[0] + v_x n$$

$$y[n] = y[0] + v_y n$$

donde v_x y v_y son los componentes de la velocidad en las direcciones x e y respectivamente, ambos desconocidos. Se desea estimar la posición del vehículo en cada instante n así como la velocidad. A pesar de que la posición inicial $(x[0], y[0])$ y la velocidad $[v_x \ v_y]^T$ son desconocidas, pueden ser modeladas como un vector aleatorio. Mostrar que $\boldsymbol{\theta}[n] = [x[n] \ y[n] \ v_x \ v_y]^T$ puede ser estimado por el estimador MMSE

$$\hat{\boldsymbol{\theta}}[n] = \mathbf{A}^n \hat{\boldsymbol{\theta}}[0]$$

donde $\hat{\boldsymbol{\theta}}[0]$ es el estimador MMSE de $[x[n] \ y[n] \ v_x \ v_y]^T$. Determinar también \mathbf{A} .

Solución Considerando que

$$x[n-1] = x[0] + v_x(n-1)$$

$$y[n-1] = y[0] + v_y(n-1),$$

se cumple que

$$x[n] = x[n-1] + v_x$$

$$y[n] = y[n-1] + v_y.$$

Por lo tanto,

$$\underbrace{\begin{bmatrix} x[n] \\ y[n] \\ v_x \\ v_y \end{bmatrix}}_{\boldsymbol{\theta}[n]} = \begin{bmatrix} x[n-1] + v_x \\ y[n-1] + v_y \\ v_x \\ v_y \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} x[n-1] \\ y[n-1] \\ v_x \\ v_y \end{bmatrix}}_{\boldsymbol{\theta}[n-1]},$$

es decir, se cumple que

$$\boldsymbol{\theta}[n] = \mathbf{A} \boldsymbol{\theta}[n-1], \quad \text{con} \quad \mathbf{A} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

y del problema de la sección 11.6.8 se concluye que el MMSE de $\boldsymbol{\theta}[n]$ es

$$\hat{\boldsymbol{\theta}}[n] = \mathbf{A}^n \hat{\boldsymbol{\theta}}[0],$$

donde $\hat{\boldsymbol{\theta}}[0]$ es el estimador MMSE de $\boldsymbol{\theta}[0]$.

11.6.10. Problema 10

Mostrar que el estimador MAP del ejemplo de la sección 11.3.1 de la PDF exponencial es consistente en el sentido bayesiano, o que con $N \rightarrow \infty$, $\hat{\theta} \rightarrow \theta$ para cualquier realización θ del parámetro desconocido. Dar un argumento de porque esto debe ser cierto en general. Sugerencia: ver el ejemplo de la PDF exponencial en la sección 9.3.

Solución En el ejemplo de la sección 11.3.1 se obtuvo que el estimador MAP es

$$\hat{\theta} = \frac{1}{\bar{x} + \frac{\lambda}{N}}.$$

Por lo tanto

$$\lim_{N \rightarrow \infty} \hat{\theta} = \frac{1}{\bar{x}}.$$

Por otro lado, para una realización dada de θ , por la ley de los grandes números se cumple que

$$\lim_{N \rightarrow \infty} \bar{x} = E(x[n]) \stackrel{(a)}{=} \frac{1}{\theta},$$

donde en (a) se empleó el resultado obtenido en la sección 9.2.1. Combinando los resultados, se concluye que $\hat{\theta} \rightarrow \theta$ con $N \rightarrow \infty$ para cualquier realización del parámetro. Como se explica en el ejemplo de la sección 9.3, el estimador $1/\bar{x}$ es el estimador MLE, y cumple que su PDF asintótica es

$$1/\bar{x} \stackrel{a}{\sim} \mathcal{N}(\theta, \theta^2/N).$$

En general, con $N \rightarrow \infty$, la PDF condicional domina a la PDF a priori y el estimador MAP se obtiene maximizando $p(\mathbf{x}|\theta)$, que es el MLE bayesiano. Dada una realización de θ , se cumple que $p(\mathbf{x}|\theta) = p(\mathbf{x}; \theta)$, por lo que $\hat{\theta}$ puede pensarse como el MLE. Como el estimador MLE es consistente (ver la sección 7.1.2), también lo es el estimador MAP.

11.6.11. Problema 11

Se considera el estimador MAP vectorial, o

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta}} p(\boldsymbol{\theta}|\mathbf{x}).$$

Mostrar que este estimador minimiza el costo de Bayes para la función de costo

$$\mathcal{C}(\boldsymbol{\epsilon}) = \begin{cases} 1 & \|\boldsymbol{\epsilon}\| > \delta \\ 0 & \|\boldsymbol{\epsilon}\| < \delta, \end{cases}$$

donde $\boldsymbol{\epsilon} = \boldsymbol{\theta} - \hat{\boldsymbol{\theta}}$, $\|\boldsymbol{\epsilon}\|^2 = \sum_{i=1}^p \epsilon_i^2$, y $\delta \rightarrow 0$.

Solución El costo de Bayes, definido por la ecuación 11.1, es

$$\mathcal{R} = \int \left[\int \mathcal{C}(\boldsymbol{\epsilon}) p(\boldsymbol{\theta}|\mathbf{x}) d\boldsymbol{\theta} \right] p(\mathbf{x}) d\mathbf{x}.$$

Notar que en este caso, el parámetro $\boldsymbol{\theta}$ es un vector. Al igual que antes, el costo de Bayes se minimiza minimizando la integral interior para cada \mathbf{x} . La integral interior es

$$\begin{aligned} g(\hat{\boldsymbol{\theta}}) &= \int \mathcal{C}(\boldsymbol{\epsilon}) p(\boldsymbol{\theta}|\mathbf{x}) d\boldsymbol{\theta} \\ &= \int_{\|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\|^2 > \delta^2} p(\boldsymbol{\theta}|\mathbf{x}) d\boldsymbol{\theta} \\ &= 1 - \int_{\|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\|^2 < \delta^2} p(\boldsymbol{\theta}|\mathbf{x}) d\boldsymbol{\theta}, \end{aligned}$$

que se minimiza maximizando

$$\int_{\|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\|^2 < \delta^2} p(\boldsymbol{\theta}|\mathbf{x}) d\boldsymbol{\theta}.$$

Teniendo en cuenta que la región de integración $\|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\|^2 < \delta^2$ es una bola de centro $\hat{\boldsymbol{\theta}}$ y radio δ , con $\delta \rightarrow 0$, la integral se maximiza eligiendo

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta}} p(\boldsymbol{\theta}|\mathbf{x}).$$

11.6.12. Problema 12

Mostrar que el estimador MAP de

$$\boldsymbol{\alpha} = \mathbf{A}\boldsymbol{\theta}$$

es

$$\hat{\boldsymbol{\alpha}} = \mathbf{A}\hat{\boldsymbol{\theta}}$$

donde $\boldsymbol{\alpha}$ es un vector $p \times 1$ y \mathbf{A} es una matriz $p \times p$ invertible. En palabras, esto significa que el estimador MAP conmuta sobre transformaciones lineales invertibles.

Solución El estimador MAP de $\boldsymbol{\theta}$ es

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta}} p(\boldsymbol{\theta}|\mathbf{x}),$$

y como

$$p(\boldsymbol{\theta}|\mathbf{x}) = \frac{p(\mathbf{x}, \boldsymbol{\theta})}{p(\mathbf{x})}$$

y el denominador $p(\mathbf{x})$ no depende de el parámetro $\boldsymbol{\theta}$, equivalentemente

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta}} p(\mathbf{x}, \boldsymbol{\theta}).$$

De forma análoga, el MAP de $\boldsymbol{\alpha}$ está dado por

$$\hat{\boldsymbol{\alpha}} = \arg \max_{\boldsymbol{\alpha}} p(\mathbf{x}, \boldsymbol{\alpha}).$$

Se expresará $p(\mathbf{x}, \boldsymbol{\alpha})$ a partir de $p(\mathbf{x}, \boldsymbol{\theta})$. Sea $\boldsymbol{\alpha} = \mathbf{A}\boldsymbol{\theta}$. Como \mathbf{A} es una matriz invertible, la transformación es uno a uno con $\boldsymbol{\theta} = \mathbf{A}^{-1}\boldsymbol{\alpha}$, y por el teorema fundamental de transformación de variables aleatorias (ecuación 11.13), se cumple que

$$p_{\mathbf{x}, \boldsymbol{\alpha}}(\mathbf{x}, \boldsymbol{\alpha}) = \frac{p_{\mathbf{x}, \boldsymbol{\theta}}(\mathbf{x}, \boldsymbol{\theta})}{|\det[J(\boldsymbol{\theta})]|} \Big|_{\boldsymbol{\theta}=\mathbf{A}^{-1}\boldsymbol{\alpha}}$$

donde la matriz jacobiana de la transformación es en este caso

$$J(\boldsymbol{\theta}) = \frac{\partial \boldsymbol{\alpha}}{\partial \boldsymbol{\theta}} = \frac{\partial \mathbf{A}\boldsymbol{\theta}}{\partial \boldsymbol{\theta}} = \mathbf{A}.$$

Por lo tanto,

$$p_{\mathbf{x}, \boldsymbol{\alpha}}(\mathbf{x}, \boldsymbol{\alpha}) = \frac{p_{\mathbf{x}, \boldsymbol{\theta}}(\mathbf{x}, \mathbf{A}^{-1}\boldsymbol{\alpha})}{|\det \mathbf{A}|}.$$

Como el denominador no depende de $\boldsymbol{\alpha}$, el MAP de $\boldsymbol{\alpha}$ se obtiene como

$$\hat{\boldsymbol{\alpha}} = \arg \max_{\boldsymbol{\alpha}} p_{\mathbf{x}, \boldsymbol{\theta}}(\mathbf{x}, \mathbf{A}^{-1}\boldsymbol{\alpha}).$$

Por lo tanto, se cumple que

$$\begin{aligned} \hat{\boldsymbol{\alpha}} &: p_{\mathbf{x}, \boldsymbol{\theta}}(\mathbf{x}, \mathbf{A}^{-1}\hat{\boldsymbol{\alpha}}) \geq p_{\mathbf{x}, \boldsymbol{\theta}}(\mathbf{x}, \mathbf{A}^{-1}\boldsymbol{\alpha}) \quad \forall \boldsymbol{\alpha} \\ \hat{\boldsymbol{\theta}} &: p_{\mathbf{x}, \boldsymbol{\theta}}(\mathbf{x}, \hat{\boldsymbol{\theta}}) \geq p_{\mathbf{x}, \boldsymbol{\theta}}(\mathbf{x}, \boldsymbol{\theta}) \quad \forall \boldsymbol{\theta}. \end{aligned}$$

Como $\mathbf{A}^{-1}\hat{\boldsymbol{\alpha}}$ y $\hat{\boldsymbol{\theta}}$ maximizan la misma función, se debe cumplir que

$$\mathbf{A}^{-1}\hat{\boldsymbol{\alpha}} = \hat{\boldsymbol{\theta}} \quad \Rightarrow \quad \hat{\boldsymbol{\alpha}} = \mathbf{A}\hat{\boldsymbol{\theta}}.$$

11.6.13. Problema 13

Si \mathbf{x} es un vector aleatorio 2×1 con PDF

$$\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}),$$

probar que la PDF de $\mathbf{x}^T \mathbf{C}^{-1} \mathbf{x}$ es $\chi^2(2)$. Sugerencia: notar que \mathbf{C}^{-1} puede ser factorizada como $\mathbf{D}^T \mathbf{D}$, donde \mathbf{D} es una transformación de blanqueado como se describió en la sección 4.5.

Solución Como se explicó en la sección 4.5, mediante la descomposición de Cholesky, la matriz \mathbf{C}^{-1} puede factorizarse como

$$\mathbf{C}^{-1} = \mathbf{D}^T \mathbf{D},$$

donde \mathbf{D} es una matriz invertible 2×2 . Sea \mathbf{y} el vector aleatorio 2×1

$$\mathbf{y} = \mathbf{D}\mathbf{x}.$$

La matriz \mathbf{D} actúa como una transformación de blanqueado sobre \mathbf{x} . Efectivamente, por un lado, el vector \mathbf{y} tiene PDF gaussiana por tratarse de una transformación lineal de un vector gaussiano. Además, la esperanza de \mathbf{y} es

$$E(\mathbf{y}) = \mathbf{D}E(\mathbf{x}) = \mathbf{0}$$

con matriz de covarianza

$$\begin{aligned} \mathbf{C}_y &= E(\mathbf{y}\mathbf{y}^T) = E[(\mathbf{D}\mathbf{x})(\mathbf{D}\mathbf{x})^T] = E(\mathbf{D}\mathbf{x}\mathbf{x}^T\mathbf{D}^T) = \mathbf{D}E(\mathbf{x}\mathbf{x}^T)\mathbf{D}^T = \mathbf{D}\mathbf{C}\mathbf{D}^T = \mathbf{D}(\mathbf{D}^T\mathbf{D})^{-1}\mathbf{D}^T \\ &= \mathbf{D}\mathbf{D}^{-1}\mathbf{D}^{T-1}\mathbf{D}^T = \mathbf{I}. \end{aligned}$$

Se concluye que

$$\mathbf{y} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}).$$

Si $\mathbf{y} = [y_1 \ y_2]^T$, y_1 e y_2 son variables aleatorias escalares gaussianas no correlacionadas, por ser diagonal la matriz de covarianza de \mathbf{y} , con $y_1 \sim \mathcal{N}(0, 1)$ y $y_2 \sim \mathcal{N}(0, 1)$. Como en el caso de variables aleatorias gaussianas la no correlación implica independencia, y_1 e y_2 son independientes. Teniendo esto en cuenta,

$$\mathbf{x}^T \mathbf{C}^{-1} \mathbf{x} = \mathbf{x}^T \mathbf{D}^T \mathbf{D} \mathbf{x} = (\mathbf{D}\mathbf{x})^T \mathbf{D} \mathbf{x} = \mathbf{y}^T \mathbf{y} = y_1^2 + y_2^2 \sim \chi^2(2),$$

ya que la suma de los cuadrados de n variables aleatorias $\mathcal{N}(0, 1)$ tiene PDF $\chi^2(n)$ (ver el capítulo 7 de [2]).

11.6.14. Problema 14

Si $\boldsymbol{\epsilon}$ es un vector aleatorio 2×1 y $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{M}_{\hat{\theta}})$, graficar la elipse de error para $P = 0.9$ para los siguientes casos:

$$\text{a. } \mathbf{M}_{\hat{\theta}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{b. } \mathbf{M}_{\hat{\theta}} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \quad \text{c. } \mathbf{M}_{\hat{\theta}} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

Solución La elipse de error es (ecuación 11.9)

$$\boldsymbol{\epsilon}^T \mathbf{M}_{\hat{\theta}}^{-1} \boldsymbol{\epsilon} = c^2.$$

y para una probabilidad P ,

$$c^2 = 2 \ln \frac{1}{1-P}.$$

Con $P = 0.9$,

$$c^2 = 2 \ln \frac{1}{1-0.9} = 2 \ln 10 \approx 4.61 \quad \Rightarrow \quad c \approx 2.15.$$

a. En este caso

$$\mathbf{M}_{\hat{\theta}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \Rightarrow \mathbf{M}_{\hat{\theta}}^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Por lo tanto,

$$\boldsymbol{\epsilon}^T \mathbf{M}_{\hat{\theta}}^{-1} \boldsymbol{\epsilon} = \boldsymbol{\epsilon}^T \boldsymbol{\epsilon} = \epsilon_a^2 + \epsilon_b^2$$

La elipse de error es entonces

$$\epsilon_a^2 + \epsilon_b^2 = c^2,$$

que en este caso es un círculo de radio c y se muestra en la figura 11.9.

b. Ahora se tiene que

$$\mathbf{M}_{\hat{\theta}} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \Rightarrow \mathbf{M}_{\hat{\theta}}^{-1} = \frac{1}{2} \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix},$$

y

$$\boldsymbol{\epsilon}^T \mathbf{M}_{\hat{\theta}}^{-1} \boldsymbol{\epsilon} = [\epsilon_{\hat{a}} \quad \epsilon_{\hat{b}}] \frac{1}{2} \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \epsilon_{\hat{a}} \\ \epsilon_{\hat{b}} \end{bmatrix} = \frac{1}{2} [\epsilon_{\hat{a}} \quad \epsilon_{\hat{b}}] \begin{bmatrix} 2\epsilon_{\hat{a}} \\ \epsilon_{\hat{b}} \end{bmatrix} = \epsilon_a^2 + \frac{\epsilon_b^2}{2},$$

por lo que la elipse de error es

$$\epsilon_a^2 + \frac{\epsilon_b^2}{2} = c^2,$$

y dividiendo entre c^2 se puede escribir en la forma estándar (ver la ecuación I.1) como

$$\frac{\epsilon_a^2}{c^2} + \frac{\epsilon_b^2}{2c^2} = 1.$$

Esto es una elipse estándar con $a^2 = c^2$ y $b^2 = 2c^2$. Está centrada en el origen y tiene semieje-mayor $b = \sqrt{2c^2} = \sqrt{2}c$ en el eje vertical y semi-eje menor $a = \sqrt{c^2} = c$ en el eje horizontal. Los focos están situados en el eje vertical a una distancia $\pm\sqrt{b^2 - a^2} = \pm\sqrt{2c^2 - c^2} = \pm c$ del origen, es decir, sus coordenadas son $F_{1,2} = (0, \pm c)$ (ver el apéndice I). La gráfica se muestra en la figura 11.9.

c. En este caso

$$\mathbf{M}_{\hat{\theta}} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \Rightarrow \mathbf{M}_{\hat{\theta}}^{-1} = \frac{1}{3} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix},$$

y

$$\boldsymbol{\epsilon}^T \mathbf{M}_{\hat{\theta}}^{-1} \boldsymbol{\epsilon} = [\epsilon_{\hat{a}} \quad \epsilon_{\hat{b}}] \frac{1}{3} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} \epsilon_{\hat{a}} \\ \epsilon_{\hat{b}} \end{bmatrix} = \frac{1}{3} [\epsilon_{\hat{a}} \quad \epsilon_{\hat{b}}] \begin{bmatrix} 2\epsilon_{\hat{a}} - \epsilon_{\hat{b}} \\ -\epsilon_{\hat{a}} + 2\epsilon_{\hat{b}} \end{bmatrix} = \frac{1}{3} (2\epsilon_a^2 - \epsilon_a\epsilon_b - \epsilon_a\epsilon_b + 2\epsilon_b^2),$$

por lo que la elipse de error es

$$\frac{2}{3} (\epsilon_a^2 - \epsilon_a\epsilon_b + \epsilon_b^2) = c^2 \quad \text{o} \quad \epsilon_a^2 - \epsilon_a\epsilon_b + \epsilon_b^2 - \frac{3c^2}{2} = 0.$$

Comparando con la ecuación I.7, que es la ecuación general implícita de una elipse, los coeficientes son en este caso $A = C = 1$, $B = -1$, $D = E = 0$ y $F = -3c^2/2$. A partir de las ecuaciones I.8 con $D = E = 0$, se obtiene que los semi-ejes mayor y menor son

$$\begin{aligned} a, b &= \frac{-\sqrt{2(B^2 - 4AC)F} \left[(A + C) \pm \sqrt{(A - C)^2 + B^2} \right]}{B^2 - 4AC} \\ &= \frac{-\sqrt{2(1 - 4) \left(\frac{-3c^2}{2} \right)} (2 \pm \sqrt{1})}{4 - 1} \\ &= \frac{1}{3} \sqrt{9c^2(2 \pm 1)} \\ &= c\sqrt{(2 \pm 1)}, \end{aligned}$$

resultando en que

$$a = \sqrt{3}c \quad \text{y} \quad b = c.$$

Además, las coordenadas del centro son

$$x_0 = \frac{2CD - BE}{B^2 - 4AC} = 0 \quad \text{y} \quad y_0 = \frac{2AE - BD}{B^2 - 4AC} = 0$$

y el ángulo de rotación es

$$\theta = \arctan \left[\frac{1}{B} \left(C - A - \sqrt{(A - C)^2 + B^2} \right) \right] = \arctan \left[- \left(-\sqrt{1} \right) \right] = \arctan(1) = \frac{\pi}{4}.$$

Se concluye que se trata de una elipse centrada en el origen rotada un ángulo $\theta = \pi/4$ con semi-eje mayor $a = \sqrt{3}c$ y semi-eje menor $b = c$. Los focos están en la recta de ángulo $\pi/4$ a una distancia $\sqrt{a^2 - b^2} = \sqrt{2}c$ del origen, es decir, tienen coordenadas $F_{1,2} = (\pm c, \pm c)$. La gráfica se muestra en la figura 11.9. A partir de la ecuación I.5 de una elipse rotada centrada en el origen y considerando que $\cos \pi/4 = \sin \pi/4 = \sqrt{2}/2$, la elipse se puede escribir como

$$\frac{\frac{1}{2}(\epsilon_{\hat{a}} + \epsilon_{\hat{b}})^2}{3c^2} + \frac{\frac{1}{2}(-\epsilon_{\hat{a}} + \epsilon_{\hat{b}})^2}{c^2} = 1 \quad \text{o} \quad \frac{(\epsilon_{\hat{a}} + \epsilon_{\hat{b}})^2}{6c^2} + \frac{(\epsilon_{\hat{a}} - \epsilon_{\hat{b}})^2}{2c^2} = 1.$$

Para graficarla, conviene emplear la ecuación I.6 de la forma paramétrica, que en este caso queda

$$\begin{aligned} x(t) &= \frac{\sqrt{2}}{2}c(\sqrt{3}\cos t - \sin t) \\ y(t) &= \frac{\sqrt{2}}{2}c(\sqrt{3}\cos t + \sin t) \end{aligned}, \quad \text{para} \quad 0 \leq t < 2\pi.$$

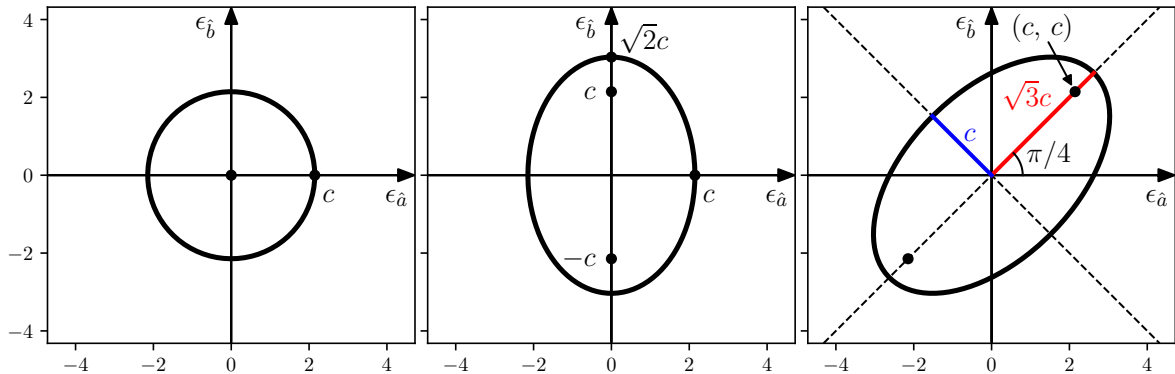


Figura 11.9: Elipses de error para una probabilidad $P = 0.9$ para los errores con matrices de covarianza dadas por a , b , y c , respectivamente.

11.6.15. Problema 15

Como caso particular del modelo lineal bayesiano, asúmase que

$$\mathbf{x} = \boldsymbol{\theta} + \mathbf{w}$$

donde $\boldsymbol{\theta}$ es un vector aleatorio 2×1 y

$$\mathbf{C}_w = \sigma^2 \mathbf{I}, \quad \mathbf{C}_\theta = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}.$$

Graficar las elipses de error si $\sigma_1^2 = \sigma_2^2$ y $\sigma_2^2 > \sigma_1^2$. ¿Qué ocurre si $\sigma_2^2 \gg \sigma_1^2$?

Solución El modelo de los datos se puede escribir como

$$\mathbf{x} = \mathbf{I}\boldsymbol{\theta} + \mathbf{w},$$

que es el modelo lineal bayesiano con $\mathbf{H} = \mathbf{I}$. De la ecuación 11.8 con $\mathbf{H} = \mathbf{I}$, la matriz de covarianza del error es

$$\mathbf{M}_{\hat{\boldsymbol{\theta}}} = (\mathbf{C}_{\theta}^{-1} + \mathbf{C}_w^{-1})^{-1},$$

y por lo tanto,

$$\mathbf{M}_{\hat{\boldsymbol{\theta}}}^{-1} = (\mathbf{C}_{\theta}^{-1} + \mathbf{C}_w^{-1}) = \begin{bmatrix} \frac{1}{\sigma_1^2} & 0 \\ 0 & \frac{1}{\sigma_2^2} \end{bmatrix} + \begin{bmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{\sigma^2} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sigma_1^2} + \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{\sigma_2^2} + \frac{1}{\sigma^2} \end{bmatrix}.$$

Con $\boldsymbol{\epsilon} = [\epsilon_1 \ \epsilon_2]^T$, se tiene que

$$\boldsymbol{\epsilon}^T \mathbf{M}_{\hat{\boldsymbol{\theta}}}^{-1} \boldsymbol{\epsilon} = \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma^2} \right) \epsilon_1^2 + \left(\frac{1}{\sigma_2^2} + \frac{1}{\sigma^2} \right) \epsilon_2^2,$$

por lo que la elipse de error para una probabilidad P es

$$\left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma^2} \right) \epsilon_1^2 + \left(\frac{1}{\sigma_2^2} + \frac{1}{\sigma^2} \right) \epsilon_2^2 = 2 \ln \frac{1}{1-P},$$

o

$$\frac{\frac{1}{\sigma_1^2} + \frac{1}{\sigma^2}}{2 \ln \frac{1}{1-P}} \epsilon_1^2 + \frac{\frac{1}{\sigma_2^2} + \frac{1}{\sigma^2}}{2 \ln \frac{1}{1-P}} \epsilon_2^2 = 1,$$

que es la ecuación de una elipse estándar centrada en el origen (ecuación I.1) con

$$a^2 = \frac{2 \ln \frac{1}{1-P}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma^2}} \quad \text{y} \quad b^2 = \frac{2 \ln \frac{1}{1-P}}{\frac{1}{\sigma_2^2} + \frac{1}{\sigma^2}}.$$

En el caso en que $\sigma_1^2 = \sigma_2^2$, $a^2 = b^2$, por lo que la superficie de error es un círculo de radio

$$a = \sqrt{\frac{2 \ln \frac{1}{1-P}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma^2}}},$$

como se muestra en la figura 11.10. En el caso en que $\sigma_2^2 > \sigma_1^2$, $b^2 > a^2$ y la superficie de error es una elipse con semi-eje mayor de largo b en el eje vertical y semi-eje menor de largo a . Finalmente, si $\sigma_2^2 \gg \sigma_1^2$, observando que

$$\lim_{\sigma_1^2 \rightarrow 0} a^2 = 0 \quad \text{y} \quad \lim_{\sigma_2^2 \rightarrow \infty} b^2 = 2\sigma^2 \ln \frac{1}{1-P},$$

la superficie de error es una elipse con semi-ejes de largo a y b con $b \gg a$, indicando que la incertidumbre se encuentra fundamentalmente en el eje vertical ϵ_2 , como se muestra en la figura 11.10.

11.6.16. Problema 16

En el ajuste de una recta a través de datos experimentales se asume el modelo

$$x[n] = A + Bn + w[n], \quad -M \leq n \leq M,$$

donde $w[n]$ es WGN con varianza σ^2 . Si el conocimiento a priori de la pendiente B y la intersección A es

$$\begin{bmatrix} A \\ B \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} A_0 \\ B_0 \end{bmatrix}, \begin{bmatrix} \sigma_A^2 & 0 \\ 0 & \sigma_B^2 \end{bmatrix} \right)$$

encontrar el estimador MMSE de A y B así como el MSE bayesiano mínimo. Asumir que A y B son independientes de $w[n]$. ¿Qué parámetro se beneficia mas del conocimiento a priori?

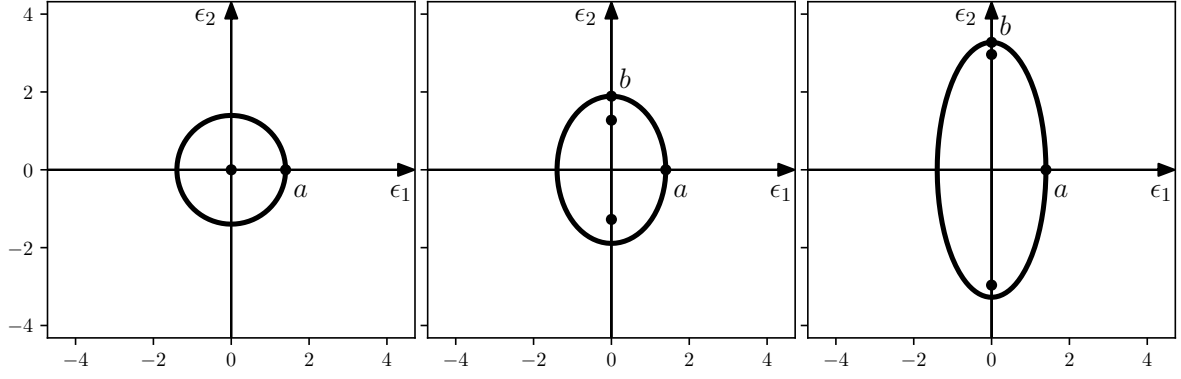


Figura 11.10: Elipses de error al variar σ_2^2 . En los tres casos, $P = 0.9$, $\sigma^2 = 10$ y $\sigma_1^2 = 1$. Además, $\sigma_2^2 = 1, 2$ y 10 respectivamente. Como σ_1^2 es igual en los tres casos, el semi-eje horizontal a de las elipses tiene el mismo largo. Al incrementar σ_2^2 , crece el semi-eje vertical b .

Solución El modelo de datos es el modelo lineal bayesiano

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}, \quad \text{con} \quad \mathbf{H} = \begin{bmatrix} 1 & -M \\ 1 & -(M-1) \\ \vdots & \vdots \\ 1 & M \end{bmatrix}, \quad \boldsymbol{\theta} = \begin{bmatrix} A \\ B \end{bmatrix} \quad \text{y} \quad \mathbf{C}_w = \sigma^2 \mathbf{I},$$

y el estimador MMSE y su matriz de covarianza están dados por las ecuaciones 10.19 y 10.20. La matriz de covarianza del estimador es

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}} = (\mathbf{C}_{\boldsymbol{\theta}}^{-1} + \mathbf{H}^T \mathbf{C}_w^{-1} \mathbf{H})^{-1}.$$

Por un lado, se tiene que

$$\mathbf{H}^T \mathbf{C}_w^{-1} \mathbf{H} = \frac{1}{\sigma^2} \mathbf{H}^T \mathbf{H} = \frac{1}{\sigma^2} \begin{bmatrix} 1 & 1 & \dots & 1 \\ -M & -(M-1) & \dots & M \end{bmatrix} \begin{bmatrix} 1 & -M \\ 1 & -(M-1) \\ \vdots & \vdots \\ 1 & M \end{bmatrix} = \frac{1}{\sigma^2} \begin{bmatrix} \sum_{n=-M}^M 1 & \sum_{n=-M}^M n \\ \sum_{n=-M}^M n & \sum_{n=-M}^M n^2 \end{bmatrix},$$

resultando en que

$$\mathbf{H}^T \mathbf{C}_w^{-1} \mathbf{H} = \frac{1}{\sigma^2} \begin{bmatrix} \frac{2M+1}{\sigma^2} & 0 \\ 0 & \frac{\sum_{n=-M}^M n^2}{\sigma^2} \end{bmatrix}.$$

Luego,

$$\mathbf{C}_{\boldsymbol{\theta}}^{-1} + \mathbf{H}^T \mathbf{C}_w^{-1} \mathbf{H} = \begin{bmatrix} \frac{1}{\sigma_A^2} & 0 \\ 0 & \frac{1}{\sigma_B^2} \end{bmatrix} + \begin{bmatrix} \frac{2M+1}{\sigma^2} & 0 \\ 0 & \frac{\sum_{n=-M}^M n^2}{\sigma^2} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sigma_A^2} + \frac{2M+1}{\sigma^2} & 0 \\ 0 & \frac{1}{\sigma_B^2} + \frac{\sum_{n=-M}^M n^2}{\sigma^2} \end{bmatrix}$$

y por lo tanto, la matriz de covarianza del estimador es

$$\mathbf{C}_{\hat{\theta}} = (\mathbf{C}_{\theta}^{-1} + \mathbf{H}^T \mathbf{C}_w^{-1} \mathbf{H})^{-1} = \begin{bmatrix} \frac{1}{\frac{1}{\sigma_A^2} + \frac{2M+1}{\sigma^2}} & 0 \\ 0 & \frac{1}{\frac{1}{\sigma_B^2} + \frac{\sum_{n=-M}^M n^2}{\sigma^2}} \end{bmatrix}. \quad (11.14)$$

El estimador es

$$\hat{\boldsymbol{\theta}} = \boldsymbol{\mu}_{\theta} + (\mathbf{C}_{\theta}^{-1} + \mathbf{H}^T \mathbf{C}_w^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C}_w^{-1} (\mathbf{x} - \mathbf{H} \boldsymbol{\mu}_{\theta}).$$

Como

$$\mathbf{H}^T \mathbf{C}_w^{-1} (\mathbf{x} - \mathbf{H} \boldsymbol{\mu}_{\theta}) = \frac{1}{\sigma^2} (\mathbf{H}^T \mathbf{x} - \mathbf{H}^T \mathbf{H} \boldsymbol{\mu}_{\theta}),$$

con

$$\mathbf{H}^T \mathbf{x} = \begin{bmatrix} 1 & \cdots & 1 \\ -M & \cdots & M \end{bmatrix} \begin{bmatrix} x[-M] \\ \vdots \\ x[M] \end{bmatrix} = \begin{bmatrix} \sum_{n=-M}^M x[n] \\ \sum_{n=-M}^M nx[n] \end{bmatrix}$$

y

$$\mathbf{H}^T \mathbf{H} \boldsymbol{\mu}_{\theta} = \begin{bmatrix} \frac{2M+1}{\sigma^2} & 0 \\ 0 & \frac{\sum_{n=-M}^M n^2}{\sigma^2} \end{bmatrix} \begin{bmatrix} A_0 \\ B_0 \end{bmatrix} = \begin{bmatrix} A_0(2M+1) \\ B_0 \sum_{n=-M}^M n^2 \end{bmatrix}$$

se obtiene que

$$\mathbf{H}^T \mathbf{C}_w^{-1} (\mathbf{x} - \mathbf{H} \boldsymbol{\mu}_{\theta}) = \frac{1}{\sigma^2} \begin{bmatrix} \sum_{n=-M}^M x[n] - A_0(2M+1) \\ \sum_{n=-M}^M nx[n] - B_0 \sum_{n=-M}^M n^2 \end{bmatrix} = \frac{1}{\sigma^2} \begin{bmatrix} \bar{x}(2M+1) - A_0(2M+1) \\ \sum_{n=-M}^M nx[n] - B_0 \sum_{n=-M}^M n^2 \end{bmatrix},$$

que se puede escribir como

$$\mathbf{H}^T \mathbf{C}_w^{-1} (\mathbf{x} - \mathbf{H} \boldsymbol{\mu}_{\theta}) = \begin{bmatrix} \frac{2M+1}{\sigma^2} (\bar{x} - A_0) \\ \frac{\sum_{n=-M}^M n^2}{\sigma^2} \left(\frac{\sum_{n=-M}^M nx[n]}{\sum_{n=-M}^M n^2} - B_0 \right) \end{bmatrix}.$$

Combinando los resultados, el estimador es

$$\boldsymbol{\theta} = \begin{bmatrix} \hat{A} \\ \hat{B} \end{bmatrix} = \begin{bmatrix} A_0 \\ B_0 \end{bmatrix} + \begin{bmatrix} \frac{1}{\frac{1}{\sigma_A^2} + \frac{2M+1}{\sigma^2}} & 0 \\ 0 & \frac{1}{\frac{1}{\sigma_B^2} + \frac{\sum_{n=-M}^M n^2}{\sigma^2}} \end{bmatrix} \begin{bmatrix} \frac{2M+1}{\sigma^2} (\bar{x} - A_0) \\ \frac{\sum_{n=-M}^M n^2}{\sigma^2} \left(\frac{\sum_{n=-M}^M nx[n]}{\sum_{n=-M}^M n^2} - B_0 \right) \end{bmatrix},$$

es decir

$$\hat{A} = A_0 + \frac{\frac{2M+1}{\sigma_A^2}}{\frac{1}{\sigma_A^2} + \frac{2M+1}{\sigma^2}} (\bar{x} - A_0)$$

$$\hat{B} = B_0 + \frac{\frac{\sum_{n=-M}^M n^2}{\sigma^2}}{\frac{1}{\sigma_B^2} + \frac{\sum_{n=-M}^M n^2}{\sigma^2}} \left(\frac{\sum_{n=-M}^M nx[n]}{\sum_{n=-M}^M n^2} - B_0 \right).$$

El MSE bayesiano mínimo es

$$\text{Bmse}(\hat{A}) = [\mathbf{C}_{\hat{\theta}}]_{11} = \frac{1}{\frac{1}{\sigma_A^2} + \frac{2M+1}{\sigma^2}}$$

$$\text{Bmse}(\hat{B}) = [\mathbf{C}_{\hat{\theta}}]_{22} = \frac{1}{\frac{1}{\sigma_B^2} + \frac{\sum_{n=-M}^M n^2}{\sigma^2}}.$$

Se observa que el MSE bayesiano mínimo decrece como $1/(2M+1)$ para el estimador \hat{A} del punto de intersección y como $1/\sum n^2$ para el estimador \hat{B} de la pendiente. El estimador \hat{A} se beneficia mas del conocimiento a priori, ya que el incremento de la cantidad de datos reduce mas el $\text{Bmse}(\hat{B})$ que el $\text{Bmse}(\hat{A})$ haciendo que el conocimiento a priori de B tenga menos influencia en el estimador que el conocimiento a priori de A .

11.6.17. Problema 17

Si $s[n] = A + Bn$ para $n = -M, -(M-1), \dots, M$ y \hat{A} y \hat{B} son los estimadores MMSE de A y B respectivamente, encontrar el estimador MMSE $\hat{\mathbf{s}}$ de $\mathbf{s} = [s[-M], s[-(M-1)], \dots, s[M]]^T$. Luego, encontrar la PDF del error $\boldsymbol{\epsilon} = \mathbf{s} - \hat{\mathbf{s}}$ si el modelo de los datos es el del problema de la sección 11.6.16. ¿Cómo varía el error con n y porqué?

Solución Por definición, los estimadores MMSE de A y B son respectivamente

$$\hat{A} = E(A|\mathbf{x}) \quad \text{y} \quad \hat{B} = E(B|\mathbf{x}).$$

Considerando que $s[n] = A + Bn$, el estimador MMSE de $s[n]$ es

$$\hat{s}[n] = E(s[n]|\mathbf{x}) = E(A + Bn|\mathbf{x}) = E(A|\mathbf{x}) + E(B|\mathbf{x})n = \hat{A} + \hat{B}n,$$

que en forma vectorial se expresa como

$$\hat{\mathbf{s}} = \mathbf{H}\hat{\boldsymbol{\theta}}, \quad \text{con} \quad \hat{\boldsymbol{\theta}} = \begin{bmatrix} \hat{A} \\ \hat{B} \end{bmatrix} \quad \text{y} \quad \mathbf{H} = \begin{bmatrix} 1 & -M \\ 1 & -(M-1) \\ \vdots & \vdots \\ 1 & M \end{bmatrix}.$$

El error de estimación es

$$\boldsymbol{\epsilon} = \mathbf{s} - \hat{\mathbf{s}} = \mathbf{H}\boldsymbol{\theta} - \mathbf{H}\hat{\boldsymbol{\theta}} = \mathbf{H}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}).$$

Considerando que $\hat{\boldsymbol{\theta}}$ es un vector gaussiano por ser una combinación lineal de los elementos del vector gaussiano \mathbf{x} (ver la ecuación 10.19) y $\boldsymbol{\theta}$ también es gaussiano, $\boldsymbol{\epsilon}$ es gaussiano, por lo que resta calcular su media y su matriz de covarianza. La media es

$$E(\boldsymbol{\epsilon}) = \mathbf{H}E(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) = \mathbf{H}[E(\boldsymbol{\theta}) - E(\hat{\boldsymbol{\theta}})] \stackrel{(a)}{=} \mathbf{H}(\boldsymbol{\mu}_{\boldsymbol{\theta}} - \boldsymbol{\mu}_{\boldsymbol{\theta}}) = \mathbf{0},$$

donde en (a) se consideró que de la ecuación 10.19,

$$E(\hat{\boldsymbol{\theta}}) = \boldsymbol{\mu}_{\boldsymbol{\theta}} + (\mathbf{C}_{\boldsymbol{\theta}}^{-1} + \mathbf{H}^T \mathbf{C}_w^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C}_w^{-1} (E(\mathbf{x}) - \mathbf{H} \boldsymbol{\mu}_{\boldsymbol{\theta}})$$

y

$$E(\mathbf{x}) = E(\mathbf{H}\boldsymbol{\theta} + \mathbf{w}) = \mathbf{H}E(\boldsymbol{\theta}) = \mathbf{H}\boldsymbol{\mu}_{\boldsymbol{\theta}},$$

por lo que

$$E(\hat{\boldsymbol{\theta}}) = \boldsymbol{\mu}_{\boldsymbol{\theta}}.$$

Notar que la esperanza se toma respecto a $p(\mathbf{x}, \boldsymbol{\theta})$. La matriz de covarianza del error es

$$E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T) = E\{\mathbf{H}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})[\mathbf{H}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})]^T\} = E[\mathbf{H}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \mathbf{H}^T] = \mathbf{H}E[(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T] \mathbf{H}^T$$

resultando en

$$E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T) = \mathbf{H}\mathbf{M}_{\hat{\boldsymbol{\theta}}}\mathbf{H}^T,$$

donde $\mathbf{M}_{\hat{\boldsymbol{\theta}}}$ es la matriz de covarianza del error $\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}$, que se calculó en el problema 11.6.16 y está dada por la ecuación 11.14. Definiendo el vector $\mathbf{m} = [-M \quad -(M-1) \quad \dots \quad M]^T$ de forma de que $\mathbf{H} = [\mathbf{1} \quad \mathbf{m}]$, se tiene que

$$\begin{aligned} E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T) &= \mathbf{H}\mathbf{M}_{\hat{\boldsymbol{\theta}}}\mathbf{H}^T \\ &= [\mathbf{1} \quad \mathbf{m}] \begin{bmatrix} \frac{1}{\sigma_A^2 + \frac{2M+1}{\sigma^2}} & 0 \\ 0 & \frac{1}{\frac{1}{\sigma_B^2} + \frac{\sum_{n=-M}^M n^2}{\sigma^2}} \end{bmatrix} \begin{bmatrix} \mathbf{1}^T \\ \mathbf{m}^T \end{bmatrix} \\ &= [\mathbf{1} \quad \mathbf{m}] \begin{bmatrix} \frac{1}{\sigma_A^2 + \frac{2M+1}{\sigma^2}} \mathbf{1}^T \\ \frac{1}{\frac{1}{\sigma_B^2} + \frac{\sum_{n=-M}^M n^2}{\sigma^2}} \mathbf{m}^T \end{bmatrix} \\ &= \frac{1}{\frac{1}{\sigma_A^2} + \frac{2M+1}{\sigma^2}} \mathbf{1}\mathbf{1}^T + \frac{1}{\frac{1}{\sigma_B^2} + \frac{\sum_{n=-M}^M n^2}{\sigma^2}} \mathbf{m}\mathbf{m}^T. \end{aligned}$$

Definiendo $\mathbf{M}_{\hat{\mathbf{s}}} = E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T)$ y considerando que el elemento i -ésimo de \mathbf{m} es $[\mathbf{m}]_i = i - (M+1)$ con $i = 1, \dots, 2M+1$, se tiene que

$$[\mathbf{M}_{\hat{\mathbf{s}}}]_{ij} = \frac{1}{\frac{1}{\sigma_A^2} + \frac{2M+1}{\sigma^2}} + \frac{[i - (M+1)][j - (M+1)]}{\frac{1}{\sigma_B^2} + \frac{\sum_{n=-M}^M n^2}{\sigma^2}}$$

por lo que el MSE bayesiano mínimo es

$$\text{Bsme}(\hat{s}[n]) = [\mathbf{M}_{\hat{s}}]_{n+(M+1), n+(M+1)} = \frac{1}{\frac{1}{\sigma_A^2} + \frac{2M+1}{\sigma^2}} + \frac{n^2}{\frac{1}{\sigma_B^2} + \frac{\sum_{n=-M}^M n^2}{\sigma^2}}$$

y se muestra en la figura 11.11. Se observa que el error crece con n . Esto es porque $\hat{s}[n] = \hat{A} + \hat{B}n$ y por lo tanto,

$$\epsilon[n] = s[n] - \hat{s}[n] = (A - \hat{A}) + (B - \hat{B})n,$$

es decir, el error de la estimación de B se amplifica con n .

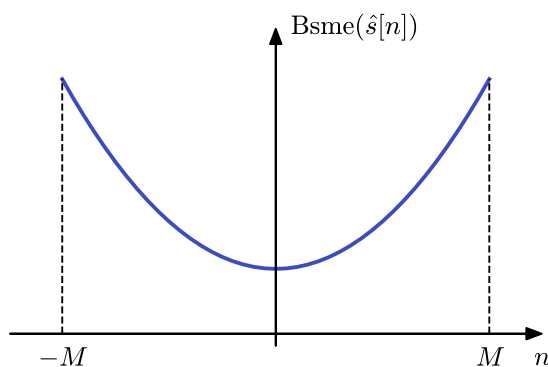


Figura 11.11: Dependencia del MSE bayesiano mínimo con n .

11.6.18. Problema 18

En este problema se deriva el filtro de Wiener no causal para la estimación de una señal en WGN. Se hace extendiendo los resultados de la sección 11.5. Allí, el estimador MMSE de la señal está dado por la ecuación 11.11. Mostrar primero que

$$(\mathbf{C}_s + \sigma^2 \mathbf{I})\hat{\mathbf{s}} = \mathbf{C}_s \mathbf{x}.$$

Luego, asumir la señal es WSS y por lo tanto

$$\begin{aligned} [\mathbf{C}_s + \sigma^2 \mathbf{I}]_{ij} &= r_{ss}[i-j] + \sigma^2 \delta[i-j] = r_{xx}[i-j] \\ [\mathbf{C}_s]_{ij} &= r_{ss}[i-j]. \end{aligned}$$

Como consecuencia, las ecuaciones a resolver para $\hat{\mathbf{s}}$ pueden escribirse como

$$\sum_{j=0}^{N-1} r_{xx}[i-j]\hat{s}[j] = \sum_{j=0}^{N-1} r_{ss}[i-j]x[j]$$

para $i = 0, \dots, N-1$. Luego, extender este resultado a el caso en que $s[n]$ se estima para $|n| \leq M$ en base a $x[n]$ para $|n| \leq M$, y tomar $M \rightarrow \infty$ para obtener

$$\sum_{j=-\infty}^{\infty} r_{xx}[i-j]\hat{s}[j] = \sum_{j=-\infty}^{\infty} r_{ss}[i-j]x[j]$$

para $-\infty < i < \infty$. Empleando la transformada de Fourier, encontrar la respuesta en frecuencia $H(f)$ del filtro de Wiener, o

$$H(f) = \frac{\hat{S}(f)}{X(f)}$$

asumiendo que las transformadas de Fourier de $x[n]$ y $\hat{s}[n]$ existen. Explicar los resultados.

Solución Para mostrar que

$$(\mathbf{C}_s + \sigma^2 \mathbf{I}) \hat{\mathbf{s}} = \mathbf{C}_s \mathbf{x},$$

se parte de la ecuación 11.11 y operando, se tiene que

$$\begin{aligned} \hat{\mathbf{s}} &= \mathbf{C}_s (\mathbf{C}_s + \sigma^2 \mathbf{I})^{-1} \mathbf{x} \\ &= [(\mathbf{C}_s + \sigma^2 \mathbf{I}) \mathbf{C}_s^{-1}]^{-1} \mathbf{x} \\ &= (\mathbf{I} + \sigma^2 \mathbf{C}_s^{-1})^{-1} \mathbf{x} \end{aligned}$$

Luego, despejando \mathbf{x} , se tiene

$$(\mathbf{I} + \sigma^2 \mathbf{C}_s^{-1}) \hat{\mathbf{s}} = \mathbf{x},$$

y al premultiplicar ambos lados de la igualdad por \mathbf{C}_s

$$\mathbf{C}_s (\mathbf{I} + \sigma^2 \mathbf{C}_s^{-1}) \hat{\mathbf{s}} = \mathbf{C}_s \mathbf{x},$$

se obtiene que

$$(\mathbf{C}_s + \sigma^2 \mathbf{I}) \hat{\mathbf{s}} = \mathbf{C}_s \mathbf{x}.$$

Esta ecuación puede escribirse en forma escalar como

$$\sum_{j=0}^{N-1} [\mathbf{C}_s + \sigma^2 \mathbf{I}]_{ij} \hat{s}[j] = \sum_{j=0}^{N-1} [\mathbf{C}_s]_{ij} x[j], \quad \text{para } i = 0, \dots, N-1.$$

donde se considera que las filas y las columnas de las matrices se numeran partiendo de 0 para simplificar la notación. Asumiendo que la señal es WSS, se cumple que

$$[\mathbf{C}_s]_{ij} = E(s[i]s[j]) = r_{ss}[i-j],$$

donde en la primera igualdad se consideró que \mathbf{s} tiene media nula y en la segunda igualdad la hipótesis de WSS. Además, como el modelo es $\mathbf{x} = \mathbf{s} + \mathbf{w}$ y \mathbf{s} y \mathbf{w} son independientes, la ACF de \mathbf{x} cumple que $r_{xx}[k] = r_{ss}[k] + r_{ww}[k]$, con $r_{ww}[k] = \sigma^2 \mathbf{I}$ y por lo tanto

$$[\mathbf{C}_s + \sigma^2 \mathbf{I}]_{ij} = r_{ss}[i-j] + \sigma^2 \delta[i-j] = r_{xx}[i-j].$$

Sustituyendo, se llega a que

$$\sum_{j=0}^{N-1} r_{xx}[i-j] \hat{s}[j] = \sum_{j=0}^{N-1} r_{ss}[i-j] x[j], \quad \text{para } i = 0, \dots, N-1.$$

Extendiendo este resultado a el caso en que $s[n]$ se estima para $|n| \leq M$ en base a $x[n]$ para $|n| \leq M$, y tomando $M \rightarrow \infty$, se obtiene que

$$\sum_{j=-\infty}^{\infty} r_{xx}[i-j] \hat{s}[j] = \sum_{j=-\infty}^{\infty} r_{ss}[i-j] x[j], \quad \text{para } -\infty < i < \infty,$$

y notando que las sumatorias son convoluciones, la ecuación se puede escribir como

$$(r_{xx} * \hat{s})[i] = (r_{ss} * x)[i] \quad \text{para } -\infty < i < \infty.$$

Aplicando la transformada de Fourier se obtiene que

$$P_{xx}(f) \hat{S}(f) = P_{ss}(f) X(f),$$

por lo que la respuesta en frecuencia del filtro de Wiener es

$$H(f) = \frac{\hat{S}(f)}{X(f)} = \frac{P_{ss}(f)}{P_{xx}(f)}.$$

Como $P_{xx}(f) = P_{ss}(f) + \sigma^2$, resulta en

$$H(f) = \frac{P_{ss}(f)}{P_{ss}(f) + \sigma^2},$$

y definiendo la SNR $\eta(f) = P_{ss}(f)/\sigma^2$, se puede expresar como

$$H(f) = \frac{\eta(f)}{\eta(f) + 1}.$$

Por lo tanto, en las regiones en frecuencia donde la SNR es alta, es decir, $\eta(f) \gg 1$ se tiene que $|H(f)| \approx 1$, y en las regiones en frecuencia donde la SNR es baja, es decir, $\eta(f) \ll 1$ se tiene que $|H(f)| \approx 0$. El filtro de Wiener deja incambiadas las regiones en frecuencia donde la SNR es alta y atenúa las regiones en frecuencia donde la SNR es pobre.

Capítulo 12

Estimadores bayesianos lineales

12.1. Estimación MMSE lineal

Asúmase que se quiere estimar un parámetro θ escalar a partir del conjunto de datos $\{x[0], \dots, x[N-1]\}$ o en forma vectorial, de $\mathbf{x} = [x[0] \dots x[N-1]]^T$. El parámetro desconocido se modela como la realización de una variable aleatoria. No se asume una PDF conjunta $p(\mathbf{x}, \theta)$ específica, solo el conocimiento de los primeros dos momentos. Se considera la clase de estimadores lineales (en realidad, afines) de la forma

$$\hat{\theta} = \sum_{n=0}^{N-1} a_n x[n] + a_N, \quad (12.1)$$

y se eligen los coeficientes a_n de forma de minimizar el MSE bayesiano

$$\text{Bmse}(\hat{\theta}) = E[(\theta - \hat{\theta})^2] \quad (12.2)$$

donde la esperanza es respecto a la PDF conjunta $p(\mathbf{x}, \theta)$. El estimador obtenido se llama estimador *lineal de error cuadrático medio mínimo* (*linear minimum mean square error*, LMMSE). Observar que el término a_N se incluyó para permitir \mathbf{x} y θ con media no nula. Téngase en cuenta que el estimador LMMSE es subóptimo a menos que el estimador MMSE sea lineal, como por ejemplo, si se aplica el modelo lineal bayesiano (ver la sección 10.4).

A continuación se derivan los coeficientes óptimos para emplear en la ecuación 12.1. Sustituyendo la ecuación 12.1 en la ecuación 12.2 y derivando respecto a a_N , se tiene que

$$\frac{\partial}{\partial a_N} E \left[\left(\theta - \sum_{n=0}^{N-1} a_n x[n] - a_N \right)^2 \right] = -2E \left[\theta - \sum_{n=0}^{N-1} a_n x[n] - a_N \right],$$

e igualando a cero se obtiene que

$$a_N = E(\theta) - \sum_{n=0}^{N-1} a_n E(x[n]). \quad (12.3)$$

Observar que a_N es nulo si las medias son nulas. Continuando, se quiere minimizar

$$\begin{aligned} \text{Bmse}(\hat{\theta}) &= E \left[\left(\theta - \sum_{n=0}^{N-1} a_n x[n] - a_N \right)^2 \right] \\ &= E \left\{ \left[\theta - \sum_{n=0}^{N-1} a_n x[n] - \left(E(\theta) - \sum_{n=0}^{N-1} a_n E(x[n]) \right) \right]^2 \right\} \\ &= E \left\{ \left[\sum_{n=0}^{N-1} a_n (x[n] - E(x[n])) - (\theta - E(\theta)) \right]^2 \right\} \end{aligned}$$

sobre el resto de los a_n . Sea $\mathbf{a} = [a_0 \dots a_{N-1}]$, se tiene que

$$\begin{aligned} \text{Bmse}(\hat{\theta}) &= E \left\{ [\mathbf{a}^T (\mathbf{x} - E(\mathbf{x})) - (\theta - E(\theta))]^2 \right\} \\ &= E [\mathbf{a}^T (\mathbf{x} - E(\mathbf{x})) (\mathbf{x} - E(\mathbf{x}))^T \mathbf{a}] - E [\mathbf{a}^T (\mathbf{x} - E(\mathbf{x})) (\theta - E(\theta))] \\ &\quad - E [(\theta - E(\theta)) (\mathbf{x} - E(\mathbf{x}))^T \mathbf{a}] + E [(\theta - E(\theta))^2] \\ &= \mathbf{a}^T \mathbf{C}_{xx} \mathbf{a} - \mathbf{a}^T \mathbf{C}_{x\theta} - \mathbf{C}_{\theta x} \mathbf{a} + C_{\theta\theta} \end{aligned} \quad (12.4)$$

donde \mathbf{C}_{xx} es la matriz $N \times N$ de covarianza de \mathbf{x} , $\mathbf{C}_{\theta x}$ es el vector $1 \times N$ de covarianza cruzada, con la propiedad de que $\mathbf{C}_{\theta x}^T = \mathbf{C}_{x\theta}$ y $C_{\theta\theta}$ es la varianza de θ . Para minimizar esta expresión, se toma el gradiente respecto a \mathbf{a} , obteniendo que

$$\frac{\partial \text{Bmse}(\hat{\theta})}{\partial \mathbf{a}} = 2\mathbf{C}_{xx} - 2\mathbf{C}_{x\theta},$$

donde se emplearon los resultados deducidos en el apéndice D. Finalmente, igualando a cero resulta en

$$\mathbf{a} = \mathbf{C}_{xx}^{-1} \mathbf{C}_{x\theta}. \quad (12.5)$$

Sustituyendo las ecuaciones 12.3 y 12.5 en la ecuación 12.1 resulta en

$$\begin{aligned} \hat{\theta} &= \mathbf{a}^T \mathbf{x} + a_N \\ &= \mathbf{C}_{x\theta}^T \mathbf{C}_{xx}^{-1} \mathbf{x} + E(\theta) - \mathbf{C}_{x\theta}^T \mathbf{C}_{xx}^{-1} E(\mathbf{x}), \end{aligned}$$

por lo que el estimador LMMSE es

$$\hat{\theta} = E(\theta) + \mathbf{C}_{\theta x} \mathbf{C}_{xx}^{-1} (\mathbf{x} - E(\mathbf{x})). \quad (12.6)$$

Observar que la forma del estimador LMMSE es idéntica al estimador MMSE para el caso en que \mathbf{x} y θ son conjuntamente gaussianos, dado por la ecuación 10.14. Esto ocurre porque en el caso gaussiano el estimador MMSE resulta ser lineal. Si la media de θ y \mathbf{x} es nula, el estimador es

$$\hat{\theta} = \mathbf{C}_{\theta x} \mathbf{C}_{xx}^{-1} \mathbf{x}. \quad (12.7)$$

El MSE bayesiano mínimo se obtiene sustituyendo la ecuación 12.5 en la 12.4, resultando en

$$\begin{aligned} \text{Bmse}(\hat{\theta}) &= \mathbf{C}_{x\theta}^T \mathbf{C}_{xx}^{-1} \mathbf{C}_{xx} \mathbf{C}_{xx}^{-1} \mathbf{C}_{x\theta} - \mathbf{C}_{x\theta}^T \mathbf{C}_{xx}^{-1} \mathbf{C}_{x\theta} - \mathbf{C}_{\theta x} \mathbf{C}_{xx}^{-1} \mathbf{C}_{x\theta} + C_{\theta\theta} \\ &= \mathbf{C}_{\theta x} \mathbf{C}_{xx}^{-1} \mathbf{C}_{x\theta} - \mathbf{C}_{\theta x} \mathbf{C}_{xx}^{-1} \mathbf{C}_{x\theta} - \mathbf{C}_{\theta x} \mathbf{C}_{xx}^{-1} \mathbf{C}_{x\theta} + C_{\theta\theta}, \end{aligned}$$

es decir,

$$\text{Bmse}(\hat{\theta}) = C_{\theta\theta} - \mathbf{C}_{\theta x} \mathbf{C}_{xx}^{-1} \mathbf{C}_{x\theta}. \quad (12.8)$$

Este es el mismo resultado que se obtiene de sustituir la ecuación 10.15 en la ecuación 11.3.

12.1.1. Ejemplo: nivel de DC en WGN con PDF a priori uniforme

Considérese el ejemplo de la introducción del capítulo 10. El modelo de los datos es

$$x[n] = A + w[n], \quad n = 0, \dots, N-1,$$

donde $A \sim \mathcal{U}[-A_0, A_0]$, $w[n]$ es WGN con varianza σ^2 , y A y $w[n]$ son independientes. Se desea estimar A . El estimador MMSE no tiene una expresión analítica debido a la integración (ver la ecuación 10.4). Para calcular el estimador LMMSE, se observa primero que $E(A) = 0$ y por lo tanto, $E(x[n]) = 0$. Como $E(\mathbf{x}) = 0$, las covarianzas son

$$\begin{aligned} \mathbf{C}_{xx} &= E(\mathbf{x}\mathbf{x}^T) \\ &= E[(A\mathbf{1} + \mathbf{w})(A\mathbf{1} + \mathbf{w})^T] \\ &= E(A^2)\mathbf{1}\mathbf{1}^T + \sigma^2\mathbf{I} \end{aligned}$$

y

$$\begin{aligned}\mathbf{C}_{\theta x} &= E(A\mathbf{x}^T) \\ &= E[A(A\mathbf{1} + \mathbf{w})^T] \\ &= E(A^2)\mathbf{1}^T,\end{aligned}$$

donde en ambos casos, en la última igualdad, se consideró que A y \mathbf{w} son independientes, y $\mathbf{1}$ es un vector $N \times 1$ de unos. Por lo tanto, de la ecuación 12.7

$$\begin{aligned}\hat{A} &= \mathbf{C}_{\theta x} \mathbf{C}_{xx}^{-1} \mathbf{x} \\ &= \sigma_A^2 \mathbf{1}^T (\sigma_A^2 \mathbf{1} \mathbf{1}^T + \sigma^2 \mathbf{I})^{-1} \mathbf{x} \\ &= \frac{\sigma_A^2}{\sigma^2} \mathbf{1}^T \left(\frac{\sigma_A^2}{\sigma^2} \mathbf{1} \mathbf{1}^T + \mathbf{I} \right)^{-1} \mathbf{x},\end{aligned}$$

donde $\sigma_A^2 = E(A^2)$. Considerando la identidad de Woodbury dada por la ecuación A.2 con $\mathbf{A} = \mathbf{I}$ y $\mathbf{u} = (\sigma_A/\sigma)\mathbf{1}$, se cumple que

$$\left(\mathbf{I} + \frac{\sigma_A^2}{\sigma^2} \mathbf{1} \mathbf{1}^T \right)^{-1} = \mathbf{I} - \frac{\frac{\sigma_A^2}{\sigma^2} \mathbf{1} \mathbf{1}^T}{1 + N \frac{\sigma_A^2}{\sigma^2}},$$

y por lo tanto,

$$\begin{aligned}\hat{A} &= \frac{\sigma_A^2}{\sigma^2} \mathbf{1}^T \left(\mathbf{I} - \frac{\mathbf{1} \mathbf{1}^T}{N + \frac{\sigma^2}{\sigma_A^2}} \right) \mathbf{x} \\ &\stackrel{(a)}{=} \frac{\sigma_A^2}{\sigma^2} \left(\mathbf{1}^T - \frac{N}{N + \frac{\sigma^2}{\sigma_A^2}} \mathbf{1}^T \right) \mathbf{x} \\ &\stackrel{(b)}{=} \frac{\sigma_A^2}{\sigma^2} \left(1 - \frac{N}{N + \frac{\sigma^2}{\sigma_A^2}} \right) N \bar{x} \\ &= \frac{\sigma_A^2}{\sigma^2} \left(\frac{\frac{\sigma^2}{\sigma_A^2}}{N + \frac{\sigma^2}{\sigma_A^2}} \right) N \bar{x} \\ &= \frac{N}{N + \frac{\sigma^2}{\sigma_A^2}} \bar{x},\end{aligned}$$

donde en (a) se consideró que $\mathbf{1}^T(\mathbf{1}\mathbf{1}^T) = N\mathbf{1}^T$ y en (b) que $\mathbf{1}^T \mathbf{x} = N\bar{x}$. Finalmente, se obtiene que

$$\hat{A} = \frac{\sigma_A^2}{\sigma_A^2 + \frac{\sigma^2}{N}} \bar{x}.$$

y teniendo en cuenta que

$$\sigma_A^2 = E(A^2) = \frac{1}{2A_0} \int_{-A_0}^{A_0} A^2 dA = \frac{1}{2A_0} \left(\frac{A^3}{3} \Big|_{-A_0}^{A_0} \right) = \frac{1}{2A_0} \frac{2A_0^3}{3} = \frac{A_0^2}{3},$$

el estimador LMMSE de A es

$$\hat{A} = \frac{\frac{A_0^2}{3}}{\frac{A_0^2}{3} + \frac{\sigma^2}{N}} \bar{x}. \quad (12.9)$$

En contraposición con el estimador MMSE, el estimador LMMSE obtenido tiene una expresión analítica. Observar además que en realidad no es necesario saber que la PDF a priori de A es uniforme. Solo alcanza con conocer su media y su varianza. Tampoco se necesita saber que $w[n]$ es gaussiano. Alcanza con saber que es blanco y conocer su varianza. De la misma forma, la hipótesis de independencia entre A y \mathbf{w} tampoco es necesaria, ya que solo se empleó que son no correlacionados. En general, todo lo que se necesita para calcular el estimador LMMSE son los dos primeros momentos de $p(\mathbf{x}, \theta)$, es decir,

$$\begin{bmatrix} E(\theta) \\ E(\mathbf{x}) \end{bmatrix} \quad y \quad \begin{bmatrix} C_{\theta\theta} & C_{\theta x} \\ C_{x\theta} & C_{xx} \end{bmatrix}.$$

Sin embargo, debe tenerse en cuenta que el estimador LMMSE de la ecuación 12.9 es subóptimo debido a que se restringió a que sea lineal. El estimador óptimo para este problema está dado por la ecuación 10.4.

12.2. Interpretación geométrica

Se discute a continuación una interpretación geométrica del estimador LMSSE, análoga a la realizada en la sección 8.2.3 para el LSE, pero ahora los vectores son variables aleatorias. En la formulación se asumirá que θ y \mathbf{x} son de media nula. En caso de no serlo, es posible definir las variables aleatorias de media nula $\theta' = \theta - E(\theta)$ y $\mathbf{x}' = \mathbf{x} - E(\mathbf{x})$ y considerar la estimación de θ' mediante una combinación lineal de \mathbf{x}' (ver también el problema de la sección 12.6.5). Se desea encontrar los coeficientes a_n de forma que

$$\hat{\theta} = \sum_{n=0}^{N-1} a_n x[n]$$

minimice

$$\text{Bmse}(\hat{\theta}) = E \left[(\theta - \hat{\theta})^2 \right].$$

Considérese a las variables aleatorias $\theta, x[0], \dots, x[N-1]$ como elementos en un espacio vectorial, como

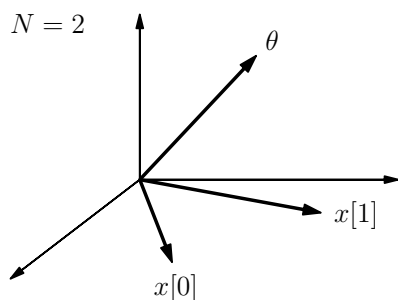


Figura 12.1: Interpretación de variables aleatorias en un espacio vectorial.

se muestra simbólicamente en la figura 12.1. Como ocurre usualmente, θ no pertenece al subespacio generado por los $x[n]$, por lo que no puede expresarse como combinación lineal de los $x[n]$. Si se pudiera, se obtendría un estimador perfecto. Defínase el largo de cada vector x como $\|x\| = \sqrt{E(x^2)}$, es decir, como la raíz cuadrada de la varianza. Vectores más largos son vectores con mayor varianza. El vector de largo cero es la variable aleatoria con varianza nula, y por lo tanto, aquella que es idénticamente cero, y en realidad es una constante en lugar de una variable aleatoria. Finalmente, para completar la descripción se requiere la noción de producto interno entre dos vectores. Recordar que si \mathbf{x}, \mathbf{y} son vectores en el espacio euclidiano \mathbb{R}^3 , el producto interno es $(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T \mathbf{y} = \|\mathbf{x}\| \|\mathbf{y}\| \cos \alpha$, donde α es el ángulo entre

los vectores. Puede mostrarse que una definición apropiada, es decir, una que satisface las propiedades de un producto interno entre vectores x e y es (ver el problema de la sección 12.6.4)

$$(x, y) = E(xy).$$

Con esta definición, se cumple que

$$(x, x) = E(x^2) = \|x\|^2,$$

que es consistente con la definición previa del largo de un vector. Además, se define que dos vectores son ortogonales si

$$(x, y) = E(xy) = 0.$$

Como los vectores son de media nula, esto es equivalente a decir que dos vectores son ortogonales si y solo si son no correlacionados (en \mathbb{R}^3 , dos vectores euclidianos son ortogonales si el ángulo entre ellos es $\alpha = 90^\circ$, de forma que $(\mathbf{x}, \mathbf{y}) = \|\mathbf{x}\|\|\mathbf{y}\|\cos\alpha = 0$).

A partir de estas ideas, se procederá a determinar el estimador LMMSE empleando el enfoque de espacio vectorial. El enfoque es útil para la conceptualización del proceso de estimación LMMSE y además será empleado para la derivación del LMMSE secuencial. Se asume que

$$\hat{\theta} = \sum_{n=0}^{N-1} a_n x[n],$$

donde $a_N = 0$ debido a la hipótesis de media nula. Los coeficientes a_n se eligen de forma de minimizar el MSE

$$\begin{aligned} E[(\theta - \hat{\theta})^2] &= E\left[\left(\theta - \sum_{n=0}^{N-1} a_n x[n]\right)^2\right] \\ &= \left\|\theta - \sum_{n=0}^{N-1} a_n x[n]\right\|^2. \end{aligned}$$

Esto significa que la minimización del MSE es equivalente a la minimización del largo al cuadrado del *vector de error* $\epsilon = \theta - \hat{\theta}$. Como se muestra en la figura 12.2, el largo del vector de error es mínimo cuando ϵ es ortogonal al subespacio generado por $\{x[0], \dots, x[N-1]\}$. Por lo tanto, se requiere que

$$\epsilon \perp x[0], \dots, x[N-1] \quad (12.10)$$

o empleando la definición de ortogonalidad

$$E[(\theta - \hat{\theta})x[n]] = 0, \quad n = 0, \dots, N-1.$$

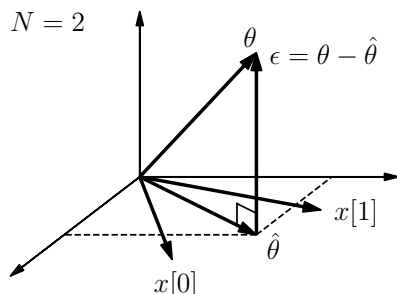


Figura 12.2: Principio de ortogonalidad para la estimación LMMSE.

Este es el *principio de ortogonalidad* o *teorema de proyección*, que indica que en la estimación de la realización de una variable aleatoria mediante una combinación lineal de las muestras de datos, el estimador

óptimo es obtenido cuando el error es ortogonal a cada muestra de datos. Empleando el principio de ortogonalidad, los coeficientes se obtienen como

$$E \left[\left(\theta - \sum_{m=0}^{N-1} a_m x[m] \right) x[n] \right] = 0, \quad n = 0, \dots, N-1$$

o

$$\sum_{m=0}^{N-1} a_m E(x[m]x[n]) = E(\theta x[n]), \quad n = 0, \dots, N-1.$$

En forma matricial, el sistema es

$$\begin{bmatrix} E(x^2[0]) & E(x[0]x[1]) & \dots & E(x[0]x[N-1]) \\ E(x[1]x[0]) & E(x^2[1]) & \dots & E(x[1]x[N-1]) \\ \vdots & \vdots & \ddots & \vdots \\ E(x[N-1]x[0]) & E(x[N-1]x[1]) & \dots & E(x^2[N-1]) \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_{N-1} \end{bmatrix} = \begin{bmatrix} E(\theta x[0]) \\ E(\theta x[1]) \\ \vdots \\ E(\theta x[N-1]) \end{bmatrix}.$$

Este sistema de ecuaciones se denomina *ecuaciones normales*. La matriz se reconoce como \mathbf{C}_{xx} y el vector del lado derecho de la igualdad como $\mathbf{C}_{x\theta}$. Por lo tanto,

$$\mathbf{C}_{xx}\mathbf{a} = \mathbf{C}_{x\theta}$$

o

$$\mathbf{a} = \mathbf{C}_{xx}^{-1} \mathbf{C}_{x\theta}. \quad (12.11)$$

El estimador LMMSE de θ es

$$\begin{aligned} \hat{\theta} &= \mathbf{a}\mathbf{x} \\ &= \mathbf{C}_{x\theta}^T \mathbf{C}_{xx}^{-1} \mathbf{x}, \end{aligned}$$

o finalmente

$$\hat{\theta} = \mathbf{C}_{\theta x} \mathbf{C}_{xx}^{-1} \mathbf{x},$$

acorde con la ecuación 12.7. El MSE bayesiano mínimo es el cuadrado del largo del error,

$$\begin{aligned} \text{Bsme}(\hat{\theta}) &= \|\epsilon\|^2 \\ &= \left\| \theta - \sum_{n=0}^{N-1} a_n x[n] \right\|^2 \\ &= E \left[\left(\theta - \sum_{n=0}^{N-1} a_n x[n] \right)^2 \right] \end{aligned}$$

para los a_n dados por la ecuación 12.11. Pero

$$\begin{aligned} \text{Bsme}(\hat{\theta}) &= E \left[\left(\theta - \sum_{n=0}^{N-1} a_n x[n] \right) \left(\theta - \sum_{m=0}^{N-1} a_m x[m] \right) \right] \\ &= E \left[\left(\theta - \sum_{n=0}^{N-1} a_n x[n] \right) \theta \right] - E \left[\left(\theta - \sum_{n=0}^{N-1} a_n x[n] \right) \sum_{m=0}^{N-1} a_m x[m] \right] \\ &= E(\theta^2) - \sum_{n=0}^{N-1} a_n E(x[n]\theta) - \sum_{m=0}^{N-1} a_m E \left[\left(\theta - \sum_{n=0}^{N-1} a_n x[n] \right) x[m] \right]. \end{aligned}$$

El último término es nulo por tratarse del principio de ortogonalidad, resultando en

$$\begin{aligned} \text{Bsme}(\hat{\theta}) &= C_{\theta\theta} - \mathbf{a}^T \mathbf{C}_{x\theta} \\ &= C_{\theta\theta} - \mathbf{C}_{x\theta}^T \mathbf{C}_{xx}^{-1} \mathbf{C}_{x\theta} \\ &= C_{\theta\theta} - \mathbf{C}_{\theta x} \mathbf{C}_{xx}^{-1} \mathbf{C}_{x\theta}, \end{aligned}$$

acorde con el resultado de la ecuación 12.8.

12.2.1. Ejemplo: estimación con vectores ortogonales

Asúmase que $x[0]$ y $x[1]$ son de media nula y no correlacionados entre si. Sin embargo, son correlacionados con θ . La situación se ilustra en la figura 12.3. Gracias a la ortogonalidad, el estimador de θ en base a $x[0]$ y $x[1]$ es la suma de las proyecciones de θ sobre $x[0]$ y $x[1]$, como se muestra en la figura 12.3, o

$$\begin{aligned}\hat{\theta} &= \hat{\theta}_0 + \hat{\theta}_1 \\ &= \left(\theta, \frac{x[0]}{\|x[0]\|} \right) \frac{x[0]}{\|x[0]\|} + \left(\theta, \frac{x[1]}{\|x[1]\|} \right) \frac{x[1]}{\|x[1]\|}\end{aligned}$$

El factor $(\theta, x[n]/\|x[n]\|)$ es el largo de la proyección y el factor $x[n]/\|x[n]\|$ es la dirección. Teniendo en cuenta que $\|x[n]\| = \sqrt{E(x^2[n])} = \sqrt{\text{var}(x[n])}$ es constante, y por lo tanto

$$\left(\theta, \frac{x[n]}{\|x[n]\|} \right) \frac{x[n]}{\|x[n]\|} = \frac{E(\theta x[n])}{\|x[n]\|} \frac{x[n]}{\|x[n]\|} = \frac{E(\theta x[n])}{E(x^2[n])} x[n] = \frac{(\theta, x[n])}{(x[n], x[n])} x[n],$$

el estimador se puede escribir como

$$\hat{\theta} = \frac{(\theta, x[0])}{(x[0], x[0])} x[0] + \frac{(\theta, x[1])}{(x[1], x[1])} x[1].$$

Observar que empleando la definición de producto interno, se tiene que

$$\begin{aligned}\hat{\theta} &= \frac{E(\theta x[0])}{E(x^2[0])} x[0] + \frac{E(\theta x[1])}{E(x^2[1])} x[1] \\ &= [E(\theta x[0]) \quad E(\theta x[1])] \begin{bmatrix} E(x^2[0]) & 0 \\ 0 & E(x^2[1]) \end{bmatrix}^{-1} \begin{bmatrix} x[0] \\ x[1] \end{bmatrix} \\ &= \mathbf{C}_{\theta x} \mathbf{C}_{xx}^{-1} \mathbf{x}.\end{aligned}$$

Claramente, la facilidad en la obtención de este resultado proviene de la ortogonalidad entre $x[0]$ y $x[1]$,

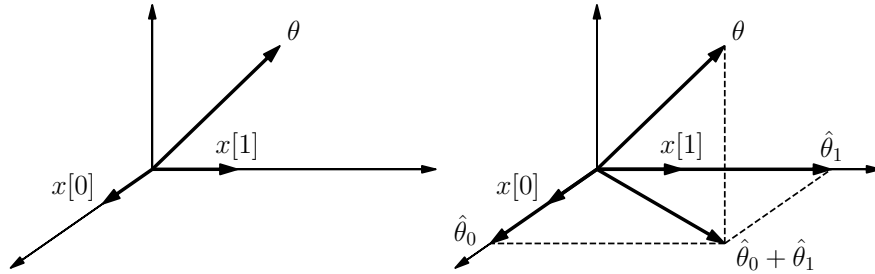


Figura 12.3: Estimación lineal mediante vectores ortogonales.

o equivalentemente, en la naturaleza diagonal de \mathbf{C}_{xx} . Si las muestras de datos no son ortogonales puede emplearse el mismo enfoque si previamente se ortogonalizan las muestras o se reemplazan los datos por muestras no correlacionadas que generan el mismo subespacio. Esto se hace en la deducción del estimador LMMSE secuencial en la sección 12.4.

12.3. El estimador LMMSE vectorial

El estimador LMMSE vectorial es una extensión directa del caso escalar. Si el parámetro es vectorial, se desea encontrar el estimador lineal que minimiza el MSE bayesiano para cada elemento. Se asume que

$$\hat{\theta}_i = \sum_{n=0}^{N-1} a_{in} x[n] + a_{iN} \quad (12.12)$$

para $i = 1, \dots, p$ y se eligen los coeficientes de forma de minimizar

$$\text{Bmse}(\hat{\theta}_i) = E \left[(\theta_i - \hat{\theta}_i)^2 \right] \quad i = 1, \dots, p,$$

donde la esperanza se toma respecto a $p(\mathbf{x}, \theta_i)$. Como en realidad se están determinando p estimadores separados, puede aplicarse la solución escalar, obteniendo de la ecuación 12.6 que

$$\hat{\theta}_i = E(\theta_i) + \mathbf{C}_{\theta_i x} \mathbf{C}_{xx}^{-1} (\mathbf{x} - E(\mathbf{x})) \quad i = 1, \dots, p$$

y el MSE bayesiano mínimo, de la ecuación 12.8, es

$$\text{Bmse}(\hat{\theta}_i) = C_{\theta_i \theta_i} - \mathbf{C}_{\theta_i x} \mathbf{C}_{xx}^{-1} \mathbf{C}_{x \theta_i} \quad i = 1, \dots, p.$$

Los estimadores LMMSE escalares pueden combinarse en un estimador vectorial como

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= \begin{bmatrix} E(\theta_1) \\ E(\theta_2) \\ \vdots \\ E(\theta_p) \end{bmatrix} + \begin{bmatrix} \mathbf{C}_{\theta_1 x} \mathbf{C}_{xx}^{-1} (\mathbf{x} - E(\mathbf{x})) \\ \mathbf{C}_{\theta_2 x} \mathbf{C}_{xx}^{-1} (\mathbf{x} - E(\mathbf{x})) \\ \vdots \\ \mathbf{C}_{\theta_p x} \mathbf{C}_{xx}^{-1} (\mathbf{x} - E(\mathbf{x})) \end{bmatrix} \\ &= \begin{bmatrix} E(\theta_1) \\ E(\theta_2) \\ \vdots \\ E(\theta_p) \end{bmatrix} + \begin{bmatrix} \mathbf{C}_{\theta_1 x} \\ \mathbf{C}_{\theta_2 x} \\ \vdots \\ \mathbf{C}_{\theta_p x} \end{bmatrix} \mathbf{C}_{xx}^{-1} (\mathbf{x} - E(\mathbf{x})) \end{aligned}$$

resultando en

$$\hat{\boldsymbol{\theta}} = E(\boldsymbol{\theta}) + \mathbf{C}_{\boldsymbol{\theta} x} \mathbf{C}_{xx}^{-1} (\mathbf{x} - E(\mathbf{x})), \quad (12.13)$$

donde ahora $\mathbf{C}_{\boldsymbol{\theta} x}$ es una matriz $p \times N$. Mediante un enfoque similar, la matriz MSE bayesiana es

$$\begin{aligned} \mathbf{M}_{\hat{\boldsymbol{\theta}}} &= E \left[(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \right] \\ &= \mathbf{C}_{\boldsymbol{\theta} \boldsymbol{\theta}} - \mathbf{C}_{\boldsymbol{\theta} x} \mathbf{C}_{xx}^{-1} \mathbf{C}_{x \boldsymbol{\theta}} \end{aligned} \quad (12.14)$$

donde $\mathbf{C}_{\boldsymbol{\theta} \boldsymbol{\theta}}$ es la matriz de covarianza $p \times p$. Como consecuencia, el MSE bayesiano mínimo es

$$\text{Bmse}(\hat{\theta}_i) = [\mathbf{M}_{\hat{\boldsymbol{\theta}}}]_{ii}. \quad (12.15)$$

Estos resultados se muestran en el problema de la sección 12.6.7. Observar que los resultados son idénticos a los del estimador MMSE para el caso gaussiano, para el cual el estimador es lineal. Notar además que para determinar el estimador LMMSE se necesitan únicamente los dos primeros momentos de la PDF.

Dos propiedades del estimador MMSE son particularmente útiles. Una de ellas es que el estimador LMMSE conmuta sobre transformaciones lineales (en realidad, afines). Es decir, si

$$\boldsymbol{\alpha} = \mathbf{A} \boldsymbol{\theta} + \mathbf{b},$$

el estimador de $\boldsymbol{\alpha}$ es

$$\hat{\boldsymbol{\alpha}} = \mathbf{A} \hat{\boldsymbol{\theta}} + \mathbf{b}, \quad (12.16)$$

donde $\hat{\boldsymbol{\theta}}$ está dado por la ecuación 12.13. La segunda propiedad establece que el estimador LMMSE de la suma de parámetros desconocidos es la suma de los estimadores individuales. Concretamente, si se desea estimar $\boldsymbol{\alpha} = \boldsymbol{\theta}_1 + \boldsymbol{\theta}_2$, se cumple que

$$\hat{\boldsymbol{\alpha}} = \hat{\boldsymbol{\theta}}_1 + \hat{\boldsymbol{\theta}}_2, \quad (12.17)$$

donde

$$\begin{aligned} \hat{\boldsymbol{\theta}}_1 &= E(\boldsymbol{\theta}_1) + \mathbf{C}_{\boldsymbol{\theta}_1 x} \mathbf{C}_{xx}^{-1} (\mathbf{x} - E(\mathbf{x})) \\ \hat{\boldsymbol{\theta}}_2 &= E(\boldsymbol{\theta}_2) + \mathbf{C}_{\boldsymbol{\theta}_2 x} \mathbf{C}_{xx}^{-1} (\mathbf{x} - E(\mathbf{x})). \end{aligned}$$

Las pruebas de estas propiedades se realizan en el problema de la sección 12.6.8.

En analogía con el BLUE, hay un teorema de Gauss-Markov para el caso bayesiano. Este indica que para datos con la *forma* del modelo lineal bayesiano, es decir, el modelo lineal bayesiano *sin la hipótesis gaussiana*, existe un estimador óptimo. La optimalidad se mide mediante el MSE bayesiano. El teorema es la aplicación del estimador LMMSE al modelo lineal bayesiano. La deducción es idéntica a la realizada en la sección 10.4, y se resume en el teorema de Gauss-Markov bayesiano:

Teorema de Gauss-Markov bayesiano Si los datos tienen la forma del modelo lineal bayesiano

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$$

donde \mathbf{x} es un vector de datos $N \times 1$, \mathbf{H} es una matriz de observación $N \times p$ conocida, $\boldsymbol{\theta}$ es un vector aleatorio $p \times 1$ de parámetros cuya realización se quiere estimar y tiene media $E(\boldsymbol{\theta})$ y matriz de covarianza $\mathbf{C}_{\theta\theta}$, y \mathbf{w} es un vector aleatorio $N \times 1$ de media nula y matriz de covarianza \mathbf{C}_w y es no correlacionado con $\boldsymbol{\theta}$ (salvo esto, la PDF conjunta $p(\mathbf{w}, \boldsymbol{\theta})$ es arbitraria), el estimador LMMSE de $\boldsymbol{\theta}$ es

$$\hat{\boldsymbol{\theta}} = E(\boldsymbol{\theta}) + \mathbf{C}_{\theta\theta}\mathbf{H}^T(\mathbf{H}\mathbf{C}_{\theta\theta}\mathbf{H}^T + \mathbf{C}_w)^{-1}(\mathbf{x} - \mathbf{H}E(\boldsymbol{\theta})) \quad (12.18)$$

$$= E(\boldsymbol{\theta}) + (\mathbf{C}_{\theta\theta}^{-1} + \mathbf{H}^T\mathbf{C}_w^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}_w^{-1}(\mathbf{x} - \mathbf{H}E(\boldsymbol{\theta})). \quad (12.19)$$

El desempeño del estimador se mide con el error de estimación $\boldsymbol{\epsilon} = \boldsymbol{\theta} - \hat{\boldsymbol{\theta}}$, cuya media es nula y su matriz de covarianza es

$$\begin{aligned} \mathbf{C}_\epsilon &= E_{\mathbf{x}, \boldsymbol{\theta}}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T) \\ &= \mathbf{C}_{\theta\theta} - \mathbf{C}_{\theta\theta}\mathbf{H}^T(\mathbf{H}\mathbf{C}_{\theta\theta}\mathbf{H}^T + \mathbf{C}_w)^{-1}\mathbf{H}\mathbf{C}_{\theta\theta} \end{aligned} \quad (12.20)$$

$$= (\mathbf{C}_{\theta\theta}^{-1} + \mathbf{H}^T\mathbf{C}_w^{-1}\mathbf{H})^{-1}. \quad (12.21)$$

La matriz de covarianza del error es también la matriz MSE bayesiana $\mathbf{M}_{\hat{\boldsymbol{\theta}}}$ cuyos elementos de la diagonal contienen el MSE bayesiano mínimo

$$\begin{aligned} [\mathbf{M}_{\hat{\boldsymbol{\theta}}}]_{ii} &= [\mathbf{C}_\epsilon]_{ii} \\ &= \text{Bmse}(\hat{\theta}_i). \end{aligned}$$

Estos resultados son idénticos a los del teorema de la sección 11.4.2 para el modelo lineal bayesiano excepto que el vector de error no es necesariamente gaussiano. El teorema de Gauss-Markov bayesiano indica que dentro de la clase de estimadores lineales, el que minimiza el MSE bayesiano para cada elemento de $\boldsymbol{\theta}$ es el dado por las ecuaciones 12.18 o 12.19. No es óptimo a menos que la esperanza condicional $E(\boldsymbol{\theta}|\mathbf{x})$ sea lineal. Ese es el caso para la PDF conjunta gaussiana.

12.4. Estimación LMMSE secuencial

En la sección 8.4 se discutió el procedimiento de mínimos cuadrados secuencial como el proceso de actualizar el estimador LSE en el tiempo a medida que están disponibles nuevas muestras. Un procedimiento análogo puede emplearse para los estimadores LMMSE. En el ejemplo de la sección 12.2.1, el estimador LMMSE se obtuvo sumando el nuevo estimador $\hat{\theta}_1$ basado en el nuevo dato $x[1]$ al viejo estimador $\hat{\theta}_0$ basado en la muestra previa $x[0]$. Esto es posible debido a que $x[0]$ y $x[1]$ son ortogonales. Cuando los datos no son ortogonales, el álgebra se vuelve mas complicada pero el enfoque es similar. Antes de establecer el resultado general se mostrarán las operaciones necesarias en el caso de la estimación del nivel de DC en WGN, que tiene la forma del modelo lineal bayesiano. La deducción se hará de forma puramente algebraica y luego se repetirán los cálculos empleando el enfoque de espacio vectorial. Se asume que el nivel A de DC tiene media nula de forma que el enfoque de espacio vectorial es aplicable.

Comenzando con la deducción algebraica, en el ejemplo de la sección 10.2.1 del cálculo del estimador MMSE del nivel de DC en WGN con PDF a priori gaussiana se encontró que con $\mu_A = 0$ (ecuación 10.7)

$$\hat{A}[N-1] = \frac{\sigma_A^2}{\sigma_A^2 + \frac{\sigma^2}{N}} \bar{x}.$$

$\hat{A}[N-1]$ es también el estimador LMMSE basado en $\{x[0], \dots, x[N-1]\}$. Esto es porque el estimador LMMSE es idéntico al estimador MMSE para el caso gaussiano debido a que es lineal con los datos. También, de la ecuación 10.9, se tiene que

$$\text{Bsme}(\hat{A}[N-1]) = \frac{\sigma_A^2 \sigma^2}{N\sigma_A^2 + \sigma^2}. \quad (12.22)$$

Para actualizar el estimador al disponer de la nueva muestra $x[N]$, se nota que

$$\begin{aligned}
\hat{A}[N] &= \frac{\sigma_A^2}{\sigma_A^2 + \frac{\sigma^2}{N+1}} \frac{1}{N+1} \sum_{n=0}^N x[n] \\
&\stackrel{(a)}{=} \frac{N\sigma_A^2}{(N+1)\sigma_A^2 + \sigma^2} \frac{1}{N} \left(\sum_{n=0}^{N-1} x[n] + x[N] \right) \\
&\stackrel{(b)}{=} \frac{N\sigma_A^2}{(N+1)\sigma_A^2 + \sigma^2} \frac{\sigma_A^2 + \frac{\sigma^2}{N}}{\sigma_A^2} \hat{A}[N-1] + \frac{\sigma_A^2}{(N+1)\sigma_A^2 + \sigma^2} x[N] \\
&\stackrel{(c)}{=} \frac{N\sigma_A^2 + \sigma^2}{(N+1)\sigma_A^2 + \sigma^2} \hat{A}[N-1] + \frac{\sigma_A^2}{(N+1)\sigma_A^2 + \sigma^2} x[N] \\
&\stackrel{(d)}{=} \hat{A}[N-1] + \left[\frac{N\sigma_A^2 + \sigma^2}{(N+1)\sigma_A^2 + \sigma^2} - 1 \right] \hat{A}[N-1] + \frac{\sigma_A^2}{(N+1)\sigma_A^2 + \sigma^2} x[N] \\
&\stackrel{(e)}{=} \hat{A}[N-1] + \frac{\sigma_A^2}{(N+1)\sigma_A^2 + \sigma^2} (x[N] - \hat{A}[N-1]), \tag{12.23}
\end{aligned}$$

donde en (a) se sacó denominador común $N+1$ en el denominador del primer factor y se multiplicó el numerador y el denominador por N , en (b) se desarrolló el término entre paréntesis y se sustituyó el factor \bar{x} empleando la ecuación de $\hat{A}[N-1]$ en (c) se canceló σ_A^2 en el numerador y denominador del primer sumando, en (d) se sumó y restó $\hat{A}[N-1]$ y en (e) se desarrolló el factor entre paréntesis rectos. De forma similar al estimador LS secuencial, se corrige el viejo estimador $\hat{A}[N-1]$ mediante una versión escalada del error de predicción $x[N] - \hat{A}[N-1]$. De las ecuaciones 12.23 y 12.22, el factor de escala o de ganancia es

$$\begin{aligned}
K[N] &= \frac{\sigma_A^2}{(N+1)\sigma_A^2 + \sigma^2} \\
&= \frac{\text{Bsme}(\hat{A}[N-1])}{\text{Bsme}(\hat{A}[N-1]) + \sigma^2}, \tag{12.24}
\end{aligned}$$

que decrece a cero con $N \rightarrow \infty$, reflejando la confianza creciente en el viejo estimador a medida que se emplean mas datos. También es posible encontrar una ecuación de actualización del MSE bayesiano mínimo, ya que de la ecuación 12.22,

$$\begin{aligned}
\text{Bsme}(\hat{A}[N]) &= \frac{\sigma_A^2 \sigma^2}{(N+1)\sigma_A^2 + \sigma^2} \\
&\stackrel{(a)}{=} \frac{N\sigma_A^2 + \sigma^2}{(N+1)\sigma_A^2 + \sigma^2} \frac{\sigma_A^2 \sigma^2}{N\sigma_A^2 + \sigma^2} \\
&\stackrel{(b)}{=} (1 - K[N]) \text{Bsme}(\hat{A}[N-1]),
\end{aligned}$$

donde en (a) se multiplicó el numerador y el denominador por $N\sigma_A^2 + \sigma^2$ y en (b) se notó que el primer factor es $1 - K[N]$ y el segundo es $\text{Bsme}(\hat{A}[N-1])$. Resumiendo los resultados, se obtuvo el siguiente estimador LMMSE secuencial.

Actualización del estimador:

$$\hat{A}[N] = \hat{A}[N-1] + K[N](x[N] - \hat{A}[N-1]), \tag{12.25}$$

donde

$$K[N] = \frac{\text{Bsme}(\hat{A}[N-1])}{\text{Bsme}(\hat{A}[N-1]) + \sigma^2}, \tag{12.26}$$

Actualización del MSE mínimo:

$$\text{Bsme}(\hat{A}[N]) = (1 - K[N]) \text{Bsme}(\hat{A}[N-1]). \tag{12.27}$$

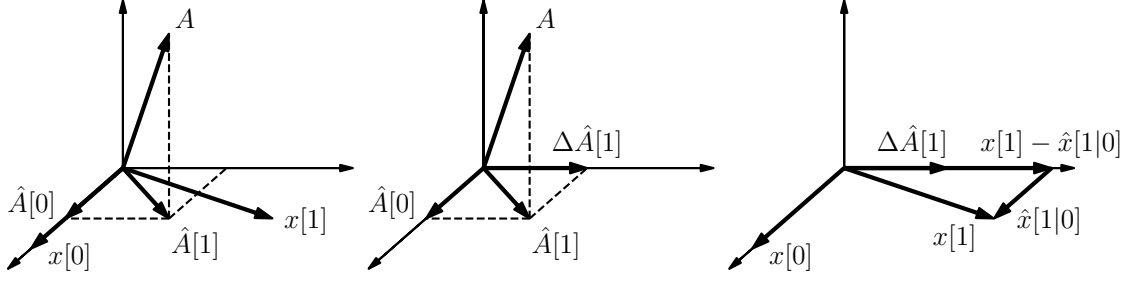


Figura 12.4: Estimación secuencial empleando el enfoque de espacio vectorial.

A continuación se derivan los mismos resultados empleando el enfoque de espacio vectorial. Asíumase que se quiere encontrar el estimador LMMSE $\hat{A}[1]$ en base a los datos $\{x[0], x[1]\}$, como se muestra en la figura 12.4. Como $x[0]$ y $x[1]$ no son ortogonales no puede sumarse el estimador basado en $x[0]$ al estimador basado en $x[1]$. Sin embargo, se puede obtener $\hat{A}[1]$ como la suma de $\hat{A}[0]$ y un componente ortogonal a $\hat{A}[0]$ como se muestra en la figura. Ese componente es $\Delta\hat{A}[1]$. Para encontrar un vector en la dirección de ese componente, se tiene en cuenta que el estimador LMMSE tiene la propiedad de que el error de estimación es ortogonal a los datos (principio de ortogonalidad). Por lo tanto, si se calcula el estimador de $x[1]$ en base a $x[0]$, que se denotará $\hat{x}[1|0]$, el error $x[1] - \hat{x}[1|0]$ será ortogonal a $x[0]$, como se muestra en la figura 12.4. Como $x[0]$ y $x[1] - \hat{x}[1|0]$ son ortogonales, se puede proyectar A sobre cada vector separadamente y sumar los resultados, de forma que

$$\hat{A}[1] = \hat{A}[0] + \Delta\hat{A}[1].$$

La idea fundamental del procedimiento es que $x[0]$ y $x[1] - \hat{x}[1|0]$ generan el mismo subespacio que $x[0]$ y $x[1]$, por lo que el estimador $\hat{A}[1]$, que es la proyección de A en dicho subespacio, es la misma en base a cualquier par de vectores. Pero $x[0]$ y $x[1] - \hat{x}[1|0]$ son ortogonales, y por lo tanto, se puede estimar A en base a $x[0]$ y $x[1] - \hat{x}[1|0]$ separadamente y sumar los resultados. Para encontrar el término de corrección $\Delta\hat{A}[1]$, se observa que el estimador LMMSE de una variable aleatoria y en base a x , ambas de media nula, por la ecuación 12.7 es

$$\hat{y} = \mathbf{C}_{yx} \mathbf{C}_{xx}^{-1} \mathbf{x} \quad \Rightarrow \quad \hat{y} = \frac{E(xy)}{E(x^2)} x, \quad (12.28)$$

por tratarse de variables aleatorias escalares. Por lo tanto, el estimador LMMSE de $x[1]$ en base a $x[0]$ es

$$\begin{aligned} \hat{x}[1|0] &= \frac{E(x[0]x[1])}{E(x^2[0])} x[0] \\ &= \frac{E[(A + w[0])(A + w[1])]}{E[(A + w[0])^2]} x[0] \\ &= \frac{E(A^2)}{E(A^2) + E(w^2[0])} x[0] \\ &= \frac{\sigma_A^2}{\sigma_A^2 + \sigma^2} x[0], \end{aligned}$$

donde se consideró que A , $w[0]$ y $w[1]$ son independientes y de media nula. El vector de error $\tilde{x}[1] = x[1] - \hat{x}[1|0]$, que tiene la dirección de $\Delta\hat{A}[1]$, representa la nueva información que $x[1]$ contribuye a la estimación de A , y por esa razón se llama *innovación*. La proyección de A sobre este vector es la

corrección buscada,

$$\begin{aligned}\Delta\hat{A}[1] &= \left(A, \frac{\tilde{x}[1]}{\|\tilde{x}[1]\|} \right) \frac{\tilde{x}[1]}{\|\tilde{x}[1]\|} \\ &= \frac{E(A, \tilde{x}[1])}{E(\tilde{x}^2[1])} \tilde{x}[1].\end{aligned}$$

Definiendo

$$K[1] = \frac{E(A, \tilde{x}[1])}{E(\tilde{x}^2[1])}, \quad (12.29)$$

se obtiene que el estimador LMMSE es

$$\hat{A}[1] = \hat{A}[0] + K[1](x[1] - \hat{x}[1|0]).$$

Otra forma de evaluar $\hat{x}[1|0]$ es notando que $x[1] = A + w[1]$. Por la propiedad de aditividad dada por la ecuación 12.17, se cumple que $\hat{x}[1|0] = \hat{A}[0] + \hat{w}[1|0]$. Como $w[1]$ no está correlacionado con $w[0]$ se cumple que $\hat{w}[1|0] = 0$, ya que el factor en el numerador del estimador (ecuación 12.28) es $E(w[0]w[1]) = E(w[0])E(w[1]) = 0$. Finalmente, $\hat{x}[1|0] = \hat{A}[0]$, y de la ecuación 12.28, $\hat{A}[0]$ es

$$\hat{A}[0] = \frac{E(Ax[0])}{E(x^2[0])} x[0] = \frac{E[A(A + w[0])]}{E[(A + w[0])^2]} x[0] = \frac{\sigma_A^2}{\sigma_A^2 + \sigma^2} x[0].$$

Por lo tanto,

$$\hat{A}[1] = \hat{A}[0] + K[1](x[1] - \hat{A}[0]),$$

y solo queda determinar la ganancia $K[1]$ para la corrección, que está definida por la ecuación 12.29. Como

$$\begin{aligned}\tilde{x}[1] &= x[1] - \hat{x}[1|0] \\ &= x[1] - \hat{A}[0] \\ &= x[1] - \frac{\sigma_A^2}{\sigma_A^2 + \sigma^2} x[0],\end{aligned}$$

la ganancia queda

$$\begin{aligned}K[1] &= \frac{E \left[A \left(x[1] - \frac{\sigma_A^2}{\sigma_A^2 + \sigma^2} x[0] \right) \right]}{E \left[\left(x[1] - \frac{\sigma_A^2}{\sigma_A^2 + \sigma^2} x[0] \right)^2 \right]} \\ &= \frac{E(Ax[1]) - \frac{\sigma_A^2}{\sigma_A^2 + \sigma^2} E(Ax[0])}{E(x^2[1]) - 2 \frac{\sigma_A^2}{\sigma_A^2 + \sigma^2} E(x[0]x[1]) + \left(\frac{\sigma_A^2}{\sigma_A^2 + \sigma^2} \right)^2 E(x^2[0])}.\end{aligned}$$

Considerando que

$$\begin{aligned}E(Ax[n]) &= E[A(A + w[n])] = E(A^2) = \sigma_A^2 \\ E(x[m]x[n]) &= E[(A + w[m])(A + w[n])] = E(A^2) = \sigma_A^2 \\ E(x^2[n]) &= E[(A + w[n])^2] = E(A^2) + E(w^2[n]) = \sigma_A^2 + \sigma^2,\end{aligned}$$

la ganancia queda

$$\begin{aligned}
K[1] &= \frac{\sigma_A^2 - \frac{\sigma_A^2}{\sigma_A^2 + \sigma^2} \sigma_A^2}{\sigma_A^2 + \sigma^2 - 2 \frac{\sigma_A^2}{\sigma_A^2 + \sigma^2} \sigma_A^2 + \left(\frac{\sigma_A^2}{\sigma_A^2 + \sigma^2} \right)^2 (\sigma_A^2 + \sigma^2)} \\
&\stackrel{(a)}{=} \frac{\sigma_A^2(\sigma_A^2 + \sigma^2) - \sigma_A^4}{(\sigma_A^2 + \sigma^2)^2 - 2\sigma_A^4 + \sigma_A^4} \\
&= \frac{\sigma_A^2 \sigma^2}{\sigma_A^4 + 2\sigma_A^2 \sigma^2 + \sigma^4 - \sigma_A^4} \\
&= \frac{\sigma_A^2}{2\sigma_A^2 + \sigma^2},
\end{aligned}$$

acorde a la ecuación 12.24 para $N = 1$. Se puede continuar con el procedimiento para determinar $\hat{A}[2], \hat{A}[3], \dots$. Los pasos del procedimiento son los siguientes:

1. encontrar el estimador LMMSE de A en base a $x[0]$, conduciendo a $\hat{A}[0]$.
2. encontrar el estimador LMMSE de $x[1]$ en base a $x[0]$, conduciendo $\hat{x}[1|0]$.
3. determinar la innovación del nuevo dato $x[1]$, que es $x[1] - \hat{x}[1|0]$.
4. sumar a $\hat{A}[0]$ el estimador LMMSE de A en base a la innovación, conduciendo $\hat{A}[1]$.
5. continuar el proceso.

En esencia, se genera el conjunto de variables aleatorias no correlacionadas o ortogonales, las innovaciones $\{x[0], x[1] - \hat{x}[1|0], x[2] - \hat{x}[2|0], 1, \dots\}$. Este procedimiento se llama ortogonalización de Gram-Schmidt (ver el problema de la sección 12.6.10). Para encontrar el estimador LMMSE de A en base a $\{x[0], \dots, x[N-1]\}$ simplemente se suman los estimadores individuales obteniendo

$$\hat{A}[N-1] = \sum_{n=0}^{N-1} K[n](x[n] - \hat{x}[n|0, \dots, n-1])$$

donde cada factor de ganancia es

$$K[n] = \frac{E[A(x[n] - \hat{x}[n|0, \dots, n-1])]}{E[(x[n] - \hat{x}[n|0, \dots, n-1])^2]}. \quad (12.30)$$

En forma secuencial, el estimador LMMSE es

$$\hat{A}[N] = \hat{A}[N-1] + K[N](x[N] - \hat{x}[N|0, \dots, N-1]). \quad (12.31)$$

Para completar el procedimiento se debe determinar $\hat{x}[N|0, \dots, N-1]$ y $K[N]$. Para calcular $\hat{x}[N|0, \dots, N-1]$, que es el estimador LMMSE de $x[N] = A + w[N]$ en base a $\{x[0], \dots, x[N-1]\}$ se emplea la propiedad de aditividad de la ecuación 12.17, resultando en

$$\hat{x}[N|0, \dots, N-1] = \hat{A}[N|0, \dots, N-1] + \hat{w}[N|0, \dots, N-1]$$

Pero $\hat{A}[N|0, \dots, N-1]$ es por definición el estimador LMMSE de A en el tiempo N en base a $\{x[0], \dots, x[N-1]\}$, o $\hat{A}[N-1]$, teniendo en cuenta que A no cambia durante el intervalo de observación. Además, $\hat{w}[N|0, \dots, N-1]$ es cero ya que $w[N]$ no está correlacionado con las muestras previas de los datos debido a que A y $w[n]$ son no correlacionados y $w[n]$ es ruido blanco ($\mathbf{C}_{w[N]\mathbf{x}} = \mathbf{0}$ en la ecuación 12.7). Por o tanto,

$$\hat{x}[N|0, \dots, N-1] = \hat{A}[N-1], \quad (12.32)$$

y la actualización de estimador es

$$\hat{A}[N] = \hat{A}[N-1] + K[N](x[N] - \hat{A}[N-1]). \quad (12.33)$$

Para encontrar la ganancia se parte de las ecuaciones 12.30 y 12.32,

$$K[N] = \frac{E \left[A(x[N] - \hat{A}[N-1]) \right]}{E \left[(x[N] - \hat{A}[N-1])^2 \right]}.$$

Para evaluar el numerador, se observa que

$$\begin{aligned} E \left[A(x[N] - \hat{A}[N-1]) \right] &\stackrel{(a)}{=} E \left[(A - \hat{A}[N-1])(x[N] - \hat{A}[N-1]) \right] \\ &= E \left[(A - \hat{A}[N-1])(A + w[N] - \hat{A}[N-1]) \right] \\ &= E \left[(A - \hat{A}[N-1])^2 \right] + E \left[w[N](A - \hat{A}[N-1]) \right] \\ &\stackrel{(b)}{=} E \left[(A - \hat{A}[N-1])^2 \right] \\ &= \text{Bsme}(\hat{A}[N-1]), \end{aligned} \quad (12.34)$$

donde en (a) se tuvo en cuenta que $x[N] - \hat{A}[N-1] = x[N] - \hat{x}[N|0, \dots, N-1]$ es la innovación de $x[N]$, que es ortogonal a $\{x[0], \dots, x[N-1]\}$ y por lo tanto, también es ortogonal a $\hat{A}[N-1]$, que por definición de LMMSE es una combinación lineal de esos datos. Por lo tanto, $E[\hat{A}[N-1](x[N] - \hat{A}[N-1])] = 0$. En (b) se consideró nuevamente que $w[N]$ no está correlacionado con A ni con las muestras previas de datos. Además, el denominador es

$$\begin{aligned} E \left[(x[N] - \hat{A}[N-1])^2 \right] &= E \left[(A + w[N] - \hat{A}[N-1])^2 \right] \\ &= E(w^2[n]) + E \left[(A - \hat{A}[N-1])^2 \right] \\ &= \sigma^2 + \text{Bsme}(\hat{A}[N-1]), \end{aligned} \quad (12.35)$$

por lo que finalmente

$$K[N] = \frac{\text{Bsme}(\hat{A}[N-1])}{\sigma^2 + \text{Bsme}(\hat{A}[N-1])}. \quad (12.36)$$

La actualización de MSE mínimo puede obtenerse notando que

$$\begin{aligned} \text{Bsme}(\hat{A}[N]) &= E \left[(A - \hat{A}[N])^2 \right] \\ &\stackrel{(a)}{=} E \left[\left(A - \hat{A}[N-1] + K[N](x[N] - \hat{A}[N-1]) \right)^2 \right] \\ &= E \left[(A - \hat{A}[N-1])^2 \right] - 2K[N]E \left[(A - \hat{A}[N-1])(x[N] - \hat{A}[N-1]) \right] \\ &\quad + K^2[N]E \left[(x[N] - \hat{A}[N-1])^2 \right] \\ &\stackrel{(b)}{=} \text{Bsme}(\hat{A}[N-1]) - 2K[N] \text{Bsme}(\hat{A}[N-1]) + K^2[N] (\sigma^2 + \text{Bsme}(\hat{A}[N-1])) \\ &\stackrel{(c)}{=} \text{Bsme}(\hat{A}[N-1]) - 2K[N] \text{Bsme}(\hat{A}[N-1]) + K[N] \text{Bsme}(\hat{A}[N-1]), \end{aligned}$$

donde en (a) se empleó la ecuación 12.33 de la actualización del estimador, en (b) se sustituyó la esperanza del segundo sumando empleando la ecuación 12.34 y la esperanza del tercer sumando empleando la ecuación 12.35 y en (c) se empleó la ecuación 12.36 que indica que $K[N](\sigma^2 + \text{Bsme}(\hat{A}[N-1])) = \text{Bsme}(\hat{A}[N-1])$. Finalmente, la actualización del MSE mínimo es

$$\text{Bsme}(\hat{A}[N]) = (1 - K[N]) \text{Bsme}(\hat{A}[N-1]).$$

12.4.1. Estimador LMMSE secuencial para el modelo lineal bayesiano

Como generalización de los resultados, en esta sección se deduce el estimador LMMSE secuencial para el modelo lineal bayesiano. El estimador LMMSE para el modelo lineal bayesiano está dado por el

teorema de Gauss-Markov bayesiano (sección 12.4). Para deducir el estimador secuencial, se asume que los datos tienen el modelo lineal bayesiano y además que \mathbf{C}_w es una matriz diagonal, es decir, las muestras de ruido $w[n]$ no están correlacionadas y tienen varianza $E(w[n]) = \sigma_n^2$. Esta hipótesis es crítica para la deducción del estimador secuencial y también fue asumida en el ejemplo anterior. De forma fortuita, el conjunto de ecuaciones obtenido sirve también para el caso en que las medias no son nulas. Por lo tanto, las ecuaciones son la implementación secuencial del estimador LMSSE vectorial de las ecuaciones 12.18 o 12.19 cuando la matriz de covarianza \mathbf{C}_w del ruido es diagonal. Para definir las ecuaciones, sea $\hat{\boldsymbol{\theta}}[n]$ el estimador LMMSE basado en las muestras $\{x[0], \dots, x[n]\}$ y $\mathbf{M}[n]$ la matriz de MSE mínimo correspondiente, que la es versión secuencial de las ecuaciones 12.20 o 12.21, o

$$\mathbf{M}[n] = E \left[(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}[n])(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}[n])^T \right].$$

Además, la matriz $(n+1) \times p$ de observación se particiona como

$$\mathbf{H}[n] = \begin{bmatrix} \mathbf{H}[n-1] \\ \mathbf{h}^T[n] \end{bmatrix} = \begin{bmatrix} n \times p \\ 1 \times p \end{bmatrix}.$$

Entonces, el estimador LMMSE secuencial es

Actualización del estimador:

$$\hat{\boldsymbol{\theta}}[n] = \hat{\boldsymbol{\theta}}[n-1] + \mathbf{K}[n](x[n] - \mathbf{h}^T[n]\hat{\boldsymbol{\theta}}[n-1]), \quad (12.37)$$

donde

$$\mathbf{K}[n] = \frac{\mathbf{M}[n-1]\mathbf{h}[n]}{\sigma_n^2 + \mathbf{h}^T\mathbf{M}[n-1]\mathbf{h}[n]}, \quad (12.38)$$

Actualización de la matriz de MSE mínimo:

$$\mathbf{M}[n] = (\mathbf{I} - \mathbf{K}[n]\mathbf{h}^T[n])\mathbf{M}[n-1]. \quad (12.39)$$

El factor de ganancia $\mathbf{K}[n]$ es un vector $p \times 1$ y la matriz de MSE mínimo tiene dimensiones $p \times p$. El algoritmo se resume en la figura 12.5. Para comenzar la recursión se necesita especificar los valores iniciales de $\hat{\boldsymbol{\theta}}[n-1]$ y $\mathbf{M}[n-1]$ de forma de determinar $\mathbf{K}[n]$ con la ecuación 12.38 y luego $\hat{\boldsymbol{\theta}}[n]$ con la ecuación 12.37. Para hacerlo, se especifica $\hat{\boldsymbol{\theta}}[-1]$ y $\mathbf{M}[-1]$ de forma de comenzar la recursión en $n = 0$. Debido a que no se observaron datos en $n = -1$, el estimador es la constante $\theta_i = a_{i0}$ (ver la ecuación 12.12). De la ecuación 12.3 se deduce que el estimador LMMSE es la media de θ_i , o

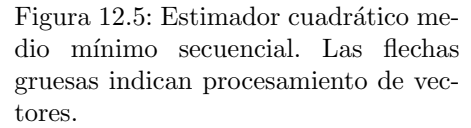
$$\hat{\boldsymbol{\theta}}[-1] = E(\boldsymbol{\theta}).$$

Como resultado, la matriz de MSE mínimo es

$$\begin{aligned} \mathbf{M}[-1] &= E \left[(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}[-1])(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}[-1])^T \right] \\ &= \mathbf{C}_{\theta\theta}. \end{aligned}$$

Algunas observaciones son las siguientes:

1. Si no se tiene conocimiento a priori de $\boldsymbol{\theta}$ puede establecerse $\mathbf{C}_{\theta\theta} \rightarrow \infty$. En ese caso, se tiene la misma forma que el LSE secuencial (ver la sección 8.4), aunque ambos enfoques son sustancialmente diferentes.
2. No se requiere inversión de matrices.
3. El factor de ganancia $\mathbf{K}[n]$ depende de la confianza en la nueva muestra, dada por σ_n^2 , contra la confianza en los datos previos, resumida en $\mathbf{M}[n-1]$.


$$\hat{\theta}_i[n] = \hat{\theta}_i[n-1] + K_i[n](x[n] - \hat{x}[n|n-1]), \quad (12.40)$$
$$\hat{\boldsymbol{\theta}}[n] = \hat{\boldsymbol{\theta}}[n-1] + \mathbf{K}[n](x[n] - \hat{x}[n|n-1]). \quad (12.41)$$
$$E \left[\hat{\theta}_i[n-1](x[n] - \hat{x}[n|n-1]) \right] = 0.$$
$$E \left[\left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}[n-1] \right) w[n] \right] = 0.$$
$$x[n] = \mathbf{h}^T[n]\boldsymbol{\theta} + w[n].$$
$$\hat{x}[n|n-1] = \mathbf{h}^T[n] \hat{\boldsymbol{\theta}}[n-1] + \hat{w}[n|n-1].$$
$$\hat{x}[n|n-1] = \mathbf{h}^T[n]\hat{\boldsymbol{\theta}}[n-1].$$
$$K_i[n] = \frac{E[\theta_i(x[n] - \hat{x}[n|n-1])]}{E[(x[n] - \hat{x}[n|n-1])^2]}.$$

Evalutando el denominador,

$$\begin{aligned}
E[(x[n] - \hat{x}[n|n-1])^2] &= E[(\mathbf{h}^T[n]\boldsymbol{\theta} + w[n] - \mathbf{h}^T[n]\hat{\boldsymbol{\theta}}[n-1])^2] \\
&= E\left[\left(\mathbf{h}^T[n](\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}[n-1]) + w[n]\right)^2\right] \\
&\stackrel{(a)}{=} E\left[\left(\mathbf{h}^T[n](\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}[n-1])\right)^2\right] + E(w^2[n]) \\
&= \mathbf{h}^T[n] E\left[\left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}[n-1]\right)\left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}[n-1]\right)^T\right] \mathbf{h}[n] + E(w^2[n]) \\
&= \mathbf{h}^T[n] \mathbf{M}[n-1] \mathbf{h}[n] + \sigma_n^2,
\end{aligned} \tag{12.42}$$

donde en (a) se empleó la propiedad 2. El numerador es

$$\begin{aligned}
E[\theta_i(x[n] - \hat{x}[n|n-1])] &\stackrel{(a)}{=} E[(\theta_i - \hat{\theta}_i[n-1])(x[n] - \hat{x}[n|n-1])] \\
&= E[(\theta_i - \hat{\theta}_i[n-1])(\mathbf{h}^T[n]\boldsymbol{\theta} + w[n] - \mathbf{h}^T[n]\hat{\boldsymbol{\theta}}[n-1])] \\
&\stackrel{(b)}{=} E[(\theta_i - \hat{\theta}_i[n-1])\left(\mathbf{h}^T[n](\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}[n-1])\right)] \\
&= E[(\theta_i - \hat{\theta}_i[n-1])(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}[n-1])^T] \mathbf{h}[n],
\end{aligned} \tag{12.43}$$

donde en (a) se empleó la propiedad 1 y en (b) la propiedad 2. Combinando los resultados,

$$K_i[n] = \frac{E[(\theta_i - \hat{\theta}_i[n-1])(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}[n-1])^T] \mathbf{h}[n]}{\mathbf{h}^T[n] \mathbf{M}[n-1] \mathbf{h}[n] + \sigma_n^2},$$

y por lo tanto,

$$\mathbf{K}[n] = \frac{E[(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}[n-1])(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}[n-1])^T] \mathbf{h}[n]}{\mathbf{h}^T[n] \mathbf{M}[n-1] \mathbf{h}[n] + \sigma_n^2},$$

resultando en

$$\mathbf{K}[n] = \frac{\mathbf{M}[n-1] \mathbf{h}[n]}{\mathbf{h}^T[n] \mathbf{M}[n-1] \mathbf{h}[n] + \sigma_n^2}. \tag{12.44}$$

Para determinar la actualización de la matriz de MSE, se observa que

$$\begin{aligned}
\mathbf{M}[n] &= E[(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}[n])(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}[n])^T] \\
&\stackrel{(a)}{=} E\left[\left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}[n-1] - \mathbf{K}[n](x[n] - \hat{x}[n|n-1])\right)\left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}[n-1] - \mathbf{K}[n](x[n] - \hat{x}[n|n-1])\right)^T\right] \\
&= E\left[\left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}[n-1]\right)\left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}[n-1]\right)^T\right] - E\left[\left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}[n-1]\right)\left(\mathbf{K}[n](x[n] - \hat{x}[n|n-1])\right)^T\right] \\
&\quad - E\left[\left(\mathbf{K}[n](x[n] - \hat{x}[n|n-1])\right)\left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}[n-1]\right)^T\right] \\
&\quad + E\left[\left(\mathbf{K}[n](x[n] - \hat{x}[n|n-1])\right)\left(\mathbf{K}[n](x[n] - \hat{x}[n|n-1])\right)^T\right] \\
&= \mathbf{M}[n-1] - E\left[\left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}[n-1]\right)(x[n] - \hat{x}[n|n-1])\right] \mathbf{K}^T[n] \\
&\quad - \mathbf{K}[n] E\left[(x[n] - \hat{x}[n|n-1])\left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}[n-1]\right)^T\right] + \mathbf{K}[n] E[(x[n] - \hat{x}[n|n-1])^2] \mathbf{K}^T[n],
\end{aligned}$$

donde en (a) se empleó la ecuación 12.41 de la actualización del estimador. De las ecuaciones 12.42 y 12.44 se tiene que

$$\mathbf{K}[n] = \frac{\mathbf{M}[n-1] \mathbf{h}[n]}{E[(x[n] - \hat{x}[n|n-1])^2]} \Rightarrow \mathbf{K}[n] E[(x[n] - \hat{x}[n|n-1])^2] = \mathbf{M}[n-1] \mathbf{h}[n].$$

y además

$$\begin{aligned} E \left[\left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}[n-1] \right) (x[n] - \hat{x}[n|n-1]) \right] &\stackrel{(a)}{=} E \left[\boldsymbol{\theta} (x[n] - \hat{x}[n|n-1]) \right] \\ &\stackrel{(b)}{=} \mathbf{M}[n-1] \mathbf{h}[n], \end{aligned}$$

donde en (a) se empleó la propiedad 1 y en (b) la ecuación 12.43. Sustituyendo estos dos resultados en la ecuación de la actualización de la matriz de MSE, se tiene que

$$\begin{aligned} \mathbf{M}[n] &= \mathbf{M}[n-1] - \mathbf{M}[n-1] \mathbf{h}[n] \mathbf{K}^T[n] - \mathbf{K}[n] \mathbf{h}^T[n] \mathbf{M}[n-1] + \mathbf{M}[n-1] \mathbf{h}[n] \mathbf{K}^T[n] \\ &= (\mathbf{I} - \mathbf{K}[n] \mathbf{h}^T[n]) \mathbf{M}[n-1]. \end{aligned}$$

Se mostrará ahora que se obtiene el mismo conjunto de ecuaciones si $\boldsymbol{\theta}$ y $x[n]$ no tienen media nula. Como la implementación secuencial de la matriz de MSE mínimo de la ecuación 12.39 debe coincidir con la implementación no secuencial dada por la ecuación 12.20 y ésta no depende de las medias, $\mathbf{M}[n]$ también debe ser independiente de las medias. Además, como $\mathbf{M}[n]$ depende de $\mathbf{K}[n]$, el vector de ganancia también tiene que ser independiente de las medias. La ecuación 12.37 de actualización del estimador es válida si las medias son nulas. Si las medias no son nulas, igual es posible aplicar la ecuación 12.37 a los datos $y[n] = y[n] - E(x[n])$ para obtener el estimador de $\boldsymbol{\alpha} = \boldsymbol{\theta} - E(\boldsymbol{\theta})$,

$$\hat{\boldsymbol{\alpha}}[n] = \hat{\boldsymbol{\alpha}}[n-1] + \mathbf{K}[n](y[n] - \mathbf{h}^T[n] \hat{\boldsymbol{\alpha}}[n-1]).$$

Pero por la propiedad conmutativa del LMMSE (ecuación 12.16), el estimador de $\boldsymbol{\theta} + \mathbf{b}$ con \mathbf{b} una constante es $\hat{\boldsymbol{\theta}} + \mathbf{b}$, y por lo tanto, $\hat{\boldsymbol{\alpha}} = \hat{\boldsymbol{\theta}} - E(\boldsymbol{\theta})$. Teniendo esto en cuenta, la ecuación de actualización de estimador es

$$\hat{\boldsymbol{\theta}}[n] - E(\boldsymbol{\theta}) = \hat{\boldsymbol{\theta}}[n-1] - E(\boldsymbol{\theta}) + \mathbf{K}[n] \left[x[n] - E(x[n]) - \mathbf{h}^T[n] (\hat{\boldsymbol{\theta}}[n-1] - E(\boldsymbol{\theta})) \right],$$

donde $\boldsymbol{\theta}[n]$ es el estimador LMMSE para medias no nulas. Operando, se tiene que

$$\hat{\boldsymbol{\theta}}[n] = \hat{\boldsymbol{\theta}}[n-1] + \mathbf{K}[n] \left[x[n] - \mathbf{h}^T[n] \hat{\boldsymbol{\theta}}[n-1] - (E(x[n]) - \mathbf{h}^T[n] E(\boldsymbol{\theta})) \right],$$

y como

$$\begin{aligned} E(x[n]) &= E(\mathbf{h}^T[n] \boldsymbol{\theta} + w[n]) \\ &= \mathbf{h}^T[n] E(\boldsymbol{\theta}), \end{aligned}$$

se obtiene que

$$\hat{\boldsymbol{\theta}}[n] = \hat{\boldsymbol{\theta}}[n-1] + \mathbf{K}[n] \left(x[n] - \mathbf{h}^T[n] \hat{\boldsymbol{\theta}}[n-1] \right),$$

que coincide con la ecuación de actualización para el caso con medias nulas.

12.4.2. Análisis de Fourier bayesiano

Se empleará el LS secuencial para calcular el estimador LMMSE del ejemplo de la sección 11.2.1. El modelo de los datos es

$$x[n] = a \cos 2\pi f_0 n + b \sin 2\pi f_0 n + w[n], \quad n = 0, \dots, N-1,$$

donde f_0 es cualquier valor (en el ejemplo de la sección 11.2.1 se empleó la hipótesis de que f_0 es un múltiplo de $1/N$), excepto 0 y $1/2$, para los cuales $\sin 2\pi f_0 n = 0$, y $w[n]$ es ruido blanco con varianza σ^2 . Se desea estimar $\boldsymbol{\theta} = [a \ b]^T$ secuencialmente. Se asume además que se aplica el modelo lineal bayesiano y que $\boldsymbol{\theta}$ tiene media $E(\boldsymbol{\theta})$ y covarianza $\sigma_{\boldsymbol{\theta}}^2 \mathbf{I}$. El estimador secuencial se inicializa como

$$\begin{aligned} \hat{\boldsymbol{\theta}}[-1] &= E(\boldsymbol{\theta}) \\ \mathbf{M}[-1] &= \mathbf{C}_{\boldsymbol{\theta}\boldsymbol{\theta}} = \sigma_{\boldsymbol{\theta}}^2 \mathbf{I}. \end{aligned}$$

Por lo tanto, de la ecuación 12.37

$$\hat{\boldsymbol{\theta}}[0] = \hat{\boldsymbol{\theta}}[-1] + \mathbf{K}[0](x[0] - \mathbf{h}^T[0]\hat{\boldsymbol{\theta}}[-1]).$$

El vector $\mathbf{h}^T[0]$ es la primer fila de la matriz de observación,

$$\mathbf{h}^T[0] = [1 \quad 0].$$

Las filas subsecuentes son

$$\mathbf{h}^T[0] = [\cos 2\pi f_0 n \quad \sin 2\pi f_0 n].$$

El vector de ganancia tiene dimensiones 2×1 y en la primer iteración es

$$\mathbf{K}[0] = \frac{\mathbf{M}[-1]\mathbf{h}[0]}{\sigma^2 + \mathbf{h}^T[0]\mathbf{M}[-1]\mathbf{h}[0]}$$

donde $\mathbf{M}[-1] = \sigma_{\theta}^2$. Luego de calcular el vector ganancia, puede calcularse $\hat{\boldsymbol{\theta}}[0]$. Finalmente, la actualización de la matriz de MSE mínimo es (ecuación 12.39)

$$\mathbf{M}[0] = (\mathbf{I} - \mathbf{K}[0]\mathbf{h}^T[0])\mathbf{M}[-1],$$

y el procedimiento continúa de forma similar para $n \geq 1$. Se realizó una simulación en computadora del cálculo del estimador LMSSE secuencial para el análisis de Fourier bayesiano. El valor verdadero del parámetro a estimar es

$$\boldsymbol{\theta} = \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \end{bmatrix},$$

y la varianza del ruido es $\sigma^2 = 2$. Los parámetros de inicialización del algoritmo son

$$\hat{\boldsymbol{\theta}}[-1] = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \text{y} \quad \mathbf{M}[-1] = 1 \times 10^2 \mathbf{I}.$$

En la figura 12.6 se muestra el código la implementación del LMMSE secuencial para el análisis de Fourier bayesiano en lenguaje Python. En las gráficas de la figura 12.7 se muestran los resultados. Se observa

```
import numpy as np

## x es el vector de datos de largo N (x.shape=(N,))
# número de parámetros
p = 2
# condiciones iniciales del LMMSE secuencial
theta_i = np.zeros((p, ))
M_i = 1e2 * np.eye(p)
# matriz de MSE mínimo
M = M_i
# estimador
hat_theta = theta_i
for n in np.arange(N):
    # fila n-ésima de la matriz de observación
    h = np.array([np.cos(2 * np.pi * f0 * n), np.sin(2 * np.pi * f0 * n)])
    # cálculo del vector ganancia
    K = (M @ h) / (var + h @ M @ h)
    # cálculo de la matriz de MSE mínimo
    M = (np.eye(p) - np.outer(K, h)) @ M
    # cálculo del estimador
    hat_theta = hat_theta + K * (x[n] - h @ hat_theta)
```

Figura 12.6: Implementación en el lenguaje Python del estimador secuencial para el análisis de Fourier bayesiano.

que los parámetros estimados se acercan al los valores verdaderos a medida que avanza n . También se observa que los componetes de la ganancia $\mathbf{K}[n]$ decrecen con n , ya que a medida que se disponen mas datos, la contribución de un nuevo dato es cada vez mas irrelevante.

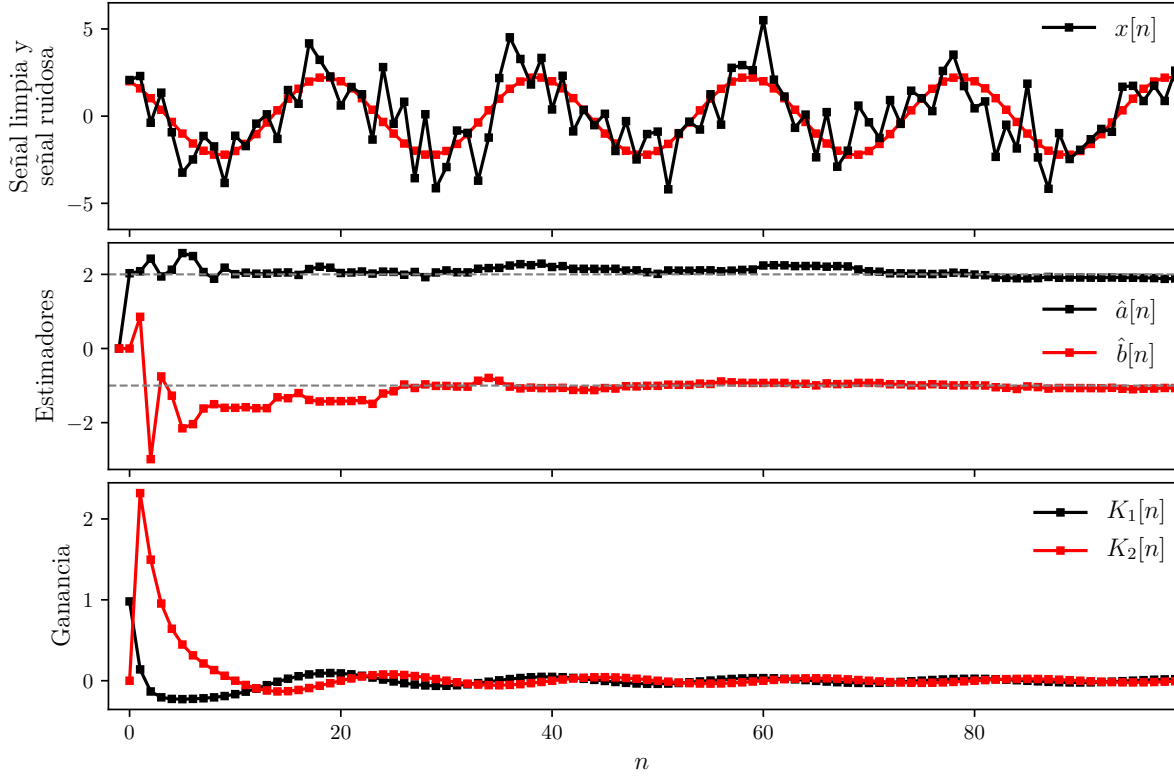


Figura 12.7: Análisis de Fourier bayesiano mediante estimación secuencial.

12.5. Ejemplos de procesamiento de señales: filtrado de Wiener

En esta sección se describe un grupo de aplicaciones del estimador LMMSE denominado genéricamente *filtro de Wiener*. Se asume que los datos $\{x[0], \dots, x[N-1]\}$ son WSS de media nula. Por lo tanto, la matriz \mathbf{C}_{xx} de covarianza $N \times N$ es Toeplitz y simétrica de la forma

$$\mathbf{C}_{xx} = \begin{bmatrix} r_{xx}[0] & r_{xx}[1] & \dots & r_{xx}[N-1] \\ r_{xx}[1] & r_{xx}[0] & \dots & r_{xx}[N-2] \\ \vdots & \vdots & \ddots & \vdots \\ r_{xx}[N-1] & r_{xx}[N-2] & \dots & r_{xx}[0] \end{bmatrix} = \mathbf{R}_{xx}$$

donde $r_{xx}[k]$ es la ACF del proceso $x[n]$ y \mathbf{R}_{xx} denota a la matriz de autocorrelación. Adicionalmente, se asume que el parámetro θ a estimar también tiene media nula. Hay tres problemas de filtrado de Wiener. Estos son

1. *Filtrado*, donde se estima $\theta = s[n]$ en base a $x[m] = s[m] + w[m]$ para $m = 0, \dots, n$. Las secuencias $s[n]$ y $w[n]$ representan los procesos señal y ruido, y el problema es filtrar el ruido de la señal. Observar que cada muestra de la señal es estimada únicamente en base a la muestra actual y pasadas de los datos, y por lo tanto, el proceso de estimación puede verse como la aplicación de un filtro causal a los datos.
2. *Suavizado*, donde se estima $\theta = s[n]$ para $n = 0, \dots, N-1$ en base al conjunto de datos $\{x[0], \dots, x[N-1]\}$, donde $x[n] = s[n] + w[n]$. En contraste al problema de filtrado, en este caso se emplean además datos futuros. Por lo tanto, en el problema de suavizado el estimador no puede obtenerse hasta que se recolectaron todos los datos.
3. *Predicción*, donde se estima $\theta = x[N-1+l]$ para l un entero positivo en base a $\{x[0], \dots, x[N-1]\}$. Este problema es referido como predicción de paso l . Un problema vinculado es la estimación de $x[n]$

en base a $\{x[0], \dots, x[n-1], x[n+1], \dots, x[N-1]\}$ y se denomina *interpolación*. Éste se explora en el problema de la sección 12.6.14.

Para resolver los tres problemas se emplea la ecuación 12.13 con $E(\boldsymbol{\theta}) = E(\mathbf{x}) = 0$, o

$$\hat{\boldsymbol{\theta}} = \mathbf{C}_{\theta x} \mathbf{C}_{xx}^{-1} \mathbf{x}, \quad (12.45)$$

y la matriz de MSE bayesiano mínimo está dada la ecuación 12.14

$$\mathbf{M}_{\hat{\boldsymbol{\theta}}} = \mathbf{C}_{\theta\theta} - \mathbf{C}_{\theta x} \mathbf{C}_{xx}^{-1} \mathbf{C}_{x\theta}. \quad (12.46)$$

Considérese primero el problema de suavizado, en el cual se quiere estimar $\boldsymbol{\theta} = \mathbf{s} = [s[0] \dots s[N-1]]^T$ en base a $\mathbf{x} = [x[0] \dots x[N-1]]^T$. Se asume la razonable hipótesis que la señal y el ruido son procesos no correlacionados, y por lo tanto,

$$r_{xx}[k] = r_{ss}[k] + r_{ww}[k].$$

De esta forma, se tiene que

$$\mathbf{C}_{xx} = \mathbf{R}_{xx} = \mathbf{R}_{ss} + \mathbf{R}_{ww}.$$

Además,

$$\mathbf{C}_{\theta x} = E(\mathbf{s}\mathbf{x}^T) = E[\mathbf{s}(\mathbf{s} + \mathbf{w})^T] = E(\mathbf{s}\mathbf{s}^T) + E(\mathbf{s}\mathbf{w}^T) = \mathbf{R}_{ss},$$

donde en la última igualdad se consideró que \mathbf{s} y \mathbf{w} son no correlacionados y \mathbf{w} tiene media nula. Por lo tanto, de la ecuación 12.45 se obtiene que el estimador de Wiener es

$$\hat{\mathbf{s}} = \mathbf{R}_{ss}(\mathbf{R}_{ss} + \mathbf{R}_{ww})^{-1} \mathbf{x}.$$

La matriz $N \times N$

$$\mathbf{W} = \mathbf{R}_{ss}(\mathbf{R}_{ss} + \mathbf{R}_{ww})^{-1}$$

es referida como *matriz de suavizado de Wiener*. De la ecuación 12.46, la matriz de MSE bayesiano mínimo es

$$\begin{aligned} \mathbf{M}_{\hat{\mathbf{s}}} &= \mathbf{R}_{ss} - \mathbf{R}_{ss}(\mathbf{R}_{ss} + \mathbf{R}_{ww})^{-1} \mathbf{R}_{ss} \\ &= (\mathbf{I} - \mathbf{W}) \mathbf{R}_{ss}. \end{aligned}$$

Como ejemplo, si $N = 1$, se quiere estimar $s[0]$ en base a $x[0] = s[0] + w[0]$. En este caso, la matriz \mathbf{W} de suavizado de Wiener es un escalar W dado por

$$W = \frac{r_{ss}[0]}{r_{ss}[0] + r_{ww}[0]} = \frac{\eta}{\eta + 1} \quad (12.47)$$

donde $\eta = r_{ss}[0]/r_{ww}[0]$ es la SNR. Teniendo en cuenta que el estimador es $\hat{s}[0] = Wx[0]$, si la SNR es alta de forma que $W \rightarrow 1$, se tiene que $\hat{s}[0] \rightarrow x[0]$ mientras que si la SNR es baja de forma que $W \rightarrow 0$, se cumple que $\hat{s}[0] \rightarrow 0$. El correspondiente MSE mínimo es

$$M_{\hat{s}} = (1 - W)r_{ss}[0] = \left(1 - \frac{\eta}{\eta + 1}\right) r_{ss}[0]$$

que para ambos extremos es 0 si la SNR es grande o $r_{ss}[0]$ si la SNR es baja. Un ejemplo numérico del suavizado de Wiener se da en la sección 11.5. Ver además el problema de la sección 12.6.15.

Se considera ahora el problema de filtrado, en el cual se quiere estimar $\theta = s[n]$ en base a $\mathbf{x} = [x[0] \dots x[n]]^T$. El problema se repite para cada valor de n hasta estimar la señal $s[n]$ completa para $n = 0, \dots, N-1$. Como antes, $x[n] = s[n] + w[n]$, donde $s[n]$ y $w[n]$ son los procesos señal y ruido respectivamente, y no están correlacionados entre si. Por lo tanto,

$$\mathbf{C}_{xx} = \mathbf{R}_{ss} + \mathbf{R}_{ww}$$

donde \mathbf{R}_{ss} y \mathbf{R}_{ww} son las matrices de autocorrelación $(n+1) \times (n+1)$. Además

$$\begin{aligned} \mathbf{C}_{\theta x} &= E(s[n][x[0] \ x[1] \ \dots \ x[n]]) \\ &= E(s[n][s[0] \ s[1] \ \dots \ s[n]]) \\ &= [r_{ss}[n] \ r_{ss}[n-1] \ \dots \ r_{ss}[0]]. \end{aligned}$$

Denotando a este vector como \mathbf{r}'_{ss} , de la ecuación 12.45 el estimador es

$$\hat{s}[n] = \mathbf{r}'_{ss}(\mathbf{R}_{ss} + \mathbf{R}_{ww})^{-1}\mathbf{x}.$$

El vector $(n+1) \times 1$ de pesos es

$$\mathbf{a} = (\mathbf{R}_{ss} + \mathbf{R}_{ww})^{-1}\mathbf{r}'_{ss}$$

considerando que

$$\hat{s}[n] = \mathbf{a}^T \mathbf{x}$$

donde $\mathbf{a} = [a_0 \dots a_n]$ como en la formulación original del estimador LMMSE escalar de la ecuación 12.1. A medida que crece n , el proceso de estimación puede verse como una operación de filtrado con un filtro de respuesta al impulso $h^{(n)}[k]$ variable en el tiempo. Sea

$$h^{(n)}[k] = a_{n-k}, \quad k = 0, \dots, n.$$

Observar que el vector $\mathbf{h} = [h^{(n)}[0] \dots h^{(n)}[n]]^T$ es el vector \mathbf{a} invertido temporalmente. Por lo tanto,

$$\begin{aligned} \hat{s}[n] &= \sum_{k=0}^n a_k x[k] \\ &= \sum_{k=0}^n h^{(n)}[n-k] x[k], \end{aligned}$$

o equivalentemente,

$$\hat{s}[n] = \sum_{k=0}^n h^{(n)}[k] x[n-k],$$

que se reconoce como un filtrado FIR variable en el tiempo. Para encontrar explícitamente la respuesta al impulso \mathbf{h} se observa que

$$(\mathbf{R}_{ss} + \mathbf{R}_{ww})\mathbf{a} = \mathbf{r}'_{ss}$$

y por lo tanto,

$$(\mathbf{R}_{ss} + \mathbf{R}_{ww})\mathbf{h} = \mathbf{r}_{ss},$$

donde $\mathbf{r}_{ss} = [r_{ss}[0] \dots r_{ss}[n]]^T$. Esto se debe a la estructura Toeplitz simétrica de la matriz $\mathbf{R}_{ss} + \mathbf{R}_{ww}$ y al hecho que de \mathbf{h} es \mathbf{a} invertido y \mathbf{r}'_{ss} es \mathbf{r}_{ss} invertido. Para ver que la equivalencia entre las dos ecuaciones es cierta, considérese los sistemas

$$\mathbf{A}\mathbf{x} = \mathbf{b} \quad \text{y} \quad \mathbf{A}\mathbf{x}' = \mathbf{b}',$$

donde \mathbf{A} es una matriz Toeplitz simétrica $(n+1) \times (n+1)$ y \mathbf{x}' y \mathbf{b}' son respectivamente los vectores \mathbf{x} y \mathbf{b} invertidos. Se mostrará que ambos sistemas son idénticos. La ecuación i -ésima de cada sistema, con $i = 0, \dots, n$, es

$$\sum_{j=0}^n a_{ij} x_j = b_i, \quad \text{y} \quad \sum_{j=0}^n a_{ij} x'_j = b'_i.$$

Se operará con la segunda ecuación con el objetivo de obtener la primera. Como \mathbf{x}' y \mathbf{b}' son respectivamente los vectores \mathbf{x} y \mathbf{b} invertidos, se cumple que $x'_j = x_{n-j}$ y $b'_i = b_{n-i}$, y por lo tanto, la ecuación del segundo sistema se puede escribir como

$$\sum_{j=0}^n a_{ij} x_{n-j} = b_{n-i} \quad \stackrel{(a)}{\Rightarrow} \quad \sum_{l=0}^n a_{n-k, n-l} x_l = b_k \quad \stackrel{(b)}{\Rightarrow} \quad \sum_{l=0}^n a_{n-k+m, n-l+m} x_l = b_k,$$

donde en (a) se realizó el cambio de variable $k = n - i$ y $l = n - j$ y en (b) se sumó el valor m al número de fila y de columna de los elementos de la matriz, lo que mantiene la igualdad debido a la naturaleza Toeplitz de \mathbf{A} . Eligiendo m de forma que $n - k + m = l$, y por lo tanto, $n - l + m = k$ se obtiene que

$$\sum_{l=0}^n a_{lk} x_l = b_k \quad \Rightarrow \quad \sum_{j=0}^n a_{ij} x_j = b_i,$$

donde en el último paso se cambió el nombre de las variables a $k = i$ y $l = j$ y se tuvo en cuenta que la matriz es simétrica, por lo que $a_{ji} = a_{ij}$. Se concluye que ambos sistemas son idénticos. Continuando con el filtro de Wiener, el sistema de ecuaciones obtenido es

$$\begin{bmatrix} r_{xx}[0] & r_{xx}[1] & \cdots & r_{xx}[n] \\ r_{xx}[1] & r_{xx}[0] & \cdots & r_{xx}[n-1] \\ \vdots & \vdots & \ddots & \vdots \\ r_{xx}[n] & r_{xx}[n-1] & \cdots & r_{xx}[0] \end{bmatrix} \begin{bmatrix} h^{(n)}[0] \\ h^{(n)}[1] \\ \vdots \\ h^{(n)}[n] \end{bmatrix} = \begin{bmatrix} r_{ss}[0] \\ r_{ss}[1] \\ \vdots \\ r_{ss}[n] \end{bmatrix}.$$

donde $r_{xx}[k] = r_{ss}[k] + r_{ww}[k]$. Estas son las *ecuaciones de filtrado de Wiener-Hopf*. Es aparente que el sistema de ecuaciones debe ser resuelto para cada n . Sin embargo, un algoritmo computacionalmente eficiente para hacerlo es la recursión de Levinson, que resuelve las ecuaciones recursivamente evitando hacerlo para cada n . Para un largo n suficientemente grande, puede mostrarse que el filtro se hace invariante en el tiempo, por lo que se necesita una única solución. Las ecuaciones de filtrado de Wiener-Hopf pueden escribirse como

$$\sum_{k=0}^n h^{(n)}[k] r_{xx}[l-k] = r_{ss}[l], \quad l = 0, \dots, n,$$

donde se empleó la propiedad $r_{xx}[-k] = r_{xx}[k]$. Con $n \rightarrow \infty$, luego de reemplazar $h^{(n)}[k]$ por su versión $h[k]$ invariante en el tiempo, se obtiene que

$$\sum_{k=0}^n h[k] r_{xx}[l-k] = r_{ss}[l], \quad l = 0, 1, \dots \quad (12.48)$$

El mismo sistema de ecuaciones se obtiene si se intenta estimar $s[n]$ en base a la muestra presente y las infinitas muestras pasadas, o en base a $x[m]$ para $m \leq n$. Éste se llama *filtro de Wiener infinito*. Para ver porque es así, el estimador en este caso es

$$\hat{s}[n] = \sum_{k=0}^{\infty} h[k] x[n-k], \quad (12.49)$$

y considerando el principio de ortogonalidad (ecuación 12.10), se cumple que

$$s[n] - \hat{s}[n] \perp \dots, x[n-1], x[n]$$

o empleando la definición de ortogonalidad

$$E[(s[n] - \hat{s}[n])x[n-l]] = 0, \quad l = 0, 1, \dots,$$

es decir

$$E(\hat{s}[n]x[n-l]) = E(s[n]x[n-l]), \quad l = 0, 1, \dots$$

Sustituyendo el estimador, se obtiene que

$$E\left(\sum_{k=0}^{\infty} h[k] x[n-k] x[n-l]\right) = E(s[n]x[n-l]),$$

por lo que el sistema de ecuaciones a resolver para obtener la respuesta al impulso del filtro de Wiener infinito es

$$\sum_{k=0}^{\infty} h[k] r_{xx}[l-k] = r_{ss}[l], \quad l = 0, 1, \dots$$

Es razonable el hecho de que las soluciones sean idénticas, ya que el problema de estimar $s[n]$ en base a $x[m]$ para $0 \leq m \leq n$ con $n \rightarrow \infty$ es lo mismo que emplear la muestra actual y las infinitas muestras pasadas. La solución de la ecuación 12.49 emplea la factorización espectral y se explora en el problema de la sección 12.6.16. A primera vista, podría pensarse que la ecuación 12.49 podría resolverse mediante

la transformada de Fourier, ya que el lado izquierdo de la igualdad es una convolución. Sin embargo no es así debido a que las ecuaciones son válidas únicamente para $l \geq 0$. Si las ecuaciones fueran válidas además para $l < 0$, la transformada de Fourier sería una estrategia viable. Ese conjunto de ecuaciones surge en el problema de suavizado, en el cual $s[n]$ es estimado en base a $\{\dots, x[-1], x[0], x[1], \dots\}$ o a $x[k]$ para todo k . En este caso, el estimador de suavizado tiene la forma

$$\hat{s}[n] = \sum_{k=-\infty}^{\infty} a_k x[k],$$

y definiendo $h[k] = a_{n-k}$, se obtiene la convolución

$$\hat{s}[n] = \sum_{k=-\infty}^{\infty} h[k] x[n-k],$$

donde $h[k]$ es la respuesta al impulso de un filtro invariante en el tiempo bilateral infinito. Las ecuaciones de Wiener-Hopf se deducen en el problema de la sección 12.6.17 y son

$$\sum_{k=-\infty}^{\infty} h[k] r_{xx}[l-k] = r_{ss}[l], \quad -\infty < l < \infty. \quad (12.50)$$

Considerando que ahora

$$(h * r_{xx})[n] = r_{ss}[n],$$

puede emplearse la transformada de Fourier para encontrar la respuesta al impulso, obteniendo que

$$\begin{aligned} H(f) &= \frac{P_{ss}(f)}{P_{xx}(f)} \\ &= \frac{P_{ss}(f)}{P_{ss}(f) + P_{ww}(f)}. \end{aligned}$$

La respuesta en frecuencia es una función real y par de la frecuencia, por lo que la respuesta al impulso debe ser también real y par. El filtro de suavizado de Wiener enfatiza las partes del espectro de los datos en donde la SNR es alta y atenúa las partes donde es baja. Esto se evidencia si se define la “SNR local” como

$$\eta(f) = \frac{P_{ss}(f)}{P_{ww}(f)},$$

que es la SNR en una banda angosta de frecuencias centrada en f . De esta forma, la respuesta en frecuencia del filtro óptimo es

$$H(f) = \frac{\eta(f)}{\eta(f) + 1}.$$

Se observa que la respuesta en frecuencia cumple que $0 < H(f) < 1$, con $H(f) \approx 0$ cuando $\eta(f) \approx 0$ (baja SNR local) y $H(f) \approx 1$ cuando $\eta(f) \rightarrow \infty$ (alta SNR local). Comparar este resultado con el obtenido para el filtro de suavizado escalar de la ecuación 12.47. Ver además el problema de la sección 12.6.18 por resultados adicionales sobre el filtro de suavizado infinito de Wiener.

Por último, se examina el problema de predicción, en el cual se quiere estimar $x[N-1+l]$ para $l \geq 1$ en base a $\mathbf{x} = [x[0] \dots x[N-1]]^T$. El estimador resultante se denomina *predictor lineal de paso l*. El estimador está dado por la ecuación 12.45 con

$$\mathbf{C}_{xx} = \mathbf{R}_{xx},$$

donde \mathbf{R}_{xx} es de dimensión $N \times N$ y

$$\begin{aligned} \mathbf{C}_{\theta x} &= E[x[N-1+l][x[0] x[1] \dots x[N-1]]] \\ &= [r_{xx}[N-1+l] \ r_{xx}[N-2+l] \ \dots \ r_{xx}[l]]. \end{aligned}$$

Denotando a este vector como $\mathbf{r}_{xx}'^T$, el estimador es

$$\begin{aligned}\hat{x}[N-1+l] &= \mathbf{r}_{xx}'^T \mathbf{R}_{xx}^{-1} \mathbf{x} \\ &= \mathbf{a}^T \mathbf{x}\end{aligned}$$

con

$$\mathbf{a} = \mathbf{R}_{xx}^{-1} \mathbf{r}_{xx}', \quad (12.51)$$

se tiene que

$$\hat{x}[N-1+l] = \sum_{k=0}^{N-1} a_k x[k].$$

Definiendo $h[N-k] = a_k$ para permitir una interpretación como filtrado, se obtiene que

$$\begin{aligned}\hat{x}[N-1+l] &= \sum_{k=0}^{N-1} h[N-k] x[k] \\ &\stackrel{(a)}{=} \sum_{m=1}^N h[m] x[N-m],\end{aligned} \quad (12.52)$$

donde en (a) se realizó el cambio de variable $m = N - k$. Se observa que la predicción de la muestra es la salida de un filtro con respuesta al impulso $h[n]$. De la ecuación 12.51, el sistema de ecuaciones a resolver es

$$\mathbf{R}_{xx} \mathbf{h} = \mathbf{r}_{xx},$$

donde \mathbf{h} y \mathbf{r}_{xx} son respectivamente los vectores \mathbf{a} y \mathbf{r}_{xx}' invertidos temporalmente. En forma explícita, el sistema de ecuaciones es

$$\begin{bmatrix} r_{xx}[0] & r_{xx}[1] & \dots & r_{xx}[N-1] \\ r_{xx}[1] & r_{xx}[0] & \dots & r_{xx}[N-2] \\ \vdots & \vdots & \ddots & \vdots \\ r_{xx}[N-1] & r_{xx}[N-2] & \dots & r_{xx}[0] \end{bmatrix} \begin{bmatrix} h[1] \\ h[2] \\ \vdots \\ h[N] \end{bmatrix} = \begin{bmatrix} r_{xx}[l] \\ r_{xx}[l+1] \\ \vdots \\ r_{xx}[N-1+l] \end{bmatrix},$$

o

$$\sum_{k=1}^N h[k] r_{xx}[m-k] = r_{xx}[m+l-1], \quad m = 1, \dots, N. \quad (12.53)$$

Estas son las ecuaciones de Wiener-Hopf para el predictor lineal de paso l en base a N muestras pasadas. Un método eficiente para resolverlas en la recursión de Levinson. Para el caso específico en que $l = 1$, el predictor de paso 1, los valores de $-h[n]$ se denominan *coeficientes de predicción lineal*, y se emplean extensivamente en modelado del habla [7]. Además, para $l = 1$, el sistema de ecuaciones es idéntico a las ecuaciones de Yule-Walker empleadas para obtener los parámetros AR de un proceso AR(N) (ver la ecuación 7.26 del ejemplo de la sección 7.5.2).

El MSE mínimo para el predictor de paso l , de la ecuación 12.46, es

$$M_{\hat{x}} = r_{xx}[0] - \mathbf{r}_{xx}'^T \mathbf{R}_{xx}^{-1} \mathbf{r}_{xx}',$$

o equivalentemente

$$\begin{aligned}M_{\hat{x}} &= r_{xx}[0] - \mathbf{r}_{xx}'^T \mathbf{a} \\ &= r_{xx}[0] - \sum_{k=0}^{N-1} a_k r_{xx}[N-1+l-k] \\ &= r_{xx}[0] - \sum_{k=0}^{N-1} h[N-k] r_{xx}[N-1+l-k] \\ &\stackrel{(a)}{=} r_{xx}[0] - \sum_{m=1}^N h[m] r_{xx}[m+(l-1)],\end{aligned} \quad (12.54)$$

donde en (a) se realizó el cambio de variable $m = N - k$.

Como ejemplo, supóngase que $x[n]$ es un proceso AR(1) con ACF

$$r_{xx}[k] = \frac{\sigma_u^2}{1 - a^2[1]} (-a[1])^{|k|}.$$

y se quiere calcular el predictor $\hat{x}[N]$ de un paso como (ecuación 12.52 con $l = 1$)

$$\hat{x}[N] = \sum_{k=1}^N h[k]x[N - k].$$

Para encontrar los coeficientes $h[k]$ del filtro, hay que resolver el sistema dado por la ecuación 12.53 con $l = 1$, que es

$$\sum_{k=1}^N h[k]r_{xx}[m - k] = r_{xx}[m], \quad m = 1, \dots, N,$$

y sustituyendo la ACF queda

$$\sum_{k=1}^N h[k](-a[1])^{|m-k|} = (-a[1])^{|m|}, \quad m = 1, \dots, N.$$

Expresado explícitamente, el sistema es

$$\begin{cases} h[1](-a[1])^0 + h[2](-a[1])^1 + \dots + h[N](-a[1])^{N-1} = (-a[1])^1 \\ h[1](-a[1])^1 + h[2](-a[1])^0 + \dots + h[N](-a[1])^{N-2} = (-a[1])^2 \\ \vdots \\ h[1](-a[1])^{N-1} + h[2](-a[1])^{N-2} + \dots + h[N](-a[1])^0 = (-a[1])^N \end{cases},$$

y es fácil verificar que la solución es

$$h[k] = \begin{cases} -a[1], & k = 1 \\ 0, & k = 2, \dots, N. \end{cases}$$

De esta forma, el predictor de un paso es

$$\hat{x}[N] = -a[1]x[N - 1]$$

y depende únicamente de la muestra previa. El resultado es esperable teniendo en cuenta que un proceso AR(1) es

$$x[n] = -a[1]x[n - 1] + u[n]$$

donde $u[n]$ es ruido blanco, por lo que la muestra a predecir es

$$x[N] = -a[1]x[N - 1] + u[N].$$

El predictor no puede predecir la muestra $u[N]$ ya que no está correlacionada con las muestras pasadas de los datos. El error de predicción es $x[N] - \hat{x}[N] = u[N]$, y el MSE mínimo es la potencia σ_u^2 del ruido. Para verificarlo, de la ecuación 12.54 se tiene que

$$\begin{aligned} M_{\hat{x}} &= r_{xx}[0] - \sum_{k=1}^N h[k]r_{xx}[k] \\ &= r_{xx}[0] + a[1]r_{xx}[1] \\ &= \frac{\sigma_u^2}{1 - a^2[1]} + a[1]\frac{\sigma_u^2}{1 - a^2[1]}(-a[1]) \\ &= \frac{\sigma_u^2}{1 - a^2[1]}(1 - a^2[1]) \\ &= \sigma_u^2. \end{aligned}$$

Estos resultados pueden generalizarse para el predictor de paso l resolviendo (ecuación 12.53)

$$\sum_{k=1}^N h[k] r_{xx}[m-k] = r_{xx}[m+l-1], \quad m = 1, \dots, N.$$

Sustituyendo la ACF del proceso AR(1) se tiene que

$$\sum_{k=1}^N h[k] (-a[1])^{|m-k|} = (-a[1])^{|m+l-1|}, \quad m = 1, \dots, N.$$

Como antes, es fácil verificar que la solución es

$$h[k] = \begin{cases} (-a[1])^l, & k = 1 \\ 0, & k = 2, \dots, N. \end{cases}$$

por lo que el predictor de paso l es

$$\hat{x}[(N-1)+l] = (-a[1])^l x[N-1],$$

y de la ecuación 12.54, el MSE mínimo es

$$\begin{aligned} M_{\hat{x}} &= r_{xx}[0] - h[1] r_{xx}[l] \\ &= \frac{\sigma_u^2}{1-a^2[1]} - (-a[1])^l \frac{\sigma_u^2}{1-a^2[1]} (-a[1])^l \\ &= \frac{\sigma_u^2}{1-a^2[1]} (1-a[1]^{2l}). \end{aligned}$$

Es interesante notar que como $|a[1]| < 1$, el predictor decae a 0 con l . Esto es razonable, ya que la correlación entre la muestra $x[(N-1)+l]$ a predecir y la muestra $x[N-1]$ empleada para la predicción es $r_{xx}[l]$. Cuando l crece, $r_{xx}[l]$ decrece a 0, y también lo hace $\hat{x}[(N-1)+l]$. Esto también se refleja en el MSE mínimo, que toma el valor mas pequeño cuando $l = 1$ y crece con l . Ver además los problemas de las secciones 12.6.19 y 12.6.20.

12.6. Problemas

12.6.1. Problema 1

Se considera el estimador cuadrático

$$\hat{\theta} = ax^2[0] + bx[0] + c$$

del parámetro escalar θ basado en una única muestra $x[0]$. Encontrar los coeficientes a , b y c que minimizan el MSE bayesiano. Si $x[0] \sim \mathcal{U}[-\frac{1}{2}, \frac{1}{2}]$, encontrar el estimador LMMSE y el estimador MMSE cuadrático si $\theta = \cos 2\pi x[0]$. Además, comparar los MSE mínimos.

Solución El MSE bayesiano está definido por la ecuación 12.2 y en este caso es

$$\text{Bsme}(\hat{\theta}) = E [(\theta - ax^2[0] - bx[0] - c)^2].$$

Para encontrar los coeficientes que lo minimizan, se deriva respecto a los coeficientes y se iguala a cero. Diferenciando respecto a a e igualando a cero, se tiene que

$$\frac{\partial \text{Bsme}(\hat{\theta})}{\partial a} = E [-2(\theta - ax^2[0] - bx[0] - c)x^2[0]] = 0, \quad (12.55)$$

resultando en que

$$\hat{a}E(x^4[0]) + \hat{b}E(x^3[0]) + \hat{c}E(x^2[0]) = E(\theta x^2[0]).$$

Repitiendo para el coeficiente b , se tiene que

$$\frac{\partial \text{Bsme}(\hat{\theta})}{\partial b} = E[-2(\theta - ax^2[0] - bx[0] - c)x[0]] = 0, \quad (12.56)$$

que resulta en la condición

$$\hat{a}E(x^3[0]) + \hat{b}E(x^2[0]) + \hat{c}E(x[0]) = E(\theta x[0]),$$

y haciendo lo mismo para c

$$\frac{\partial \text{Bsme}(\hat{\theta})}{\partial c} = E[-2(\theta - ax^2[0] - bx[0] - c)] = 0, \quad (12.57)$$

obteniendo que

$$\hat{a}E(x^2[0]) + \hat{b}E(x[0]) + \hat{c} = E(\theta).$$

Se concluye que los coeficientes que minimizan el MSE bayesiano satisfacen el sistema de ecuaciones

$$\begin{bmatrix} E(x^4[0]) & E(x^3[0]) & E(x^2[0]) \\ E(x^3[0]) & E(x^2[0]) & E(x[0]) \\ E(x^2[0]) & E(x[0]) & 1 \end{bmatrix} \begin{bmatrix} \hat{a} \\ \hat{b} \\ \hat{c} \end{bmatrix} = \begin{bmatrix} E(\theta x^2[0]) \\ E(\theta x[0]) \\ E(\theta) \end{bmatrix}. \quad (12.58)$$

Considerando ahora que $x[0] \sim \mathcal{U}[-\frac{1}{2}, \frac{1}{2}]$ y $\theta = \cos 2\pi x[0]$, se calculará el estimador MMSE cuadrático. Como $x[0]$ tiene PDF simétrica, los momentos impares son nulos, es decir,

$$E(x[0]) = E(x^3[0]) = 0.$$

El momento segundo es

$$E(x^2[0]) = \int_{-\frac{1}{2}}^{\frac{1}{2}} x^2 dx = \frac{x^3}{3} \Big|_{-\frac{1}{2}}^{\frac{1}{2}} = \frac{1}{3} \left(\frac{1}{8} + \frac{1}{8} \right) = \frac{1}{3} \times \frac{1}{4} = \frac{1}{12},$$

y el momento cuarto es

$$E(x^4[0]) = \int_{-\frac{1}{2}}^{\frac{1}{2}} x^4 dx = \frac{x^5}{5} \Big|_{-\frac{1}{2}}^{\frac{1}{2}} = \frac{1}{5} \left(\frac{1}{32} + \frac{1}{32} \right) = \frac{1}{5} \times \frac{1}{16} = \frac{1}{80}.$$

Por otro lado, con $\theta = \cos 2\pi x[0] = g(x[0])$ se tiene que

$$E(\theta) = \int_{-\infty}^{\infty} \theta p_{x[0]}(x) dx = \int_{-\frac{1}{2}}^{\frac{1}{2}} \cos 2\pi x dx = 0,$$

donde en la última igualdad se consideró que se está integrando una función sinusoidal en un período. Además,

$$E(\theta x[0]) = \int_{-\frac{1}{2}}^{\frac{1}{2}} x \cos 2\pi x dx = 0,$$

donde en la última igualdad se tuvo en cuenta que el integrando es una función impar. Finalmente,

$$\begin{aligned} E(\theta x^2[0]) &= \int_{-\frac{1}{2}}^{\frac{1}{2}} x^2 \cos 2\pi x dx \\ &\stackrel{(a)}{=} \frac{1}{2\pi} x \sin 2\pi x \Big|_{-\frac{1}{2}}^{\frac{1}{2}} - \frac{1}{\pi} \int_{-\frac{1}{2}}^{\frac{1}{2}} x \sin 2\pi x dx \\ &\stackrel{(b)}{=} \frac{1}{2\pi^2} x \cos 2\pi x \Big|_{-\frac{1}{2}}^{\frac{1}{2}} - \frac{1}{2\pi^2} \int_{-\frac{1}{2}}^{\frac{1}{2}} \cos 2\pi x dx \\ &= \frac{1}{2\pi^2} \left[\frac{1}{2} \cos \pi + \frac{1}{2} \cos(-\pi) \right] \\ &= \frac{1}{2\pi^2} \left[-\frac{1}{2} - \frac{1}{2} \right] \\ &= -\frac{1}{2\pi^2}, \end{aligned}$$

donde en (a) se aplicó integración por partes $\int u dv = uv - \int v du$ con $u = x^2$ y $dv = \cos 2\pi x dx$ y por lo tanto, $du = 2x dx$ y $v = \frac{1}{2\pi} \sin 2\pi x$ y en (b) se aplicó nuevamente integración por partes con $u = x$ y $dv = -\sin 2\pi x dx$ y por lo tanto, $du = dx$ y $v = \frac{1}{2\pi} \cos 2\pi x$. Sustituyendo estos resultados en 12.58, se obtiene que los coeficientes del estimador MMSE cuadrático cumplen que

$$\begin{bmatrix} \frac{1}{80} & 0 & \frac{1}{12} \\ 0 & \frac{1}{12} & 0 \\ \frac{1}{12} & 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{a} \\ \hat{b} \\ \hat{c} \end{bmatrix} = \begin{bmatrix} -\frac{1}{2\pi^2} \\ 0 \\ 0 \end{bmatrix}.$$

De la segunda ecuación se obtiene inmediatamente que $\hat{b} = 0$. Continuando con la resolución del sistema,

$$\left\{ \begin{array}{l} \frac{1}{80}\hat{a} + \frac{1}{12}\hat{c} = -\frac{1}{2\pi^2} \\ \frac{1}{12}\hat{a} + \hat{c} = 0 \end{array} \right\} \Rightarrow \hat{c} = -\frac{1}{12}\hat{a} \Rightarrow \frac{1}{40}\hat{a} - \frac{1}{6 \times 12}\hat{a} = -\frac{1}{\pi^2} \Rightarrow \hat{a} = -\frac{90}{\pi^2}$$

y finalmente calculando c se obtiene que

$$\hat{a} = -\frac{90}{\pi^2}, \quad \hat{b} = 0, \quad \hat{c} = \frac{15}{2\pi^2}.$$

Se calculará el MSE bayesiano mínimo. Para eso, se tiene que cuenta que

$$\text{Bmse}(\hat{\theta}) = E[(\theta - \hat{\theta})^2] = E[(\theta - \hat{\theta})(\theta - \hat{\theta})] = E[\theta(\theta - \hat{\theta})] - E[\hat{\theta}(\theta - \hat{\theta})].$$

Pero

$$\begin{aligned} E[\hat{\theta}(\theta - \hat{\theta})] &= E[(\hat{a}x^2[0] + \hat{b}x[0] + \hat{c})(\theta - \hat{a}x^2[0] - \hat{b}x[0] - \hat{c})] \\ &= \hat{a}E[x^2[0](\theta - \hat{a}x^2[0] - \hat{b}x[0] - \hat{c})] + \hat{b}E[x[0](\theta - \hat{a}x^2[0] - \hat{b}x[0] - \hat{c})] \\ &\quad + \hat{c}E[(\theta - \hat{a}x^2[0] - \hat{b}x[0] - \hat{c})] \\ &= 0, \end{aligned}$$

ya que las esperanzas son nulas por la condición de minimización dada por las ecuaciones 12.55, 12.56 y 12.57. Por lo tanto, el MSE bayesiano mínimo queda

$$\begin{aligned} \text{Bmse}(\hat{\theta}) &= E[\theta(\theta - \hat{\theta})] \\ &= E[\theta(\theta - \hat{a}x^2[0] - \hat{b}x[0] - \hat{c})] \\ &= E(\theta^2) - \hat{a}E(\theta x^2[0]) - \hat{b}E(\theta x[0]) - \hat{c}E(\theta). \end{aligned}$$

Como

$$\begin{aligned} E(\theta^2) &= \int_{-\frac{1}{2}}^{\frac{1}{2}} \cos^2 2\pi x dx \\ &\stackrel{(a)}{=} \int_{-\frac{1}{2}}^{\frac{1}{2}} \left(\frac{1}{2} + \frac{1}{2} \cos 4\pi x \right) dx \\ &= \frac{1}{2}x \Big|_{-\frac{1}{2}}^{\frac{1}{2}} + \frac{1}{8\pi} \sin 4\pi x \Big|_{-\frac{1}{2}}^{\frac{1}{2}} \\ &= \frac{1}{2} \left[\frac{1}{2} - \left(-\frac{1}{2} \right) \right] \\ &= \frac{1}{2} \end{aligned}$$

donde en (a) se empleó la identidad trigonométrica $\cos 2\theta = 2\cos^2 \theta - 1$, y teniendo en cuenta los resultados anteriores, resulta en

$$\text{Bmse}(\hat{\theta}) = \frac{1}{2} - \frac{90}{\pi^2} \times \frac{1}{2\pi^2} = \frac{1}{2} - \frac{45}{\pi^4} \approx 0.038.$$

El estimador LMMSE es

$$\hat{\theta} = bx[0] + c,$$

es decir, el caso anterior con $a = 0$. Del sistema de la ecuación 12.58 con $a = 0$, la condición de minimización del MSE bayesiano es

$$\begin{bmatrix} E(x^2[0]) & E(x[0]) \\ E(x[0]) & 1 \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{c} \end{bmatrix} = \begin{bmatrix} E(\theta x[0]) \\ E(\theta) \end{bmatrix}.$$

y con los valores de las esperanzas calculados previamente, el sistema queda

$$\begin{bmatrix} \frac{1}{12} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{c} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

y por lo tanto, los coeficientes y el estimador son

$$\hat{b} = \hat{c} = 0 \quad \Rightarrow \quad \hat{\theta} = 0.$$

El error bayesiano mínimo es

$$\text{Bsme}(\hat{\theta}) = E[(\theta - \hat{\theta})^2] = E(\theta^2) = \frac{1}{2}.$$

12.6.2. Problema 2

Se consideran los datos

$$x[n] = Ar^n + w[n], \quad n = 0, \dots, N-1,$$

donde A es el parámetro a estimar, r es una constante conocida y $w[n]$ es ruido blanco de media nula y varianza σ^2 . El parámetro A se modela como una variable aleatoria de media μ_A y varianza σ_A^2 , y es independiente de $w[n]$. Encontrar el estimador LMMSE de A y el MSE bayesiano mínimo.

Solución Los datos cumplen con el modelo lineal bayesiano

$$\mathbf{x} = \mathbf{h}A + \mathbf{w}, \quad \text{con} \quad \mathbf{h} = [1 \ r \ \dots \ r^{N-1}]^T \quad \text{y} \quad \mathbf{C}_w = \sigma^2 \mathbf{I},$$

por lo que el estimador LMMSE está dado por la ecuación 12.19 y el MSE bayesiano mínimo por la ecuación 12.21. El MSE bayesiano mínimo es

$$\begin{aligned} \text{Bmse}(\hat{A}) &= \frac{1}{\frac{1}{\sigma_A^2} + \mathbf{h}^T \mathbf{C}_w^{-1} \mathbf{h}} \\ &= \frac{1}{\frac{1}{\sigma_A^2} + \frac{1}{\sigma^2} \sum_{n=0}^{N-1} r^{2n}}. \end{aligned}$$

El estimador LMMSE es

$$\begin{aligned} \hat{A} &= \mu_A + \frac{\mathbf{h}^T \mathbf{C}_w^{-1} (\mathbf{x} - \mathbf{h}\mu_A)}{\frac{1}{\sigma_A^2} + \mathbf{h}^T \mathbf{C}_w^{-1} \mathbf{h}} \\ &= \mu_A + \frac{\frac{1}{\sigma^2} \sum_{n=0}^{N-1} r^n (x[n] - \mu_A r^n)}{\frac{1}{\sigma_A^2} + \frac{1}{\sigma^2} \sum_{n=0}^{N-1} r^{2n}}. \end{aligned}$$

12.6.3. Problema 3

Un vector gaussiano $\mathbf{x} = [x_1 \ x_2]^T$ tiene media nula y matriz de covarianza \mathbf{C}_{xx} . Si x_2 se estima linealmente en base a x_1 , encontrar el estimador que minimiza el MSE bayesiano. Además, encontrar el MSE mínimo y probar que es nulo si y solo si \mathbf{C}_{xx} es singular. Extender los resultados para mostrar que si la matriz de covarianza de un vector aleatorio gaussiano $N \times 1$ de media nula no es definida positiva, cualquier variable aleatoria puede ser perfectamente estimada mediante una combinación lineal de las otras. Sugerencia: notar que

$$E \left[\left(\sum_{i=1}^N a_i x_i \right)^2 \right] = \mathbf{a}^T \mathbf{C}_{xx} \mathbf{a}.$$

Solución Si se estima x_2 como

$$\hat{x}_2 = a_1 x_1,$$

el MSE bayesiano es

$$\begin{aligned} \text{Bmse}(\hat{x}_2) &= E[(x_2 - \hat{x}_2)^2] \\ &= E[(x_2 - a_1 x_1)^2] \\ &= E(x_2^2) - 2a_1 E(x_1 x_2) + a_1^2 E(x_1^2). \end{aligned}$$

Diferenciando respecto al coeficiente e igualando a cero, se tiene que el coeficiente que minimiza el MSE cumple que

$$\frac{\partial \text{Bmse}(\hat{x}_2)}{\partial a_1} = -2E(x_1 x_2) + 2a_1 E(x_1^2) = 0,$$

resultando en

$$a_1 = \frac{E(x_1 x_2)}{E(x_1^2)},$$

y el estimador es

$$\hat{x}_2 = \frac{E(x_1 x_2)}{E(x_1^2)} x_1.$$

Sustituyendo el resultado en el MSE bayesiano se obtiene que el valor mínimo es

$$\begin{aligned} \text{Bmse}(\hat{x}_2) &= E(x_2^2) - 2 \frac{E(x_1 x_2)}{E(x_1^2)} E(x_1 x_2) + \frac{E^2(x_1 x_2)}{E^2(x_1^2)} E(x_1^2) \\ &= E(x_2^2) - \frac{E^2(x_1 x_2)}{E(x_1^2)} \end{aligned}$$

Estos resultados podrían haberse obtenido directamente de las ecuaciones 12.6 y 12.8. Se observa que ocurre que

$$\text{Bmse}(\hat{x}_2) = E(x_2^2) - \frac{E^2(x_1 x_2)}{E(x_1^2)} = 0$$

si

$$E(x_1^2)E(x_2^2) = E^2(x_1 x_2).$$

Teniendo en cuenta que

$$\mathbf{C}_{xx} = E(\mathbf{x}\mathbf{x}^T) = \begin{bmatrix} E(x_1^2) & E(x_1 x_2) \\ E(x_2 x_1) & E(x_2^2) \end{bmatrix} \Rightarrow \det(\mathbf{C}_{xx}) = E(x_1^2)E(x_2^2) - E^2(x_1 x_2),$$

se cumple que $\text{Bmse}(\hat{x}_2) = 0$ si $\det(\mathbf{C}_{xx}) = 0$, o si \mathbf{C}_{xx} es singular. En ese caso, x_2 se estima de forma perfecta en base a x_1 . Efectivamente, teniendo en cuenta que el coeficiente de correlación es

$$\rho^2 = \frac{E^2(x_1 x_2)}{E(x_1^2)E(x_2^2)},$$

la condición implica que $|\rho| = 1$, indicando que x_1 y x_2 están perfectamente correlacionadas.

En el caso general de un vector \mathbf{x} $N \times 1$ de variables aleatorias

$$\mathbf{x} = [x_1, \dots, x_N]^T$$

con matriz de covarianza \mathbf{C}_{xx} , el estimador de la variable aleatoria x_i en base al resto es

$$\hat{x}_i = \sum_{\substack{j=1 \\ j \neq i}}^N a_j x_j,$$

y el MSE bayesiano es

$$\begin{aligned} \text{Bmse}(\hat{x}_i) &= E \left[\left(x_i - \sum_{\substack{j=1 \\ j \neq i}}^N a_j x_j \right)^2 \right] \\ &\stackrel{(a)}{=} E \left[\left(\sum_{j=1}^N b_j x_j \right)^2 \right] \\ &\stackrel{(b)}{=} \mathbf{b}^T \mathbf{C}_{xx} \mathbf{b}, \end{aligned}$$

donde en (a) se definió b_j como

$$b_j = \begin{cases} -a_j & j \neq i \\ 1 & j = i \end{cases}$$

y en (b) se tuvo en cuenta que

$$\mathbf{b}^T \mathbf{C}_{xx} \mathbf{b} = \mathbf{b}^T E(\mathbf{x} \mathbf{x}^T) \mathbf{b} = E(\mathbf{b}^T \mathbf{x} \mathbf{x}^T \mathbf{b}) = E[(\mathbf{b}^T \mathbf{x})(\mathbf{b}^T \mathbf{x})^T] = E[(\mathbf{b}^T \mathbf{x})^2] = E \left[\left(\sum_{j=1}^N b_j x_j \right)^2 \right].$$

Si \mathbf{C}_{xx} no es definida positiva, existe $\mathbf{b} \neq \mathbf{0}$ que cumple que

$$\text{Bmse}(\hat{x}_i) = \mathbf{b}^T \mathbf{C}_{xx} \mathbf{b} = 0.$$

En ese caso, cualquier variable aleatoria puede ser estimada de forma perfecta a partir del resto eligiendo $a_j = -b_j$ para $j \neq i$ en el estimador.

12.6.4. Problema 4

Un producto interno (x, y) entre dos vectores x e y de un espacio vectorial debe satisfacer las siguientes propiedades:

- $(x, x) \geq 0$, y $(x, x) = 0$ si y solo si $x = 0$.
- $(x, y) = (y, x)$.
- $(c_1 x_1 + c_2 x_2, y) = c_1(x_1, y) + c_2(x_2, y)$.

Probar que la definición $(x, y) = E(xy)$ para las variables aleatorias x e y de media nula satisface dichas propiedades.

Solución

- $(x, x) = E(x^2) \geq 0$. Además, $(x, x) = E(x^2) = 0$ si y solo si $x = 0$ constante en lugar de una variable aleatoria.

b. $(x, y) = (y, x)$, ya que $E(xy) = E(yx)$ porque

$$E(xy) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyp(x, y) dx dy.$$

c. $(c_1x_1 + c_2x_2, y) = c_1(x_1, y) + c_2(x_2, y)$ también es cierto por la propiedad de linealidad de la esperanza, es decir,

$$(c_1x_1 + c_2x_2, y) = E[(c_1x_1 + c_2x_2)y] = c_1E(x_1y) + c_2E(x_2y) = c_1(x_1, y) + c_2(x_2, y).$$

12.6.5. Problema 5

Si se consideran variables aleatorias de media no nula, un enfoque razonable para definir el producto interno entre x e y es como $(x, y) = \text{cov}(x, y)$. Con esta definición x e y son ortogonales si y solo si son no correlacionadas. Para este producto interno, ¿cuales de las propiedades dadas en el problema de la sección 12.6.5 no se cumplen?

Solución

- a. $(x, x) = \text{cov}(x, x) = \sigma^2 \geq 0$, y $(x, x) = \sigma^2 = 0$ si x es cualquier constante en lugar de una variable aleatoria. Por lo tanto, no se cumple que $(x, x) = 0$ si y solo si $x = 0$.
- b. $(x, y) = (y, x)$ se cumple ya que $\text{cov}(x, y) = \text{cov}(y, x)$.
- c. $(c_1x_1 + c_2x_2, y) = c_1(x_1, y) + c_2(x_2, y)$ se cumple ya que

$$\begin{aligned} (c_1x_1 + c_2x_2, y) &= \text{cov}(c_1x_1 + c_2x_2, y) \\ &= E[(c_1x_1 + c_2x_2)y] - E(c_1x_1 + c_2x_2)E(y) \\ &= c_1E(x_1y) + c_2E(x_2y) - c_1E(x_1)E(y) - c_2E(x_2)E(y) \\ &= c_1[E(x_1y) - E(x_1)E(y)] + c_2[E(x_2y) - E(x_2)E(y)] \\ &= c_1\text{cov}(x_1, y) + c_2\text{cov}(x_2, y) \\ &= c_1(x_1, y) + c_2(x_2, y). \end{aligned}$$

12.6.6. Problema 6

Se observan los datos $x[n] = s[n] + w[n]$ para $n = 0, \dots, N-1$, donde $s[n]$ y $w[n]$ son procesos WSS de media nula no correlacionados entre si. Las ACF son

$$\begin{aligned} r_{ss}[k] &= \sigma_s^2 \delta[k] \\ r_{ww}[k] &= \sigma^2 \delta[k]. \end{aligned}$$

Determinar el estimador LMMSE de $\mathbf{s} = [s[0] \dots s[N-1]]^T$ en base a $\mathbf{x} = [x[0] \dots x[N-1]]^T$ y la correspondiente matriz de MSE mínimo.

Solución Considerando que los procesos tienen media nula, el estimador está dado por la ecuación 12.45 y es

$$\hat{\mathbf{s}} = \mathbf{C}_{sx} \mathbf{C}_{xx}^{-1} \mathbf{x}.$$

Luego, como

$$\mathbf{C}_{sx} = E(\mathbf{s}\mathbf{x}^T) = E[\mathbf{s}(\mathbf{s} + \mathbf{w})^T] = E(\mathbf{s}\mathbf{s}^T) = \sigma_s^2 \mathbf{I}$$

y

$$\mathbf{C}_{xx} = E(\mathbf{x}\mathbf{x}^T) = E[(\mathbf{s} + \mathbf{w})(\mathbf{s} + \mathbf{w})^T] = E(\mathbf{s}\mathbf{s}^T) + E(\mathbf{w}\mathbf{w}^T) = (\sigma_s^2 + \sigma^2) \mathbf{I},$$

el estimador es

$$\hat{\mathbf{s}} = \sigma_s^2 \mathbf{I} [(\sigma_s^2 + \sigma^2) \mathbf{I}]^{-1} \mathbf{x} = \frac{\sigma_s^2}{\sigma_s^2 + \sigma^2} \mathbf{x},$$

que puede expresarse como

$$\hat{\mathbf{s}} = \frac{\eta}{\eta + 1} \mathbf{x},$$

donde $\eta = \sigma_s^2/\sigma^2$ es la SNR. La matriz de MSE mínimo está dada por la ecuación 12.14 y es

$$\begin{aligned} \mathbf{M}_{\hat{\mathbf{s}}} &= \mathbf{C}_{ss} - \mathbf{C}_{sx} \mathbf{C}_{xx}^{-1} \mathbf{C}_{x\theta} \\ &= \sigma_s^2 \mathbf{I} - \frac{\sigma_s^4}{\sigma_s^2 + \sigma^2} \mathbf{I} \\ &= \frac{\sigma_s^4 + \sigma_s^2 \sigma^2 - \sigma_s^4}{\sigma_s^2 + \sigma^2} \mathbf{I}, \end{aligned}$$

resultando en

$$\mathbf{M}_{\hat{\mathbf{s}}} = \frac{\sigma_s^2 \sigma^2}{\sigma_s^2 + \sigma^2} \mathbf{I}.$$

12.6.7. Problema 7

Deducir la matriz MSE bayesiana para el estimador LMMSE vectorial dada por la ecuación 12.14 así como el MSE bayesiano mínimo de la ecuación 12.15. Notar el promediado de las PDF implicadas debido a la operador esperanza.

Solución Por definición, la matriz MSE bayesiana es

$$\mathbf{M}_{\hat{\boldsymbol{\theta}}} = E \left[(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \right].$$

Como en el caso vectorial el estimador LMMSE es (ecuación 12.13)

$$\hat{\boldsymbol{\theta}} = E(\boldsymbol{\theta}) + \mathbf{C}_{\theta x} \mathbf{C}_{xx}^{-1} (\mathbf{x} - E(\mathbf{x})),$$

por lo que sustituyendo se tiene que

$$\begin{aligned} \mathbf{M}_{\hat{\boldsymbol{\theta}}} &= E \left\{ [\boldsymbol{\theta} - E(\boldsymbol{\theta}) - \mathbf{C}_{\theta x} \mathbf{C}_{xx}^{-1} (\mathbf{x} - E(\mathbf{x}))] [\boldsymbol{\theta} - E(\boldsymbol{\theta}) - \mathbf{C}_{\theta x} \mathbf{C}_{xx}^{-1} (\mathbf{x} - E(\mathbf{x}))]^T \right\} \\ &= E \left[(\boldsymbol{\theta} - E(\boldsymbol{\theta})) (\boldsymbol{\theta} - E(\boldsymbol{\theta}))^T \right] - E \left\{ (\boldsymbol{\theta} - E(\boldsymbol{\theta})) [\mathbf{C}_{\theta x} \mathbf{C}_{xx}^{-1} (\mathbf{x} - E(\mathbf{x}))]^T \right\} \\ &\quad - E \left\{ [\mathbf{C}_{\theta x} \mathbf{C}_{xx}^{-1} (\mathbf{x} - E(\mathbf{x}))] (\boldsymbol{\theta} - E(\boldsymbol{\theta}))^T \right\} + E \left\{ [\mathbf{C}_{\theta x} \mathbf{C}_{xx}^{-1} (\mathbf{x} - E(\mathbf{x}))] [\mathbf{C}_{\theta x} \mathbf{C}_{xx}^{-1} (\mathbf{x} - E(\mathbf{x}))]^T \right\} \\ &= E \left[(\boldsymbol{\theta} - E(\boldsymbol{\theta})) (\boldsymbol{\theta} - E(\boldsymbol{\theta}))^T \right] - E \left[(\boldsymbol{\theta} - E(\boldsymbol{\theta})) (\mathbf{x} - E(\mathbf{x}))^T \right] \mathbf{C}_{xx}^{-1} \mathbf{C}_{\theta x}^T \\ &\quad - \mathbf{C}_{\theta x} \mathbf{C}_{xx}^{-1} E \left[(\mathbf{x} - E(\mathbf{x})) (\boldsymbol{\theta} - E(\boldsymbol{\theta}))^T \right] + \mathbf{C}_{\theta x} \mathbf{C}_{xx}^{-1} E \left[(\mathbf{x} - E(\mathbf{x})) (\mathbf{x} - E(\mathbf{x}))^T \right] \mathbf{C}_{xx}^{-1} \mathbf{C}_{\theta x}^T \\ &= \mathbf{C}_{\theta\theta} - \mathbf{C}_{\theta x} \mathbf{C}_{xx}^{-1} \mathbf{C}_{x\theta} - \mathbf{C}_{\theta x} \mathbf{C}_{xx}^{-1} \mathbf{C}_{x\theta} + \mathbf{C}_{\theta x} \mathbf{C}_{xx}^{-1} \mathbf{C}_{xx} \mathbf{C}_{xx}^{-1} \mathbf{C}_{\theta x}^T \end{aligned}$$

resultando en

$$\mathbf{M}_{\hat{\boldsymbol{\theta}}} = \mathbf{C}_{\theta\theta} - \mathbf{C}_{\theta x} \mathbf{C}_{xx}^{-1} \mathbf{C}_{x\theta}.$$

Partiendo de la definición, el MSE bayesiano mínimo es

$$\begin{aligned} \text{Bmse}(\hat{\theta}_i) &= E \left[(\theta_i - \hat{\theta}_i)^2 \right] \\ &= \int \int (\theta_i - \hat{\theta}_i)^2 p(\mathbf{x}, \theta_i) d\theta_i d\mathbf{x} \\ &= \int \int (\theta_i - \hat{\theta}_i)^2 \left[\int \cdots \int p(\mathbf{x}, \boldsymbol{\theta}) d\theta_1 \cdots d\theta_{i-1} d\theta_{i+1} \cdots d\theta_p \right] d\theta_i d\mathbf{x} \\ &\stackrel{(a)}{=} \int \cdots \int (\theta_i - \hat{\theta}_i)^2 p(\mathbf{x}, \boldsymbol{\theta}) d\boldsymbol{\theta} d\mathbf{x} \\ &= \int \cdots \int [(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T]_{ii} p(\mathbf{x}, \boldsymbol{\theta}) d\boldsymbol{\theta} d\mathbf{x} \\ &= E \left[(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \right]_{ii} \\ &= [\mathbf{M}_{\hat{\boldsymbol{\theta}}}]_{ii}, \end{aligned}$$

donde en (a) se tuvo en cuenta de que $\hat{\theta}_i$ depende únicamente de \mathbf{x} .

12.6.8. Problema 8

Probar la propiedad conmutativa y adictiva del estimador LMMSE dadas por las ecuaciones 12.16 y 12.17.

Solución La propiedad conmutativa indica que si

$$\boldsymbol{\alpha} = \mathbf{A}\boldsymbol{\theta} + \mathbf{b},$$

el estimador de $\boldsymbol{\alpha}$ es

$$\hat{\boldsymbol{\alpha}} = \mathbf{A}\hat{\boldsymbol{\theta}} + \mathbf{b},$$

donde $\hat{\boldsymbol{\theta}}$ está dado por la ecuación 12.13. Para probarlo, se parte observando que de la misma ecuación, el estimador de $\boldsymbol{\alpha}$ es

$$\hat{\boldsymbol{\alpha}} = E(\boldsymbol{\alpha}) + \mathbf{C}_{\alpha x} \mathbf{C}_{xx}^{-1} (\mathbf{x} - E(\mathbf{x})). \quad (12.59)$$

Además

$$E(\boldsymbol{\alpha}) = E(\mathbf{A}\boldsymbol{\theta} + \mathbf{b}) = \mathbf{A}E(\boldsymbol{\theta}) + \mathbf{b},$$

y

$$\mathbf{C}_{\alpha x} = E(\boldsymbol{\alpha} \mathbf{x}^T) = E[(\mathbf{A}\boldsymbol{\theta} + \mathbf{b}) \mathbf{x}^T] = \mathbf{A}E(\boldsymbol{\theta} \mathbf{x}^T) + \mathbf{b}E(\mathbf{x}^T) = \mathbf{A}\mathbf{C}_{\theta x},$$

donde en la última igualdad se consideró que por hipótesis, $E(\mathbf{x}^T) = 0$. Sustituyendo estos resultados en el estimador de $\boldsymbol{\alpha}$, se tiene que

$$\begin{aligned} \hat{\boldsymbol{\alpha}} &= \mathbf{A}E(\boldsymbol{\theta}) + \mathbf{b} + \mathbf{A}\mathbf{C}_{\theta x} \mathbf{C}_{xx}^{-1} (\mathbf{x} - E(\mathbf{x})) \\ &= \mathbf{A} [E(\boldsymbol{\theta}) + \mathbf{C}_{\theta x} \mathbf{C}_{xx}^{-1} (\mathbf{x} - E(\mathbf{x}))] + \mathbf{b} \\ &= \mathbf{A}\hat{\boldsymbol{\theta}} + \mathbf{b}. \end{aligned}$$

La propiedad aditiva indica que si

$$\boldsymbol{\alpha} = \boldsymbol{\theta}_1 + \boldsymbol{\theta}_2,$$

el estimador de $\boldsymbol{\alpha}$ es

$$\hat{\boldsymbol{\alpha}} = \hat{\boldsymbol{\theta}}_1 + \hat{\boldsymbol{\theta}}_2.$$

Nuevamente, el estimador de $\boldsymbol{\alpha}$ está dado por la ecuación 12.59 con

$$E(\boldsymbol{\alpha}) = E(\boldsymbol{\theta}_1) + E(\boldsymbol{\theta}_2)$$

y

$$\mathbf{C}_{\alpha x} = E(\boldsymbol{\alpha} \mathbf{x}^T) = E[(\boldsymbol{\theta}_1 + \boldsymbol{\theta}_2) \mathbf{x}^T] = E(\boldsymbol{\theta}_1 \mathbf{x}^T) + E(\boldsymbol{\theta}_2 \mathbf{x}^T) = \mathbf{C}_{\theta_1 x} + \mathbf{C}_{\theta_2 x},$$

y por lo tanto,

$$\begin{aligned} \hat{\boldsymbol{\alpha}} &= E(\boldsymbol{\theta}_1) + E(\boldsymbol{\theta}_2) + (\mathbf{C}_{\theta_1 x} + \mathbf{C}_{\theta_2 x}) \mathbf{C}_{xx}^{-1} (\mathbf{x} - E(\mathbf{x})) \\ &= E(\boldsymbol{\theta}_1) + \mathbf{C}_{\theta_1 x} \mathbf{C}_{xx}^{-1} (\mathbf{x} - E(\mathbf{x})) + E(\boldsymbol{\theta}_2) + \mathbf{C}_{\theta_2 x} \mathbf{C}_{xx}^{-1} (\mathbf{x} - E(\mathbf{x})) \\ &= \hat{\boldsymbol{\theta}}_1 + \hat{\boldsymbol{\theta}}_2. \end{aligned}$$

12.6.9. Problema 9

Derivar el estimador LMMSE secuencial análogo a las ecuaciones 12.25-12.27 pero para el caso donde $\mu_A \neq 0$. Sugerencia: emplear un enfoque algebraico basado en los resultados del ejemplo de la sección 10.2.1.

Solución En el ejemplo de la sección 10.2.1 se encontró que el estimador MMSE del nivel de DC en WGN con PDF a priori gaussiana está dado por la ecuación 10.7. Como el estimador es lineal con los datos, el estimador LMMSE debe coincidir. Por lo tanto, el estimador LMMSE secuencial en base a las muestras $\{x[0], \dots, x[N-1]\}$ es

$$\hat{A}[N-1] = \frac{\sigma_A^2}{\sigma_A^2 + \frac{\sigma^2}{N}} \bar{x} + \frac{\frac{\sigma^2}{N}}{\sigma_A^2 + \frac{\sigma^2}{N}} \mu_A.$$

Al observar la nueva muestra $x[N]$ es estimador es

$$\begin{aligned} \hat{A}[N] &= \frac{\sigma_A^2}{\sigma_A^2 + \frac{\sigma^2}{N+1}} \frac{1}{N+1} \sum_{n=0}^N x[n] + \frac{\frac{\sigma^2}{N+1}}{\sigma_A^2 + \frac{\sigma^2}{N+1}} \mu_A \\ &= \frac{N\sigma_A^2}{(N+1)\sigma_A^2 + \sigma^2} \frac{1}{N} \left(\sum_{n=0}^{N-1} x[n] + x[N] \right) + \frac{\sigma^2}{(N+1)\sigma_A^2 + \sigma^2} \mu_A \\ &= \frac{N\sigma_A^2}{(N+1)\sigma_A^2 + \sigma^2} \left(\bar{x} + \frac{x[N]}{N} \right) + \frac{\sigma^2}{(N+1)\sigma_A^2 + \sigma^2} \mu_A, \end{aligned}$$

donde \bar{x} es la media muestral de $\{x[0], \dots, x[N-1]\}$, al igual que en $\hat{A}[N-1]$. Continuando,

$$\begin{aligned} \hat{A}[N] &= \frac{N}{(N+1)\sigma_A^2 + \sigma^2} \left(\sigma_A^2 \bar{x} + \frac{\sigma^2}{N} \mu_A \right) + \frac{\sigma_A^2}{(N+1)\sigma_A^2 + \sigma^2} x[N] \\ &\stackrel{(a)}{=} \frac{N}{(N+1)\sigma_A^2 + \sigma^2} \left(\sigma_A^2 + \frac{\sigma^2}{N} \right) \hat{A}[N-1] + \frac{\sigma_A^2}{(N+1)\sigma_A^2 + \sigma^2} x[N] \\ &= \frac{N\sigma_A^2 + \sigma^2}{(N+1)\sigma_A^2 + \sigma^2} \hat{A}[N-1] + \frac{\sigma_A^2}{(N+1)\sigma_A^2 + \sigma^2} x[N] \\ &\stackrel{(b)}{=} \hat{A}[N-1] + \left[\frac{N\sigma_A^2 + \sigma^2}{(N+1)\sigma_A^2 + \sigma^2} - 1 \right] \hat{A}[N-1] + \frac{\sigma_A^2}{(N+1)\sigma_A^2 + \sigma^2} x[N] \\ &= \hat{A}[N-1] - \frac{\sigma_A^2}{(N+1)\sigma_A^2 + \sigma^2} \hat{A}[N-1] + \frac{\sigma_A^2}{(N+1)\sigma_A^2 + \sigma^2} x[N], \end{aligned}$$

resultando en

$$\hat{A}[N] = \hat{A}[N-1] + \frac{\sigma_A^2}{(N+1)\sigma_A^2 + \sigma^2} (x[N] - \hat{A}[N-1]).$$

Se obtuvo que la actualización del estimador coincide con el caso en que $\mu_A = 0$ (ecuación 12.23). Además, del ejemplo de la sección 10.2.1, el MSE bayesiano mínimo del estimador MMSE está dado por la ecuación 10.9, y como no depende de μ_A , coincide con $\text{Bsme}(\hat{A}[N-1])$, por lo que la deducción continúa como en el caso con $\mu_A = 0$ de la sección 12.4. Se concluye que las ecuaciones 12.25-12.27 son válidas para el caso en que $\mu_A \neq 0$.

12.6.10. Problema 10

Dado un conjunto de vectores $\{x_1, x_2, \dots, x_n\}$, el proceso de ortogonalización de Gram-Schmidt encuentra un nuevo conjunto de vectores $\{e_1, e_2, \dots, e_n\}$ ortonormales (ortogonales y de largo unidad) o $(e_i, e_j) = \delta_{ij}$. El procedimiento es

- $e_1 = \frac{x_1}{\|x_1\|}$
- $z_2 = x_2 - (x_2, e_1)e_1$
 $e_2 = \frac{z_2}{\|z_2\|}$
- y así sucesivamente,

o en general para $n \geq 2$

$$z_n = x_n - \sum_{i=0}^{n-1} (x_n, e_i) e_i$$

$$e_n = \frac{z_n}{\|z_n\|}.$$

Dar una interpretación geométrica del procedimiento. Para los vectores euclidianos

$$\begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

encontrar tres vectores ortonormales empleando el procedimiento de Gram-Schmidt. El producto interno se define como $(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T \mathbf{y}$ para los vectores euclidianos \mathbf{x} e \mathbf{y} .

Solución La interpretación geométrica se muestra en la figura 12.8, donde se asume que ya se calculó e_1 y e_2 a partir de x_1 y x_2 . En el paso $n = 3$ se calcula e_3 a partir de x_3 , e_1 y e_2 . Para eso, se obtiene z_3 como

$$z_3 = x_3 - [(x_3, e_1)e_1 + (x_3, e_2)e_2].$$

Pero por la definición geométrica de producto interno¹,

$$(x_3, e_i) = \|x_3\| \|e_i\| \cos \theta_i = \|x_3\| \cos \theta_i$$

donde θ_i es el ángulo formado entre x_3 y e_i y en la segunda igualdad se consideró que $\|e_i\| = \sqrt{(e_i, e_i)} = 1$. Por lo tanto, el número (x_3, e_i) es el largo de la proyección de x_3 sobre e_i , y el vector $(x_3, e_i)e_i$ es el

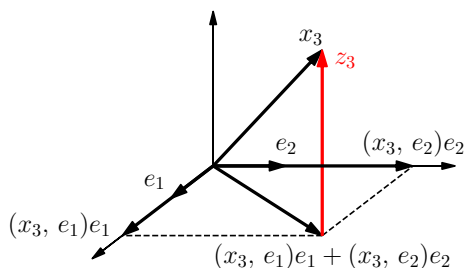


Figura 12.8: Interpretación geométrica del procedimiento de Gram-Schmidt.

vector de largo la proyección de x_3 sobre e_i con la dirección de e_i . En resumen, los vectores $(x_3, e_1)e_1$ y $(x_3, e_2)e_2$ son los vectores formados por la proyección de x_3 sobre e_1 y e_2 respectivamente, como se muestra en la figura. De esta forma, $(x_3, e_1)e_1 + (x_3, e_2)e_2$ es la proyección de x_3 en el subespacio generado por $\{e_1, e_2\}$. Por lo tanto, el vector $z_3 = x_3 - [(x_3, e_1)e_1 + (x_3, e_2)e_2]$ es ortogonal al subespacio generado por $\{e_1, e_2\}$ o ortogonal a e_1 y e_2 , como se aprecia en la figura. Finalmente, $e_3 = z_3 / \|z_3\|$ tiene la misma dirección de z_3 y por lo tanto es ortogonal a e_1 y e_2 y tiene norma unidad.

Se aplica a continuación el procedimiento de Gram-Schmidt a los vectores

$$x_1 = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}, \quad x_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad x_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

El primer paso es

$$e_1 = \frac{x_1}{\|x_1\|} = \frac{1}{\sqrt{1^2 + 1^2 + 2^2}} \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}.$$

¹Por la equivalencia entre la definición geométrica y la definición algebraica $(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T \mathbf{y}$ de producto interno ver por ejemplo https://en.wikipedia.org/wiki/Dot_product#Equivalence_of_the_definitions.

En el segundo paso hay que calcular primero $z_2 = x_2 - (e_1, x_2)e_1$. Considerando que

$$(e_1, x_2) = [1 \quad 0 \quad 1] \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} = \frac{3}{\sqrt{6}},$$

se obtiene que

$$z_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} - \frac{3}{\sqrt{6}} \times \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1/2 \\ -1/2 \\ 0 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}.$$

Luego,

$$e_2 = \frac{z_2}{\|z_2\|} = \frac{1}{\frac{\sqrt{2}}{2}} \times \frac{1}{2} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}.$$

En el tercer paso, se comienza calculando

$$z_3 = x_3 - (x_3, e_1)e_1 - (x_3, e_2)e_2,$$

donde

$$(x_3, e_1) = [1 \quad 1 \quad 1] \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} = \frac{4}{\sqrt{6}} \quad \text{y} \quad (x_3, e_2) = [1 \quad 1 \quad 1] \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} = 0.$$

Por lo tanto,

$$z_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \frac{4}{\sqrt{6}} \times \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \frac{2}{3} \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1/3 \\ 1/3 \\ -1/3 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}.$$

Finalmente

$$e_3 = \frac{z_3}{\|z_3\|} = \frac{1}{\frac{\sqrt{3}}{3}} \times \frac{1}{3} \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}.$$

El conjunto de vectores ortonormales obtenido es

$$e_1 = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}, \quad e_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \quad e_3 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}.$$

12.6.11. Problema 11

Sea \mathbf{x} el vector compuesto por tres variables aleatorias de media nula con matriz de covarianza

$$\mathbf{C}_{xx} = \begin{bmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{bmatrix}.$$

Si $\mathbf{y} = \mathbf{Ax}$, determinar la matriz \mathbf{A} 3×3 de forma tal que la matriz de covarianza de \mathbf{y} sea \mathbf{I} , o equivalentemente, que las variables aleatorias $\{y_1, y_2, y_3\}$ sean no correlacionadas y tengan varianza unidad. Emplear el procedimiento de ortogonalización de Gram-Schmidt del problema de la sección 12.6.10 para hacerlo. ¿Qué es interesante sobre \mathbf{A} ? Finalmente, relacionar \mathbf{C}_{xx} con \mathbf{A} . Notar que \mathbf{A} puede verse como un transformación de blanqueado (ver la sección 4.5).

Solución Aplicando el procedimiento de Gram-Schmidt a los elementos del vector \mathbf{x} , se obtiene un conjunto de variables aleatorias ortonormales, y por lo tanto, la matriz de covarianza resultante es la matriz identidad. Siguiendo el procedimiento descrito en el problema de la sección 12.6.10 y considerando que

$$\mathbf{C}_{xx} = E(\mathbf{xx}^T) = \begin{bmatrix} E(x_1^2) & E(x_1x_2) & E(x_1x_3) \\ E(x_2x_1) & E(x_2^2) & E(x_2x_3) \\ E(x_3x_1) & E(x_3x_2) & E(x_3^2) \end{bmatrix} = \begin{bmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{bmatrix},$$

se comienza calculando y_1 como

$$y_1 = \frac{x_1}{\|x_1\|} = \frac{x_1}{\sqrt{E(x_1^2)}} = x_1.$$

Luego, se calcula y_2 comenzando con

$$z_2 = x_2 - (x_2, y_1)y_1 = x_2 - (x_2, x_1)x_1 = x_2 - \rho x_1,$$

por lo que

$$y_2 = \frac{z_2}{\|z_2\|} = \frac{z_2}{\sqrt{E(z_2^2)}} = \frac{x_2 - \rho x_1}{\sqrt{E[(x_2 - \rho x_1)^2]}} = \frac{x_2 - \rho x_1}{\sqrt{E(x_2^2) - 2\rho E(x_1x_2) + \rho^2 E(x_1^2)}} = \frac{x_2 - \rho x_1}{\sqrt{1 - 2\rho^2 + \rho^2}},$$

resultando en

$$y_2 = \frac{x_2 - \rho x_1}{\sqrt{1 - \rho^2}}.$$

Finalmente, para calcular y_3 se tiene que

$$\begin{aligned} z_3 &= x_3 - (x_3, y_1)y_1 - (x_3, y_2)y_2 \\ &= x_3 - E(x_3y_1)y_1 - E(x_3y_2)y_2 \\ &= x_3 - E(x_3x_1)x_1 - E\left[x_3\left(\frac{x_2 - \rho x_1}{\sqrt{1 - \rho^2}}\right)\right]\frac{x_2 - \rho x_1}{\sqrt{1 - \rho^2}} \\ &= x_3 - E(x_3x_1)x_1 - \frac{1}{1 - \rho^2}[E(x_3x_2) - \rho E(x_3x_1)](x_2 - \rho x_1) \\ &= x_3 - \rho^2 x_1 - \frac{1}{1 - \rho^2}(\rho - \rho^3)(x_2 - \rho x_1) \\ &= x_3 - \rho^2 x_1 - \rho(x_2 - \rho x_1), \end{aligned}$$

resultando en

$$z_3 = x_3 - \rho x_2,$$

por lo que

$$y_3 = \frac{z_3}{\|z_3\|} = \frac{z_3}{\sqrt{E(z_3^2)}} = \frac{x_3 - \rho x_2}{\sqrt{E[(x_3 - \rho x_2)^2]}} = \frac{x_3 - \rho x_2}{\sqrt{E(x_3^2) - 2\rho E(x_2x_3) + \rho^2 E(x_2^2)}} = \frac{x_3 - \rho x_2}{\sqrt{1 - 2\rho^2 + \rho^2}},$$

obteniendo que

$$y_3 = \frac{x_3 - \rho x_2}{\sqrt{1 - \rho^2}}.$$

En resumen,

$$y_1 = x_1, \quad y_2 = \frac{x_2 - \rho x_1}{\sqrt{1 - \rho^2}}, \quad y_3 = \frac{x_3 - \rho x_2}{\sqrt{1 - \rho^2}},$$

que en forma matricial se expresa como

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ -\frac{\rho}{\sqrt{1 - \rho^2}} & \frac{1}{\sqrt{1 - \rho^2}} & 0 \\ 0 & -\frac{\rho}{\sqrt{1 - \rho^2}} & \frac{1}{\sqrt{1 - \rho^2}} \end{bmatrix}}_{\mathbf{A}} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$

Se observa que la matriz \mathbf{A} es triangular inferior. Esto es así en general debido al procedimiento de Gram-Schmidt. Efectivamente, en el procedimiento y_1 depende solo de x_1 , y_2 depende de x_1 y x_2 y así sucesivamente.

Por construcción, la matriz de covarianza de $\mathbf{y} = [y_1 \ y_2 \ y_3]^T$ cumple que $\mathbf{C}_{yy} = \mathbf{I}$. Además,

$$\mathbf{C}_{yy} = E(\mathbf{y}\mathbf{y}^T) = E[(\mathbf{A}\mathbf{x})(\mathbf{A}\mathbf{x})^T] = \mathbf{A}E(\mathbf{x}\mathbf{x}^T)\mathbf{A}^T = \mathbf{A}\mathbf{C}_{xx}\mathbf{A}^T.$$

Por lo tanto

$$\mathbf{A}\mathbf{C}_{xx}\mathbf{A}^T = \mathbf{I} \quad \Rightarrow \quad \mathbf{C}_{xx} = \mathbf{A}^{-1}\mathbf{A}^{T-1} = (\mathbf{A}^T\mathbf{A})^{-1} \quad \Rightarrow \quad \mathbf{C}_{xx}^{-1} = \mathbf{A}^T\mathbf{A}.$$

Esto indica que la matriz \mathbf{A} de blanqueado provee la descomposición de Cholesky de \mathbf{C}_{xx}^{-1} .

12.6.12. Problema 12

Para el estimador LMMSE secuencial de un parámetro escalar explicar que ocurre si $\sigma_n^2 \rightarrow 0$ para algún n . ¿Se obtienen los mismos resultados para el caso vectorial?

Solución El estimador LMMSE secuencial está dado por las ecuaciones 12.37-12.39. Teniendo en cuenta que se considera el caso escalar, se tiene que

$$K[n] = \frac{M[n-1]h[n]}{\sigma_n^2 + h[n]M[n-1]h[n]} \xrightarrow{\sigma_n^2 \rightarrow 0} \frac{M[n-1]h[n]}{h[n]M[n-1]h[n]} = \frac{1}{h[n]},$$

y el estimador se actualiza como

$$\hat{\theta}[n] = \hat{\theta}[n-1] + K[n](x[n] - h[n]\hat{\theta}[n-1]) \xrightarrow{\sigma_n^2 \rightarrow 0} \frac{x[n]}{h[n]}.$$

Esto indica que se descartan todas las muestras previas y se emplea únicamente la muestra actual, ya que al no ser ruidosa, es una observación perfecta. Además, el MSE mínimo se actualiza como

$$M[n] = (1 - K[n]h[n])M[n-1] \xrightarrow{\sigma_n^2 \rightarrow 0} 0.$$

En el caso vectorial se tiene que

$$\mathbf{K}[n] = \frac{\mathbf{M}[n-1]\mathbf{h}[n]}{\sigma_n^2 + \mathbf{h}^T\mathbf{M}[n-1]\mathbf{h}[n]} \xrightarrow{\sigma_n^2 \rightarrow 0} \frac{\mathbf{M}[n-1]\mathbf{h}[n]}{\mathbf{h}^T\mathbf{M}[n-1]\mathbf{h}[n]},$$

donde el numerador es un vector $p \times 1$ y el denominador es un escalar, por lo que la expresión no puede simplificarse. Se concluye que el resultado en el caso vectorial no es análogo al obtenido para el caso escalar. En el caso vectorial se necesitan p observaciones sin ruido para estimar a $\boldsymbol{\theta}$ de forma perfecta.

12.6.13. Problema 13

Encontrar el estimador LMMSE secuencial para el problema descrito en el ejemplo de la sección 12.1.1. Verificar que el estimador secuencial es idéntico al estimador por lotes dado por la ecuación 12.9. Sugerencia: resolver para la secuencia $1/K[n]$.

Solución El ejemplo de la sección 12.1.1 consiste en la estimación del nivel de DC en WGN con PDF a priori uniforme, por lo que los datos verifican el modelo lineal bayesiano

$$\mathbf{x} = \mathbf{H}\mathbf{A} + \mathbf{w}, \quad \text{con} \quad \mathbf{H} = [1 \ \dots \ 1]^T,$$

por lo que el estimador LMMSE secuencial está dado por las ecuaciones 12.37-12.39 para el caso escalar con $\mathbf{h}[n] = 1$ y $\sigma_n^2 = \sigma^2$. De esta forma, la actualización del estimador es

$$\hat{A}[n] = \hat{A}[n-1] + K[n](x[n] - \hat{A}[n-1]),$$

donde

$$K[n] = \frac{M[n-1]}{\sigma^2 + M[n-1]},$$

y la actualización del de MSE mínimo es

$$M[n] = (1 - K[n])M[n-1].$$

Se buscará resolver la secuencia $1/K[n]$. Para eso, se parte substituyendo la ganancia en la ecuación de actualización del MSE mínimo

$$M[n] = \left(1 - \frac{M[n-1]}{\sigma^2 + M[n-1]}\right) M[n-1] = \frac{\sigma^2 M[n-1]}{\sigma^2 + M[n-1]} = \sigma^2 K[n],$$

y por lo tanto

$$M[n-1] = \sigma^2 K[n-1].$$

Sustituyendo este resultado en la ecuación de la ganancia, se obtiene que

$$K[n] = \frac{K[n-1]}{1 + K[n-1]},$$

o

$$\frac{1}{K[n]} = \frac{1 + K[n-1]}{K[n-1]} = \frac{1}{K[n-1]} + 1.$$

Por lo tanto, se tiene que

$$\begin{aligned} \frac{1}{K[1]} &= \frac{1}{K[0]} + 1 \\ \frac{1}{K[2]} &= \frac{1}{K[1]} + 1 = \frac{1}{K[0]} + 2, \end{aligned}$$

y así sucesivamente, y en general, se cumple que

$$\frac{1}{K[n]} = \frac{1}{K[0]} + n. \quad (12.60)$$

Las condiciones iniciales del estimador son (ver la sección 12.4.1)

$$\hat{A}[-1] = E(A) = 0 \quad \text{y} \quad M[-1] = E[(A - \hat{A}[-1])^2] = E(A^2) \stackrel{(a)}{=} \frac{A_0^2}{3},$$

donde la igualdad (a) se obtuvo en el ejemplo 12.1.1. Por lo tanto,

$$K[0] = \frac{M[-1]}{\sigma^2 + M[-1]} = \frac{\frac{A_0^2}{3}}{\sigma^2 + \frac{A_0^2}{3}} \quad \Rightarrow \quad \frac{1}{K[0]} = \frac{\sigma^2}{\frac{A_0^2}{3}} + 1$$

y substituyendo en la ecuación 12.60 se obtiene que

$$\frac{1}{K[n]} = \frac{\sigma^2}{\frac{A_0^2}{3}} + n + 1.$$

Finalmente,

$$K[n] = \frac{\frac{A_0^2}{3}}{\sigma^2 + \frac{(n+1)A_0^2}{3}} \quad (12.61)$$

que es la solución de la recursión de la ganancia. A continuación se resolverá la recursión del estimador, para lo cual se empleará además que

$$1 - K[n] = 1 - \frac{\frac{A_0^2}{3}}{\sigma^2 + \frac{(n+1)A_0^2}{3}} = \frac{\sigma^2 + \frac{nA_0^2}{3}}{\sigma^2 + \frac{(n+1)A_0^2}{3}}. \quad (12.62)$$

Observando que la recursión del estimador se puede escribir como

$$\hat{A}[n] = (1 - K[n])\hat{A}[n-1] + K[n]x[n],$$

se comienza calculando

$$\hat{A}[0] = (1 - K[0])\hat{A}[-1] + K[0]x[0] = K[0]x[0] = \frac{\frac{A_0^2}{3}}{\sigma^2 + \frac{A_0^2}{3}}x[0],$$

donde se sustituyó $K[0]$ empleando la ecuación 12.61 evaluada en $n = 0$ y se tuvo en cuenta que $\hat{A}[-1] = 0$, como se estableció previamente. En los pasos que siguen se sustituye $K[n]$ y $(1 - K[n])$ empleando las ecuaciones 12.61 y 12.62 respectivamente evaluadas en el valor de n correspondiente. Luego,

$$\begin{aligned} \hat{A}[1] &= (1 - K[1])\hat{A}[0] + K[1]x[1] \\ &= \left(\frac{\sigma^2 + \frac{A_0^2}{3}}{\sigma^2 + \frac{2A_0^2}{3}} \right) \left(\frac{\frac{A_0^2}{3}}{\sigma^2 + \frac{A_0^2}{3}}x[0] \right) + \frac{\frac{A_0^2}{3}}{\sigma^2 + \frac{2A_0^2}{3}}x[1] \\ &= \frac{\frac{A_0^2}{3}}{\sigma^2 + \frac{2A_0^2}{3}}(x[0] + x[1]) \\ &= \frac{\frac{2A_0^2}{3}}{\sigma^2 + \frac{2A_0^2}{3}} \left(\frac{x[0] + x[1]}{2} \right) \\ &= \frac{\frac{A_0^2}{3}}{\frac{\sigma^2}{2} + \frac{A_0^2}{3}} \left(\frac{x[0] + x[1]}{2} \right). \end{aligned} \quad (12.63)$$

Continuando con $n = 2$,

$$\begin{aligned}
\hat{A}[2] &= (1 - K[2])\hat{A}[1] + K[2]x[2] \\
&\stackrel{(a)}{=} \left(\frac{\sigma^2 + \frac{2A_0^2}{3}}{\sigma^2 + \frac{3A_0^2}{3}} \right) \left(\frac{\frac{A_0^2}{3}}{\sigma^2 + \frac{2A_0^2}{3}} (x[0] + x[1]) \right) + \frac{\frac{A_0^2}{3}}{\sigma^2 + \frac{3A_0^2}{3}} x[2] \\
&= \frac{\frac{A_0^2}{3}}{\sigma^2 + \frac{3A_0^2}{3}} (x[0] + x[1] + x[2]) \\
&= \frac{\frac{3A_0^2}{3}}{\sigma^2 + \frac{3A_0^2}{3}} \left(\frac{x[0] + x[1] + x[2]}{3} \right) \\
&= \frac{\frac{A_0^2}{3}}{\frac{\sigma^2}{3} + \frac{A_0^2}{3}} \left(\frac{x[0] + x[1] + x[2]}{3} \right),
\end{aligned}$$

donde en (a) se substituyó $\hat{A}[1]$ por la ecuación 12.63. En general, se cumple que

$$\hat{A}[n] = \frac{\frac{A_0^2}{3}}{\frac{\sigma^2}{n+1} + \frac{A_0^2}{3}} \bar{x},$$

que coincide con el estimador por lotes de la ecuación 12.9.

12.6.14. Problema 14

En este problema se examina la interpolación de una muestra de datos. Se asume que se dispone del conjunto de datos $\{x[n-M], \dots, x[n-1], x[n+1], \dots, x[n+M]\}$ y se desea estimar o interpolar $x[n]$. Se supone que los datos y $x[n]$ son la realización de un proceso WSS de media nula. Sea el estimador LMMSE de $x[n]$

$$\hat{x}[n] = \sum_{\substack{k=-M \\ k \neq 0}}^M a_k x[n-k].$$

Encontrar el conjunto de ecuaciones lineales a resolver para obtener los coeficientes de ponderación empleando el principio de ortogonalidad. Luego, probar que $a_{-k} = a_k$ y explicar porque esto debe cumplirse.

Solución El principio de ortogonalidad (ecuación 12.10) indica que en este caso

$$x[n] - \hat{x}[n] \perp x[n-M], \dots, x[n-1], x[n+1], \dots, x[n+M],$$

o empleando la definición de ortogonalidad,

$$E[(x[n] - \hat{x}[n])x[n-l]] = 0, \quad l = -M, \dots, M, \quad l \neq 0,$$

es decir,

$$E(\hat{x}[n]x[n-l]) = E(x[n]x[n-l]), \quad l = -M, \dots, M, \quad l \neq 0.$$

Substituyendo el estimador, la condición queda

$$E \left(\sum_{\substack{k=-M \\ k \neq 0}}^M a_k x[n-k] x[n-l] \right) = E(x[n]x[n-l]), \quad l = -M, \dots, M, \quad l \neq 0,$$

o

$$\sum_{\substack{k=-M \\ k \neq 0}}^M a_k E(x[n-k]x[n-l]) = E(x[n]x[n-l]), \quad l = -M, \dots, M, \quad l \neq 0,$$

resultando en

$$\sum_{\substack{k=-M \\ k \neq 0}}^M a_k r_{xx}[l-k] = r_{xx}[l], \quad l = -M, \dots, M, \quad l \neq 0. \quad (12.64)$$

En forma matricial, el sistema de ecuaciones queda

$$\begin{bmatrix} r_{xx}[0] & \dots & r_{xx}[-M+1] & r_{xx}[-M-1] & \dots & r_{xx}[-2M] \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ r_{xx}[M-1] & \dots & r_{xx}[0] & r_{xx}[-2] & \dots & r_{xx}[-M-1] \\ r_{xx}[M+1] & \dots & r_{xx}[2] & r_{xx}[0] & \dots & r_{xx}[-M+1] \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ r_{xx}[2M] & \dots & r_{xx}[M+1] & r_{xx}[M-1] & \dots & r_{xx}[0] \end{bmatrix} \begin{bmatrix} a_{-M} \\ \vdots \\ a_{-1} \\ a_1 \\ \vdots \\ a_M \end{bmatrix} = \begin{bmatrix} r_{xx}[-M] \\ \vdots \\ r_{xx}[-1] \\ r_{xx}[1] \\ \vdots \\ r_{xx}[M] \end{bmatrix}$$

Notar que como se cumple que $r_{xx}[-k] = r_{xx}[k]$, la matriz es simétrica, pero no es Toeplitz. Para probar que $a_{-k} = a_k$ se comienza realizando el cambio de variable $k' = -k$ en la ecuación 12.64, obteniendo que

$$\sum_{\substack{k'=-M \\ k' \neq 0}}^M a_{-k'} r_{xx}[l+k'] = r_{xx}[l], \quad l = -M, \dots, M, \quad l \neq 0,$$

y luego el cambio de variable $l' = -l$, resultando en

$$\sum_{\substack{k'=-M \\ k' \neq 0}}^M a_{-k'} r_{xx}[-l'+k'] = r_{xx}[-l'], \quad l' = -M, \dots, M, \quad l' \neq 0,$$

y considerando que $r_{xx}[-k] = r_{xx}[k]$, el sistema de ecuaciones queda

$$\sum_{\substack{k'=-M \\ k' \neq 0}}^M a_{-k'} r_{xx}[l'-k'] = r_{xx}[l'], \quad l' = -M, \dots, M, \quad l' \neq 0.$$

Se concluye que como este sistema es el mismo que el de la ecuación 12.64 y tiene una solución única, se cumple que $a_{-k} = a_k$.

12.6.15. Problema 15

Considérese el filtro de suavizado de Wiener para una única muestra dado por la ecuación 12.47. Reescribir W y $M_{\hat{s}}$ como una función del coeficiente de correlación

$$\rho = \frac{\text{cov}(s[0], x[0])}{\sqrt{\text{var}(s[0]) \text{var}(x[0])}}$$

entre $s[0]$ y $x[0]$. Explicar los resultados.

Solución El coeficiente de correlación es

$$\begin{aligned}\rho &\stackrel{(a)}{=} \frac{E(s[0]x[0])}{\sqrt{E(s^2[0])E(x^2[0])}} \\ &= \frac{E[s[0](s[0] + w[0])]}{\sqrt{E(s^2[0])E[(s[0] + w[0])^2]}} \\ &\stackrel{(b)}{=} \frac{E(s^2[0])}{\sqrt{E(s^2[0])[E(s^2[0]) + E(w^2[0])]}} \\ &= \frac{r_{ss}[0]}{\sqrt{r_{ss}[0](r_{ss}[0] + r_{ww}[0])}},\end{aligned}$$

donde en (a) se consideró que $x[0]$ y $s[0]$ tienen media nula y en (b) que $s[0]$ y $w[0]$ son no correlacionados y tienen media nula. Elevando al cuadrado ambos lados de la igualdad se tiene que

$$\begin{aligned}\rho^2 &= \frac{r_{ss}^2[0]}{r_{ss}[0](r_{ss}[0] + r_{ww}[0])} \\ &= \frac{r_{ss}[0]}{r_{ss}[0] + r_{ww}[0]},\end{aligned}$$

y comparando el resultado con la ecuación 12.47 se concluye que

$$W = \rho^2 \quad \text{y} \quad M_{\hat{s}} = (1 - \rho^2)r_{ss}[0],$$

y el estimador LMMSE es

$$\hat{s}[0] = \rho^2 x[0].$$

Considerando que la SNR es $\eta = r_{ss}[0]/r_{ww}[0]$, se cumple que

$$\rho^2 = \frac{\eta}{\eta + 1}.$$

Una SNR alta ($\eta \rightarrow \infty$) implica que $\rho \rightarrow 1$, es decir, las muestras $s[0]$ y $x[0]$ están fuertemente correlacionadas. El estimador en ese caso es $\hat{s}[0] = x[0]$. Si la SNR es baja ($\eta \rightarrow 0$) implica que $\rho \rightarrow 0$, y las muestras $s[0]$ y $x[0]$ tienden a ser no correlacionadas. En ese caso, el estimador es $\hat{s}[n] = 0$.

12.6.16. Problema 16

En este problema se explora la solución de las ecuaciones de Wiener-Hopf en base a la muestra actual e infinitas muestras pasadas, o la solución de

$$\sum_{k=0}^{\infty} h[k]r_{xx}[l-k] = r_{ss}[l], \quad l \geq 0$$

mediante el uso de la transformada z . Primero se define la transformada z unilateral como

$$\begin{aligned}[\mathcal{X}(z)]_+ &= \left[\sum_{n=-\infty}^{\infty} x[n]z^{-n} \right]_+ \\ &= \sum_{n=0}^{\infty} x[n]z^{-n},\end{aligned}$$

que es la transformada z habitual pero omitiendo las potencias positivas de z . Luego, las ecuaciones de Wiener-Hopf se escriben como

$$(h * r_{xx})[n] - r_{ss}[n] = 0, \quad n \geq 0, \quad (12.65)$$

donde $h[n]$ se restringe a ser causal. La transformada z habitual (bilateral) del lado izquierdo de la igualdad es

$$\mathcal{H}(z)\mathcal{P}_{xx}(z) - \mathcal{P}_{ss}(z)$$

de forma de que para satisfacer a condición 12.65 se debe cumplir que

$$[\mathcal{H}(z)\mathcal{P}_{xx}(z) - \mathcal{P}_{ss}(z)]_+ = 0.$$

Mediante el teorema de factorización espectral [14], si $\mathcal{P}_{xx}(z)$ no tiene ceros en la circunferencia unidad, puede ser factorizado como

$$\mathcal{P}_{xx}(z) = \mathcal{B}(z)\mathcal{B}(z^{-1}),$$

donde $\mathcal{B}(z)$ es la transformada z de una secuencia casual y $\mathcal{B}(z^{-1})$ es la transformada z de una secuencia anticausal. Por lo tanto

$$[\mathcal{H}(z)\mathcal{B}(z)\mathcal{B}(z^{-1}) - \mathcal{P}_{ss}(z)]_+ = 0.$$

Sea $\mathcal{G}(z) = \mathcal{H}(z)\mathcal{B}(z)$ de forma que

$$\left[\mathcal{B}(z^{-1}) \left(\mathcal{G}(z) - \frac{\mathcal{P}_{ss}(z)}{\mathcal{B}(z^{-1})} \right) \right]_+ = 0.$$

Notando que $\mathcal{G}(z)$ es la transformada z de una secuencia causal, mostrar que

$$\mathcal{H}(z) = \frac{1}{\mathcal{B}(z)} \left[\frac{\mathcal{P}_{ss}}{\mathcal{B}(z^{-1})} \right]_+ = 0.$$

Por un ejemplo del cálculo involucrado en determinar el filtro de Wiener, ver [14].

Solución El sistema a resolver está dado por la ecuación 12.48, que se puede escribir como

$$(h * r_{xx})[n] - r_{ss}[n] = 0, \quad n \geq 0,$$

donde $h[n]$ se restringe a ser causal. Para satisfacer esta condición, se debe cumplir que

$$[\mathcal{H}(z)\mathcal{P}_{xx}(z) - \mathcal{P}_{ss}(z)]_+ = 0.$$

Realizado la factorización espectral

$$\mathcal{P}_{xx}(z) = \mathcal{B}(z)\mathcal{B}(z^{-1}),$$

donde $\mathcal{B}(z)$ es la transformada z de una secuencia casual y $\mathcal{B}(z^{-1})$ es la transformada z de una secuencia anticausal, la condición queda

$$[\mathcal{H}(z)\mathcal{B}(z)\mathcal{B}(z^{-1}) - \mathcal{P}_{ss}(z)]_+ = 0.$$

Definiendo $\mathcal{G}(z) = \mathcal{H}(z)\mathcal{B}(z)$, que es la transformada z de una secuencia causal por ser $\mathcal{H}(z)$ y $\mathcal{B}(z)$ las transformadas de secuencias causales, la condición se expresa como

$$[\mathcal{G}(z)\mathcal{B}(z^{-1}) - \mathcal{P}_{ss}(z)]_+ = 0.$$

o

$$\left[\mathcal{B}(z^{-1}) \left(\mathcal{G}(z) - \frac{\mathcal{P}_{ss}(z)}{\mathcal{B}(z^{-1})} \right) \right]_+ = 0.$$

Esto implica que la secuencia

$$\mathcal{Z}^{-1} \left\{ \mathcal{B}(z^{-1}) \left(\mathcal{G}(z) - \frac{\mathcal{P}_{ss}(z)}{\mathcal{B}(z^{-1})} \right) \right\} = \mathcal{Z}^{-1} \{ \mathcal{B}(z^{-1}) \} * \mathcal{Z}^{-1} \left\{ \mathcal{G}(z) - \frac{\mathcal{P}_{ss}(z)}{\mathcal{B}(z^{-1})} \right\} = 0, \quad n \geq 0.$$

Pero $\mathcal{B}(z^{-1})$ es la transformada z de una secuencia anticausal, por lo que para cumplir la condición alcanza que

$$\mathcal{Z}^{-1} \left\{ \mathcal{G}(z) - \frac{\mathcal{P}_{ss}(z)}{\mathcal{B}(z^{-1})} \right\} = 0, \quad n \geq 0,$$

es decir, esta secuencia debe ser anticausal y además tomar valor 0 en $n = 0$. La condición equivalente es

$$\left[\mathcal{G}(z) - \frac{\mathcal{P}_{ss}(z)}{\mathcal{B}(z^{-1})} \right]_+ = 0$$

o

$$[\mathcal{G}(z)]_+ = \left[\frac{\mathcal{P}_{ss}(z)}{\mathcal{B}(z^{-1})} \right]_+.$$

Como $\mathcal{G}(z)$ es la transformada z de una secuencia causal, se cumple que

$$[\mathcal{G}(z)]_+ = \mathcal{G}(z) = \mathcal{H}(z)\mathcal{B}(z),$$

y por lo tanto,

$$\mathcal{H}(z)\mathcal{B}(z) = \left[\frac{\mathcal{P}_{ss}(z)}{\mathcal{B}(z^{-1})} \right]_+,$$

resultando en

$$\mathcal{H}(z) = \frac{1}{\mathcal{B}(z)} \left[\frac{\mathcal{P}_{ss}(z)}{\mathcal{B}(z^{-1})} \right]_+.$$

En la sección 4.7 de [14] se resuelven las ecuaciones de Wiener-Hopf del filtro de Wiener infinito mediante el teorema de factorización espectral.

12.6.17. Problema 17

Derivar el filtro de suavizado infinito de Wiener dado por la ecuación 12.50 asumiendo que

$$\hat{s}[n] = \sum_{k=-\infty}^{\infty} h[k]x[n-k]$$

y luego empleando el principio de ortogonalidad.

Solución Para el suavizado de Wiener infinito, el principio de ortogonalidad es

$$E[(s[n] - \hat{s}[n])x[n-l]] = 0, \quad -\infty < l < \infty.$$

Sustituyendo el estimador, se tiene que

$$E \left[\left(s[n] - \sum_{k=-\infty}^{\infty} h[k]x[n-k] \right) x[n-l] \right] = 0, \quad -\infty < l < \infty,$$

o

$$E(s[n]x[n-l]) - \sum_{k=-\infty}^{\infty} h[k]E(x[n-k]x[n-l]) = 0, \quad -\infty < l < \infty,$$

resultando en

$$r_{ss}[l] - \sum_{k=-\infty}^{\infty} h[k]r_{xx}[l-k] = 0, \quad -\infty < l < \infty.$$

12.6.18. Problema 18

Para el filtro de suavizado infinito de Wiener mostrar que el MSE bayesiano mínimo es

$$M_{\hat{s}} = r_{ss}[0] - \sum_{k=-\infty}^{\infty} h[k]r_{ss}[k].$$

Luego, empleando técnicas de transformada de Fourier, mostrar que puede escribirse como

$$M_{\hat{s}} = \int_{-\frac{1}{2}}^{\frac{1}{2}} P_{ss}(1 - H(f)) df$$

donde

$$H(f) = \frac{P_{ss}(f)}{P_{ss}(f) + P_{ww}(f)}.$$

Evaluar el filtro de suavizado y el MSE mínimo si

$$P_{ss}(f) = \begin{cases} P_0 & |f| \leq \frac{1}{4} \\ 0 & \frac{1}{4} < |f| \leq \frac{1}{2} \end{cases}$$

$$P_{ww}(f) = \sigma^2$$

y explicar los resultados.

Solución Por definición, el MSE bayesiano mínimo es

$$\begin{aligned} M_{\hat{s}} &= E[(s[n] - \hat{s}[n])^2] \\ &= E[s[n](s[n] - \hat{s}[n])] - E[\hat{s}[n](s[n] - \hat{s}[n])] \\ &\stackrel{(a)}{=} E(s^2[n]) - E(s[n]\hat{s}[n]) \\ &= E(s^2[n]) - \sum_{k=-\infty}^{\infty} h[k]E(s[n]x[n-k]), \end{aligned}$$

donde en (a) se tuvo en cuenta que el segundo sumando es nulo, ya que

$$E[\hat{s}[n](s[n] - \hat{s}[n])] = \sum_{k=-\infty}^{\infty} h[k]E[x[n-k](s[n] - \hat{s}[n])] = 0$$

por tratarse del principio de ortogonalidad. Se concluye que

$$M_{\hat{s}} = r_{ss}[0] - \sum_{k=-\infty}^{\infty} h[k]r_{ss}[k].$$

Además, como

$$r_{ss}[n] = \int_{-\frac{1}{2}}^{\frac{1}{2}} P_{ss}(f) e^{j2\pi f n} df,$$

evaluando en $n = 0$ se cumple que

$$r_{ss}[0] = \int_{-\frac{1}{2}}^{\frac{1}{2}} P_{ss}(f) df.$$

De forma similar,

$$\sum_{k=-\infty}^{\infty} h[k]r_{ss}[n-k] = (h * r_{ss})[n] = \int_{-\frac{1}{2}}^{\frac{1}{2}} H(f)P_{ss}(f)e^{j2\pi f n} df,$$

por lo que evaluando en $n = 0$

$$\sum_{k=-\infty}^{\infty} h[k]r_{ss}[k] = (h * r_{ss})[0] = \int_{-\frac{1}{2}}^{\frac{1}{2}} H(f)P_{ss}(f) df.$$

Sustituyendo, se obtiene que

$$\begin{aligned} M_{\hat{s}} &= \int_{-\frac{1}{2}}^{\frac{1}{2}} P_{ss}(f) df - \int_{-\frac{1}{2}}^{\frac{1}{2}} H(f)P_{ss}(f) df \\ &= \int_{-\frac{1}{2}}^{\frac{1}{2}} P_{ss}(f)(1 - H(f)) df, \end{aligned}$$

y aplicando la transformada de Fourier a las ecuaciones de Wiener-Hopf, calculadas en el problema de la sección 12.6.17, se tiene que

$$P_{ss}(f) - H(f)P_{xx}(f) = 0$$

por lo que

$$H(f) = \frac{P_{ss}(f)}{P_{xx}(f)} = \frac{P_{ss}(f)}{P_{ss}(f) + P_{ww}(f)}.$$

Con las PSD dadas, la respuesta en frecuencia del filtro de Wiener es

$$H(f) = \frac{P_{ss}(f)}{P_{ss}(f) + P_{ww}(f)} = \begin{cases} \frac{P_0}{P_0 + \sigma^2} & |f| \leq \frac{1}{4} \\ 0 & \frac{1}{4} < |f| \leq \frac{1}{2} \end{cases},$$

y el MSE bayesiano mínimo queda

$$\begin{aligned} M_{\hat{s}} &= \int_{-\frac{1}{4}}^{\frac{1}{4}} P_0 \left(1 - \frac{P_0}{P_0 + \sigma^2}\right) df \\ &= \int_{-\frac{1}{4}}^{\frac{1}{4}} \frac{P_0 \sigma^2}{P_0 + \sigma^2} df \\ &= \frac{P_0 \sigma^2}{2(P_0 + \sigma^2)}. \end{aligned}$$

Considerando que la SNR en función de la frecuencia es

$$\eta(f) = \frac{P_{ss}(f)}{P_{ww}(f)} = \begin{cases} \frac{P_0}{\sigma^2} & |f| \leq \frac{1}{4} \\ 0 & \frac{1}{4} < |f| \leq \frac{1}{2} \end{cases},$$

es decir, la SNR es nula sobre $f > 1/4$. En el caso en que $P_0 \gg \sigma^2$, el filtro de Wiener tiene respuesta en frecuencia

$$H(f) = \begin{cases} 1 & |f| \leq \frac{1}{4} \\ 0 & \frac{1}{4} < |f| \leq \frac{1}{2} \end{cases},$$

permitiendo pasar sin alteración las frecuencias en donde hay presencia de señal y eliminando las frecuencias en donde únicamente hay ruido. En ese caso además el MSE bayesiano mínimo es

$$M_{\hat{s}} = \frac{\sigma^2}{2},$$

que es la potencia del ruido luego del filtrado.

12.6.19. Problema 19

Sea $x[n]$ un proceso aleatorio WSS de media nula. Se quiere predecir $x[n]$ en base a $\{x[n-1], \dots, x[n-N]\}$. Emplear el principio de ortogonalidad para derivar el estimador LMMSE o predictor. Explicar porque las ecuaciones a resolver son las mismas que las 12.53 con $l = 1$, es decir, porque son independientes de n . Además, derivar nuevamente el MSE mínimo (ecuación 12.54) empleando también el principio de ortogonalidad.

Solución El estimador LMMSE de $x[n]$ en base a $\{x[n-1], \dots, x[n-N]\}$ es

$$\hat{x}[n] = \sum_{k=1}^N h[k]x[n-k],$$

y el principio de ortogonalidad indica que

$$E[(x[n] - \hat{x}[n])x[n-m]] = 0, \quad m = 1, \dots, N,$$

y sustituyendo el estimador y operando se obtiene que

$$E(x[n]x[n-m]) - \sum_{k=1}^N h[k]E(x[n-k]x[n-m]) = 0, \quad m = 1, \dots, N,$$

o

$$r_{xx}[m] - \sum_{k=1}^N h[k]r_{xx}[m-k] = 0, \quad m = 1, \dots, N,$$

que coincide con el sistema de ecuaciones 12.53 con $l = 1$. Resolviendo este sistema de ecuaciones, se obtienen los coeficientes $h[k]$ del estimador, que no dependen de n debido a que la autocorrelación de depende de n por ser $x[n]$ es un proceso WSS. El MSE mínimo es

$$\begin{aligned} M_{\hat{x}} &= E[(x[n] - \hat{x}[n])^2] \\ &= E[x[n](x[n] - \hat{x}[n])] - E[\hat{x}[n](x[n] - \hat{x}[n])] \\ &\stackrel{(a)}{=} E(x^2[n]) - \sum_{k=1}^N h[k]E(x[n]x[n-k]), \end{aligned}$$

resultando en

$$M_{\hat{x}} = r_{xx}[0] - \sum_{k=1}^N h[k]r_{xx}[k],$$

donde en (a) se tuvo en cuenta que el segundo sumando es nulo, ya que

$$E[\hat{x}[n](x[n] - \hat{x}[n])] = \sum_{k=1}^N h[k]E[x[n-k](x[n] - \hat{x}[n])] = 0$$

por el principio de ortogonalidad. El MSE mínimo obtenido coincide con la ecuación 12.54 con $l = 1$.

12.6.20. Problema 20

Se considera el proceso AR(N)

$$x[n] = - \sum_{k=1}^N a[k]x[n-k] + u[n]$$

donde $u[n]$ es ruido blanco con varianza σ_u^2 . Probar que el predictor lineal óptimo de un paso de $x[n]$ es

$$\hat{x}[n] = - \sum_{k=1}^N a[k]x[n-k].$$

Además, encontrar el MSE mínimo. Sugerencia: comparar las ecuaciones a resolver con las ecuaciones de Yule-Walker.

Solución En el problema anterior se encontró que el predictor óptimo es

$$\hat{x}[n] = \sum_{k=1}^N h[k]x[n-k],$$

y los coeficientes cumplen el sistema de ecuaciones de Wiener-Hopf para el predictor

$$r_{xx}[m] = \sum_{k=1}^N h[k]r_{xx}[m-k], \quad m = 1, \dots, N.$$

Por otro lado, la autocorrelación de un proceso $AR(N)$ cumple las ecuaciones de Yule-Walker (ver el capítulo 3 de [5]), que indican que

$$r_{xx}[m] = \begin{cases} -\sum_{k=1}^N a[k]r_{xx}[m-k], & m \geq 1 \\ -\sum_{k=1}^N a[k]r_{xx}[k] + \sigma_u^2, & m = 0. \end{cases}$$

Comparando las ecuaciones de Wiener-Hopf y las de Yule-Walker para $m \geq 1$, se concluye que se debe cumplir que

$$h[k] = -a[k],$$

por lo que el predictor óptimo es

$$\hat{x}[n] = -\sum_{k=1}^N a[k]x[n-k],$$

que es lo que se quería mostrar. Del problema anterior, el MSE mínimo es

$$\begin{aligned} M_{\hat{x}} &= r_{xx}[0] - \sum_{k=1}^N h[k]r_{xx}[k] \\ &= r_{xx}[0] + \sum_{k=1}^N a[k]r_{xx}[k] \\ &= \sigma_u^2, \end{aligned}$$

donde la última igualdad se obtiene de las ecuaciones de Yule-Walker con $m = 0$.

Capítulo 13

Filtros de Kalman

13.1. Introducción

En este capítulo se discute una generalización importante del filtro de Wiener. La importancia de la extensión es que admite señales y ruidos vectoriales que además pueden ser no estacionarios, en contraste con el filtro de Wiener, que solo admite señales y ruidos escalares estacionarios. La generalización se llama *filtro de Kalman*, y puede ser pensado como el estimador MMSE secuencial de una señal contaminada con ruido, donde la señal está caracterizada mediante un modelo dinámico o de estados. Es una generalización del estimador MMSE secuencial de la sección 12.4 que permite que los parámetros desconocidos evolucionen en el tiempo de acuerdo a un modelo dinámico. Si la señal y el ruido son conjuntamente gaussianos, el filtro de Kalman es el estimador MMSE óptimo, y en caso contrario, es el estimador LMMSE óptimo.

13.2. Modelos de señales dinámicas

Se adopta ahora la notación $s[n]$ como la señal a estimar, en contraposición a la notación $\theta[n]$ habitual, y se considerará que es la realización de un proceso aleatorio de media nula. En caso de no ser de media nula, puede estimarse $s[n] - E(s[n])$, y luego se suma $E(s[n])$ al estimador. Se asumirá siempre que la media es conocida.

Un modelo simple para $s[n]$ que permite especificar la correlación entre muestras es el proceso de *Gauss-Markov de primer orden*

$$s[n] = as[n-1] + u[n], \quad n \geq 0 \quad (13.1)$$

donde $u[n]$ es WGN con varianza σ_u^2 , $s[-1] \sim \mathcal{N}(\mu_s, \sigma_s^2)$, y $s[-1]$ es independiente de $u[n]$ para todo $n \geq 0$. El ruido $u[n]$ se llama *ruido de excitación*, ya que $s[n]$ puede verse como la salida de un sistema lineal invariante en el tiempo excitado con $u[n]$. El modelo de la ecuación 13.1 se llama *modelo de estados* o *modelo dinámico*. La salida actual $s[n]$ depende únicamente del estado del sistema en el instante previo, o $s[n-1]$, y la muestra actual de la entrada $u[n]$. El estado es $s[n-1]$ y resume el efecto de todas las muestras pasadas sobre el sistema. La ecuación 13.1 será referida como modelo de Gauss-Markov, entendiendo que es de primer orden.

El modelo de señal 13.1 recuerda a un proceso AR(1), excepto que la señal comienza en $n = 0$ y por lo tanto, puede no ser WSS. En breve se verá que cuando $n \rightarrow \infty$, el efecto de la condición inicial es insignificante y el proceso es en realidad WSS y puede considerarse como un proceso AR(1) de parámetro $a[1] = -a$. Una típica realización de $s[n]$ se muestra en la figura 13.1 con parámetros $a = 0.98$, $\sigma_u^2 = 0.1$, $\mu_s = 5$ y $\sigma_s^2 = 1$. Nótese que la media comienza alrededor de 5 y va decreciendo a cero. Ese comportamiento se debe al transitorio de la respuesta del sistema al valor inicial $s[-1] \approx 5$. Además, se observa que las muestras están fuertemente correlacionadas. Estas consideraciones pueden cuantificarse determinando la media y la covarianza de $s[n]$. De esta forma, el proceso queda completamente caracterizado estadísticamente por tratarse de un proceso gaussiano.

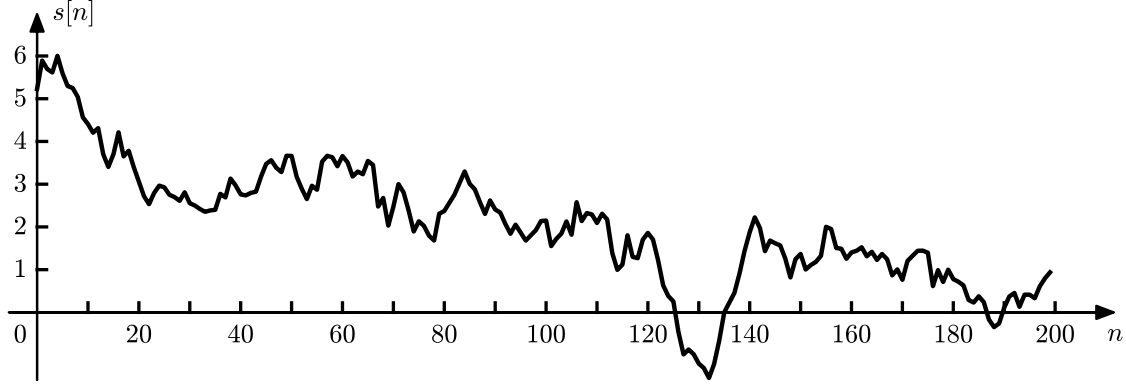


Figura 13.1: Realización de un proceso de Gauss-Markov de primer orden.

Se comenzará expresando $s[n]$ como función de la condición inicial y las entradas como

$$\begin{aligned} s[0] &= as[-1] + u[0] \\ s[1] &= as[0] + u[1] = a^2s[-1] + au[0] + u[1] \\ s[2] &= as[1] + u[2] = a^3s[-1] + a^2u[0] + au[1] + u[2], \end{aligned}$$

y así sucesivamente, y en general se cumple que

$$s[n] = a^{n+1}s[-1] + \sum_{k=0}^n a^k u[n-k]. \quad (13.2)$$

Se observa que $s[n]$ es un función lineal de la condición inicial y el ruido de excitación desde el instante inicial hasta el instante actual. Como esas variables aleatorias son todas independientes y gaussianas, $s[n]$ es gaussiana, y puede mostrarse que además se trata de un proceso aleatorio gaussiano, como se hace en el problema de la sección 13.8.1. De la ecuación 13.2, la media es

$$E(s[n]) = a^{n+1}E(s[-1]) = a^{n+1}\mu_s. \quad (13.3)$$

La covarianza entre las muestras $s[m]$ y $s[n]$, de las ecuaciones 13.2 y 13.3, es

$$\begin{aligned} c_s[m, n] &= E[(s[m] - E(s[m]))(s[n] - E(s[n]))] \\ &= E\left[\left(a^{m+1}s[-1] + \sum_{k=0}^m a^k u[m-k] - a^{m+1}\mu_s\right)\left(a^{n+1}s[-1] + \sum_{l=0}^n a^l u[n-l] - a^{n+1}\mu_s\right)\right] \\ &= E\left[\left(a^{m+1}(s[-1] - \mu_s) + \sum_{k=0}^m a^k u[m-k]\right)\left(a^{n+1}(s[-1] - \mu_s) + \sum_{l=0}^n a^l u[n-l]\right)\right] \\ &= a^{m+n+2}\sigma_s^2 + \sum_{k=0}^m \sum_{l=0}^n a^{k+l} E(u[m-k]u[n-l]). \end{aligned}$$

Pero teniendo en cuenta que $r_{uu}[k] = \sigma_u^2 \delta[k]$, se cumple que

$$E(u[m-k]u[n-l]) = r_{uu}[(m-k) - (n-l)] = r_{uu}[l - (n-m+k)] = \sigma_u^2 \delta[l - (n-m+k)]$$

o

$$E(u[m-k]u[n-l]) = \begin{cases} \sigma_u^2 & l = n - m + k \\ 0 & \text{en otro caso.} \end{cases}$$

Si $m \geq n$, la condición $l = n - m + k$, o $k = l + m - n$, implica que cuando $l = 0$, $k = m - n > 0$, y cuando $l = n$, $k = m$, y son los límites de la sumatoria. Por lo tanto, para $m > n$ la covarianza queda

$$\begin{aligned} c_s[m, n] &= a^{m+n+2} \sigma_s^2 + \sigma_u^2 \sum_{k=m-n}^m a^{k+(n-m+k)} \\ &\stackrel{(a)}{=} a^{m+n+2} \sigma_s^2 + \sigma_u^2 \sum_{l=0}^n a^{2(l+m-n)+n-m} \\ &\stackrel{(b)}{=} a^{m+n+2} \sigma_s^2 + \sigma_u^2 a^{m-n} \sum_{k=0}^n a^{2k}, \end{aligned} \quad (13.4)$$

donde en (a) se realizó el cambio de variable $l = k - (m - n)$ y en (b) se renombró l como k . Además, $c_s[m, n] = c_s[n, m]$ si $m < n$. Notar que la varianza es

$$\begin{aligned} \text{var}(s[n]) &= c_s[n, n] \\ &= a^{2n+2} \sigma_s^2 + \sigma_u^2 \sum_{k=0}^n a^{2k}. \end{aligned} \quad (13.5)$$

Claramente, $s[n]$ no es WSS ya que la media depende de n y la covarianza depende de m y n y no de la diferencia $m - n$. Sin embargo, cuando $n \rightarrow \infty$, de las ecuaciones 13.3 y 13.4, se ve que

$$\begin{aligned} E(s[n]) &\rightarrow 0 \\ c_s[m, n] &\rightarrow \frac{\sigma_u^2 a^{m-n}}{1 - a^2} \end{aligned}$$

ya que $|a| < 1$, condición necesaria para la estabilidad del proceso. En el cálculo del límite de la covarianza se empleó que

$$\sum_{k=0}^n a^{2k} = \frac{1 - a^{2(n+1)}}{1 - a^2} \xrightarrow{n \rightarrow \infty} \frac{1}{1 - a^2}.$$

Se obtuvo que con $n \rightarrow \infty$, la media es nula y la covarianza es la ACF con

$$r_{ss}[k] = c_s[m, n]|_{m-n=k} = \frac{\sigma_u^2}{1 - a^2} a^k \quad k \geq 0, \quad (13.6)$$

que se reconoce como la ACF de un proceso AR(1). Es incluso posible que $s[n]$ sea WSS para $n \geq 0$ si las condiciones iniciales son elegidas apropiadamente (ver el problema de la sección 13.8.2). De la covarianza y la ACF, puede verse además que mediante la elección de a puede controlarse el grado de correlación entre las muestras. Si $|a| \rightarrow 1$, las muestras son fuertemente correlacionadas y con $a \rightarrow 0$ son no correlacionadas. Esto también se deduce de la ecuación 13.1. Ver también el problema de la sección 13.8.3.

La media y la varianza de un proceso de Gauss-Markov pueden expresarse recursivamente, lo cual es útil para la extensión de los resultados al proceso de Gauss-Markov vectorial. A partir de la ecuación 13.1, la media es

$$E(s[n]) = aE(s[n-1]) + E(u[n]),$$

es decir,

$$E(s[n]) = aE(s[n-1]), \quad (13.7)$$

y la varianza es

$$\begin{aligned} \text{var}(s[n]) &= E[(s[n] - E(s[n]))^2] \\ &= E[(as[n-1] + u[n] - aE(s[n-1]))^2] \\ &\stackrel{(a)}{=} a^2 E[(s[n-1] - E(s[n-1]))^2] + E(u^2[n]), \end{aligned}$$

donde en (a) se empleó que $u[n]$ y $s[n-1]$ no están correlacionados, o $E(u[n]s[n-1]) = 0$, ya que de la ecuación 13.2, $s[n-1]$ depende de las muestras $\{s[-1], u[0], \dots, u[n-1]\}$, y por hipótesis son independientes de $u[n]$. Finalmente, la varianza resulta en

$$\text{var}(s[n]) = a^2 \text{var}(s[n-1]) + \sigma_u^2. \quad (13.8)$$

Las ecuaciones 13.7 y 13.8 se llaman *ecuaciones de propagación de la media y la varianza*. En el problema de la sección 13.8.4 se explora una ecuación de propagación de la covarianza. Observar que de la ecuación 13.8, la varianza decrece debido a $as[n-1]$ ($|a| < 1$) y crece debido a $u[n]$. En estado estacionario, o con $n \rightarrow \infty$, ambos efectos se compensan conduciendo a la varianza $\sigma_u^2/(1-a^2)$. Para ver esto, tomar $n \rightarrow \infty$ en 13.8, que implica que $\text{var}(s[n]) = \text{var}(s[n-1])$, y resolver.

A continuación se extienden los resultados a un proceso de Gauss-Markov de orden p , dado por

$$s[n] = - \sum_{k=1}^p a[k]s[n-k] + u[n]. \quad (13.9)$$

Se observa que ahora el estado del sistema en el instante n está dado por $\{s[n-1], \dots, s[n-p]\}$, ya que las p muestras previas junto con la entrada actual $u[n]$ determinan la salida. Por lo tanto, se define el *vector de estado* como

$$\mathbf{s}[n-1] = \begin{bmatrix} s[n-p] \\ s[n-p+1] \\ \vdots \\ s[n-1] \end{bmatrix}.$$

Con esta definición, la ecuación 13.9 puede escribirse como

$$\begin{bmatrix} s[n-p+1] \\ s[n-p+2] \\ \vdots \\ s[n-1] \\ s[n] \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ -a[p] & -a[p-1] & -a[p-2] & \dots & -a[1] \end{bmatrix}}_{\mathbf{A}} \begin{bmatrix} s[n-p] \\ s[n-p+1] \\ \vdots \\ s[n-2] \\ s[n-1] \end{bmatrix} + \underbrace{\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}}_{\mathbf{B}} u[n]$$

donde las $(p-1)$ ecuaciones adicionales son solamente identidades. Empleando la definición de vector de estado, se tiene que

$$\mathbf{s}[n] = \mathbf{A}\mathbf{s}[n-1] + \mathbf{B}u[n],$$

donde \mathbf{A} es una matriz $p \times p$ no singular denominada *matriz de transición de estados* y \mathbf{B} es un vector $p \times 1$, conduciendo al modelo en la forma de la ecuación 13.1, en donde la señal vectorial $\mathbf{s}[n]$ se calcula a partir de su valor previo, que es el vector de estado, y la entrada. Esto se denomina modelo de Gauss-Markov vectorial. Un último nivel de generalización permite que la entrada $u[n]$ sea un vector $r \times 1$, de forma tal que, como se muestra en la figura 13.2, se tiene un modelo de señal vectorial que es la salida de un sistema lineal e invariante en el tiempo excitado por una entrada vectorial. En resumen, el modelo

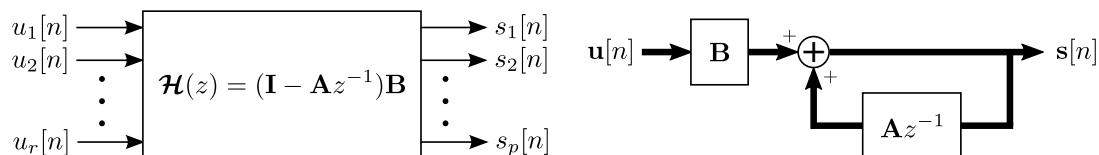


Figura 13.2: Sistema del modelo de Gauss-Markov vectorial. Se muestra el modelo multi-entrada-multi-salida y el modelo vectorial equivalente.

general de Gauss-Markov vectorial toma la forma

$$\mathbf{s}[n] = \mathbf{A}\mathbf{s}[n-1] + \mathbf{B}\mathbf{u}[n] \quad n > 0, \quad (13.10)$$

donde \mathbf{A} y \mathbf{B} son matrices constantes de dimensiones $p \times p$ y $p \times r$ respectivamente, $\mathbf{s}[n]$ es la señal vectorial $p \times 1$ y $\mathbf{u}[n]$ es el ruido de excitación vectorial $r \times 1$. En la figura 13.2, la matriz $\mathbf{H}(z)$ de dimensiones $p \times r$ es la función del sistema, que se obtiene aplicando la transformada z a la ecuación 13.10,

$$\mathbf{S}(z) = \mathbf{A}z^{-1}\mathbf{S}(z) + \mathbf{B}\mathbf{U}(z) \Rightarrow (\mathbf{I} - \mathbf{A}z^{-1})\mathbf{S}(z) = \mathbf{B}\mathbf{U}(z) \Rightarrow \mathbf{S}(z) = (\mathbf{I} - \mathbf{A}z^{-1})^{-1}\mathbf{B}\mathbf{U}(z).$$

Las hipótesis estadísticas del modelo son

1. La entrada $\mathbf{u}[n]$ es una secuencia WGN vectorial, es decir, $\mathbf{u}[n]$ es una secuencia de vectores conjuntamente gaussianos de media nula no correlacionados entre si. Por lo tanto, $E(\mathbf{u}[n]) = \mathbf{0}$ y

$$E(\mathbf{u}[m]\mathbf{u}^T[n]) = \mathbf{0}.$$

Además, la covarianza de $\mathbf{u}[n]$ es

$$E(\mathbf{u}[n]\mathbf{u}^T[n]) = \mathbf{Q},$$

donde \mathbf{Q} es una matriz $r \times r$ definida positiva. Observar que las muestras vectoriales de la entrada son independientes debido a la hipótesis de que son conjuntamente gaussianas.

2. El estado inicial $\mathbf{s}[-1]$ es un vector aleatorio con PDF

$$\mathbf{s}[-1] \sim \mathcal{N}(\boldsymbol{\mu}_s, \mathbf{C}_s)$$

y es independiente de $\mathbf{u}[n]$ para $n \geq 0$.

Para completar el análisis del modelo de Gauss-Markov vectorial, se derivarán las propiedades estadísticas. Los cálculos son una extensión de los realizados para el modelo escalar. Primero se determina una expresión explícita para $\mathbf{s}[n]$. De la ecuación 13.10 se tiene que

$$\begin{aligned} \mathbf{s}[0] &= \mathbf{A}\mathbf{s}[-1] + \mathbf{B}\mathbf{u}[0] \\ \mathbf{s}[1] &= \mathbf{A}\mathbf{s}[0] + \mathbf{B}\mathbf{u}[1] = \mathbf{A}^2\mathbf{s}[-1] + \mathbf{A}\mathbf{B}\mathbf{u}[0] + \mathbf{B}\mathbf{u}[1] \\ \mathbf{s}[2] &= \mathbf{A}\mathbf{s}[1] + \mathbf{B}\mathbf{u}[2] = \mathbf{A}^3\mathbf{s}[-1] + \mathbf{A}^2\mathbf{B}\mathbf{u}[0] + \mathbf{A}\mathbf{B}\mathbf{u}[1] + \mathbf{B}\mathbf{u}[2], \end{aligned}$$

y así sucesivamente, y en general se cumple que

$$\mathbf{s}[n] = \mathbf{A}^{n+1}\mathbf{s}[-1] + \sum_{k=0}^n \mathbf{A}^k \mathbf{B}\mathbf{u}[n-k]. \quad (13.11)$$

donde $\mathbf{A}^0 = \mathbf{I}$. Se observa que $\mathbf{s}[n]$ es una función de la condición inicial y las entradas del ruido de excitación. Como resultado, $\mathbf{s}[n]$ es un proceso aleatorio gaussiano. Queda determinar la media y la covarianza. De la ecuación 13.11, la media es

$$E(\mathbf{s}[n]) = \mathbf{A}^{n+1}E(\mathbf{s}[-1]) = \mathbf{A}^{n+1}\boldsymbol{\mu}_s. \quad (13.12)$$

La covarianza es

$$\begin{aligned} \mathbf{C}_s[m, n] &= E[(\mathbf{s}[m] - E(\mathbf{s}[m]))(\mathbf{s}[n] - E(\mathbf{s}[n]))^T] \\ &= E \left[\left(\mathbf{A}^{m+1}(\mathbf{s}[-1] - \boldsymbol{\mu}_s) + \sum_{k=0}^m \mathbf{A}^k \mathbf{B}\mathbf{u}[m-k] \right) \left(\mathbf{A}^{n+1}(\mathbf{s}[-1] - \boldsymbol{\mu}_s) + \sum_{l=0}^n \mathbf{A}^l \mathbf{B}\mathbf{u}[n-l] \right)^T \right] \\ &= \mathbf{A}^{m+1}E[(\mathbf{s}[-1] - \boldsymbol{\mu}_s)(\mathbf{s}[-1] - \boldsymbol{\mu}_s)^T] \mathbf{A}^{n+1T} + \sum_{k=0}^m \sum_{l=0}^n \mathbf{A}^k \mathbf{B} E(\mathbf{u}[m-k]\mathbf{u}^T[n-l]) \mathbf{B}^T \mathbf{A}^{lT}. \end{aligned}$$

Pero $E[(\mathbf{s}[-1] - \boldsymbol{\mu}_s)(\mathbf{s}[-1] - \boldsymbol{\mu}_s)^T] = \mathbf{C}_s$ y

$$\begin{aligned} E(\mathbf{u}[m-k]\mathbf{u}^T[n-l]) &= \mathbf{Q}\delta[l - (n-m+k)] \\ &= \begin{cases} \mathbf{Q} & l = n-m+k \\ \mathbf{0} & \text{en otro caso.} \end{cases} \end{aligned}$$

Se concluye que para $m \geq n$

$$\mathbf{C}_s[m, n] = \mathbf{A}^{m+1} \mathbf{C}_s \mathbf{A}^{n+1T} + \sum_{k=m-n}^m \mathbf{A}^k \mathbf{B} \mathbf{Q} \mathbf{B}^T \mathbf{A}^{n-m+kT} \quad (13.13)$$

y para $m < n$

$$\mathbf{C}_s[m, n] = \mathbf{C}_s^T[m, n].$$

Finalmente, la matriz de covarianza de $\mathbf{s}[n]$ es

$$\begin{aligned} \mathbf{C}[n] &= \mathbf{C}_s[n, n] \\ &= \mathbf{A}^{n+1} \mathbf{C}_s \mathbf{A}^{n+1T} + \sum_{k=0}^n \mathbf{A}^k \mathbf{B} \mathbf{Q} \mathbf{B}^T \mathbf{A}^{kT}. \end{aligned} \quad (13.14)$$

Además, las ecuaciones de propagación de la media y la covarianza son

$$E(\mathbf{s}[n]) = \mathbf{A} E(\mathbf{s}[n-1]) \quad (13.15)$$

$$\mathbf{C}[n] = \mathbf{A} \mathbf{C}[n-1] \mathbf{A}^T + \mathbf{B} \mathbf{Q} \mathbf{B}^T \quad n \geq 0, \quad (13.16)$$

las cuales se obtienen de la ecuación 13.10, y se deducen en el problema de la sección 13.8.5. Análogamente al caso escalar, la matriz de covarianza decrece debido al término $\mathbf{A} \mathbf{C}[n-1] \mathbf{A}^T$ y crece debido al término $\mathbf{B} \mathbf{Q} \mathbf{B}^T$. En el problema de la sección 13.8.6 se muestra que para que el proceso sea estable, todos los valores propios de \mathbf{A} deben ser de magnitud menor que 1. Propiedades del estado estacionario similares a las del caso escalar pueden obtenerse a partir de las ecuaciones 13.12 y 13.14. Cuando $n \rightarrow \infty$, la media tiende a cero,

$$E(\mathbf{s}[n]) = \mathbf{A}^{n+1} \boldsymbol{\mu}_s \rightarrow \mathbf{0},$$

ya que los valores propios de \mathbf{A} tienen magnitud menor a 1. Además, como se hace en el problema de la sección 13.8.7, de la ecuación 13.14 puede demostrarse que con $n \rightarrow \infty$,

$$\mathbf{A}^{n+1} \mathbf{C}_s \mathbf{A}^{n+1T} \rightarrow \mathbf{0}$$

y por lo tanto

$$\mathbf{C}[n] \rightarrow \mathbf{C} = \sum_{k=0}^{\infty} \mathbf{A}^k \mathbf{B} \mathbf{Q} \mathbf{B}^T \mathbf{A}^{kT}. \quad (13.17)$$

Es interesante notar que la matriz de covarianza en estado estacionario es también la solución de 13.16. Efectivamente, estableciendo $\mathbf{C}[n-1] = \mathbf{C}[n] = \mathbf{C}$ en dicha ecuación, se tiene que

$$\begin{aligned} \sum_{k=0}^{\infty} \mathbf{A}^k \mathbf{B} \mathbf{Q} \mathbf{B}^T \mathbf{A}^{kT} &= \mathbf{A} \left(\sum_{k=0}^{\infty} \mathbf{A}^k \mathbf{B} \mathbf{Q} \mathbf{B}^T \mathbf{A}^{kT} \right) \mathbf{A}^T + \mathbf{B} \mathbf{Q} \mathbf{B}^T \\ &= \sum_{k=0}^{\infty} \mathbf{A}^{k+1} \mathbf{B} \mathbf{Q} \mathbf{B}^T \mathbf{A}^{k+1T} + \mathbf{B} \mathbf{Q} \mathbf{B}^T \\ &\stackrel{(a)}{=} \sum_{l=1}^{\infty} \mathbf{A}^l \mathbf{B} \mathbf{Q} \mathbf{B}^T \mathbf{A}^{lT} + \mathbf{A}^0 \mathbf{B} \mathbf{Q} \mathbf{B}^T \mathbf{A}^{0T} \\ &\stackrel{(b)}{=} \sum_{k=0}^{\infty} \mathbf{A}^k \mathbf{B} \mathbf{Q} \mathbf{B}^T \mathbf{A}^{kT}, \end{aligned}$$

donde en (a) se realizó el cambio de variable $l = k+1$ en la sumatoria y se pre-multiplicó y post-multiplicó el segundo término por $\mathbf{A}^0 = \mathbf{A}^{0T} = \mathbf{I}$, y en (b) se agregó el segundo término a la sumatoria, a la cual se le renombró el índice l a k . Se concluye que la covarianza en estado estacionario cumple que

$$\mathbf{C} = \mathbf{A} \mathbf{C} \mathbf{A}^T + \mathbf{B} \mathbf{Q} \mathbf{B}^T$$

Esta ecuación se llama *ecuación de Lyapunov* y la solución está dada por la ecuación 13.17. A continuación se resume el modelo y sus propiedades.

Teorema: modelo de Gauss-Markov vectorial El modelo de Gauss-Markov para una señal vectorial $\mathbf{s}[n]$ $p \times 1$ es

$$\mathbf{s}[n] = \mathbf{A}\mathbf{s}[n-1] + \mathbf{B}\mathbf{u}[n] \quad n \geq 0.$$

Las matrices \mathbf{A} y \mathbf{B} tienen dimensiones $p \times p$ y $p \times r$ respectivamente y son conocidas, y se asume además que los valores propios de \mathbf{A} son todos de magnitud menor a 1. El ruido vectorial $\mathbf{u}[n]$ de excitación tienen dimensiones $p \times 1$ y se trata de WGN vectorial, es decir, $\mathbf{u}[n] \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$ y los $\mathbf{u}[n]$ son independientes. La condición inicial $\mathbf{s}[-1]$ es un vector aleatorio $p \times 1$ con PDF $\mathbf{s}[-1] \sim \mathcal{N}(\boldsymbol{\mu}_s, \mathbf{C}_s)$ e independiente de los $\mathbf{u}[n]$. Por lo tanto, el proceso es gaussiano con media

$$E(\mathbf{s}[n]) = \mathbf{A}^{n+1} \boldsymbol{\mu}_s$$

y covarianza cruzada para $m \geq n$

$$\begin{aligned} \mathbf{C}_s[m, n] &= E[(\mathbf{s}[m] - E(\mathbf{s}[m]))(\mathbf{s}[n] - E(\mathbf{s}[n]))^T] \\ &= \mathbf{A}^{m+1} \mathbf{C}_s \mathbf{A}^{n+1T} + \sum_{k=m-n}^m \mathbf{A}^k \mathbf{B} \mathbf{Q} \mathbf{B}^T \mathbf{A}^{n-m+kT} \end{aligned}$$

y para $m < n$

$$\mathbf{C}_s[m, n] = \mathbf{C}_s^T[n, m].$$

La matriz de covarianza es

$$\mathbf{C}[n] = \mathbf{C}_s[n, n] = \mathbf{A}^{n+1} \mathbf{C}_s \mathbf{A}^{n+1T} + \sum_{k=0}^n \mathbf{A}^k \mathbf{B} \mathbf{Q} \mathbf{B}^T \mathbf{A}^{kT}.$$

Las ecuaciones de propagación de la media y la covarianza son

$$\begin{aligned} E(\mathbf{s}[n]) &= \mathbf{A}E(\mathbf{s}[n-1]) \\ \mathbf{C}[n] &= \mathbf{A}\mathbf{C}[n-1]\mathbf{A}^T + \mathbf{B}\mathbf{Q}\mathbf{B}^T \quad n \geq 0. \end{aligned}$$

A pesar de que en la definición del modelo de Gauss-Markov se asumió que las matrices \mathbf{A} , \mathbf{B} y \mathbf{Q} no dependen de n , no hay ningún impedimento para que sea así, y pueden deducirse ecuaciones similares para la media y la covarianza. Sin embargo, una diferencia es que puede no alcanzarse nunca un estado estacionario.

13.3. Filtro de Kalman escalar

El modelo de señal de Gauss-Markov escalar presentado en la sección anterior tiene la forma

$$s[n] = as[n-1] + u[n] \quad n \geq 0.$$

A continuación se describe un estimador MMSE secuencial que permite estimar $s[n]$ en base a los datos $\{x[0], \dots, x[n]\}$ a medida que n crece. Esta operación se llama *filtrado*. El enfoque consiste en calcular el estimador $\hat{s}[n]$ en base al estimador de la muestra previa $\hat{s}[n-1]$, por lo que es de naturaleza recursiva. Este es el *filtro de Kalman*. Puede ser aplicado a una señal de Gauss-Markov escalar así como a su extensión vectorial. Adicionalmente, los datos, que en la discusión previa consistieron en la secuencia escalar $\{x[0], \dots, x[n]\}$, pueden extenderse a observaciones vectoriales $\{\mathbf{x}[0], \dots, \mathbf{x}[n]\}$. Los tres niveles de generalización en orden jerárquico son:

1. estado escalar – observación escalar $s[n-1], x[n]$
2. estado vectorial – observación escalar $\mathbf{s}[n-1], x[n]$
3. estado vectorial – observación vectorial $\mathbf{s}[n-1], \mathbf{x}[n]$.

En esta sección se discute el primer caso, dejando los otros para la sección 13.5.

Se considera la ecuación de estados escalar y la ecuación de observación escalar

$$\begin{aligned} s[n] &= as[n-1] + u[n] \\ x[n] &= s[n] + w[n] \end{aligned}$$

donde $u[n]$ es ruido gaussiano de media nula con muestras independientes y $E(u^2[n]) = \sigma_u^2$, $w[n]$ es ruido gaussiano de media nula con muestras independientes y $E(w^2[n]) = \sigma_w^2$. Se asume además que $s[-1]$, $u[n]$ y $w[n]$ son independientes. Por último, se asume que $s[-1] \sim \mathcal{N}(\mu_s, \sigma_s^2)$. El ruido $w[n]$ difiere de ser ruido blanco solo en que se permite que la varianza cambien en el tiempo. Para simplificar la deducción, se asume que $\mu_s = 0$, de forma de que de acuerdo a la ecuación 13.3 $E(s[n]) = 0$ para $n \geq 0$. Mas adelante se considerará el caso en que el valor inicial de la media no es nulo. Se desea estimar $s[n]$ en base a las observaciones $\{\mathbf{x}[0], \dots, \mathbf{x}[n]\}$, o filtrar $x[n]$ para producir $\hat{s}[n]$. De forma mas general, el estimador $\hat{s}[n]$ en base a las observaciones $\{\mathbf{x}[0], \dots, \mathbf{x}[m]\}$ se denota como $\hat{s}[n|m]$. El criterio de optimalidad es el MSE bayesiano mínimo, o

$$E[(s[n] - \hat{s}[n|n])^2],$$

donde la esperanza se toma respecto a $p(x[0], \dots, x[n], s[n])$. Pero el estimador MMSE es la media de la PDF a posteriori (ver el capítulo 10), o

$$\hat{s}[n|n] = E(s[n]|x[0], \dots, x[n]). \quad (13.18)$$

De la ecuación 10.14 de la media condicional de variables aleatorias conjuntamente gaussianas, en el caso de media nula, se cumple que

$$\hat{s}[n|n] = \mathbf{C}_{\theta x} \mathbf{C}_{xx}^{-1} \mathbf{x}, \quad (13.19)$$

ya que $\theta = s[n]$ y $\mathbf{x} = [x[0] \dots x[n]]^T$ son conjuntamente gaussianas y de media nula. Debido a la hipótesis de estadística gaussiana de la señal y el ruido, el estimador MMSE es lineal, por lo que es algebraicamente idéntico al estimador LMMSE (ver las ecuaciones 10.14 y 12.6). Las propiedades algebraicas permiten emplear el enfoque de espacio vectorial para encontrar el estimador, de forma análoga a la deducción del estimador LMMSE secuencial realizada en la sección 12.4. La restricción implícita de linealidad no quita generalidad debido a que se sabe que el estimador óptimo es lineal. Además, si las hipótesis de estadística gaussiana no se cumplen, el estimador continúa siendo válido, pero solo puede decirse que se trata del estimador LMMSE óptimo. Retornando a la estimación secuencial de la ecuación 13.19, se observa que si $x[n]$ no estuviera correlacionado con $\{\mathbf{x}[0], \dots, \mathbf{x}[n-1]\}$, de la ecuación 13.18 y el principio de ortogonalidad (ver el ejemplo de la sección 12.2.1) se cumpliría que

$$\begin{aligned} \hat{s}[n|n] &= E(s[n]|x[0], \dots, x[n-1]) + E(s[n]|x[n]) \\ &= \hat{s}[n|n-1] + E(s[n]|x[n]), \end{aligned}$$

que tiene la forma secuencial buscada. Sin embargo, los $x[n]$ son correlacionados debido a su dependencia con $s[n]$. De la discusión de la sección 12.4 del estimador LMMSE secuencial, se puede emplear la interpretación de espacio vectorial para determinar la corrección del viejo estimador $\hat{s}[n|n-1]$ debido a la observación de $x[n]$. Antes de hacerlo, se consideran dos propiedades del estimador MMSE que serán empleadas en la deducción.

1. El estimador MMSE de θ en base a dos vectores de datos no correlacionados, asumiendo estadística conjuntamente gaussiana, es

$$\begin{aligned} \hat{\theta} &= E(\theta|\mathbf{x}_1, \mathbf{x}_2) \\ &= E(\theta|\mathbf{x}_1) + E(\theta|\mathbf{x}_2), \end{aligned}$$

si θ es de media nula (ver la ecuación 11.4).

2. El estimador MMSE es aditivo en el sentido en que si $\theta = \theta_1 + \theta_2$, se cumple que

$$\begin{aligned} \hat{\theta} &= E(\theta|\mathbf{x}) \\ &= E(\theta_1 + \theta_2|\mathbf{x}) \\ &= E(\theta_1|\mathbf{x}) + E(\theta_2|\mathbf{x}), \end{aligned}$$

propiedad que se obtiene inmediatamente de la linealidad de la esperanza.

Con estas propiedades se comienza la deducción de las ecuaciones 13.26–13.30. Sea $\mathbf{X}[n] = [x[0] \dots x[n]]^T$ (se emplea la notación \mathbf{X} porque $\mathbf{x}[n]$ se empleó para denotar observaciones vectoriales) y sea $\tilde{x}[n]$ la innovación. Recordar que la innovación es la parte de $x[n]$ que no está correlacionada con las muestras previas $\{x[0], \dots, x[n-1]\}$ (ver la sección 12.4) o

$$\tilde{x}[n] = x[n] - \hat{x}[n|n-1].$$

Esto es porque $\hat{x}[n|n-1]$ es el estimador MMSE de $x[n]$ en base a $\{x[0], \dots, x[n-1]\}$ y por el principio de ortogonalidad, el error de estimación $\tilde{x}[n]$ es ortogonal (no correlacionado) con los datos. El conjunto de datos $\mathbf{X}[n]$, $\tilde{x}[n]$ es equivalente al conjunto de datos original, ya que $x[n]$ puede recuperarse del primer conjunto como

$$\begin{aligned} x[n] &= \tilde{x}[n] + \hat{x}[n|n-1] \\ &= \tilde{x}[n] + \sum_{k=0}^{n-1} a_k x[k] \end{aligned}$$

donde los a_k son los coeficientes óptimos del estimador MMSE de $x[n]$ en base a $\{x[0], \dots, x[n-1]\}$. Por lo tanto, la ecuación 13.19 puede escribirse como

$$\hat{s}[n|n] = E(s[n]|\mathbf{X}[n-1], \tilde{x}[n]),$$

y como $\mathbf{X}[n-1]$ y $\tilde{x}[n]$ son no correlacionados, de la propiedad 1 se tiene que

$$\hat{s}[n|n] = E(s[n]|\mathbf{X}[n-1]) + E(s[n]|\tilde{x}[n]).$$

Pero $E(s[n]|\mathbf{X}[n-1])$ es la predicción de $s[n]$ en base a los datos previos, y se denota como $\hat{s}[n|n-1]$. De la ecuación 13.1 y la propiedad 2, la predicción es

$$\begin{aligned} \hat{s}[n|n-1] &= E(s[n]|\mathbf{X}[n-1]) \\ &= E(as[n-1] + u[n]|\mathbf{X}[n-1]) \\ &\stackrel{(a)}{=} aE(s[n-1]|\mathbf{X}[n-1]) \\ &= a\hat{s}[n-1|n-1], \end{aligned}$$

donde en (a) se tuvo en cuenta que $E(u[n]|\mathbf{X}[n-1]) = 0$. Esto es porque

$$E(u[n]|\mathbf{X}[n-1]) = E(u[n]) = 0$$

debido a que $u[n]$ es independiente de $\{x[0], \dots, x[n-1]\}$, ya que $u[n]$ es independiente de todos los $w[n]$, y de la ecuación 13.2, $s[0], \dots, s[n-1]$ son combinación lineal de las variables aleatorias $\{u[0], \dots, u[n-1], s[-1]\}$, que también son independientes de $u[n]$. Por lo tanto, se tiene que

$$\hat{s}[n|n] = \hat{s}[n|n-1] + E(s[n]|\tilde{x}[n]) \quad (13.20)$$

donde

$$\hat{s}[n|n-1] = a\hat{s}[n-1|n-1].$$

Para determinar $E(s[n]|\tilde{x}[n])$, se observa que se trata del estimador MMSE de $s[n]$ en base a $\tilde{x}[n]$. Por lo tanto, por la hipótesis de estadística gaussiana, es lineal, y además como se trata de variables aleatorias escalares y $s[n]$ es de media nula, toma la forma (ver la ecuación 12.28)

$$\begin{aligned} E(s[n]|\tilde{x}[n]) &= \frac{E(s[n]\tilde{x}[n])}{E(\tilde{x}^2[n])} \tilde{x}[n] \\ &= K[n]\tilde{x}[n] \\ &= K[n](x[n] - \hat{x}[n|n-1]), \end{aligned}$$

donde

$$K[n] = \frac{E(s[n]\tilde{x}[n])}{E(s^2[n])}. \quad (13.21)$$

Pero $x[n] = s[n] + w[n]$, por lo que por la propiedad 2,

$$\begin{aligned}\hat{x}[n|n-1] &= \hat{s}[n|n-1] + \hat{w}[n|n-1] \\ &= \hat{s}[n|n-1],\end{aligned}$$

ya que $\hat{w}[n|n-1] = 0$ debido a que $w[n]$ es independiente de $\{x[0], \dots, x[n-1]\}$. Así,

$$E(s[n]|\tilde{x}[n]) = K[n](x[n] - \hat{s}[n|n-1]),$$

y la ecuación 13.20 queda

$$\hat{s}[n|n] = \hat{s}[n|n-1] + K[n](x[n] - \hat{s}[n|n-1]) \quad (13.22)$$

donde

$$\hat{s}[n|n-1] = a\hat{s}[n-1|n-1]. \quad (13.23)$$

Falta solo determinar el factor de ganancia $K[n]$. De la ecuación 13.21 el factor de ganancia es

$$K[n] = \frac{E[s[n](x[n] - \hat{s}[n|n-1])]}{E[(x[n] - \hat{s}[n|n-1])^2]},$$

donde se empleó nuevamente que $\hat{x}[n|n-1] = \hat{s}[n|n-1]$. Para evaluar esta expresión se necesitan los siguientes resultados:

1.

$$E[s[n](x[n] - \hat{s}[n|n-1])] = E[(s[n] - \hat{s}[n|n-1])(x[n] - \hat{s}[n|n-1])].$$

2.

$$E[w[n](s[n] - \hat{s}[n|n-1])] = 0.$$

El primer resultado proviene de que la innovación $\tilde{x}[n]$ no está correlacionada con las muestras pasadas de datos $\{x[0], \dots, x[n-1]\}$ y por lo tanto, $x[n] - \hat{s}[n|n-1]$ no está correlacionada con $\hat{s}[n|n-1]$. Efectivamente, la innovación

$$\tilde{x}[n] = x[n] - \hat{x}[n|n-1] = x[n] - \hat{s}[n|n-1] \quad (13.24)$$

no está correlacionada con las muestras pasadas de los datos y $\hat{s}[n|n-1]$ es una combinación lineal de las muestras pasadas $\{x[0], \dots, x[n-1]\}$ de los datos, concluyendo que $E[\hat{s}[n|n-1](x[n] - \hat{s}[n|n-1])] = 0$. El segundo resultado proviene de que $s[n]$ y $w[n]$ son no correlacionados y $w[n]$ no está correlacionado con las muestras pasadas de los datos. Partiendo del primer resultado, la ganancia se puede escribir como

$$\begin{aligned}K[n] &= \frac{E[(s[n] - \hat{s}[n|n-1])(x[n] - \hat{s}[n|n-1])]}{E[(x[n] - \hat{s}[n|n-1])^2]} \\ &\stackrel{(a)}{=} \frac{E[(s[n] - \hat{s}[n|n-1])(s[n] + w[n] - \hat{s}[n|n-1])]}{E[(s[n] + w[n] - \hat{s}[n|n-1])^2]} \\ &= \frac{E[(s[n] - \hat{s}[n|n-1])^2] + E[(s[n] - \hat{s}[n|n-1])w[n]]}{E(w^2[n]) + E[(s[n] - \hat{s}[n|n-1])^2] + 2E[(s[n] - \hat{s}[n|n-1])w[n]]} \\ &\stackrel{(b)}{=} \frac{E[(s[n] - \hat{s}[n|n-1])^2]}{\sigma_n^2 + E[(s[n] - \hat{s}[n|n-1])^2]},\end{aligned} \quad (13.25)$$

donde en (a) se consideró que $x[n] = s[n] + w[n]$ y en (b) se empleó el segundo resultado. Pero el numerador es el MSE mínimo que se obtiene al estimar $s[n]$ en base a los datos previos o el error mínimo de predicción de un paso. Esto se denotará como $M[n|n-1]$, de forma que

$$K[n] = \frac{M[n|n-1]}{\sigma_n^2 + M[n|n-1]}.$$

Para evaluar la ganancia se necesita una expresión del error mínimo de predicción. Éste es

$$\begin{aligned}
M[n|n-1] &= E[(s[n] - \hat{s}[n|n-1])^2] \\
&\stackrel{(a)}{=} E[(as[n-1] + u[n] - a\hat{s}[n-1|n-1])^2] \\
&= E[(a(s[n-1] - \hat{s}[n-1|n-1]) + u[n])^2] \\
&\stackrel{(b)}{=} a^2 E[(s[n-1] - \hat{s}[n-1|n-1])^2] + E(u^2[n]),
\end{aligned}$$

donde en (a) se emplearon las ecuaciones 13.1 y 13.23 y en (b) se consideró que

$$E[(s[n-1] - \hat{s}[n-1|n-1])u[n]] = 0$$

ya que $s[n-1]$ depende de $\{u[0], \dots, u[n-1], s[-1]\}$, que son independientes de $u[n]$ y $\hat{s}[n-1|n-1]$ depende de las muestras pasadas de datos, que también son independientes de $u[n]$. Se concluye que

$$M[n|n-1] = a^2 M[n-1|n-1] + \sigma_u^2.$$

Finalmente, se requiere una recursión para $M[n|n]$. Partiendo de la definición se tiene que

$$\begin{aligned}
M[n|n] &= E[(s[n] - \hat{s}[n|n])^2] \\
&\stackrel{(a)}{=} E[(s[n] - \hat{s}[n|n-1] - K[n](x[n] - \hat{s}[n|n-1]))^2] \\
&= E[(s[n] - \hat{s}[n|n-1])^2] - 2K[n]E[(s[n] - \hat{s}[n|n-1])(x[n] - \hat{s}[n|n-1])] \\
&\quad + K^2[n]E[(x[n] - \hat{s}[n|n-1])^2].
\end{aligned}$$

donde en (a) se empleó la ecuación 13.22. Observando que en la última igualdad el primer sumando es $M[n|n-1]$, la esperanza del segundo sumando es el numerador de $K[n]$, como se ve en la ecuación 13.25, que también resulta en $M[n|n-1]$ y la esperanza del tercer sumando es el denominador de $K[n]$, como se ve en la ecuación 13.25, y es $M[n|n-1]/K[n]$, resulta en

$$\begin{aligned}
M[n|n] &= M[n|n-1] - 2K[n]M[n|n-1] + K[n]M[n|n-1] \\
&= (1 - K[n])M[n|n-1].
\end{aligned}$$

Esto completa la derivación del filtro de Kalman con estado escalar–observación escalar. Si bien la deducción es tediosa, las ecuaciones finales son simples e intuitivas, y se resumen a continuación. Para $n \geq 0$

Predicción:

$$\hat{s}[n|n-1] = a\hat{s}[n-1|n-1]. \quad (13.26)$$

Predicción del MSE mínimo:

$$M[n|n-1] = a^2 M[n-1|n-1] + \sigma_u^2. \quad (13.27)$$

Ganancia de Kalman:

$$K[n] = \frac{M[n|n-1]}{\sigma_n^2 + M[n|n-1]}. \quad (13.28)$$

Corrección:

$$\hat{s}[n|n] = \hat{s}[n|n-1] + K[n](x[n] - \hat{s}[n|n-1]). \quad (13.29)$$

MSE mínimo:

$$M[n|n] = (1 - K[n])M[n|n-1]. \quad (13.30)$$

Si bien la deducción se realizó para $\mu_s = 0$ de forma que $E(s[n]) = 0$, las mismas ecuaciones resultan si $\mu_s \neq 0$, como se muestra para el caso vectorial en la sección 13.5. Por lo tanto, para inicializar las ecuaciones se usa $\hat{s}[-1| -1] = E(s[-1]) = \mu_s$ y $M[-1| -1] = \sigma_s^2$, que es la estimación de $s[-1]$ sin observar datos. En la figura 13.3 se muestra un diagrama del filtro de Kalman. Es interesante notar que el modelo dinámico de la señal es una parte integral del estimador. Observar que la salida del bloque de ganancia puede verse como un estimador $\hat{u}[n]$ de $u[n]$. De las ecuaciones 13.26 y 13.29 el estimador de la señal es

$$\hat{s}[n|n] = a\hat{s}[n-1|n-1] + \hat{u}[n].$$

donde $\hat{u}[n] = K[n](x[n] - \hat{s}[n|n-1])$. Si este estimador es aproximadamente $u[n]$, se cumple que $\hat{s}[n|n]$, como se quiere.

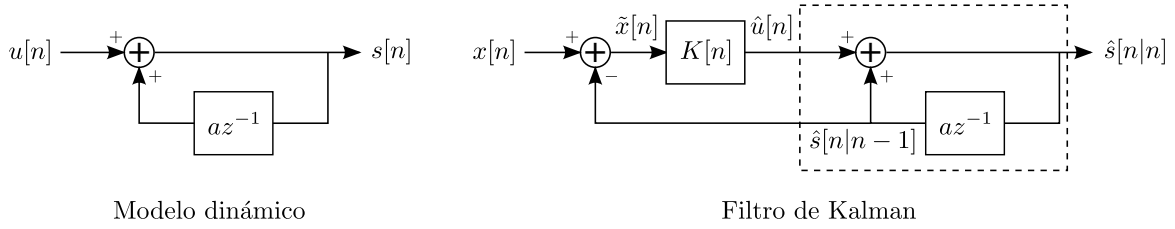


Figura 13.3: Filtro de Kalman con estado escalar–observación escalar y su relación con el modelo dinámico.

Debe notarse que se obtiene el mismo conjunto de ecuaciones si $E(s[-1]) \neq 0$. En ese caso, la media de $s[n]$ no es nula, ya que $E(s[n]) = a^{n+1}E(s[-1])$ (ver la ecuación 13.3). Esta extensión se explora en el problema de la sección 13.8.13. Recordar que el motivo de la hipótesis de media nula fue para emplear el principio de ortogonalidad. A continuación se discuten algunas propiedades del filtro de Kalman.

1. El filtro de Kalman extiende al estimador MMSE secuencial visto en la sección 12.4.1 para el caso en que el parámetro desconocido evoluciona en el tiempo de acuerdo al modelo dinámico. En la sección 12.4.1 se dedujeron las ecuaciones del estimador LMMSE secuencial (12.37–12.39), que tienen forma idéntica al estimador MMSE secuencial con estadística gaussiana. En particular, para el caso escalar con $h[n] = 1$ y reemplazando $\theta[n]$ por $s[n]$, esas ecuaciones quedan

$$\hat{s}[n] = \hat{s}[n-1] + K[n](x[n] - \hat{s}[n-1]) \quad (13.31)$$

$$K[n] = \frac{M[n-1]}{\sigma_n^2 + M[n-1]} \quad (13.32)$$

$$M[n] = (1 - K[n])M[n-1]. \quad (13.33)$$

El filtro de Kalman se reduce a estas ecuaciones cuando el parámetro a estimar no evoluciona en el tiempo, situación que implica asumir que el ruido de excitación es nulo, o $\sigma_u^2 = 0$, y $a = 1$. La ecuación de estado es, de 13.1, $s[n] = s[n-1]$, o explícitamente, $s[n] = s[-1] = \theta$. El parámetro a estimar es una constante, que se modela como la realización de una variable aleatoria. Por lo tanto, la predicción es $\hat{s}[n|n-1] = \hat{s}[n-1|n-1]$ con MSE mínimo $M[n|n-1] = M[n-1|n-1]$, lo que indica que se puede omitir la etapa de predicción del filtro de Kalman. La predicción es el último estimador de $s[n]$. Las etapas de corrección se reducen a las ecuaciones 13.31–13.33 definiendo $\hat{s}[n|n] = \hat{s}[n]$, $\hat{s}[n|n-1] = \hat{s}[n-1]$, $M[n|n] = M[n]$ y $M[n|n-1] = M[n-1]$.

2. No se requiere inversión de matrices, en comparación con la estimación por lotes de $\theta = s[n]$ como

$$\hat{\theta} = \mathbf{C}_{\theta x} \mathbf{C}_{xx}^{-1} \mathbf{x},$$

donde se requiere la inversión de \mathbf{C}_{xx} para cada muestra de $s[n]$ a estimar. Además, la dimensión de la matriz es $(n+1) \times (n+1)$, creciente con n .

3. El filtro de Kalman es un filtro lineal variante en el tiempo. Efectivamente, de las ecuaciones 13.26 y 13.29, se tiene que

$$\begin{aligned}\hat{s}[n|n] &= a\hat{s}[n-1|n-1] + K[n](x[n] - a\hat{s}[n-1|n-1]) \\ &= a(1 - K[n])\hat{s}[n-1|n-1] + K[n]x[n],\end{aligned}$$

que es un filtro recursivo de primer orden con coeficientes variables como se muestra en la figura 13.4.

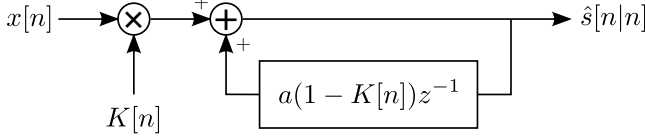


Figura 13.4: Filtro de Kalman con un filtro lineal variante en el tiempo.

4. El filtro de Kalman provee su propia medida de desempeño. Con la ecuación 13.30 se obtiene el MSE mínimo como parte integral del estimador. Además, la medida del error puede realizarse antes de recolectar los datos. Esto es porque $M[n|n]$ depende de las ecuaciones 13.27 y 13.28, que no dependen de los datos. Mas adelante se verá que en el filtro de Kalman extendido, el MSE debe calcularse en línea.
5. La etapa de predicción incrementa el error, mientras que la etapa de corrección lo decrementa. Existe una interacción interesante entre el MSE mínimo de predicción y el MSE mínimo. Si con $n \rightarrow \infty$ se alcanza un estado estacionario, $M[n|n]$ y $M[n-1|n-1]$ se hacen constantes, con $M[n|n-1] > M[n-1|n-1]$, como se muestra en el problema de la sección 13.8.14. Por lo tanto, el error crece luego de la etapa de predicción. Cuando se obtiene una nueva muestra de datos, se corrige el estimador, y el error decrece de acuerdo a la ecuación 13.30, ya que $K[n] < 1$. En la figura 13.5 se muestra un ejemplo en el cual $a = 0.99$, $\sigma_u^2 = 0.1$, $\sigma_n^2 = 0.9^{n+1}$ y $M[-1| -1] = 1$.

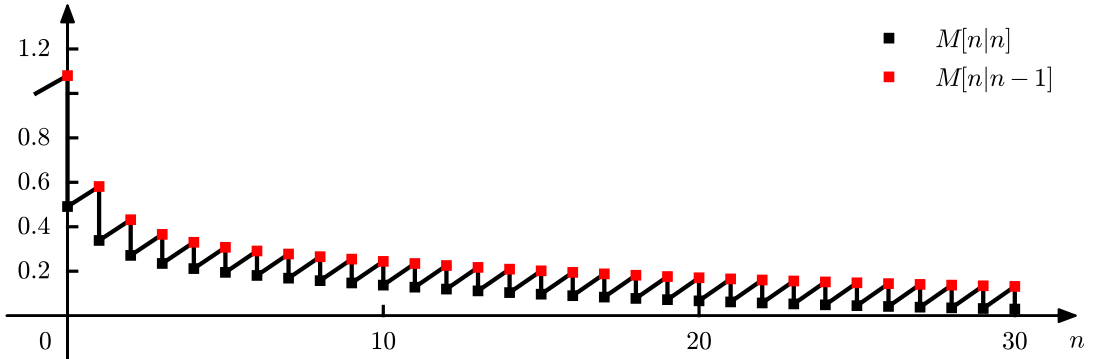


Figura 13.5: Predicción $M[n|n-1]$ y corrección $M[n|n]$ del MSE mínimo.

6. La predicción es una parte integral del filtro de Kalman. La ecuación 13.26 indica que para encontrar el mejor estimador filtrado de $s[n]$ se emplean predicciones. Se puede obtener la mejor predicción de un paso de $s[n]$ en base a $\{x[0], \dots, x[n-1]\}$ con la ecuación 13.26. Si se desea encontrar la mejor predicción de dos pasos, puede hacerse fácilmente observando que se trata del mejor estimador de $s[n+1]$ en base a $\{x[0], \dots, x[n-1]\}$. Para encontrarlo, se toma $\sigma_n^2 \rightarrow \infty$, lo que implica que $x[n]$ es una muestra tan ruidosa que el filtro de Kalman no la va a usar. De esta forma, $\hat{s}[n+1|n-1]$ es la predicción óptima de dos pasos. Para evaluarla, se observa que en el paso n la predicción es (ecuación 13.26)

$$\hat{s}[n|n-1] = a\hat{s}[n-1|n-1].$$

Al tomar $\sigma_n^2 \rightarrow \infty$, de la ecuación 13.28, se cumple que $K[n] \rightarrow 0$ y de la ecuación 13.29, el estimador es

$$\hat{s}[n|n] = \hat{s}[n|n-1] = a\hat{s}[n-1|n-1].$$

En la paso $n+1$, la predicción es

$$\hat{s}[n+1|n] = a\hat{s}[n|n] = a^2\hat{s}[n-1|n-1],$$

pero como no se empleó la muestra $x[n]$ por ser ruidosa, $\hat{s}[n+1|n] = \hat{s}[n+1|n-1]$, resultando en que el predictor óptimo de dos pasos es

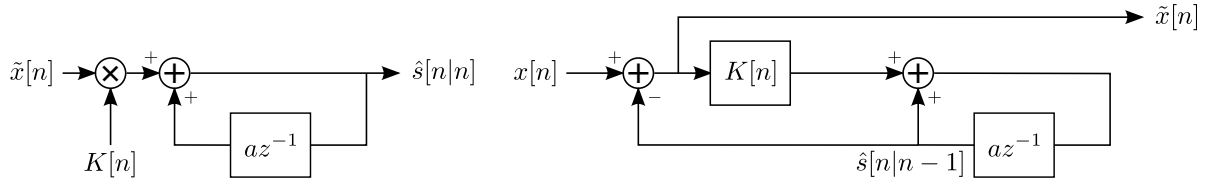
$$\hat{s}[n+1|n-1] = a^2\hat{s}[n-1|n-1].$$

Este resultado se puede generalizar para el predictor de paso l , como se hace en el problema de la sección 13.8.15.

7. El filtro de Kalman es excitado por la secuencia no correlacionada de innovación y en estado estacionario puede verse como un filtro de blanqueado. Observar que de las ecuaciones 13.26 y 13.29 se tiene que

$$\hat{s}[n|n] = a\hat{s}[n-1|n-1] + K[n](x[n] - \hat{s}[n|n-1])$$

de forma que la entrada del filtro de Kalman es la secuencia de innovación $\tilde{x}[n] = x[n] - \hat{x}[n|n-1] = x[n] - \hat{s}[n|n-1]$ (ver la ecuación 13.24), como se muestra en la figura 13.6. De la discusión basada en



Filtro de Kalman excitado por la innovación

Intepretación de blanqueado del filtro de Kalman

Figura 13.6: Filtro de Kalman excitado por la secuencia de innovación e interpretación del filtro de Kalman como filtro de blanqueado cuando se consideran a los datos como entrada y a la secuencia de innovación como salida.

el enfoque espacio vectorial, se sabe que $\tilde{x}[n]$ no está correlacionada con $\{x[0], \dots, x[n-1]\}$, lo que se traduce en una secuencia de variables aleatorias no correlacionadas, como se muestra en la figura 13.7. Alternativamente, se se ve a $\tilde{x}[n]$ como la salida y el filtro alcanza un estado estacionario, se trata de un filtro lineal invariante de blanqueado, como se muestra en la figura 13.6. Esto se discute con mas detalle en la sección siguiente.

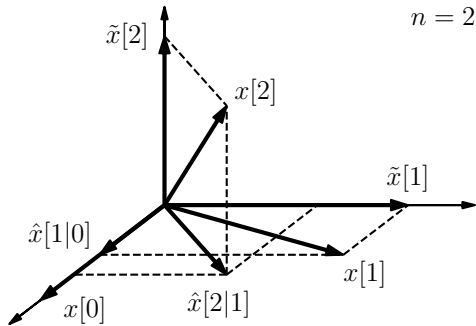


Figura 13.7: Ortogonalidad de la secuencia de innovación $\tilde{x}[n]$. La estimación $\hat{x}[1|0]$ de $x[1]$ en base a $x[0]$ es la proyección de $x[1]$ sobre $x[0]$, por lo que la innovación $\tilde{x}[1] = x[1] - \hat{x}[1|0]$ es ortogonal a $x[0]$. De la misma forma, la estimación $\hat{x}[2|1]$ de $x[2]$ en base a $\{x[0], x[1]\}$ es la proyección de $x[2]$ sobre el espacio vectorial generado por $\{x[0], x[1]\}$, por lo que la innovación $\tilde{x}[2] = x[2] - \hat{x}[2|1]$ es ortogonal a $x[0]$ y $\tilde{x}[1]$.

8. El filtro de Kalman es óptimo en el sentido de que minimiza el MSE bayesiano para cada estimación $\hat{s}[n]$. En el caso en que no se cumpla la hipótesis gaussiana, aún es el estimador MMSE lineal óptimo, como se describió en el capítulo 12.

Estas propiedades son válidas también para el caso de estado vectorial, a excepción de la propiedad 2 si las observaciones son vectoriales.

13.4. Comparación entre los filtros de Wiener y de Kalman

El filtro de Wiener causal de largo infinito descrito en la sección 12.5 produce un estimador de $s[n]$ como la salida de un filtro lineal invariante en el tiempo, o

$$\hat{s}[n] = \sum_{k=0}^{\infty} h[k]x[n-k].$$

El estimador de $s[n]$ es en base a la muestra actual e infinitas muestras pasadas de los datos. Para determinar analíticamente la respuesta al impulso $h[k]$ del filtro se necesita asumir la señal $s[n]$ y el ruido $w[n]$ son procesos WSS, de forma tal que las ecuaciones de Wiener-Hopf puedan resolverse, como se hace en el problema de la sección 12.6.16. En la formulación del filtro de Kalman, no es necesario imponer que la señal y el ruido son WSS. La varianza de $w[n]$ puede cambiar con n , y $s[n]$ es WSS solo cuando $n \rightarrow \infty$. Adicionalmente, el filtro de Kalman produce estimadores en base a las muestras de datos desde 0 a n , y no en base a las infinitas muestras pasadas, como se asume en el filtro de Wiener de largo infinito. Sin embargo, ambos filtros son iguales cuando $n \rightarrow \infty$ si $\sigma_n^2 = \sigma^2$. Esto ocurre debido a que $s[n]$ se aproxima a un proceso estacionario AR(1), como se explicó en la sección 13.2, y el estimador es en base a la muestra actual e infinitas muestras pasadas de los datos. Como el filtro de Kalman se aproxima a un filtro lineal invariante en el tiempo, sean $K[n] \rightarrow K[\infty]$, $M[n|n] \rightarrow M[\infty]$ y $M[n|n-1] \rightarrow M_p[\infty]$, donde $M_p[\infty]$ es el error de predicción de un paso en estado estacionario. Para encontrar el filtro de Kalman en estado estacionario, se necesita primero encontrar $M[\infty]$. Partiendo de la ecuación 13.30, se tiene que

$$\begin{aligned} M[\infty] &= (1 - K[\infty])M_p[\infty] \\ &\stackrel{(a)}{=} \left(1 - \frac{M_p[\infty]}{\sigma^2 + M_p[\infty]}\right) M_p[\infty] \\ &= \frac{\sigma^2 M_p[\infty]}{\sigma^2 + M_p[\infty]} \\ &\stackrel{(b)}{=} \frac{\sigma^2(a^2 M[\infty] + \sigma_u^2)}{a^2 M[\infty] + \sigma_u^2 + \sigma^2}, \end{aligned} \quad (13.34)$$

donde en (a) se substituyó $K[\infty]$ empanando la ecuación 13.28 y en (b) se substituyó $M_p[\infty]$ empleando la ecuación 13.27. Esta ecuación, que debe resolverse para $M[\infty]$, se llama *ecuación de Riccati* de estado estacionario, y se observa que es cuadrática en $M[\infty]$. Luego de obtener $M[\infty]$, se obtiene $M_p[\infty]$ con la ecuación 13.27 y $K[\infty]$ con la ecuación 13.28. De esta forma, el filtro de Kalman en estado estacionario toma la forma de un filtro recursivo de primer orden como se muestra en la figura 13.4, pero con $K[n]$ reemplazando por $K[\infty]$. Un fomra sencilla de resolver numéricamente la ecuación de Riccati, es ejecutar el filtro de Kalman hasta que converja, como se hizo para generar la figura 13.5. El filtro en estado estacionario, de las ecuaciones 13.29 y 13.26, tendrá la forma

$$\begin{aligned} \hat{s}[n|n] &= a\hat{s}[n-1|n-1] + K[\infty](x[n] - a\hat{s}[n-1|n-1]) \\ &= a(1 - K[\infty])\hat{s}[n-1|n-1] + K[\infty]x[n]. \end{aligned} \quad (13.35)$$

Aplicando la transformada z , se tiene que

$$\hat{\mathcal{S}}(z) = a(1 - K[\infty])z^{-1}\hat{\mathcal{S}}(z) + K[\infty]\mathcal{X}(z),$$

por lo que su función de transferencia en estado estacionario es

$$\mathcal{H}_{\infty}(z) = \frac{\hat{\mathcal{S}}(z)}{\mathcal{X}(z)} = \frac{K[\infty]}{1 - a(1 - K[\infty])z^{-1}}. \quad (13.36)$$

Como ejemplo, si $a = 0.9$, $\sigma_u^2 = 1$ y $\sigma^2 = 1$, resolviendo el polinomio de segundo grado dado por la ecuación 13.34 se obtiene que $M[\infty] = 0.5974$. Luego, de la ecuación 13.27, $M_p[\infty] = 1.4839$, y de la ecuación 13.28, $K[\infty] = 0.5974$. La respuesta en frecuencia es

$$\begin{aligned} H_\infty(f) &= \mathcal{H}_\infty(e^{j2\pi f}) \\ &= \frac{K[\infty]}{1 - a(1 - K[\infty])e^{-j2\pi f}}, \end{aligned}$$

y se muestra en la figura 13.8 para los valores indicados previamente, junto con la PSD de la señal en estado estacionario, que por tratarse de un proceso AR(1), es

$$P_{ss}(f) = \frac{\sigma_u^2}{|1 - ae^{-j2\pi f}|^2}.$$

Se obtendrían los mismos resultados si se resolvieran las ecuaciones de Wiener-Hopf para el filtro causal de Wiener de largo infinito.

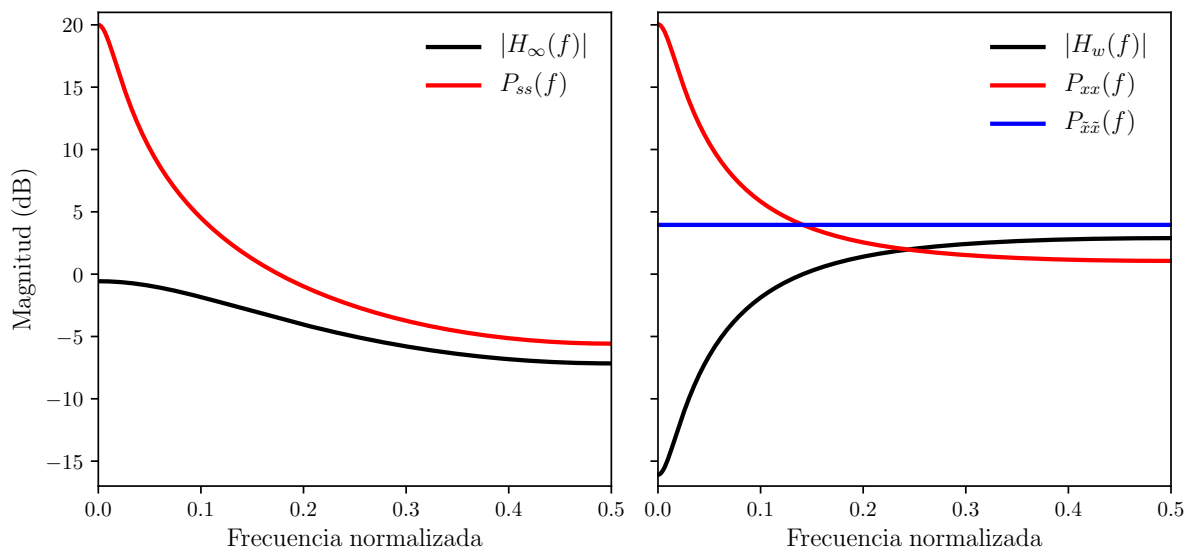


Figura 13.8: PSD de la señal y respuesta en magnitud del filtro de Kalman en estado estacionario, y propiedad de blanqueo del filtro de Kalman en estado estacionario.

Finalmente, la propiedad de blanqueo discutida en la sección 13.3 puede ser verificada para este ejemplo. La innovación es

$$\begin{aligned} \tilde{x}[n] &= x[n] - \hat{s}[n|n-1] \\ &= x[n] - a\hat{s}[n-1|n-1]. \end{aligned}$$

Aplicando la transformada z , se tiene que

$$\begin{aligned} \tilde{\mathcal{X}}(z) &= \mathcal{X}(z) - az^{-1}\hat{\mathcal{S}}(z) \\ &\stackrel{(a)}{=} \mathcal{X}(z) - az^{-1}\mathcal{X}(z)\mathcal{H}_\infty(z) \\ &= \mathcal{X}(z)[1 - az^{-1}\mathcal{H}_\infty(z)], \end{aligned}$$

donde en (a) se tuvo en cuenta que en estado estacionario, $\hat{s}[n|n]$ es la salida del filtro con función de transferencia $\mathcal{H}_\infty(z)$ excitado por $x[n]$ y por lo tanto, $\hat{\mathcal{S}}(z) = \mathcal{X}(z)\mathcal{H}_\infty(z)$ (ver la ecuación 13.36). Por lo

tanto, la función de transferencia que relaciona la entrada $x[n]$ con la salida $\tilde{x}[n]$ es

$$\begin{aligned}\mathcal{H}_w(z) &= \frac{\tilde{\mathcal{X}}(z)}{\mathcal{X}(z)} \\ &= 1 - az^{-1}\mathcal{H}_\infty(z) \\ &\stackrel{(a)}{=} 1 - \frac{az^{-1}K[\infty]}{1 - a(1 - K[\infty])z^{-1}} \\ &= \frac{1 - az^{-1}}{1 - a(1 - K[\infty])z^{-1}},\end{aligned}$$

donde en (a) se empleó la ecuación 13.36. Por lo tanto, la respuesta en frecuencia del filtro de blanqueado es

$$H_w(f) = \frac{1 - ae^{-j2\pi f}}{1 - a(1 - K[\infty])e^{-j2\pi f}}.$$

Por otro lado, la PSD de $x[n]$ es

$$\begin{aligned}P_{xx}(f) &= P_{ss}(f) + \sigma^2 \\ &= \frac{\sigma_u^2}{|1 - ae^{-j2\pi f}|^2} + \sigma^2 \\ &= \frac{\sigma_u^2 + \sigma^2|1 - ae^{-j2\pi f}|^2}{|1 - ae^{-j2\pi f}|^2}.\end{aligned}$$

Tanto la magnitud de la respuesta en frecuencia del filtro de blanqueado como la PSD de $x[n]$ se muestran en la figura 13.8, junto con la PSD de la salida $\tilde{x}[n]$, que es

$$P_{\tilde{x}\tilde{x}}(f) = |H_w(f)|^2 P_{xx}(f),$$

que se observa que es constante, $P_{\tilde{x}\tilde{x}}(f) = \sigma_{\tilde{x}}^2$, donde $\sigma_{\tilde{x}}^2$ es la varianza de la innovación. En general, se cumple que

$$P_{xx}(f) = \frac{\sigma_{\tilde{x}}^2}{|H_w(f)|^2}.$$

13.5. Filtro de Kalman vectorial

El filtro de Kalman con estado escalar–observación escalar puede generalizarse fácilmente. Las dos generalizaciones consisten en reemplazar $s[n]$ por $\mathbf{s}[n]$, donde $\mathbf{s}[n]$ obedece el modelo de Gauss-Markov descrito en el teorema de la sección 13.2, y reemplazar la observación escalar $x[n]$ por la observación vectorial $\mathbf{x}[n]$. La primera generalización produce el filtro de Kalman con estado vectorial–observación escalar, mientras que la segunda conduce a la forma mas general posible, el filtro de Kalman con estado vectorial–observación vectorial.

13.5.1. Estado vectorial–observación escalar

Si el estado es vectorial, tanto en el caso de observación vectorial como escalar, del teorema de la sección 13.2, el modelo de datos es

$$\mathbf{s}[n] = \mathbf{A}\mathbf{s}[n-1] + \mathbf{B}\mathbf{u}[n], \quad n \geq 0,$$

donde \mathbf{A} y \mathbf{B} son matrices $p \times p$ y $p \times r$ respectivamente, $\mathbf{u}[n]$ es WGN vectorial con $\mathbf{u}[n] \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$, $\mathbf{s}[-1] \sim \mathcal{N}(\boldsymbol{\mu}_s, \mathbf{C}_s)$, y $\mathbf{s}[-1]$ es independiente de los $\mathbf{u}[n]$. El filtro de Kalman con estado vectorial–observación escalar asume que las observaciones siguen el modelo lineal bayesiano descrito en la sección 10.4, con la hipótesis adicional de que la matriz de covarianza del ruido es diagonal. De esta forma, para la muestra n -ésima se tiene que

$$x[n] = \mathbf{h}^T[n]\mathbf{s}[n] + w[n] \quad (13.37)$$

donde $\mathbf{h}[n]$ es un vector $p \times 1$ conocido y $w[n]$ es ruido gaussiano no correlacionado, de media nula y varianza σ_n^2 , independiente de $\mathbf{s}[-1]$ y $\mathbf{u}[n]$. El modelo de datos de la ecuación 13.37 se denomina *ecuación de observación o ecuación de medida*. En la sección 13.7.1 se da un ejemplo que consiste en el seguimiento de los coeficientes de un filtro FIR aleatoriamente variable en el tiempo. El filtro de Kalman con esta configuración se deriva de forma exactamente igual al caso con estado escalar. La deducción se incluye en la sección siguiente para el caso mas general con observación vectorial. A continuación se resumen los resultados.

Predicción:

$$\hat{\mathbf{s}}[n|n-1] = \mathbf{A}\hat{\mathbf{s}}[n-1|n-1]. \quad (13.38)$$

Matriz de predicción del MSE mínimo: $(p \times p)$

$$\mathbf{M}[n|n-1] = \mathbf{A}\mathbf{M}[n-1|n-1]\mathbf{A}^T + \mathbf{B}\mathbf{Q}\mathbf{B}^T. \quad (13.39)$$

Matriz ganancia de Kalman: $(p \times 1)$

$$\mathbf{K}[n] = \frac{\mathbf{M}[n|n-1]\mathbf{h}[n]}{\sigma_n^2 + \mathbf{h}^T[n]\mathbf{M}[n|n-1]\mathbf{h}[n]}^{-1}. \quad (13.40)$$

Corrección:

$$\hat{\mathbf{s}}[n|n] = \hat{\mathbf{s}}[n|n-1] + \mathbf{K}[n](x[n] - \mathbf{h}^T[n]\hat{\mathbf{s}}[n|n-1]). \quad (13.41)$$

Matriz del MSE mínimo: $(p \times p)$

$$\mathbf{M}[n|n] = (\mathbf{I} - \mathbf{K}[n]\mathbf{h}^T[n])\mathbf{M}[n|n-1]. \quad (13.42)$$

donde las matrices de error cuadrático medio se definen como

$$\begin{aligned} \mathbf{M}[n|n] &= E[(\mathbf{s}[n] - \hat{\mathbf{s}}[n|n])(\mathbf{s}[n] - \hat{\mathbf{s}}[n|n])^T] \\ \mathbf{M}[n|n-1] &= E[(\mathbf{s}[n] - \hat{\mathbf{s}}[n|n-1])(\mathbf{s}[n] - \hat{\mathbf{s}}[n|n-1])^T] \end{aligned}$$

Para inicializar la recursión se emplea $\hat{\mathbf{s}}[-1|-1] = E(\mathbf{s}[-1]) = \boldsymbol{\mu}_s$ y $\mathbf{M}[-1|-1] = \mathbf{C}_s$. En este caso, tampoco se necesita invertir matrices, pero esto ya no es cierto en el caso de observación vectorial. Si $\mathbf{A} = \mathbf{I}$ y $\mathbf{B} = \mathbf{0}$, las ecuaciones son idénticas a las del estimador LMMSE secuencial, dadas por las ecuaciones 12.37–12.39. Esto se debe a que los modelos de señal son idénticos, siendo la señal constante en el tiempo.

13.5.2. Estado vectorial–observación vectorial

En el siguiente teorema se plantea el filtro de Kalman con estado vectorial–observación vectorial y luego se incluye la deducción, que es totalmente análoga a la del caso escalar dada en la sección 13.3.

Teorema: Filtro de Kalman vectorial La señal $\mathbf{s}[n]$ vectorial de dimensiones $p \times 1$ varía en el tiempo de acuerdo al modelo de Gauss-Markov,

$$\mathbf{s}[n] = \mathbf{A}\mathbf{s}[n-1] + \mathbf{B}\mathbf{u}[n], \quad n \geq 0,$$

donde \mathbf{A} y \mathbf{B} son matrices conocidas de dimensiones $p \times p$ y $p \times r$ respectivamente. El ruido de excitación $\mathbf{u}[n]$ tiene PDF $\mathbf{u}[n] \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$ y las muestras son independientes entre si, es decir, $E(\mathbf{u}[m]\mathbf{u}^T[n]) = \mathbf{0}$ si $m \neq n$. Esto significa que $\mathbf{u}[n]$ es WGN vectorial. El estado vectorial inicial $\mathbf{s}[-1]$ tiene PDF $\mathbf{s}[-1] \sim \mathcal{N}(\boldsymbol{\mu}_s, \mathbf{C}_s)$ y es independiente de $\mathbf{u}[n]$.

Los vectores $\mathbf{x}[n]$ de observación tienen dimensiones $M \times 1$ y se modelan mediante el modelo lineal bayesiano

$$\mathbf{x}[n] = \mathbf{H}[n]\mathbf{s}[n] + \mathbf{w}[n] \quad n \geq 0$$

donde $\mathbf{H}[n]$ es una matriz de observación $M \times p$ conocida que puede variar en el tiempo, y $\mathbf{w}[n]$ es un vector $M \times 1$ de ruido con PDF $\mathbf{w}[n] \sim \mathcal{N}(\mathbf{0}, \mathbf{C}[n])$ y las muestras son independientes entre si, es decir $E(\mathbf{w}[m]\mathbf{w}^T[n]) = \mathbf{0}$ si $m \neq n$. Si $\mathbf{C}[n]$ no dependiera de n , $\mathbf{w}[n]$ sería WGN vectorial.

El estimado MMSE de $\mathbf{s}[n]$ en base a $\{\mathbf{x}[0], \dots, \mathbf{x}[n]\}$ puede ser calculado secuencialmente mediante la siguiente recursión:

Predicción:

$$\hat{\mathbf{s}}[n|n-1] = \mathbf{A}\hat{\mathbf{s}}[n-1|n-1]. \quad (13.43)$$

Matriz de predicción del MSE mínimo: $(p \times p)$

$$\mathbf{M}[n|n-1] = \mathbf{A}\mathbf{M}[n-1|n-1]\mathbf{A}^T + \mathbf{B}\mathbf{Q}\mathbf{B}^T. \quad (13.44)$$

Matriz ganancia de Kalman: $(p \times M)$

$$\mathbf{K}[n] = \mathbf{M}[n|n-1]\mathbf{H}^T[n](\mathbf{C}[n] + \mathbf{H}[n]\mathbf{M}[n|n-1]\mathbf{H}^T[n])^{-1}. \quad (13.45)$$

Corrección:

$$\hat{\mathbf{s}}[n|n] = \hat{\mathbf{s}}[n|n-1] + \mathbf{K}[n](\mathbf{x}[n] - \mathbf{H}[n]\hat{\mathbf{s}}[n|n-1]). \quad (13.46)$$

Matriz del MSE mínimo: $(p \times p)$

$$\mathbf{M}[n|n] = (\mathbf{I} - \mathbf{K}[n]\mathbf{H}[n])\mathbf{M}[n|n-1]. \quad (13.47)$$

La recursión es inicializada con $\hat{\mathbf{s}}[-1|-1] = \boldsymbol{\mu}_s$ y $\mathbf{M}[-1|-1] = \mathbf{C}_s$.

Previamente a la deducción se realizan algunas observaciones. Los comentarios realizados sobre el filtro de Kalman con estado escalar-observación escalar son válidos para el caso vectorial, con la excepción de que ahora se necesita invertir matrices. Es necesario invertir una matriz $M \times M$ para obtener la ganancia de Kalman. Si la dimensión p del estado vectorial es menor a la dimensión M del vector de observación, existe una implementación mas eficiente del filtro. Hay que mencionar además que el filtro de Kalman dado por el teorema previo no es el mas general. Sin embargo, es apropiado para una gran cantidad de problemas prácticos. Pueden realizarse extensiones permitiendo que las matrices \mathbf{A} , \mathbf{B} y \mathbf{Q} sean variables con n . Afortunadamente, se obtiene el mismo conjunto de ecuaciones del teorema previo si se reemplaza \mathbf{A} por $\mathbf{A}[n]$, \mathbf{B} por $\mathbf{B}[n]$ y \mathbf{Q} por $\mathbf{Q}[n]$. Además, es posible extender los resultados a ruido de observación coloreado y modelos de señal con componentes determinísticos además del ruido de excitación. Finalmente, se han derivado ecuaciones de suavizado en base al método de Kalman.

Deducción del filtro de Kalman vectorial A continuación se deriva el filtro de Kalman con estado vectorial-observación vectorial, siendo el filtro con estado vectorial-observación escalar un caso especial. Se asumirá que $\boldsymbol{\mu}_s = \mathbf{0}$, de forma de que todas las variables aleatorias tendrán media nula (ver la ecuación 13.12 de la media de $\mathbf{s}[n]$). Con esta hipótesis se puede aplicar el enfoque de espacio vectorial. Puede mostrarse que si $\boldsymbol{\mu}_s \neq \mathbf{0}$ se obtiene el mismo conjunto de ecuaciones. La razón de esto solo una generalización directa de lo análogo para el estimador LMMSE para el modelo lineal bayesiano, discutido al final de la sección 12.4.1 (ver además el problema de la sección 13.8.13).

Se emplearán las siguientes propiedades del estimador MMSE:

1. El estimador MMSE de $\boldsymbol{\theta}$ en base a dos muestras no correlacionadas \mathbf{x}_1 y \mathbf{x}_2 conjuntamente gaussianas es

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= E(\boldsymbol{\theta}|\mathbf{x}_1, \mathbf{x}_2) \\ &= E(\boldsymbol{\theta}|\mathbf{x}_1) + E(\boldsymbol{\theta}|\mathbf{x}_2) \end{aligned}$$

si $\boldsymbol{\theta}$ es de media nula (ver la ecuación 11.4).

2. El estimador MMSE es lineal en el sentido de que si $\boldsymbol{\theta} = \mathbf{A}_1\boldsymbol{\theta}_1 + \mathbf{A}_2\boldsymbol{\theta}_2$, entonces

$$\begin{aligned}\hat{\boldsymbol{\theta}} &= E(\boldsymbol{\theta}|\mathbf{x}) \\ &= E(\mathbf{A}_1\boldsymbol{\theta}_1 + \mathbf{A}_2\boldsymbol{\theta}_2|\mathbf{x}) \\ &= \mathbf{A}_1E(\boldsymbol{\theta}_1|\mathbf{x}) + \mathbf{A}_2E(\boldsymbol{\theta}_2|\mathbf{x}) \\ &= \mathbf{A}_1\hat{\boldsymbol{\theta}}_1 + \mathbf{A}_2\hat{\boldsymbol{\theta}}_2.\end{aligned}$$

La deducción es idéntica a la del filtro con estado escalar-observación escalar con los ajustes obvios de las cantidades vectoriales. Se asume $\boldsymbol{\mu}_s = \mathbf{0}$ de forma de que todos los vectores aleatorios son de media nula. El estimador MMSE de $\mathbf{s}[n]$ en base a $\{\mathbf{x}[0], \dots, \mathbf{x}[n]\}$ es la media de la PDF a posteriori (ver el capítulo 10)

$$\hat{\mathbf{s}}[n|n] = E(\mathbf{s}[n]|\mathbf{x}[0], \dots, \mathbf{x}[n]). \quad (13.48)$$

Pero $\{\mathbf{s}[n], \mathbf{x}[0], \dots, \mathbf{x}[n]\}$ son conjuntamente gaussianos, ya que de la ecuación 13.11, $\mathbf{s}[n]$ depende linealmente de $\{\mathbf{s}[-1], \mathbf{u}[0], \dots, \mathbf{u}[n]\}$ y $\mathbf{x}[n]$ depende linealmente de $\mathbf{s}[n]$ y $\mathbf{w}[n]$. Por lo tanto, se tiene que $\{\mathbf{s}[n], \mathbf{x}[0], \dots, \mathbf{x}[n]\}$ son una combinación lineal de $S = \{\mathbf{s}[-1], \mathbf{u}[0], \dots, \mathbf{u}[n], \mathbf{w}[0], \dots, \mathbf{w}[n]\}$ y cada vector aleatorio de S son independientes entre si. Como resultado, los vectores en S son conjuntamente gaussianos y cualquier combinación lineal también produce vectores aleatorios conjuntamente gaussianos. Por lo tanto, de la ecuación 10.14 con medias nulas, se tiene que

$$\hat{\mathbf{s}}[n|n] = \mathbf{C}_{\theta x} \mathbf{C}_{xx}^{-1} \mathbf{x},$$

donde $\boldsymbol{\theta} = \mathbf{s}[n]$ y $\mathbf{x} = [\mathbf{x}^T[0], \dots, \mathbf{x}^T[n]]^T$, y es lineal en \mathbf{x} . Para determinar un algoritmo recursivo en el tiempo se recurre al enfoque de espacio vectorial. Sea

$$\mathbf{X}[n] = [\mathbf{x}^T[0] \quad \mathbf{x}^T[1] \quad \dots \quad \mathbf{x}^T[n-1]]^T$$

y sea $\tilde{\mathbf{x}}[n] = \mathbf{x}[n] - \hat{\mathbf{x}}[n|n-1]$ la innovación o el error incurrido en la estimación lineal de $\mathbf{x}[n]$ en base a $\mathbf{X}[n-1]$. De esta forma, se puede reemplazar la ecuación 13.48 por

$$\hat{\mathbf{s}}[n|n] = E(\mathbf{s}[n]|\mathbf{X}[n-1], \tilde{\mathbf{x}}[n])$$

ya que $\mathbf{x}[n]$ es recuperable de $\mathbf{X}[n-1]$ y $\tilde{\mathbf{x}}[n]$, como se explicó en la sección 13.3 para el caso escalar. Como $\mathbf{X}[n-1]$ y $\tilde{\mathbf{x}}[n]$ son no correlacionados, de la propiedad 1 se tiene que

$$\begin{aligned}\hat{\mathbf{s}}[n|n] &= E(\mathbf{s}[n]|\mathbf{X}[n-1]) + E(\mathbf{s}[n]|\tilde{\mathbf{x}}[n]) \\ &= \mathbf{s}[n|n-1] + E(\mathbf{s}[n]|\tilde{\mathbf{x}}[n]).\end{aligned}$$

De la ecuación 13.10, la predicción de la muestra es

$$\begin{aligned}\hat{\mathbf{s}}[n|n-1] &= E(\mathbf{A}\mathbf{s}[n-1] + \mathbf{B}\mathbf{u}[n]|\mathbf{X}[n-1]) \\ &= \mathbf{A}E(\mathbf{s}[n-1]|\mathbf{X}[n-1]) + \mathbf{B}E(\mathbf{u}[n]|\mathbf{X}[n-1]) \\ &= \mathbf{A}\hat{\mathbf{s}}[n-1|n-1],\end{aligned}$$

donde en la última igualdad se tuvo en cuenta que el segundo sumando es nulo. Esto es porque

$$E(\mathbf{u}[n]|\mathbf{X}[n-1]) = E(\mathbf{u}[n]) = \mathbf{0}$$

ya que $\mathbf{X}[n-1]$ depende de $\{\mathbf{s}[-1], \mathbf{u}[0], \dots, \mathbf{u}[n-1], \mathbf{w}[0], \dots, \mathbf{w}[n-1]\}$, que son todos independientes de $\mathbf{u}[n]$. Por lo tanto, se tiene que

$$\hat{\mathbf{s}}[n|n] = \mathbf{s}[n|n-1] + E(\mathbf{s}[n]|\tilde{\mathbf{x}}[n]), \quad (13.49)$$

donde

$$\hat{\mathbf{s}}[n|n-1] = \mathbf{A}\hat{\mathbf{s}}[n-1|n-1]. \quad (13.50)$$

Para determinar $E(\mathbf{s}[n]|\tilde{\mathbf{x}}[n])$, se observa que se trata del estimador MMSE de $\mathbf{s}[n]$ en base a $\{\tilde{\mathbf{x}}[n]\}$. Como $\mathbf{s}[n]$ y $\tilde{\mathbf{x}}[n]$ son conjuntamente gaussianos, de la ecuación 10.14 se tiene que

$$E(\mathbf{s}[n]|\tilde{\mathbf{x}}[n]) = \mathbf{C}_{s\tilde{x}} \mathbf{C}_{\tilde{x}\tilde{x}}^{-1} \tilde{\mathbf{x}}[n].$$

Si $\mathbf{K}[n] = \mathbf{C}_{s\tilde{x}} \mathbf{C}_{\tilde{x}\tilde{x}}^{-1}$ es la matriz de ganancia, esta ecuación se reescribe como

$$E(\mathbf{s}[n]|\tilde{\mathbf{x}}[n]) = \mathbf{K}[n](\mathbf{x}[n] - \hat{\mathbf{x}}[n|n-1]).$$

Pero como

$$\mathbf{x}[n] = \mathbf{H}[n]\mathbf{s}[n] + \mathbf{w}[n],$$

por la propiedad 2 se tiene que

$$\begin{aligned}\hat{\mathbf{x}}[n|n-1] &= \mathbf{H}[n]\hat{\mathbf{s}}[n|n-1] + \hat{\mathbf{w}}[n|n-1] \\ &= \mathbf{H}[n]\hat{\mathbf{s}}[n|n-1],\end{aligned}\tag{13.51}$$

ya que $\hat{\mathbf{w}}[n|n-1] = \mathbf{0}$ debido a la independencia de $\mathbf{w}[n]$ con $\{\mathbf{x}[0], \dots, \mathbf{x}[n-1]\}$. De esta forma

$$E(\mathbf{s}[n]|\tilde{\mathbf{x}}[n]) = \mathbf{K}[n](\mathbf{x}[n] - \mathbf{H}[n]\hat{\mathbf{s}}[n|n-1])$$

y sustituyendo en la ecuación 13.49, se tiene que

$$\hat{\mathbf{s}}[n|n] = \hat{\mathbf{s}}[n|n-1] + \mathbf{K}[n](\mathbf{x}[n] - \mathbf{H}[n]\hat{\mathbf{s}}[n|n-1]),\tag{13.52}$$

donde

$$\hat{\mathbf{s}}[n|n-1] = \mathbf{A}\hat{\mathbf{s}}[n-1|n-1].$$

La matriz de ganancia $\mathbf{K}[n]$ es

$$\begin{aligned}\mathbf{K}[n] &= \mathbf{C}_{s\tilde{x}} \mathbf{C}_{\tilde{x}\tilde{x}}^{-1} \\ &= E(\mathbf{s}[n]\tilde{\mathbf{x}}^T[n])E(\tilde{\mathbf{x}}[n]\tilde{\mathbf{x}}^T[n])^{-1}\end{aligned}$$

Para evaluar esta expresión se necesitan los siguientes resultados:

1.

$$E[\mathbf{s}[n](\mathbf{x}[n] - \mathbf{H}[n]\hat{\mathbf{s}}[n|n-1])^T] = E[(\mathbf{s}[n] - \hat{\mathbf{s}}[n|n-1])(\mathbf{x}[n] - \mathbf{H}[n]\hat{\mathbf{s}}[n|n-1])^T]$$

2.

$$E[\mathbf{w}[n](\mathbf{s}[n] - \hat{\mathbf{s}}[n|n-1])^T] = \mathbf{0}.$$

El primer resultado proviene de que la innovación $\tilde{\mathbf{x}}[n] = \mathbf{x}[n] - \hat{\mathbf{x}}[n|n-1] = \mathbf{x}[n] - \mathbf{H}[n]\hat{\mathbf{s}}[n|n-1]$ (ver la ecuación 13.51) no está correlacionada con las muestras previas de los datos, y por lo tanto tampoco con $\mathbf{s}[n|n-1]$, que es una combinación lineal de las muestras previas de los datos. El segundo resultado se debe a que $\mathbf{s}[n] - \hat{\mathbf{s}}[n|n-1]$ es una combinación lineal de $\{\mathbf{s}[-1], \mathbf{u}[0], \dots, \mathbf{u}[n], \mathbf{w}[0], \dots, \mathbf{w}[n-1]\}$, que es independiente de $\mathbf{w}[n]$. Se examina primero $\mathbf{C}_{s\tilde{x}}$, que es

$$\begin{aligned}\mathbf{C}_{s\tilde{x}} &= E[\mathbf{s}[n](\mathbf{x}[n] - \hat{\mathbf{x}}[n|n-1])^T] \\ &\stackrel{(a)}{=} E[\mathbf{s}[n](\mathbf{x}[n] - \mathbf{H}[n]\hat{\mathbf{s}}[n|n-1])^T] \\ &\stackrel{(b)}{=} E[(\mathbf{s}[n] - \hat{\mathbf{s}}[n|n-1])(\mathbf{x}[n] - \mathbf{H}[n]\hat{\mathbf{s}}[n|n-1])^T] \\ &\stackrel{(c)}{=} E[(\mathbf{s}[n] - \hat{\mathbf{s}}[n|n-1])(\mathbf{H}[n]\mathbf{s}[n] + \mathbf{w}[n] - \mathbf{H}[n]\hat{\mathbf{s}}[n|n-1])^T] \\ &\stackrel{(d)}{=} E[(\mathbf{s}[n] - \hat{\mathbf{s}}[n|n-1])(\mathbf{s}[n] - \hat{\mathbf{s}}[n|n-1])^T] \mathbf{H}^T[n] \\ &= \mathbf{M}[n|n-1]\mathbf{H}^T[n],\end{aligned}\tag{13.53}$$

donde en (a) se empleó la ecuación 13.51, en (b) el resultado 1, en (c) se sustituyó $\mathbf{x}[n]$ por el modelo lineal bayesiano de la hipótesis y en (d) se empleó el resultado 2. Por otro lado, $\mathbf{C}_{\tilde{x}\tilde{x}}$ es

$$\begin{aligned}\mathbf{C}_{\tilde{x}\tilde{x}} &= E[(\mathbf{x}[n] - \hat{\mathbf{x}}[n|n-1])(\mathbf{x}[n] - \hat{\mathbf{x}}[n|n-1])^T] \\ &= E[(\mathbf{x}[n] - \mathbf{H}[n]\hat{\mathbf{s}}[n|n-1])(\mathbf{x}[n] - \mathbf{H}[n]\hat{\mathbf{s}}[n|n-1])^T] \\ &= E[(\mathbf{H}[n]\mathbf{s}[n] + \mathbf{w}[n] - \mathbf{H}[n]\hat{\mathbf{s}}[n|n-1])(\mathbf{H}[n]\mathbf{s}[n] + \mathbf{w}[n] - \mathbf{H}[n]\hat{\mathbf{s}}[n|n-1])^T] \\ &\stackrel{(a)}{=} \mathbf{H}[n]E[(\mathbf{s}[n] - \hat{\mathbf{s}}[n|n-1])(\mathbf{s}[n] - \hat{\mathbf{s}}[n|n-1])^T] \mathbf{H}^T[n] + E(\mathbf{w}[n]\mathbf{w}^T[n]) \\ &= \mathbf{H}[n]\mathbf{M}[n|n-1]\mathbf{H}^T[n] + \mathbf{C}[n],\end{aligned}\tag{13.55}$$

donde en (a) se empleó el resultado 2. Empleando estos resultados, la matriz de ganancia de Kalman es

$$\mathbf{K}[n] = \mathbf{M}[n|n-1]\mathbf{H}^T[n] (\mathbf{H}[n]\mathbf{M}[n|n-1]\mathbf{H}^T[n] + \mathbf{C}[n])^{-1}.$$

Para evaluarla, hay que calcular $\mathbf{M}[n|n-1]$, y es

$$\begin{aligned} \mathbf{M}[n|n-1] &= E[(\mathbf{s}[n] - \hat{\mathbf{s}}[n|n-1])(\mathbf{s}[n] - \hat{\mathbf{s}}[n|n-1])^T] \\ &\stackrel{(a)}{=} E[(\mathbf{A}\mathbf{s}[n-1] + \mathbf{B}\mathbf{u}[n] - \mathbf{A}\hat{\mathbf{s}}[n-1|n-1])(\mathbf{A}\mathbf{s}[n-1] + \mathbf{B}\mathbf{u}[n] - \mathbf{A}\hat{\mathbf{s}}[n-1|n-1])^T] \\ &= E\left\{[\mathbf{A}(\mathbf{s}[n-1] - \hat{\mathbf{s}}[n-1|n-1]) + \mathbf{B}\mathbf{u}[n]][\mathbf{A}(\mathbf{s}[n-1] - \hat{\mathbf{s}}[n-1|n-1]) + \mathbf{B}\mathbf{u}[n]]^T\right\} \\ &\stackrel{(b)}{=} \mathbf{A}E[(\mathbf{s}[n-1] - \hat{\mathbf{s}}[n-1|n-1])(\mathbf{s}[n-1] - \hat{\mathbf{s}}[n-1|n-1])^T]\mathbf{A}^T + \mathbf{B}E(\mathbf{u}[n]\mathbf{u}[n]^T)\mathbf{B}^T, \end{aligned}$$

donde en (a) se sustituyó $\mathbf{s}[n]$ por el modelo de Gauss-Markov de hipótesis y además se empleó la ecuación 13.50 y en (b) se consideró que

$$E[(\mathbf{s}[n-1] - \hat{\mathbf{s}}[n-1|n-1])\mathbf{u}^T[n]] = \mathbf{0},$$

ya que $\mathbf{s}[n-1] - \hat{\mathbf{s}}[n-1|n-1]$ depende de $\{\mathbf{s}[-1], \mathbf{u}[0], \dots, \mathbf{u}[n-1], \mathbf{w}[0], \dots, \mathbf{w}[n-1]\}$, que son todos independientes de $\mathbf{u}[n]$. Se concluye que

$$\mathbf{M}[n|n-1] = \mathbf{A}\mathbf{M}[n-1|n-1]\mathbf{A}^T + \mathbf{B}\mathbf{Q}\mathbf{B}^T.$$

Finalmente, se necesita encontrar la recursión para $\mathbf{M}[n|n]$. Empleando la ecuación 13.52, se tiene que

$$\begin{aligned} \mathbf{M}[n|n] &= E[(\mathbf{s}[n] - \hat{\mathbf{s}}[n|n])(\mathbf{s}[n] - \hat{\mathbf{s}}[n|n])^T] \\ &= E[(\mathbf{s}[n] - \hat{\mathbf{s}}[n|n-1] - \mathbf{K}[n](\mathbf{x}[n] - \mathbf{H}[n]\hat{\mathbf{s}}[n|n-1])) \\ &\quad \times (\mathbf{s}[n] - \hat{\mathbf{s}}[n|n-1] - \mathbf{K}[n](\mathbf{x}[n] - \mathbf{H}[n]\hat{\mathbf{s}}[n|n-1]))^T] \\ &= E[(\mathbf{s}[n] - \hat{\mathbf{s}}[n|n-1])(\mathbf{s}[n] - \hat{\mathbf{s}}[n|n-1])^T] \\ &\quad - E[(\mathbf{s}[n] - \hat{\mathbf{s}}[n|n-1])(\mathbf{x}[n] - \mathbf{H}[n]\hat{\mathbf{s}}[n|n-1])^T]\mathbf{K}^T[n] \\ &\quad - \mathbf{K}[n]E[(\mathbf{x}[n] - \mathbf{H}[n]\hat{\mathbf{s}}[n|n-1])(\mathbf{s}[n] - \hat{\mathbf{s}}[n|n-1])^T] \\ &\quad + \mathbf{K}[n]E[(\mathbf{x}[n] - \mathbf{H}[n]\hat{\mathbf{s}}[n|n-1])(\mathbf{x}[n] - \mathbf{H}[n]\hat{\mathbf{s}}[n|n-1])^T]\mathbf{K}^T[n]. \end{aligned}$$

El primer sumando es $\mathbf{M}[n|n-1]$, la segunda esperanza es $\mathbf{C}_{s\hat{x}}$, como indica la ecuación 13.53, la tercera esperanza es $\mathbf{C}_{s\hat{x}}^T$, y la última esperanza, como indica la ecuación 13.55, es $\mathbf{C}_{\hat{x}\hat{x}}$. Por lo tanto,

$$\begin{aligned} \mathbf{M}[n|n] &= \mathbf{M}[n|n-1] - \mathbf{C}_{s\hat{x}}\mathbf{K}^T[n] - \mathbf{K}[n]\mathbf{C}_{s\hat{x}}^T + \mathbf{K}[n]\mathbf{C}_{\hat{x}\hat{x}}\mathbf{K}^T[n] \\ &\stackrel{(a)}{=} \mathbf{M}[n|n-1] - \mathbf{C}_{s\hat{x}}\mathbf{K}^T[n] - \mathbf{K}[n]\mathbf{C}_{s\hat{x}}^T + \mathbf{C}_{s\hat{x}}\mathbf{K}^T[n] \\ &\stackrel{(b)}{=} \mathbf{M}[n|n-1] - \mathbf{K}[n]\mathbf{H}[n]\mathbf{M}[n|n-1] \\ &= (\mathbf{I} - \mathbf{K}[n]\mathbf{H}[n])\mathbf{M}[n|n-1], \end{aligned}$$

donde en (a) se consideró que como $\mathbf{K}[n] = \mathbf{C}_{s\hat{x}}\mathbf{C}_{\hat{x}\hat{x}}^{-1}$, se cumple que $\mathbf{K}[n]\mathbf{C}_{\hat{x}\hat{x}} = \mathbf{C}_{s\hat{x}}$, y en (b) se tuvo en cuenta que, como indica la ecuación 13.54, $\mathbf{C}_{s\hat{x}}^T = (\mathbf{M}[n|n-1]\mathbf{H}^T[n])^T = \mathbf{H}[n]\mathbf{M}[n|n-1]$. Esto concluye la deducción.

13.6. Filtro de Kalman extendido

Es usual encontrar problemas en donde la ecuación de estados y/o la ecuación de observación es no lineal. Un ejemplo sencillo que se explora con cierto detalle en la sección 13.7.2 es el seguimiento de un vehículo. Para este problema, las observaciones o medidas son estimadores de la distancia $\hat{R}[n]$ y el ángulo $\hat{\beta}[n]$. Si el estado del vehículo es la posición (r_x, r_y) en coordenadas cartesianas (se asume que el

vehículo se mueve en el plano x - y), los valores verdaderos de la distancia y el ángulo se relacionan con los parámetros desconocidos como

$$R[n] = \sqrt{r_x^2[n] + r_y^2[n]}$$

$$\beta[n] = \arctan \frac{r_y[n]}{r_x[n]}.$$

Sin embargo, debido a errores de medición se dispone de los estimadores $\hat{R}[n]$ y $\hat{\beta}[n]$, que se asume son los valores verdaderos de la distancia y el rango pero con ruido de medición,

$$\begin{aligned}\hat{R}[n] &= R[n] + w_R[n] \\ \hat{\beta}[n] &= \beta[n] + w_\beta[n],\end{aligned}\tag{13.56}$$

o

$$\begin{aligned}\hat{R}[n] &= \sqrt{r_x^2[n] + r_y^2[n]} + w_R[n] \\ \hat{\beta}[n] &= \arctan \frac{r_y[n]}{r_x[n]} + w_\beta[n],\end{aligned}$$

Claramente, las ecuaciones no pueden expresarse en la forma del modelo lineal como

$$\mathbf{x}[n] = \mathbf{H}[n]\boldsymbol{\theta}[n] + \mathbf{w}[n],$$

donde $\boldsymbol{\theta}[n] = [r_x[n] \ r_y[n]]^T$. La ecuación de observación es no lineal.

Un ejemplo de ecuación de estado no lineal ocurre si se asume que el vehículo viaja en cierta dirección a velocidad fija conocida y se elijen coordenadas polares, distancia y ángulo, para representar el estado. Notar que en ese caso la ecuación de observación está dada por 13.56 y es lineal. Ignorando el ruido de excitación, la ecuación de estado es

$$\begin{aligned}r_x[n] &= v_x n \Delta + r_x[0] \\ r_y[n] &= v_y n \Delta + r_y[0],\end{aligned}\tag{13.57}$$

donde (v_x, v_y) es la velocidad, que es conocida, Δ es el período de muestreo y $r_x[0]$, $r_y[0]$ es la posición inicial. Alternativamente, la ecuación de estado puede escribirse como

$$\begin{aligned}r_x[n] &= r_x[n-1] + v_x \Delta \\ r_y[n] &= r_y[n-1] + v_y \Delta,\end{aligned}$$

por lo que la distancia es

$$\begin{aligned}R[n] &= \sqrt{(r_x[n-1] + v_x \Delta)^2 + (r_y[n-1] + v_y \Delta)^2} \\ &= \sqrt{r_x^2[n-1] + 2r_x[n-1]v_x \Delta + v_x^2 \Delta^2 + r_y^2[n-1] + 2r_y[n-1]v_y \Delta + v_y^2 \Delta^2} \\ &= \sqrt{r_x^2[n-1] + r_y^2[n-1] + 2\Delta(v_x r_x[n-1] + v_y r_y[n-1]) + (v_x^2 + v_y^2) \Delta^2} \\ &= \sqrt{R^2[n-1] + 2R[n-1]\Delta(v_x \cos \beta[n-1] + v_y \sin \beta[n-1]) + (v_x^2 + v_y^2) \Delta^2},\end{aligned}$$

donde en la última igualdad se empleó que

$$\begin{aligned}r_x[n] &= R[n] \cos \beta[n] \\ r_y[n] &= R[n] \sin \beta[n].\end{aligned}$$

De la misma forma, el ángulo es

$$\begin{aligned}\beta[n] &= \arctan \frac{r_y[n-1] + v_y \Delta}{r_x[n-1] + v_x \Delta} \\ &= \arctan \frac{R[n-1] \sin \beta[n-1] + v_y \Delta}{R[n-1] \cos \beta[n-1] + v_x \Delta}.\end{aligned}$$

Se observa que la distancia y el ángulo en el instante actual son fuertemente no lineales respecto a la distancia y el ángulo en el instante previo. *La ecuación de estados es no lineal.* En general, el problema a resolver es la estimación secuencial del estado en donde la ecuación de estado y/o la ecuación de observación es no lineal. Por lo tanto, en lugar del modelo lineal de filtro de Kalman

$$\begin{aligned}\mathbf{s}[n] &= \mathbf{A}\mathbf{s}[n-1] + \mathbf{B}\mathbf{u}[n] \\ \mathbf{x}[n] &= \mathbf{H}\mathbf{s}[n] + \mathbf{w}[n]\end{aligned}$$

se tendrá

$$\mathbf{s}[n] = \mathbf{a}(\mathbf{s}[n-1]) + \mathbf{B}\mathbf{u}[n] \quad (13.58)$$

$$\mathbf{x}[n] = \mathbf{h}(\mathbf{s}[n]) + \mathbf{w}[n], \quad (13.59)$$

donde \mathbf{a} es una función de dimensión p ($\mathbf{a} : \mathbb{R}^p \rightarrow \mathbb{R}^p$) y \mathbf{h} es una función de dimensión M ($\mathbf{h} : \mathbb{R}^p \rightarrow \mathbb{R}^M$). Las dimensiones del resto de las matrices y vectores son las mismas que antes. Ahora $\mathbf{a}(\mathbf{s}[n-1])$ representa el modelo físico verdadero de la evolución del estado, mientras que $\mathbf{u}[n]$ representa los errores de modelado, entradas imprevistas, etc. De la misma forma, $\mathbf{h}(\mathbf{s}[n])$ representa la transformación de esas variables de estado a las observaciones ideales (sin ruido). Para este problema, la estimación MMSE es intratable. Lo único que se puede hacer es encontrar una solución aproximada basada en la linealización de \mathbf{a} y \mathbf{h} . El resultado de esta linealización y la aplicación subsecuente del filtro de Kalman lineal de las ecuaciones 13.43-13.47 resulta en el *filtro de Kalman extendido*. No tiene propiedades de optimalidad y su desempeño dependerá de la exactitud de la linealización. Por tratarse de una linealización dinámica no es posible determinar su desempeño de antemano.

Para la deducción, se linealiza $\mathbf{a}(\mathbf{s}[n-1])$ en torno al estimador de $\mathbf{s}[n-1]$, es decir, en torno a $\hat{\mathbf{s}}[n-1|n-1]$. De la misma forma, se linealiza $\mathbf{h}(\mathbf{s}[n])$ en torno al estimador de $\mathbf{s}[n]$ en base a los datos previos $\hat{\mathbf{s}}[n|n-1]$, ya que de la ecuación 13.46 se necesita la ecuación de observación linealizada para determinar $\hat{\mathbf{s}}[n|n]$. Esto es porque el término $\mathbf{H}[n]\hat{\mathbf{s}}[n|n-1]$ del la ecuación 13.46 del caso lineal es $\mathbf{h}(\hat{\mathbf{s}}[n|n-1])$ en el caso no lineal. Una expansión de Taylor de primer orden conduce a

$$\begin{aligned}\mathbf{a}(\mathbf{s}[n-1]) &\approx \mathbf{a}(\hat{\mathbf{s}}[n-1|n-1]) + \left. \frac{\partial \mathbf{a}}{\partial \mathbf{s}[n-1]} \right|_{\mathbf{s}[n-1]=\hat{\mathbf{s}}[n-1|n-1]} (\mathbf{s}[n-1] - \hat{\mathbf{s}}[n-1|n-1]) \\ \mathbf{h}(\mathbf{s}[n]) &\approx \mathbf{h}(\hat{\mathbf{s}}[n|n-1]) + \left. \frac{\partial \mathbf{h}}{\partial \mathbf{s}[n]} \right|_{\mathbf{s}[n]=\hat{\mathbf{s}}[n|n-1]} (\mathbf{s}[n] - \hat{\mathbf{s}}[n|n-1]).\end{aligned}$$

Denotando a los Jacobianos como

$$\begin{aligned}\mathbf{A}[n-1] &= \left. \frac{\partial \mathbf{a}}{\partial \mathbf{s}[n-1]} \right|_{\mathbf{s}[n-1]=\hat{\mathbf{s}}[n-1|n-1]} \\ \mathbf{H}[n] &= \left. \frac{\partial \mathbf{h}}{\partial \mathbf{s}[n]} \right|_{\mathbf{s}[n]=\hat{\mathbf{s}}[n|n-1]},\end{aligned}$$

las ecuaciones de estado y de observación linealizadas, de las ecuaciones 13.58 y 13.59, son

$$\begin{aligned}\mathbf{s}[n] &= \mathbf{a}(\hat{\mathbf{s}}[n-1|n-1]) + \mathbf{A}[n-1](\mathbf{s}[n-1] - \hat{\mathbf{s}}[n-1|n-1]) + \mathbf{B}\mathbf{u}[n] \\ \mathbf{x}[n] &= \mathbf{h}(\hat{\mathbf{s}}[n|n-1]) + \mathbf{H}[n](\mathbf{s}[n] - \hat{\mathbf{s}}[n|n-1]) + \mathbf{w}[n]\end{aligned}$$

o

$$\mathbf{s}[n] = \mathbf{A}[n-1]\mathbf{s}[n-1] + \mathbf{B}\mathbf{u}[n] + (\mathbf{a}(\hat{\mathbf{s}}[n-1|n-1]) - \mathbf{A}[n-1]\hat{\mathbf{s}}[n-1|n-1]) \quad (13.60)$$

$$\mathbf{x}[n] = \mathbf{H}[n]\mathbf{s}[n] + \mathbf{w}[n] + (\mathbf{h}(\hat{\mathbf{s}}[n|n-1]) - \mathbf{H}[n]\hat{\mathbf{s}}[n|n-1]). \quad (13.61)$$

Las ecuaciones difieren de las ecuaciones estándar en que ahora \mathbf{A} es variante en el tiempo y ambas ecuaciones tienen términos conocidos sumados a ellas. A continuación se demostrará que el filtro de Kalman lineal para este modelo, que es el filtro de Kalman extendido, es

Predicción:

$$\hat{\mathbf{s}}[n|n-1] = \mathbf{a}(\hat{\mathbf{s}}[n-1|n-1]). \quad (13.62)$$

Matriz de predicción del MSE mínimo: $(p \times p)$

$$\mathbf{M}[n|n-1] = \mathbf{A}[n-1]\mathbf{M}[n-1|n-1]\mathbf{A}^T[n-1] + \mathbf{BQB}^T. \quad (13.63)$$

Matriz ganancia de Kalman: $(p \times M)$

$$\mathbf{K}[n] = \mathbf{M}[n|n-1]\mathbf{H}^T[n](\mathbf{C}[n] + \mathbf{H}[n]\mathbf{M}[n|n-1]\mathbf{H}^T[n])^{-1}. \quad (13.64)$$

Corrección:

$$\hat{\mathbf{s}}[n|n] = \hat{\mathbf{s}}[n|n-1] + \mathbf{K}[n](\mathbf{x}[n] - \mathbf{h}(\hat{\mathbf{s}}[n|n-1])). \quad (13.65)$$

Matriz del MSE mínimo: $(p \times p)$

$$\mathbf{M}[n|n] = (\mathbf{I} - \mathbf{K}[n]\mathbf{H}[n])\mathbf{M}[n|n-1], \quad (13.66)$$

donde

$$\mathbf{A}[n-1] = \left. \frac{\partial \mathbf{a}}{\partial \mathbf{s}[n-1]} \right|_{\mathbf{s}[n-1] = \hat{\mathbf{s}}[n-1|n-1]}$$

$$\mathbf{H}[n] = \left. \frac{\partial \mathbf{h}}{\partial \mathbf{s}[n]} \right|_{\mathbf{s}[n] = \hat{\mathbf{s}}[n|n-1]}.$$

Notar que en contraste con el filtro de Kalman lineal, la ganancia y las matrices MSE deben ser calculadas durante la ejecución del algoritmo, ya que no puede calcularse de antemano porque dependen de los estimadores de los estados a través de $\mathbf{A}[n-1]$ y $\mathbf{H}[n]$. Además, el término matriz MSE en este caso no es apropiado, ya que no se está implementando el estimador MMSE sino una aproximación de éste.

Para derivar las ecuaciones del filtro de Kalman extendido, se determinará primero las ecuaciones cuando el modelo de estado tiene una entrada $\mathbf{v}[n]$ determinística conocida como en la ecuación 13.60,

$$\mathbf{s}[n] = \mathbf{As}[n-1] + \mathbf{Bu}[n] + \mathbf{v}[n] \quad n \geq 0. \quad (13.67)$$

La presencia de $\mathbf{v}[n]$ como entrada produce un componente determinístico en la salida, por lo que ahora la media de $\mathbf{s}[n]$ no es nula. Se asume que $E(\mathbf{s}[-1]) = \mathbf{0}$, de forma de que si $\mathbf{v}[n] = \mathbf{0}$, $E(\mathbf{s}[n]) = \mathbf{0}$. Por lo tanto, el efecto de la entrada determinística es producir una señal vectorial con media no nula

$$\mathbf{s}[n] = \mathbf{s}'[n] + E(\mathbf{s}[n]).$$

donde $\mathbf{s}'[n]$ es el valor de $\mathbf{s}[n]$ cuando $\mathbf{v}[n] = \mathbf{0}$. Para la señal $\mathbf{s}'[n] = \mathbf{s}[n] - E(\mathbf{s}[n])$ de media nula, la ecuación de estado se puede escribir como

$$\mathbf{s}'[n] = \mathbf{As}'[n-1] + \mathbf{Bu}[n].$$

Observar que de la ecuación 13.67, la media cumple que

$$E(\mathbf{s}[n]) = \mathbf{A}E(\mathbf{s}[n-1]) + \mathbf{v}[n]. \quad (13.68)$$

De la misma forma, la ecuación de observación, por ahora sin el sumando determinístico de la ecuación 13.61, es

$$\begin{aligned} \mathbf{x}[n] &= \mathbf{H}[n]\mathbf{s}[n] + \mathbf{w}[n] \\ &= \mathbf{H}[n](\mathbf{s}'[n] + E(\mathbf{s}[n])) + \mathbf{w}[n] \\ &= \mathbf{H}[n]\mathbf{s}'[n] + \mathbf{w}[n] + \mathbf{H}[n]E(\mathbf{s}[n]), \end{aligned}$$

y definiendo

$$\mathbf{x}'[n] = \mathbf{x}[n] - \mathbf{H}[n]E(\mathbf{s}[n])$$

se obtiene la ecuación de observación usual

$$\mathbf{x}'[n] = \mathbf{H}[n]\mathbf{s}'[n] + \mathbf{w}[n].$$

El filtro de Kalman para $\mathbf{s}'[n]$ puede obtenerse a partir de las ecuaciones 13.43-13.47 reemplazando $\mathbf{x}[n]$ por $\mathbf{x}'[n]$, y luego, el estimador MMSE de $\mathbf{s}[n]$ puede obtenerse de forma directa empleando las relaciones

$$\begin{aligned}\hat{\mathbf{s}}'[n|n-1] &= \hat{\mathbf{s}}[n|n-1] - E(\mathbf{s}[n]) \\ \hat{\mathbf{s}}'[n-1|n-1] &= \hat{\mathbf{s}}[n-1|n-1] - E(\mathbf{s}[n-1]).\end{aligned}$$

De la ecuación 13.43, la ecuación de predicción para $\mathbf{s}'[n]$ es

$$\hat{\mathbf{s}}'[n|n-1] = \mathbf{A}\hat{\mathbf{s}}'[n-1|n-1]$$

por lo que para $\mathbf{s}[n]$ queda

$$\hat{\mathbf{s}}[n|n-1] - E(\mathbf{s}[n]) = \mathbf{A}(\hat{\mathbf{s}}[n-1|n-1] - E(\mathbf{s}[n-1])),$$

o

$$\hat{\mathbf{s}}[n|n-1] = \mathbf{A}\hat{\mathbf{s}}[n-1|n-1] + [E(\mathbf{s}[n]) - \mathbf{A}E(\mathbf{s}[n-1])],$$

y empleando la ecuación 13.68 se reduce a

$$\hat{\mathbf{s}}[n|n-1] = \mathbf{A}\hat{\mathbf{s}}[n-1|n-1] + \mathbf{v}[n].$$

Para la corrección, de la ecuación 13.46, se tiene que

$$\hat{\mathbf{s}}'[n|n] = \hat{\mathbf{s}}'[n|n-1] + \mathbf{K}[n](\mathbf{x}'[n] - \mathbf{H}[n]\hat{\mathbf{s}}'[n|n-1])$$

o

$$\hat{\mathbf{s}}[n|n] - E(\mathbf{s}[n]) = \hat{\mathbf{s}}[n|n-1] - E(\mathbf{s}[n]) + \mathbf{K}[n](\mathbf{x}[n] - \mathbf{H}[n]E(\mathbf{s}[n]) - \mathbf{H}[n](\hat{\mathbf{s}}[n|n-1] - E(\mathbf{s}[n]))),$$

que se reduce a

$$\hat{\mathbf{s}}[n|n] = \hat{\mathbf{s}}[n|n-1] + \mathbf{K}[n](\mathbf{x}[n] - \mathbf{H}[n]\hat{\mathbf{s}}[n|n-1]),$$

que es la ecuación usual. Tanto la ganancia de Kalman como las matrices de MSE no cambian ya que el estimador MMSE no incurre en ningún error adicional debido a constantes conocidas aditivas. Se concluye que en el caso en que la ecuación de estado tiene una entrada conocida como en la ecuación 13.67, lo único que cambia es la ecuación de predicción.

Retornando a la deducción del filtro de Kalman extendido, hay que considerar también el caso en que la ecuación tiene un componente $\mathbf{z}[n]$ aditivo determinístico conocido,

$$\mathbf{x}[n] = \mathbf{H}[n]\mathbf{s}[n] + \mathbf{w}[n] + \mathbf{z}[n].$$

Con $\mathbf{x}'[n] = \mathbf{x}[n] - \mathbf{z}[n]$ se tiene la ecuación de observación usual. Por lo tanto, las ecuaciones de predicción y corrección reemplazando \mathbf{A} por $\mathbf{A}[n]$ quedan

$$\begin{aligned}\hat{\mathbf{s}}[n|n-1] &= \mathbf{A}[n-1]\hat{\mathbf{s}}[n-1|n-1] + \mathbf{v}[n] \\ \hat{\mathbf{s}}[n|n] &= \hat{\mathbf{s}}[n|n-1] + \mathbf{K}[n](\mathbf{x}[n] - \mathbf{z}[n] - \mathbf{H}[n]\hat{\mathbf{s}}[n|n-1]).\end{aligned}$$

Pero de las ecuaciones 13.60 y 13.61 de estado y observación linealizadas, los términos aditivos determinísticos conocidos son

$$\begin{aligned}\mathbf{v}[n] &= \mathbf{a}(\hat{\mathbf{s}}[n-1|n-1]) - \mathbf{A}[n-1]\hat{\mathbf{s}}[n-1|n-1] \\ \mathbf{z}[n] &= \mathbf{h}(\hat{\mathbf{s}}[n|n-1]) - \mathbf{H}[n]\hat{\mathbf{s}}[n|n-1]\end{aligned}$$

resultando en que las ecuaciones de corrección y predicción del filtro de Kalman extendido son

$$\begin{aligned}\hat{\mathbf{s}}[n|n-1] &= \mathbf{a}(\hat{\mathbf{s}}[n-1|n-1]) \\ \hat{\mathbf{s}}[n|n] &= \hat{\mathbf{s}}[n|n-1] + \mathbf{K}[n](\mathbf{x}[n] - \mathbf{h}(\hat{\mathbf{s}}[n|n-1])),\end{aligned}$$

como indican las ecuaciones 13.62 y 13.65.

13.7. Ejemplos de procesamiento de señales

13.7.1. Estimación de canal variante en el tiempo

Muchos canales de transmisión pueden ser caracterizados por ser lineales pero no invariantes en el tiempo. En ese caso, la relación entrada-salida puede modelarse como

$$y[n] = \sum_{k=0}^{p-1} h_n[k]v[n-k], \quad (13.69)$$

que no es otra cosa que un filtro FIR con coeficientes variables en el tiempo. Para diseñar canales de comunicación efectivos, es necesario conocer los coeficientes del filtro. Por lo tanto, el problema consiste en estimar $h_n[k]$ en base a la salida del canal contaminada con ruido

$$x[n] = \sum_{k=0}^{p-1} h_n[k]v[n-k] + w[n]. \quad (13.70)$$

donde $w[n]$ es ruido de observación. Un problema similar fue discutido en el ejemplo de la sección 4.4.3, excepto que en ese caso se asumió que los coeficientes eran invariantes en el tiempo. Como consecuencia, en ese caso se pudo aplicar el modelo lineal para estimar los coeficientes determinísticos. Sin embargo, no es posible extender ese enfoque al problema actual debido a que hay muchos parámetros a estimar. Por ejemplo, si $p = 2$ y se asume que $v[n] = 0$ para $n < 0$, las observaciones son

$$\begin{aligned} x[0] &= h_0[0]v[0] + h_0[1]v[-1] + w[0] = h_0v[0] + w[0] \\ x[1] &= h_1[0]v[1] + h_1[1]v[0] + w[1] \\ x[2] &= h_2[0]v[2] + h_2[1]v[1], \end{aligned}$$

y así sucesivamente. Se observa que para $n \geq 1$ hay dos nuevos parámetros a estimar por cada nueva muestra de datos. Por lo tanto no es posible estimar los coeficientes del filtro, incluso sin ruido de observación. Una forma de enfrentar el problema es asumir que los coeficientes no cambian rápidamente de muestra a muestra. Estadísticamente, la variación lenta implica un alto grado de correlación de los coeficientes del filtro entre muestras. Esta observación conduce naturalmente a modelar a los coeficientes como variables aleatorias cuya variación se describe mediante el modelo de Gauss-Markov. El uso de dicho modelo permite imponer el grado de correlación en el tiempo entre valores de los coeficientes sucesivos. Por lo tanto, se supone que el vector de estados es

$$\mathbf{h}[n] = \mathbf{A}\mathbf{h}[n-1] + \mathbf{u}[n],$$

donde $\mathbf{h}[n] = [h_n[0] \dots h_n[p-1]]^T$, \mathbf{A} es una matriz $p \times p$ conocida y $\mathbf{u}[n]$ es un vector de WGN con matriz de covarianza \mathbf{Q} . Una observación respecto a la notación es que ahora $\mathbf{h}[n]$ se refiere a la señal $\mathbf{s}[n]$ y no al vector de observación, como en la ecuación 13.37. Una hipótesis típica que ayuda a simplificar el modelo es que los coeficientes no están correlacionados entre sí (*uncorrelated scattering*), y por lo tanto, son independientes debido a la hipótesis de que son conjuntamente gaussianos. Con esta hipótesis, las matrices \mathbf{A} , \mathbf{Q} y \mathbf{C}_h , que es la matriz de covarianza de $\mathbf{h}[-1]$, son diagonales. El modelo de Gauss-Markov se convierte entonces en p modelos escalares independientes. De la ecuación 13.70, el modelo de observación es

$$x[n] = \underbrace{[v[n] \dots v[n-p+1]]}_{\mathbf{v}^T[n]} \mathbf{h}[n] + w[n],$$

donde se asume que $w[n]$ es WGN de varianza σ^2 y la secuencia $v[n]$ se asume conocida, ya que la entrada del filtro puede ser elegida. El estimador MMSE de los coeficientes puede calcularse recursivamente en el tiempo empleando el filtro de Kalman para estado vectorial y observación escalar. Con los cambios de

notación obvios, de las ecuaciones 13.38-13.42, el estimador recursivo es

$$\begin{aligned}\hat{\mathbf{h}}[n|n-1] &= \mathbf{A}\hat{\mathbf{h}}[n-1|n-1] \\ \mathbf{M}[n|n-1] &= \mathbf{A}\mathbf{M}[n-1|n-1]\mathbf{A}^T + \mathbf{B}\mathbf{Q}\mathbf{B}^T \\ \mathbf{K}[n] &= \frac{\mathbf{M}[n|n-1]\mathbf{v}[n]}{\sigma_n^2 + \mathbf{v}^T[n]\mathbf{M}[n|n-1]\mathbf{v}[n]} \\ \hat{\mathbf{h}}[n|n] &= \hat{\mathbf{h}}[n|n-1] + \mathbf{K}[n](x[n] - \mathbf{v}^T[n]\hat{\mathbf{h}}[n|n-1]) \\ \mathbf{M}[n|n] &= (\mathbf{I} - \mathbf{K}[n]\mathbf{v}^T[n])\mathbf{M}[n|n-1]\end{aligned}$$

y se inicializa con $\hat{\mathbf{h}}[-1|-1] = \boldsymbol{\mu}_h$, $\mathbf{M}[-1|-1] = \mathbf{C}_h$.

Como ejemplo, se considera la implementación del filtro de Kalman para la estimación de un canal con $p = 2$ coeficientes. Se asume un modelo de estados con

$$\mathbf{A} = \begin{bmatrix} 0.99 & 0 \\ 0 & 0.999 \end{bmatrix} \quad \mathbf{Q} = \begin{bmatrix} 0.0001 & 0 \\ 0 & 0.0001 \end{bmatrix}.$$

La PDF del valor inicial es $h[-1] \sim \mathcal{N}(\boldsymbol{\mu}_{h[-1]}, \mathbf{C}_{h[-1]})$ con

$$\boldsymbol{\mu}_{h[-1]} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \mathbf{C}_{h[-1]} = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix}.$$

En la figura 13.9 se muestra parte del código en lenguaje Python para generar el proceso $\mathbf{h}[n]$ de Gauss-

```
import numpy as np

# generación del proceso de Gauss-Markov h[n] = Ah[n-1] + u[n] de dimensión p=2.
# N es el número total de muestras a generar: h[0], ..., h[N-1]
# valor inicial h[-1]: variable aleatoria de distribución normal multivariada con
# media mu_hi y matriz de covarianza C_hi.
# media de h[-1]
mu_hi = [1, 1]
# covarianza de h[-1]
C_hi = [[0.1, 0], [0, 0.1]]
# u[n] es WGN con matriz de covarianza Q
# media de u[n]
mu_u = [0, 0]
# covarianza de u[n]
Q = [[0.0001, 0], [0, 0.0001]]
# matriz de transición de estados
A = np.array([[0.99, 0], [0, 0.999]])
h = np.zeros((2, N)) # matriz para almacenar los valores de h[n].
# generación de h[-1]
h_i = np.random.multivariate_normal(mu_hi, C_hi, 1)[0]
h_prev = h_i
for n in np.arange(N):
    h_prev = A @ h_prev + np.random.multivariate_normal(mu_u, Q, 1)[0]
    h[:, n] = h_prev
```

Figura 13.9: Implementación en el lenguaje Python para la generación del proceso de Gauss-Markov $\mathbf{h}[n]$.

Markov y en la figura 13.10 se muestra una realización particular, en donde se observa que $h_n[0]$ decae a cero mientras que $h_n[1]$ permanece mas o menos constante. Esto es porque la media de los coeficientes son nulos en estado estacionario, como indica la ecuación 13.12, y que en este caso, $\mathbf{A}^n \rightarrow \mathbf{0}$. Esto es porque los valores propios de \mathbf{A} son los elementos de la diagonal, ambos menores que 1 (ver el problema de la sección 13.8.6). Debido al menor valor \mathbf{A}_{11} respecto a \mathbf{A}_{22} , $\mathbf{h}_n[0]$ decrece mas rápidamente que $\mathbf{h}_n[1]$. Para esta realización de los coeficientes y la entrada $v[n]$ mostrada en la figura 13.11, se obtiene la salida $y[n]$, dada por la ecuación 13.69, y cuando se agrega ruido de observación con $\sigma^2 = 0.1$ se obtiene la salida $x[n]$ del canal (ecuación 13.70). Tanto la señal $y[n]$ como la señal $x[n]$ también se muestran en la figura 13.11. Luego se aplica el filtro de Kalman con $\hat{\mathbf{h}}[-1|-1] = \mathbf{0}$ y $\mathbf{M}[-1|-1] = 100\mathbf{I}$, valores que fueron elegidos para reflejar poco conocimiento sobre el estado inicial. En el desarrollo teórico del

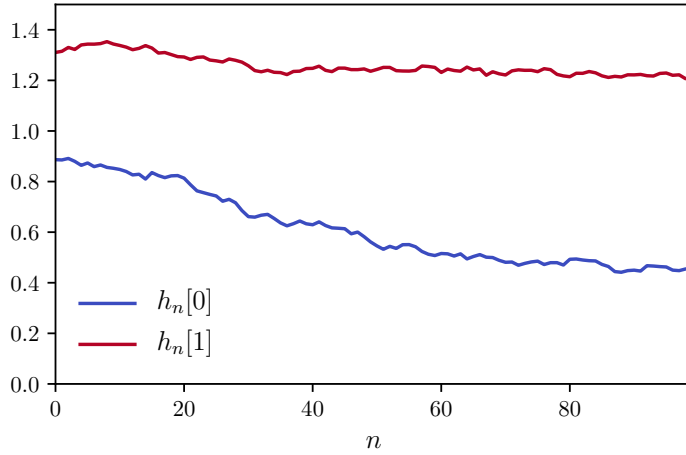


Figura 13.10: Realización de la evolución de los coeficientes $h_n[0]$ y $h_n[1]$ del canal, que se trata de un proceso de Gauss-Markov.

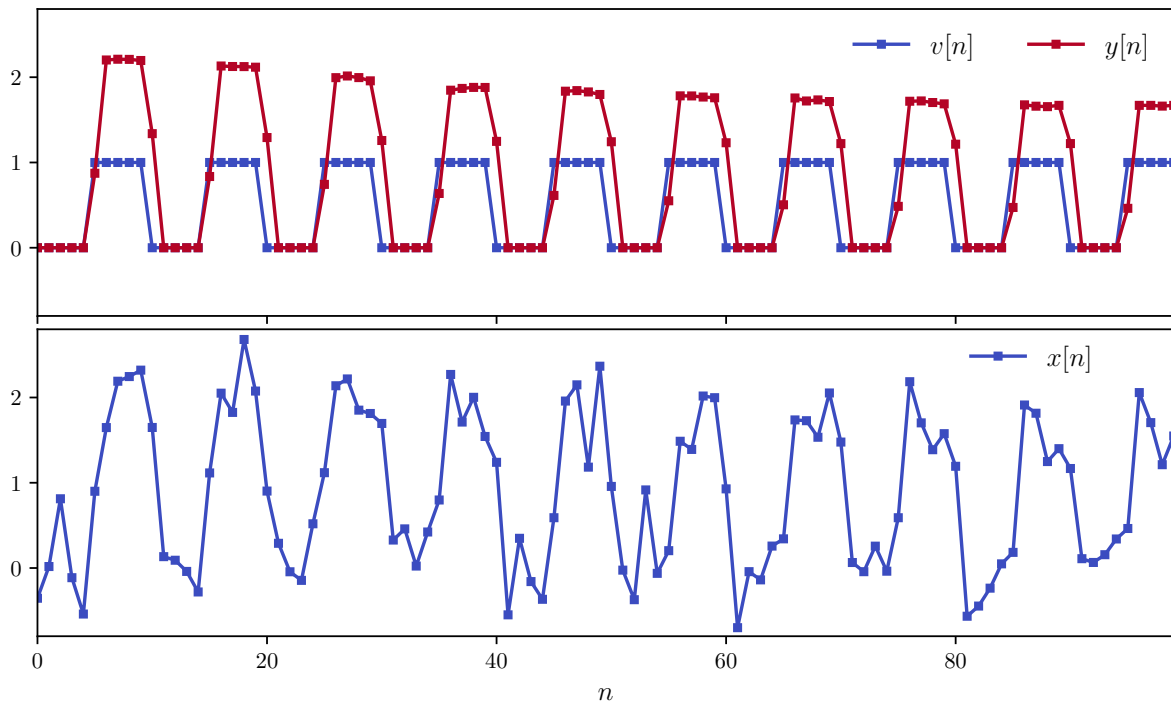


Figura 13.11: Señal $v[n]$ de entrada al canal y señal $y[n]$ de salida. También se muestra la salida $x[n]$ de canal, que es la señal $y[n]$ con WGN aditivo con potencia $\sigma^2 = 0.01$ en este ejemplo.

filtro de Kalman, el estado inicial está dado por la media de $\mathbf{s}[-1]$. En la práctica, esto en general no se conoce, por lo que se eligen valores arbitrarios con una matriz de MSE inicial grande para evitar sesgar el filtro de Kalman al estado inicial asumido. En la figura 13.12 se muestra parte del código en el lenguaje Python de la implementación del filtro de Kalman. Los coeficientes estimados se muestran en la figura 13.13. Se observa que luego de un transitorio inicial, el estimador se acerca al valor verdadero del parámetro y lo sigue de cerca. En la figura 13.14 se muestra la evolución de la ganancia de Kalman. Ambos componentes aparentan alcanzar un estado estacionario periódico, si bien este comportamiento es distinto al comportamiento de estado estacionario usual, ya que $\mathbf{v}[n]$ cambia en el tiempo, por lo que nunca se alcanza un verdadero estado estacionario. Además, se observa que en ocasiones la ganancia de Kalman es nula, como por ejemplo, $[\mathbf{K}]_1 = K_1[n]$ para $0 \leq n \leq 4$. Esto es porque $\mathbf{v}[n]$ es un vector nulo debido a que la entrada $v[n]$ es nula, por lo que las observaciones contienen solo ruido. El filtro

```

import numpy as np

# Filtro de kalman para estimación de los p=2 coeficientes del canal.
# h_est es  $\hat{h}[n|n]$ , la estimación de los coeficientes en el paso n
# y M_est es  $M[n|n]$ , el MSE mínimo en el paso n.
#  $v[n]$  es la entrada al canal. Tiene largo N+1 y el primer elemento es  $v[-1]=0$ .
# A y Q son la matrices correspondientes al modelo de estado.
# var_w es la varianza del ruido de observación.
#  $x[n]$  es el vector de observación, de largo N.

# condiciones iniciales del filtro de Kalman
#  $h[-1|-1]$ 
mu_h = [[0], [0]]
#  $M[-1|-1]$ 
C_h = 100 * np.eye(p)
# inicialización de la recursión
h_est = mu_h
M_est = C_h
for n in np.arange(N):
    # cálculo de la predicción  $\hat{h}[n|n-1]$ 
    h_pred = A @ h_est
    # cálculo de la predicción del MSE mínimo  $M[n|n-1]$ 
    M_pred = A @ M_est @ A.T + Q
    # construcción del vector  $v[n]$  de dimensión (1, 2)
    vn = np.array([[v[n+1]], [v[n]]])
    # cálculo de la ganancia de Kalman
    K = (M_pred @ vn) / (var_w + vn.T @ M_pred @ vn)
    # corrección  $\hat{h}[n|n]$ 
    h_est = h_pred + K * (x[n] - vn.T @ h_pred)
    # cálculo del MSE mínimo  $M[n|n]$ 
    M_est = (np.eye(p) - K @ vn.T) @ M_pred

```

Figura 13.12: Implementación en el lenguaje Python del filtro de Kalman con estado vectorial y observación escalar para la estimación de los coeficientes $\mathbf{h}[n]$ del canal.

de Kalman ignora esas muestras forzando a que la ganancia sea nula. Finalmente, el MSE mínimo se muestra en la figura 13.15 y se observa que decrece monótonamente, aunque este comportamiento no es general en el filtro de Kalman.

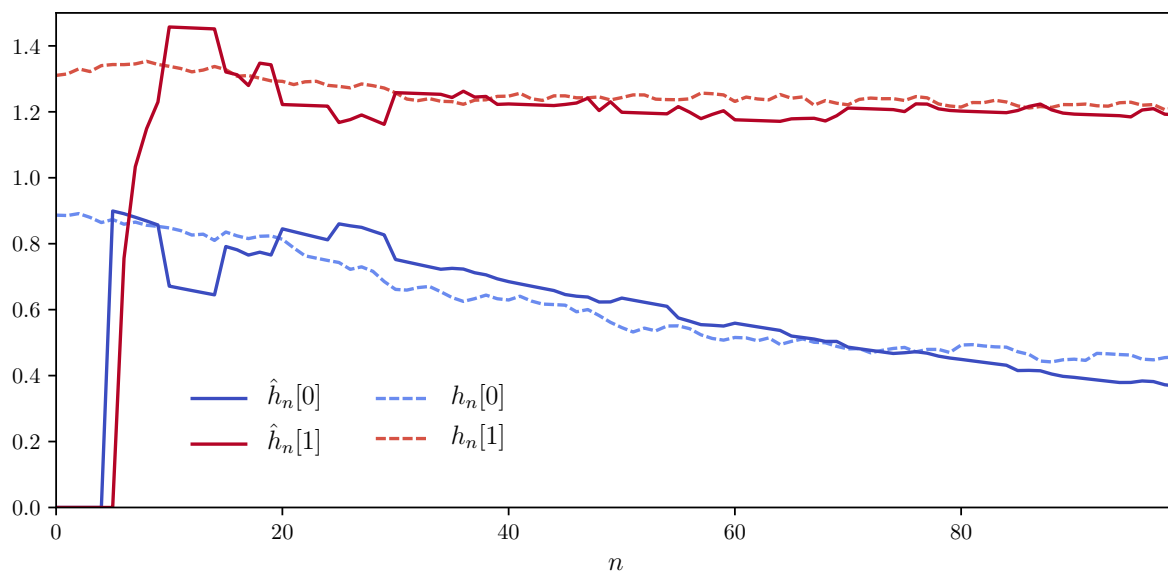


Figura 13.13: Estimación de los coeficientes $\mathbf{h}[n]$ del canal mediante el filtro de Kalman.

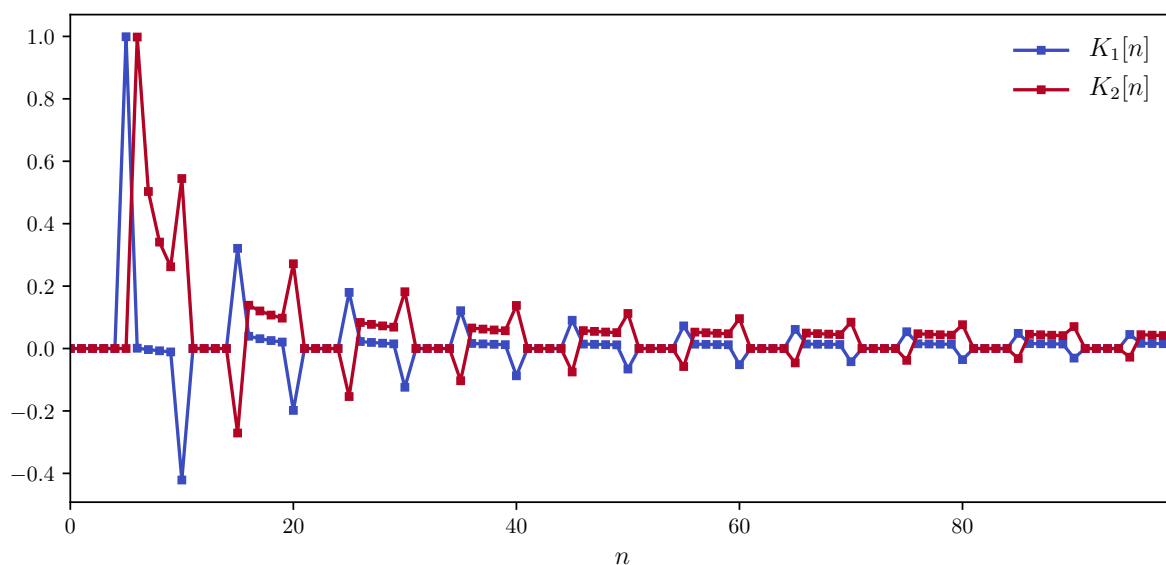


Figura 13.14: Evolución de la ganancia de Kalman $\mathbf{K}[n]$.

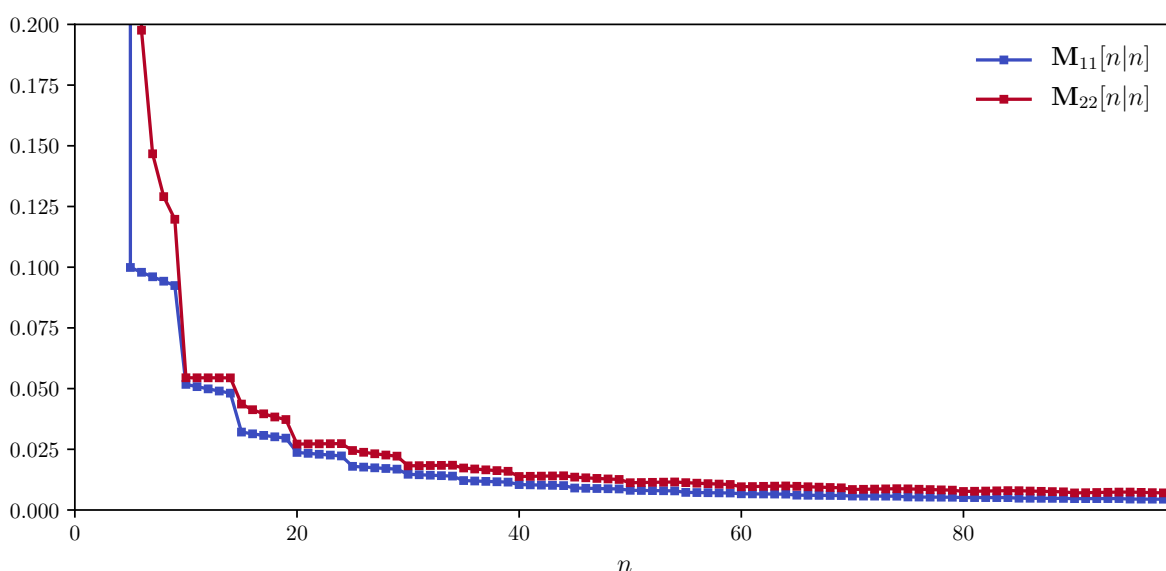


Figura 13.15: Evolución del MSE mínimo del filtro de Kalman.

13.7.2. Seguimiento de un vehículo

En este ejemplo se usa un filtro de Kalman extendido para el seguimiento de la posición y velocidad de un vehículo que se mueve en cierta dirección y velocidad nominal conocida. Las medidas son versiones ruidosas de la distancia y el ángulo. Para obtener un modelo de la dinámica del vehículo, se asume que la velocidad es constante, perturbada únicamente por ráfagas de viento o pequeñas correcciones de la velocidad, como ocurre por una aeronave. Esas perturbaciones se modelan como entradas ruidosas, de forma que los componentes de velocidad en la dirección x e y son

$$\begin{aligned} v_x[n] &= v_x[n-1] + u_x[n] \\ v_y[n] &= v_y[n-1] + u_y[n]. \end{aligned} \tag{13.71}$$

Sin las perturbaciones $u_x[n]$, $u_y[n]$ las velocidades serían constantes, y por lo tanto, el vehículo se movería en una línea recta. De las ecuaciones de movimiento, la posición en el instante n es

$$\begin{aligned} r_x[n] &= r_x[n-1] + v_x[n-1]\Delta \\ r_y[n] &= r_y[n-1] + v_y[n-1]\Delta, \end{aligned} \quad (13.72)$$

donde Δ es el intervalo de tiempo entre muestras. La señal vectorial se elige consistiendo en los componentes de posición y velocidad,

$$\mathbf{s}[n] = \begin{bmatrix} r_x[n] \\ r_y[n] \\ v_x[n] \\ v_y[n] \end{bmatrix}$$

y de las ecuaciones 13.71 y 13.72 cumple que

$$\underbrace{\begin{bmatrix} r_x[n] \\ r_y[n] \\ v_x[n] \\ v_y[n] \end{bmatrix}}_{\mathbf{s}[n]} = \underbrace{\begin{bmatrix} 1 & 0 & \Delta & 0 \\ 0 & 1 & 0 & \Delta \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} r_x[n-1] \\ r_y[n-1] \\ v_x[n-1] \\ v_y[n-1] \end{bmatrix}}_{\mathbf{s}[n-1]} + \underbrace{\begin{bmatrix} 0 \\ 0 \\ u_x[n] \\ u_y[n] \end{bmatrix}}_{\mathbf{u}[n]}. \quad (13.73)$$

Las medidas son observaciones ruidosas de la distancia y el ángulo,

$$\begin{aligned} R[n] &= \sqrt{r_x^2[n] + r_y^2[n]} \\ \beta[n] &= \arctan \frac{r_y[n]}{r_x[n]}, \end{aligned}$$

o

$$\begin{aligned} \hat{R}[n] &= R[n] + w_R[n] \\ \hat{\beta}[n] &= \beta[n] + w_\beta[n]. \end{aligned} \quad (13.74)$$

En términos generales, la ecuación de observación 13.74 es

$$\mathbf{x}[n] = \mathbf{h}(\mathbf{s}[n]) + \mathbf{w}[n],$$

donde \mathbf{h} es la función

$$\mathbf{h}(\mathbf{s}[n]) = \begin{bmatrix} \sqrt{r_x^2[n] + r_y^2[n]} \\ \arctan \frac{r_y[n]}{r_x[n]} \end{bmatrix}.$$

Desafortunadamente, el vector de medida es no lineal en los parámetros de la señal, y por lo tanto, para estimar la señal se necesita aplicar el filtro de Kalman extendido, dado por las ecuaciones 13.62-13.66. Como la ecuación 13.73 de estado es lineal, solo se necesita determinar

$$\mathbf{H}[n] = \left. \frac{\partial \mathbf{h}}{\partial \mathbf{s}[n]} \right|_{\mathbf{s}[n] = \hat{\mathbf{s}}[n|n-1]}$$

ya que $\mathbf{A}[n]$ es la matriz \mathbf{A} de la ecuación 13.73. Diferenciando la ecuación de observación para obtener el jacobiano,

$$\frac{\partial \mathbf{h}}{\partial \mathbf{s}[n]} = \begin{bmatrix} \frac{\partial R[n]}{\partial r_x[n]} & \frac{\partial R[n]}{\partial r_y[n]} & \frac{\partial R[n]}{\partial v_x[n]} & \frac{\partial R[n]}{\partial v_y[n]} \\ \frac{\partial \beta[n]}{\partial r_x[n]} & \frac{\partial \beta[n]}{\partial r_y[n]} & \frac{\partial \beta[n]}{\partial v_x[n]} & \frac{\partial \beta[n]}{\partial v_y[n]} \end{bmatrix},$$

donde

$$\frac{\partial R[n]}{\partial r_x[n]} = \frac{\partial (r_x^2[n] + r_y^2[n])^{\frac{1}{2}}}{\partial r_x[n]} = \frac{1}{2} (r_x^2[n] + r_y^2[n])^{-\frac{1}{2}} 2r_x[n] = \frac{r_x[n]}{R[n]}$$

y análogamente para la derivada respecto a $r_y[n]$, y además

$$\frac{\partial \beta[n]}{\partial r_x[n]} = \frac{\partial \arctan \frac{r_y[n]}{r_x[n]}}{\partial r_x[n]} = \frac{1}{1 + \frac{r_y^2[n]}{r_x^2[n]}} \left(-\frac{r_y[n]}{r_x^2[n]} \right) = -\frac{r_y[n]}{R[n]}.$$

y

$$\frac{\partial \beta[n]}{\partial r_y[n]} = \frac{\partial \arctan \frac{r_y[n]}{r_x[n]}}{\partial r_y[n]} = \frac{1}{1 + \frac{r_y^2[n]}{r_x^2[n]}} \left(\frac{1}{r_x[n]} \right) = \frac{r_x[n]}{R[n]}.$$

Como también

$$\frac{\partial R[n]}{\partial v_x[n]} = \frac{\partial R[n]}{\partial v_y[n]} = \frac{\partial \beta[n]}{\partial v_x[n]} = \frac{\partial \beta[n]}{\partial v_y[n]} = 0,$$

se concluye que

$$\frac{\partial \mathbf{h}}{\partial \mathbf{s}[n]} = \begin{bmatrix} \frac{r_x[n]}{R[n]} & \frac{r_y[n]}{R[n]} & 0 & 0 \\ -\frac{r_y[n]}{R[n]} & \frac{r_x[n]}{R[n]} & 0 & 0 \end{bmatrix}.$$

Finalmente, se necesita especificar las covarianzas del ruido de excitación y el ruido de observación. Si se asume que las ráfagas de viento y las correcciones de velocidad ocurren de forma equiprobable en cualquier dirección con la misma magnitud, parece razonable asignar la misma varianza a $u_x[n]$ y $u_y[n]$ y además, asumir que son independientes. Denominando σ_u^2 a la varianza común, la matriz de covarianza del ruido de excitación es

$$\mathbf{Q} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_u^2 & 0 \\ 0 & 0 & 0 & \sigma_u^2 \end{bmatrix}.$$

El valor exacto de σ_u^2 depende de la posibilidad de cambio de los componentes de la velocidad de una muestra a otra, ya que $u_x[n] = v_x[n] - v_x[n-1]$. Esto es la aceleración multiplicada por Δ , y debería poder derivarse de la física del vehículo. Para especificar las varianzas del ruido de medición, se observa que el error de medición puede pensarse como el error de estimación de $\hat{R}[n]$ y $\hat{\beta}[n]$. De esta forma si $\hat{R}[n]$ y $\hat{\beta}[n]$ fueran el MLE, si se dispone de un registro de datos largo o la SNR es grande, se puede asumir que los estimadores alcanzan la CRLB para establecer las varianzas. Por simplicidad, se asume que los errores de estimación son independientes y que las varianzas no cambian, aunque esto no es siempre válido. De esta forma, se tiene la matriz de covarianza del ruido de observación es

$$\mathbf{C}[n] = \mathbf{C} = \begin{bmatrix} \sigma_R^2 & 0 \\ 0 & \sigma_\beta^2 \end{bmatrix}.$$

En resumen, las ecuaciones del filtro de Kalman extendido para este problema, de las ecuaciones 13.62-13.66, son

$$\begin{aligned} \hat{\mathbf{s}}[n|n-1] &= \mathbf{A}\hat{\mathbf{s}}[n-1|n-1] \\ \mathbf{M}[n|n-1] &= \mathbf{A}[n-1]\mathbf{M}[n-1|n-1]\mathbf{A}^T[n-1] + \mathbf{Q} \\ \mathbf{K}[n] &= \mathbf{M}[n|n-1]\mathbf{H}^T[n](\mathbf{C} + \mathbf{H}[n]\mathbf{M}[n|n-1]\mathbf{H}^T[n])^{-1} \\ \hat{\mathbf{s}}[n|n] &= \hat{\mathbf{s}}[n|n-1] + \mathbf{K}[n](\mathbf{x}[n] - \mathbf{h}(\hat{\mathbf{s}}[n|n-1])) \\ \mathbf{M}[n|n] &= (\mathbf{I} - \mathbf{K}[n]\mathbf{H}[n])\mathbf{M}[n|n-1], \end{aligned}$$

donde

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & \Delta & 0 \\ 0 & 1 & 0 & \Delta \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{Q} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_u^2 & 0 \\ 0 & 0 & 0 & \sigma_u^2 \end{bmatrix}, \quad \mathbf{x}[n] = \begin{bmatrix} \hat{R}[n] \\ \hat{\beta}[n] \end{bmatrix}, \quad \mathbf{h}(\mathbf{s}[n]) = \begin{bmatrix} \sqrt{r_x^2[n] + r_y^2[n]} \\ \arctan \frac{r_y[n]}{r_x[n]} \end{bmatrix},$$

$$\mathbf{H}[n] = \begin{bmatrix} \frac{r_x[n]}{\sqrt{r_x^2[n] + r_y^2[n]}} & \frac{r_y[n]}{\sqrt{r_x^2[n] + r_y^2[n]}} & 0 & 0 \\ -\frac{r_y[n]}{\sqrt{r_x^2[n] + r_y^2[n]}} & \frac{r_x[n]}{\sqrt{r_x^2[n] + r_y^2[n]}} & 0 & 0 \end{bmatrix}_{\mathbf{s}[n]=\hat{\mathbf{s}}[n|n-1]}$$

y $\mathbf{C} = \begin{bmatrix} \sigma_R^2 & 0 \\ 0 & \sigma_\beta^2 \end{bmatrix}$,

y las condiciones iniciales son $\hat{\mathbf{s}}[-1|-1] = \boldsymbol{\mu}_s$ y $\mathbf{M}[-1|-1] = \mathbf{C}_s$. En la figura 13.16 se muestra el pseudocódigo en lenguaje Python de la implementación del filtro de Kalman extendido. Como ejemplo,

```
import numpy as np
from numpy.linalg import inv

def fun_H(s):
    """
    Implementación del cálculo de la matriz H[n]
    :param s: vector de estados
    """
    rx = s[0]
    ry = s[1]
    R2 = rx ** 2 + ry ** 2
    return np.array([[rx / np.sqrt(R2), ry / np.sqrt(R2), 0, 0], [-ry / R2, rx / R2, 0, 0]])

def fun_h(s):
    """
    Implementación del cálculo de h(s[n])
    :param s: vector de estados
    """
    rx = s[0]
    ry = s[1]
    return [np.sqrt(rx ** 2 + ry ** 2), np.arctan2(ry, rx)]

# Implementación del filtro de Kalman extendido.
# p es el número de elementos de la señal, N es el número de muestras,
# A es la matriz (p, p) de transición de estados, Q es la matriz (p, p) de ruido de
# medición, C es la matriz (2, 2) de ruido de observación,
# s_est_i es el vector (p, ) de estado inicial hat{s}[-1|-1] y C_s_i es la matriz (p, p)
# de MSE mínimo inicial M[-1|-1], x una matriz (2, N) con los vectores de observación
# en las columnas.
s_est = s_est_i
M_est = C_s_i
for n in np.arange(N):
    # predicción del estimador
    s_pred = A @ s_est
    # predicción de la matriz de MSE mínimo
    M_pred = A @ M_est @ A.T + Q
    # cálculo de H
    H = fun_H(s_pred)
    # cálculo de la ganancia de Kalman
    K = M_pred @ H.T @ inv(C + H @ M_pred @ H.T)
    # corrección del estimador
    s_est = s_pred + K @ (x[:, n] - fun_h(s_pred))
    # corrección de la matriz de MSE mínimo
    M_est = (np.eye(p) - K @ H) @ M_pred
```

Figura 13.16: Parte de la implementación en el lenguaje Python del filtro de Kalman extendido para la estimación de la posición y velocidad del vehículo.

supóngase que la trayectoria ideal del vehículo es la línea recta dada por

$$\begin{aligned} r_x[n] &= 10 - 0.2n \\ r_y[n] &= -5 + 0.2n \end{aligned}$$

para $n = 0, \dots, N - 1$ con $N = 100$, donde se asumió que $\Delta = 1$ por conveniencia, y se muestra en la figura 13.17. De la ecuación 13.57, esta trayectoria asume que $v_x = -0.2$ y $v_y = 0.2$. Para contemplar un movimiento de vehículo mas realista, se introduce ruido de excitación, por lo que el estado del vehículo

es descrito por la ecuación 13.73 con $\sigma_u^2 = 0.0001$. Con un estado inicial

$$s[-1] = \begin{bmatrix} 10 \\ -5 \\ -0.2 \\ 0.2 \end{bmatrix},$$

que es idéntico al estado inicial de la trayectoria rectilínea ideal, una realización de la posición $[r_x[n] \ r_y[n]]^T$ del vehículo se muestra en la figura 13.17. Para generar esta realización se empleó la ecuación 13.73. Ob-

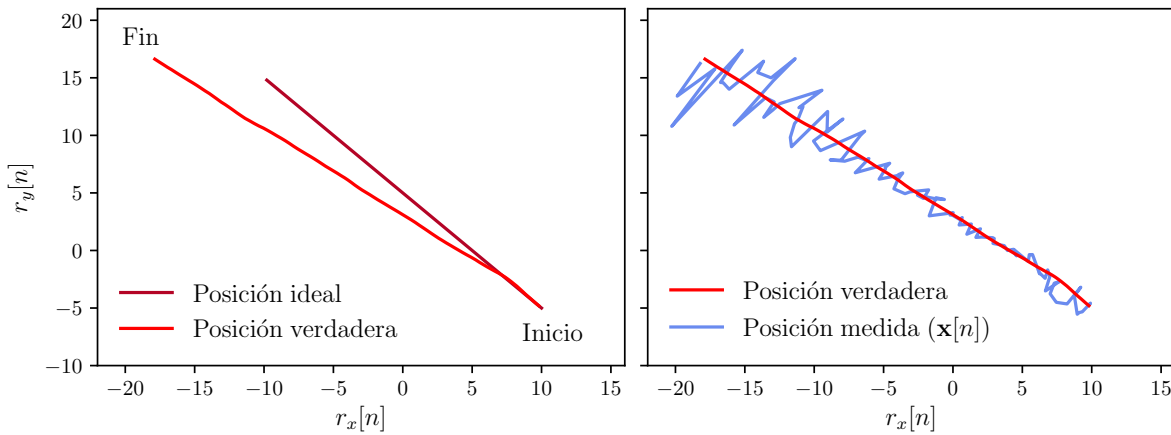


Figura 13.17: Trayectoria ideal, trayectoria verdadera y trayectoria observada.

servar que a medida que transcurre el tiempo, la trayectoria verdadera se desvía gradualmente de la línea recta. Puede demostrarse que las varianzas de $v_x[n]$ y $v_y[n]$ eventualmente crecen a infinito (ver el problema de la sección 13.8.22), haciendo que $r_x[n]$ y $r_y[n]$ se hagan rápidamente no acotados. Por lo tanto, este modelo es válido solo para una parte de la trayectoria. La distancia y el ángulo verdadero se muestran en la figura 13.18. Se asume que el ruido de medición es $\sigma_R^2 = 0.1$ y $\sigma_\beta^2 = 0.01$, donde β se mide en radianes. Con estos valores, la distancia y el ángulo medido, que son las observaciones $\mathbf{x}[n]$, también se muestran en la figura 13.18. En la figura 13.17 se compara la trayectoria verdadera con la trayectoria contaminada con ruido obtenida como

$$\begin{aligned} \hat{r}_x[n] &= \hat{R}[n] \cos \hat{\beta}[n] \\ \hat{r}_y[n] &= \hat{R}[n] \sin \hat{\beta}[n]. \end{aligned}$$

En la figura 13.17 se compara la trayectoria verdadera con la trayectoria contaminada con ruido obtenida como

$$\begin{aligned} \hat{r}_x[n] &= \hat{R}[n] \cos \hat{\beta}[n] \\ \hat{r}_y[n] &= \hat{R}[n] \sin \hat{\beta}[n]. \end{aligned}$$

Para emplear un filtro de Kalman extendido se debe especificar un estimador del estado inicial. En la práctica es poco probable que se disponga del conocimiento de la velocidad y la posición. Por lo tanto, como ilustración se elige el estado inicial arbitrario

$$\hat{\mathbf{s}}[-1|-1] = \begin{bmatrix} 5 \\ 5 \\ 0 \\ 0 \end{bmatrix}.$$

Como el estado inicial está bastante apartado del verdadero, para no sesgar el filtro de Kalman extendido a este estado inicial, se asume un MSE inicial grande, $\mathbf{M}[-1|-1] = 100\mathbf{I}$. Los resultados del filtro de

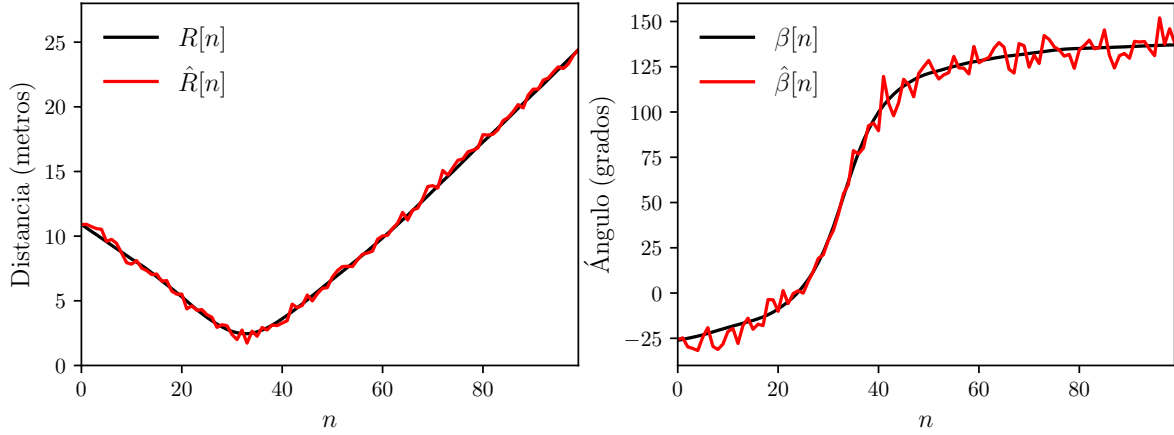


Figura 13.18: Distancia y ángulo verdaderos y observados.

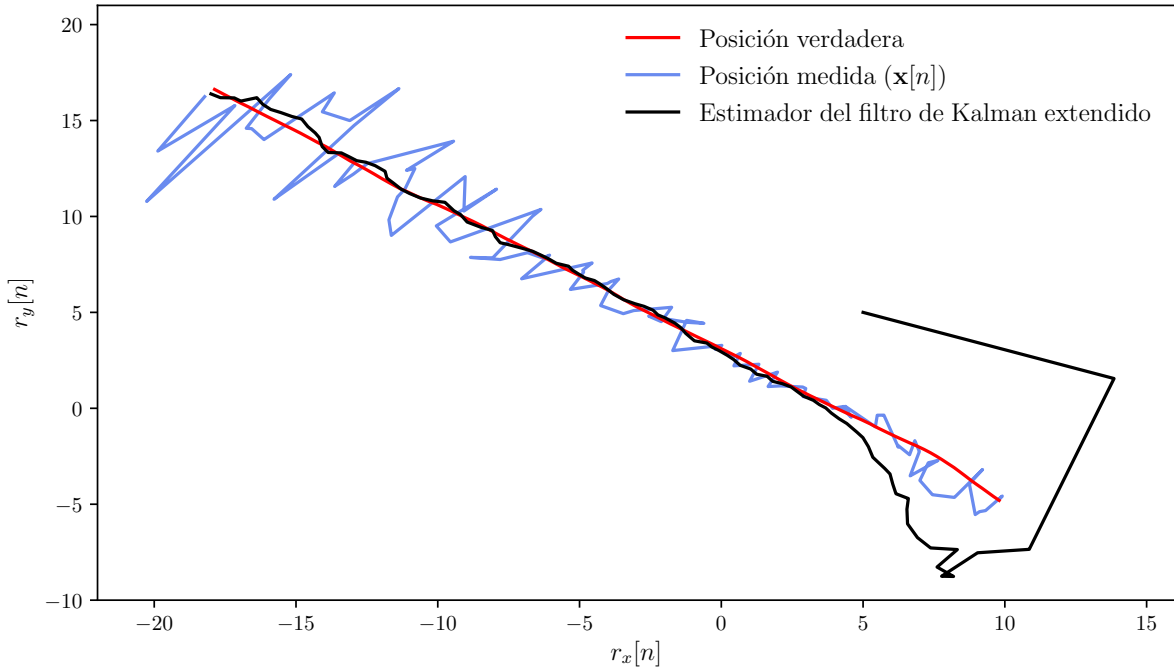


Figura 13.19: Resultado de la estimación de la posición por el filtro de Kalman extendido.

Kalman extendido se muestran en la figura 13.19. Inicialmente, debido al pobre estimador inicial, el error es grande. Esto se refleja en las curvas del MSE de la figura 13.20, que en realidad, debido a la linealización, son solo estimadores del error. Sin embargo, luego de alrededor de 20 muestras, el filtro alcanza la posición del vehículo. Es interesante notar que el MSE mínimo no decrece monótonamente como en el ejemplo previo. Por el contrario, crece a partir de cierto instante. Esto se explica contrastando el filtro de Kalman con el estimador LMMSE secuencial discutido en el capítulo 12. En el caso del último, se estima el mismo parámetro a medida que se reciben datos. Como consecuencia, el MSE mínimo decrece, o, en el peor caso, se mantendrá constante. Con el filtro de Kalman, cada vez que se recibe una muestra, se está estimando un nuevo parámetro. La incertidumbre creciente del nuevo parámetro debido al creciente ruido de excitación puede ser suficientemente grande para sobrepasar el conocimiento obtenido por la observación de la nueva muestra, provocando que el MSE crezca, como se ilustra en el problema de

la sección 13.8.23. Como última observación, en esta simulación, el filtro de Kalman extendido parece tolerante a los errores producidos por la linealización debidos al pobre estado inicial.

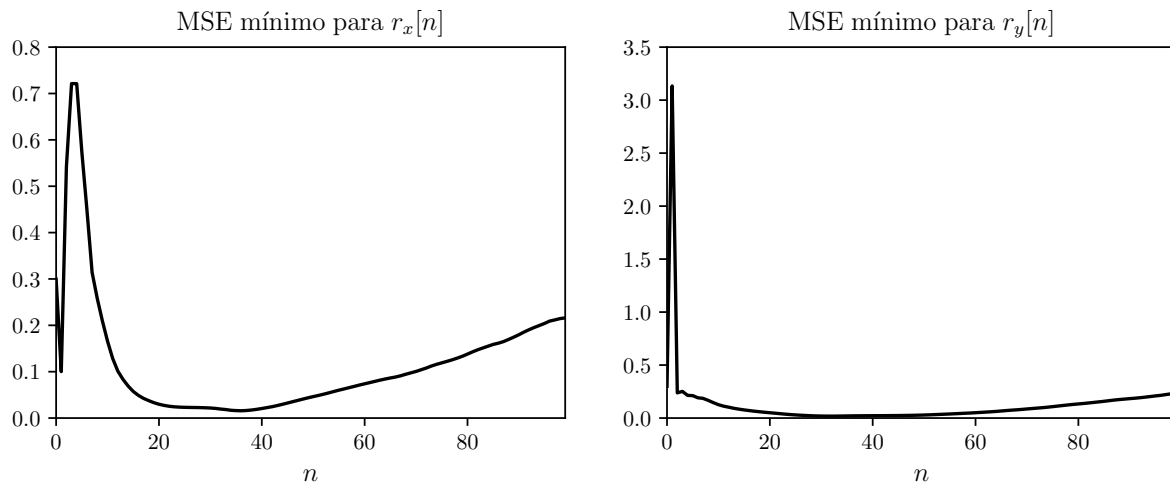


Figura 13.20: MSE mínimo de $r_x[n]$ y $r_y[n]$ del filtro de Kalman extendido.

13.8. Problemas

13.8.1. Problema 1

Un proceso es gaussiano si para muestras arbitrarias $\{s[n_1], s[n_2], \dots, s[n_k]\}$, cualquier vector $\mathbf{s} = [s[n_1] \ s[n_2] \ \dots \ s[n_k]]^T$ de dimensiones $k \times 1$ está distribuido de acuerdo a un PDF gaussiana multivariada. Si $s[n]$ está dada por 13.2, probar que es un proceso aleatorio gaussiano.

Solución Sea el vector

$$\mathbf{s} = [s[n_1] \ s[n_2]] \dots s[n_k]]^T,$$

donde las muestras están dadas por la ecuación 13.2,

$$s[n_i] = a^{n_i+1} s[-1] + \sum_{k=0}^{n_i} a^{n_i-k} u[n_i].$$

Si $n_1 < n_2 < \dots < n_k$, se cumple que

$$\mathbf{s} = \begin{bmatrix} s[n_1] \\ s[n_2] \\ s[n_3] \\ \vdots \\ s[n_k] \end{bmatrix} = \begin{bmatrix} a^{n_1+1} & a^{n_1} & \dots & 1 & 0 & 0 & \dots & 0 \\ a^{n_2+1} & a^{n_2} & \dots & a & 1 & 0 & \dots & 0 \\ a^{n_3+1} & a^{n_3} & \dots & a^2 & a & 1 & \dots & 0 \\ \vdots & \vdots & & & \ddots & \ddots & \ddots & \vdots \\ a^{n_k+1} & a^{n_k} & a^{n_k-1} & \dots & \dots & a^2 & a & 1 \end{bmatrix} \begin{bmatrix} s[-1] \\ u[0] \\ u[1] \\ \vdots \\ u[n_k] \end{bmatrix}.$$

Como las variables aleatorias $\{s[-1], u[0], \dots, u[n_k]\}$ son gaussianas e independientes, son conjuntamente gaussianas, por lo que las variables aleatorias $\{s[n_1], s[n_2], \dots, s[n_k]\}$ también son conjuntamente gaussianas por tratarse de una transformación lineal de $\{s[-1], u[0], \dots, u[n_k]\}$. Se concluye que el vector $\mathbf{s} = [s[n_1] \ s[n_2]] \dots s[n_k]]^T$ tiene PDF gaussiana multivariada.

13.8.2. Problema 2

Se considera un proceso de Gauss-Markov escalar. Mostrar que si $\mu_s = 0$ y $\sigma_s^2 = \sigma_u^2/(1 - a^2)$, el proceso es WSS para $n \geq 0$ y explicar porque ocurre esto.

Solución De la ecuación 13.3 y considerando que $\mu_s = 0$, la media del proceso es

$$E(s[n]) = a^{n+1}\mu_s = 0 \quad n \geq 0,$$

es decir, la media es constante para todo $n \geq 0$. De la ecuación 13.4 la covarianza es

$$\begin{aligned} c[m, n] &= a^{m+n+2}\sigma_s^2 + \sigma_u^2 a^{m-n} \sum_{k=0}^n a^{2k} \\ &\stackrel{(a)}{=} a^{m+n+2} \frac{\sigma_u^2}{1-a^2} + \sigma_u^2 a^{m-n} \frac{1-a^{2(n+1)}}{1-a^2} \\ &= \frac{\sigma_u^2}{1-a^2} a^{m-n} (a^{2n+2} + 1 - a^{2n+2}) \\ &= \frac{\sigma_u^2}{1-a^2} a^{m-n}, \end{aligned}$$

donde en (a) se substituyó σ_s^2 por el valor dado y se expresó la serie geométrica de la sumatoria de forma analítica. Se observa que la covarianza depende solo de la diferencia $m - n$. Se concluye que con la estadística de $s[-1]$ indicada el proceso es WSS. Esto ocurre porque dicha estadística es la de $s[n]$ en estado estacionario, o cuando $n \rightarrow \infty$, por lo que el proceso se encuentra en estado estacionario en $n = -1$.

13.8.3. Problema 3

Graficar la media, varianza, y la covarianza en estado estacionario de un proceso de Gauss-Markov si $a = 0.98$, $\sigma_u^2 = 0.1$, $\mu_s = 5$, y $\sigma_s^2 = 1$. ¿Cual es la PSD del proceso en estado estacionario?

Solución De la ecuación 13.3, la media del proceso es

$$E(s[n]) = a^{n+1}\mu_s,$$

y de la ecuación 13.5, la varianza es

$$\begin{aligned} \text{var}(s[n]) &= a^{2n+2}\sigma_s^2 + \sigma_u^2 \sum_{k=0}^n a^{2k} \\ &= a^{2n+2}\sigma_s^2 + \frac{1-a^{2n+2}}{1-a^2}\sigma_u^2. \end{aligned}$$

En la figura 13.21 se muestra $E(s[n])$ y $\text{var}(s[n])$.

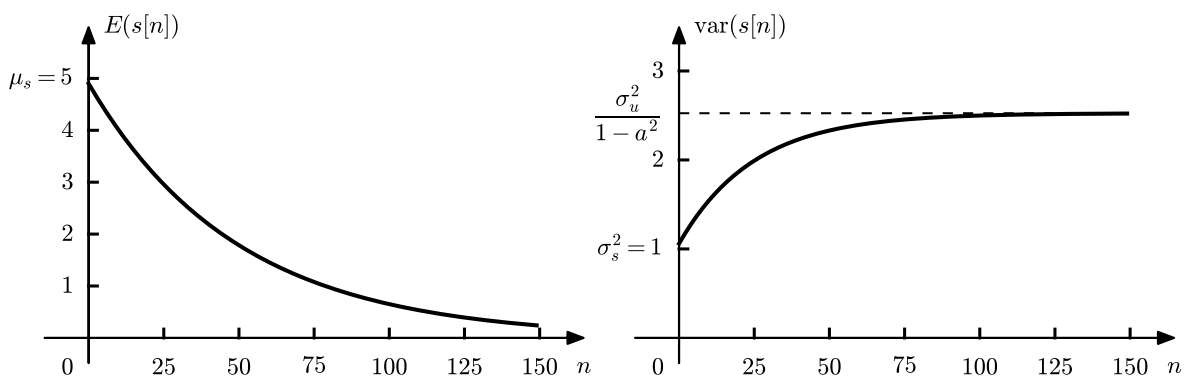


Figura 13.21: Evolución de la media y la varianza del proceso de Gauss-Markov.

La covarianza en estado estacionario, es la ACF, que está dada por la ecuación 13.6

$$r_{ss}[k] = \frac{\sigma_u^2}{1-a^2} a^{|k|}.$$

Como el proceso de Gauss-Markov tiende a un proceso AR(1) si $a < 1$, la PSD en estado estacionario es la PSD de un proceso AR(1). Si bien la PSD de un proceso AR(1) es conocida, igualmente se calculará. Partiendo de la transformada de Fourier de la ACF y operando, se tiene que

$$\begin{aligned} P_{ss}(f) &= \sum_{k=-\infty}^{\infty} r_{ss}[k] e^{-j2\pi f k} \\ &= \frac{\sigma_u^2}{1-a^2} \sum_{k=-\infty}^{\infty} a^{|k|} e^{-j2\pi f k} \\ &= \frac{\sigma_u^2}{1-a^2} \left(\sum_{k=0}^{\infty} a^k e^{-j2\pi f k} + \sum_{k=-\infty}^{-1} a^{-k} e^{-j2\pi f k} \right) \\ &= \frac{\sigma_u^2}{1-a^2} \left(\sum_{k=0}^{\infty} a^k e^{-j2\pi f k} + \sum_{k=1}^{\infty} a^k e^{j2\pi f k} \right) \\ &= \frac{\sigma_u^2}{1-a^2} \left[\sum_{k=0}^{\infty} (ae^{-j2\pi f})^k + \sum_{k=0}^{\infty} (ae^{j2\pi f})^k - 1 \right] \\ &= \frac{\sigma_u^2}{1-a^2} \left(\frac{1}{1-ae^{-j2\pi f}} + \frac{1}{1-ae^{j2\pi f}} - 1 \right) \\ &= \frac{\sigma_u^2}{1-a^2} \left[\frac{1-ae^{j2\pi f} + 1-ae^{-j2\pi f} + (1-ae^{-j2\pi f})(1-ae^{j2\pi f})}{(1-ae^{-j2\pi f})(1-ae^{j2\pi f})} \right]. \end{aligned}$$

Teniendo en cuenta que

$$(1-ae^{-j2\pi f})(1-ae^{j2\pi f}) = 1-ae^{j2\pi f}-ae^{-j2\pi f}+a^2 = 1-2a\cos 2\pi f+a^2,$$

la PSD queda

$$\begin{aligned} P_{ss}(f) &= \frac{\sigma_u^2}{1-a^2} \left[\frac{2-2a\cos 2\pi f - (1-2a\cos 2\pi f + a^2)}{1-2a\cos 2\pi f + a^2} \right] \\ &= \left(\frac{\sigma_u^2}{1-a^2} \right) \left(\frac{1-a^2}{1-2a\cos 2\pi f + a^2} \right), \end{aligned}$$

resultando en

$$P_{ss}(f) = \frac{\sigma_u^2}{1-2a\cos 2\pi f + a^2}.$$

En la figura 13.22 se muestra $r_{ss}[k]$ y $P_{ss}(f)$.

13.8.4. Problema 4

Para un proceso Gauss-Markov escalar, derivar la ecuación de propagación de la covarianza, esto es, una fórmula que relacione $c_s[m, n]$ con $c_s[n, n]$ para $m \geq n$. Para hacerlo, mostrar primero que

$$c_s[m, n] = a^{m-n} \text{var}(s[n])$$

para $m \geq n$.

Solución La covarianza es

$$c[m, n] = E[(s[m] - E(s[m]))(s[n] - E(s[n]))].$$

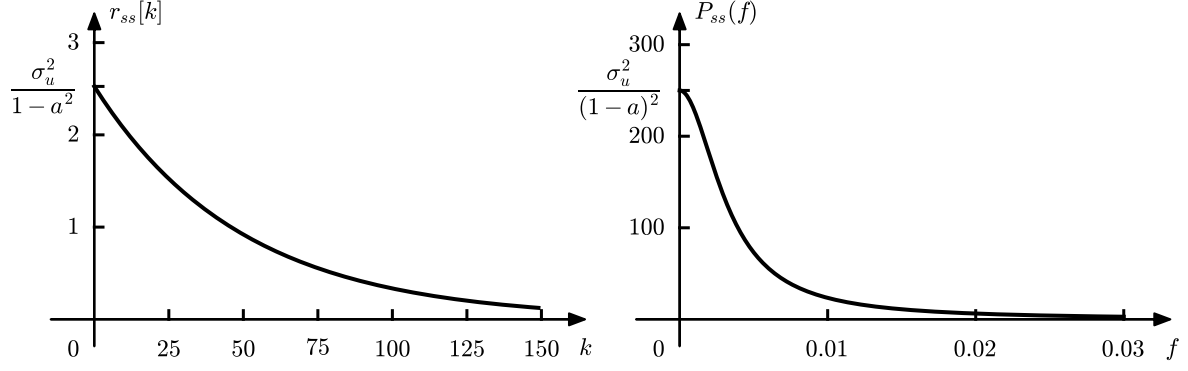


Figura 13.22: ACF y PSD del proceso de Gauss-Markov en estado estacionario. La PSD se grafica solo hasta la frecuencia 0.03, pero continúa acercándose a 0 a medida que la frecuencia crece a 0.5.

Observando que

$$\begin{aligned} s[n+1] &= as[n] + u[n+1] \\ s[n+2] &= as[n+1] + u[n+2] = a^2s[n] + au[n+1] + u[n+2] \\ s[n+3] &= as[n+2] + u[n+3] = a^3s[n] + a^2u[n+1] + au[n+2] + u[n+3], \end{aligned}$$

y en general se cumple que

$$s[n+l] = a^l s[n] + \sum_{k=1}^l a^{l-k} u[n+k].$$

Tomando $m = n + l$, se tiene que

$$s[m] = a^{m-n} s[n] + \sum_{k=1}^{m-n} a^{m-n-k} u[n+k].$$

Por otro lado, de la ecuación 13.7

$$E(s[m]) = aE(s[m-1]) = a^l E(s[m-l]),$$

y tomando $l = m - n$, se tiene que

$$E(s[m]) = a^{m-n} E(s[n]).$$

Sustituyendo estos resultados en la covarianza, se tiene que

$$\begin{aligned} c_s[m, n] &= E \left[\left(a^{m-n} s[n] + \sum_{k=1}^{m-n} a^{m-n-k} u[n+k] - a^{m-n} E(s[n]) \right) (s[n] - E(s[n])) \right] \\ &\stackrel{(a)}{=} a^{m-n} E[(s[n] - E(s[n]))^2] + \sum_{k=1}^{m-n} a^{m-n-k} E[u[n+k](s[n] - E(s[n]))] \\ &= a^{m-n} \text{var}(s[n]), \end{aligned}$$

donde en (a) se consideró que $u[n+k]$ es independiente de $s[n]$ para todo $k \geq 1$. Teniendo en cuenta que $c_s[n, n] = \text{var}(s[n])$, se concluye que

$$c_s[m, n] = a^{m-n} c_s[n, n], \quad m \geq n,$$

y además $c_s[m, n] = c_s[n, m]$ si $m < n$.

La misma deducción puede realizarse de forma mas directa empleando resultados obtenidos previamente. De la ecuación 13.4, se tiene que

$$\begin{aligned} c_s[m, n] &= a^{m+n+2} \sigma_s^2 + \sigma_u^2 a^{m-n} \sum_{k=0}^n a^{2k} \\ &= a^{m-n} \left(a^{2n+2} \sigma_s^2 + \sigma_u^2 \sum_{k=0}^n a^{2k} \right) \\ &\stackrel{(a)}{=} a^{m-n} \text{var}(s[n]), \end{aligned}$$

donde en (a) se empleó la ecuación 13.5.

13.8.5. Problema 5

Verificar las ecuaciones 13.15 y 13.16 de la propagación de la media y la covarianza de un proceso de Gauss-Markov vectorial.

Solución La ecuación 13.15 se obtiene directamente de la ecuación 13.10 tomando esperanza. Para obtener la ecuación de propagación de la covarianza, se parte de la covarianza y se sustituye 13.10,

$$\begin{aligned} \mathbf{C}[n] &= E[(\mathbf{s}[n] - E(\mathbf{s}[n]))(\mathbf{s}[n] - E(\mathbf{s}[n]))^T] \\ &= E\left\{[\mathbf{A}(\mathbf{s}[n-1] - E(\mathbf{s}[n-1])) + \mathbf{B}\mathbf{u}[n]] [\mathbf{A}(\mathbf{s}[n-1] - E(\mathbf{s}[n-1])) + \mathbf{B}\mathbf{u}[n]]^T\right\} \\ &\stackrel{(a)}{=} \mathbf{A}E[(\mathbf{s}[n-1] - E(\mathbf{s}[n-1]))(\mathbf{s}[n-1] - E(\mathbf{s}[n-1]))^T] \mathbf{A}^T + \mathbf{B}E(\mathbf{u}[n]\mathbf{u}[n]^T)\mathbf{B}^T \\ &= \mathbf{A}\mathbf{C}[n-1]\mathbf{A}^T + \mathbf{B}\mathbf{Q}\mathbf{B}^T, \end{aligned}$$

donde en (a) se consideró que $\mathbf{s}[n-1]$ y $\mathbf{u}[n]$ son no correlacionados y $\mathbf{u}[n]$ es de media nula.

13.8.6. Problema 6

Mostrar que la media de un proceso de Gauss-Markov vectorial en general crece con n si algún valor propio de \mathbf{A} es mayor que 1 en magnitud. ¿Qué ocurre con la media en estado estacionario si todos los valores propios son menor que 1 en magnitud? Por razones de simplicidad, asumir que \mathbf{A} es simétrica de forma que puede escribirse como $\mathbf{A} = \sum_{i=1}^p \lambda_i \mathbf{v}_i \mathbf{v}_i^T$, donde \mathbf{v}_i es el i -ésimo vector propio de \mathbf{A} y λ_i es el valor propio correspondiente.

Solución Si la matriz \mathbf{A} es simétrica, admite una descomposición de la forma $\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^T$, donde $\mathbf{D} = \text{diag}(\lambda_1, \dots, \lambda_p)$ es una matriz diagonal formada con los valores propios de \mathbf{A} y $\mathbf{P} = [\mathbf{v}_1 \dots \mathbf{v}_p]$ es una matriz ortonormal cuyas columnas son los vectores propios de \mathbf{A} (ver el teorema espectral en el apéndice H). Por lo tanto, \mathbf{A} se puede escribir como

$$\mathbf{A} = [\mathbf{v}_1 \dots \mathbf{v}_p] \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \lambda_p \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_p^T \end{bmatrix} = [\mathbf{v}_1 \dots \mathbf{v}_p] \begin{bmatrix} \lambda_1 \mathbf{v}_1^T \\ \vdots \\ \lambda_p \mathbf{v}_p^T \end{bmatrix} = \sum_{i=1}^p \lambda_i \mathbf{v}_i \mathbf{v}_i^T.$$

La media del proceso está dada por la ecuación 13.12 y es

$$E(\mathbf{s}[n]) = \mathbf{A}^{n+1} \boldsymbol{\mu}_s.$$

Considerando la descomposición de \mathbf{A} , se observa que

$$\mathbf{A}^2 = (\mathbf{P}\mathbf{D}\mathbf{P}^T)(\mathbf{P}\mathbf{D}\mathbf{P}^T) \stackrel{(a)}{=} \mathbf{P}\mathbf{D}\mathbf{D}\mathbf{P}^T = \mathbf{P}\mathbf{D}^2\mathbf{P}^T,$$

donde en (a) se consideró que como \mathbf{P} es una matriz ortogonal, se cumple que $\mathbf{P}^T \mathbf{P} = \mathbf{I}$. En general se cumple que

$$\mathbf{A}^{n+1} = \mathbf{P} \mathbf{D}^{n+1} \mathbf{P}^T, \quad \text{con} \quad \mathbf{D}^{n+1} = \text{diag}(\lambda_1^{n+1}, \dots, \lambda_p^{n+1}),$$

o

$$\mathbf{A}^{n+1} = \sum_{i=1}^p \lambda_i^{n+1} \mathbf{v}_i \mathbf{v}_i^T.$$

Por lo tanto, se cumple que

$$E(\mathbf{s}[n]) = \left(\sum_{i=1}^p \lambda_i^{n+1} \mathbf{v}_i \mathbf{v}_i^T \right) \boldsymbol{\mu}_s \xrightarrow{n \rightarrow \infty} \begin{cases} \mathbf{0} \boldsymbol{\mu}_s = \mathbf{0} & |\lambda_i| < 1 \\ \infty \boldsymbol{\mu}_s = \infty & |\lambda_i| > 1 \end{cases} \quad \begin{matrix} i = 1, \dots, p \\ \text{para algún } i. \end{matrix}$$

13.8.7. Problema 7

Mostrar que

$$\mathbf{A}^n \mathbf{C}_s \mathbf{A}^{nT} \rightarrow \mathbf{0}$$

cuando $n \rightarrow \infty$ si todos los valores propios de \mathbf{A} son menores a 1 en magnitud. Como en el problema 13.8.6, asumir que \mathbf{A} es simétrica. Sugerencia: examinar $\mathbf{e}_i^T \mathbf{A}^n \mathbf{C}_s \mathbf{A}^{nT} \mathbf{e}_j = [\mathbf{A}^n \mathbf{C}_s \mathbf{A}^{nT}]_{ij}$, donde \mathbf{e}_i es un vector de ceros excepto que el elemento i -ésimo vale 1.

Solución Como la matriz \mathbf{A} es simétrica, se puede descomponer empleando el teorema espectral como en el problema 13.8.6, y se cumple que

$$\mathbf{A}^n = \mathbf{P} \mathbf{D}^n \mathbf{P}^T.$$

Por lo tanto,

$$\mathbf{A}^n \mathbf{C}_s \mathbf{A}^{nT} = \mathbf{P} \mathbf{D}^n \mathbf{P}^T \mathbf{C}_s \mathbf{P} \mathbf{D}^n \mathbf{P}^T,$$

y el elemento (i, j) de esta matriz es

$$[\mathbf{A}^n \mathbf{C}_s \mathbf{A}^{nT}]_{ij} = \mathbf{e}_i^T \mathbf{A}^n \mathbf{C}_s \mathbf{A}^{nT} \mathbf{e}_j = \underbrace{\mathbf{e}_i^T \mathbf{P} \mathbf{D}^n \mathbf{P}^T}_{\mathbf{b}^T} \mathbf{C}_s \underbrace{\mathbf{P} \mathbf{D}^n \mathbf{P}^T}_{\mathbf{a}} \mathbf{e}_j = \mathbf{b}^T \mathbf{P}^T \mathbf{C}_s \mathbf{P} \mathbf{a}.$$

Teniendo en cuenta que

$$\mathbf{a} = \mathbf{D}^n \mathbf{P}^T \mathbf{e}_j = \mathbf{D}^n [\mathbf{P}^T]_j,$$

donde $[\mathbf{P}^T]_j$ denota la columna j -ésima de \mathbf{P}^T , el elemento i -ésimo de \mathbf{a} es

$$\mathbf{a}_i = \lambda_i^n [\mathbf{P}^T]_{ij} = \lambda_i^n [\mathbf{P}]_{ji} \xrightarrow{n \rightarrow \infty} 0, \quad \text{si} \quad |\lambda_i| < 1,$$

o

$$\mathbf{a} \xrightarrow{n \rightarrow \infty} \mathbf{0}, \quad \text{si} \quad |\lambda_i| < 1, \quad i = 1, \dots, p.$$

Empleando el mismo argumento, se cumple que

$$\mathbf{b} \xrightarrow{n \rightarrow \infty} \mathbf{0}, \quad \text{si} \quad |\lambda_i| < 1, \quad i = 1, \dots, p,$$

Se concluye que

$$[\mathbf{A}^n \mathbf{C}_s \mathbf{A}^{nT}]_{ij} = \mathbf{b}^T \mathbf{P}^T \mathbf{C}_s \mathbf{P} \mathbf{a} \xrightarrow{n \rightarrow \infty} 0, \quad \text{para todo } i, j.$$

13.8.8. Problema 8

Se considera la ecuación recursiva en diferencias

$$r[n] = - \sum_{k=1}^p a[k] r[n-k] + u[n] + \sum_{k=1}^q b[k] u[n-k], \quad n \geq 0, \quad (13.75)$$

donde $u[n]$ es WGN con varianza σ_u^2 y $q < p$. En estado estacionario, sería un proceso ARMA. Se define el vector de estados como

$$\mathbf{s}[n-1] = \begin{bmatrix} r[n-p] \\ r[n-p+1] \\ \vdots \\ r[n-1] \end{bmatrix},$$

donde $s[-1] \sim \mathcal{N}(\mu_s, \mathbf{C}_s)$ y es independiente de $u[n]$. También se define la secuencia vectorial de ruido de excitación como

$$\mathbf{u}[n] = \begin{bmatrix} u[n-q] \\ u[n-q+1] \\ \vdots \\ u[n] \end{bmatrix}.$$

Reescribir el proceso en la forma de la ecuación 13.10. Explicar porque no se trata de un proceso de Gauss-Markov vectorial examinando las hipótesis sobre $\mathbf{u}[n]$.

Solución Considerando el vector de estados y el vector de ruido de excitación dados, el proceso se puede escribir de la forma de la ecuación 13.10,

$$\mathbf{s}[n] = \mathbf{A}\mathbf{s}[n-1] + \mathbf{B}\mathbf{u}[n],$$

como

$$\underbrace{\begin{bmatrix} r[n-p+1] \\ r[n-p+2] \\ \vdots \\ r[n-1] \\ r[n] \end{bmatrix}}_{\mathbf{s}[n]} = \underbrace{\begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ -a[p] & -a[p-1] & -a[p-2] & \dots & -a[-1] \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} r[n-p] \\ r[n-p+1] \\ \vdots \\ r[n-2] \\ r[n-1] \end{bmatrix}}_{\mathbf{s}[n-1]} + \underbrace{\begin{bmatrix} 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 \\ b[q] & b[q-1] & \dots & b[1] & 1 \end{bmatrix}}_{\mathbf{B}} \underbrace{\begin{bmatrix} u[n-q] \\ u[n-q+1] \\ \vdots \\ u[n-1] \\ u[n] \end{bmatrix}}_{\mathbf{u}[n]},$$

donde \mathbf{A} tiene dimensiones $p \times p$ y \mathbf{B} tiene dimensiones $p \times q$. Como las muestras de ruido de excitación del proceso cumplen que $u[n] \sim \mathcal{N}(0, \sigma_u^2)$, el vector de ruido cumple que $\mathbf{u}[n] \sim \mathcal{N}(\mathbf{0}, \sigma_u^2 \mathbf{I})$. Sin embargo, las muestras vectoriales están correlacionadas entre si, ya que tienen muestras de ruido en común. Efectivamente, sean $\mathbf{u}[n]$ y $\mathbf{u}[n-k]$ con $k \leq q$. De esta forma, se tiene que

$$\begin{aligned} \mathbf{u}[n]\mathbf{u}^T[n-k] &= \begin{bmatrix} u[n-q] \\ u[n-q+1] \\ \vdots \\ u[n-k] \\ \vdots \\ u[n] \end{bmatrix} \begin{bmatrix} u[n-q-k] & \dots & u[n-q] & u[n-q+1] & \dots & u[n-k] \end{bmatrix} \\ &= \begin{bmatrix} u[n-q]u[n-q-k] & \dots & u^2[n-q] & u[n-q]u[n-q+1] & \dots & u[n-q]u[n-k] \\ u[n-q+1]u[n-q-k] & \dots & u[n-q+1]u[n-q] & u^2[n-q+1] & \dots & u[n-q+1]u[n-k] \\ \vdots & & \vdots & \vdots & & \vdots \\ u[n-k]u[n-q-k] & \dots & u[n-k]u[n-q] & u[n-k]u[n-q+1] & \dots & u^2[n-k] \\ \vdots & & \vdots & \vdots & & \vdots \\ u[n]u[n-q-k] & \dots & u[n]u[n-q] & u[n]u[n-q+1] & \dots & u[n]u[n-k] \end{bmatrix}, \end{aligned}$$

y por lo tanto,

$$E(\mathbf{u}[n]\mathbf{u}^T[n-k]) = \begin{bmatrix} 0 & \cdots & \sigma_u^2 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \sigma_u^2 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & \sigma_u^2 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 \end{bmatrix} \neq \mathbf{0}.$$

Se concluye que $\mathbf{u}[n]$ no es una secuencia de WGN vectorial, por lo que el proceso no es de Gauss-Markov.

13.8.9. Problema 9

Para el problema de la sección 13.8.8, mostrar que el proceso puede expresarse alternativamente mediante las ecuaciones en diferencias

$$s[n] = - \sum_{k=1}^p a[k]s[n-k] + u[n] \quad (13.76)$$

$$r[n] = s[n] + \sum_{k=1}^q b[k]s[n-k] \quad (13.77)$$

para $n \geq 0$. Asumiendo que $\mathbf{s}[-1] = [s[-p] \ s[-p+1] \ \dots \ s[-1]]^T \sim \mathcal{N}(\boldsymbol{\mu}_s, \mathbf{C}_s)$ y se observa $x[n] = r[n] + w[n]$, donde $w[n]$ es WGN con varianza σ_w^2 , y $\mathbf{s}[-1]$, $u[n]$ y $w[n]$ son independientes, mostrar como configurar un filtro de Kalman con estado vectorial-observación escalar.

Solución Partiendo de las ecuaciones en diferencias dadas, de la ecuación 13.76 se tiene que

$$s[n-k] = - \sum_{l=1}^p a[l]s[n-l-k] + u[n-k],$$

y sustituyendo este resultado y la ecuación 13.76 en la ecuación 13.77, se tiene que

$$\begin{aligned} r[n] &= - \sum_{k=1}^p a[k]s[n-k] + u[n] + \sum_{k=1}^q b[k] \left(- \sum_{l=1}^p a[l]s[n-l-k] + u[n-k] \right) \\ &= - \sum_{k=1}^p a[k]s[n-k] + u[n] - \sum_{k=1}^q b[k] \sum_{l=1}^p a[l]s[n-l-k] + \sum_{k=1}^q b[k]u[n-k] \\ &\stackrel{(a)}{=} - \sum_{k=1}^p a[k]s[n-k] + u[n] - \sum_{k=1}^p a[k] \sum_{l=1}^q b[l]s[n-k-l] + \sum_{k=1}^q b[k]u[n-k] \\ &= - \sum_{k=1}^p a[k] \left(s[n-k] + \sum_{l=1}^q b[l]s[n-k-l] \right) + u[n] + \sum_{k=1}^q b[k]u[n-k] \end{aligned}$$

donde en (a) se intercambi6 el orden de las sumatorias y se intercambi6 el nombre de las variables k y l en el t6rmino con la doble sumatoria. Finalmente, teniendo en cuenta que el t6rmino entre par6ntesis es $r[n-k]$, se obtiene que

$$r[n] = - \sum_{k=1}^p a[k]r[n-k] + u[n] + \sum_{k=1}^q b[k]u[n-k],$$

que es la ecuaci6n 13.75.

Para configurar un filtro de Kalman para estimar $s[n]$, se observa que la ecuación 13.76 se puede escribir en la forma de la ecuación 13.10 como

$$\underbrace{\begin{bmatrix} s[n-p+1] \\ s[n-p+2] \\ \vdots \\ s[n-1] \\ s[n] \end{bmatrix}}_{\mathbf{s}[n]} = \underbrace{\begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ -a[p] & -a[p-1] & -a[p-2] & \dots & -a[-1] \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} s[n-p] \\ s[n-p+1] \\ \vdots \\ s[n-2] \\ s[n-1] \end{bmatrix}}_{\mathbf{s}[n-1]} + \underbrace{\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}}_{\mathbf{B}} u[n].$$

Además, a partir de la definición del vector de estado $\mathbf{s}[n]$, la ecuación 13.77 puede escribirse como

$$r[n] = \underbrace{\begin{bmatrix} 0 & \dots & 0 & b[q] & \dots & b[1] & 1 \end{bmatrix}}_{\mathbf{h}^T} \begin{bmatrix} s[n-p+1] \\ \vdots \\ s[n-q-1] \\ s[n-q] \\ \vdots \\ s[n-1] \\ s[n] \end{bmatrix},$$

por lo que las observaciones se pueden escribir como

$$x[n] = \mathbf{h}^T \mathbf{s}[n] + w[n].$$

El filtro de Kalman está dado por las ecuaciones 13.38-13.42 con $\mathbf{h}[n] = \mathbf{h}$, $\mathbf{Q} = \sigma_n^2$ y las matrices \mathbf{A} y \mathbf{B} definidas previamente. Para la inicialización se toma $\mathbf{s}[-1| -1] = \boldsymbol{\mu}_s$ y $\mathbf{M}[-1| -1] = \mathbf{C}_s$.

13.8.10. Problema 10

Se asume que se observa $x[n] = A + w[n]$ para $n = 0, 1, \dots$, donde A es la realización de una variable aleatoria con PDF $\mathcal{N}(0, \sigma_A^2)$ y $w[n]$ es WGN con varianza σ^2 . Empleando un filtro de Kalman con estado escalar-observación escalar, encontrar un estimador secuencial de A en base a $\{x[0], \dots, x[n]\}$, o \hat{A} . Resolver explícitamente para $\hat{A}[n]$, la ganancia de Kalman y el MSE mínimo.

Solución En este caso, el modelo es de Gauss-Markov de primer orden dado por la ecuación 13.1 con $s[n] = A$, $a = 1$ y $\sigma_u^2 = 0$ y la observación es $x[n] = A + w[n]$, donde el ruido de observación tiene varianza $\sigma_n^2 = \sigma^2$ constante. El estimador secuencial es el filtro de Kalman con estado escalar-observación escalar dado por las ecuaciones 13.26-13.30. Como con $a = 1$ y $\sigma_u^2 = 0$ se tiene que $\hat{A}[n|n-1] = \hat{A}[n-1|n-1]$ y $M[n|n-1] = M[n-1|n-1]$, modificando la notación, las ecuaciones del filtro de Kalman se reducen a

$$\begin{aligned} K[n] &= \frac{M[n-1]}{\sigma^2 + M[n-1]} \\ \hat{A}[n] &= \hat{A}[n-1] + K[n](x[n] - \hat{A}[n-1]) \\ M[n] &= (1 - K[n])M[n-1]. \end{aligned}$$

La inicialización se realiza con $\hat{A}[-1] = 0$ y $M[-1] = \sigma_A^2$. Sustituyendo la ganancia de Kalman en el MSE mínimo, se tiene que

$$M[n] = \left(1 - \frac{M[n-1]}{\sigma^2 + M[n-1]}\right) M[n-1] = \frac{\sigma^2 M[n-1]}{\sigma^2 + M[n-1]}.$$

Comenzando con la recursión para $M[n]$, se tiene que

$$M[0] = \frac{\sigma^2 M[-1]}{\sigma^2 + M[-1]} = \frac{\sigma^2 \sigma_A^2}{\sigma^2 + \sigma_A^2}.$$

Luego, para $n = 1$,

$$M[1] = \frac{\sigma^2 M[0]}{\sigma^2 + M[0]} = \frac{\sigma^2 \left(\frac{\sigma^2 \sigma_A^2}{\sigma^2 + \sigma_A^2} \right)}{\sigma^2 + \frac{\sigma^2 \sigma_A^2}{\sigma^2 + \sigma_A^2}} = \frac{\sigma^4 \sigma_A^2}{\sigma^4 + 2\sigma^2 \sigma_A^2} = \frac{\sigma^2 \sigma_A^2}{\sigma^2 + 2\sigma_A^2}.$$

Continuando con $n = 2$,

$$M[2] = \frac{\sigma^2 M[1]}{\sigma^2 + M[1]} = \frac{\sigma^2 \left(\frac{\sigma^2 \sigma_A^2}{\sigma^2 + 2\sigma_A^2} \right)}{\sigma^2 + \frac{\sigma^2 \sigma_A^2}{\sigma^2 + 2\sigma_A^2}} = \frac{\sigma^4 \sigma_A^2}{\sigma^4 + 3\sigma^2 \sigma_A^2} = \frac{\sigma^2 \sigma_A^2}{\sigma^2 + 3\sigma_A^2},$$

y en el caso general, el MSE mínimo queda

$$M[n] = \frac{\sigma^2 \sigma_A^2}{\sigma^2 + (n+1)\sigma_A^2}.$$

Por lo tanto, la ganancia de Kalman es

$$K[n] = \frac{M[n-1]}{\sigma^2 + M[n-1]} = \frac{\frac{\sigma^2 \sigma_A^2}{\sigma^2 + n\sigma_A^2}}{\sigma^2 + \frac{\sigma^2 \sigma_A^2}{\sigma^2 + n\sigma_A^2}} = \frac{\sigma^2 \sigma_A^2}{\sigma^4 + (n+1)\sigma_A^2},$$

resultando en

$$K[n] = \frac{\sigma_A^2}{\sigma^2 + (n+1)\sigma_A^2}.$$

El estimador se puede escribir como

$$\begin{aligned} \hat{A}[n] &= \hat{A}[n-1] + K[n](x[n] - \hat{A}[n-1]) \\ &= \hat{A}[n-1] + \frac{\sigma_A^2}{\sigma^2 + (n+1)\sigma_A^2} (x[n] - \hat{A}[n-1]) \\ &= \frac{1}{\sigma^2 + (n+1)\sigma_A^2} \left[(\sigma^2 + (n+1)\sigma_A^2) \hat{A}[n-1] + \sigma_A^2 (x[n] - \hat{A}[n-1]) \right] \\ &= \frac{1}{\sigma^2 + (n+1)\sigma_A^2} \left[(\sigma^2 + n\sigma_A^2) \hat{A}[n-1] + \sigma_A^2 x[n] \right] \end{aligned}$$

resultando en

$$\hat{A}[n] = \frac{\sigma_A^2}{\sigma^2 + (n+1)\sigma_A^2} \left(\frac{\sigma^2 + n\sigma_A^2}{\sigma_A^2} \hat{A}[n-1] + x[n] \right).$$

Comenzando con la recursión con $\hat{A}[-1] = 0$, se tiene que

$$\hat{A}[0] = \frac{\sigma_A^2}{\sigma^2 + \sigma_A^2} \left(\frac{\sigma^2}{\sigma_A^2} \hat{A}[-1] + x[0] \right) = \frac{\sigma_A^2}{\sigma^2 + \sigma_A^2} x[0].$$

Luego,

$$\begin{aligned} \hat{A}[1] &= \frac{\sigma_A^2}{\sigma^2 + 2\sigma_A^2} \left(\frac{\sigma^2 + \sigma_A^2}{\sigma_A^2} \hat{A}[0] + x[1] \right) \\ &= \frac{\sigma_A^2}{\sigma^2 + 2\sigma_A^2} \left(\frac{\sigma^2 + \sigma_A^2}{\sigma_A^2} \frac{\sigma_A^2}{\sigma^2 + \sigma_A^2} x[0] + x[1] \right) \\ &= \frac{\sigma_A^2}{\sigma^2 + 2\sigma_A^2} (x[0] + x[1]). \end{aligned}$$

Continuando, para $n = 2$ se tiene que

$$\begin{aligned}\hat{A}[2] &= \frac{\sigma_A^2}{\sigma^2 + 3\sigma_A^2} \left(\frac{\sigma^2 + 2\sigma_A^2}{\sigma_A^2} \hat{A}[1] + x[2] \right) \\ &= \frac{\sigma_A^2}{\sigma^2 + 3\sigma_A^2} \left(\frac{\sigma^2 + 2\sigma_A^2}{\sigma_A^2} \frac{\sigma_A^2}{\sigma^2 + 2\sigma_A^2} (x[0] + x[1]) + x[2] \right) \\ &= \frac{\sigma_A^2}{\sigma^2 + 3\sigma_A^2} (x[0] + x[1] + x[2]),\end{aligned}$$

y así sucesivamente. En el caso general, se tiene que

$$\hat{A}[n] = \frac{\sigma_A^2}{\sigma^2 + (n+1)\sigma_A^2} \sum_{k=0}^n x[k]$$

resultando en

$$\hat{A}[n] = \frac{\sigma_A^2}{\frac{\sigma^2}{n+1} + \sigma_A^2} \left(\frac{1}{n+1} \sum_{k=0}^n x[k] \right).$$

Como es de esperar, los resultados coinciden con los obtenidos en la sección 12.4.

13.8.11. Problema 11

En este problema se implementa el filtro de Kalman con estado escalar–observación escalar (ecuaciones 13.26–13.30). Se recomienda la solución usando una computadora. Si $a = 0.9$, σ_u^2 , $\mu_s = 0$, $\sigma_s = 1$, calcular la ganancia de Kalman y el MSE mínimo si

- $\sigma_n^2 = (0.9)^n$
- $\sigma_n^2 = 1$
- $\sigma_n^2 = (1.1)^n$

Explicar los resultados. Empleando una computadora, generar una realización de la señal y el ruido y aplicar el filtro para estimar la señal en los tres casos. Graficar la señal y la estimación del filtro de Kalman.

Solución Empleando una computadora, se calculó la ganancia de Kalman y el MSE mínimo a partir de la recursión dada por las ecuaciones 13.27, 13.28 y 13.30,

$$\begin{aligned}M[n|n-1] &= a^2 M[n-1|n-1] + \sigma_u^2 \\ K[n] &= \frac{M[n|n-1]}{\sigma_n^2 + M[n|n-1]} \\ M[n|n] &= (1 - K[n])M[n|n-1],\end{aligned}$$

donde la inicialización se realiza con $M[-1|-1] = \sigma_s^2$. En la figura 13.23 se muestran los resultados para los tres casos. En el caso a. con $\sigma_n^2 = (0.9)^n$, las observaciones $x[n]$ son cada vez menos ruidosas cuando crece n , por lo que el estimador MMSE debería tender a $\hat{s}[n|n] = x[n]$. Efectivamente es lo que ocurre. Se observa que $K[n] \rightarrow 1$, y por la ecuación 13.29, eso implica que $\hat{s}[n|n] \rightarrow x[n]$. De forma acorde, el estimador de la señal mejora con n ya que $M[n|n] \rightarrow 0$. Para el caso b. con $\sigma_n^2 = (1.1)^n$, se observa que $K[n] \rightarrow 0$. Esto se debe a que las muestras son cada vez mas ruidosas, por lo que el estimador considera cada vez menos a las observaciones cuando crece n . Además, con $K[n] \rightarrow 0$, de la ecuaciones 13.29 y 13.26, se cumple que

$$\hat{s}[n|n] \rightarrow \hat{s}[n|n-1] = a\hat{s}[n-1|n-1] = 0.9\hat{s}[n-1|n-1],$$

lo que implica que $\hat{s}[n|n] \rightarrow 0 = E(s[n])$. De esta forma, el MSE es

$$M[n|n] = E[(s[n] - \hat{s}[n|n])^2] \rightarrow E(s^2[n]) = \frac{\sigma_u^2}{1-a},$$

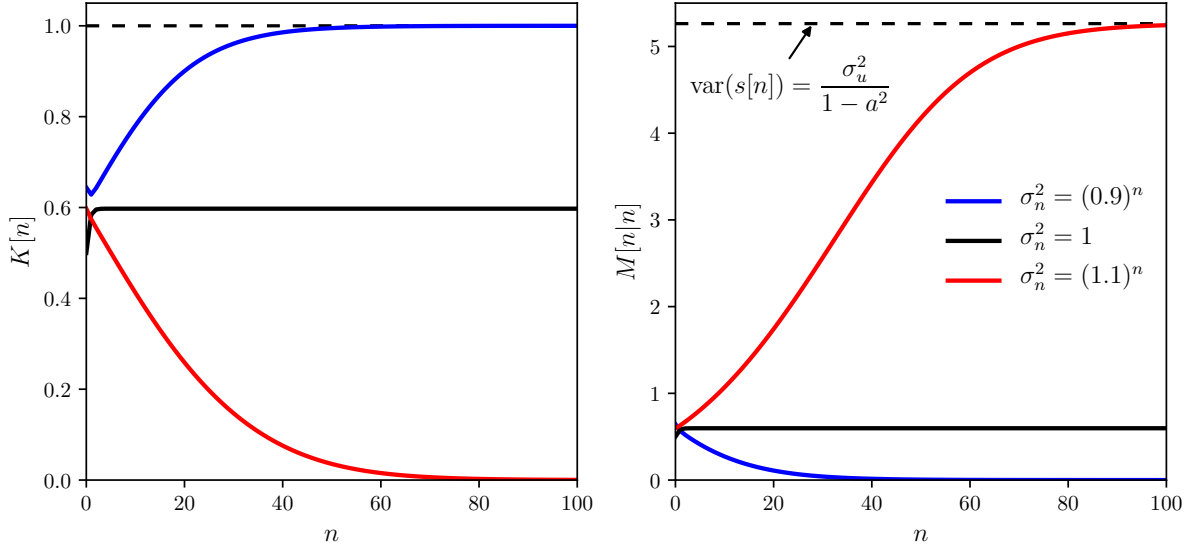


Figura 13.23: Ganancia de Kalman y MSE mínimo del filtro de Kalman para la estimación de un proceso de Gauss-Markov de primer orden con ruido de observación de varianza σ_n^2 .

que es la varianza en estado estacionario del proceso $s[n]$, como indica la ecuación 13.6 de la ACF en estado estacionario evaluada en $k = 0$. Finalmente, en el caso b. con $\sigma_n^2 = 1$, se observa que luego de par de iteraciones se alcanza un estado estacionario. Este es el filtro de Kalman de estado estacionario o filtro de Wiener que se analizó en la sección 13.4. El MMSE en estado estacionario se obtiene resolviendo el polinomio de la ecuación 13.34, y para los valores dados es $M[\infty] = K[\infty] = 0.5974$, acorde con la figura. En la figura 13.24 se muestra el resultado del filtrado de una realización del proceso en los tres casos.

13.8.12. Problema 12

Para el filtro de Kalman con estado escalar-observación escalar, se asume que $\sigma_n^2 = 0$ para todo n , de forma de que se observa directamente $s[n]$. Encontrar la secuencia de innovación. ¿Es blanca?

Solución La secuencia de innovación es

$$\tilde{x}[n] = x[n] - \hat{x}[n|n-1].$$

Como en este caso se observa directamente $s[n]$, se cumple que $x[n] = s[n]$. Además, siempre se cumple que $\hat{x}[n|n-1] = \hat{s}[n|n-1]$ (ver la deducción del filtro de Kalman en la sección 13.3). De esta forma, la innovación es

$$\tilde{x}[n] = s[n] - \hat{s}[n|n-1].$$

Como $\sigma_n^2 = 0$, de la ecuación 13.28 se obtiene que $K[n] = 1$ para todo n , y la ecuación 13.29 queda

$$\hat{s}[n|n] = x[n] = s[n],$$

y de la ecuación 13.26 se tiene que

$$\hat{s}[n|n-1] = a\hat{s}[n-1|n-1] = as[n-1].$$

Se concluye que la innovación es

$$\tilde{x}[n] = s[n] - as[n-1] = u[n],$$

que es el ruido de excitación, y por lo tanto, es blanca.

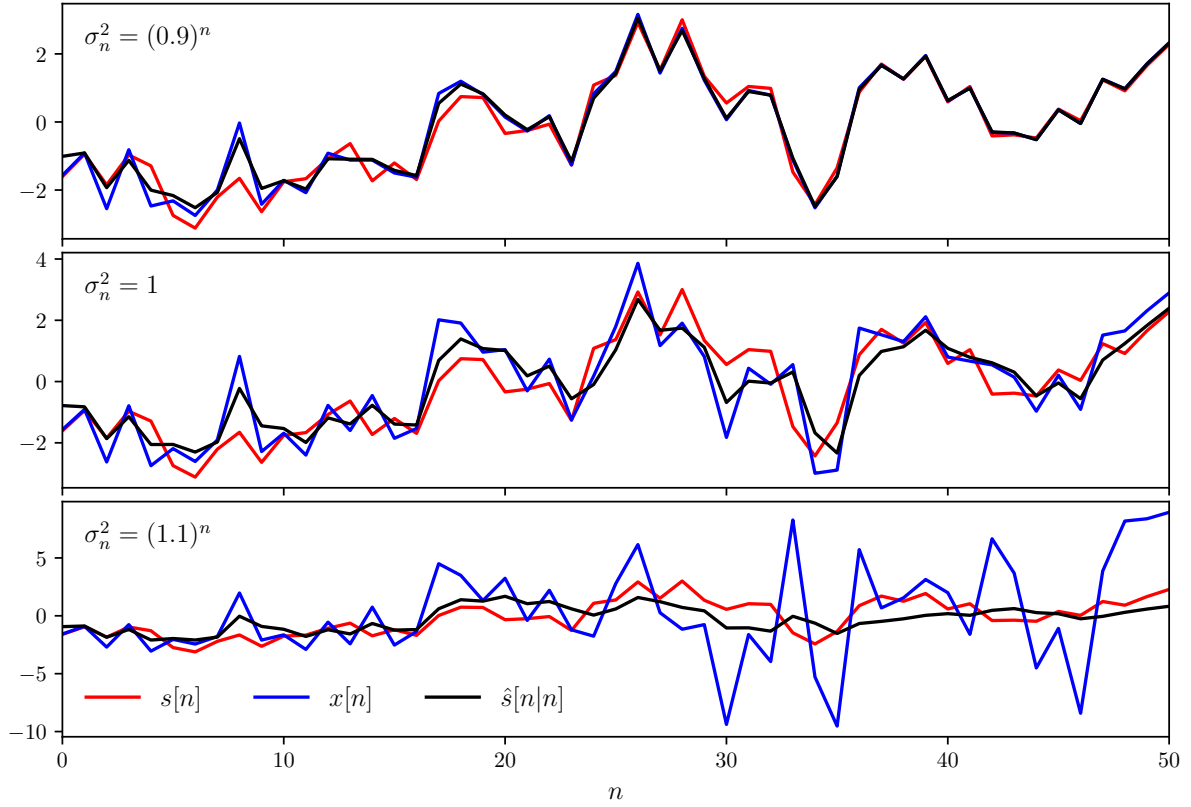


Figura 13.24: Estimación de un proceso de Gauss-Markov de primer orden con ruido de observación σ_n^2 mediante el filtro de Kalman. $s[n]$ es el proceso a estimar y $x[n]$ es la señal observada, $x[n] = s[n] + w[n]$, donde $w[n]$ es ruido no correlacionado con PDF $\mathcal{N}(0, \sigma_n^2)$. El ruido se generó como $w[n] = \sigma_n * v[n]$, donde $v[n]$ es WGN de varianza unidad. El proceso $s[n]$ a estimar es el mismo en los tres casos y el ruido $w[n]$ se generó a partir de la misma realización de $v[n]$ en los tres casos. La estimación del filtro de Kalman es $\hat{s}[n|n]$.

13.8.13. Problema 13

En este problema se muestra que se obtiene el mismo conjunto de ecuaciones para el filtro de Kalman con estado escalar-observación escalar si $E(s[-1]) \neq 0$. Para hacerlo, sean $s'[n] = s[n] - E(s[n])$ y $x'[n] = x[n] - E(x[n])$, de forma que las ecuaciones 13.26–13.30 se cumplen para $s'[n]$. Determinar las ecuaciones para $s[n]$. Recordar que el estimador MMSE de $\theta + c$ para una constante c es $\hat{\theta} + c$, donde $\hat{\theta}$ es el estimador MMSE de θ .

Solución El estimador MMSE de $s'[n] = s[n] - E(s[n])$ es $\hat{s}'[n] = \hat{s}[n] - E(s[n])$, donde $\hat{s}[n]$ es el estimador MMSE de $s[n]$, y como $E(s[n])$ es conocido, el MSE mínimo es el mismo para los estimadores de $s'[n]$ y $s[n]$. Por lo tanto, las ecuaciones 13.27 y 13.30 de $M[n|n-1]$ y $M[n|n]$ son las mismas para $s'[n]$ y $s[n]$. Como consecuencia, la ecuación 13.28 de $K[n]$ tampoco cambia. La ecuación de predicción es

$$\hat{s}'[n|n-1] = a\hat{s}'[n-1|n-1],$$

y para $s[n]$ queda

$$\hat{s}[n|n-1] - E(s[n]) = a[\hat{s}[n-1|n-1] - E(s[n-1])].$$

Pero la recursión de la media, de la ecuación 13.7 es,

$$E(s[n]) = aE(s[n-1])$$

por lo que la ecuación de predicción queda

$$\hat{s}[n|n-1] = a\hat{s}[n-1|n-1],$$

que es la misma que la ecuación de predicción de $s'[n]$. La ecuación de corrección 13.29 es

$$\hat{s}'[n|n] = \hat{s}'[n|n-1] + K[n](x'[n] - \hat{s}'[n|n-1]).$$

y para $s[n]$ queda

$$\hat{s}[n|n] - E(s[n]) = \hat{s}[n|n-1] - E(s[n]) + K[n](x[n] - E(x[n]) - (\hat{s}[n|n-1] - E(s[n]))).$$

Teniendo en cuenta que

$$x[n] = s[n] + w[n] \quad \Rightarrow \quad E(x[n]) = E(s[n])$$

se obtiene que

$$\hat{s}[n|n] = \hat{s}[n|n-1] + K[n](x[n] - \hat{s}[n|n-1]),$$

que también es la misma que la ecuación de corrección de $s'[n]$.

13.8.14. Problema 14

Probar que para el filtro de Kalman con estado escalar-observación escalar se cumple que

$$M[n|n-1] > M[n-1|n-1]$$

para n suficientemente largo o en estado estacionario. ¿Porqué esto es razonable?

Solución Si con $n \rightarrow \infty$ se alcanza un estado estacionario, $M[n|n]$ y $M[n|n-1]$ son constantes. De la ecuación 13.30 se observa que

$$M[n|n] = (1 - K[n])M[n|n-1] < M[n|n-1],$$

ya que $K[n] < 1$ (ver la ecuación 13.28). Pero en estado estacionario, $M[n|n] \approx M[n-1|n-1]$, y por lo tanto

$$M[n|n-1] > M[n-1|n-1].$$

Esto es razonable, ya que es más fácil predecir a $s[n-1]$ que a $s[n]$ en base a $\{x[0], \dots, x[n-1]\}$, y como consecuencia, $M[n-1|n-1] < M[n|n-1]$.

13.8.15. Problema 15

Probar que el predictor óptimo de paso l para un proceso $s[n]$ de Gauss-Markov escalar es

$$\hat{s}[n+l|n] = a^l \hat{s}[n|n]$$

donde los estimadores $\hat{s}[n|n]$ y $\hat{s}[n+l|n]$ son en base a $\{x[0], \dots, x[n]\}$.

Solución Se dispone del estimador $\hat{s}[n|n]$ de $s[n]$ en base a $\{x[0], \dots, x[n]\}$. Se asume que $\sigma_m^2 \rightarrow \infty$ para $m = n+1, \dots, n+l$. De esta forma, el filtro de Kalman no emplea las muestras $\{x[n+1], \dots, x[n+l]\}$ para estimar $\{s[n+1], \dots, s[n+l]\}$ por ser ruidosas, ya que de la ecuación 13.28, $K[m] \rightarrow 0$ para $m = n+1, \dots, n+l$. En el paso $n+1$, de la ecuación 13.26, la predicción es

$$\hat{s}[n+1|n] = a\hat{s}[n|n].$$

Como $\sigma_{n+1}^2 \rightarrow \infty$, de la ecuación 13.28, $K[n+1] \rightarrow 0$, por lo que el estimador es

$$\hat{s}[n+1|n+1] = \hat{s}[n+1|n] = a\hat{s}[n|n].$$

Como el filtro de Kalman no emplea la muestra $x[n+1]$, se cumple que $\hat{s}[n+1|n+1] = \hat{s}[n+1|n]$, resultando en que

$$\hat{s}[n+1|n] = a\hat{s}[n|n].$$

De forma análoga, en el paso $n+2$, la predicción es

$$\hat{s}[n+2|n+1] = a\hat{s}[n+1|n+1]$$

y como $K[n+2] \rightarrow 0$, el estimador es

$$\hat{s}[n+2|n+2] = \hat{s}[n+2|n+1] = a\hat{s}[n+1|n+1] = a^2\hat{s}[n|n],$$

donde en la última igualdad se empleó el estimador obtenido en el paso $n+1$. Luego, como el filtro de Kalman no emplea las muestras $\{x[n+1], x[n+2]\}$ por ser infinitamente ruidosas, $\hat{s}[n+2|n+2] = \hat{s}[n+2|n]$, concluyendo que

$$\hat{s}[n+2|n] = a^2\hat{s}[n|n].$$

Continuando de forma análoga en los pasos sucesivos, en el paso $n+l$ se obtiene que

$$\hat{s}[n+l|n] = a^l\hat{s}[n|n].$$

13.8.16. Problema 16

Encontrar la función de transferencia del filtro de Kalman con estado escalar-observación escalar en estado estacionario si $a = 0.8$, $\sigma_u^2 = 1$ y $\sigma^2 = 1$. Dar su forma en el dominio del tiempo como una ecuación en recurrencia.

Solución La función de transferencia del filtro de Kalman en estado estacionario está dada por la ecuación 13.36 y es

$$\mathcal{H}_\infty(z) = \frac{K[\infty]}{1 - a(1 - K[\infty])z^{-1}},$$

donde para obtener $K[\infty]$, primero hay que obtener $M[\infty]$ empleando la ecuación 13.34, que es

$$M[\infty] = \frac{\sigma^2(a^2M[\infty] + \sigma_u^2)}{a^2M[\infty] + \sigma_u^2 + \sigma^2},$$

o

$$a^2M^2[\infty] + [\sigma_u^2 + \sigma^2(1 - a^2)]M[\infty] - \sigma_u^2\sigma^2 = 0.$$

Con los valores $a = 0.8$, $\sigma_u^2 = 1$ y $\sigma^2 = 1$, el polinomio es

$$0.64M^2[\infty] + 1.36M[\infty] - 1 = 0$$

y las raíces son -2.7031 y 0.5781. Como el MSE debe ser positivo, se concluye que $M[\infty] = 0.5781$. Luego, de la ecuación 13.27, se tiene que

$$M_p[\infty] = a^2M[\infty] + \sigma_u^2,$$

que para los valores dados es

$$M_p[\infty] = 0.64 \times 0.5781 + 1 = 1.3700.$$

Finalmente, de la ecuación 13.28 se tiene que

$$K[\infty] = \frac{M_p[\infty]}{\sigma^2 + M_p[\infty]},$$

que queda

$$K[\infty] = \frac{1.3700}{1 + 1.3700} = 0.5781.$$

Por lo tanto, la función de transferencia es

$$\mathcal{H}_\infty(z) = \frac{0.5781}{1 - 0.8(1 - 0.5781)z^{-1}} = \frac{0.5781}{1 - 0.3376z^{-1}}.$$

La ecuación en recurrencia en el dominio del tiempo en estado estacionario esta dada por 13.35

$$\hat{s}[n|n] = a(1 - K[\infty])\hat{s}[n-1|n-1] + K[\infty]x[n].$$

Sustituyendo por los valores queda

$$\hat{s}[n|n] = 0.3376\hat{s}[n-1|n-1] + 0.5781x[n].$$

13.8.17. Problema 17

Para el filtro de Kalman con estado escalar-observación escalar, sea $a = 0.9$, $\sigma_u^2 = 1$ y $\sigma^2 = 1$. Encontrar la ganancia y el MSE mínimo en estado estacionario mediante la ejecución del filtro hasta la convergencia, es decir, mediante el cálculo recursivo de las ecuaciones 13.27, 13.28 y 13.30. Comparar los resultados con los obtenidos en la sección 13.4.

Solución El resultado de la ejecución del filtro se muestra en la figura 13.25, donde se grafica la evolución de $K[n]$ y $M[n|n]$. Las condiciones iniciales empleadas fueron $s[-1] = 0$ y $M[-1| - 1] = 1$. Se observa que en este caso se cumple que $K[n] = M[n|n]$. Efectivamente, sustituyendo la ecuación 13.28 en la ecuación 13.30, se tiene que

$$\begin{aligned} M[n|n] &= \left(1 - \frac{M[n|n-1]}{\sigma^2 + M[n|n-1]}\right) M[n|n-1] \\ &= \frac{\sigma^2 M[n|n-1]}{\sigma^2 + M[n|n-1]}, \end{aligned}$$

y como $\sigma^2 = 1$, $M[n|n]$ coincide con $K[n]$ (ver la ecuación 13.28). En la sección 13.4 se encontró que con

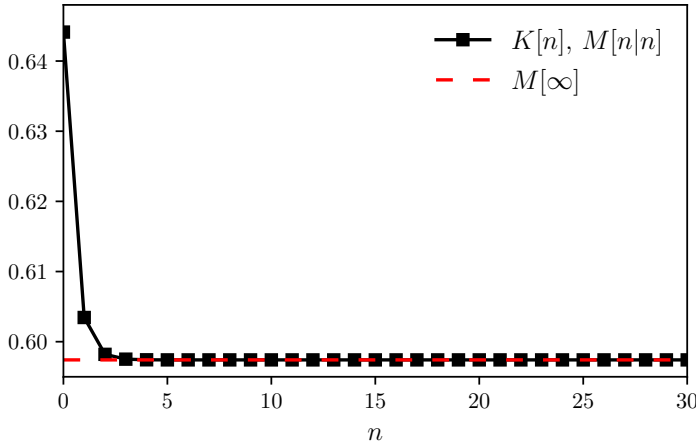


Figura 13.25: Evolución de la ganancia $K[n]$ y el MSE mínimo $M[n|n]$ y comparación con el MSE mínimo $M[\infty]$ en estado estacionario.

estos valores, el MSE mínimo en estado estacionario es $M[\infty] = 0.5974$, y como se aprecia en la figura, $M[n|n]$ alcanza ese valor rápidamente.

13.8.18. Problema 18

Asúmase que se observan los datos

$$x[k] = Ar^k + w[k]$$

para $k = 0, \dots, n$, donde A es la realización de un variable aleatoria con PDF $\mathcal{N}(\mu_a, \sigma_A^2)$, $0 < r < 1$, y las muestras $w[k]$ son WGN de varianza σ^2 . Adicionalmente se asume que A es independiente de las muestras $w[k]$. Encontrar el estimador MMSE secuencial de A en base a $\{x[0], \dots, x[n]\}$.

Solución Para este problema, la ecuación de estado y la ecuación de observación son

$$\begin{aligned} A[n] &= A[n-1] \\ x[n] &= A[n]r^n + w[n]. \end{aligned}$$

Si bien el estado y la observación son escalares, el factor que multiplica al estado en la ecuación de observación, r^n , es variable en el tiempo. Por lo tanto, se empleará el filtro de Kalman mas general (ver la sección 13.5), que considera este caso. De esta forma, el estimador MMSE secuencial está dado por las ecuaciones 13.43-13.47 con

$$\mathbf{s}[n] = A[n], \quad \mathbf{A} = 1, \quad \mathbf{Q} = 0, \quad \mathbf{H}[n] = r^n \quad \mathbf{C}[n] = \sigma^2,$$

donde se observa que todas las variables son escalares. Se obtiene que el filtro de Kalman es

$$\begin{aligned} \hat{A}[n|n-1] &= \hat{A}[n-1|n-1] \\ M[n|n-1] &= M[n-1|n-1] \\ K[n] &= \frac{M[n|n-1]r^n}{\sigma^2 + r^{2n}M[n|n-1]} \\ \hat{s}[n|n] &= \hat{s}[n|n-1] + K[n](x[n] - r^n\hat{A}[n|n-1]) \\ M[n|n] &= (1 - K[n]r^n)M[n|n-1]. \end{aligned}$$

En este caso, se puede omitir la etapa de predicción, y cambiando la notación, la recursión queda

$$\begin{aligned} K[n] &= \frac{M[n-1]r^n}{\sigma^2 + r^{2n}M[n-1]} \\ \hat{s}[n] &= \hat{s}[n-1] + K[n](x[n] - r^n\hat{A}[n-1]) \\ M[n] &= (1 - K[n]r^n)M[n-1]. \end{aligned}$$

Los valores para la inicialización son $\hat{A}[-1] = \mu_A$ y $M[-1] = \sigma_A^2$.

13.8.19. Problema 19

Se considera el filtro de Kalman con estado vectorial–observación vectorial, en el cual se asume que $\mathbf{H}[n]$ es invertible. Si una observación particular no es ruidosa de forma tal que $\mathbf{C}[n] = \mathbf{0}$, encontrar $\hat{\mathbf{s}}[n|n]$ y explicar los resultados. ¿Qué ocurre si $\mathbf{C}[n] \rightarrow \infty$?

Solución Si la muestra n -ésima no es ruidosa, se cumple que $\mathbf{C}[n] = \mathbf{0}$ y de la ecuación 13.45, la ganancia de Kalman queda

$$\begin{aligned} \mathbf{K}[n] &= \mathbf{M}[n|n-1]\mathbf{H}^T[n](\mathbf{H}[n]\mathbf{M}[n|n-1])\mathbf{H}^T[n])^{-1} \\ &= \mathbf{M}[n|n-1]\mathbf{H}^T[n]\mathbf{H}^{T-1}[n]\mathbf{M}^{-1}[n|n-1]\mathbf{H}^{-1}[n] \\ &= \mathbf{H}^{-1}[n]. \end{aligned}$$

Sustituyendo este resultado en la ecuación 13.46, la estimación es

$$\begin{aligned} \hat{\mathbf{s}}[n|n] &= \hat{\mathbf{s}}[n|n-1] + \mathbf{H}^{-1}[n](\mathbf{x}[n] - \mathbf{H}[n]\hat{\mathbf{s}}[n|n-1]) \\ &= \mathbf{H}^{-1}[n]\mathbf{x}[n]. \end{aligned}$$

Se observa que la estimación, por no depender de $\hat{\mathbf{s}}[n|n-1]$, ignora todas las observaciones previas y solo considera la observación actual. Además, como la observación no tiene ruido, en la ecuación de observación $\mathbf{w}[n] = \mathbf{0}$ y por lo tanto,

$$\mathbf{x}[n] = \mathbf{H}[n]\mathbf{s}[n] \quad \Rightarrow \quad \mathbf{s}[n] = \mathbf{H}^{-1}[n]\mathbf{x}[n] = \hat{\mathbf{s}}[n|n],$$

por lo que la estimación es exacta.

En el caso en que $\mathbf{C}[n] \rightarrow \infty$, la ganancia de Kalman cumple que $\mathbf{K}[n] \rightarrow \mathbf{0}$ y la estimación es

$$\hat{\mathbf{s}}[n|n] = \hat{\mathbf{s}}[n|n-1],$$

lo que implica que la observación es ignorada por el estimador.

13.8.20. Problema 20

Probar que el predictor óptimo de paso l para un proceso $\mathbf{s}[n]$ de Gauss-Markov es

$$\hat{\mathbf{s}}[n+l|n] = \mathbf{A}^l \hat{\mathbf{s}}[n|n]$$

donde $\hat{\mathbf{s}}[n|n]$ y $\hat{\mathbf{s}}[n+l|n]$ son en base a $\{\mathbf{x}[0], \dots, \mathbf{x}[n]\}$. Emplear el filtro de Kalman de estado vectorial-observación vectorial.

Solución La deducción es completamente análoga al caso escalar del problema de la sección 13.8.15. Se dispone del estimador $\hat{\mathbf{s}}[n|n]$ de $\mathbf{s}[n]$ en base a $\{\mathbf{x}[0], \dots, \mathbf{x}[n]\}$. Se asume que $\mathbf{C}[m] \rightarrow \infty$ para $m = n+1, \dots, n+l$. De esta forma, de la ecuación 13.45, $\mathbf{K}[m] \rightarrow \mathbf{0}$ para $m = n+1, \dots, n+l$, lo que implica que el filtro de Kalman no emplea las muestras $\{\mathbf{x}[n+1], \dots, \mathbf{x}[n+l]\}$ para estimar $\{\mathbf{s}[n+1], \dots, \mathbf{s}[n+l]\}$, como indica la ecuación 13.46. En estas condiciones, las ecuaciones del filtro de Kalman quedan

$$\begin{aligned}\hat{\mathbf{s}}[n|n-1] &= \mathbf{A} \hat{\mathbf{s}}[n-1|n-1] \\ \hat{\mathbf{s}}[n|n] &= \hat{\mathbf{s}}[n|n-1],\end{aligned}$$

es decir,

$$\hat{\mathbf{s}}[n|n] = \mathbf{A} \hat{\mathbf{s}}[n-1|n-1].$$

De esta forma, se cumple que

$$\hat{\mathbf{s}}[n+l|n+l] = \mathbf{A}^l \hat{\mathbf{s}}[n|n],$$

y como no se emplean las muestras $\{\mathbf{x}[n+1], \dots, \mathbf{x}[n+l]\}$, se concluye que

$$\hat{\mathbf{s}}[n+l|n] = \mathbf{A}^l \hat{\mathbf{s}}[n|n].$$

13.8.21. Problema 21

En este problema, se configura un filtro de Kalman extendido para la aplicación de seguimiento de frecuencia. Específicamente, se desea seguir la frecuencia de una senoide en ruido. Se asume que la frecuencia se modela como

$$f_0[n] = a f_0[n-1] + u[n], \quad n \geq 0,$$

donde $u[n]$ es WGN de varianza σ_u^2 y $f_0[-1]$ tiene PDF $\mathcal{N}(\mu_{f_0}, \sigma_{f_0}^2)$ y es independiente de $u[n]$. Los datos observados son

$$x[n] = \cos(2\pi f_0[n]) + w[n], \quad n \geq 0,$$

donde $w[n]$ es WGN de varianza σ^2 y es independiente de $u[n]$ y $f_0[-1]$. Escribir las ecuaciones del filtro de Kalman extendido para este problema.

Solución En este problema, el estado y la observación son escalares. Además, la ecuación de estado es lineal y la ecuación de observaciones es no lineal de la forma

$$x[n] = h(f_0[n]) + w[n], \quad \text{con} \quad h(f_0[n]) = \cos(2\pi f_0[n]),$$

y

$$H[n] = \left. \frac{\partial h}{\partial f_0[n]} \right|_{f_0[n] = \hat{f}_0[n|n-1]} = -2\pi \sin(2\pi f_0[n]) \Big|_{f_0[n] = \hat{f}_0[n|n-1]},$$

resultando en que

$$H[n] = -2\pi \sin(2\pi \hat{f}_0[n|n-1]).$$

Con estos valores y

$$\mathbf{A}[n-1] = a, \quad \mathbf{Q} = \sigma_u^2, \quad \mathbf{B} = 1, \quad \mathbf{C}[n] = \sigma^2,$$

las ecuaciones 13.62-13.66 del filtro de Kalman extendido quedan

$$\begin{aligned}\hat{f}_0[n|n-1] &= a\hat{f}_0[n-1|n-1] \\ M[n|n-1] &= a^2 M[n-1|n-1] + \sigma_u^2 \\ K[n] &= \frac{-2\pi \sin(2\pi \hat{f}_0[n|n-1]) M[n|n-1]}{\sigma^2 + 4\pi^2 \sin^2(2\pi \hat{f}_0[n|n-1]) M[n|n-1]} \\ \hat{f}_0[n|n] &= \hat{f}_0[n|n-1] + K[n] [x[n] - \cos(2\pi \hat{f}_0[n|n-1])] \\ M[n|n] &= [1 - 2\pi \sin(2\pi \hat{f}_0[n|n-1]) K[n]] M[n|n-1],\end{aligned}$$

y la inicialización se realiza con $\hat{f}_0[-1|-1] = \mu_{f_0}$ y $M[-1|-1] = \sigma_{f_0}^2$.

13.8.22. Problema 22

Para el modelo de la posición del vehículo del ejemplo de la sección 13.7.2 se tiene que

$$v_x[n] = v_x[n-1] + u_x[n],$$

donde $u_x[n]$ es WGN de varianza σ_u^2 . Encontrar la varianza de $v_x[n]$ para mostrar que crece con n . ¿Cuál sería un modelo mas apropiado para $v_x[n]$?

Solución Resolviendo la recursión, se observa que

$$\begin{aligned}v_x[0] &= v_x[-1] + u_x[0] \\ v_x[1] &= v_x[0] + u_x[1] = v_x[-1] + u_x[0] + u_x[1]\end{aligned}$$

y así sucesivamente. En el caso general se tiene que

$$v_x[n] = v_x[-1] + \sum_{k=0}^n u_x[k].$$

Considerando que como $u_x[n]$ es WGN, las muestras son independientes, y también son independientes de $v[-1]$. Por lo tanto, se cumple que

$$\begin{aligned}\text{var}(v_x[n]) &= \text{var}(v_x[-1]) + \sum_{k=0}^n \text{var}(u_x[k]) \\ &= \sigma_v^2 + (n+1)\sigma_u^2.\end{aligned}$$

Se observa que la varianza de $v_x[n]$ crece con n , y en particular,

$$\lim_{n \rightarrow \infty} \text{var}(v_x[n]) = \lim_{n \rightarrow \infty} \sigma_v^2 + (n+1)\sigma_u^2 = \infty.$$

Un modelo mas apropiado sería

$$v_x[n] = av_x[n-1] + u_x[n],$$

con $a < 1$ pero cercano a 1. De esta forma, resolviendo la recursión se tiene que

$$\begin{aligned}v_x[0] &= av_x[-1] + u_x[0] \\ v_x[1] &= av_x[0] + u_x[1] = a(av_x[-1] + u_x[0]) + u_x[1] = a^2 v_x[-1] + au_x[0] + u_x[1] \\ v_x[2] &= av_x[1] + u_x[2] = a(a^2 v_x[-1] + au_x[0] + u_x[1]) + u_x[2] = a^3 v_x[-1] + a^2 u_x[0] + au_x[1] + u_x[2]\end{aligned}$$

y así sucesivamente. En el caso general se cumple que

$$v_x[n] = a^{n+1} v_x[-1] + \sum_{k=0}^n a^k u_x[n-k].$$

Por lo tanto, considerando nuevamente la independencia de las variables aleatorias, se cumple que

$$\begin{aligned}\text{var}(v_x[n]) &= a^{2(n+1)} \text{var}(v_x[-1]) + \sum_{k=0}^n a^{2k} \text{var}(u_x[n-k]) \\ &= a^{2(n+1)} \sigma_v^2 + \sigma_u^2 \sum_{k=0}^n a^{2k}.\end{aligned}$$

De esta forma, se cumple que

$$\lim_{n \rightarrow \infty} \text{var}(v_x[n]) = \lim_{n \rightarrow \infty} a^{2(n+1)} \sigma_v^2 + \sigma_u^2 \sum_{k=0}^n a^{2k} = \frac{\sigma_u^2}{1-a^2},$$

es decir, la varianza de $v_x[n]$ tiende a una constante cuando $n \rightarrow \infty$.

13.8.23. Problema 23

Para el filtro de Kalman con estado escalar-observación escalar, encontrar una expresión que relacione $M[n|n]$ con $M[n-1|n-1]$. Sea $a = 0.9$, $\sigma_u^2 = 1$, y $\sigma_n^2 = n+1$. Si $M[-1|-1] = 1$, determinar $M[n|n]$ para $n \geq 0$. Explicar los resultados.

Solución De las ecuaciones 13.26-13.30 se tiene que

$$\begin{aligned}M[n|n-1] &= a^2 M[n-1|n-1] + \sigma_u^2 \\ K[n] &= \frac{M[n|n-1]}{\sigma_n^2 + M[n|n-1]} \\ M[n|n] &= (1 - K[n])M[n|n-1].\end{aligned}$$

Por lo tanto

$$\begin{aligned}M[n|n] &= \left(1 - \frac{M[n|n-1]}{\sigma_n^2 + M[n|n-1]}\right) M[n|n-1] \\ &= \frac{\sigma_n^2 M[n|n-1]}{\sigma_n^2 + M[n|n-1]} \\ &= \frac{\sigma_n^2 (a^2 M[n-1|n-1] + \sigma_u^2)}{\sigma_n^2 + a^2 M[n-1|n-1] + \sigma_u^2},\end{aligned}$$

Con los valores dados, la recursión es

$$M[n|n] = \frac{(n+1)(0.81M[n-1|n-1] + 1)}{n+2 + 0.81M[n-1|n-1]},$$

con $M[-1|-1] = 1$. En la figura 13.26 se muestra la evolución del MSE. Como las muestras observadas son cada vez mas ruidosas debido al incremento de la varianza del ruido de observación como $\sigma_n^2 = n+1$, el MSE crece.

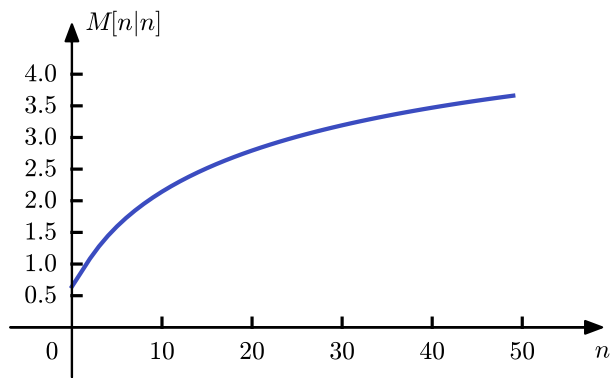


Figura 13.26: Evolución de MSE $M[n|n]$ con n en el filtro de Kalman.

Apéndice A

Identidad matricial de Woodbury

La *identidad matricial de Woodbury* o *lema de inversión de matrices* establece que¹

$$(\mathbf{A} + \mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{DA}^{-1}\mathbf{B} + \mathbf{C}^{-1})^{-1}\mathbf{DA}^{-1}, \quad (\text{A.1})$$

donde \mathbf{A} es $n \times n$, \mathbf{B} es $n \times m$, \mathbf{C} es $m \times m$ y \mathbf{D} es $m \times n$, y las matrices inversas de la identidad existen.

Una forma de demostrar el lema es multiplicar $(\mathbf{A} + \mathbf{BCD})$ por la matriz inversa establecida por el lema y obtener la matriz identidad,

$$\begin{aligned} (\mathbf{A} + \mathbf{BCD})(\mathbf{A} + \mathbf{BCD})^{-1} &\stackrel{(a)}{=} (\mathbf{A} + \mathbf{BCD}) [\mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{DA}^{-1}\mathbf{B} + \mathbf{C}^{-1})^{-1}\mathbf{DA}^{-1}] \\ &= \mathbf{AA}^{-1} + \mathbf{BCDA}^{-1} - [\mathbf{AA}^{-1}\mathbf{B}(\mathbf{DA}^{-1}\mathbf{B} + \mathbf{C}^{-1})^{-1}\mathbf{DA}^{-1} \\ &\quad + \mathbf{BCDA}^{-1}\mathbf{B}(\mathbf{DA}^{-1}\mathbf{B} + \mathbf{C}^{-1})^{-1}\mathbf{DA}^{-1}] \\ &= \mathbf{I} + \mathbf{BCDA}^{-1} \\ &\quad - [\mathbf{B}(\mathbf{DA}^{-1}\mathbf{B} + \mathbf{C}^{-1})^{-1}\mathbf{DA}^{-1} + \mathbf{BCDA}^{-1}\mathbf{B}(\mathbf{DA}^{-1}\mathbf{B} + \mathbf{C}^{-1})^{-1}\mathbf{DA}^{-1}] \\ &= \mathbf{I} + \mathbf{BCDA}^{-1} - [(\mathbf{B} + \mathbf{BCDA}^{-1}\mathbf{B})(\mathbf{DA}^{-1}\mathbf{B} + \mathbf{C}^{-1})^{-1}\mathbf{DA}^{-1}] \\ &= \mathbf{I} + \mathbf{BCDA}^{-1} - [\mathbf{BC}(\mathbf{C}^{-1} + \mathbf{DA}^{-1}\mathbf{B})(\mathbf{DA}^{-1}\mathbf{B} + \mathbf{C}^{-1})^{-1}\mathbf{DA}^{-1}] \\ &= \mathbf{I} + \mathbf{BCDA}^{-1} - \mathbf{BCDA}^{-1} \\ &= \mathbf{I}, \end{aligned}$$

y como en (a) se empleó la identidad que se quería demostrar, se concluye que efectivamente se cumple.

Otra forma de demostrar la identidad es la siguiente,

$$\begin{aligned} \mathbf{A}^{-1} &= (\mathbf{A} + \mathbf{BCD})^{-1}(\mathbf{A} + \mathbf{BCD})\mathbf{A}^{-1} \\ &= (\mathbf{A} + \mathbf{BCD})^{-1}(\mathbf{I} + \mathbf{BCDA}^{-1}) \\ &= (\mathbf{A} + \mathbf{BCD})^{-1} + (\mathbf{A} + \mathbf{BCD})^{-1}\mathbf{BCDA}^{-1} \\ &\stackrel{(a)}{=} (\mathbf{A} + \mathbf{BCD})^{-1} + (\mathbf{C}^{-1}\mathbf{B}^{-1}\mathbf{A} + \mathbf{D})^{-1}\mathbf{DA}^{-1} \\ &\stackrel{(b)}{=} (\mathbf{A} + \mathbf{BCD})^{-1} + \mathbf{A}^{-1}\mathbf{B}(\mathbf{C}^{-1} + \mathbf{DA}^{-1}\mathbf{B})^{-1}\mathbf{DA}^{-1}, \end{aligned}$$

donde en (a) se notó que

$$(\mathbf{A} + \mathbf{BCD})^{-1}\mathbf{BC} = [(\mathbf{BC})^{-1}\mathbf{A} + (\mathbf{BC})^{-1}\mathbf{BCD}]^{-1} = (\mathbf{C}^{-1}\mathbf{B}^{-1}\mathbf{A} + \mathbf{D})^{-1}$$

y en (b) que

$$(\mathbf{C}^{-1}\mathbf{B}^{-1}\mathbf{A} + \mathbf{D})^{-1} = (\mathbf{B}^{-1}\mathbf{A})^{-1} [\mathbf{C}^{-1} + \mathbf{D}(\mathbf{B}^{-1}\mathbf{A})^{-1}]^{-1} = \mathbf{A}^{-1}\mathbf{B}(\mathbf{C}^{-1} + \mathbf{DA}^{-1}\mathbf{B})^{-1},$$

y despejando se obtiene la ecuación A.2.

¹Ver https://en.wikipedia.org/wiki/Woodbury_matrix_identity

En el caso particular en que $\mathbf{B} = \mathbf{u}$ es un vector $n \times 1$, $\mathbf{C} = 1$ y $\mathbf{D} = \mathbf{u}^T$, se cumple que

$$(\mathbf{A} + \mathbf{u}\mathbf{u}^T)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{u}(\mathbf{u}^T\mathbf{A}^{-1}\mathbf{u} + 1)^{-1}\mathbf{u}^T\mathbf{A}^{-1},$$

resultando en

$$(\mathbf{A} + \mathbf{u}\mathbf{u}^T)^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{u}\mathbf{u}^T\mathbf{A}^{-1}}{1 + \mathbf{u}^T\mathbf{A}^{-1}\mathbf{u}}. \quad (\text{A.2})$$

Apéndice B

Sumatoria finita de algunas secuencias sinusiodales

Se demostrará que

$$\begin{aligned}\frac{1}{N^{i+1}} \sum_{n=0}^{N-1} n^i \sin(4\pi f_0 n + 2\phi) &\approx 0 \\ \frac{1}{N^{i+1}} \sum_{n=0}^{N-1} n^i \cos(4\pi f_0 n + 2\phi) &\approx 0\end{aligned}\tag{B.1}$$

para $i = 0, 1, 2$.

En el problema de la sección 3.6.7 se demostraron ambas aproximaciones para el caso $i = 0$. Se demostrará a continuación las aproximaciones para el caso $i = 1$. Una forma de hacerlo, es partir de la ecuación

$$\frac{1}{N} \sum_{n=0}^{N-1} \sin(\alpha n + \beta) = \frac{\sin[\alpha(N-1)/2 + \beta] \sin(\alpha N/2)}{N \sin(\alpha/2)},$$

deducida en el problema de la sección 3.6.7, donde $\alpha = 4\pi f_0$ y $\beta = 2\phi$, y multiplicar por $1/N$ y diferenciar respecto a α ambos lados de la igualdad [15]. Previo a hacerlo, conviene notar que el lado derecho de la igualdad se puede expresar como

$$\frac{1}{N} \sum_{n=0}^{N-1} \sin(\alpha n + \beta) = \frac{\cos(\alpha/2 - \beta) - \cos[\alpha(2N-1)/2 + \beta]}{2N \sin(\alpha/2)},$$

donde se empleó que $2 \sin \theta \sin \varphi = \cos(\varphi - \theta) - \cos(\theta + \varphi)$. Al multiplicar por $1/N$ y diferenciar respecto a α , el lado izquierdo de la igualdad queda

$$\frac{\partial}{\partial \alpha} \left[\frac{1}{N^2} \sum_{n=0}^{N-1} \sin(\alpha n + \beta) \right] = \frac{1}{N^2} \sum_{n=0}^{N-1} n \cos(\alpha n + \beta)$$

y el lado derecho queda

$$\begin{aligned}
\frac{\partial}{\partial \alpha} \left[\frac{\cos(\alpha/2 - \beta) - \cos[\alpha(2N-1)/2 + \beta]}{2N^2 \sin(\alpha/2)} \right] &= \frac{\left\{ -\frac{1}{2} \sin\left(\frac{\alpha}{2} - \beta\right) + \frac{2N-1}{2} \sin\left[\frac{\alpha(2N-1)}{2} + \beta\right] \right\} 2N^2 \sin\left(\frac{\alpha}{2}\right)}{4N^4 \sin^2\left(\frac{\alpha}{2}\right)} \\
&\quad - \frac{\left\{ \cos\left(\frac{\alpha}{2} - \beta\right) - \cos\left[\frac{\alpha(2N-1)}{2} + \beta\right] \right\} N^2 \cos\left(\frac{\alpha}{2}\right)}{4N^4 \sin^2\left(\frac{\alpha}{2}\right)} \\
&\stackrel{(a)}{=} \frac{-\sin(\alpha/2 - \beta) \sin(\alpha/2) - \cos(\alpha/2 - \beta) \cos(\alpha/2)}{4N^2 \sin^2(\alpha/2)} \\
&\quad - \frac{\sin[\alpha(2N-1)/2 + \beta] \sin(\alpha/2)}{4N^2 \sin^2(\alpha/2)} \\
&\quad + \frac{\cos[\alpha(2N-1)/2 + \beta] \cos(\alpha/2)}{4N^2 \sin^2(\alpha/2)} \\
&\quad + \frac{2N \sin[\alpha(2N-1)/2 + \beta] \sin(\alpha/2)}{4N^2 \sin^2(\alpha/2)}
\end{aligned}$$

donde en (a) se hicieron algunas simplificaciones, se desarrollaron las factorizaciones y se reordenaron los sumandos del numerador. Los dos primeros sumandos del numerador son

$$-\sin(\alpha/2 - \beta) \sin(\alpha/2) - \cos(\alpha/2 - \beta) \cos(\alpha/2) = -\cos(\beta),$$

igualdad que proviene de la identidad trigonométrica $\cos(\alpha \pm \beta) = \cos \alpha \cos \beta \mp \sin \alpha \sin \beta$. De la misma forma, el tercer y cuarto sumando valen

$$-\sin[\alpha(2N-1)/2 + \beta] \sin(\alpha/2) + \cos[\alpha(2N-1)/2 + \beta] \cos(\alpha/2) = \cos(\alpha N + \beta),$$

y combinando estos dos resultados, se tiene que los primeros cuatro sumandos valen

$$-[\cos(\beta) - \cos(\alpha N + \beta)] = -2 \sin(\alpha N/2 + \beta) \sin(\alpha N/2),$$

donde nuevamente se usó que $2 \sin \theta \sin \varphi = \cos(\varphi - \theta) - \cos(\theta + \varphi)$ con $\varphi = \alpha N/2 + \beta$ y $\theta = \alpha N/2$. Por lo tanto, se llega a que

$$\frac{\partial}{\partial \alpha} \left[\frac{\cos(\alpha/2 - \beta) - \cos[\alpha(2N-1)/2 + \beta]}{2N^2 \sin(\alpha/2)} \right] = \frac{N \sin[\alpha(2N-1)/2 + \beta] \sin(\alpha/2) - \sin(\alpha N/2 + \beta) \sin(\alpha N/2)}{2N^2 \sin^2(\alpha/2)}.$$

Se concluye que

$$\frac{1}{N^2} \sum_{n=0}^{N-1} n \cos(\alpha n + \beta) = \frac{N \sin[\alpha(2N-1)/2 + \beta] \sin(\alpha/2) - \sin(\alpha N/2 + \beta) \sin(\alpha N/2)}{2N^2 \sin^2(\alpha/2)}. \quad (\text{B.2})$$

Notar que esta expresión se puede escribir como

$$\frac{1}{N^2} \sum_{n=0}^{N-1} n \cos(\alpha n + \beta) = \frac{1}{2} \frac{\sin[\alpha(2N-1)/2 + \beta]}{N \sin(\alpha/2)} - \frac{1}{2} \frac{\sin(\alpha N/2 + \beta)}{N \sin(\alpha/2)} \frac{\sin(\alpha N/2)}{N \sin(\alpha/2)} \approx 0,$$

ya que todos los términos son de la forma del término mostrado en la gráfica de la figura 3.2, que toma valores pequeños si α no es cercano a 0 ni a π .

Para demostrar las ecuaciones B.1 para $i = 2$, se procede de forma análoga al caso para $i = 1$, pero la deducción se omite porque requiere una cantidad considerable de cuentas.

Otra forma de obtener el resultado, equivalente a la deducción del problema 3.6.7, es la siguiente

$$\begin{aligned}
\frac{1}{N^2} \sum_{n=0}^{N-1} n \cos(\alpha n + \beta) &= \operatorname{Re} \left\{ \frac{1}{N^2} \sum_{n=0}^{N-1} n e^{j(\alpha n + \beta)} \right\} \\
&= \frac{1}{N^2} \operatorname{Re} \left\{ e^{j\beta} \sum_{n=0}^{N-1} n e^{j\alpha n} \right\} \\
&\stackrel{(a)}{=} \frac{1}{N^2} \operatorname{Re} \left\{ e^{j\beta} e^{j\alpha} \frac{1 - N e^{j\alpha(N-1)} + (N-1) e^{j\alpha N}}{(1 - e^{j\alpha})^2} \right\} \\
&= \frac{1}{N^2} \operatorname{Re} \left\{ e^{j\beta} e^{j\alpha} \frac{1 - N e^{j\alpha N} e^{-j\alpha} + N e^{j\alpha N} - e^{j\alpha N}}{[e^{j\alpha/2} (e^{-j\alpha/2} - e^{j\alpha/2})]^2} \right\} \\
&= \frac{1}{N^2} \operatorname{Re} \left\{ e^{j\beta} e^{j\alpha} \frac{N e^{j\alpha N} (1 - e^{-j\alpha}) + (1 - e^{j\alpha N})}{e^{j\alpha} [2j \sin(\alpha/2)]^2} \right\} \\
&= \frac{1}{N^2} \operatorname{Re} \left\{ e^{j\beta} \frac{-N e^{j\alpha N} e^{-j\alpha/2} (e^{j\alpha/2} - e^{-j\alpha/2}) + e^{j\alpha N/2} (e^{j\alpha N/2} - e^{-j\alpha N/2})}{4 \sin^2(\alpha/2)} \right\} \\
&= \frac{1}{N^2} \operatorname{Re} \left\{ \frac{-N e^{j[\alpha(2N-1)/2 + \beta]} (e^{j\alpha/2} - e^{-j\alpha/2}) + e^{j(\alpha N/2 + \beta)} (e^{j\alpha N/2} - e^{-j\alpha N/2})}{4 \sin^2(\alpha/2)} \right\} \\
&= \frac{1}{N^2} \operatorname{Re} \left\{ \frac{-N e^{j[\alpha(2N-1)/2 + \beta]} 2j \sin(\alpha/2) + e^{j(\alpha N/2 + \beta)} 2j \sin(\alpha N/2)}{4 \sin^2(\alpha/2)} \right\}
\end{aligned}$$

donde en (a) se empleó que¹

$$\sum_{n=0}^{N-1} n z^n = z \frac{1 - N z^{N-1} + (N-1) z^N}{(1 - z^N)}.$$

Finalmente, al tomar la parte real se obtiene que

$$\frac{1}{N^2} \sum_{n=0}^{N-1} n \cos(\alpha n + \beta) = \frac{N \sin[\alpha(2N-1)/2 + \beta] \sin(\alpha/2) - \sin(\alpha N/2 + \beta) \sin(\alpha N/2)}{2N^2 \sin^2(\alpha/2)},$$

coincidiendo con la ecuación B.2.

¹Ver https://en.wikipedia.org/wiki/List_of_mathematical_series#Low-order_polylogarithms

Apéndice C

Momento cuarto de la media muestral

En esta sección se calculará el momento cuarto de la media muestral,

$$E(\hat{A}^4) = E\left[\left(\frac{1}{N} \sum_{n=0}^{N-1} x[n]\right)^4\right] = \frac{1}{N^4} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} E(x[i]x[j]x[k]x[l]),$$

asumiendo que las variables aleatorias $x[n]$ son idénticamente distribuidas. Teniendo esto en cuenta, la sumatoria cuádruple puede calcularse como

$$\begin{aligned} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} E(x[i]x[j]x[k]x[l]) &= (\# \text{ de sumandos con todos los índices iguales}) E(x^4[i]) \\ &\quad + (\# \text{ de sumandos con tres índices iguales}) E(x^3[i]x[j]) \\ &\quad + (\# \text{ de sumandos con dos pares de índices iguales}) E(x^2[i]x^2[j]) \\ &\quad + (\# \text{ de sumandos con dos índices iguales}) E(x^2[i]x[j]x[k]) \\ &\quad + (\# \text{ de sumandos sin índices iguales}) E(x[i]x[j]x[k]x[l]), \end{aligned}$$

ya que como las variables aleatorias $x[n]$ son idénticamente distribuidas, no importa el valor específico del índice, es decir, por ejemplo, $E(x^3[0]x[1]) = E(x^3[2]x[3])$. Por lo tanto, se calculará la cantidad de sumandos de cada tipo.

- *Cantidad de sumandos con todos los índices iguales:* la condición de este caso es que $i = j = k = l$, y como los índices toman los valores desde 0 a $N - 1$, la cantidad de sumandos es N .
- *Cantidad de sumandos con tres índices iguales:* considérese primero el caso en que tres índices específicos son iguales y toman el valor r y el restante toma el valor s , por ejemplo $i = j = k = r$ y $l = s$. La cantidad de pares posibles de valores de índices (r, s) con $r \neq s$ es $N(N - 1)$. Además, la cantidad de combinaciones de tres índices de un total de cuatro índices es $\binom{4}{3}$ (ijk, ijl, ikl, jkl). Por lo tanto,

$$\# \text{ de sumandos con tres índices iguales} = \binom{4}{3} N(N - 1) = 4N(N - 1).$$

- *Cantidad de sumandos con dos pares de índices iguales:* nuevamente, considérese primero el caso en que dos índices específicos son iguales y toman el valor r y los otros dos índices toman el valor s , por ejemplo $i = j = r$ y $k = l = s$. La cantidad de pares posibles de valores de índices (r, s) con $r \neq s$ es $N(N - 1)$. Además, la cantidad de combinaciones de dos pares de índices iguales $\binom{4}{2}/2$ (ij, ik, il), donde el factor de 2 se debe a que si ij son iguales ya implica que kl son iguales. Por lo tanto,

$$\# \text{ de sumandos con dos pares de índices iguales} = \frac{\binom{4}{2}}{2} N(N - 1) = 3N(N - 1).$$

- *Cantidad de sumandos con dos índices iguales y los otros dos distintos:* considérese el caso en que dos índices específicos son iguales y toman el valor r y los otros dos índices toman los valores s y t , por ejemplo $i = j = r$ y $k = s$ y $l = t$. La cantidad de posibles de valores de índices (r, s, t) distintos es $N(N-1)(N-2)$. Además, la cantidad de combinaciones de dos pares de índices iguales es $\binom{4}{2} (ij, ik, il, jk, jl, kl)$. Por lo tanto,

$$\# \text{ de sumandos con dos índices iguales} = \binom{4}{2} N(N-1)(N-2) = 6N(N-1)(N-2).$$

- *Cantidad de sumandos con todos los índices distintos:* finalmente, la cantidad de sumandos con todos los índices distintos es $N(N-1)(N-2)(N-3)$.

Para verificar estos resultados, puede confirmarse que la suma de la cantidad de cada tipo de sumando sea igual a la cantidad de sumandos de la sumatoria cuádruple, que es N^4 ,

$$N + 4N(N-1) + 3N(N-1) + 6N(N-1)(N-2) + N(N-1)(N-2)(N-3) = \\ N [1 + (4N-4) + (3N-3) + (6N^2-18N+12) + (N^3-6N^2+11N-6)] = N^4.$$

Finalmente, se concluye que

$$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} E(x[i]x[j]x[k]x[l]) = NE(x^4[i]) + 4N(N-1)E(x^3[i]x[j]) + 3N(N-1)E(x^2[i]x^2[j]) \\ + 6N(N-1)(N-2)E(x^2[i]x[j]x[k]) \\ + N(N-1)(N-2)(N-3)E(x[i]x[j]x[k]x[l]). \quad (\text{C.1})$$

En el caso en que además las variables son independientes, el resultado es

$$\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} E(x[i]x[j]x[k]x[l]) = NE(x^4[i]) + 4N(N-1)E(x^3[i])E(x[i]) + 3N(N-1)E^2(x^2[i]) \\ + 6N(N-1)(N-2)E(x^2[i])E^2(x[i]) \\ + N(N-1)(N-2)(N-3)E^4(x[i]) \\ = Nm_4 + 4N(N-1)m_3m_1 + 3N(N-1)m_2^2 \\ + 6N(N-1)(N-2)m_2m_1^2 + N(N-1)(N-2)(N-3)m_1^4,$$

donde m_i es el momento i -ésimo de $x[n]$.

Apéndice D

Derivadas de algunas funciones escalares multivariadas

En esta sección, se deducen algunas derivadas de funciones escalares de varias variables. Particularmente, se deducen las identidades

$$\frac{\partial \mathbf{b}^T \boldsymbol{\theta}}{\partial \boldsymbol{\theta}} = \mathbf{b}, \quad \frac{\partial^2 \mathbf{b}^T \boldsymbol{\theta}}{\partial \boldsymbol{\theta}^2} = \mathbf{0}, \quad \frac{\partial \boldsymbol{\theta}^T \mathbf{A} \boldsymbol{\theta}}{\partial \boldsymbol{\theta}} = 2\mathbf{A}\boldsymbol{\theta}, \quad \frac{\partial^2 \boldsymbol{\theta}^T \mathbf{A} \boldsymbol{\theta}}{\partial \boldsymbol{\theta}^2} = 2\mathbf{A},$$

donde \mathbf{b} y $\boldsymbol{\theta}$ son vectores $p \times 1$ y \mathbf{A} es una matriz $p \times p$ simétrica.

Considérese la función $f(\boldsymbol{\theta}) : \mathbb{R}^p \rightarrow \mathbb{R}$. Las derivadas primera y segunda de la función $f(\boldsymbol{\theta})$ respecto a $\boldsymbol{\theta}$ se definen como

$$\frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \frac{\partial f(\boldsymbol{\theta})}{\partial \theta_1} \\ \vdots \\ \frac{\partial f(\boldsymbol{\theta})}{\partial \theta_p} \end{bmatrix}, \quad \frac{\partial^2 f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2} = \begin{bmatrix} \frac{\partial^2 f(\boldsymbol{\theta})}{\partial \theta_1^2} & \cdots & \frac{\partial^2 f(\boldsymbol{\theta})}{\partial \theta_1 \partial \theta_p} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f(\boldsymbol{\theta})}{\partial \theta_p \theta_1} & \cdots & \frac{\partial^2 f(\boldsymbol{\theta})}{\partial \theta_p^2} \end{bmatrix},$$

donde la matriz $\partial^2 f(\boldsymbol{\theta})/\partial \boldsymbol{\theta}^2$ de dimensión $p \times p$ es referida como matriz Hessiana de $f(\boldsymbol{\theta})$. Una forma de calcular las derivadas es desarrollar la expresión matricial y derivar respecto a cada elemento θ_i de $\boldsymbol{\theta}$. Por ejemplo,

$$\mathbf{b}^T \boldsymbol{\theta} = \sum_{i=1}^p b_i \theta_i \quad \Rightarrow \quad \frac{\partial \mathbf{b}^T \boldsymbol{\theta}}{\partial \theta_i} = b_i \quad \Rightarrow \quad \frac{\partial \mathbf{b}^T \boldsymbol{\theta}}{\partial \boldsymbol{\theta}} = \mathbf{b}.$$

Otra forma de hacerlo es en base a la expansión de Taylor para funciones escalares de varias variables, y es la que se empleará en esta sección. La expansión de Taylor de $f(\boldsymbol{\theta})$ es¹,

$$f(\boldsymbol{\theta} + d\boldsymbol{\theta}) = f(\boldsymbol{\theta}) + d\boldsymbol{\theta}^T \frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} + \frac{1}{2} d\boldsymbol{\theta}^T \frac{\partial^2 f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2} d\boldsymbol{\theta} + \mathcal{O}(\|d\boldsymbol{\theta}\|^3). \quad (\text{D.1})$$

Defínase $f(\boldsymbol{\theta}) = \mathbf{b}^T \boldsymbol{\theta}$. De esta forma,

$$f(\boldsymbol{\theta} + d\boldsymbol{\theta}) = \mathbf{b}^T (\boldsymbol{\theta} + d\boldsymbol{\theta}) = \mathbf{b}^T \boldsymbol{\theta} + \mathbf{b}^T d\boldsymbol{\theta} = \mathbf{b}^T \boldsymbol{\theta} + d\boldsymbol{\theta}^T \mathbf{b},$$

y asociando términos con la ecuación D.1, se tiene que

$$f(\boldsymbol{\theta}) = \mathbf{b}^T \boldsymbol{\theta}, \quad \frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{b}, \quad \frac{\partial^2 f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2} = \mathbf{0},$$

¹Ver por ejemplo https://mathinsight.org/taylors_theorem_multivariable_introduction

que es lo que se quería demostrar. Ahora, definiendo $f(\boldsymbol{\theta}) = \boldsymbol{\theta}^T \mathbf{A} \boldsymbol{\theta}$, se tiene que

$$\begin{aligned}
 f(\boldsymbol{\theta} + d\boldsymbol{\theta}) &= (\boldsymbol{\theta} + d\boldsymbol{\theta})^T \mathbf{A} (\boldsymbol{\theta} + d\boldsymbol{\theta}) \\
 &= \boldsymbol{\theta}^T \mathbf{A} \boldsymbol{\theta} + d\boldsymbol{\theta}^T \mathbf{A} \boldsymbol{\theta} + \boldsymbol{\theta}^T \mathbf{A} d\boldsymbol{\theta} + d\boldsymbol{\theta}^T \mathbf{A} d\boldsymbol{\theta} \\
 &\stackrel{(a)}{=} \boldsymbol{\theta}^T \mathbf{A} \boldsymbol{\theta} + d\boldsymbol{\theta}^T (\mathbf{A} + \mathbf{A}^T) \boldsymbol{\theta} + d\boldsymbol{\theta}^T \mathbf{A} d\boldsymbol{\theta} \\
 &\stackrel{(b)}{=} \boldsymbol{\theta}^T \mathbf{A} \boldsymbol{\theta} + 2d\boldsymbol{\theta}^T \mathbf{A} \boldsymbol{\theta} + d\boldsymbol{\theta}^T \mathbf{A} d\boldsymbol{\theta},
 \end{aligned}$$

donde en (a) se empleó que como los sumandos son escalares, $\boldsymbol{\theta}^T \mathbf{A} d\boldsymbol{\theta} = (\boldsymbol{\theta}^T \mathbf{A} d\boldsymbol{\theta})^T = d\boldsymbol{\theta}^T \mathbf{A}^T \boldsymbol{\theta}$ y en (b) se empleó que \mathbf{A} es simétrica y por lo tanto, $\mathbf{A}^T = \mathbf{A}$. Nuevamente asociando términos con la ecuación D.1, se obtiene que

$$f(\boldsymbol{\theta}) = \boldsymbol{\theta}^T \mathbf{A} \boldsymbol{\theta}, \quad \frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = 2\mathbf{A} \boldsymbol{\theta}, \quad \frac{\partial^2 f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2} = 2\mathbf{A},$$

concluyendo la demostración.

Apéndice E

Transformada discreta de Fourier real

La transformada discreta de Fourier real (*real discrete Fourier transform*, RDFT) indica que cualquier señal real $x[n]$, con $n = 0, \dots, N-1$, puede representarse como la suma de cosenos y senos de frecuencias $f_k = k/N$, con $k = 0, \dots, N/2$, como,

$$x[n] = \sum_{k=0}^{N/2} a_k \cos\left(\frac{2\pi kn}{N}\right) + \sum_{k=0}^{N/2} b_k \sin\left(\frac{2\pi kn}{N}\right). \quad (\text{E.1})$$

Los coeficientes a_k y b_k corresponden a la amplitud de los cosenos y los senos respectivamente de frecuencia f_k . La ecuación E.1 es referida como *RDFT inversa* o *ecuación de síntesis*. Los coeficientes a_k y b_k se obtienen mediante la *RDFT* o *ecuaciones de análisis*,

$$a_k = \begin{cases} \frac{1}{N} \sum_{n=0}^{N-1} x[n] \cos\left(\frac{2\pi kn}{N}\right), & \text{si } k = 0 \text{ o } k = \frac{N}{2} \\ \frac{2}{N} \sum_{n=0}^{N-1} x[n] \cos\left(\frac{2\pi kn}{N}\right), & \text{en otro caso.} \end{cases} \quad (\text{E.2})$$
$$b_k = \begin{cases} \frac{1}{N} \sum_{n=0}^{N-1} x[n] \sin\left(\frac{2\pi kn}{N}\right), & \text{si } k = 0 \text{ o } k = \frac{N}{2} \\ \frac{2}{N} \sum_{n=0}^{N-1} x[n] \sin\left(\frac{2\pi kn}{N}\right), & \text{en otro caso.} \end{cases}$$

En este apéndice se deducen las ecuaciones de análisis y síntesis de la RDFT a partir de la transformada discreta de Fourier compleja convencional, mucho mas común en la bibliografía. Una discusión detallada sobre la RDFT y su relación con la DFT puede encontrarse en los capítulos 8 y 31 de [16].

Considérese una señal $x[n]$, $n = 0, \dots, N-1$. La DFT de $x[n]$ se define como (ver por ejemplo, el capítulo 8 de [6])

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-j2\pi kn/N}, \quad k = 0, \dots, N-1, \quad (\text{E.3})$$

y la DFT inversa es

$$x[n] = \frac{1}{N} \sum_{k=0}^{N-1} X[k] e^{j2\pi kn/N}, \quad n = 0, \dots, N-1. \quad (\text{E.4})$$

El objetivo es deducir a_k y b_k en función de $X[k]$. Para hacerlo, se llevará la ecuación E.4 de la DFT inversa a la forma de la ecuación E.1 de la RDFT inversa considerando que $x[n]$ es real. En general $X[k]$ toma valores complejos, pero en el caso en que $x[n]$ es real, cumple las siguientes propiedades:

- $X[0]$ y $X[N/2]$ son reales. Efectivamente,

$$X[0] = \sum_{n=0}^{N-1} x[n], \quad X[N/2] = \sum_{n=0}^{N-1} x[n]e^{-j\pi n} = \sum_{n=0}^{N-1} x[n](-1)^n$$

son reales por ser $x[n]$ real.

- $X[N-k] = X^*[k]$. Para verlo, se parte de la definición de la DFT evaluada en $k = N-k$,

$$\begin{aligned} X[N-k] &= \sum_{n=0}^{N-1} x[n]e^{-j2\pi(N-k)n/N} \\ &= \sum_{n=0}^{N-1} x[n]e^{j2\pi kn/N}e^{-j2\pi n} \\ &= \sum_{n=0}^{N-1} x[n]e^{j2\pi kn/N} \\ &\stackrel{(a)}{=} \left(\sum_{n=0}^{N-1} x[n]e^{-j2\pi kn/N} \right)^* \\ &= X^*[k], \end{aligned}$$

en donde en (a) se consideró que $x[n]$ es real.

Partiendo de la DFT inversa, se observa que

$$\begin{aligned} x[n] &= \frac{1}{N} \sum_{k=0}^{N-1} X[k]e^{j2\pi kn/N} \\ &= \frac{1}{N} \left(\sum_{k=0}^{N/2} X[k]e^{j2\pi kn/N} + \sum_{k=N/2+1}^{N-1} X[k]e^{j2\pi kn/N} \right). \end{aligned}$$

Se operará con la segunda sumatoria haciendo el cambio de variable $l = N-k$. De esta forma, se tiene que

$$\begin{aligned} \sum_{k=N/2+1}^{N-1} X[k]e^{j2\pi kn/N} &= \sum_{l=N/2-1}^{-1} X[N-l]e^{j2\pi(N-l)n/N} \\ &= \sum_{l=1}^{N/2-1} X^*[l]e^{-j2\pi ln/N}, \end{aligned}$$

y por lo tanto,

$$\begin{aligned} x[n] &= \frac{1}{N} \left(\sum_{k=0}^{N/2} X[k]e^{j2\pi kn/N} + \sum_{k=1}^{N/2-1} X^*[k]e^{-j2\pi kn/N} \right) \\ &\stackrel{(a)}{=} \frac{1}{N} \sum_{k=0}^{N/2} \left(\tilde{X}[k]e^{j2\pi kn/N} + \tilde{X}^*[k]e^{-j2\pi kn/N} \right) \\ &= \frac{1}{N} \sum_{k=0}^{N/2} 2 \operatorname{Re} \left(\tilde{X}[k]e^{j2\pi kn/N} \right), \end{aligned}$$

donde en (a) se agregaron los sumandos $X^*[0]$ y $X^*[N/2]e^{-j\pi n}$ en la segunda sumatoria correspondientes a $k=0$ y $k=N/2$ respectivamente y se notó que coinciden con los sumandos correspondientes a $k=0$ y $k=N/2$ de la primera sumatoria ya que $X[0]$ y $X[N/2]$ son reales, es decir,

$$X^*[0] = X[0], \quad \text{y} \quad X^*[N/2]e^{-j\pi n} = X[N/2]e^{j\pi n},$$

y se definió

$$\tilde{X}[k] = \begin{cases} \frac{X[k]}{2}, & \text{si } k = 0 \text{ o } k = \frac{N}{2} \\ X[k], & \text{en otro caso} \end{cases}$$

para mantener la igualdad. Continuando el razonamiento, como

$$\tilde{X}[k]e^{j2\pi kn/N} = \left[\text{Re}(\tilde{X}[k]) + j \text{Im}(\tilde{X}[k]) \right] \left[\cos\left(\frac{2\pi kn}{N}\right) + j \sin\left(\frac{2\pi kn}{N}\right) \right],$$

se tiene que

$$\text{Re}(\tilde{X}[k]e^{j2\pi kn/N}) = \text{Re}(\tilde{X}[k]) \cos\left(\frac{2\pi kn}{N}\right) - \text{Im}(\tilde{X}[k]) \sin\left(\frac{2\pi kn}{N}\right),$$

resultando en

$$x[n] = \frac{2}{N} \sum_{k=0}^{N/2} \text{Re}(\tilde{X}[k]) \cos\left(\frac{2\pi kn}{N}\right) - \frac{2}{N} \sum_{k=0}^{N/2} \text{Im}(\tilde{X}[k]) \sin\left(\frac{2\pi kn}{N}\right).$$

Comparando este resultado con la ecuación E.1, se observa que

$$a_k = \frac{2}{N} \text{Re}(\tilde{X}[k]), \quad b_k = -\frac{2}{N} \text{Im}(\tilde{X}[k]),$$

y sustituyendo por la definición de la DFT dada por la ecuación E.3,

$$X[k] = \sum_{n=0}^{N-1} x[n] \cos\left(\frac{2\pi kn}{N}\right) - j \sum_{n=0}^{N-1} x[n] \sin\left(\frac{2\pi kn}{N}\right),$$

se obtienen las ecuaciones E.2 de la RDFT.

Apéndice F

PDF condicional de variables aleatorias gaussianas

En esta sección se deriva la PDF condicional de variables aleatorias conjuntamente gaussianas en el caso bivariado y luego se generaliza el resultado para el caso multivariado.

F.1. Caso escalar

Sea $[x \ y]^T$ un vector gaussiano de media $\boldsymbol{\mu} = [E(x) \ E(y)]^T$ y matriz de covarianza

$$\mathbf{C} = \begin{bmatrix} \text{var}(x) & \text{cov}(x, y) \\ \text{cov}(y, x) & \text{var}(y) \end{bmatrix},$$

Previo a deducir la PDF condicional, se demostrará que

$$E(x | y) = E(x) + \frac{\text{cov}(x, y)}{\text{var}(y)}(y - E(y)).$$

Para hacerlo, se considera la variable aleatoria¹

$$z = x - \frac{\text{cov}(x, y)}{\text{var}(y)}y,$$

que por ser una combinación lineal de variables aleatorias conjuntamente normales que incluye a y , es conjuntamente normal con y (ver el capítulo 6 de [2]). Además, z no está correlacionada con y . Efectivamente,

$$\text{cov}(y, z) = E(yz) - E(y)E(z),$$

con

$$E(yz) = E\left[y\left(x - \frac{\text{cov}(x, y)}{\text{var}(y)}y\right)\right] = E(xy) - \frac{\text{cov}(x, y)}{\text{var}(y)}E(y^2)$$

y

$$E(y)E(z) = E(y)E\left(x - \frac{\text{cov}(x, y)}{\text{var}(y)}y\right) = E(x)E(y) - \frac{\text{cov}(x, y)}{\text{var}(y)}E^2(y),$$

y por lo tanto, restando estas dos ecuaciones se obtiene que

$$\begin{aligned} \text{cov}(y, z) &= E(xy) - E(x)E(y) - \frac{\text{cov}(x, y)}{\text{var}(y)}(E(y^2) - E^2(y)) \\ &= \text{cov}(x, y) - \frac{\text{cov}(x, y)}{\text{var}(y)}\text{var}(y) \\ &= 0. \end{aligned}$$

¹Esta demostración está basada en <https://math.stackexchange.com/questions/2167643/conditional-expectation-of-a-joint-normal-distribution>

Se concluye que como y y z son conjuntamente normales y no correlacionadas, son independientes (ver el capítulo 6 de [2]). Se calculará a continuación la esperanza condicional,

$$\begin{aligned}
E(x|y) &\stackrel{(a)}{=} E\left(x - \frac{\text{cov}(x, y)}{\text{var}(y)}y + \frac{\text{cov}(x, y)}{\text{var}(y)}y \mid y\right) \\
&= E\left(x - \frac{\text{cov}(x, y)}{\text{var}(y)}y \mid y\right) + \frac{\text{cov}(x, y)}{\text{var}(y)}E(y|y) \\
&\stackrel{(c)}{=} E\left(x - \frac{\text{cov}(x, y)}{\text{var}(y)}y\right) + \frac{\text{cov}(x, y)}{\text{var}(y)}y \\
&= E(x) - \frac{\text{cov}(x, y)}{\text{var}(y)}(E(y) - y),
\end{aligned}$$

donde en (a) se sumó y se restó el mismo valor en la variable aleatoria condicionada y en (c) se consideró que la variable condicionada en la primer esperanza es z , y como es independiente de y se cumple que $E(z|y) = E(z)$ y además, en la segunda esperanza se tomó en cuenta que $E(y|y) = y$.

A continuación se deriva la PDF condicional

$$p(x|y) = \frac{p(x, y)}{p(y)},$$

donde

$$p(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_x)^2}{\sigma_x^2} - 2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \frac{(y-\mu_y)^2}{\sigma_y^2}\right]\right\}$$

con $\mu_x = E(x)$, $\mu_y = E(y)$, $\sigma_x^2 = \text{var}(x)$, $\sigma_y^2 = \text{var}(y)$ y $\rho = \text{cov}(x, y)/\sqrt{\text{var}(x)\text{var}(y)}$, y

$$p(y) = \frac{1}{\sqrt{2\pi}\sigma_y} \exp\left[-\frac{1}{2\sigma_y^2}(y-\mu_y)^2\right].$$

Sustituyendo, se tiene que

$$\begin{aligned}
p(x|y) &= \frac{\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_x)^2}{\sigma_x^2} - 2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \frac{(y-\mu_y)^2}{\sigma_y^2}\right]\right\}}{\frac{1}{\sqrt{2\pi}\sigma_y} \exp\left[-\frac{1}{2\sigma_y^2}(y-\mu_y)^2\right]} \\
&= \frac{1}{\sqrt{2\pi}\sigma_x\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_x)^2}{\sigma_x^2} - 2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \frac{(y-\mu_y)^2}{\sigma_y^2}\right] + \frac{1}{2\sigma_y^2}(y-\mu_y)^2\right\} \\
&= \frac{1}{\sqrt{2\pi}(1-\rho^2)\sigma_x^2} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_x)^2}{\sigma_x^2} - 2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \frac{(y-\mu_y)^2}{\sigma_y^2} - \frac{1-\rho^2}{\sigma_y^2}(y-\mu_y)^2\right]\right\} \\
&= \frac{1}{\sqrt{2\pi}(1-\rho^2)\sigma_x^2} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_x)^2}{\sigma_x^2} - 2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \rho^2\frac{(y-\mu_y)^2}{\sigma_y^2}\right]\right\} \\
&= \frac{1}{\sqrt{2\pi}(1-\rho^2)\sigma_x^2} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{x-\mu_x}{\sigma_x} - \rho\frac{y-\mu_y}{\sigma_y}\right]^2\right\}.
\end{aligned}$$

resultando en que

$$p(x|y) = \frac{1}{\sqrt{2\pi}(1-\rho^2)\sigma_x^2} \exp\left\{-\frac{1}{2(1-\rho^2)\sigma_x^2}\left[x-\mu_x-\rho\frac{\sigma_x}{\sigma_y}(y-\mu_y)\right]^2\right\}.$$

Esta PDF se reconoce como una PDF gaussiana de media $\mu_x + \rho \frac{\sigma_x}{\sigma_y}(y - \mu_y)$ y varianza $(1 - \rho^2)\sigma_x^2$, es decir

$$x | y \sim \mathcal{N}\left(\mu_x + \rho \frac{\sigma_x}{\sigma_y}(y - \mu_y), (1 - \rho^2)\sigma_x^2\right).$$

De forma acorde al resultado obtenido previamente, la esperanza condicional es

$$\begin{aligned} E(x | y) &= \mu_x + \rho \frac{\sigma_x}{\sigma_y}(y - \mu_y) \\ &= E(x) + \frac{\text{cov}(x, y)}{\sqrt{\text{var}(x) \text{var}(y)}} \sqrt{\frac{\text{var}(x)}{\text{var}(y)}} (y - E(y)) \\ &= E(x) + \frac{\text{cov}(x, y)}{\text{var}(y)} (y - E(y)), \end{aligned}$$

y la varianza condicional es

$$\begin{aligned} \text{var}(x | y) &= (1 - \rho^2)\sigma_x^2 \\ &= \left[1 - \frac{\text{cov}^2(x, y)}{\text{var}(x) \text{var}(y)}\right] \text{var}(x) \\ &= \text{var}(x) - \frac{\text{cov}^2(x, y)}{\text{var}(y)}. \end{aligned}$$

F.2. Caso multivariado

Como generalización del resultado anterior, se considera el caso en que las variables aleatorias en la esperanza condicional son vectores.

Teorema (PDF condicional de gaussiana multivariada) Si \mathbf{x} y \mathbf{y} son conjuntamente gaussianas, donde \mathbf{x} es $k \times 1$ y \mathbf{y} es $l \times 1$, con vector de media $[E(\mathbf{x})^T E(\mathbf{y})^T]^T$ y matriz de covarianza particionada

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_{xx} & \mathbf{C}_{xy} \\ \mathbf{C}_{yx} & \mathbf{C}_{yy} \end{bmatrix}$$

tal que

$$p(\mathbf{x}, \mathbf{y}) = \frac{1}{(2\pi)^{\frac{k+l}{2}} \det^{\frac{1}{2}}(\mathbf{C})} \exp \left[-\frac{1}{2} \begin{bmatrix} \mathbf{x} - E(\mathbf{x}) \\ \mathbf{y} - E(\mathbf{y}) \end{bmatrix}^T \mathbf{C}^{-1} \begin{bmatrix} \mathbf{x} - E(\mathbf{x}) \\ \mathbf{y} - E(\mathbf{y}) \end{bmatrix} \right],$$

la PDF condicional $p(\mathbf{y}|\mathbf{x})$ es también gaussiana con

$$\begin{aligned} E(\mathbf{y}|\mathbf{x}) &= E(\mathbf{y}) + \mathbf{C}_{yx} \mathbf{C}_{xx}^{-1} (\mathbf{x} - E(\mathbf{x})) \\ \mathbf{C}_{y|x} &= \mathbf{C}_{yy} - \mathbf{C}_{yx} \mathbf{C}_{xx}^{-1} \mathbf{C}_{xy}. \end{aligned}$$

A continuación, se incluye una demostración en la que se asume que la matriz de covarianza \mathbf{C} es invertible. Sin embargo, las expresiones de la media y la covarianza condicional son válidas incluso en el caso en que \mathbf{C} es singular.

Demostración Por definición, la PDF condicional es

$$\begin{aligned} p(\mathbf{y}|\mathbf{x}) &= \frac{p(\mathbf{x}, \mathbf{y})}{p(\mathbf{x})} \\ &= \frac{\frac{1}{(2\pi)^{\frac{k+l}{2}} \det^{\frac{1}{2}}(\mathbf{C})} \exp \left[-\frac{1}{2} \begin{bmatrix} \mathbf{x} - E(\mathbf{x}) \\ \mathbf{y} - E(\mathbf{y}) \end{bmatrix}^T \mathbf{C}^{-1} \begin{bmatrix} \mathbf{x} - E(\mathbf{x}) \\ \mathbf{y} - E(\mathbf{y}) \end{bmatrix} \right]}{\frac{1}{(2\pi)^{\frac{k}{2}} \det^{\frac{1}{2}}(\mathbf{C}_{xx})} \exp \left[-\frac{1}{2} (\mathbf{x} - E(\mathbf{x}))^T \mathbf{C}_{xx}^{-1} (\mathbf{x} - E(\mathbf{x})) \right]} \end{aligned}$$

El determinante de la matriz de covarianza puede ser evaluado como

$$\det(\mathbf{C}) = \det \left(\begin{bmatrix} \mathbf{C}_{xx} & \mathbf{C}_{xy} \\ \mathbf{C}_{yx} & \mathbf{C}_{yy} \end{bmatrix} \right) = \det(\mathbf{C}_{xx}) \det(\mathbf{C}_{yy} - \mathbf{C}_{yx} \mathbf{C}_{xx}^{-1} \mathbf{C}_{xy}),$$

como se demuestra en el apéndice G (ver la ecuación G.1) y por lo tanto,

$$\frac{\det(\mathbf{C})}{\det(\mathbf{C}_{xx})} = \det(\mathbf{C}_{yy} - \mathbf{C}_{yx} \mathbf{C}_{xx}^{-1} \mathbf{C}_{xy}).$$

De esta forma, la PDF condicional queda

$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{(2\pi)^{\frac{1}{2}} \det^{\frac{1}{2}}(\mathbf{C}_{yy} - \mathbf{C}_{yx} \mathbf{C}_{xx}^{-1} \mathbf{C}_{xy})} \exp \left(-\frac{1}{2} Q \right),$$

donde

$$Q = \begin{bmatrix} \mathbf{x} - E(\mathbf{x}) \\ \mathbf{y} - E(\mathbf{y}) \end{bmatrix}^T \mathbf{C}^{-1} \begin{bmatrix} \mathbf{x} - E(\mathbf{x}) \\ \mathbf{y} - E(\mathbf{y}) \end{bmatrix} - (\mathbf{x} - E(\mathbf{x}))^T \mathbf{C}_{xx}^{-1} (\mathbf{x} - E(\mathbf{x}))$$

Para evaluar Q hay que invertir la matriz de covarianza \mathbf{C} . Esto puede hacerse empleando el resultado de la inversa de una matriz en bloques de la ecuación G.3 del apéndice G, que indica que

$$\begin{bmatrix} \mathbf{C}_{xx} & \mathbf{C}_{xy} \\ \mathbf{C}_{yx} & \mathbf{C}_{yy} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{C}_{xx}^{-1} + \mathbf{C}_{xx}^{-1} \mathbf{C}_{xy} (\mathbf{C}_{yy} - \mathbf{C}_{yx} \mathbf{C}_{xx}^{-1} \mathbf{C}_{xy})^{-1} \mathbf{C}_{yx} \mathbf{C}_{xx}^{-1} & -\mathbf{C}_{xx}^{-1} \mathbf{C}_{xy} (\mathbf{C}_{yy} - \mathbf{C}_{yx} \mathbf{C}_{xx}^{-1} \mathbf{C}_{xy})^{-1} \\ -(\mathbf{C}_{yy} - \mathbf{C}_{yx} \mathbf{C}_{xx}^{-1} \mathbf{C}_{xy})^{-1} \mathbf{C}_{yx} \mathbf{C}_{xx}^{-1} & (\mathbf{C}_{yy} - \mathbf{C}_{yx} \mathbf{C}_{xx}^{-1} \mathbf{C}_{xy})^{-1} \end{bmatrix}$$

y definiendo

$$\mathbf{B} = \mathbf{C}_{yy} - \mathbf{C}_{yx} \mathbf{C}_{xx}^{-1} \mathbf{C}_{xy}$$

se obtiene que

$$\mathbf{C}^{-1} = \begin{bmatrix} \mathbf{C}_{xx}^{-1} + \mathbf{C}_{xx}^{-1} \mathbf{C}_{xy} \mathbf{B}^{-1} \mathbf{C}_{yx} \mathbf{C}_{xx}^{-1} & -\mathbf{C}_{xx}^{-1} \mathbf{C}_{xy} \mathbf{B}^{-1} \\ -\mathbf{B}^{-1} \mathbf{C}_{yx} \mathbf{C}_{xx}^{-1} & \mathbf{B}^{-1} \end{bmatrix}.$$

Puede comprobarse que la matriz inversa puede factorizarse como

$$\mathbf{C}^{-1} = \begin{bmatrix} \mathbf{I} & -\mathbf{C}_{xx}^{-1} \mathbf{C}_{xy} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{C}_{xx}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{C}_{yx} \mathbf{C}_{xx}^{-1} & \mathbf{I} \end{bmatrix},$$

por lo tanto, definiendo $\bar{\mathbf{x}} = \mathbf{x} - E(\mathbf{x})$ y $\bar{\mathbf{y}} = \mathbf{y} - E(\mathbf{y})$, se tiene que

$$\begin{aligned} Q &= [\bar{\mathbf{x}}^T \quad \bar{\mathbf{y}}^T]^T \begin{bmatrix} \mathbf{I} & -\mathbf{C}_{xx}^{-1} \mathbf{C}_{xy} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{C}_{xx}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{C}_{yx} \mathbf{C}_{xx}^{-1} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \bar{\mathbf{x}} \\ \bar{\mathbf{y}} \end{bmatrix} - \bar{\mathbf{x}}^T \mathbf{C}_{xx}^{-1} \bar{\mathbf{x}} \\ &= [\bar{\mathbf{x}}^T \quad \bar{\mathbf{y}}^T - \bar{\mathbf{x}}^T \mathbf{C}_{xx}^{-1} \mathbf{C}_{xy}] \begin{bmatrix} \mathbf{C}_{xx}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}^{-1} \end{bmatrix} \begin{bmatrix} \bar{\mathbf{x}} \\ \bar{\mathbf{y}} - \mathbf{C}_{yx} \mathbf{C}_{xx}^{-1} \bar{\mathbf{x}} \end{bmatrix} - \bar{\mathbf{x}}^T \mathbf{C}_{xx}^{-1} \bar{\mathbf{x}} \\ &= [\bar{\mathbf{x}}^T \quad \bar{\mathbf{y}}^T - \bar{\mathbf{x}}^T \mathbf{C}_{xx}^{-1} \mathbf{C}_{xy}] \begin{bmatrix} \mathbf{C}_{xx}^{-1} \bar{\mathbf{x}} \\ \mathbf{B}^{-1} (\bar{\mathbf{y}} - \mathbf{C}_{yx} \mathbf{C}_{xx}^{-1} \bar{\mathbf{x}}) \end{bmatrix} - \bar{\mathbf{x}}^T \mathbf{C}_{xx}^{-1} \bar{\mathbf{x}} \\ &= \bar{\mathbf{x}}^T \mathbf{C}_{xx}^{-1} \bar{\mathbf{x}} + (\bar{\mathbf{y}}^T - \bar{\mathbf{x}}^T \mathbf{C}_{xx}^{-1} \mathbf{C}_{xy}) \mathbf{B}^{-1} (\bar{\mathbf{y}} - \mathbf{C}_{yx} \mathbf{C}_{xx}^{-1} \bar{\mathbf{x}}) - \bar{\mathbf{x}}^T \mathbf{C}_{xx}^{-1} \bar{\mathbf{x}} \\ &\stackrel{(a)}{=} (\bar{\mathbf{y}} - \mathbf{C}_{yx} \mathbf{C}_{xx}^{-1} \bar{\mathbf{x}})^T \mathbf{B}^{-1} (\bar{\mathbf{y}} - \mathbf{C}_{yx} \mathbf{C}_{xx}^{-1} \bar{\mathbf{x}}), \end{aligned}$$

donde en (a) se tuvo en cuenta que $\mathbf{C}_{xy}^T = \mathbf{C}_{yx}$. Finalmente, sustituyendo las definiciones de $\bar{\mathbf{x}}$, $\bar{\mathbf{y}}$ y \mathbf{B} , se obtiene que

$$Q = [\mathbf{y} - (E(\mathbf{y}) + \mathbf{C}_{yx} \mathbf{C}_{xx}^{-1} (\mathbf{x} - E(\mathbf{x})))^T [\mathbf{C}_{yy} - \mathbf{C}_{yx} \mathbf{C}_{xx}^{-1} \mathbf{C}_{xy}]^{-1} [\mathbf{y} - (E(\mathbf{y}) + \mathbf{C}_{yx} \mathbf{C}_{xx}^{-1} (\mathbf{x} - E(\mathbf{x})))],$$

concluyendo que la media y la covarianza de la PDF condicional están dadas por las ecuaciones 10.14 y 10.15 respectivamente.

Apéndice G

Propiedades de matrices en bloques

Se considera la matriz en bloques

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}$$

donde \mathbf{A} , \mathbf{B} , \mathbf{C} , y \mathbf{D} son matrices de dimensión $n \times n$, $n \times m$, $m \times n$ y $m \times m$ respectivamente.

G.1. Determinante

En esta sección se verá que ¹

$$\det \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} = \det(\mathbf{A}) \times \det(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}) = \det(\mathbf{D}) \times \det(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}). \quad (\text{G.1})$$

Para hacerlo, se considera primero que

$$\det \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} = \det(\mathbf{A}) \times \det(\mathbf{D}) = \det \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{0} & \mathbf{D} \end{bmatrix}.$$

Esto se deduce a partir de la descomposición

$$\begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{C} & \mathbf{I}_m \end{bmatrix} \begin{bmatrix} \mathbf{I}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{D} \end{bmatrix}$$

(para la primer igualdad) y de que $\det(\mathbf{AB}) = \det(\mathbf{A}) \times \det(\mathbf{B})$. Teniendo esto en cuenta, para demostrar la ecuación G.1, alcanza con considerar las descomposiciones

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{C} & \mathbf{I}_m \end{bmatrix} \begin{bmatrix} \mathbf{I}_n & \mathbf{A}^{-1}\mathbf{B} \\ \mathbf{0} & \mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_n & \mathbf{B} \\ \mathbf{0} & \mathbf{D} \end{bmatrix} \begin{bmatrix} \mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C} & \mathbf{0} \\ \mathbf{D}^{-1}\mathbf{C} & \mathbf{I}_m \end{bmatrix},$$

y tomar el determinante. Observar que para realizar la primera descomposición es necesario que \mathbf{A} sea invertible y para realizar la segunda, es necesario que \mathbf{D} sea invertible.

G.2. Matriz inversa

La matriz inversa a la matriz en bloques es ²

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} (\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} & -(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1} \\ -\mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} & \mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1} \end{bmatrix} \quad (\text{G.2})$$

¹Ver https://en.wikipedia.org/wiki/Determinant#Block_matrices

²Ver https://en.wikipedia.org/wiki/Invertible_matrix#Blockwise_inversion

Para mostrarlo, alcanza con multiplicar la matriz original por la matriz indicada como la inversa y ver que el resultado es la matriz identidad. Efectivamente,

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} \begin{bmatrix} (\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} & -(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1} \\ -\mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} & \mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}$$

donde

$$\begin{aligned} \mathbf{A}_{11} &= \mathbf{A}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} = (\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} = \mathbf{I} \\ \mathbf{A}_{12} &= -\mathbf{A}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1} + \mathbf{B}[\mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1}] \\ &= (-\mathbf{A} + \mathbf{B}\mathbf{D}^{-1}\mathbf{C})(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1} + \mathbf{B}\mathbf{D}^{-1} = -\mathbf{B}\mathbf{D}^{-1} + \mathbf{B}\mathbf{D}^{-1} = \mathbf{0} \\ \mathbf{A}_{21} &= \mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} - \mathbf{D}\mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} \\ &= \mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} - \mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} = \mathbf{0} \\ \mathbf{A}_{22} &= -\mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1} + \mathbf{D}[\mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1}] \\ &= (-\mathbf{C} + \mathbf{C})(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1} + \mathbf{D}\mathbf{D}^{-1} = \mathbf{I}, \end{aligned}$$

que es la matriz identidad. De la misma forma, puede demostrarse que también se cumple que

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} \\ -(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1} & (\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} \end{bmatrix} \quad (\text{G.3})$$

Igualando las ecuaciones G.2 y G.3 elemento a elemento se obtienen las siguientes identidades

$$\begin{aligned} (\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} &= \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1} \\ (\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1} &= \mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} \\ \mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} &= (\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1} \\ \mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1} &= (\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}. \end{aligned}$$

En particular, la primera es la identidad matricial de Woodbury, demostrada en el apéndice A.

G.3. Matriz diagonal en bloques

Inversa Una matriz diagonal en bloques es invertible si y solo si cada bloque de la diagonal es invertible, y en ese caso, la matriz inversa también es diagonal a bloques dada por

$$\begin{bmatrix} \mathbf{A}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{A}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{A}_n \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}_1^{-1} & 0 & \cdots & 0 \\ 0 & \mathbf{A}_2^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{A}_n^{-1} \end{bmatrix}.$$

Para mostrarlo, alcanza con multiplicar la matriz por la inversa y observar que el resultado es la matriz identidad.

Apéndice H

Conceptos básicos de álgebra lineal

En este apéndice se incluyen algunos conceptos básicos de álgebra lineal y propiedades de matrices [17].

H.1. Conceptos básicos

Dimensión La dimensión de un espacio vectorial V es el número de vectores que forman una base de V . Se denota como $\dim(V)$.

Rango El rango de una transformación lineal $L : V \rightarrow W$ entre dos espacios vectoriales V y W , denotado como $\text{rank}(L)$, se define como la dimensión de la imagen de la transformación,

$$\text{rank}(L) = \dim(\text{Im}(L)).$$

Si \mathbf{A} es una matriz $m \times n$, $f(\mathbf{x}) = \mathbf{A}\mathbf{x}$ define una transformación lineal en $\mathbb{R}^n \rightarrow \mathbb{R}^m$. El rango de una matriz \mathbf{A} es la dimensión del espacio vectorial generado por sus columnas y corresponde al número de columnas linealmente independientes de \mathbf{A} . A su vez, coincide con la dimensión del espacio vectorial generado por las filas de \mathbf{A} . Algunas consideraciones son las siguientes:

- El rango de una matriz $m \times n$ es un entero no-negativo que cumple que

$$\text{rank}(\mathbf{A}) \leq \min(m, n).$$

Una matriz que tiene rango $\min(m, n)$ se dice que es de *rango completo*.

- Una matriz \mathbf{A} cuadrada ($m = n$) es invertible si y solo si \mathbf{A} es de rango n , es decir, \mathbf{A} es de rango completo.
- Si \mathbf{A} y \mathbf{C} son matrices invertibles,

$$\text{rank}(\mathbf{ABC}) = \text{rank}(\mathbf{B}).$$

Núcleo El núcleo de una transformación lineal es el conjunto de vectores del dominio de la transformación que se mapean en el vector nulo. Es decir, dada una transformación lineal $L : V \rightarrow W$ entre dos espacios vectoriales V y W , el núcleo, denotado como $\ker(L)$, es el conjunto de elementos $\mathbf{v} \in V$ que cumplen que $L(\mathbf{v}) = \mathbf{0}$, es decir,

$$\ker(L) = \{\mathbf{v} \in V \mid L(\mathbf{v}) = \mathbf{0}\}.$$

Si la transformación está representada por una matriz \mathbf{A} $m \times n$ el núcleo son los vectores \mathbf{x} que cumplen que $\mathbf{Ax} = \mathbf{0}$, es decir,

$$\ker(\mathbf{A}) = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{Ax} = \mathbf{0}\}.$$

Algunas consideraciones son las siguientes

- La dimensión de $\ker(\mathbf{A})$ se llama *nulidad* de \mathbf{A} y se denota como $\text{nul}(\mathbf{A})$,

$$\text{nul}(\mathbf{A}) = \dim(\ker(\mathbf{A})).$$

H.2. Propiedades y teoremas

Teorema del rango El teorema del rango o teorema del rango-nulidad relaciona las dimensiones del núcleo y la imagen de una transformación lineal con las dimensiones de su dominio¹. Dada una transformación lineal $L : V \rightarrow W$ entre dos espacios vectoriales V y W , el teorema del rango establece que

$$\text{rank}(L) + \text{nul}(L) = \dim(V).$$

Si la transformación está representada por una matriz \mathbf{A} $m \times n$, el teorema indica que

$$\text{rank}(\mathbf{A}) + \text{nul}(\mathbf{A}) = n.$$

- **Ejemplo:** si \mathbf{A} es una matriz $m \times n$, se cumple que

$$\text{rank}(\mathbf{A}^T \mathbf{A}) = \text{rank}(\mathbf{A}).$$

Para verlo², se mostrará primero que $\text{nul}(\mathbf{A}^T \mathbf{A}) = \text{nul}(\mathbf{A})$, lo que implica mostrar que dado un vector \mathbf{x} $n \times 1$ se cumple que

$$\mathbf{A}^T \mathbf{A} \mathbf{x} = \mathbf{0} \quad \Leftrightarrow \quad \mathbf{A} \mathbf{x} = \mathbf{0}.$$

El recíproco (\Leftarrow) es obvio. Para probar el directo, por hipótesis se cumple que $\mathbf{A}^T \mathbf{A} \mathbf{x} = \mathbf{0}$ y por lo tanto, multiplicando por \mathbf{x}^T se tiene que

$$\mathbf{x}^T \mathbf{A}^T \mathbf{A} \mathbf{x} = (\mathbf{A} \mathbf{x})^T (\mathbf{A} \mathbf{x}) = \|\mathbf{A} \mathbf{x}\|^2 = 0 \quad \Rightarrow \quad \mathbf{A} \mathbf{x} = \mathbf{0}.$$

Esto indica que $\ker(\mathbf{A}^T \mathbf{A}) = \ker(\mathbf{A})$ y por lo tanto, $\text{nul}(\mathbf{A}^T \mathbf{A}) = \text{nul}(\mathbf{A})$. Por otro lado, $\mathbf{A}^T \mathbf{A}$ es una matriz $n \times n$ y teniendo en cuenta que \mathbf{A} es una matriz $m \times n$, por el teorema del rango, se cumple que

$$\begin{aligned} \text{rank}(\mathbf{A}^T \mathbf{A}) + \text{nul}(\mathbf{A}^T \mathbf{A}) &= n \\ \text{rank}(\mathbf{A}) + \text{nul}(\mathbf{A}) &= n \end{aligned} \quad \Rightarrow \quad \text{rank}(\mathbf{A}^T \mathbf{A}) + \text{nul}(\mathbf{A}^T \mathbf{A}) = \text{rank}(\mathbf{A}) + \text{nul}(\mathbf{A})$$

y como $\text{nul}(\mathbf{A}^T \mathbf{A}) = \text{nul}(\mathbf{A})$ como se demostró previamente, se concluye que $\text{rank}(\mathbf{A}^T \mathbf{A}) = \text{rank}(\mathbf{A})$.

Matriz diagonalizable Sea \mathbf{A} una matriz simétrica real $n \times n$. Se cumple que:

- los valores propios $\lambda_1, \dots, \lambda_n$ de \mathbf{A} son reales. Además, existen n vectores propios ortonormales $\mathbf{v}_1, \dots, \mathbf{v}_n$ correspondientes a cada valor propio.
- *Teorema espectral:* sea la matriz diagonal $\mathbf{D} = \text{diag}(\lambda_1, \dots, \lambda_n)$ formada con los valores propios de \mathbf{A} y la matriz $\mathbf{P} = [\mathbf{v}_1 \dots \mathbf{v}_n]$ cuyas columnas son los vectores propios de \mathbf{A} . La matriz \mathbf{P} es ortonormal por lo que cumple que $\mathbf{P}^{-1} = \mathbf{P}^T$ y además

$$\mathbf{A} = \mathbf{P} \mathbf{D} \mathbf{P}^{-1} = \mathbf{P} \mathbf{D} \mathbf{P}^T.$$

Las matrices que cumplen esta propiedad se llaman *matrices similares*³.

- el rango de \mathbf{A} es el número de valores propios distintos de cero. Además, los vectores propios correspondientes a valores propios distintos de cero son una base del subespacio generado por las columnas de \mathbf{A} .
- Esto es consecuencia de que $\mathbf{A} = \mathbf{P} \mathbf{D} \mathbf{P}^{-1}$ y como \mathbf{P} es de rango completo n , $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{D})$. Como \mathbf{D} es una matriz diagonal con los valores propios como elementos, su rango es la cantidad de valores propios distintos de cero.

¹Ver https://en.wikipedia.org/wiki/Rank-nullity_theorem

²La demostración está basada en <https://math.stackexchange.com/questions/215145/rank-of-product-of-a-matrix-and-its-transpose/805845#805845>

³Ver: https://en.wikipedia.org/wiki/Matrix_similarity

- $\det(\mathbf{A}) = \lambda_1 \times \cdots \times \lambda_n$, ya que

$$\det(\mathbf{A}) = \det(\mathbf{PDP}^{-1}) = \det(\mathbf{P}) \det(\mathbf{D}) \det(\mathbf{P}^{-1}) = \det(\mathbf{P}) \det(\mathbf{D}) \frac{1}{\det(\mathbf{P})} = \det(\mathbf{D}).$$

- $\text{tr}(\mathbf{A}) = \lambda_1 + \cdots + \lambda_n$, ya que

$$\begin{aligned} \text{tr}(\mathbf{A}) &= \sum_{i=1}^n [\mathbf{PDP}^{-1}]_{ii} \\ &\stackrel{(a)}{=} \sum_{i=1}^n \sum_{j=1}^n \mathbf{P}_{ij} [\mathbf{DP}^{-1}]_{ji} \\ &\stackrel{(b)}{=} \sum_{i=1}^n \sum_{j=1}^n \mathbf{P}_{ij} \mathbf{D}_{jj} \mathbf{P}_{ji}^{-1} \\ &= \sum_{j=1}^n \mathbf{D}_{jj} \sum_{i=1}^n \mathbf{P}_{ji}^{-1} \mathbf{P}_{ij} \\ &= \sum_{j=1}^n \mathbf{D}_{jj} \mathbf{I}_{jj} \\ &= \text{tr}(\mathbf{D}), \end{aligned}$$

donde en (a) se consideró que

$$[\mathbf{AB}]_{ii} = \sum_{j=1}^n \mathbf{A}_{ij} \mathbf{B}_{ji}$$

y en (b) que si \mathbf{D} es una matriz diagonal,

$$[\mathbf{DA}]_{ij} = \mathbf{D}_{ii} \mathbf{A}_{ij}.$$

Los dos últimos resultados son generales y se cumplen para cualquier matriz cuadrada.

Algunas propiedades de matrices cuadradas

- *Traza del producto de matrices:* sean las matrices \mathbf{A} y \mathbf{B} de dimensiones $n \times m$ y $m \times n$ respectivamente. Se cumple que $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$. Efectivamente,

$$\text{tr}(\mathbf{AB}) = \sum_{i=1}^n [\mathbf{AB}]_{ii} = \sum_{i=1}^n \sum_{j=1}^m \mathbf{A}_{ij} \mathbf{B}_{ji} = \sum_{j=1}^m \sum_{i=1}^n \mathbf{B}_{ji} \mathbf{A}_{ij} = \sum_{j=1}^m [\mathbf{BA}]_{jj} = \text{tr}(\mathbf{BA}).$$

- La traza es invariable a permutaciones cíclicas. Por ejemplo

$$\text{tr}(\mathbf{ABCD}) = \text{tr}(\mathbf{BCDA}) = \text{tr}(\mathbf{CDAB}) = \text{tr}(\mathbf{DABC}),$$

Esto es consecuencia de la propiedad anterior.

- $\text{tr}(\mathbf{A}) = \sum_{i=1}^n \lambda_i$ y $\det(\mathbf{A}) = \prod_{i=1}^n \lambda_i$. Efectivamente, sea \mathbf{A} una matriz $n \times n$. Por definición, su polinomio característico es

$$\begin{aligned} p(\lambda) = \det(\mathbf{A} - \lambda \mathbf{I}) &= \begin{vmatrix} a_{11} - \lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{21} - \lambda & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} - \lambda \end{vmatrix} \\ &\stackrel{(a)}{=} (a_{11} - \lambda) \times (a_{22} - \lambda) \times \cdots \times (a_{nn} - \lambda) + b\lambda + c \\ &\stackrel{(b)}{=} (-1)^n \lambda^n + (-1)^{n-1} \lambda^{n-1} (a_{11} + \cdots + a_{nn}) + \cdots + \det(\mathbf{A}), \end{aligned}$$

donde en (a) el primer sumando es el producto de la diagonal y se tuvo en cuenta que los demás sumandos del determinante son de grado 1 o 0 en λ , y en (b) se desarrolló el primer sumando considerando solamente los coeficientes de λ^n , λ^{n-1} y el término independiente. Se observa que el coeficiente que multiplica al término de grado $n-1$ es $\text{tr}(\mathbf{A})$ y el término independiente es $\det(\mathbf{A})$. Por otro lado, considerando al polinomio característico factorizado en las raíces $\lambda_1, \dots, \lambda_n$, se tiene que

$$\begin{aligned} p(\lambda) &= (\lambda_1 - \lambda) \times \cdots \times (\lambda_n - \lambda) \\ &= (-1)^n \lambda^n + (-1)^{n-1} \lambda^{n-1} (\lambda_1 + \cdots + \lambda_n) + \cdots + \lambda_1 \times \cdots \times \lambda_n. \end{aligned}$$

Finalmente, igualando los coeficientes de los términos de grado $n-1$ e independiente en ambos polinomios se obtiene que

$$\text{tr}(\mathbf{A}) = \sum_{i=1}^n \lambda_i, \quad \det(\mathbf{A}) = \prod_{i=1}^n \lambda_i.$$

Apéndice I

Ecuación analítica de la elipse

Una elipse puede ser definida como un lugar geométrico en el plano euclideo. Dados dos puntos F_1 y F_2 denominados *focos*, la elipse es el conjunto de puntos P tal que la suma de las distancias $|PF_1|$ y $|PF_2|$ es constante igual a $2a$, es decir

$$E = \{P \in \mathbb{R}^2 \mid |PF_2| + |PF_1| = 2a\}.$$

I.1. Elipse estándar

La ecuación de una elipse estándar centrada en el origen de ancho $2a$ y altura $2b$ es

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1. \quad (\text{I.1})$$

Los focos son $F_1 = (c, 0)$ y $F_2 = (-c, 0)$, donde $c = \sqrt{a^2 - b^2}$ asumiendo que $a > b$. Los parámetros se muestran en la figura I.1. La ecuación paramétrica estándar es

$$\begin{aligned} x(t) &= a \cos t \\ y(t) &= b \sin t \end{aligned}, \quad \text{para } 0 \leq t < 2\pi. \quad (\text{I.2})$$

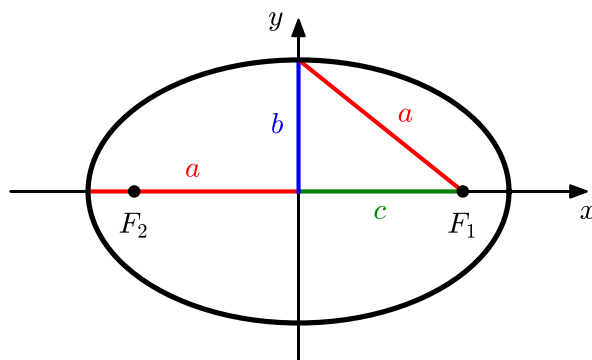


Figura I.1: Parámetros de una elipse. F_1 y F_2 son los focos, a es el semieje mayor y b el semieje menor.

I.2. Elipse rotada centrada en el origen

Se deduce a continuación la ecuación de una elipse centrada en el origen con el eje rotado un ángulo θ respecto al eje x como se muestra en la figura I.2. En el eje de coordenadas (x', y') rotado un ángulo θ , la elipse es la elipse estándar y está dada por la ecuación

$$\frac{x'^2}{a^2} + \frac{y'^2}{b^2} = 1.$$

Para determinar la ecuación de la elipse en el eje (x, y) , hay que expresar las coordenadas en el eje

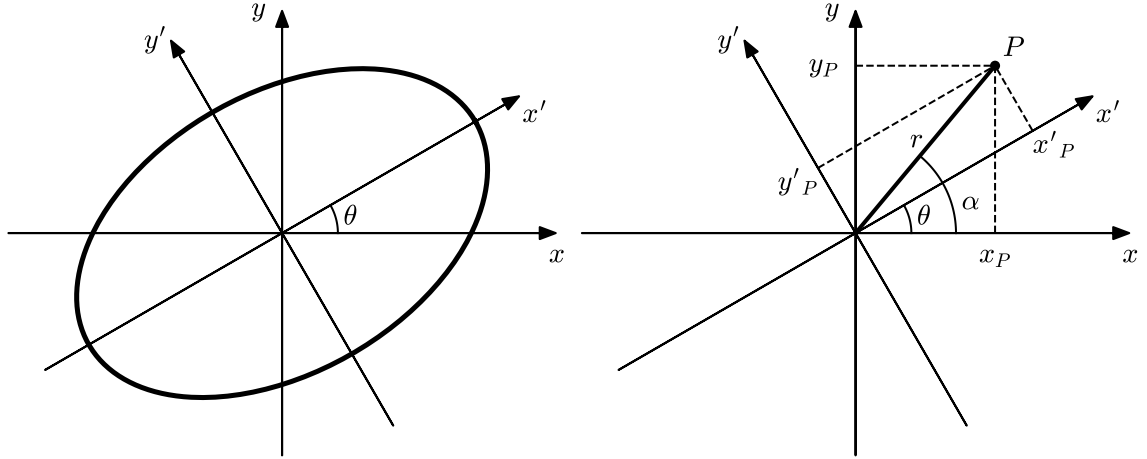


Figura I.2: Elipse rotada un ángulo θ respecto al eje x (izquierda) y rotación del eje de coordenadas (derecha).

(x', y') en función de las coordenadas en el eje (x, y) . Se considera el caso general de la rotación de un eje de coordenadas¹ mostrado en la figura I.2. Sea el punto P con coordenadas (x_P, y_P) en el eje original. Para determinar las coordenadas (x'_P, y'_P) en un eje (x', y') rotado un ángulo θ respecto al eje original, se parte considerando las coordenadas polares (r, α) del punto P . De esta forma, las coordenadas en los dos sistemas son

$$\begin{aligned} x_P &= r \cos \alpha & y & & x'_P &= r \cos(\alpha - \theta) \\ y_P &= r \sin \alpha & & & y'_P &= r \sin(\alpha - \theta). \end{aligned}$$

Empleando las identidades trigonométricas $\cos(\alpha \pm \beta) = \cos \alpha \cos \beta \mp \sin \alpha \sin \beta$ y $\sin(\alpha \pm \beta) = \sin \alpha \cos \beta \pm \cos \alpha \sin \beta$ se obtiene que las coordenadas en el eje rotado son

$$\begin{aligned} x'_P &= r \cos \alpha \cos \theta + r \sin \alpha \sin \theta \\ y'_P &= r \sin \alpha \cos \theta - r \cos \alpha \sin \theta \end{aligned}$$

y considerando que $x_P = r \cos \alpha$ y $y_P = r \sin \alpha$ se obtiene que

$$\begin{aligned} x'_P &= x_P \cos \theta + y_P \sin \theta \\ y'_P &= -x_P \sin \theta + y_P \cos \theta. \end{aligned} \tag{I.3}$$

Además, la transformación inversa es

$$\begin{aligned} x_P &= x'_P \cos \theta - y'_P \sin \theta \\ y_P &= x'_P \sin \theta + y'_P \cos \theta, \end{aligned} \tag{I.4}$$

la cual se deduce de forma análoga pensando al eje (x', y') como el original y entonces el eje (x, y) está rotado un ángulo $-\theta$ por lo que alcanza con reemplazar θ por $-\theta$ en la ecuación I.3. En el caso de la elipse rotada, de la ecuación I.3, las coordenadas del eje (x', y') rotado en función de las coordenadas en el eje (x, y) original son

$$\begin{aligned} x' &= x \cos \theta + y \sin \theta \\ y' &= -x \sin \theta + y \cos \theta, \end{aligned}$$

y sustituyendo x' y y' en la ecuación de la elipse estándar se obtiene que la ecuación de una elipse centrada en el origen y rotada un ángulo θ es

$$\frac{(x \cos \theta + y \sin \theta)^2}{a^2} + \frac{(-x \sin \theta + y \cos \theta)^2}{b^2} = 1. \tag{I.5}$$

¹Ver https://en.wikipedia.org/wiki/Rotation_of_axes.

Operando, se obtiene que la ecuación implícita es

$$(a^2 \sin^2 \theta + b^2 \cos^2 \theta)x^2 + 2xy(b^2 - a^2) \sin \theta \cos \theta + (a^2 \cos^2 \theta + b^2 \sin^2 \theta)y^2 - a^2b^2 = 0.$$

Considerando que la ecuación implícita de una elipse genérica es

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0,$$

con $B^2 - 4AC < 0$, para una elipse centrada en el origen rotada un ángulo θ y ejes de largo $2a$ y $2b$ se tiene que

$$\begin{aligned} A &= a^2 \sin^2 \theta + b^2 \cos^2 \theta \\ B &= 2(b^2 - a^2) \sin \theta \cos \theta \\ C &= a^2 \cos^2 \theta + b^2 \sin^2 \theta \\ D &= 0 \\ E &= 0 \\ F &= -a^2b^2. \end{aligned}$$

Para obtener la ecuación paramétrica, se parte considerando que en el eje rotado la ecuación paramétrica es (ver la ecuación I.2)

$$\begin{aligned} x'(t) &= a \cos t \\ y'(t) &= b \sin t \end{aligned}, \quad \text{para} \quad 0 \leq t < 2\pi.$$

y transformando el eje de coordenadas al eje original mediante la ecuación I.4, se obtiene que

$$\begin{aligned} x(t) &= a \cos t \cos \theta - b \sin t \sin \theta \\ y(t) &= a \cos t \sin \theta + b \sin t \cos \theta \end{aligned}, \quad \text{para} \quad 0 \leq t < 2\pi. \quad (\text{I.6})$$

I.3. Ecuación general

La ecuación implícita general de una elipse es²

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0, \quad (\text{I.7})$$

con $B^2 - 4AC < 0$. Para una elipse de centro (x_0, y_0) rotada un ángulo θ y ejes de largo $2a$ y $2b$ se tiene que

$$\begin{aligned} A &= a^2 \sin^2 \theta + b^2 \cos^2 \theta \\ B &= 2(b^2 - a^2) \sin \theta \cos \theta \\ C &= a^2 \cos^2 \theta + b^2 \sin^2 \theta \\ D &= -2Ax_0 - By_0 \\ E &= -Bx_0 - 2Cy_0 \\ F &= Ax_0^2 + Bx_0y_0 + Cy_0^2 - a^2b^2. \end{aligned}$$

Estas expresiones pueden derivarse a partir de la ecuación estándar

$$\frac{x'^2}{a^2} + \frac{y'^2}{b^2} = 1$$

y aplicando una transformación afín del eje de coordenadas (x', y')

$$\begin{aligned} x' &= (x - x_0) \cos \theta + (y - y_0) \sin \theta \\ y' &= -(x - x_0) \sin \theta + (y - y_0) \cos \theta, \end{aligned}$$

²Ver <https://en.wikipedia.org/wiki/Ellipse>.

como se realizó para la elipse rotada centrada en el origen. Los parámetros de la ecuación implícita pueden obtenerse a partir de los coeficientes como

$$\begin{aligned}
 a, b &= \frac{-\sqrt{2[AE^2 + CD^2 - BDE + (B^2 - 4AC)F]} \left[(A + C) \pm \sqrt{(A - C)^2 + B^2} \right]}{B^2 - 4AC} \\
 x_0 &= \frac{2CD - BE}{B^2 - 4AC} \\
 y_0 &= \frac{2AE - BD}{B^2 - 4AC} \\
 \theta &= \begin{cases} \arctan \left[\frac{1}{B} (C - A - \sqrt{(A - C)^2 + B^2}) \right] & \text{si } B \neq 0 \\ 0 & \text{si } B = 0, A < C \\ \frac{\pi}{2} & \text{si } B = 0, A > C. \end{cases}
 \end{aligned} \tag{I.8}$$

Bibliografía

- [1] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*. Prentice Hall, 1st ed., 1993.
- [2] A. Papoulis and S. U. Pillai, *Probability, random variables, and stochastic processes*. Tata McGraw-Hill Education, 4th ed., 2002.
- [3] B. Porat and B. Friedlander, “Computation of the exact information matrix of Gaussian time series with stationary random components,” *IEEE transactions on acoustics, speech, and signal processing*, vol. 34, no. 1, pp. 118–130, 1986.
- [4] S. S. Haykin, *Adaptive filter theory*. Pearson Education Limited, 5th ed., 2014.
- [5] M. H. Hayes, *Statistical Digital Signal Processing and Modeling*. Wiley, 1st ed., 1996.
- [6] A. V. Oppenheim and R. W. Schaffer, *Discrete-Time Signal Processing*. Prentice Hall, 3rd ed., 2009.
- [7] J. Makhoul, “Linear prediction: A tutorial review,” *Proceedings of the IEEE*, vol. 63, no. 4, pp. 561–580, 1975.
- [8] B. Porat and B. Friedlander, “The exact Cramer-Rao bound for Gaussian autoregressive processes,” *IEEE transactions on aerospace and electronic systems*, no. 4, pp. 537–542, 1987.
- [9] P. Stoica and P. Babu, “The Gaussian data assumption leads to the largest Cramér-Rao bound [lecture notes],” *IEEE Signal Processing Magazine*, vol. 28, no. 3, pp. 132–133, 2011.
- [10] G. Casella and R. L. Berger, *Statistical Inference*. Cengage Learning, 2nd ed., 2001.
- [11] R. M. Norton, “The double exponential distribution: Using calculus to find a maximum likelihood estimator,” *The American Statistician*, vol. 38, no. 2, pp. 135–136, 1984.
- [12] M. Feder and E. Weinstein, “Parameter estimation of superimposed signals using the em algorithm,” *IEEE Transactions on acoustics, speech, and signal processing*, vol. 36, no. 4, pp. 477–489, 1988.
- [13] L. L. Scharf, *Statistical signal processing*. Addison-Wesley Reading, MA, 1991.
- [14] S. J. Orfanidis, *Optimum signal processing: An introduction*. Collier Macmillan, 2nd ed., 1988.
- [15] P. Stoica, R. L. Moses, B. Friedlander, and T. Soderstrom, “Maximum likelihood estimation of the parameters of multiple sinusoids from noisy measurements,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 3, pp. 378–392, 1989.
- [16] S. W. Smith, *The Scientist and Engineer’s Guide to Digital Signal Processing*. California Technical Pub., 1st ed., 1997.
- [17] D. Khoshnevisan, “Module 3: Some linear algebra,” 2014. University of Utah. Lecture Notes: Linear Models (Math 6010-1). <https://www.math.utah.edu/~davar/math6010/2014/SomeLinearAlgebra.pdf>. Last Checked: 2020-07-14.