UNIVERSIDAD DE CANTABRIA

Departamento de
Ingeniería de Comunicaciones

Técnicas polinómicas en procesado de señal: aplicaciones en interpolación, modelado y comunicaciones

Tesis Doctoral presentada para la obtención del título de
Doctor Ingeniero de Telecomunicación

Autor: Jesús María Ibáñez Díaz
Directores: Carlos J. Pantaleón Prieto
Ignacio Santamaría Caballero

Santander, mayo de 2004

Certificado del director de la tesis

D. Ignacio Santamaría Caballero, Profesor Titular de la Universidad de Cantabria en el Área de Teoría de la Señal y Comunicaciones,

HACE CONSTAR:

Que la tesis titulada "**Técnicas polinómicas en procesado de señal: aplicaciones en interpolación, modelado y comunicaciones**" ha sido realizada por D. **Jesús María Ibáñez Díaz** en el Departamento de Ingeniería de Comunicaciones de la Universidad de Cantabria bajo mi dirección, y que reúne las condiciones exigidas a los trabajos de Doctorado.

Santander, mayo de 2004

Fdo.: Ignacio Santamaría Caballero

A Carlos
y a David.

Agradecimientos

En primer lugar, quiero dar las gracias a un gran amigo, al director de esta Tesis, a Carlos Pantaleón. Las gracias por haberme honrado con su amistad y cariño y porque, si existe alguna idea valiosa en este trabajo, sin duda surgió de su brillante y especial pensamiento. Así mismo, quiero agradecer a Ignacio Santamaría, también amigo, también director de Tesis, su eterna paciencia, comprensión y ayuda. Esto no hubiera llegado a ningún sitio (aceptable) sin ti.

Cómo no mencionar a Luis VIELVA, por sus permanentes y acertadas lecciones de estilo y por su apoyo constante y sin "cotas". De igual forma, quiero dejar constancia de la ayuda y amistad que he recibido durante tantos años de todos los miembros del Grupo de Tratamiento Avanzado de Señal, y en especial de Marce, David Blanco, David Luengo, Jose, Rafa, Diego y Carlos Calderón; y de Marta, la más querida del grupo, aunque no pertenezca a él.

A mi familia, sobre todo a mis padres, deciros que, aunque lo disimulo muy bien, os quiero y os agradezco todo lo que habéis hecho por mi durante todos y cada uno de los días de mi vida; entre otras muchas cosas por soportarme, que no es poco. También a vosotros, mis amigos de siempre, os agradezco vuestra preocupación, tanto cuando me preguntabais por la Tesis como cuando no lo hacíais, sin duda tanta compasión os honra.

Y por último, Geni, a ti no voy a agradecerte ni tu apoyo, ni tu ayuda en la corrección del texto de esta Tesis; es tanto lo que te debo que es imposible dar cuenta de todo ello. Dada mi inutilidad a la hora de expresarlo, únicamente deseo ser capaz de corresponderte cada día.

Resumen

La interpolación funcional es un problema al que se ha enfrentado la ciencia desde la antigüedad, surgiendo a lo largo de la Historia multitud de técnicas que han posibilitado su resolución y aplicación en áreas como el procesado y análisis de señales, las comunicaciones, el análisis numérico, el diseño asistido por ordenador, etc. Sin duda, las técnicas polinómicas han sido las más utilizadas en este ámbito, impulsadas, en las últimas décadas, por la aparición de los splines. A pesar de ello, estas técnicas polinómicas de interpolación no han tenido un impacto importante en el procesado de señal y las comunicaciones, donde el Teorema de Muestreo sigue desempeñando el papel protagonista.

Tras llevar a cabo una revisión de las técnicas y los conceptos teóricos relativos a la aproximación funcional, la interpolación y el modelado, la presente Tesis analiza las prestaciones de las técnicas polinómicas de interpolación (en especial de los splines) cuando se tienen en cuenta parámetros de diseño reales, como pueden ser el sobremuestreo, la precisión finita y el coste computacional. El hecho de que, en ciertos casos prácticos, las prestaciones de dichas técnicas sean superiores a las obtenidas mediante los métodos convencionales basados en el Teorema de Muestreo, sirve de motivación y justificación para la búsqueda, en esta Tesis, de técnicas polinómicas novedosas aplicables a la interpolación-reconstrucción de señales, al modelado de sistemas no lineales y a las comunicaciones digitales.

Como primera aportación, se propone una nueva familia de splines, denominada α-splines, que incrementa la flexibilidad aportada por la interpolación spline, al permitir una transición parametrizada entre splines de órdenes arbitrarios. Así, se han desarrollado las expresiones de las bases α-spline y los algoritmos de interpolación correspondientes. Dada la difusión e importancia de la interpolación lineal y de la interpolación spline cúbica, se estudia en detalle un caso particular de la interpolación α-spline que proporciona una solución de compromiso entre ambas técnicas, pues a su reducido coste computacional se une la posibilidad de controlar ciertas características de la interpolación, como la suavidad, la varianza o el área residual.

Dadas las ventajas aportadas por la familia α-spline, se amplía su uso al

problema del modelado no lineal. En concreto, se propone una extensión del modelo lineal a tramos canónico mediante el empleo de la base α-spline como función de activación, obteniéndose un modelo que preserva las propiedades de simplicidad y capacidad de aproximación originales y que aporta las características de derivabilidad y suavidad propias de la interpolación α-spline. Además de la presentación del modelo α-spline y de su metodología de entrenamiento, en esta Tesis se propone su empleo en el modelado de transistores y en el modelado de amplificadores de potencia de radiofrecuencia. En este último caso se hace uso de medidas temporales entrada-salida obtenidas mediante una técnica de caracterización novedosa, basada en submuestreo, que permite conocer el comportamiento del amplificador cuando se excita con señales de gran ancho de banda y envolvente no constante. Sin duda este es un aspecto de extrema importancia en los actuales sistemas de comunicaciones digitales de banda ancha que emplean modulaciones complejas.

Finalmente, se abordan algunos aspectos del problema de las comunicaciones digitales aprovechando el conocido paralelismo entre las mismas y la interpolación funcional. Se propone, dadas sus especiales características, la utilización de las bases spline, y de las bases α-spline desarrolladas en esta Tesis, como filtros conformadores, posibilitando implementar sistemas adaptados y libres de interferencia entre símbolos con una destacable simplicidad tanto computacional como hardware. Esta metodología es susceptible de ser extendida a pulsos de forma arbitraria, lo que abre la puerta a nuevos diseños. Así mismo, se propone una generalización de la teoría de parejas biortogonales que incluye la restricción de respuesta frecuencial global positiva, lo que permite su descomposición en dos filtros adaptados. Esta técnica permite una implementación FIR en transmisión y recepción manteniendo las condiciones de ISI nula y adaptación, tal y como se comprueba en varias aplicaciones prácticas presentadas.

En definitiva, esta Tesis aporta una serie de técnicas polinómicas en el ámbito del procesado de señal, propone afrontar determinados problemas desde una perspectiva distinta a la proporcionada por el Teorema de Muestreo, y presenta, en ese contexto, numerosas aplicaciones prácticas de interpolación-reconstrucción de señales, de modelado de dispositivos no lineales y de comunicaciones digitales.

Summary

Science has tackled the functional interpolation problem since the earliest times, and a number of techniques have been developed in different applications such as signal processing and analysis, communications, numerical analysis, computer aided design, etc. Among all the interpolation techniques, those based on polynomials are probably the most commonly applied. In particular, during the last decades, polynomial interpolation techniques have played an important role due to the development of splines. However, polynomial interpolation techniques have not attained a similar impact within the signal processing community, maybe due to the prevalence of the Sampling Theorem.

Firstly, this thesis reviews the techniques and theoretical concepts related with functional approximation, interpolation and modelling. Secondly, it analyzes the performance of polynomial interpolation techniques (mainly splines) when real design constrains (e.g., oversampling, finite precision or computational burden), are taken into account. It is shown that, in some practical applications, spline-based techniques offer better performance than conventional methods based on the Sampling Theorem. This fact motivates this work and justifies the search of novel polynomial techniques as well as their application to different problems in signal interpolation-reconstruction, nonlinear system modelling and digital communications.

As a first contribution of this thesis, we propose a new family of splines, denoted as α-splines, which allows a parameterized and soft transition between splines of arbitrary order, therefore increasing the flexibility of spline interpolation. The α-spline basis and its corresponding interpolation algorithms are developed for this new family. Particular attention is paid to the α-spline basis that allows a smooth transition between linear and cubic splines: it has a low computational cost and, at the same time, it allows to control some interpolation characteristics like smoothness, variance and residual area.

Due to the advantages of α-spline interpolation, we extend and generalize the α-spline basis to nonlinear modelling problems. Specifically, this thesis proposes an extension of the canonical piecewise linear model that incorporates the α-spline basis as activation function. The new model preserves the simplicity

and approximation accuracy of the original piecewise linear model and provides new advantages such as the smoothness and derivability of α-spline interpolation. We describe the α-spline model, its training algorithm and present two different applications in communications: MESFET transistor modeling and high power amplifier modeling. In the latter case, we use a novel measurement system that acquires input-output temporal signal obtained using subsampling techniques: this permits a complete characterization of the amplifier for wideband or nonconstant envelope input signals, which are typically used in modern digital communications systems (for instance, OFDM).

Finally, we exploit the well-known parallelism between communications and functional interpolation in some digital communications problems. Specifically, this thesis explores the use of the spline basis (as well as the α-spline basis) as pulse shaping filters, producing ISI-free and matched systems with a low hardware and computational complexity. The proposed methodology can be easily extended to deal with arbitrary shaped pulses. Furthermore, when the receiver uses some oversampling, in addition to ISI-free and matched receiver it is possible an all-FIR implementation. To this end, in the thesis we present a generalization of the biorthogonal partners theory that incorporates additional constrains ensuring that the transmit and receive filters global response is a feasible autocorrelation function; then it can be factored into a pair of matched filters. This technique allows an all-FIR implementation of the transmitter and receiver while keeping the ISI free and matched conditions.

In summary, this thesis proposes some new polynomial techniques and tries to encourage their use in practice as an alternative to the "bandlimited" point of view derived from the Sampling Theorem. Moreover, it describes several practical applications of the proposed techniques in signal interpolation-reconstruction, nonlinear devices modeling and digital communications.

Índice general

Resumen I

Summary III

1. Introducción 1
1.1. Muestreo e interpolación 1
1.2. Planteamiento y motivación de la Tesis 3
1.3. Organización de la Tesis 7

2. Conceptos teóricos 9
2.1. Introducción 9
2.2. Espacio de señal 9
2.3. Aproximación, interpolación y modelado 13
2.3.1. Polinomios como interpoladores y aproximadores 15
2.4. Aproximación 16
2.4.1. Aproximación de mínimo error cuadrático 17
2.4.2. Aproximación uniforme o minimax 20
2.5. Interpolación 21
2.5.1. El Teorema de Muestreo 22
2.5.2. Técnicas de interpolación polinómicas 22
2.5.3. Divergencia de la interpolación polinómica 23
2.5.4. Splines 24
2.6. Muestreo y reconstrucción en espacios de Hilbert 30
2.6.1. Revisión del Teorema de Muestreo 30
2.6.2. Muestreo en espacios invariantes frente a desplazamientos 33
2.7. Modelado 36
2.7.1. Técnicas polinómicas de modelado: consideraciones generales 37

3. Técnicas polinómicas en procesado de señal **41**
3.1. Introducción 41
3.2. Interpolación sinc frente a la interpolación spline 42
3.2.1. Modelo de la señal a interpolar 42
3.2.2. Simulaciones 46
3.2.3. Discusión de los resultados 53
3.3. Muestreo simultáneo mediante filtros de Lagrange 54
3.3.1. Compensación de retardo fraccionario 55
3.3.2. Filtro de Lagrange de retardo fraccionario 56
3.3.3. Análisis del error de interpolación 57
3.3.4. Simulaciones 61
3.3.5. Discusión de los resultados 63
3.4. Conclusiones 64

4. α-splines **65**
4.1. Introducción 65
4.2. Familia α-spline 66
4.2.1. Propiedades 68
4.3. Base α-spline 70
4.4. Interpolación α-spline 74
4.4.1. Solución matricial 74
4.4.2. Solución mediante filtrado digital 76
4.4.3. Interpolación rápida mediante α-splines 80
4.5. α-spline cardinal 82
4.6. Suavidad de la interpolación α-spline 84
4.6.1. Concepto de suavidad 85
4.6.2. Suavidad de la interpolación α-spline de señales deterministas periódicas 86
4.6.3. Suavidad de la interpolación α-spline de señales aleatorias 90
4.7. Varianza de la interpolación α-spline 95
4.8. Área residual de la interpolación α-spline 98
4.9. Comportamiento de la interpolación 100
4.10. Generalización para muestreo no uniforme 102
4.10.1. Base α-spline no uniforme 102
4.10.2. Interpolación α-spline no uniforme 104
4.11. Simulaciones 107
4.11.1. Interpolación α-spline de tonos 107
4.11.2. Interpolación α-spline de ruido blanco 110
4.12. Aplicaciones de los α-splines 110
4.12.1. Control de suavidad, varianza o área residual en la interpolación α-spline 110
4.12.2. Interpolación α-spline de señales no estacionarias 111
4.12.3. Interpolación no uniforme α-spline 112

4.13. Conclusiones 116

5. Modelado α-spline 117
5.1. Introducción 117
5.2. Modelo α-spline 119
5.2.1. El modelo lineal a tramos canónico 119
5.2.2. Suavizado de la función valor absoluto 120
5.2.3. El modelo αPWL 123
5.2.4. Suavidad del modelo αPWL 124
5.2.5. Entrenamiento del modelo 128
5.2.6. Estimación del parámetro de suavidad 131
5.2.7. Prestaciones 133
5.3. Modelado de amplificadores de potencia 133
5.3.1. Estado del arte 133
5.3.2. Sistema de caracterización basado en submuestreo . . . 135
5.3.3. Medidas experimentales 138
5.3.4. Resultados y comparación de modelos 143
5.4. Modelo gran señal de un transistor HEMT 154
5.4.1. Resultados del modelo αPWL 155
5.5. Conclusiones 158

6. Técnicas Interpolatorias en Comunicaciones 161
6.1. Introducción 161
6.2. Conceptos básicos 163
6.2.1. Interferencia entre símbolos 163
6.2.2. Filtro adaptado 164
6.3. Comunicaciones digitales vistas como problema de interpolación 165
6.4. Sistemas de comunicación basados en splines 167
6.4.1. Interpolación spline 168
6.4.2. Filtro spline cúbico para comunicaciones libres de ISI y adaptadas 170
6.4.3. Aspectos relativos a la implementación 173
6.5. Sistemas de comunicación α-spline 176
6.6. Diseño de sistemas de comunicación adaptados y sin ISI 178
6.6.1. Solución libre de ISI 180
6.6.2. Solución adaptada 182
6.6.3. Aspectos relativos a la implementación 183
6.7. Implementación FIR en sistemas con sobremuestreo 184
6.7.1. Introducción 184
6.7.2. Planteamiento del problema 185
6.7.3. Parejas biortogonales de fase lineal 188
6.7.4. Parejas biortogonales semidefinidas positivas 191
6.7.5. Ejemplos de diseño 193

6.8. Conclusiones 195

7. Conclusiones, aportaciones y líneas futuras 199
7.1. Conclusiones y aportaciones 199
7.2. Líneas futuras 204
7.2.1. Implementación FIR causal de la interpolación spline y α-spline 204
7.2.2. Interpolación rápida α-spline en la generación de gráficos 3D 204
7.2.3. Interpolación adaptativa con control de suavidad 205
7.2.4. Sistemas de comunicaciones 205
7.2.5. Máquina de Vectores Soporte con núcleo α-spline 205

A. Filtro de retardo fraccionario de Lagrange 207
A.1. Diseño de filtro con respuesta frecuencial máximamente plana . 207

B. Base de Riesz α-spline 209
B.1. Cotas Riesz de la base α-spline 209

C. Varianza y suavidad de interpolación α-spline 211
C.1. Autocorrelaciones del filtro directo α-spline 211
C.2. Varianza de señales deterministas periódicas 212
C.3. Varianza de señales aleatorias 215
C.3.1. Varianza de la interpolación α-spline de señales blancas 216
C.3.2. Varianza de la interpolación α-spline de señales aleatorias senoidales 216
C.4. Área residual de señales deterministas periódicas 217
C.5. Área residual de señales aleatorias 220
C.5.1. Señales blancas 221
C.5.2. Señales aleatorias senoidales 222

D. Expresiones de la base α-spline no uniforme 223
D.1. Expresiones de la base α-spline no uniforme 223

E. Modelado α-spline 229
E.1. Comportamiento del término de regularización 229

F. Publicaciones de la Tesis 231
F.1. Revistas 231
F.2. Congresos internacionales 232
F.3. Congresos nacionales 233

G. Símbolos y abreviaturas **235**
G.1. Simbología 235
G.2. Abreviaturas 236

Bibliografía **239**

Índice de figuras

1.1. Complejidad de distintas técnicas interpolatorias. 4

2.1. a) Aproximación, b) interpolación y c) modelado. 14
2.2. Esquema de muestreo y reconstrucción dictado por el Teorema de Muestreo, donde $h(t) = \varphi(-t) = \varphi(t) = \mathrm{sinc}(t)$ [Unser00]. . . 32
2.3. Esquemas de muestreo: a) aproximación mínimocuadrática; b) interpolación. 35

3.1. La señal de voz: a) espectro promedio, b) filtro de ponderación sofométrica. 47
3.2. Esquema de la interpolación de señales estrictamente limitadas en banda. 48
3.3. Interpolación de señales estrictamente limitadas en banda. Valores para los que se igualan las prestaciones de la interpolación spline y sinc (línea continua); y spline y sinc truncada (línea discontinua). 49
3.4. Interpolación sinc y spline de señales blancas con frecuencia máxima distribuida exponencialmente. 50
3.5. Interpolación sinc y spline de señales blancas con frecuencia máxima distribuida uniformemente. 51
3.6. Interpolación de la señal de voz. Filtro antisolapamiento ideal. . 52
3.7. Interpolación de la señal de voz. Filtro antisolapamiento de orden ocho. 53
3.8. $C_N(D)$ en función del orden del filtro para un retardo fraccional de $d = 0{,}5$. 61
3.9. Número de bits significativos para ciertos órdenes del filtro en función del factor de sobremuestreo. Simulación en línea sólida y teórica en línea discontinua. 62
3.10. Esquema de la interpolación multitasa. 63

4.1. Base α-spline, $\beta_{\alpha}^{[p\to p+q]}(t)$, para $p = 1, 3, 5$, $q = 2$ y α variando entre 0 y 1. 68
4.2. Base α-spline, $\beta_{\alpha}(t)$, para distintos valores del parámetro α. . . 71
4.3. Derivadas de la función base α-spline, para distintos valores del parámetro α. 73
4.4. Esquema de interpolación α-spline. 77
4.5. Interpolación α-spline mediante filtrado digital. 78
4.6. α-spline cardinal: respuesta al impulso. 83
4.7. α-spline cardinal: respuesta en frecuencia. 84
4.8. Derivada segunda de la interpolación α-spline determinista: a) derivación continua, b) derivación discreta. 86
4.9. Región (área sombreada) donde la suavidad de la interpolación α-spline decrece con α. 90
4.10. Potencia de tensión de la interpolación α-spline de señales blancas de potencia unidad y su derivada frente a α. 94
4.11. Comportamiento respecto a α de la potencia de tensión (izda.), varianza (centro), área residual (dcha.) y las correspondientes derivadas (abajo) de la interpolación α-spline de señales blancas de varianza unidad. 101
4.12. Regiones (áreas sombreadas) donde la suavidad (izquierda), la varianza (centro) y el área residual (derecha) de la interpolación α-spline decrecen con α. 102
4.13. Base α-spline (arriba) y su tercera derivada (abajo) correspondientes al vector de nodos $[-2, -0{,}4, 0{,}8, 1{,}4, 2]$ para $\alpha = 0{,}2$ (izda.) y $\alpha = 0{,}7$ (dcha.). 103
4.14. Interpolación y bases α-spline para un conjunto de nodos dado 105
4.15. Interpolación α-spline de una señal multitono y su segunda derivada para $\alpha = 0{,}3$ y $\alpha = 0{,}8$. 109
4.16. Interpolación α-spline de una señal multitono para distintos valores de α. De izquierda a derecha: potencia de tensión, varianza y área residual. 109
4.17. Interpolación α-spline de señales blancas (varianza uno). De izquierda a derecha: potencia de tensión, varianza y área residual. 110
4.18. Interpolación α-spline (arriba) y su derivada segunda (abajo) utilizando múltiples valores de α. 111
4.19. Interpolación no uniforme α-spline ($\alpha = 0{,}328$) frente a técnicas NNR, PWL y spline cúbica. 113
4.20. Comparativa de técnicas de interpolación no uniforme aplicadas a la velocidad de un flujo turbulento. 115

5.1. Función de suavizado α-spline 122

5.2. Funciones de suavizado (arriba) y sus dos primeras derivadas para una curvatura $K = 4$ ($\alpha = 0{,}5$). 123
5.3. Función base $\varphi(\langle \boldsymbol{\nu}, \mathbf{x} \rangle - \beta, \alpha)$ en un espacio bidimensional y una frontera definida por la recta $x_1 - 0{,}25x_2 = 0$. 124
5.4. Región de integración para un espacio de entrada bidimensional. 127
5.5. Esquema de caracterización temporal de HPA de radiofrecuencia: a) tradicional, b) propuesto. 138
5.6. Banco de caracterización de amplificadores mediante técnica de submuestreo. 139
5.7. Densidad espectral de potencia de entrada (arriba) y salida (abajo) para una señal OFDM ($BW_i = 1$ MHz, $P_i = 0{,}3$ dBm). . . . 141
5.8. Amplitud de la envolvente compleja para una señal OFDM con $BW_i = 1$ MHz y $P_i = 0{,}3$ dBm. 142
5.9. Gráfica entrada-salida de la amplitud de la envolvente compleja de una señal OFDM ($BW_i = 1$ MHz, $P_i = 0{,}3$ dBm). 142
5.10. Gráfica entrada-salida de la amplitud de la envolvente compleja de una señal DSB-AM ($BW_i = 40$ KHz, $P_i = -1{,}6$ dBm). . . . 143
5.11. Constelación a la salida del amplificador para una señal de entrada OFDM-64QAM ($BW_i = 10$ MHz, $P_i = 3$ dBm). 144
5.12. Modelo polinómico de envolvente. 147
5.13. Información mutua cuadrática entre $y_I[n]$ y $x_I[n-k]$ en función del retardo k para una señal de entrada OFDM-QPSK con $P_i = 9$ dBm y diferentes anchos de banda. 150
5.14. Señal de salida medida (línea de puntos) y estimada por el modelo αPWL para una señal de entrada OFDM-QPSK de $BW_i = 6$ MHz y $P_i = 3$ dBm. 151
5.15. SER para una potencia de entrada de $P_i = 0$ dBm (izquierda) y $P_i = 9$ dBm (derecha). 153
5.16. ΔACPR para una potencia de entrada de $P_i = 0$ dBm (izquierda) y $P_i = 9$ dBm (derecha). 154
5.17. Curvas experimentales y generadas por el modelo αPWL para el transistor HEMT (polarización: $V_{gs0} = 0$ V, $V_{ds0} = 2$ V y $P_o = 1$ mW): a) corriente de drenador, I_{ds}, b) derivada con respecto a v_{gs}, c) derivada con respecto a v_{ds}. 158

6.1. Sistema de comunicaciones digitales banda base: a) esquema general, b) modelo equivalente. 162
6.2. Filtro de Nyquist con factor de *roll-off* $r = 0$, 0,5 y 1: a) respuesta al impulso, b) respuesta en frecuencia. 164
6.3. Esquema de receptor de comunicaciones digiales banda base. . 165
6.4. Paralelismo interpolación-comunicaciones: a) sistema equivalente de comunicaciones digitales banda base, b) sistema interpolatorio. 167

6.5. Filtro conformador $\varphi_m(t/T)$ de grado $m = 3, 5, 9$ y ∞: a) respuesta al impulso, b) respuesta en frecuencia. 170
6.6. Sistema de comunicaciones basado en splines: a) esquema no adaptado, b) esquema adaptado. 171
6.7. Densidad espectral de potencia de los esquemas de modulación basados en splines frente al sistema NRZ. 172
6.8. Diagrama de ojo de conformación spline frente a coseno alzado. 173
6.9. Degradación de BER debido al truncamiento de los filtros digitales transmisor y receptor. 174
6.10. Implementación del transmisor spline cúbico: a) binario no adaptado, b) multinivel no adaptado, c) adaptado. 175
6.11. Sistema adaptado de comunicaciones basado en α-splines. . . . 177
6.12. Densidad espectral de potencia de la señal transmitida usando conformación α-spline. 178
6.13. Diagramas de ojo de comunicaciones con conformación α-spline. 179
6.14. Esquema de comunicaciones libre de ISI. 180
6.15. Esquema de comunicaciones adaptado y libre de ISI. 183
6.16. Modelo de un sistema de comunicaciones digitales paso bajo trabajando con un factor de sobremuestreo M. 185
6.17. Respuesta en frecuencia de un filtro conformador PSD-BP basado en el filtro B-spline frente a un filtro SQRC de la misma longitud. Sobremuestreo $M = 2$. 194
6.18. Respuesta al impulso (izquierda) y respuesta en frecuencia (derecha) del filtro BP-PSD de 25 coeficientes (línea continua), del filtro SQRC inicial de 19 coeficientes (puntos) y del filtro SQRC de 25 coeficientes (discontinua). Sobremuestreo $M = 3$. 196

C.1. Varianza de la interpolación α-spline de señales blancas (izquierda) y su derivada respecto a α (derecha). 217
C.2. Esquema de cálculo del área residual α-spline. 218
C.3. Área residual de la interpolación α-spline de señales blancas (izquierda) y su derivada respecto a α (derecha). 222

Índice de tablas

3.1. Coeficientes del filtro de retardo fraccionario de Lagrange. . . . 57
3.2. Comparación del número de coeficientes del FDF de Lagrange y del filtro polifase (entre paréntesis) requeridos para obtener una determinada resolución. 63

4.1. Filtros directos α-spline con coeficientes potencia de dos. 82
4.2. Valores simulados y analíticos de la potencia de tensión, varianza y área residual de la interpolación α-spline de una señal multitono para dos valores de α. 108
4.3. Interpolación no uniforme α-spline ($\alpha = 0{,}328$) frente a técnicas S/H, NNR, PWL y spline cúbica. 113
4.4. Interpolación no uniforme de la velocidad de un flujo turbulento ($\alpha = 0{,}289$). 115

5.1. Comparativa de los resultados de modelar el HEMT Phillips D02AH (4×30 μm). 157
5.2. Coste computacional del entrenamiento de los distintos modelos del HEMT Phillips D02AH (4×30 μm). 159

Capítulo 1

Introducción

1.1. Muestreo e interpolación

La interpolación, entendida como la búsqueda de una "curva geométrica que debe pasar por una serie puntos dados",[1] ya era conocida y utilizada por antiguas civilizaciones con el fin de realizar predicciones astronómicas que les permitieran confeccionar calendarios. Hoy en día, la interpolación desempeña un papel crucial en áreas como el procesado y análisis de señales, las comunicaciones, el análisis numérico, el diseño asistido por ordenador, etc.

Varios siglos antes de nuestra era, los babilonios construyeron su calendario en base a las llamadas efemérides, consistentes en datos acerca de la posición y el movimiento del sol, la luna y otros cuerpos celestes. Sin embargo, estas informaciones eran imposibles de obtener en determinadas ocasiones (ya fuera por cuestiones atmosféricas o porque durante ciertos periodos de tiempo los cuerpos no eran visibles), lo que motivó que recurrieran a técnicas de interpolación lineal (e incluso más complejas) para completar las ausencias de datos.

Desde entonces y hasta la revolución científica ocurrida en la Europa del siglo XVII, de forma más o menos dispersa en el espacio y en el tiempo, se desarrollaron y aplicaron técnicas interpolatorias. Así, en la Grecia clásica del siglo II a.c., Hiparco de Rodas utilizó interpolación lineal para calcular la posición de los cuerpos celestes, técnica que mejoró Ptolomeo cuatro siglos después. Posteriormente, existe constancia de la utilización de interpolación de segundo orden con idénticos fines en la China del siglo VII y una versión no equiespaciada en esa misma época en la India. A partir del siglo XI, los matemáticos árabes y persas tomaron el relevo en la aplicación de esos mismos conocimientos. Ya en el siglo XIV, un matemático chino describió una regla de interpolación basada

[1] *"To describe a geometrical curve which shall pass through any given points..."* frase extraída de la carta enviada por Newton a Oldenburg en 1676.

en diferencias finitas que puede ser considerada precursora de la formulación interpolatoria moderna.

Quizás sea Newton, en el siglo XVII, la figura clave en el desarrollo de las técnicas interpolatorias, las cuales fueron refinadas por distintos matemáticos europeos durante el siglo XVIII y principios del XIX, hasta la aparición, en 1812, de la fórmula de Newton-Gauss, que sentó las bases de posteriores teorías relativas al muestreo y la reconstrucción.

Durante el siglo XIX, gracias a los trabajos de matemáticos como Lagrange y Hermite, se desarrollaron multitud de técnicas interpolatorias basadas en polinomios, pero las características oscilatorias de estos últimos dificultaron su aplicación a la aproximación funcional. En 1946, Schoenberg describió una función base que permitía generar splines para interpolación. Con el tiempo, los splines se han convertido en piedra angular de la teoría de la aproximación, interpolación multidimensional, análisis numérico, estadística, diseño asistido por ordenador, gráficos por ordenador, diseño de fuentes tipográficas, etc.

El teorema de muestreo de Shannon, publicado en 1948 tomando como base trabajos previos de Nyquist y Whittaker y reformulando conceptos desarrollados con anterioridad por otros matemáticos (entre los que se encuentran Borel y Kotelnikov), estableció los principios de la reconstrucción de una función limitada en banda a partir de sus muestras, lo que constituye un caso particular de interpolación en el que se utiliza la función $\mathrm{sinc}(t)$ como función interpoladora. Su trabajo ha desempeñado y desempeña un papel crucial en el procesado de señal, las comunicaciones y el análisis numérico.

Trabajos más recientes introducen la interpolación polinómica en el ámbito del procesado de señal. Es a finales de los años 70 cuando irrumpe con fuerza el uso de splines en la interpolación de imágenes digitales, siendo la interpolación spline cúbica la más popular. Sin embargo, no ha sido hasta finales de la década de los 90 cuando estas técnicas han sido aplicadas al resto de campos del procesado de señal, gracias a los trabajos de M. Unser y otros autores en los que se resolvía el problema de la interpolación spline mediante técnicas de filtrado digital.

Actualmente, dentro del campo del tratamiento de señal, existe una gran actividad investigadora relativa a la interpolación, aplicando y extendiendo los principios y conocimientos que los matemáticos conocían desde varias décadas atrás. En este sentido, existen multitud de obras que recapitulan los avances y extensiones que ha experimentado el Teorema de Muestreo desde su formulación por Shannon [Jerri77, Unser00, Vaidyanathan01a]. Entre dichas extensiones, merecen comentarse el muestreo multidimensional, el muestreo a partir del valor de una función y de sus derivadas, el muestreo de procesos estocásticos, el muestreo de funciones paso banda, el muestreo de señales multibanda, el muestreo de señales no limitadas en banda y el muestreo no uniforme.

1.2. Planteamiento y motivación de la Tesis

A pesar de la evolución que la teoría de muestreo e interpolación ha experimentado en los últimos años, el Teorema de Muestreo, en su forma original, constituye el estado del arte en el ámbito actual del procesado de señal y las comunicaciones. En efecto, proporciona una interpretación frecuencial simple y elegante al problema de la interpolación, pero presenta ciertos inconvenientes, entre los que se encuentra la longitud infinita de la función de reconstrucción $\mathrm{sinc}(t)$. Así mismo, el Teorema de Muestreo, y sus posteriores extensiones, imponen una serie de restricciones de difícil cumplimiento en la práctica. En primer lugar, o bien las señales a muestrear no están limitadas en banda o bien los filtros antisolapamiento distan mucho de ser ideales, siendo habituales sistemas de adquisición con filtros tan simples como los de primer orden, formados por una resistencia y un condensador. En segundo lugar, existen situaciones en las que el criterio de limitación en el ancho de banda de la señal no parece el más adecuado. Por ejemplo, la reconstrucción de una señal a partir de sus muestras siguiendo el Teorema de Muestreo puede llevar a una solución muy sobreamortiguada o con una potencia excesivamente alta [Kashyap96, Waele00]. Por ello, un criterio como el de máxima suavidad puede resultar más apropiado en determinadas ocasiones, como cuando las señales describen el movimiento de una partícula o en la interpolación de imágenes. En tercer lugar, en ciertas aplicaciones no es posible obtener o no interesa trabajar con muestras equiespaciadas de la señal (muestreo no uniforme). Para estas señales resultan de poca utilidad el concepto de ancho de banda y el Teorema de Muestreo tal y como Shannon lo formuló. Aunque existen en la literatura numerosos trabajos sobre muestreo no uniforme, éstos se caracterizan, en general, por su complejidad y por un pobre comportamiento frente al ruido [Yen56, Jerri77, Shoumekh88, Ouderaa88].

Entre las técnicas de interpolación más habituales se encuentran la interpolación lineal, la interpolación mediante polinomios de Lagrange, la interpolación *spline* y la interpolación mediante la función $\mathrm{sinc}(t)$ derivada del Teorema de Muestreo. Como muestra la Figura 1.1, estos métodos abarcan todo el rango de complejidad, desde la interpolación lineal (la más simple, que equivale a la interpolación spline de orden uno y al polinomio de Lagrange de grado uno), hasta la interpolación $\mathrm{sinc}(t)$ (equivalente a la interpolación spline de orden infinito), pasando por los polinomios de Lagrange y splines de grados crecientes.

Es importante recalcar que las prestaciones de unas y otras técnicas interpolatorias dependen básicamente de la señal a interpolar: por ejemplo, la interpolación $\mathrm{sinc}(t)$ proporciona la solución exacta cuando las muestras a interpolar pertenecen a una señal estrictamente limitada en banda; mientras que la interpolación spline cúbica reconstruye perfectamente señales pertenecientes al espacio de señal spline cúbico. Por lo tanto, hablar de prestaciones de una u otra técnica es, en todo caso, una cuestión relativa. Obviamente, siguiendo los postulados del Teorema de Muestreo, la función $\mathrm{sinc}(t)$ proporciona la so-

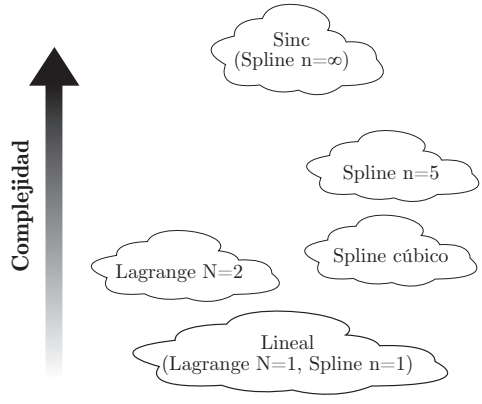


Figura 1.1: Complejidad de distintas técnicas interpolatorias.

lución exacta a la interpolación de señales paso bajo, pero su longitud infinita y su caída relativamente lenta impiden en la práctica alcanzar el resultado óptimo. Curiosamente, y a pesar de ello, en el ámbito del procesado de señal y las comunicaciones es la metodología utilizada en la inmensa mayoría de las aplicaciones (recurriendo a versiones truncadas de la función $\mathrm{sinc}(t)$ u otras aproximaciones). El resto de técnicas son empleadas de forma mucho menos habitual, salvo la interpolación spline cúbica en el procesado de imagen y la interpolación lineal cuando la sencillez es un requisito que predomina sobre la precisión.

La presente Tesis pretende analizar las prestaciones de las técnicas polinómicas de interpolación (en especial de los splines y de otras técnicas novedosas) cuando se consideran parámetros de diseño reales, pues estos últimos determinarán cuál de las técnicas comparadas proporciona los mejores resultados. En concreto, los parámetros de diseño contemplados son:

- Sobremuestreo. Habitualmente, las limitaciones prácticas propias de la implementación de los filtros antisolapamiento obligan a trabajar con factores de sobremuestreo altos, lo que implica que el contenido espectral de las señales discretas sea, fundamentalmente, paso bajo.

- Precisión. Todos los sistemas de procesado digital de señal poseen precisión finita, siendo particularmente limitada en el caso de las etapas de conversión analógico-digital y digital-analógico.

- Coste computacional. Al igual que ocurre con la precisión, los sistemas digitales poseen una capacidad de cómputo limitada, hecho este que imposibilita, en determinadas ocasiones, la aplicación exacta de las técnicas

de interpolación, recurriéndose a soluciones de compromiso basadas habitualmente en el truncamiento de la solución ideal.

En problemas de interpolación y/o reconstrucción de señales, cuando se tienen en cuenta consideraciones prácticas (sobremuestreo, filtros antisolapamiento en sistemas de adquisición, etc.) o existe una limitada capacidad de cómputo, las técnicas de interpolación polinómica ofrecen soluciones en muchos casos mejores que la técnica convencional de interpolación basada en el Teorema de Muestreo. Esta es la idea que motiva inicialmente esta Tesis y que se discute en profundidad en el Capítulo 3 mediante una serie de ejemplos: la interpolación en sistemas de adquisición reales y la implementación de un sistema de muestreo simultáneo basado en filtros polinómicos (Lagrange) de retardo fraccionario.[2]

Una vez planteado el interés de las técnicas polinómicas (no suficientemente destacado, en general, en el campo del tratamiento de señal, salvo los trabajos de Unser y otros autores relativos a la aplicación de los splines al procesado de imagen) queda justificada la búsqueda de técnicas de interpolación más flexibles y que permitan un compromiso entre prestaciones y coste computacional. Así, en el Capítulo 4 se propone una nueva familia de splines, los α-splines, que posibilitan una transición gradual y parametrizada entre splines de orden entero. Entre las ventajas de esta técnica se encuentran:

- Adaptación de la solución a la señal de entrada, posibilitando, por ejemplo, el control de la varianza, del sobredisparo o de la suavidad de la señal interpolada para ajustarse a los requerimientos de la aplicación.

- Las funciones base α-spline constituyen una base de Riesz y forman partición de la unidad, lo que asegura que la representación α-spline de una señal es estable y que el espacio definido por dichas bases es un subespacio cerrado de L_2.

- Al igual que las técnicas convencionales basadas en splines, la técnica propuesta también permite una implementación mediante filtrado digital, resultando muy eficiente desde el punto de vista computacional.

- Es posible encontrar valores del parámetro de control que permiten implementaciones simples y eficaces en sistemas hardware de alta velocidad.

- La extensión de dicha técnica al muestreo no uniforme posibilita su empleo en multitud de aplicaciones en las que la interpolación tradicional proporciona soluciones de gran complejidad, muy sensibles al ruido y poco flexibles.

[2] Estrictamente, esta técnica no es una contribución propia de la presente Tesis, aunque sí lo es su aplicación al problema concreto y el análisis de sus prestaciones.

Motivado por las ventajas de la nueva familia de splines, se extienden las técnicas de interpolación α-spline al problema de modelado. En concreto, se propone una extensión del modelo lineal a tramos canónico mediante el empleo de la base α-spline como función de activación, proporcionando un modelo que preserva las propiedades de simplicidad y capacidad de aproximación originales y que aporta las características de derivabilidad y suavidad propias de la interpolación α-spline. Además de la presentación del modelo α-spline y de su metodología de entrenamiento, en esta Tesis se propone su empleo en dos aplicaciones bien diferenciadas que se comentan a continuación.

En primer lugar se lleva a cabo el modelado α-spline de amplificadores de potencia en base a medidas temporales de entrada-salida obtenidas mediante la técnica de submuestreo. Se trata de una técnica novedosa de caracterización de amplificadores que, basada en la conversión directa de las señales de radiofrecuencia mediante un solapamiento espectral controlado, posibilita extraer las características del amplificador cuando se excita con señales de entrada de gran ancho de banda y envolvente no constante (es decir, con señales similares a las presentes en los modernos sistemas de comunicaciones digitales). Es más, dicha técnica requiere un montaje sensiblemente más simple que los esquemas de caracterización convencionales y evita la distorsión introducida por las etapas de conversión de frecuencia propias de estos últimos.

Una segunda aplicación del modelo α-spline es el modelado de la corriente de drenador de transistores HEMT sometidos a iluminación óptica. La particularidad de esta aplicación es el interés que posee el valor de la derivada de la corriente de drenador con respecto a las tensiones de polarización, pues resulta determinante en el comportamiento de los dispositivos construidos empleando dicho transistor. Por esta razón, es importante obtener modelos sencillos que, además de llevar a cabo un ajuste preciso de la función a modelar, proporcionen una función derivable y suave que reproduzca fielmente el comportamiento de las derivadas. En este sentido, el modelo α-spline propuesto constituye una opción idónea.

Finalmente, en esta Tesis se aborda el problema de las comunicaciones digitales aprovechando el paralelismo existente entre las mismas y la interpolación funcional; paralelismo que, a pesar de resultar evidente, no ha sido, a nuestro juicio, explotado suficientemente. Es posible considerar el *criterio de Nyquist* de comunicaciones sin interferencia intersimbólica como un caso particular de interpolación funcional en el que la expansión en serie de funciones base se restringe a muestras equiespaciadas y funciones limitadas en banda. En aquellas aplicaciones en las que la restricción en banda sea prescindible, la consideración de las comunicaciones desde el punto de vista de la interpolación abre las puertas a nuevos diseños, tal y como se presenta en esta Tesis.

1.3. Organización de la Tesis

La presente Tesis se organiza de la siguiente manera:

- El Capítulo 2 lleva a cabo un repaso de las distintas técnicas de interpolación, aproximación y modelado existentes en la literatura, prestando especial atención a las polinómicas y, más concretamente, a las basadas en splines. Así mismo se presenta la terminología y notación a utilizar en el resto de la Tesis.

- En el Capítulo 3 se presenta una serie de ejemplos que pretenden servir de motivación al resto de la Tesis y que demuestran que, bajo ciertas condiciones prácticas, las técnicas de interpolación polinómicas proporcionan resultados similares (o incluso mejores) a los proporcionados por la interpolación limitada en banda mediante la función núcleo $\mathrm{sinc}(t)$ (o versiones truncadas de la misma); es más, cuando el coste computacional es un factor a considerar, dichas técnicas polinómicas ofrecen ventajas sustanciales. En concreto, en el primer ejemplo considerado se compara el método tradicional de interpolación de señales limitadas en banda frente a la interpolación spline cúbica, considerando distintas resoluciones (número de bits) y distintos contenidos espectrales de las señales a interpolar. Un segundo ejemplo dentro del Capítulo 3 presenta una solución al problema del retardo fraccionario basada en los polinomios de Lagrange. Puesto que, en los sistemas reales, las señales adquiridas se encuentran sobremuestreadas con el fin de facilitar el diseño y construcción de los filtros antisolapamiento, los interpoladores de Lagrange reultan ser una solución atractiva. Además de comparar sus prestaciones con las aportadas por otras técnicas habituales de procesado de señal (como la interpolación sinc truncada y las técnicas multitasa implementadas mediante estructuras polifásicas), se realiza un estudio del error de la interpolación Lagrangiana obteniéndose expresiones que permiten determinar la longitud del filtro necesaria para obtener o mantener una resolución determinada para un factor de sobremuestreo dado.

- El Capítulo 4 presenta una nueva familia de splines, que denominamos α-splines, particularizada, en primer lugar, para muestreo equiespaciado. Se trata de una familia parametrizada de splines que se sitúa entre el spline de orden p y el de orden $p+q$, de forma que el parámetro permite controlar el grado de influencia de una u otra solución. Una vez presentadas las expresiones de la función base y sus derivadas, y tras demostrar que constituyen partición de la unidad y forman una base de Riesz, el capítulo se centra en el caso particular $p = 1$ y $q = 2$, que representa la transición parametrizada entre el spline lineal y el cúbico. Se trata del

caso de mayor interés práctico, por lo que se presenta el método de interpolación correspondiente y las expresiones cerradas de la suavidad, la varianza y el área residual de dicha interpolación α-spline. Finalmente, se generalizan los resultados para muestreo no equiespaciado, se presenta una serie de simulaciones que demuestran la veracidad de los resultados teóricos previamente obtenidos y se muestran diversas aplicaciones de utilidad práctica de la interpolación α-spline.

- A partir de la base α-spline, en el Capítulo 5 se propone una extensión del modelo canónico lineal a tramos (CPWL) manteniendo su simplicidad y precisión, y aportando las características de suavidad y derivabilidad. Para ello se tiene en cuenta el hecho de que la base α-spline con $\alpha = 0$, al igual que el modelo CPWL, proporciona funciones lineales a tramos; y que el parámetro α posibilita controlar la suavidad de una solución que, en todos lo casos, posee continuidad hasta la segunda derivada. Por lo tanto, se propone sustituir la función de activación valor absoluto del modelo CPWL por una función basada en el núcleo α-spline, generándose así modelos derivables, con transiciones suaves entre las distintas regiones del espacio de entrada y con una suavidad controlable gracias al parámetro α. En este ámbito, se presentan dos aplicaciones: el modelado α-spline de amplificadores de potencia basado en medidas temporales de entrada-salida obtenidas mediante técnicas de submuestreo, y el modelado α-spline de gran señal de transistores HEMT sometidos a iluminación óptica.

- En el Capítulo 6 se aporta una visión novedosa de las comunicaciones digitales, que, haciendo uso del paralelismo existente entre las mismas y el proceso interpolatorio, permite afrontar desde otra vertiente el diseño de sistemas de comunicaciones adaptados y libres de interferencia entre símbolos. De esta manera, se describen las técnicas para el diseño de sistemas de comunicaciones mediante el uso de splines para, posteriormente, generalizar a pulsos conformadores de forma arbitraria. Se demostrará cómo los sistemas así diseñados resultan, bajo determinadas circunstancias, más eficientes y sencillos de implementar que los habitualmente empleados. Adicionalmente, se propone la aplicación de la teoría de las parejas de filtros biortogonales para diseñar sistemas de comunicaciones adaptados y sin ISI, implementados exclusivamente con filtros de respuesta al impulso finita (FIR: *Finite Impulse Response*).

- Por último, el Capítulo 7 resume las principales aportaciones del presente trabajo de Tesis, así como las conclusiones que del mismo se derivan y las futuras líneas de trabajo que pueden completar y ampliar lo aquí presentado.

Capítulo 2

Conceptos teóricos

2.1. Introducción

En este capítulo se presentan las técnicas y los conceptos teóricos relativos a las materias tratadas en esta Tesis. Más concretamente, se profundizará en los conceptos y la teoría asociada a la aproximación funcional, la interpolación y el modelado, con especial atención a las técnicas polinómicas. Así mismo, se presentará la teoría de muestreo en espacios de Hilbert que engloba la formulación tradicional de interpolación-aproximación-muestreo.

2.2. Espacio de señal

En esta sección se presentan los conceptos relativos al espacio de señal [Davis75, Higgins96, Moon00], definiendo los espacios vectoriales, los espacios métricos, los espacios vectoriales normados y los espacios normados con producto interno, lo que permitirá introducir el concepto de proyección. Resulta necesario, además, presentar algunos conceptos básicos sobre espacios vectoriales.

Definición 2.1 (Espacio Lineal) *Un espacio lineal o espacio vectorial lineal, X, es un conjunto de elementos (vectores), $\mathbf{x}$, $\mathbf{y}$,..., para los que se definen las operaciones de suma y producto escalar: dados dos elementos, $\mathbf{x}, \mathbf{y} \in X$, existe el elemento suma como $\mathbf{x} + \mathbf{y} \in X$; y dados un elemento $\mathbf{x} \in X$ y un escalar α, existe un único producto escalar $\alpha\mathbf{x} \in X$. Ambas operaciones cumplen las siguientes propiedades:*

I) $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$;

II) $\mathbf{x} + (\mathbf{y} + \mathbf{z}) = (\mathbf{x} + \mathbf{y}) + \mathbf{z}$;

III) *existe un único elemento* $\mathbf{0} \in X$ *tal que* $\mathbf{x} + \mathbf{0} = \mathbf{x}$ *para todo* $\mathbf{x} \in X$*;*

IV) *para cada* $\mathbf{x} \in X$ *existe un único inverso,* $-\mathbf{x}$*, tal que* $\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$*;*

V) $\alpha(\beta\mathbf{x}) = (\alpha\beta)\mathbf{x}$ *para todo* $\alpha, \beta \in \mathbb{R}$ *y* $\mathbf{x} \in X$*;*

VI) $\alpha(\mathbf{x} + \mathbf{y}) = \alpha\mathbf{x} + \alpha\mathbf{y}$*;*

VII) $(\alpha + \beta)\mathbf{x} = \alpha\mathbf{x} + \beta\mathbf{x}$*;*

VIII) $1(\mathbf{x}) = \mathbf{x}$.

Un ejemplo, denotado por L_2, es el espacio de funciones o señales cuadráticamente integrables en el intervalo $[a, b]$, es decir, el conjunto de señales que cumplen $\int_a^b f^2(t)dt < \infty$.

Definición 2.2 (Independencia Lineal) *Dado un subconjunto* T *de un espacio lineal* X*, se dice que* T *es linealmente independiente si para cada subconjunto finito no vacío de* T *(por ejemplo* $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m\}$*) el único conjunto de escalares que satisfacen la ecuación*

$$c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \cdots + c_m\mathbf{x}_m = 0$$

es la solución trivial $c_1 = c_2 = \cdots = c_m = 0$.

En base al concepto de independencia lineal, es posible determinar la dimensión de un espacio.

Definición 2.3 (Dimensión) *Un espacio vectorial lineal* X *posee dimensión* n *si es posible encontrar como máximo* n *vectores linealmente independientes.*

Cuando se trabaja con señales (o funciones) pertenecientes a un espacio, resulta de interés disponer de una medida de "distancia" entre ellas que sea práctica desde el punto de vista matemático y significativa desde el punto de vista físico. Surge así el concepto de métrica y la definición de espacios métricos.

Definición 2.4 (Espacio Métrico) *Un espacio métrico es un espacio lineal con una métrica asociada* d*, y se denota* $(X, d(\cdot,\cdot))$*,. La métrica es una función utilizada para evaluar la distancia entre dos elementos pertenecientes al espacio* X *y que debe cumplir las siguientes propiedades para todo* $\mathbf{x}$*,* $\mathbf{y} \in X$*:*

I) $d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x})$*;*

II) $d(\mathbf{x}, \mathbf{y}) \geq 0$*;*

III) $d(\mathbf{x}, \mathbf{y}) = 0$ *si y sólo si* $\mathbf{x} = \mathbf{y}$*;*

IV) *para todo* $\mathbf{x}$*,* $\mathbf{y}$*,* $\mathbf{z} \in X$*,* $d(\mathbf{x}, \mathbf{z}) \leq d(\mathbf{x}, \mathbf{y}) + d(\mathbf{y}, \mathbf{z})$.

En el ámbito del procesado de señal, es habitual representar las secuencias discretas como vectores (de dimensión finita). Es más, es posible extender dicho planteamiento y considerar las señales continuas como vectores de dimensión infinita. En relación con los espacios vectoriales, es común hablar sobre la longitud y la dirección de vectores. Así, el concepto matemático asociado a la longitud de un vector es la norma, lo que nos conduce a considerar los espacios lineales normados; mientras que el ángulo entre dos vectores se asocia con el producto interno, que conducirá a los espacios dotados de producto interno.

Definición 2.5 (Espacio Lineal Normado) *Un espacio lineal, X, se dice que es un espacio lineal normado si para cada elemento, $\mathbf{x}$, del espacio está definido un número real, designado por $\|x\|$ y denominado norma de $\mathbf{x}$. Dicho espacio lineal normado se denota por $(X, \|\cdot\|)$, y su norma debe poseer las siguientes propiedades:*

I) $\|\mathbf{x}\| \geq 0$ *(positividad);*

II) $\|\mathbf{x}\| = 0$ *si y sólo si* $\mathbf{x} = \mathbf{0}$ *(concreción);*

III) $\|\alpha\mathbf{x}\| = |\alpha| \cdot \|\mathbf{x}\|$ *para todo escalar* α *(homogeneidad);*

IV) $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ *(desigualdad triangular).*

Es fácil observar la similitud entre norma y métrica; de hecho, una puede definirse en función de la otra.[1] Por ejemplo, si $\|\mathbf{x}\|$ es una norma, entonces $d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|$ es una métrica. De forma alternativa, si d es una métrica de un espacio lineal, entonces es posible definir una norma como $\|\mathbf{x}\| = d(\mathbf{x}, \mathbf{0})$, es decir, es la distancia del vector $\mathbf{x}$ al origen del espacio vectorial.

En espacios vectoriales n-dimensionales se emplean habitualmente las siguientes normas:

- Norma l_1: $\|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i|$.
- Norma l_p: $\|\mathbf{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}$.
- Norma l_∞: $\|\mathbf{x}\|_\infty = \operatorname*{máx}_{i=1,\ldots,n} |x_i|$.

El equivalente de las normas anteriores en el caso de funciones definidas en el intervalo $[a, b]$ son:

- Norma L_1: $\|x(t)\|_1 = \int_a^b |x(t)| dt$.
- Norma L_p: $\|x(t)\|_p = \left(\int_a^b |x(t)|^p dt\right)^{1/p}$.

[1] Siendo estrictos, la definición de norma precisa que estén definidas la multiplicación por un escalar y la suma, mientras que en la definición de métrica esto no es necesario.

- Norma L_∞: $\|x(t)\|_\infty = \sup_{t\in[a,b]} |x(t)|$.

Definición 2.6 (Espacio dotado de Producto Interno) *Un espacio vectorial lineal, X, se dice que es un espacio dotado de producto interno si para cada dos elementos, $\mathbf{x}$ e $\mathbf{y}$, se define un número real, designado por $\langle \mathbf{x}, \mathbf{y}\rangle$ y denominado producto interno, con las siguientes propiedades:*

I) $\langle \mathbf{x}+\mathbf{z}, \mathbf{y}\rangle = \langle \mathbf{x}, \mathbf{y}\rangle + \langle \mathbf{z}, \mathbf{y}\rangle$ *(linealidad);*

II) $\langle \mathbf{x}, \mathbf{y}\rangle = \langle \mathbf{y}, \mathbf{x}\rangle$ *(simetría);*

III) $\langle \alpha\mathbf{x}, \mathbf{y}\rangle = \alpha \langle \mathbf{x}, \mathbf{y}\rangle$ *para todo escalar α (homogeneidad);*

IV) $\langle \mathbf{x}, \mathbf{x}\rangle > 0$ *si $\mathbf{x} \neq \mathbf{0}$ y $\langle \mathbf{x}, \mathbf{x}\rangle = 0$ si y sólo si $\mathbf{x} = \mathbf{0}$ (positividad).*

Algunos ejemplos de productos internos son:

- Producto interno euclídeo en espacio n-dimensional:

$$\langle \mathbf{x}, \mathbf{x}\rangle = \sum_{i=1}^{n} x_i y_i$$

- Producto interno de funciones definidas en $[a, b]$:

$$\langle x(t), y(t)\rangle = \int_a^b x(t)y(t)dt$$

Una cuestión de interés es el hecho de que el operador convolución es un producto interno, pues

$$y(t_0) = (x(t) * h(t))\,|_{t=t_0} = \int_{-\infty}^{\infty} x(\tau)h(t_0-\tau)d\tau = \langle x(t), h(t_0-t)\rangle\,,$$

es decir, la operación de filtrado (y la obtención del valor de la salida en un instante de tiempo fijado) es equivalente a evaluar un producto interno.

Definición 2.7 (Espacio Completo) *Un espacio, X, se dice que es completo si cada serie de Cauchy* [2] *converge en X.*

En concreto, si X es un espacio lineal normado completo y para cualquier $\epsilon > 0$ es posible encontrar un $N > 0$ tal que

$$\|\mathbf{x}_m - \mathbf{x}_n\| \leq \epsilon, \text{ con } m, n \geq N,$$

entonces existe un elemento $\mathbf{x} \in X$ tal que

$$\lim_{n\to\infty} \|\mathbf{x} - \mathbf{x}_n\| = 0.$$

[2] Una secuencia $\{\mathbf{x}_n\}$ de un espacio métrico X es una serie de Cauchy si cumple que para cualquier $\epsilon > 0$ existe un $N > 0$ tal que $d(\mathbf{x}_n, \mathbf{x}_m) < \epsilon$ para cualquier $m, n > N$.

Definición 2.8 (Espacio de Banach) *Un espacio lineal normado completo se denomina espacio de Banach.*

Definición 2.9 (Espacio de Hilbert) *Un espacio lineal normado completo con un producto interno se denomina espacio de Hilbert.*

Algunos ejemplos de espacio de Hilbert son:

- El conjunto de todas las secuencias de energía finita $\sum_i |x_i|^2 < \infty$, con las operaciones habituales de suma y producto escalar y con el producto interno definido por $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_i x_i y_i$, constituye un espacio de Hilbert, y se denota como l_2.

- El conjunto de todas las funciones definidas en el intervalo $[a, b]$, cuadráticamente integrables, con el producto interno definido por $\langle f, g \rangle = \int_a^b f(x)g(x)dx$ y la norma $\|f\|^2 = \langle f, f \rangle = \int_a^b |f(x)|^2 dx$, constituye un espacio de Hilbert, y se denota como $L_2[a, b]$.

2.3. Aproximación, interpolación y modelado

El tratamiento del muestreo desde el punto de vista de interpolación y/o aproximación hace necesario introducir una serie de definiciones que aclaren la terminología a utilizar a lo largo de esta Tesis.

El problema de la aproximación funcional consiste en encontrar una función $P(x)$ de la forma

$$P(x) = \sum_{k=0}^{n} a_k \Phi_k(x)$$

que se "parezca" lo más posible a una función $f(x)$ dada explícitamente, cuyo valor, y posiblemente el de sus derivadas, es conocido en todo punto del intervalo $[a, b]$. El parecido entre la función objetivo $f(x)$ y la solución $P(x)$ se evalúa en función de cierta métrica, que determinará la técnica de aproximación a utilizar. En la Figura 2.1.a se puede observar un polinomio de tercer orden que aproxima a la función $\tan(x)$.

Por otro lado, la interpolación consiste en encontrar, partiendo de un conjunto de N pares de valores (x_i, y_i), una función $P(x)$ de la forma

$$P(x) = \sum_{k=0}^{n} a_k \Phi_k(x)$$

que cumpla

$$P(x_i) = y_i, \ i = 1, 2, \ldots, N. \tag{2.1}$$

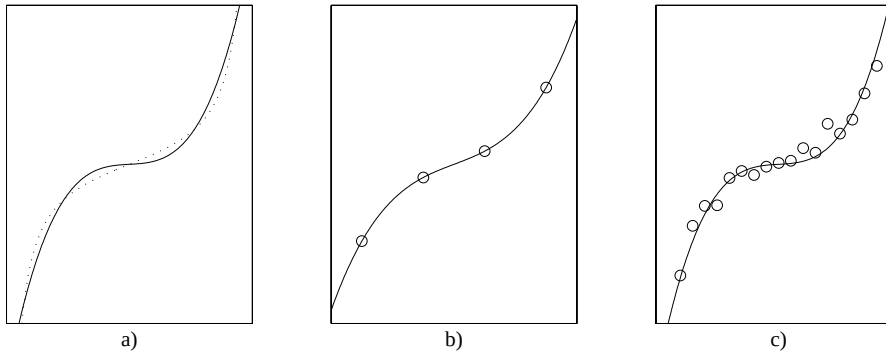


Figura 2.1: a) Aproximación, b) interpolación y c) modelado.

Este proceso se puede observar en Figura 2.1.b, donde un polinomio de tercer orden interpola cuatro puntos. En la interpolación, el comportamiento de la solución entre los puntos de interpolación depende de la función base usada (polinomios en el ejemplo).

Con el término modelado nos referiremos al proceso que en la literatura también se denomina aproximación discreta o aproximación de datos. Partiendo de un conjunto de N pares de valores (x_i, y_i), se busca una función $P(x)$ de la forma

$$P(x) = \sum_{k=0}^{n} a_k \Phi_k(x)$$

que aproxime a los datos, es decir,

$$P(x_i) = y_i + \epsilon_i, \ i = 1, 2, \ldots, N;$$

siendo el error de modelado, ϵ_i, mínimo (en función de una métrica determinada).

En la Figura 2.1.c se modela un conjunto de puntos ruidosos (cuyo número es superior a los grados de libertad de la solución) mediante un polinomio de tercer orden.

Desde un punto de vista conceptual, la única diferencia entre la aproximación y el modelado, tal y como se definen en esta Tesis, es que en el modelado se desconoce el valor de la función fuera de los puntos dato de que se dispone, los cuales, además, suelen poseer cierto nivel de ruido. En cuanto a la aproximación y la interpolación, es posible establecer un único marco teórico basado en espacios de Hilbert que englobe ambos conceptos, como se verá posteriormente.

2.3.1. Polinomios como interpoladores y aproximadores

A lo largo de este capítulo se trabajará con el espacio (lineal) de los polinomios de orden n con coeficientes reales definido como

$$\mathcal{P}_n = \left\{ p_n(x) = \sum_{i=0}^{n} c_i x^i; \quad c_0, \ldots, c_n, x \in \mathbb{R} \right\}.$$

Existen múltiples razones, ampliamente tratadas en la bibliografía (por ejemplo, [Schumaker81, Ralston78]), que justifican la intensiva utilización de los polinomios como funciones interpoladoras, aproximadoras o modeladoras.

Es bien conocido que una línea puede hacerse pasar por dos puntos, una parábola por tres puntos, un polinomio cúbico por cuatro, y así sucesivamente. Este hecho se expresa más formalmente en un antiguo teorema [Davis75]:

Teorema 2.1 (Interpolación Polinómica) *Si se consideran $n+1$ puntos distintos $x_0, \ldots, x_n$ y $n+1$ valores $y_0, \ldots, y_n$ existe un único polinomio de grado n, $p_n(x) \in \mathcal{P}_n$, para el cual*

$$p_n(x_i) = y_i, \; i = 0, 1, \ldots, n;$$

lo cual sugiere la utilización de funciones polinómicas como interpoladoras, pues siempre será posible encontrar el polinomio de grado n que pasa por $n+1$ puntos.

Por otro lado, el teorema de aproximación de Weierstrass establece que siempre es posible realizar una aproximación polinómica de una función continua tan exacta como se desee sin más que aumentar el grado del polinomio interpolador tanto como sea necesario. De forma más rigurosa:

Teorema 2.2 (Aproximación de Weierstrass) *Sea una función continua en el intervalo $[a, b]$, es decir, $f(x) \in C[a, b]$. Dado un $\epsilon > 0$ es posible encontrar un polinomio $p_n(x)$ (de grado suficientemente alto) para el cual*

$$|f(x) - p_n(x)| \leq \epsilon, \qquad a \leq x \leq b.$$

Sin embargo, este teorema debe ser considerado con cautela pues, por un lado, el grado del polinomio es generalmente muy alto (incluso para obtener un nivel de precisión relativamente bajo) y, por otro lado, no especifica el método para encontrar el polinomio óptimo.

La extraordinaria difusión de los polinomios también se debe, en gran medida, a la posibilidad que brindan para evaluar el valor en un punto de una función analítica arbitraria en base a su desarrollo en serie de Taylor, el cual, como es bien conocido, se basa en funciones polinómicas.

Teorema 2.3 (Taylor) *Dada una función continua y con derivadas continuas hasta orden $n+1$ en el intervalo $[a,b]$, es decir, $f(x) \in C^{n+1}[a,b]$, y un punto $x_0 \in [a,b]$ entonces para todo $a \leq x \leq b$,*

$$\begin{aligned} f(x) =& f(x_0) + f^{(1)}(x_0)(x-x_0) + \frac{f^{(2)}(x_0)}{2!}(x-x_0)^2 + \cdots \\ &+ \frac{f^{(n)}(x_0)}{n!}(x-x_0)^n + \frac{1}{n!}\int_{x_0}^{x} f^{(n+1)}(t)(x-t)^n dt. \end{aligned}$$

Otra característica interesante de los polinomios es que su derivación, su integración y el escalado o desplazamiento de la variable independiente proporcionan como resultado polinomios. Más concretamente, si $p(x)$ y $q(x)$ son polinomios, entonces también lo son $p(x+\alpha)$, $p(\alpha x)$, $dp(x)/dx$, $\int p(x)dx$ y $p(q(x))$.

Si añadimos la facilidad con la que las funciones polinómicas pueden ser evaluadas (simplemente mediante operaciones aritméticas que se realizan de forma directa en computadores digitales) y manipuladas (por ejemplo, derivación o integración), es razonable considerar dichas funciones como candidatas para ser utlizadas en interpolación, aproximación o modelado. A modo de ejemplo, en el ámbito del tratamiento de señal los polinomios se han empleado en aplicaciones de clasificación, apoyándose en el teorema de Cover [Haykin94], para realizar un mapeo no lineal del espacio de entrada a un nuevo espacio de mayor dimensión con el fin de poder llevar a cabo la clasificación de forma más simple. Mientras que los filtros lineales son de utilidad en gran número de aplicaciones y resultan simples tanto desde el punto de vista conceptual como de la implementación, existen ciertas situaciones prácticas que precisan del procesado no lineal de las señales que intervienen. En [Mathews91, Pantaleón95b] se incorporan polinomios a filtros adaptativos no lineales y al modelado no lineal. Así, se proponen sistemas no lineales basados en serie de Volterra [Schetzen80] truncada o en ecuaciones en diferencias no lineales recursivas. Es posible encontrar otras aplicaciones de los polinomios en [Mathews00].

En las próximas secciones se llevará a cabo un repaso de las distintas técnicas de aproximación, interpolación y modelado, con especial énfasis en aquellas que emplean polinomios.

2.4. Aproximación

Dada una función conocida, el problema de aproximación consiste en encontrar una función "más simple" que la original que permita determinar (posiblemente con cierto error) valores de la función dada. La simplicidad de la solución se puede referir a la facilidad para ser evaluada, para ser integrada o derivada o cualquier otra característica que pueda resultar interesante. La

aproximación tiene, fundamentalmente, un interés teórico, pues, como se ha comentado, es preciso conocer explícitamente la función a aproximar. Un ejemplo de aplicación es la integración o derivación numérica o la evaluación de funciones en máquinas calculadoras. Piénsese, por ejemplo, en el cálculo del valor de la función $\cos(x)$ en un punto x_0: si se encuentra una función polinómica que aproxime, con un error determinado, la función en el entorno de x_0, podremos obtener su valor realizando únicamente multiplicaciones y sumas.

A pesar de estas limitaciones, la aproximación tiene un interés adicional, pues si una familia de funciones (polinómicas, racionales, serie de Fourier, etc.) proporciona una aproximación adecuada, entonces, muy probablemente, será una familia de funciones que proporcione buenas prestaciones a la hora de interpolar o modelar datos.

De una manera formal, se entiende por aproximación, también denominada aproximación funcional en la literatura matemática, al proceso por el cual dada una función $f(x)$ de forma explícita, se busca una función $P(x)$, perteneciente a la clase de funciones $\{\Phi_k(x)\}$, que se *parezca* lo más posible a $f(x)$, es decir:

$$P(x) = \sum_{k=0}^{n} a_k \Phi_k(x) \approx f(x). \tag{2.2}$$

Nos referiremos únicamente a aproximaciones de tipo lineal, similares a la expresión (2.2), en las que la aproximación, $P(x)$, es una combinación lineal de las funciones aproximadoras, $\{\Phi_k(x)\}$. El análisis de aproximaciones basadas en combinaciones no lineales de las funciones base es de una complejidad muy superior a la del caso lineal, y su estudio se aparta del objetivo de esta Tesis. El parecido entre la función objetivo, $f(x)$, y la solución, $P(x)$, se evalúa en función de cierta métrica, que determinará el tipo de aproximación.

Además de una elección apropiada de las funciones base $\{\Phi_k(x)\}$, el punto crucial a la hora de resolver un problema de aproximación es el criterio a usar para obtener las constantes a_k en (2.2). En base a dicho criterio es posible realizar una clasificación del tipo de aproximación. Entre otras muchas, sin duda los dos tipos más importantes son la aproximación de mínimo error cuadrático y la aproximación uniforme.

2.4.1. Aproximación de mínimo error cuadrático

En la aproximación de mínimo errror cuadrático el objetivo es minimizar la integral del cuadrado de la diferencia entre $f(x)$ y su aproximación en un intervalo $[a, b]$, es decir, buscar el mínimo de la expresión del error

$$\varepsilon = \int_a^b \left(f(x) - P(x)\right)^2 dx. \tag{2.3}$$

Este proceso de aproximación funcional equivale a la búsqueda de la función perteneciente al espacio de señal definido por $\{\Phi_k(x)\}$ más próxima (en norma L_2) a la función a aproximar $f(x)$, es decir, minimizar (2.3) equivale a minimizar la norma L_2. Definido así el error, para obtener los $n+1$ coeficientes a_k de (2.2) se deberá resolver el sistema compuesto por las $n+1$ ecuaciones normales

$$\frac{\partial \varepsilon}{\partial a_k} = 0, \quad k = 0, 1, \ldots, n. \tag{2.4}$$

En ocasiones es interesante minimizar el error de forma ponderada, utilizando para ello una función de peso, cuyo cometido será asignar distintos grados de importancia al error de aproximación en las distintas zonas del intervalo.

Definición 2.10 (Función de peso) *Una función integrable, $w(x)$, es una función de peso en un intervalo I si $w(x) \geq 0$ para todo x en I y además $w(x)$ no es idénticamente nula en ningún subintervalo de I.*

Así, incluyendo en (2.3) la función de ponderación $w(x)$ se obtiene una expresión más general del error cuadrático:

$$\varepsilon = \int_a^b w(x) \left(f(x) - P(x)\right)^2 dx.$$

En este punto es conveniente introducir el concepto de ortogonalidad de funciones, propiedad interesante ya que permite calcular la aproximación de menor error cuadrático aplicando

$$a_k = \frac{1}{\alpha_k} \int_a^b w(x) f(x) \Phi_k(x) dx,$$

sin necesidad de resolver el sistema de ecuaciones normales (2.4).

Definición 2.11 (Ortogonalidad) *Dos funciones $f(x)$ y $g(x)$ son ortogonales en el intervalo $[a, b]$ con respecto a la función de peso $w(x)$ si su producto escalar es nulo, es decir, si*

$$\langle f, g \rangle = \int_a^b w(x) f(x) g(x) dx = 0.$$

Definición 2.12 (Conjunto Ortogonal de Funciones) *Se dice que el conjunto de funciones $\{\Phi_0, \ldots, \Phi_n\}$ es un conjunto ortogonal de funciones en el intervalo $[a, b]$ con respecto a la función de peso $w(x)$ si*

$$\int_a^b w(x) \Phi_j(x) \Phi_k(x) dx = \begin{cases} 0, & j \neq k; \\ \alpha_k > 0, & j = k. \end{cases}$$

Si, además, $\alpha_k = 1$ para cada $k = 0, 1, \ldots, n$, entonces se dice que el conjunto de funciones es ortonormal.

Algunos ejemplos de conjuntos ortogonales de funciones son:

Polinomios de Legendre, denotados por $\{P_n(x)\}$, son ortogonales en $[-1,1]$ con respecto a la función de peso $w(x)=1$, siendo su valor

$$\begin{aligned} P_0(x) &= 1, \\ P_1(x) &= x, \\ P_{k+1}(x) &= \frac{(2k+1)xP_k(x) - kP_{k-1}(x)}{k+1}. \end{aligned}$$

Polinomios de Tchebycheff, son ortonormales en $[-1,1]$ con respecto a la función $w(x)=\frac{1}{\sqrt{1-x^2}}$. Se denotan por $\{T_n(x)\}$ y se obtienen mediante la recursión

$$\begin{aligned} T_0(x) &= 1, \\ T_1(x) &= x, \\ T_{k+1}(x) &= 2xT_k(x) - T_{k-1}(x), \quad k=2,3,\ldots; \end{aligned}$$

o bien, de forma explícita, según la expresión

$$T_k(x) = \cos\left(k \arccos x\right), \quad k=1,2,\ldots.$$

Además de utilizarse como funciones base aproximadoras, estos polinomios se emplean para determinar la localización óptima de los puntos de interpolación (coincidiendo con las raíces de los polinomios de Tchebycheff) con el fin de obtener una solución cercana a la aproximación minimax.

Polinomios de Hermite, $\{H_n(x)\}$, son ortonormales en $(-\infty,\infty)$ con respecto a $w(x)=e^{-x^2}$,

$$\begin{aligned} H_0(x) &= 1, \\ H_1(x) &= 2x, \\ H_{k+1}(x) &= 2xH_k(x) - 2kH_{k-1}(x), \quad k=1,2\ldots. \end{aligned}$$

Polinomios Trigonométricos, son ortonormales en $[-\pi,\pi]$ con respecto a la función $w(x)=1$. Se denotan por $\{\Phi_n(x)\}$ y siguen la expresión

$$\begin{aligned} \Phi_0(x) &= 1, \\ \Phi_k(x) &= \frac{1}{\sqrt{\pi}}\cos kx, \quad k=1,2,\ldots,n, \\ \Phi_{n+k}(x) &= \frac{1}{\sqrt{\pi}}\sin kx, \quad k=1,2,\ldots,n. \end{aligned}$$

Dada una función $f(x)$ contínua en $[-\pi, \pi]$, la aproximación de menor error cuadrático usando los polinomios trigonométricos es $S_n(x) = \sum_{k=0}^{2n-1} a_k \Phi_k(x)$. En el límite, cuando $n \to \infty$, $S_n(x)$ es el desarrollo en serie de Fourier de $f(x)$, de trascendental importancia en el ámbito del procesado de señal.

Funciones Sinc, el conjunto de funciones $\operatorname{sinc}(x-k)$ es ortonormal en $(-\infty, \infty)$ con respecto a $w(x) = 1$, siendo

$$\operatorname{sinc}(x-k) = \frac{\sin(\pi(x-k))}{\pi(x-k)} \quad k \in \mathbb{Z}.$$

Éste es el conjunto de funciones base definidas en el Teorema de Muestreo.

2.4.2. Aproximación uniforme o minimax

En la aproximación uniforme se minimiza el valor máximo de la distancia entre la función a aproximar, $f(x)$, y la solución en un intervalo $[a, b]$, es decir, se minimiza la expresión del error

$$\varepsilon = \max_{x \in [a,b]} |f(x) - P(x)|,$$

lo que equivale a minimizar la norma L_∞. Habitualmente, este tipo de aproximación también es conocida como aproximación de mínimo error máximo o aproximación minimax.

Desde el punto de vista del procesado de señal, la norma L_2, que se corresponde con la aproximación de menor error cuadrático, es la más utilizada a la hora de comparar el parecido entre señales (principalmente porque de su definición se derivan expresiones analíticamente tratables). No obstante, la norma L_∞, que define la aproximación uniforme, también tiene gran interés, pues proporciona una medida del valor máximo de la diferencia entre dos señales.

Dada una función continua, la existencia de una aproximación uniforme tan buena como se desee está asegurada por el teorema de Weierstrass (Teorema 2.2), aunque su obtención no resulta, en general, sencilla. De hecho, en la práctica se recurre a soluciones subóptimas, entre las que la más utilizada es, sin duda, la aproximación basada en la interpolación en los puntos de Tchebycheff, como establece el siguiente teorema ([Cheney82] y [Davis75]):

Teorema 2.4 (Fejér) *Sean $x_1, x_2, \ldots, x_n$ los ceros del polinomio de Tchebycheff de orden n, $T_n(x)$, es decir, $\left\{x_j = \cos\left(\frac{2j-1}{2n}\pi\right)\right\}$. Sea $f(x)$ continua en [-1,1] y $P_n(x)$ un polinomio de grado n tal que $P_n(x_k) = f(x_k)$. Entonces $lim_{n\to\infty} P_n(x) = f(x)$ uniformemente en [-1,1].*

El concepto fundamental de este teorema es que la aproximación en los puntos de Tchebycheff proporciona un polinomio muy próximo al polinomio minimax, el cual, de entre todos los polinomios de orden n, es el que presenta la menor desviación máxima respecto a la función deseada. El procedimiento práctico [Press89] consiste en obtener los N coeficientes interpolatorios de Tchebycheff, c_j con $j = 1, \ldots, N$, tales que

$$c_j = \frac{2}{N} \sum_{k=1}^{N} f(x_k) T_{j-1}(x_k),$$

de tal manera que la fórmula de aproximación

$$f(x) \approx \left[\sum_{k=0}^{N} c_k T_{k-1}(x) \right] - \frac{1}{2} c_1 \tag{2.5}$$

es exacta en puntos de x correspondientes a los N ceros de $T_N(x)$. Esta solución obtenida no tiene que ser, necesariamente, la mejor aproximación de orden N de la función $f(x)$ en el sentido minimax. Pero si suponemos N suficientemente grande, entonces, en virtud de Teorema 2.4, la expresión en (2.5) es virtualmente una aproximación perfecta de $f(x)$.

El hecho de que los coeficientes c_k decrezcan rápidamente, lo cual es el caso típico, provoca que, al truncar la aproximación a $m \ll N$ términos, el error esté dominado por el término $c_{m+1} T_m(x)$, una función oscilatoria con $m + 1$ valores extremos distribuidos uniformemente en el intervalo $[-1, 1]$. De ahí que la aproximación de Tchebycheff se encuentre muy próxima a la mejor aproximación uniforme, que presenta la característica de ser equioscilante.

2.5. Interpolación

Las técnicas de interpolación tienen por objeto la búsqueda de una función que pase por una serie de puntos dados. La existencia o no de una función (señal) subyacente de la que provienen los datos (muestras) dependerá de cada aplicación concreta, pero es claro que el conocimiento de cierta información relativa al comportamiento deseado de la solución entre los puntos de interpolación (ya sea por las características propias de la señal subyacente o por suposiciones relativas a qué es lo que se considera como buen comportamiento de la solución) resulta de gran importancia a la hora de seleccionar la técnica interpolatoria a usar. Dicho de otra forma, sin esa información todas las técnicas interpolatorias proporcionan resultados válidos, pasando a marcar diferencias aspectos referentes a la implementación práctica.

2.5.1. El Teorema de Muestreo

En 1949, C. E. Shannon publicó el artículo "*Communication in presence of noise*", sentando las bases de la teoría de la información [Shannon49, Shannon98]. Con el fin de formular su teoría, era necesario un mecanismo que permitiera convertir una señal continua analógica en una secuencia de números, lo que le condujo a establecer el clásico teorema de muestreo, basándose, en gran medida, en el trabajo previo de H. Nyquist[3] [Nyquist28]:

Teorema 2.5 (Teorema de Muestreo) *Una función $f(x)$ que no contenga frecuencias iguales o superiores a ω_{max}(rad/s) queda completamente determinada dado su valor en una serie de puntos equiespaciados $T = \pi/\omega_{max}$ (s).*

Como complemento al Teorema de Muestreo, la fórmula de reconstrucción (interpolación)

$$f(x) = \sum_{k\in\mathbb{Z}} f(kT)\,\mathrm{sinc}(x/T - k), \tag{2.6}$$

permite obtener la función original, $f(x)$, a partir de sus muestras equiespaciadas, $f(kT)$. Dicha reconstrucción será exacta si y sólo si la señal original está limitada en banda, es decir, si su frecuencia máxima es $\omega_{max} \leq \pi/T$, siendo este límite conocido como la *frecuencia de Nyquist*.

2.5.2. Técnicas de interpolación polinómicas

Como se recoge en [Meijering02], los polinomios se han utilizado desde antiguo para interpolar datos . Quizá sea la interpolación de Lagrange la más común de todas estas técnicas (basada en los trabajos previos de Newton y Waring), mientras que la interpolación de Hermite se aplica de forma habitual cuando se busca un polinomio que interpole tanto el valor de una función como sus derivadas.

El problema de la interpolación polinómica de Lagrange consiste en:

Definición 2.13 (Interpolación de Lagrange) *Dada una función real $f(x)$ definida en [a,b] y $n+1$ puntos distintos $x_0, x_1, \ldots, x_n$ en dicho intervalo, determinar un polinomio $P_n \in \mathcal{P}_n$ tal que*

$$P_n(x_k) = f(x_k), \quad k = 1, 2, \ldots, n.$$

[3] El propio C. E. Shannon recoge en su artículo las contribuciones de M. Whittaker [Whittaker29], H. Nyquist [Nyquist28] y D. Gabor [Gabor46]. En [Jerri77] se cita una serie de autores que, además de los ya comentados, llevaron a cabo con anterioridad trabajos relativos a la teoría del muestreo, como A. L. Cauchy, W. L. Ferrar [Ferrar27] y V. A. Kotelnikov [Kotelnikov33]

El problema de interpolación de Lagrange tiene solución única de la forma

$$P_n(x) = \sum_{k=0}^{n} f(x_k) L_{n,k}(x), \tag{2.7}$$

donde el k-ésimo polinomio de Lagrange de orden n es

$$L_{n,k}(x) = \prod_{\substack{i=0 \\ i \neq k}}^{n} \frac{x - x_i}{x_k - x_i}, \quad k = 0, 1, \ldots, n, \tag{2.8}$$

y el error de interpolación se puede expresar como

$$e(x) = \left. \frac{f^{(n+1)}(x)}{(n+1)!} \right|_{x=\xi(x)} \prod_{k=0}^{n} |x - x_k|, \tag{2.9}$$

siendo $\xi(x)$ un punto del intervalo $[a, b]$.

La aplicación directa de (2.7) y (2.8) plantea dos problemas: el primero de ellos es de precisión numérica, mientras que el segundo consiste en que los cálculos realizados para la obtención del polinomio interpolador de grado n no pueden ser utilizados para obtener el polinomio interpolador de grado $n + 1$. Ambas cuestiones se solventan utilizando el algoritmo de Neville [Press89], que se basa en la construcción recursiva de polinomios de grado creciente que interpolan en más puntos cada vez.

2.5.3. Divergencia de la interpolación polinómica

A la vista del teorema de Weierstrass, es habitual utilizar técnicas interpolatorias en aproximación funcional. Por ejemplo, es posible realizar la aproximación minimax de una función mediante la interpolación polinómica de ciertas muestras de dicha función. Sin embargo, esta metodología puede conducir a resultados que divergen conforme se incrementa el grado del polinomio (el número de puntos de interpolación). De hecho, existen muchos ejemplos en la literatura [Isaacson66, Cheney82] que demuetran que la interpolación Lagrangiana diverge,[4] incluso cuando se emplean puntos de interpolación no equiespaciados (por ejemplo, en los puntos de Tchebycheff). Este fenómeno de no convergencia puede considerarse como una manifestación del carácter oscilatorio de los polinomios [Schumaker81]. Si los polinomios son forzados a seguir una curva en un intervalo, entonces responderán oscilando, efecto éste que se torna más notorio conforme aumenta el grado del polinomio. Para órdenes bajos puede ser aceptable pero, desafortunadamente, es preciso utilizar polinomios de alto

[4] Por ejemplo, Runge propuso la función $(x^2 + 1)^{-1}$ en el intervalo $[-5, 5]$; y Bernstein $|x|$ en $[-1, 1]$.

orden con el fin de obtener una alta precisión. Desde otro punto de vista, la escasa flexibilidad y el carácter oscilatorio de los polinomios se debe a una de sus características, que previamente ha sido considerada como una virtud: los polinomios son suaves. De hecho, son demasiado suaves como para "seguir" a ciertas funciones. En concreto, del estudio del error de aproximación polinómica se deduce que, si las derivadas de la función a aproximar crecen rápidamente, el error de aproximación crecerá conforme aumenta el grado del polinomio.

En definitiva, el teorema de Weierstrass no puede utilizarse para justificar la aproximación polinómica basada en la interpolación en ciertos puntos de la función objetivo. Por otro lado, difícilmente la señal, o función, o el fenómeno físico subyacente, posee las características de un polinomio de alto orden, lo cual también induce a considerar otro tipo de técnicas interpolatorias.

2.5.4. Splines

En las secciones anteriores se ha constatado que para una aproximación precisa son necesarios polinomios de grado relativamente alto, lo cual plantea serias desventajas debido a su comportamiento oscilatorio. Sin embargo, tal y como ya se comentó en el Apartado 2.3.1, los polinomios poseen una serie de virtudes que animan a su utilización como interpoladores-aproximadores.

Los splines surgen como una técnica que utiliza polinomios aprovechando sus indudables ventajas a la vez que solventa la problemática de su divergencia. El procedimiento básico consiste en dividir el intervalo en un conjunto de subintervalos y construir un polinomio distinto (y de grado relativamente bajo) en cada uno de ellos; es decir, se trata de una interpolación/aproximación polinómica a tramos (*piecewise polinomial*). La interpolación a tramos más simple es la interpolación lineal a tramos (*piecewise linear*) consistente en unir un conjunto de puntos mediante una serie de segmentos rectos. Esta aproximación por funciones lineales ofrece la desventaja de la no derivabilidad en los extremos de los subintervalos, es decir, la solución no es "suave". Para solventarlo se podría optar por llevar a cabo una interpolación de Hermite en cada uno de los subintervalos, fijando el valor de las derivadas en los nodos hasta un orden adecuado. Sin embargo, esto requeriría del conocimiento del valor en los nodos de las derivadas de la función a aproximar, lo cual no siempre es posible.

La interpolación mediante splines es un tipo de interpolación polinómica a tramos consistente en aplicar un polinomio interpolatorio de grado m en cada subintervalo (esto es, cada uno de ellos une dos nodos consecutivos) de manera que los $m-1$ grados de libertad sobrantes pueden ser utilizados para forzar la continuidad de las derivadas hasta orden $m-1$. Existe gran cantidad de literatura relativa a los splines [Schoenberg46, Ahlberg67, Boor78, Schumaker81, Dierckx95]. Una definición formal de spline puede expresarse de la siguiente forma [Dierckx95]:

Definición 2.14 *Dada una función $f(x)$ definida en $[a,b]$ y un conjunto de nodos $a = x_0 < x_1 < \cdots < x_n = b$, $S(x)$ es una función spline de orden m si cumple con las condiciones siguientes:*

I) *En el subintervalo $[x_j, x_{j+1}]$, para cada $j = 0, 1, \ldots, n-1$, $S(x)$ es un polinomio de orden m y se denota $S_j(x)$.*

II) *$S(x_j) = f(x_j)$ para cada $j = 0, 1, \ldots, n-1$.*

III) *$S_{j+1}^{(k)}(x_j) = S_j^{(k)}(x_j)$ para cada $j = 0, 1, \ldots, n-1$ y $k = 1, \ldots, m-1$.*

Un resultado destacable, debido a I. J. Schoenberg [Schoenberg46], consiste en el hecho de que todo spline de orden m puede ser expresado como combinación lineal de ciertas funciones base

$$S(x) = \sum_{i=-m+1}^{n+m-1} c[i]\beta_{i,m}(x), \tag{2.10}$$

siendo $\beta_{i,m}(x)$ la denominada función B-spline (función base spline o función spline básica) de orden m asociada a los nodos $x_i, \ldots, x_{i+m}$, cuyo valor es

$$\beta_{i,m}(x) = (x_{i+m+1} - x_i) \sum_{j=0}^{m+1} \frac{(x_{i+m+1} - x_i)_+^m}{\prod\limits_{\substack{k=0 \\ k \neq j}}^{m+1} (x_{i+j} - x_{i+k})},$$

en cuya definición aparece la función potencia truncada:

$$(x-c)_+^m = \begin{cases} (x-c)^m, & x \geq c; \\ 0, & x < c. \end{cases}$$

Conviene hacer notar que en (2.10) es imperativo introducir $2(m-1)$ nodos adicionales (concretamente $x_{-m+1}, \ldots, x_{-1} \leq b$ a la izquierda del intervalo y $x_{n+1}, \ldots, x_{n+m-1} \geq a$ a la derecha) con el fin de hacer dicha representación única. En función de la elección de dichos nodos adicionales nos encontraremos ante diversas condiciones de frontera (periódica, especular, libre,...).

Entre las propiedades de las bases spline se encuentran:

- Positividad:

$$\beta_{i,m}(x) \geq 0, \quad \forall x.$$

- Soporte finito:

$$\beta_{i,m}(x) = 0, \quad x \notin [x_i, x_{i+m}].$$

- Valores de frontera:

$$\beta_{i,m}^{(1)}(x_i) = \beta_{i,m}^{(1)}(x_{i+m}) = 0, \quad l = 0, \ldots, m-2.$$

- Recursión [Boor78]:

$$\beta_{i,m}(x) = \frac{x - x_i}{x_{i+m-1} - x_i}\beta_{i,m-1}(x) + \frac{x_{i+m} - x}{x_{i+m} - x_{i+1}}\beta_{i+1,m-1}(x); \quad (2.11a)$$

$$\beta_{i,1}(x) = \begin{cases} 1, & x \in [x_i, x_{i+1}]; \\ 0, & x \notin [x_i, x_{i+1}]. \end{cases} \quad (2.11b)$$

- Derivación:

$$\beta_{i,m}^{(1)}(x) = (m-1)\left[\frac{\beta_{i,m-1}(x)}{x_{i+m-1} - x_i} - \frac{\beta_{i+1,m-1}(x)}{x_{i+m} - x_{i+1}}\right]$$

- Partición de la unidad en el intervalo $[a, b]$:

$$\sum_{i=-m+1}^{n+m-1} \beta_{i,m}(x) = 1, \quad x \in [a, b].$$

Una propiedad de las funciones spline de gran importancia surge del siguiente teorema:

Teorema 2.6 *De todas las funciones, $g(x)$, con derivada segunda continua que interpolan a una función $f(x)$ en los puntos $a = x_1 \leq x_2 \leq \ldots \leq x_{n-1} \leq x_n = b$, la función spline cúbica es la única que minimiza*

$$\int_a^b \left(g^{(2)}(x)\right)^2 dx. \quad (2.12)$$

Demostración 2.1 *La demostración de este teorema se puede encontrar en [Boor78].*

Se trata de la propiedad de la interpolación spline cúbica para proporcionar la interpolación más suave, puesto que (2.12) puede ser considerada como una aproximación de la energía de tensión (*strain energy*). Este hecho constituye una de las razones que han impulsado su utilización en numerosas aplicaciones; e, incluso, ha motivado su nombre, pues el spline cúbico aproxima la posición de una tira flexible (o de una curva a mano alzada o *draftman's spline*) que se fuerza a pasar por unos puntos dados.

Entre las aplicaciones de los splines en el tratamiento de señal, sin duda son aquellas relacionadas con el procesado de imagen las que mayor impacto

han tenido, gracias en gran medida a los trabajos de M. Unser. En concreto, y tal y como se describe en [Unser99], los splines se emplean en aplicaciones de transformación geométrica de imágenes, ampliación, compresión, procesamiento multiescala, detección de bordes y contornos, etc.

Hasta nuestros días, los splines no se han empleado con igual intensidad en el resto de áreas del procesado de señal. Entre las excepciones se encuentran aplicaciones relativas al diseño de filtros [Pang94]; el muestreo spline en sistemas de adquisición no ideales [Unser94] y sistemas multicanal[Unser98]; y la generación de señales spline cuadráticas [Kamada95] aplicada a la reproducción de sonido de alta fidelidad.

A modo de resumen, se enumeran alguna de las ventajas más importantes de la representación de funciones mediante combinación lineal de B-splines [Unser97, Unser99, Neubauer04]:

- El soporte finito de las funciones B-spline posibilita una evaluación exacta de la señal a partir de los coeficientes de la expansión, en contraste con el soporte infinito de la función $\mathrm{sinc}(x)$ asociada al Teorema de Muestreo.
- La representación spline de una señal es computacionalmente más eficiente que las técnicas basadas en la función $\mathrm{sinc}(x)$.
- La representación spline engloba la representación basada en la función $\mathrm{sinc}(x)$, pues este último es el caso particular para splines de orden infinito [Unser92].
- La función B-spline de orden m es derivable hasta orden m.
- La integración y derivación de funciones spline se realiza de forma sencilla a partir de los coeficientes de la expansión y mediante simples manipulaciones de funciones B-spline.
- Los splines cúbicos (es decir, de orden $m = 3$) poseen la propiedad de *mínima curvatura* o, de forma equivalente, de *máxima suavidad*.

La principal desventaja de la representación spline es su no invarianza respecto a desplazamientos no enteros de las señales. En esta línea, en los últimos años han aparecido diversos trabajos que evalúan el error producido por dichos desplazamientos [Blu99a, Blu99b, Wang00, Neubauer04].

2.5.4.1. Algoritmo de interpolación spline

Para interpolar mediante una función spline una serie de datos es preciso llevar a cabo dos procedimientos. El primero de ellos, conocido como problema directo, consiste en obtener los coeficientes spline de la combinación lineal dada por (2.10). El segundo, llamado problema indirecto, consiste en evaluar la función spline en los puntos necesarios.

Respecto al segundo de los problemas, el indirecto, y debido al soporte compacto de las bases splines, la evaluación de la función spline en un punto $x \in [x_j, x_j + 1)$ se lleva a cabo mediante la suma

$$S(x) = \sum_{i=j-m+1}^{j} c[i]\beta_{i,m}(x), \tag{2.13}$$

que depende únicamente de $m+1$ coeficientes y de los valores de las $m+1$ bases splines, que pueden ser calculados muy eficientemente (y de forma estable) mediante (2.11). Ésta es una de las grandes virtudes de la representación spline, en contraste con el soporte infinito de la función base del Teorema de Muestreo (2.6). Por ello, la representación de una señal en el espacio spline es más eficiente computacionalmente que su representación basada en la función $\mathrm{sinc}(x)$.

Para la obtención de los coeficientes splines, problema directo, se pueden plantear dos estrategias diferentes. Originalmente la solución del problema directo se basada en inversion y multiplicación matricial [Schoenberg46, Boor78]. En dichas referencias se aprovecha el soporte compacto de las bases spline que intervienen en la representación spline (2.13), pues este hecho provoca que las matrices asociadas al problema tengan características de banda (únicamente las m diagonales centrales son no nulas), posibilitando la utilización de algoritmos eficientes como la eliminación Gaussiana o la factorización LU [Golub89].

Quizá haya sido este enfoque matemático el que, durante años, ha limitado la aplicación de los splines en el ámbito del procesado de señal. Sin embargo, los trabajos desarrollados por M. Unser [Unser93a, Unser93b], en los que se particularizaba la teoría spline para conjuntos de nodos equiespaciados y se proponían técnicas para determinar los coeficientes spline mediante filtrado lineal, posibilitaron y promovieron la utilización de los splines en el mundo del procesado de señal.

En efecto, si se consideran nodos equiespaciados, es decir, si $x_n = n, \forall\, n \in \mathbb{Z}$, y los valores de la función a interpolar en esos puntos son $f[n] = f(x)|_{x=n}$, entonces la representación spline (2.10) resulta

$$S(x) = \sum_{n\in\mathbb{Z}} c[n]\beta_m(x-n), \tag{2.14}$$

donde la base spline, $\beta_m(x)$, idéntica en todos los intervalos, obedece a la expresión

$$\beta_m(x) = \frac{1}{m!}\sum_{j=0}^{m+1}\binom{m+1}{j}(-1)^j\left(x-j+\frac{m+1}{2}\right)_+^m.$$

A la vista de (2.14), se puede interpretar la expansión spline como la convolución de los coeficientes spline, $c[n]$, con la función base spline, $\beta_m(x)$, resultando éste el punto de conexión entre la teoría spline y el procesado de

señal. Así, es evidente que los coeficientes spline que verifiquen la condición de interpolación dada en la Definición 2.14.2 se obtienen mediante filtrado inverso

$$c[n] = (b_m)^{-1}[n] * f[n],$$

donde $(b_m)^{-1}[n]$ es el filtro inverso de la base spline de orden m discretizada, es decir,

$$(b_m)^{-1}[n] \overset{z}{\longleftrightarrow} 1/B_m(z),$$

siendo $b_m[n]$ el núcleo spline discreto

$$b_m[n] = \beta_m(x)\,|_{x=n} \overset{z}{\longleftrightarrow} B_m(z) = \sum_{n\in\mathbb{Z}} \beta_m(n)\, z^{-n}.$$

Por lo tanto, la solución al problema directo, es decir, el cálculo de los coeficientes spline, se puede llevar a cabo mediante el filtrado de las muestras de la función (o, mejor, de la señal) con el denominado filtro directo spline, $(b_m)^{-1}[n]$, un filtro todo polos que puede ser implementado muy eficientemente utilizando una conexión en serie de filtros recursivos [Unser93b]. Este es el procedimiento que se utilizará a lo largo de esta Tesis, proporcionándose una descripción más detallada del mismo (y del método matricial alternativo) en el Capítulo 4.

2.5.4.2. Spline suavizado o regularizado

Cuando los datos de partida son ruidosos, un ajuste perfecto (es decir, la interpolación) puede no ser deseable, pues podría provocar sobreajuste u oscilaciones. Una solución consiste en relajar la restricción interpolatoria en la Definición 2.14.1 y buscar la función spline, $S(x)$, que minimice

$$\sum_j \left(f(x_j) - S(x_j)\right)^2 + \lambda \int_{-\infty}^{\infty} \left(S^{(m)}(x)\right)^2 dx. \tag{2.15}$$

Éste es un problema regularizado de mínimos cuadrados en el que el primer término cuantifica el error entre el modelo, $S(x)$, y los datos, $f(x_j)$, mientras que el segundo término impone una restricción de suavidad sobre la solución. La elección de un valor concreto del factor de regularización λ depende de la información disponible *a priori*: en base al conocimiento de la varianza de ruido en los datos, o de la suavidad de la función original (si es que es conocida). La solución al problema planteado en (2.15) se denomina spline suavizado (*smoothing spline*), pues lleva a cabo un suavizado de los datos.

2.6. Muestreo y reconstrucción en espacios de Hilbert

Muestreo es el proceso consistente en representar funciones de una variable continua mediante secuencias de números. En este sentido, el Teorema de Muestreo (Teorema 2.5) proporciona el enfoque clásico de procesado de señal a este problema; mientras que otras técnicas de interpolación (o, de forma equivalente, de muestreo y reconstrucción), entre las que se encuentran la interpolación de Lagrange, de Hermite y las basadas en splines, no han encontrado, hasta el momento, una amplia difusión en el tratamiento de señal. Es evidente que existen puntos en común entre las diversas técnicas, y que es deseable un marco formal que las englobe. En esta sección se revisa el Teorema de Muestreo con el fin de enunciarlo en términos de espacio de señal y de proyecciones para, posteriormente, presentar un marco general de muestreo y reconstrucción en espacios de señal invariantes con los desplazamientos. En todo caso, se considerará únicamente muestreo de tipo uniforme, en el que las muestras se toman a intervalos equiespaciados. Así mismo, se utilizará como variable independiente el tiempo, denotado por t (n para el caso de la variable independiente discreta), con el fin de emplear una notación propia de procesado de señal. Existen numerosas referencias en los últimos años relativas a estos conceptos que han supuesto un enorme impacto en el área del procesado de señal, encontrándose entre las más destacadas [Jerri77, Higgins96, Unser98, Unser00, Vaidyanathan01a].

2.6.1. Revisión del Teorema de Muestreo

El Teorema de Muestreo establece que una función limitada en banda, $x(t)$, resulta completamente caracterizada por sus muestras, $x[n]$, tomadas a una determinada frecuencia [Shannon49, Jerri77]. Junto con el procedimiento de muestreo de la señal continua, del teorema también se desprende el método de reconstrucción, que posibilita la recuperación, sin error, de la señal original a partir de sus muestras. Este teorema es aplicable siempre que la señal sea limitada en banda. En caso contrario, la práctica habitual consiste en aplicar un filtro paso bajo previo al muestreo con el fin de evitar el solapamiento espectral. La solución óptima es el filtro paso bajo ideal (con respuesta frecuencial $H(\Omega) = \mathrm{rect}(\Omega T)$ y respuesta al impulso $h(t) = \mathrm{sinc}(t/T)/T$) que elimina completamente el solapamiento sin introducir distorsión en la banda de interés. En este apartado se proporciona una interpretación geométrica en el espacio de Hilbert del paradigma de muestreo tradicional. Para comprender este nuevo punto de vista se considera, por simplicidad notacional, un periodo de muestreo unidad, $T = 1$, y se trabaja con señales pertenecientes al espacio de Hilbert L_2 consistente en todas las funciones cuadráticamente integrables en el sentido de Lebesgue (es decir, las señales de energía finita). La correspondiente norma L_2

es

$$\|x(t)\| = \left(\int_a^b |x(t)|^2 dt \right)^{1/2} = \sqrt{\langle x(t), x(t) \rangle} = \sqrt{\text{Energía}(x(t))},$$

inducida por el producto interno L_2

$$\langle x(t), y(t) \rangle = \int_{-\infty}^{\infty} x(t) y(t) dt.$$

Si se supone que la señal que se desea muestrear pertenece a L_2, un espacio de señal considerablemente más amplio que el subespacio de las señales limitadas en banda (al que denotaremos por V) entonces será necesario llevar a cabo cierta aproximación para representar una señal no limitada en banda en el espacio limitado en banda V. Haciendo una analogía, L_2 es a V lo que $\mathbb{R}$ (los números reales) es a $\mathbb{Z}$ (los enteros).

La naturaleza discreta de V se manifiesta claramente si se reescribe (2.6) con $T = 1$ como

$$x(t) = \sum_{n \in \mathbb{Z}} c[n] \varphi(t - n),$$

siendo $\varphi(t - n) = \text{sinc}(t - n)$ las funciones base y $c[n]$ ciertos coeficientes. Es sencillo comprobar que las funciones base son ortonormales, pues

$$\langle \varphi(t - n), \varphi(t - k) \rangle = \int_{-\infty}^{\infty} \varphi(t - n) \varphi(t - k) dt = \text{sinc}\,(n - k) = \delta[n - k].$$

Esta propiedad ortonormal simplifica enormemente la implementación del proceso de aproximación consistente en proyectar una señal $x(t) \in L_2$ sobre el espacio V. En concreto, se define el operador proyección ortogonal $P_V : L_2 \to V$ como

$$P_V(x(t)) = \sum_{n \in \mathbb{Z}} \langle x(t), \varphi(t - n) \rangle \, \varphi(t - n), \tag{2.16}$$

donde el producto interno $c[k] = \langle x(t), \varphi(t - k) \rangle$ representa la contribución de la señal $x(t)$ según la dirección representada por $\varphi(t - k)$. Por lo tanto, debido a la ortogonalidad de las funciones base, este problema de aproximación se encuentra desacoplado; es decir, es posible evaluar la contribución de la señal según la dirección de una de las funciones base independientemente de las demás. El teorema de proyección [Kreyszig78] asegura que esta operación de proyección está bien definida y que proporciona la aproximación con menor error de $x(t)$ en V, es decir,

$$\tilde{x}(t) = P_V(x(t)) = \arg \min_{y(t) \in V} \|x(t) - y(t)\|^2 .$$

El producto interno en (2.16) es equivalente a filtrar la señal de entrada con el filtro paso bajo ideal y muestrear posteriormente. De forma más genérica,

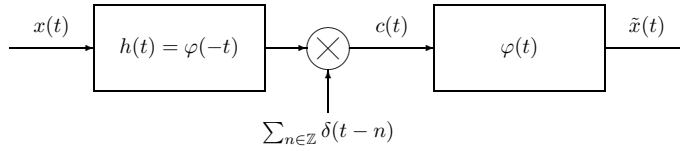


Figura 2.2: Esquema de muestreo y reconstrucción dictado por el Teorema de Muestreo, donde $h(t) = \varphi(-t) = \varphi(t) = \mathrm{sinc}(t)$ [Unser00].

cualquier combinación de prefiltrado y muestreo puede ser reescrita en términos de productos internos

$$(h(t) * x(t))|_{t=n} = \int x(t)h(n-t)dt = \langle x(t), h(n-t)\rangle\,;$$

es decir, las funciones de análisis, $h(n-t)$, se corresponden con versiones desplazadas e invertidas de la respuesta al impulso $h(t)$ (que puede ser arbitraria). En el caso del Teorema de Muestreo, resulta que $h(t) = h(-t) = \mathrm{sinc}(t)$ y que $h(t) * x(t)$ representa la versión filtrada paso bajo ideal de la señal $x(t)$.

En resumen, el paradigma del Teorema de Muestreo con un prefiltro paso bajo ideal coincide con la aproximación $\tilde{x}(t) = P_V(x(t))$, que resulta ser la proyección ortogonal de la señal de entrada sobre V (el espacio de las señales limitadas en banda). En otras palabras, $\tilde{x}(t)$ es la aproximación mínimo cuadrática de $x(t)$ en V. A la luz de esta interpretación geométrica, es obvio que $x(t) \in V$ si y sólo si $x(t) = P_V(x(t))$. Finalmente, y tal como representa el esquema de la Figura 2.2, el esquema de muestreo-reconstrucción puede contemplarse en tres etapas: primeramente, la señal continua de entrada, $x(t)$, es prefiltrada con un filtro antisolapamiento con respuesta al impulso $h(t) = \varphi(-t)$; seguidamente, el proceso de muestreo proporciona la representación discreta $c(t) = \sum_{n\in\mathbb{Z}} c[n]\delta(t-n)$; y, por último, la señal reconstruida $\tilde{x}(t) = \sum_{n\in\mathbb{Z}} c[n]\varphi(t-n)$ se obtiene mediante el filtrado de $c(t)$ con el filtro $\varphi(t)$. En caso de trabajar de acuerdo al Teorema de Muestreo, tanto el prefiltro como el postfiltro son paso bajo ideales, $h(t) = \varphi(t) = \mathrm{sinc}(t)$.[5]

Si la función a procesar no está limitada en banda, por ejemplo, si $x(t)$ pertenece a L_2, el procedimiento consiste en utilizar un prefiltro (filtro paso bajo ideal) que evite el solapamiento espectral (como muestra la Figura 2.2). De esta forma, la señal reconstruida, $\tilde{x}(t)$, no es más que la proyección ortogonal de $x(t)$ en el subespacio de las funciones limitadas en banda, lo que sugiere interpretar el Teorema de Muestreo como si de un procedimiento de aproximación se tratara. Este punto de vista conduce a la formulación de la teoría de muestreo

[5] Sin embargo, desde un punto de vista más amplio, los filtros pueden elegirse libremente sin más que considerar la restricción de biortogonalidad, es decir, $\langle \varphi(t-n), h(k-t)\rangle = \delta[n-k]$, tal y como se comentará posteriormente.

de los splines de menor error cuadrático [Unser92], así como de un conjunto de señales más amplio, las generadas mediante traslaciones enteras de una base o núcleo generador, $\varphi(x)$ [Aldroubi94]. De forma similar, Ogawa desarrolla su teoría de muestreo generalizada para aproximar funciones en espacios vectoriales de dimensión finita [Owaga86]. El principio de aproximación funcional con mínimo error (proyección ortogonal) es utilizado también en la teoría de *wavelets*. Todas estas representaciones comparten la característica de que los coeficientes discretos de la expansión se obtienen mediante el producto interno (convolución) de la función (señal) de entrada, $x(t)$, y una función (filtro) de análisis apropiado; lo que es equivalente a prefiltrar la señal y muestrearla posteriormente. De esta formulación se ocupa el siguiente apartado.

2.6.2. Muestreo en espacios invariantes frente a desplazamientos

Es posible ampliar el marco descrito en la sección anterior para incluir otras clases de funciones generadoras. A pesar de que es posible considerar cualquier conjunto arbitrario de señales como funciones base, es deseable obtener un esquema de muestreo y reconstrucción que resulte práctico y que mantenga el principio de invarianza ante desplazamientos del Teorema de Muestreo. Esto se consigue reemplazando la función $\mathrm{sinc}(t)$ por una *función generadora*, $\varphi(t)$, que define su propio espacio de señal

$$V(\varphi) = \left\{ s(t) = \sum_{n=\in\mathbb{Z}} c[n]\varphi(t-n) \ : \ c[n] \in l_2 \right\}.$$

Esto significa que cualquier función $s(t) \in V(\varphi)$ puede expresarse como combinación lineal cuyos coeficientes, $c[n]$, no tienen que ser necesariamente las muestras de la señal; es decir,

$$s(t) = \sum_{n\in\mathbb{Z}} c[n]\varphi(t-n). \tag{2.17}$$

Por otro lado, la *función generadora* $\varphi(t)$ puede ser distinta de $\mathrm{sinc}(t)$, es más, es posible trabajar con funciones más simples y más eficientes desde el punto de vista computacional (por ejemplo, las bases spline).

Para que este modelo de representación de señales tenga sentido es necesario imponer al menos tres condiciones [Unser00]. En primer lugar, la secuencia de coeficientes debe poseer energía finita; es decir, $c[n] \in l_2$:

$$\sum_{n\in\mathbb{Z}} |c[n]|^2 = ||c||_{l_2}^2 < \infty;$$

y, en segundo lugar, la representación (2.17) tiene que ser estable y no ambigua. Estas dos primeras condiciones son equivalentes a que la familia de funciones $\{\varphi(t-n)\}_{n\in\mathbb{Z}}$ forme una base de Riesz del espacio $V(\varphi)$.

Definición 2.15 (Base Riesz) *El conjunto de señales* $\{\varphi(t-n)\}_{n\in\mathbb{Z}}$ *es una base de Riesz si existen dos constantes estrictamente positivas* $0 < A, B < +\infty$ *tales que*

$$\forall\ c[n] \in l_2, \quad A||c||_{l_2}^2 \leq \left\| \sum_{n\in\mathbb{Z}} c[n]\varphi(t-n) \right\|^2 \leq B||c||_{l_2}^2, \tag{2.18}$$

o, de forma equivalente, si

$$A \leq \sum_{k\in\mathbb{Z}} |\Phi(\Omega - 2\pi k)|^2 \leq B, \tag{2.19}$$

donde $\Phi(\Omega) = \mathrm{TF}\,\{\varphi(t)\}$.

Una consecuencia directa de la desigualdad inferior de (2.18) es que si la señal es nula ($s_\alpha(t) = \sum_n c[n]\varphi(t-n) = 0$) entonces necesariamente los coeficientes, $c[n]$, deben ser nulos. Por lo tanto, las funciones base son linealmente independientes, lo que también significa que toda señal $s(t) \in V(\varphi)$ está totalmente determinada por sus coeficientes $c[k] \in l_2$. La cota superior de (2.18) implica que la norma L_2 de la señal $s(t)$ es finita, por lo que $V(\varphi)$ es un subespacio válido de L_2. Nótese que las bases son ortonormales si y sólo si $A = B = 1$, en cuyo caso existe una equivalencia entre la norma en el dominio continuo y la norma en el dominio discreto (relación de Parseval).

Es preciso imponer una tercera y última condición a la *función generadora*, consistente en que (2.17) sea capaz de aproximar cualquier señal de forma tan precisa como sea necesario sin más que seleccionar un periodo de muestreo suficientemente pequeño. Como se demuestra en [Unser00], esta condición equivale a que las funciones base formen *partición de la unidad*.

Definición 2.16 (Partición de la unidad) *El conjunto* $\{\varphi(t-n)\}_{n\in\mathbb{Z}}$ *forma partición de la unidad si*

$$\sum_{n\in\mathbb{Z}} \varphi(t-n) = 1, \quad \forall\ t \in \mathbb{R}.$$

En la práctica, esta última condición es la más difícil de cumplir en la búsqueda de *funciones generadoras*. Ejemplos de funciones base admisibles son la propia del Teorema de Muestreo, $\mathrm{sinc}(t)$; las funciones base spline, $\beta\,(t)$; y las funciones de escala de la teoría de *wavelets*.

El siguiente paso consiste en obtener los coeficientes $c[n]$ en (2.17) de forma que el modelo de señal sea una representación adecuada de una señal de entrada

$x(t) \in L_2$. La solución óptima, en el sentido de menor error cuadrático, es la proyección ortogonal dada por

$$P_{V(\varphi)}(x(t)) = \sum_{n\in\mathbb{Z}} \left\langle x(t), \overset{\circ}{\varphi}(t-n) \right\rangle \varphi(t-n), \tag{2.20}$$

donde $\overset{\circ}{\varphi}(t-n)$ es la función base dual de $\varphi(t-n)$. Esta expresión es muy similar a (2.16) salvo por el hecho de que las funciones de análisis y síntesis ($\overset{\circ}{\varphi}(t-n)$ y $\varphi(t-n)$, respectivamente) no son idénticas. El conjunto de bases duales

$$\left\{\overset{\circ}{\varphi}(t-n)\right\}_{n\in\mathbb{Z}} \text{ con } \overset{\circ}{\varphi}(t-n) \in V(\varphi)$$

es único y queda determinado por la condición de biortogonalidad

$$\left\langle \overset{\circ}{\varphi}(t-n), \varphi(t-k) \right\rangle = \delta[n-k].$$

De hecho, las funciones base duales pueden evaluarse de forma simple en el dominio de la frecuencia como

$$\overset{\circ}{\Phi}(\Omega) = \frac{\Phi(\Omega)}{\sum_{k\in\mathbb{Z}} |\Phi(\Omega+2\pi k)|^2}, \tag{2.21}$$

siendo $\overset{\circ}{\Phi}(\Omega) = \mathrm{TF}\left\{\overset{\circ}{\varphi}(t)\right\}$ y $\Phi(\Omega) = \mathrm{TF}\{\varphi(t)\}$. Nótese que la condición de base de Riesz (2.19) garantiza que esta solución siempre esté bien definida.

De forma similar a lo comentado en la interpretación geométrica del Teorema de Muestreo, el algoritmo de muestreo y reconstrucción dado por (2.20) admite una interpretación basada en conceptos de procesado de señal, tal y como muestra la Figura 2.3.a. En efecto, el procedimiento es similar al dictado por el Teorema de Muestreo excepto por el hecho de que el filtro de análisis (filtro antisolapamiento) óptimo en el sentido de mínimos cuadrados, $\overset{\circ}{\varphi}(-t)$,

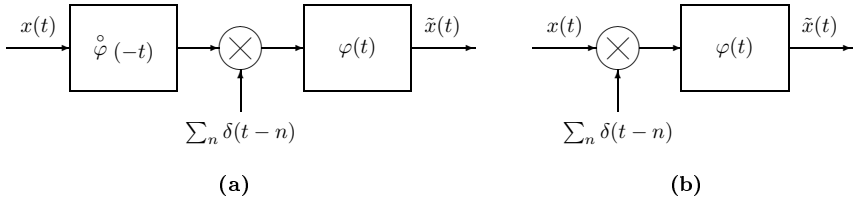


Figura 2.3: Esquemas de muestreo: a) aproximación mínimocuadrática; b) interpolación.

queda completamente especificado por la elección de la *función generadora* (o filtro reconstructor), $\varphi(t)$; pues su respuesta frecuencial viene dada por (2.21). Si $\varphi(t)$ fuera ortonormal, entoces él mismo sería su propia función de análisis (es decir, $\overset{\circ}{\varphi}(t) = \varphi(t)$), por lo que el prefiltro poseería una respuesta al impulso consistente en una inversión temporal del filtro reconstructor. En cualquier caso, esta condición de ortonormalidad sólo se cumple en ciertas ocasiones (por ejemplo, al considerar como *función generadora* la función $\mathrm{sinc}(t)$, o la función pulso rectangular de duración unitaria), aunque en otras muchas no ocurre, sin perder utilidad por este motivo (un ejemplo evidente son las funciones base spline).

En definitiva, una *base generadora*, $\varphi(x)$, define el subespacio de señal destino y determina de forma unívoca cuál debe ser el prefiltro, $\overset{\circ}{\varphi}(x)$, a aplicar a la señal, de forma que, si ésta no pertenece al subespacio considerado, el error cuadrático que se cometa sea el menor posible. Si el prefiltro no es el adecuado aparecerán errores adicionales en forma de distorsión y solapamiento (*aliasing*). En este sentido, si no se utiliza prefiltro estaremos ante una situación de interpolación estricta (Figura 2.3.b), de tal manera que, si la señal $x(t)$ no pertenece al subespacio definido por la *función generadora*, el resultado no será el óptimo, es decir, existirá otra señal perteneciente al subespacio V_φ más próxima a $x(t)$ que la obtenida mediante la interpolación de sus muestras.

2.7. Modelado

El problema del modelado, también conocido como ajuste de curva (*curve fitting*), puede ser plantedo de la siguiente forma: dado un conjunto de valores y_i, $i = 1, \ldots, N$, de la variable dependiente y, correspondientes a los valores $\mathbf{x}_i$, $i = 1, \ldots, N$, de la variable, o vector, independiente $\mathbf{x}$, determinar una función $\hat{f}(\mathbf{x}) \equiv f(\mathbf{x}; \Theta)$ de forma conocida y con un vector de n parámetros, Θ, que se ajuste a los y_i de tal forma que $\hat{f}(\mathbf{x}_i) \approx y_i$. Esta última condición debe ser especificada en función de una cierta métrica.

Existen diversos motivos que justifican acometer un problema de modelado, entre los que se encuentran los siguientes:

Estimación de parámetros: la forma de la función modeladora $f(x)$ puede estar determinada por la aplicación o el proceso a modelar, en cuyo caso, los parámetros Θ_k, $k = 1, \ldots, n$ posiblemente tendrán un significado físico. El objetivo consistirá en estimar dichos parámetros de forma tan precisa como sea posible a partir de los datos disponibles.

Suavizado de datos: si el conjunto de datos de partida y_i fuera suficientemente preciso, podría bastar con determinar una función interpolatoria $f(x)$ de tal forma que $f(x_i) = y_i$ para $i = 1, \ldots, N$. Sin embargo, en la

mayoría de las aplicaciones los valores y_i están corrompidos por errores de medida, en cuyo caso sería deseable que el proceso de modelado suavice dichos errores.

Representación funcional: la representación de un conjunto discreto de pares de datos (x_i, y_i) mediante una función $f(x)$ tiene ciertas ventajas, entre las que se encuentran el hecho de que se puede obtener de forma inmediata el valor en cualquier punto x dentro del rango o determinar el valor de la derivada, llevar a cabo la integral, etc.

Compresión de datos: si los resultados deben ser almacenados o transmitidos, resulta útil que el número de parámetros Θ_k sea significativamente menor que el número de datos (modelado parsimonioso).

2.7.1. Técnicas polinómicas de modelado: consideraciones generales

Existe en la literatura gran cantidad y diversidad de técnicas de modelado, siendo ejemplos recientes las redes neuronales (*Neural Networks*) [Haykin94] y las máquinas de vectores soporte (*Support Vector Machines*) [Vapnik00]. Sin embargo, esta sección, y el resto de la Tesis, se centrará en las técnicas de modelado polinómicas. Para realizar una clasificación de las distintas metodologías es útil establecer una representación del modelo en forma de combinación lineal de funciones base, pues resulta aplicable a todas las técnicas a considerar. De esta manera [Cherkassky96], el modelo se puede expresar como

$$\hat{f}(\mathbf{x}) = \sum_{j=0}^{M} a_j B_j(\mathbf{x}, \mathbf{p}_j),$$

donde

1. $\mathbf{x}$ es el vector de variables de entrada.
2. a_j son los coeficientes de la expansión que deberán ser determinados a partir de los datos.
3. $B_j(\mathbf{x}, \mathbf{p}_j)$ son las funciones base.
4. $\mathbf{p}_j$ son los parámetros de cada función base.
5. M es el número de funciones base del modelo.

Los métodos basados en esta representación son conocidos como métodos diccionario, pues el conjunto de funciones base (diccionario) utilizado es lo que diferencia a los distintos métodos. Es posible realizar una clasificación de los mismos de la siguiente manera:

1. **Métodos no adaptativos** utilizan funciones base preestablecidas (y con sus parámetros $\mathbf{p}_j$ prefijados), de forma que únicamente los coeficientes a_j se ajustan a los datos utilizando, habitualmente, el criterio de mínimos cuadrados, resultando un problema lineal.

   - Métodos paramétricos globales, como por ejemplo la regresión polinómica. Este tipo de modelado tiene una flexibilidad muy limitada y proporciona modelos precisos únicamente cuando la función subyacente al proceso a modelar es similar a la preestablecida. Por contra, a su favor juega su sencillez, la necesidad de pocos datos para la obtención de los parámetros, la facilidad de interpretación y la rapidez de computación.

   - Métodos paramétricos locales, como la regresión lineal a tramos (PWL) o los splines. Estos métodos extienden los principios del modelado paramétrico global mediante su aplicación local a regiones o tramos del espacio de entrada y la imposición de ciertas condiciones de continuidad y/o suavidad. La aplicabilidad de este tipo de modelado se limita a problemas de baja dimensión en el espacio de entrada (típicamente una o dos variables de entrada) pues el número de coeficientes o parámetros del modelo crece exponencialmente con la dimensión, lo cual, además de resultar problemático, requiere de un número exponencialmente creciente de datos para su evaluación de forma robusta. Es el conocido problema de "la maldición de la dimensionalidad" ("*the curse of dimensionality*"). Los splines se utilizan frecuentemente en problemas de modelado y ajuste de curvas (*spline curve fitting*). El método más simple basado en splines considera prefijados los puntos de ruptura o nodos, lo que lo convierte en un problema lineal respecto de los coeficientes de las funciones base spline, que se resuelve de manera simple. Se trata, por tanto, de un método de modelado paramétrico local no adaptativo. Sus prestaciones (relativas a la precisión del ajuste) dependen enormemente del número de nodos considerado y de sus posiciones (que se deben prefijar de antemano). El planteamiento y resolución de este problema se contempla en [Boor78, Dierckx95].

2. **Métodos adaptativos**, en los cuales, además de los coeficientes a_j, también se ajustan a los datos las funciones base y/o sus parámetros $\mathbf{p}_j$. Se trata, por tanto, de un problema no lineal, de considerable dificultad y en el que la estrategia de optimización a utilizar adquiere gran importancia. Así mismo, y de forma similar a los métodos no adaptativos, la elección de un conjunto de funciones base (diccionario) adecuado adquiere gran importancia. Estos métodos adaptativos también pueden clasificarse en:

- Globales, que utilizan funciones base definidas en todo el dominio de $\mathbf{x}$. Entre ellos se encuentran las redes neuronales (como el perceptrón multicapa) y la regresión por búsqueda de proyección (PPR: *Projection Pursuit Regression*) así como otros más proximos al ámbito de la estadística como los modelos aditivos o el producto tensorial de splines (MARS: *Multivariate Adaptative Regression Splines*).
- Locales, que hacen uso de funciones base locales (en el dominio del espacio de entrada) como en el caso de las redes neuronales que utilizan funciones base radiales o la regresión por vecino más cercano (NNR: *Nearest Neighbors Regression*), el método adaptativo regularizado, el CART, etc. Si se utilizan modelos spline en los que la posición de los nodos puede adaptarse para ajustarse a los datos, entonces se plantea un problema no lineal que se engloba dentro del grupo de métodos locales adaptativos. La obtención de una posición óptima en el modelado spline no es una tarea trivial. Numerosos autores [Boor78, Jupp78, Dierckx95, Friedman91, Yiu01] se han planteado el problema no lineal en el que la posición de los nodos debe ser optimizada, resultando algoritmos de gran coste computacional y con una elevada sensibilidad ante mínimos locales.

Capítulo 3

Técnicas polinómicas en procesado de señal: consideraciones prácticas

3.1. Introducción

En la actualidad, son numerosas las técnicas polinómicas que se emplean con éxito en aplicaciones de procesado de señal [Mathews00]. Sin embargo, dichas aplicaciones se restringen, principalmente, al procesado de imagen (de la mano de los splines) y al modelado no lineal, no habiendo encontrado gran difusión en otros campos como la reconstrucción/interpolación de señales y las comunicaciones digitales.

De hecho, y motivado en gran medida por la simplicidad, elegancia y enorme difusión e impacto del Teorema de Muestreo, en numerosas ocasiones se asumen las condiciones que impone (limitación en banda) y se considera su metodología (basada en la función interpolatoria $\operatorname{sinc}(t)$) como primera, y en la mayoría de los casos única, aproximación a la resolución de los problemas de procesado de señal. En este capítulo se presentarán una serie de ejemplos que demuestran que, bajo determinadas condiciones prácticas, existen técnicas polinómicas que proporcionan resultados iguales o incluso superiores (en términos de precisión y requerimientos computacionales) que los proporcionados por la aplicación "automática" del Teorema de Muestreo. Dichos ejemplos pretenden servir de motivación de esta Tesis, justificando la búsqueda de nuevas técnicas polinómicas en procesado de señal.

En primer lugar se realizará un estudio comparativo entre el método "tradicional" de interpolación de señales limitadas en banda y la interpolación spline

cúbica. En la comparación se considera la dependencia con el tipo de señal a interpolar (más concretamente, con su contenido espectral) y el efecto de la precisión finita de las muestras debida a la resolución limitada de los conversores analógico-digitales reales. Así mismo, se tomará en cuenta la degradación que introduce la utilización de filtros antisolapamiento realizables y el efecto del sobremuestreo.

Como segundo ejemplo que ilustra las ventajas en ciertas aplicaciones del procesado polinómico, se presenta una solución al problema del retardo fraccionario basada en la interpolación de Lagrange, obteniéndose, de nuevo bajo condiciones prácticas habituales, resultados superiores a la interpolación $\operatorname{sinc}(t)$.

3.2. Interpolación sinc frente a la interpolación spline

En este apartado se realiza una comparación, bajo condiciones prácticas reales, entre el método de interpolación tradicionalmente empleado en el procesado de señal (basado en el Teorema de Muestreo y, por tanto, en la función $\operatorname{sinc}(t)$) y la interpolación spline cúbica. Así, se tomará en consideración la disminución que provoca en el rendimiento de las distintas técnicas la precisión finita de los valores de las muestras motivada por la cuantificación introducida por los conversores analógico-digitales. Igualmente, se tendrá en cuenta el efecto de la utilización de filtros antisolapamiento realizables en lugar de los filtros paso bajo ideales que exige el Teorema de Muestreo. Por último, resulta evidente que, en función del tipo de señal a interpolar, más concretamente, de su contenido espectral, las prestaciones de los distintos métodos de interpolación variarán. Por ello es preciso considerar, en primer lugar, un modelo de las señales a interpolar que permita llevar a cabo un estudio preciso de los distintos métodos y extraer conclusiones significativas.

3.2.1. Modelo de la señal a interpolar

A la hora de establecer modelos o familias de señales para evaluar el comportamiento de las distintas técnicas de interpolación-reconstrucción se tomará como característica diferenciadora su densidad espectral de potencia (DEP), pues de ello depende, en gran medida, el error cometido en la interpolación.

3.2.1.1. Señales blancas con frecuencia máxima variable

El Teorema de Muestreo se aplica a señales limitadas en banda. El modelo más sencillo de señales que pertenecen a esta familia es el que considera señales con densidad espectral de potencia constante en la banda $[-f_c, f_c]$ y nula para

frecuencias superiores, es decir,

$$S_x(f) = \begin{cases} \frac{1}{2f_c}, & |f| \leq f_c; \\ 0, & |f| > f_c. \end{cases} \tag{3.1}$$

Dicha señal es muestreada a f_s muestras por segundo, por lo que la potencia de señal solapada (*aliasing*) se obtiene como

$$P_a = 2\int_{\frac{f_s}{2}}^{\infty} S_x(f)df,$$

por lo que, considerando señales con DEP como en (3.1), la anterior expresión resulta

$$P_a = \begin{cases} 0, & f_c \leq \frac{f_s}{2}; \\ 1 - \frac{f_s}{2f_c}, & f_c > \frac{f_s}{2}. \end{cases} \tag{3.2}$$

En muchas aplicaciones prácticas, existe cierto desconocimiento acerca de la máxima frecuencia de las señales de entrada, o, dicho de otro modo, la frecuencia máxima de la señal puede variar en un cierto rango. Por ello, consideraremos que las señales con las que el sistema de muestreo-interpolación debe trabajar tienen una densidad espectral de potencia plana (3.1) con una frecuencia de corte f_c modelada como una variable aleatoria, a la que denotaremos $\mathbf{\Phi}$, con una determinada función densidad de probabilidad, $f_{\mathbf{\Phi}}(\phi)$ con $\phi \in [0, \infty)$.

Dado el modelo de señal presentado anteriormente, la potencia de solapamiento espectral es, a su vez, una variable aleatoria, $\mathbf{P_a}$, que toma valores en función de la frecuencia de corte, $\mathbf{\Phi}$; es decir, $\mathbf{P_a}$ es una función de variable aleatoria:

$$\mathbf{P_a} = g\left(\mathbf{\Phi}\right), \tag{3.3}$$

donde, teniendo en cuenta (3.2),

$$g(\phi) = \begin{cases} 0, & 0 < \phi \leq \frac{f_s}{2}; \\ 1 - \frac{f_s}{2\phi}, & \phi > \frac{f_s}{2}. \end{cases}$$

La potencia media de solapamiento espectral vendrá dada por la esperanza matemática de (3.3), es decir,

$$P_a = \mathrm{E}\left[g\left(\mathbf{\Phi}\right)\right] = \int_0^{\infty} g(\phi) f_{\mathbf{\Phi}}(\phi)d\phi = \int_{f_s/2}^{\infty} \left(1 - \frac{f_s}{2\phi}\right) f_{\mathbf{\Phi}}(\phi)d\phi.$$

Finalmente se obtiene

$$P_a = 1 - F_{\mathbf{\Phi}}(f_s/2) - \frac{f_s}{2}\int_{f_s/2}^{\infty} \frac{f_{\mathbf{\Phi}}(\phi)}{\phi}d\phi. \tag{3.4}$$

El proceso de interpolación-reconstrucción mediante la función $\mathrm{sinc}(t)$ impone la limitación en ancho de banda de las señales a muestrear, lo que supone en el modelo de señal considerado $f_{\mathbf{\Phi}}(\phi) = 0$ para todo $\phi > f_s/2$, resultando una potencia de *aliasing* nula, según (3.4).

Aunque no es posible hablar de una densidad espectral de potencia representativa del conjunto de señales planteadas, pues no se cumple la condición de estacionariedad, sí que es posible estimar cómo se distribuye en media la potencia de señal con la frecuencia a partir de su dependencia con la frecuencia de corte, $\mathbf{\Phi}$. Para ello, partiendo de (3.1) y considerando una frecuencia de corte $f_c = \phi$:

$$S_x(f,\phi) = \begin{cases} \frac{1}{2\phi}, & |f| \leq \phi; \\ 0, & |f| > \phi. \end{cases} \tag{3.5}$$

Esta expresión nos indica que la DEP a una frecuencia determinada, f, es una variable aleatoria, o, más concretamente, una función de la variable aleatoria $\mathbf{\Phi}$; por lo tanto, es posible obtener su valor medio:

$$S_x(f) = \mathrm{E}\left[S_x(f,\mathbf{\Phi})\right] = \int_{-\infty}^{\infty} S_x(f,\phi) f_{\mathbf{\Phi}}(\phi) d\phi = \int_{|f|}^{\infty} \frac{f_{\mathbf{\Phi}}(\phi)}{2\phi} d\phi. \tag{3.6}$$

A partir de esta expresión es posible calcular la potencia de aliasing de manera alternativa a (3.4) sin más que evaluar la potencia por encima de la mitad de la frecuencia de muestreo, es decir,

$$P_a = 2\int_{f_s/2}^{\infty} S_x(f) df = \int_{f_s/2}^{\infty} \left(\int_{|f|}^{\infty} \frac{f_{\mathbf{\Phi}}(\phi)}{\phi} d\phi \right) df, \tag{3.7}$$

resultado que, obviamente, coincide con el obtenido mediante (3.4).

3.2.1.2. Ejemplos de señales blancas con frecuencia máxima variable

A. **Uniforme**

La frecuencia máxima de la señal, ϕ, se modela mediante una variable aleatoria, $\mathbf{\Phi}$, uniformemente distribuida en $[0, \phi_{\max}]$. En este caso, la función densidad de probabilidad y la función distribución toman la forma

$$f_{\mathbf{\Phi}}(\phi) = \begin{cases} \frac{1}{\phi_{\max}}, & 0 \leq \phi \leq \phi_{\max}; \\ 0, & \text{en otro caso}; \end{cases} \tag{3.8}$$

$$F_{\mathbf{\Phi}}(\phi) = \begin{cases} 0, & \phi < 0; \\ \frac{\phi}{\phi_{\max}}, & 0 \leq \phi \leq \phi_{max}; \\ 1, & \phi > \phi_{\max}. \end{cases} \tag{3.9}$$

Utilizando (3.4) y las expresiones precedentes de la función densidad de probabilidad y función distribución uniformes se obtiene la potencia de *aliasing*

$$P_a = \begin{cases} 1 - \frac{f_s}{2\phi_{\max}}\left(1 - \ln \frac{f_s}{2\phi_{\max}}\right), & \phi_{\max} > f_s/2; \\ 0, & \phi_{max} \leq f_s/2. \end{cases} \tag{3.10}$$

Considerando que la máxima frecuencia de la señal de entrada es $\phi_{\max}$ y denotando por R al mínimo factor de sobremuestreo

$$R = \frac{f_s}{2\phi_{\max}},$$

entonces la expresión de la potencia de *aliasing* puede escribirse en función de R como

$$P_a = \begin{cases} 1 - R\left(1 - \ln R\right), & R < 1; \\ 0, & R \geq 1. \end{cases}$$

Utilizando (3.6) y (3.8) se obtiene el espectro promedio

$$S_x(f) = \frac{1}{2\phi_{\max}} \ln \frac{\phi_{\max}}{f}, \quad f \leq \phi_{\max};$$

o bien

$$S_x(f) = -R \ln\left(2Rf\right), \quad f \leq \frac{1}{2R}.$$

B. **Exponencial**

Si se considera que la frecuencia máxima de la señal, ϕ, sigue una distribución exponencial de parámetro c, entonces

$$f_{\mathbf{\Phi}}(\phi) = \frac{1}{c} e^{-\frac{\phi}{c}}, \qquad \phi \geq 0; \tag{3.11}$$

$$F_{\mathbf{\Phi}}(\phi) = 1 - e^{-\frac{\phi}{c}}, \quad \phi \geq 0. \tag{3.12}$$

por lo que, utilizando (3.4), la potencia de aliasing resulta

$$P_a = e^{-\frac{f_s}{2c}} - \frac{f_s}{2c} \mathrm{E}_i\left(\frac{f_s}{2c}\right), \tag{3.13}$$

donde E_i representa la función integral exponencial

$$\mathrm{E}_i(x) = \int_x^\infty \frac{e^{-t}}{t} dt.$$

Teniendo en cuenta que la media de la variable aleatoria exponencial es c y denotando por R al valor medio de sobremuestreo,

$$R = \frac{f_s}{2c},$$

entonces la expresión de la potencia de *aliasing* queda

$$P_a = e^{-R} - R\mathrm{E}_i(R).$$

Por otra parte, el espectro promedio se obtiene utilizando (3.6) y (3.11)

$$S_x(f) = \frac{1}{2c}\mathrm{E}_i\left(\frac{f}{c}\right) = R\mathrm{E}_i(2Rf).$$

3.2.1.3. Señal de Voz

En varios de los ejemplos que se presentan a continuación se emplea la señal de voz. En aplicaciones tales como la telefonía, esta señal es muestreada con 8 bits de resolución a 8000 muestras por segundo. La voz es una señal no estacionaria (vocales, consonantes y silencios se suceden en el tiempo con características estadísticas totalmente distintas), por lo que hablar de densidad espectral de potencia carece de significado. Sin embargo, al promediar la densidad espectral de potencia de multitud de señales de voz de distintos y variados locutores se obtiene un espectro medio (en la banda $[0, 11025]$ Hz) como el mostrado en la Figura 3.1.a. Por otro lado, cuando se trabaja con este tipo de señales es necesario tener en cuenta la forma en la que el oído humano percibe los sonidos, resultando que son percibidos de forma muy distinta en función de su frecuencia. Para tener en cuenta estos efectos se utiliza el llamado filtro de ponderación sofométrica, que reproduce la sensibilidad del oído, cuya respuesta en frecuencia está definida en la recomendación ITU-T G223 y se representa en la Figura 3.1.b.

Como se puede observar, la voz es fundamentalmente paso bajo, aunque contiene potencia no despreciable a frecuencias medias y altas, es decir, no es adecuado considerar esta señal como estrictamente limitada en banda, lo que implica que, si no se utiliza un filtro antisolapamiento, se producirá un cierto solapamiento espectral.

3.2.2. Simulaciones

Las simulaciones llevadas a cabo para comparar las prestaciones de la interpolación mediante la función sinc y la interpolación spline cúbica siguen el esquema de la Figura 3.2, en el cual se dispone de M muestras por segundo de la señal a interpolar, $x[m]$, siendo M un valor elevado ($M > 100$). Dicha

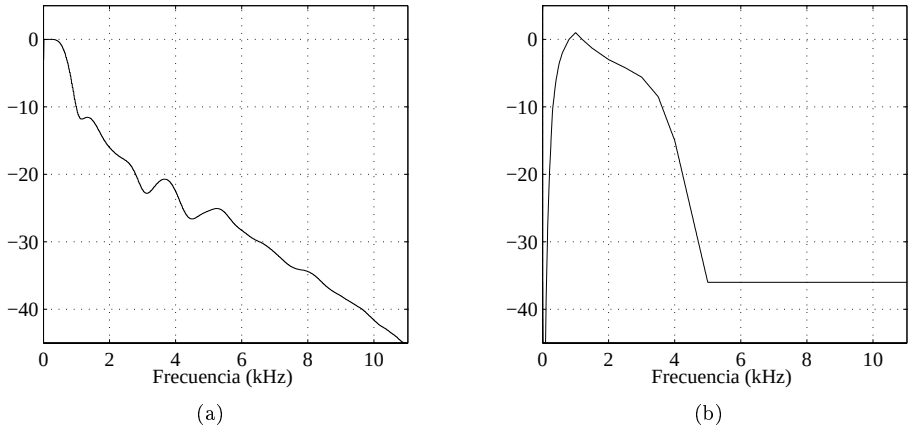


Figura 3.1: La señal de voz: a) espectro promedio, b) filtro de ponderación sofométrica.

señal es filtrada por un filtro paso bajo ideal con frecuencia de corte $\omega_c = \pi/M$ (implementado mediante técnicas FFT-IFFT) y posteriormente diezmada por un factor M, obteniéndose así una secuencia, $x_s[n]$, muestreada a $f_s = 1$. La secuencia original $x[m]$, muestreada a $f_s = M$, servirá de referencia a la hora de comparar las distintas técnicas interpolatorias aplicadas sobre $x_s[n]$.

El proceso de cuantificación se modela mediante un ruido aditivo uniforme y blanco, $n_q[n]$, de media nula y varianza equivalente a la provocada por un cuantificador con b bits de resolución. Para el caso de señales cuya amplitud siga una distribución Gaussiana de media nula y potencia σ_x^2, el ruido de cuantificación poseerá una potencia [1] [Jayant84]

$$\sigma_n^2 = \sigma_x^2 10^{\frac{7-6b}{10}}, \tag{3.14}$$

mientras que cuando se trabaje con señales con distribución Laplaciana (utilizada habitualmente para modelar la señal de voz), la varianza del ruido de cuantificación será[2]

$$\sigma_n^2 = \sigma_x^2 10^{\frac{12-6b}{10}}. \tag{3.15}$$

Finalmente, la interpolación sinc se realiza mediante un proceso de FFT, relleno con ceros e IFFT, mientras que la interpolación spline cúbica se lleva a cabo considerando condiciones de contorno que impliquen periodicidad, pues tanto la señal de referencia, $x[m]$, como la señal interpolada, $x_{\text{sinc}}[m]$, son periódicas.

[1] Se considera una probabilidad de sobrecarga de $P(|\mathbf{X}| > x_{sc}) = 5 \cdot 10^{-5}$, lo que implica, para el caso Gaussiano, que el valor de sobrecarga del cuantificador es $x_{\text{sc}} = 4\sigma_x$.

[2] Para una señal Laplaciana con $P(|\mathbf{X}| > x_{\text{sc}}) = 5 \cdot 10^{-5}$ resulta $x_{\text{sc}} = 7\sigma_x$.

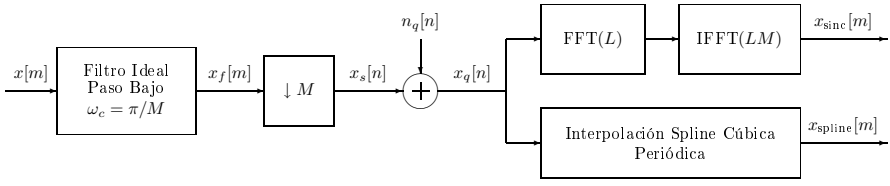


Figura 3.2: Esquema de la interpolación de señales estrictamente limitadas en banda.

3.2.2.1. Resultados con señales blancas estrictamente limitadas en banda

En primer lugar, se considera que la señal a interpolar, $x[m]$, es una secuencia blanca estrictamente limitada en banda y con un cierto factor de sobremuestreo, R, que representa la relación entre la máxima frecuencia presente en la señal, f_c, y la frecuencia de Nyquist, $f_s/2$; es decir,

$$R = \frac{f_s}{2f_c}. \tag{3.16}$$

Este tipo de señales se generan mediante el filtrado ideal paso bajo (con frecuencia de corte $f_c = \frac{f_s}{2R}$) de secuencias Gaussianas y blancas. Posteriormente, la señal $x[m]$ atraviesa el filtro antisolapamiento ideal (que en este caso no tendrá efecto alguno siempre que $R \geq 1$), se muestrea y se cuantifica, obteniéndose muestras ruidosas de la señal original, $x_q[n]$, que serán interpoladas mediante los dos métodos considerados (sinc y spline) con el fin de obtener el error cometido con cada uno de ellos para distintos valores de sobremuestreo.

En las simulaciones realizadas se han empleado $N = 100$ secuencias de LM muestras (con $L = 1000$ y $M = 100$) para cada valor de sobremuestreo, $R \in [1, 32]$ y para cada resolución considerada, $b \in [2, 16]$. En Figura 3.3 se representa la curva en el plano $R-b$ donde las prestaciones de la interpolación sinc y la interpolación spline son idénticas en términos de error cuadrático medio. Puntos del plano $R-b$ a la izquierda de la curva presentan mejores prestaciones para la interpolación spline. A modo de ejemplo, para una resolución típica de $b = 8$ bits, la interpolación spline obtiene una solución más precisa siempre que el factor de sobremuestreo sea superior a $R = 2{,}2$; hecho éste bastante habitual en la práctica, dada la no idealidad de los filtros antisolapamiento.

Es necesario comentar que, en la región $R - b$ donde la interpolación spline supera a la interpolación sinc, la mejora no es muy sustancial (entre 0,5 y 1 dB), mientras que las prestaciones de la interpolación spline se degradan considerablemente en cuanto consideramos puntos $R - b$ relativamente alejados de la curva y por debajo de la misma. Otro aspecto muy importante a la hora de comparar ambas técnicas de reconstrucción es el coste computacional de cada

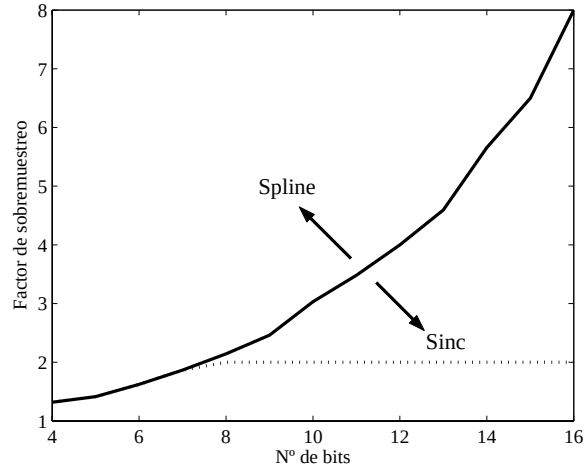


Figura 3.3: Interpolación de señales estrictamente limitadas en banda. Valores para los que se igualan las prestaciones de la interpolación spline y sinc (línea continua); y spline y sinc truncada (línea discontinua).

una de las dos interpolaciones, siendo éste un factor decisivo en muchos casos prácticos. En este sentido, la interpolación spline requiere del orden de 9 multiplicaciones y sumas por cada muestra, mientras que una interpolación sinc precisa de la realización de FFT's de longitud L e IFFT's de longitud LM, lo que, además de requerir una gran cantidad de operaciones, resulta impracticable en sistemas en tiempo real en los que el retardo admisible sea pequeño, al tratarse de un algoritmo bloque. En la práctica se recurre a implementar la interpolación sinc mediante filtros FIR cuya respuesta impulsiva es la función sinc truncada (o una aproximación de la misma), lo cual disminuye sus prestaciones incluso para filtros de longitudes considerables. En la Figura 3.3 se representa con línea discontinua los valores de sobremuestreo y número de bits para los cuales se obtienen idénticas prestaciones al interpolar mediante splines cúbicos y mediante un filtro con forma de sinc truncada en el que se consideran 80 lóbulos. Se observa cómo, en este caso, para valores de sobremuestreo superiores a $R = 2$ la interpolación spline proporciona resultados iguales o superiores, con un coste computacional seis veces inferior.

3.2.2.2. Resultados con señales blancas con frecuencia máxima variable

Utilizando el modelo de señal presentado en el Apartado 3.2.1.1 (señales blancas con frecuencia máxima siguiendo una distribución exponencial), las si-

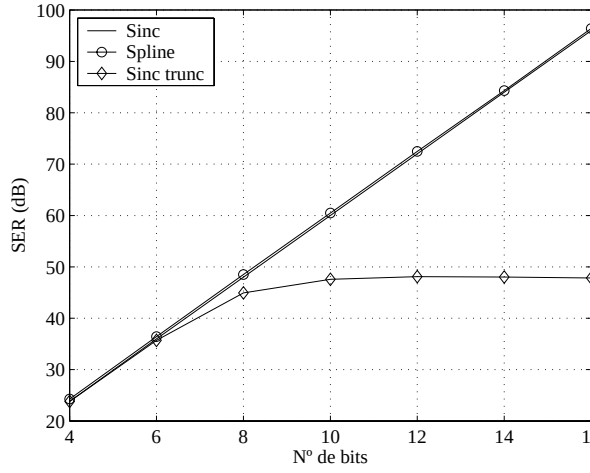


Figura 3.4: Interpolación sinc y spline de señales blancas con frecuencia máxima distribuida exponencialmente.

mulaciones siguen el esquema de la Figura 3.2. En esta ocasión, a la hora de generar las señales a interpolar, $x[m]$, se utilizan secuencias blancas Gaussianas filtradas con un filtro paso bajo ideal con frecuencia de corte dada por la distribución correspondiente, es decir, por la variable aleatoria $\mathbf{\Phi}$. El parámetro de la función densidad de probabilidad (el parámetro c en la variable aleatoria exponencial) se escoge de tal manera que la potencia de solapamiento dada por la fórmula correspondiente ((3.13) en este caso) sea inferior a la potencia del ruido de cuantificación generado por un cuantificador de b bits (3.14).

En la Figura 3.4 se muestra la relación señal a error (SER: *Signal to Error Ratio*) de la interpolación sinc y spline de señales cuya frecuencia de corte sigue una distribución exponencial para distintos bits de resolución. Se observa cómo la interpolación spline (mucho menos costosa computacionalmente hablando) obtiene resultados idénticos que la interpolación sinc ideal. Así mismo, se comprueba que, para resoluciones superiores a 8 bits, las prestaciones de la interpolación spline son mucho mejores que las de la interpolación mediante un filtro sinc truncado a 40 lóbulos laterales, que implica un coste computacional 4 veces superior a la interpolación spline.

En la Figura 3.5 se compara la SER obtenida por la interpolación sinc, la interpolación spline y la interpolación sinc truncada cuando las señales a interpolar son blancas estrictamente limitadas en banda cuya frecuencia máxima, $\mathbf{\Phi}$, sigue una distribución uniforme en $[0, f_s/2]$. Se puede observar que el rendimiento de la interpolación spline se degrada de forma leve y progresiva respecto a la interpolación sinc ideal conforme aumenta el número de bits.

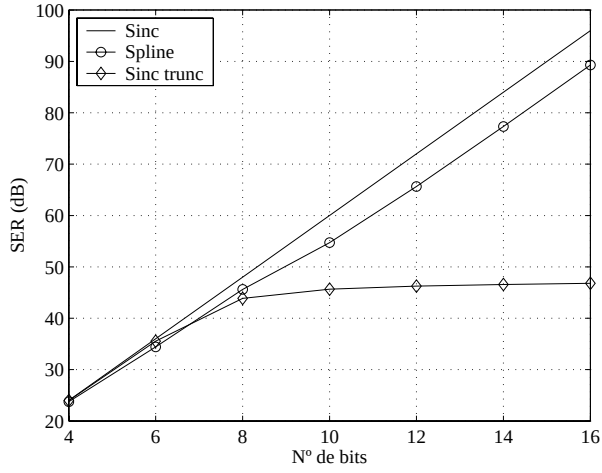


Figura 3.5: Interpolación sinc y spline de señales blancas con frecuencia máxima distribuida uniformemente.

Dicha degradación se sitúa en torno a los 2 dB para 8 bits y en torno a 5 dB para resoluciones de 16 bits. Sin embargo, al interpolar este tipo de señales con filtros sinc truncados las prestaciones obtenidas rápidamente alcanzan un valor de saturación, que en el caso presentado en la Figura 3.5 se sitúa en torno a 47 dB.

3.2.2.3. Voz

En este apartado se realiza la interpolación de señales con densidad espectral de potencia como la mostrada en Figura 3.1.a muestreadas a $f_s = 8000$ muestras por segundo y con b bits de resolución. Con el fin de comparar sus prestaciones se emplean tres métodos distintos: interpolación sinc ideal mediante FFT-relleno con ceros-IFFT; interpolación spline cúbica; e interpolación mediante filtro FIR con respuesta al impulso truncada a 20 lóbulos. En la Figura 3.6 se presenta la relación señal a error obtenida por las tres técnicas para distintos bits de cuantificación. Así mismo se puede observar la SER cuando se pondera el error de muestreo-interpolación mediante el filtro sofométrico de la Figura 3.1.b.

Como se puede observar, la interpolación spline cúbica de señales de voz alcanza un valor máximo de SER de 27 dB. Sin embargo, al aplicar el filtro perceptual, una parte importante del ruido de interpolación es eliminada, pues éste, para el spline cúbico, es más importante en las altas frecuencias, allí donde el oído humano es menos sensible. De esta forma, si consideramos el filtro sofométrico, la interpolación spline cúbica ofrece las mismas prestaciones que

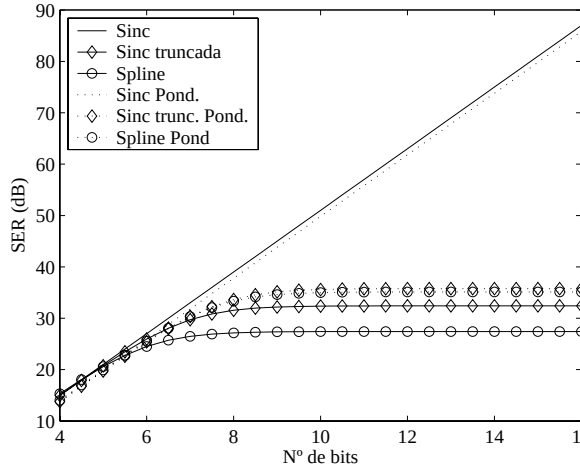


Figura 3.6: Interpolación de la señal de voz. Filtro antisolapamiento ideal.

la interpolación sinc truncada a 20 lóbulos (que tiene un coste computacional doble). Cabe resaltar que, para una resolución de 8 bits (típica en aplicaciones telefónicas), el rendimiento de la interpolación spline cúbica es únicamente 4 dB inferior a la interpolación sinc ideal sin truncar.

En las simulaciones anteriores se ha considerado un filtro antisolapamiento paso bajo ideal con frecuencia de corte $f_c = f_s/2$. Este tipo de filtro es el óptimo para la interpolación basada en sincs, pero, evidentemente, es irrealizable desde el punto de vista práctico. Un filtro real utilizado en sistemas comerciales de adquisición de voz (por ejemplo, en tarjetas de sonido de ordenadores personales) consiste en un filtro de orden ocho, frecuencia de corte $f_c = 0{,}4 f_s$ y respuesta en frecuencia

$$|H_a(f)| = \frac{1}{1 + (f/f_c)^{16}}.$$

Utilizando este filtro antisolapamiento, las prestaciones de los tres métodos de interpolación son las mostradas en la Figura 3.7. Puede observarse que la interpolación sinc ideal alcanza una relación señal a error máxima de 60 dB motivado por el *aliasing* residual debido al filtro no ideal. Dado que dicho *aliasing* afecta en mayor medida a las frecuencias próximas a $f_s/2 = 4000$ Hz, el filtro sofométrico reduce su efecto, alcanzándose un nivel máximo de 70 dB. En cuanto a la interpolación spline cúbica, los resultados obtenidos son sensiblemente mejores que los correspondientes al filtro ideal, pues para todas las resoluciones consideradas su rendimiento supera al de la interpolación sinc de 20 lóbulos. En concreto, para 8 bits de resolución la interpolación spline obtiene

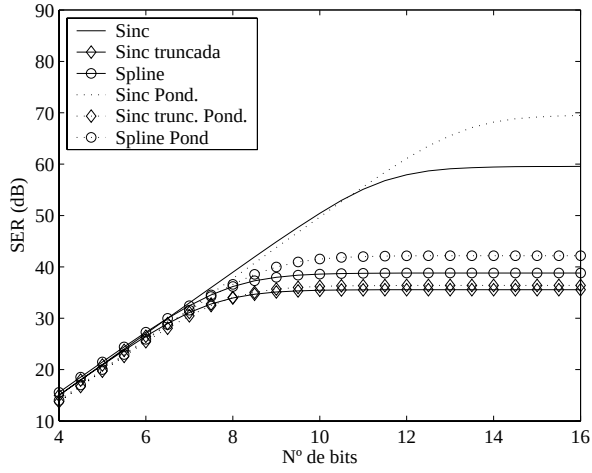


Figura 3.7: Interpolación de la señal de voz. Filtro antisolapamiento de orden ocho.

la misma SER que la interpolación sinc ideal y supera en 2,5 dB a la obtenida con la interpolación sinc truncada.

3.2.3. Discusión de los resultados

A la vista de los resultados de las simulaciones previas, es evidente que, bajo determinadas circunstancias, que se presentan en la práctica de forma muy habitual (en cuanto a factores de sobremuesreo o número de bits de cuantificación considerados), la interpolación spline cúbica proporciona resultados superiores a la interpolación basada en la función sinc. Si la capacidad de cómputo fuera un bien escaso (y lo es en gran número de aplicaciones) las prestaciones de la interpolación sinc se degradarían gravemente al truncar el filtro interpolador. Sin embargo, dada la eficiencia computacional de la interpolación spline, ésta mantendría sus prestaciones, al requerir únicamente cinco multiplicaciones-acumulaciones (MAC) por muestra para obtener los coeficientes spline y cuatro MAC's para obtener cada valor interpolado. En definitiva, ante un compromiso entre prestaciones y coste computacional, la interpolación spline proporciona en muchas ocasiones mejores resultados que la interpolación derivada del Teorema de Muestreo.

En las simulaciones presentadas se ha llevado a cabo la interpolación spline a partir de las muestras de la señal filtrada paso bajo, que no son más que la proyección de la señal original en el espacio generado por la función base sinc. Si en su lugar se utilizaran las muestras resultantes de filtrar la señal original con

el prefiltro spline adecuado (o, al menos, con un filtro realizable que aproxime su respuesta) los resultados obtenidos serían superiores, ya que en ese caso se obtendría la aproximación spline de menor error cuadrático de la señal original.

De hecho, como se discutió en el Capítulo 2, dado un espacio de señal definido por un conjunto de funciones base generadoras (la función sinc, la base spline cúbica o cualquier otra) la solución óptima (con el criterio del menor error cuadrático medio) al problema de muestreo y reconstrucción consiste, en primer lugar, en la proyección ortogonal de la señal en el espacio de representación considerado [Unser92, Unser94, Unser00]. Esto implica la utilización de un filtro adecuado a cada base, por ejemplo, el filtro paso bajo ideal para el caso de la interpolación sinc.

3.3. Muestreo simultáneo mediante filtros de Lagrange

Retardar una señal una fracción del periodo de muestreo es una necesidad habitual en muchas aplicaciones del procesado de señal, entre las que se encuentran: remuestreo con factores irracionales [Ramstad84], sincronización en modems digitales [Gardner93, Erup93], estimación del periodo fundamental de la voz (*pitch*) [Kroon91], conformación de haz [Murphy97], modelado en tiempo discreto de tubos acústicos [Välimäki95], o sistemas de adquisición multicanal [Barwicz90].

En particular, un sistema de adquisición de múltiples canales de entrada de bajo coste típicamente está formado por un multiplexor analógico seguido por un único dispositivo de muestreo y retención (*sample-and-hold*) y un conversor analógico-digital (ADC). Esta arquitectura introduce un retardo fraccionario entre los distintos canales que deberá ser corregido en aquellas aplicaciones que requieran el muestreo simultáneo de las señales de entrada.

Existe una solución de tipo hardware obvia, a la vez que costosa, basada en la utilización de un *sample-and-hold* por cada canal. En contraposición, es posible aplicar técnicas de procesado digital de señal que, manteniendo la simplicidad y bajo coste del hardware, permiten obtener un muestreo simultáneo efectivo de las distintas señales, a costa, evidentemente, de una mayor complejidad software.

En este apartado nuevamente mostramos que una técnica de diseño de filtros de retardo fraccionario basada en interpolación polinómica (interpolación de Lagrange) permite obtener mejores resultados, en términos de prestaciones y de coste computacional, que técnicas de diseño basadas en interpolación de señales limitadas en banda [Luengo97a, Luengo97b, Luengo00]. Dichas técnicas polinómicas han sido aplicadas con éxito en sistemas de monitorización en tiempo real de turbo-generadores hidroeléctricos con el fin de obte-

ner un muestreo simultáneo efectivo de señales adquiridas no simultáneamente [Pantaleón95a, Ibáñez98, Cobo98, López-Higuera98, López-Higuera99].

3.3.1. Compensación de retardo fraccionario

El problema planteado puede ser descrito de la siguiente manera: se dispone de una señal, $x(t)$, muestreada uniformemente a una frecuencia $f_s = 1/T_s$, es decir, $x[n] = x(nT_s)$. El objetivo consiste en calcular el valor de dicha señal en puntos equiespaciados distintos a los instantes de muestreo o, de forma equivalente, obtener la secuencia $x_i[n-D] = x(nT_s - D)$, donde $D \in \mathbb{R}$. La solución, en forma de filtrado digital, se puede expresar como

$$y[n] = x_i[n-D] = h[n] * x[n] = \sum_{k=-\infty}^{+\infty} h[k]x[n-k]. \tag{3.17}$$

La solución exacta a este problema es un filtro ideal paso todo que introduzca un retardo D, es decir, un filtro con respuesta en frecuencia

$$H\left(e^{j\omega}\right) = e^{-j\omega D}, \ |\omega| < \pi;$$

y cuya respuesta al impulso es una versión retardada y muestreada del núcleo interpolador definido por el Teorema de Muestreo

$$h_{\text{ideal}}[n] = \text{sinc}\,[n-D]. \tag{3.18}$$

Es posible, por tanto, obtener la señal retardada como

$$x_i[n-D] = h_{\text{ideal}}[n] * x[n] = \sum_{k=-\infty}^{+\infty} \text{sinc}\,[k-D]x[n-k]. \tag{3.19}$$

Resulta evidente la dificultad práctica de este procedimiento, pues el filtro (3.18) es anticausal y de longitud infinita. Es claro que el Teorema de Muestreo nos aporta la solución ideal, pero, en esta ocasión, su aplicación directa es imposible, por lo que resulta necesario aproximar el filtro ideal. Este sistema admite una implementación muy eficiente utilizando estructuras polifásicas [Crochiere83], aunque presenta el inconveniente de admitir únicamente retardos racionales($D = N/M$) pudiendo requerir cargas computacionales prohibitivas cuando los factores de interpolación y diezmado resulten elevados.

Una alternativa consiste en utilizar otro tipo de filtros de retardo fraccional (FDF's: *Fractional Delay Filters*), diseñados de acuerdo a otro criterio o empleando otra técnica de interpolación [Laakso96]. Por otra parte, en sistemas de adquisición reales, las señales de entrada habitualmente se encuentran sobremuestreadas, con el fin de relajar las especificaciones y facilitar el diseño de los filtros antisolapamiento. Por tanto, únicamente una pequeña porción

del espectro, la zona de bajas frecuencias, contiene información útil. En este caso los interpoladores de Lagrange se convierten en la solución más atractiva, debido a su respuesta frecuencial máximamente plana y a su capacidad de aproximar el filtro paso todo ideal en bajas frecuencias utilizando filtros con pocos coeficientes [Laakso96].

3.3.2. Filtro de Lagrange de retardo fraccionario

En las ecuaciones (2.7) y (2.8) se presentó la interpolación de Lagrange, $P_n(x)$, de una función $f(x)$ en $n+1$ puntos distintos. Si en dichas ecuaciones consideramos la variable independiente t en lugar de x, y denotamos la señal a interpolar como $x(t)$ en lugar de $f(x)$ y la señal interpolada como $x_i(t)$ en lugar de $P_n(x)$ obtendremos una notación más convencional desde el punto de vista del procesado de señal, pudiendo reescribir la señal interpolada como

$$x_i(t) = \sum_{k=0}^{N} x(t_k) L_{N,k}(t),$$

donde el k-ésimo polinomio de Lagrange de orden N es

$$L_{N,k}(t) = \prod_{\substack{i=0 \\ i \neq k}}^{N} \frac{t - t_i}{t_k - t_i}, \quad k = 0, 1, \ldots, N.$$

Considerando que la señal $x(t)$ ha sido muestreada uniformemente, el valor de $x(nT_s - D)$ puede obtenerse mediante interpolación de Lagrange utilizando $N+1$ muestras. Escribiendo las expresiones en forma de filtrado digital (3.17) en el que interviene un filtro FIR, obtenemos

$$y[n] = x_i[n - D] = \sum_{k=0}^{N} h_L[k] x[n-k], \tag{3.20}$$

donde $h_L[n]$ representa los coeficientes del filtro de retardo fraccionario de Lagrange de orden N

$$h_L[n] = \prod_{\substack{i=0 \\ i \neq k}}^{N} \frac{D - i}{k - i} = (-1)^{N-n} \binom{D}{n} \binom{D - n - 1}{N - n}. \tag{3.21}$$

En la Tabla 3.1 se muestran los coeficientes de (3.21) para $N = 0$, 1, 2 y 3. Como se puede observar, el filtro de retardo fraccionario de Lagrange de orden cero se corresponde con un esquema interpolatorio en el que se asocia a un punto el valor de la muestra más cercana. Igualmente se comprueba que el filtro

de orden uno equivale a la interpolación lineal. Además, como se demuestra en el Apéndice A.1, el resultado obtenido mediante interpolación de Lagrange coincide con el que proporciona la técnica de diseño siguiendo el criterio de respuesta en frecuencia máximamente plana [Oetken79].

3.3.3. Análisis del error de interpolación

El filtro interpolador de Lagrange de retardo fraccionario proporciona una solución exacta en $\omega = 0$, pero su comportamiento se va degradando conforme aumenta la frecuencia de la señal a retardar. En este apartado se presenta una cota del error de interpolación en función del orden del filtro, del factor de sobremuestreo de la señal y del retardo fraccionario [Luengo00]. Dicha cota podrá ser utilizada para asegurar que el filtro diseñado mantenga una precisión determinada.

3.3.3.1. Planteamiento del problema

Se considera una señal continua y analógica de entrada $x(t)$ limitada en banda en el rango $[0, f_0]$ Hz (periodo mínimo $T_0 = 1/f_0$) y muestreada a $f_S > 2f_0$ Hz (factor de sobremuestreo $R = f_S/2f_0 = T_0/2T_S$) mediante un conversor analógico-digital con una cierta resolución. Dicha señal de entrada debe ser retrasada un cierto retardo D compuesto, en general, por un número entero de muestras y un retardo fraccionario. Este problema, como se ha comentado en secciones anteriores, puede ser resuelto mediante un filtro interpolador de Lagrange de orden N ($N + 1$ coeficientes) que proporcionará un retardo total

$$D = \frac{N}{2} + d,$$

donde $N/2$ es el retraso de un filtro FIR causal, que proporciona un retardo fraccional en el caso de ser N impar, y d es el retardo fraccional adicional. Este retardo d tomará valores en el rango $[-0{,}5,\ 0{,}5]$ con el fin de obtener un región de interpolación en torno al centro de simetría del filtro ($N/2$). Al restringir el rango de interpolación a dicha región (entre los dos coeficientes centrales para N impar y alrededor del coeficiente central para N par) se obtienen los mejores resultados, tal y como se demuestra en [Laakso96].

N	**1**	**2**	**3**	**4**
$\mathbf{h_L[0]}$	1	$1-D$	$(D-1)(D-2)/2$	$-(D-1)(D-2)(D-3)/6$
$\mathbf{h_L[1]}$		D	$D(D-2)$	$D(D-2)(D-3)/2$
$\mathbf{h_L[2]}$			$(D-1)(D-2)/2$	$-D(D-1)(D-3)/2$
$\mathbf{h_L[3]}$				$D(D-1)(D-2)/6$

Tabla 3.1: Coeficientes del filtro de retardo fraccionario de Lagrange.

El objetivo consiste en obtener expresiones que relacionen el error de interpolación con el factor de sobremuestreo, R, y con el orden del filtro, N. De esta manera, si el error de interpolación no excede el nivel máximo del ruido de cuantificación impuesto por el ADC, entonces se preservará la resolución del conversor. Será posible, por tanto, asegurar que la señal ha sido retardada sin pérdida de resolución.

3.3.3.2. Cota superior del error

Con el fin de garantizar que, dado un cierto factor de sobremuestreo y un determinado orden del filtro, se mantenga la resolución del ADC, es necesario encontrar las condiciones bajo las cuales se produce el mayor error de interpolación. De esta manera será posible obtener una cota superior del error y desarrollar ecuaciones útiles a la hora de diseñar el filtro de retardo fraccionario.

El error de interpolación viene dado por la diferencia entre la señal interpolada usando el filtro interpolador ideal (3.19), denotada por x_i, y la señal interpolada usando el filtro de Lagrange de retardo fraccional (3.20), x_L,

$$e[n] = x_i[n-D] - x_L[n-D] = x[n] * (h_i[n] - h_L[n]),$$

donde $h_i[n]$ es el filtro retardador paso bajo ideal (3.18) y $h_L[n]$ el filtro de retardo fraccionario de Lagrange (3.21).

El error de interpolación depende de dos factores: la señal de entrada y la diferencia entre el filtro ideal y el utilizado, siendo este último factor dependiente del retardo fraccional y de la frecuencia (o, de forma equivalente, del factor de sobremuestreo). Dado que la respuesta del filtro interpolador se aparta de la ideal de forma monótona conforme aumenta la frecuencia, como se demuestra en el Apéndice A.1, el error máximo tiene lugar cuando la señal de entrada es una sinusoide cuya frecuencia es la máxima permitida [Ramstad84]

$$x[n] = A\cos(\omega_0 n).$$

El caso peor, en lo que respecta al retardo fraccional, se produce cuando la distancia entre el punto de interpolación y la muestra más cercana sea máxima [Laakso96], es decir, cuando $d = 0$ para N impar y $d = 0{,}5$ para N par. Sin embargo, y por simplicidad, en las siguientes expresiones se mantendrá D como parámetro.

En estas condiciones, el error de interpolación, e_I, se puede escribir como

$$e_I[n] = A\left(\cos[(n-D)\omega_0] - \sum_{k=0}^{N} h_L[k]\cos[(n-k)\omega_0]\right).$$

Al tratarse de un sistema lineal e invariante en el tiempo, el error, obviamente, podrá expresarse como un señal sinusoidal de la misma frecuencia que

la entrada y con una cierta amplitud y fase:

$$e_I[n] = A_{e_I} \cos(\omega_0 n + \phi_e).$$

En esta expresión la fase es irrelevante para el propósito de obtener una cota superior del error. Esta cota puede obtenerse de forma directa tomando el valor absoluto de la fórmula del residuo de Lagrange (2.9)

$$|e_I[n]| = \frac{\left|x^{(N+1)}(\xi(t))\right|}{(N+1)!} \prod_{k=0}^{N} |t - t_k|, \tag{3.22}$$

donde $\xi(t)$ es un punto sin especificar perteneciente al intervalo de interpolación, en este caso $|t - t_k| = |D - k|T_s$. Teniendo en cuenta que el valor absoluto de la $(N+1)$-ésima derivada de una sinusoide siempre es menor o igual a $(2\pi f_0)^{N+1}$, es posible obtener una cota superior a (3.22) de la forma

$$|e_I[n]| \leq \frac{A}{(N+1)!} \left(\frac{2\pi f_0}{f_s}\right)^{N+1} \prod_{k=0}^{N} |D - k|.$$

Reordenando términos, se puede calcular una cota superior de la amplitud de la señal de error

$$A_{e_I} \leq A \left(\frac{\pi}{RC_N(D)}\right)^{N+1}, \tag{3.23}$$

donde $C_N(D)$ es una función que no depende del factor de sobremuestreo, R, y cuyo valor viene dado por

$$C_N(D) = \left(\frac{(N+1)!}{\prod_{k=0}^{N} |D - k|}\right)^{\frac{1}{N+1}}.$$

En el siguiente apartado se relaciona la amplitud del error con la resolución del conversor analógico-digital con el fin de obtener las ecuaciones de diseño que se buscan.

3.3.3.3. Ecuaciones de diseño

Se considera un ADC bipolar con $b + 1$ bits significativos y un rango de entrada de $[-x_m, x_m]$ o, en su caso, un ADC unipolar con b bits significativos y un rango $[0, x_m]$. Con el fin de evitar que la etapa de compensación de retardo fraccionario introduzca distorsión adicional, el error máximo de interpolación deberá ser menor que el error introducido por el cuantificador del ADC.

El error de cuantificación puede alcanzar un valor máximo de 1/2 bit menos significativo (LSB: *Least Significant Bit*), es decir, su valor está comprendido

en el rango $[-\Delta/2, \Delta/2]$, siendo Δ la distancia entre dos niveles contiguos de cuantificación y cuyo valor es

$$\Delta = \frac{2x_m}{2^{b+1}},$$

por lo que la amplitud del error de cuantificación, A_{e_Q}, está acotada por

$$A_{e_Q} \leq \frac{x_m}{2^{b+1}}. \tag{3.24}$$

El objetivo del proceso de interpolación es obtener una señal retardada de forma que la amplitud del error de interpolación sea menor que el error de cuantificación. Por tanto, considerando que la señal de entrada cubre el rango completo del ADC ($A = x_m$ para el caso bipolar y $A = x_m/2$ para el caso unipolar) y comparando las expresiones (3.23) y (3.24) se obtiene la relación

$$A_{e_I} \leq A\left(\frac{\pi}{RC_N(D)}\right)^{N+1} \leq A_{e_Q} \leq \frac{A}{2^{b+1}},$$

a partir de la cual es posible obtener, por ejemplo, el valor del factor de sobremuestreo necesario para alcanzar una precisión determinada usando un filtro de retardo fraccionario de Lagrange de cierto orden

$$R \geq \frac{\pi}{C_N(D)} 2^{\frac{b+1}{N+1}}. \tag{3.25}$$

De forma inversa, es posible obtener la resolución, $b+1$ bits, alcanzada utilizando cierto factor de sobremuestreo y cierto orden del filtro

$$b+1 \leq (N+1)\log_2\left(\frac{C_N(D)R}{\pi}\right), \tag{3.26}$$

es decir, esta expresión asegura que para unos valores concretos de D, R, y N la resolución que se obtendrá será mayor que la proporcionada por un ADC de $b+1$ bits de precisión.

La obtención de una expresión que proporcione el orden del filtro para un factor de sobremuestreo y una cierta resolución es más complicada debido a la dependencia de la función $C_N(D)$ con el orden del filtro N. Sin embargo, al dibujar dicha función frente a N para el caso de mayor error ($d = 0$ para N impar y $d = 0{,}5$ para N par, es decir, $D_{\text{peor}} = \lfloor N/2 \rfloor + 1/2$) se puede observar, Figura 3.8, que el máximo es $\sqrt{8}$ para $N = 1$ y que tiende a 2 conforme N crece.

Por tanto, y dado que el escenario típico de trabajo implica diseñar para el caso peor, estos dos valores pueden ser utilizados como cotas superior e inferior para estimar el valor de N

$$\frac{b+1}{\log_2(\sqrt{8}R/\pi)} \leq N+1 \leq \frac{b+1}{\log_2(2R/\pi)}. \tag{3.27}$$

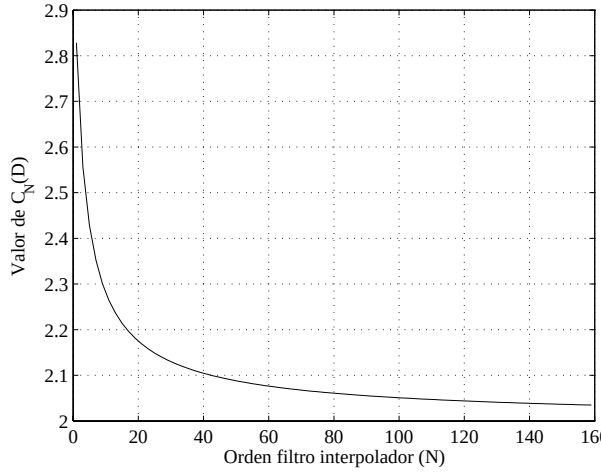


Figura 3.8: $C_N(D)$ en función del orden del filtro para un retardo fraccional de $d = 0{,}5$.

Estas cotas superior e inferior, que podrán consistir en números no enteros, pueden ser usadas para obtener iterativamente el orden necesario del filtro. El proceso consistirá en:

1. Inicializar N al mínimo valor entero mayor o igual que el término de la izquierda de la desigualdad (3.27).

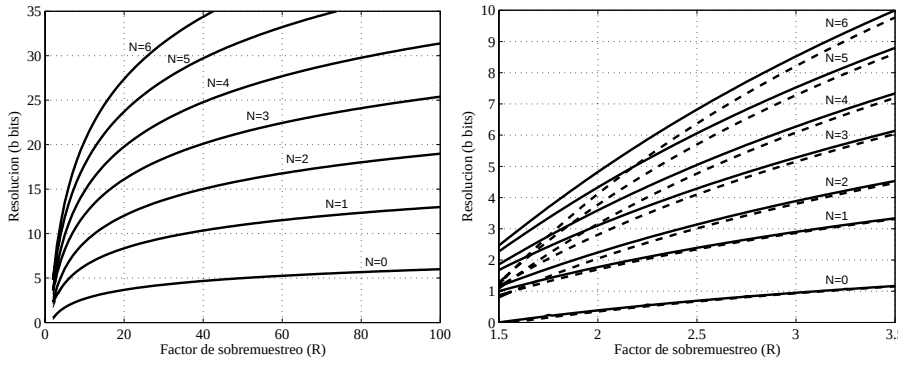
2. Calcular la resolución obtenida por medio de (3.26). Si se cumplen los requisitos de diseño se finaliza.

3. Hacer $N = N + 1$ y volver al paso 2.

Las ecuaciones (3.25)–(3.27) son válidas para el caso de conversores unipolares sin más que sustituir $(b + 1)$ por b.

3.3.4. Simulaciones

Con el fin de validar las expresiones previas, se ha realizado una serie de simulaciones en las cuales se ha generado una señal sinusoidal discreta $x[n]$ de fase aleatoria y amplitud $A = 1$ y factores de sobremuestreo en el rango de 1 a 100. Dicha señal ha sido filtrada utilizando filtros interpoladores de Lagrange de diversos órdenes (N entre 0 y 6) y considerando el retardo que introduce más error ($d = 0$ para los filtros de orden impar y $d = 0{,}5$ para los de orden par). Los resultados se han comparado con una réplica idealmente retardada $x[n - D]$ mostrándose en la Figura 3.9 los resultados de dicha simulación (en



(a) R entre $[1, 100]$. **(b)** Detalle.

Figura 3.9: Número de bits significativos para ciertos órdenes del filtro en función del factor de sobremuestreo. Simulación en línea sólida y teórica en línea discontinua.

forma de resolución efectiva: bits) así como la resolución teórica obtenida de (3.26).

La cota superior obtenida es tan precisa que ambos conjuntos de curvas (teóricas y simuladas) son indistinguibles para factores de sobremuestreo medios y altos (Figura 3.9a). Dichas curvas muestran que un interpolador lineal ($N = 1$) trabajando con un factor de sobremuestreo tan alto como $R = 100$ apenas permite mantener una resolución de 13 bits, mientras que la utilización de un filtro de sexto orden con R=10 posibilita una resolución de 20 bits.

En la Figura 3.9b se considera la aproximación para valores de factores de sobremuestreo bajos. Dado que las ecuaciones de diseño ofrecen una cota superior del error de interpolación, dicho error es sobreestimado y, por tanto, las curvas teóricas subestiman la verdadera resolución obtenida.

Por último, se compara el coste computacional del filtro de retardo fraccionario de Lagrange con un esquema multitasa (mostrado en la Figura 3.10 para un retardo $D = r/M$) e implementado mediante su descomposición polifase en M ramas. El filtro paso bajo ha sido diseñado usando la técnica de Parcks McClellan con una banda de paso de $[0, \pi/(MR)]$, una banda de rechazo de $[(2\pi - \pi/R)/M, 2\pi]$, un rizado en banda $r_p = 2^{-(b+1)}$ y un rechazo de $A = 6{,}02(b + 1) + 1{,}76$ dB. La Tabla 3.2 muestra el número de coeficientes del filtro interpolador de Lagrange necesarios para cinco resoluciones distintas y para seis valores de R y los compara con el número de coeficientes de la correspondiente rama del filtro polifase requerido para obtener el retardo. Los valores del filtro polifase se corresponden a $M = 16$, y se observa que el coste

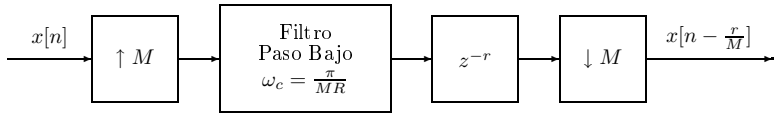


Figura 3.10: Esquema de la interpolación multitasa.

	Resolución ($b+1$ **bits**)				
R	8	12	16	24	32
2	17(6)	27(9)	38(12)	60(18)	82(24)
5	4(4)	7(6)	9(8)	14(12)	18(15)
10	3(3)	4(5)	6(7)	9(10)	12(13)
25	2(3)	3(5)	4(6)	6(10)	8(12)
50	2(3)	3(5)	4(6)	5(10)	7(12)
100	2(3)	2(5)	3(6)	4(9)	6(12)

Tabla 3.2: Comparación del número de coeficientes del FDF de Lagrange y del filtro polifase (entre paréntesis) requeridos para obtener una determinada resolución.

computacional de los filtros de retardo fraccionario de Lagrange es inferior a la estructura polifase cuando los factores de sobremuestreo son mayores que diez.

Así mismo, se ha comprobado que el número de coeficientes del filtro polifase se incrementa cuando, debido al valor deseado del retardo, $D = r/M$, se precisen valores superiores de M. Sin embargo, la longitud del filtro interpolador de Lagrange no depende del valor del retardo (las expresiones propuestas proporcionan la longitud mínima considerando el caso peor), por lo que, en estos casos, se incrementarán las ventajas computacionales de esta técnica.

3.3.5. Discusión de los resultados

El muestreo simultáneo de diversos canales adquiridos mediante un multiplexor y un único ADC, así como el retardo arbitrario de una señal pueden ser implementados utilizando filtros de Lagrange de retardo fraccionario. El rendimiento de este método es función del factor de sobremuestreo de las señales de entrada y del orden del filtro interpolador. En aquellas aplicaciones donde resulte habitual un factor de sobremuestreo elevado, los filtros de retardo fraccionario de Lagrange resultan ventajosos debido a su suavidad a bajas frecuencias y al bajo orden necesario para obtener una aproximación suficientemente buena al filtro ideal.

En esta sección se ha obtenido una cota del error de interpolación del fil-

tro de retardo fraccionario de Lagrange, mediante la cual se han desarrollado expresiones que relacionan el número de bits significativos con el factor de sobremuestreo, el orden del filtro y el retardo fraccionario. Se ha demostrado que, para un determinado factor de sobremuestreo, es posible alcanzar cualquier resolución deseada incrementando el orden del filtro; o, de forma inversa, para un determinado orden del filtro, incrementando el factor de sobremuestreo.

Los resultados obtenidos han sido validados mediante simulaciones, comprobándose que la cota proporciona un valor muy próximo al valor real, especialmente cuando el factor de sobremuestreo es superior a cinco. Por otro lado, el coste computacional del método propuesto se ha comparado con una alternativa basada en técnicas convencionales multitasa implementadas mediante su estructura polifase, obteniéndose un número inferior de coeficientes cuando el factor de sobremuestreo es superior a diez. Es más, dado que el coste computacional del retardador de Lagrange, para una resolución y sobremuestreo dados, es independiente del valor de retardo, D, resulta que el filtro de retardo fraccionario de Lagrange proporcionará una solución computacionalmente mucho menos costosa que las técnicas tradicionales en aquellas aplicaciones en las que el retardo sea un número no racional (o racional con denominador elevado).

3.4. Conclusiones

Con el objeto de motivar la búsqueda de nuevas técnicas polinómicas en procesado de señal que esta Tesis lleva a cabo, en el presente capítulo se han planteado dos ejemplos que muestran que, en aplicaciones prácticas, las técnicas de interpolación polinómicas pueden ofrecer prestaciones superiores al considerar simultáneamente precisión y gasto computacional.

En concreto, en el primer ejemplo se ha comprobado la superioridad (bajo los condicionantes prácticos comentados) de la interpolación spline cúbica frente a la interpolación basada en el truncamiento de la función sinc. De igual forma, en el segundo ejemplo se ha demostrado la utilidad de los interpoladores de retardo fraccionario basados en los polinomios de Lagrange.

Capítulo 4

α-splines

4.1. Introducción

La interpolación spline se aplica de forma habitual en problemas de reconstrucción de señales (principalmente de imágenes). El orden de la interpolación spline se escoge en función de las características que se deseen en la solución, por ejemplo, splines lineales (orden uno) para aplicaciones simples en las que no se precise continuidad en las derivadas, splines cúbicos (orden tres) para obtener máxima suavidad en la solución, etc. Sin embargo, únicamente es posible escoger entre órdenes enteros, y las limitaciones prácticas (fundamentalmente el coste computacional) restringen la elección a valores relativamente bajos. En esta Tesis se propone una nueva familia de splines [Ibáñez04], los α-splines, dependientes de un parámetro que posibilita una transición suave entre la interpolación spline de dos órdenes cualesquiera y que aporta, por tanto, una flexibilidad adicional a la interpolación spline. De esta forma, si la interpolación spline de cierto orden proporciona determinadas propiedades que puedan ser complementadas por las aportadas por la interpolación spline de un orden distinto, entonces será posible controlar la preponderancia de unas y otras características gracias al parámetro de control de la interpolación α-spline.

Sin duda es la interpolación spline cúbica la técnica spline más habitual. Sin embargo, la solución obtenida puede producir oscilaciones o sobredisparo entre muestras, generando de esta manera una señal reconstruida que no preserva ciertas propiedades estadísticas de los datos originales [Kashyap96, Waele00], ni otras propiedades interesantes como pueden ser la monotonicidad o la convexidad. Estos efectos indeseables son minimizados en el caso de aproximaciones lineales a tramos, a costa, obviamente, de discontinuidad en las derivadas. Por esta razón, resulta especialmente interesante el caso particular de la familia α-spline propuesta en el que se transiciona entre el spline lineal y el spline cú-

bico. La interpolación mediante splines cúbicos y la interpolación lineal a tramos (splines de orden uno) tienen características y propiedades diferenciadas y, en ciertos aspectos, opuestas. Mientras la primera es ampliamente utilizada y presenta las ventajas de su derivabilidad y suavidad, la segunda proporciona una solución simple y un comportamiento robusto frente al ruido en los datos. Ciertamente, aquellos aspectos que resultan positivos en una de las soluciones se tornan negativos en la otra, y viceversa. Por ello, cobra un interés especial la interpolación α-spline que se sitúa entre el spline cúbico y el spline de orden uno, de forma que el parámetro permite controlar el grado de influencia de una u otra solución. Dicho parámetro puede ser fijado, por ejemplo, para obtener la aproximación más suave que mantenga cierto grado de sobredisparo o que preserve la varianza de las muestras originales.

En este capítulo se presenta esta nueva familia de funciones núcleo de interpolación denominada α-spline, considerando, en primer lugar, muestreo uniforme. Además de las expresiones de las funciones base α-spline, se prestará especial atención a las propiedades del espacio de señal generado por ellas. Dado el interés de los splines lineales y cúbicos, se particularizarán los resultados para la transición entre orden uno y orden tres, desarrollándose las expresiones de la función base y sus derivadas; la metodología de interpolación; y las expresiones cerradas de la suavidad, la varianza y el área residual de la interpolación. Finalmente, se generalizan los resultados para muestreo no equiespaciado.

4.2. Familia α-spline

Los splines de orden p se definieron originalmente [Boor78] como una señal a tramos, constituida en cada uno de ellos por un polinomio de orden p y con continuidad hasta la derivada $p-1$. Por otro lado, la base spline (B-spline) de orden p, $\beta_p(t)$, se representa en la literatura como la convolución de $p+1$ pulsos rectangulares idénticos de duración unidad, $\beta_0(t)$.[1] Dicha base spline desempeña un papel fundamental, pues su forma y características determina las propiedades de los splines que genera. Así, se dispone de un conjunto infinito pero discreto de bases spline, desde orden cero hasta orden infinito, que proporcionan soluciones que abarcan desde señales escalonadas (*zero-order hold*) hasta señales estrictamente paso bajo (sinc), pasando, entre otras, por soluciones lineales a tramos y por los splines cúbicos.

En la práctica se utilizan mayoritariamente splines lineales (orden 1) o splines cúbicos en función de la suavidad o la complejidad requerida. La obtención de una familia de splines que no estén restringidos a órdenes enteros supondría una mayor flexibilidad a la hora de encontrar la solución que mejor se adapte a los requisitos de cada aplicación. En esta línea, en [Horbelt00] se presentan

[1] Se considera, salvo que se diga expresamente lo contrario, muestreo uniforme con periodo de muestreo $T=1$.

los llamados núcleos convolucionales B-spline (*B-spline convolution kernels*), formados a partir de la convolución de dos bases spline de orden p_1 y p_2 y anchuras h_1 y h_2. Sin embargo, no existe un desarrollo amplio de los mismos, y su utilización se limita a aplicaciones de imagen.

De forma paralela al trabajo de [Horbelt00] y con un objetivo similar, esta Tesis presenta una nueva familia de bases spline, denominadas α-splines y denotadas como $\beta_\alpha^{[p\to p+q]}(t)$, que supone una transición continua entre la base spline de orden p y la base de orden $p+q$: la base α-spline de orden $[p \to p+q]$ se define como la convolución de $p+1$ pulsos rectangulares de duración unidad, $\beta_0(t)$, y q pulsos rectangulares de duración $0 \leq \alpha \leq 1$, es decir,

$$\beta_\alpha^{[p\to p+q]}(t) = \overbrace{\beta_0(t) * \cdots * \beta_0(t)}^{p+1} * \overbrace{p_\alpha(t) * \cdots * p_\alpha(t)}^{q}, \tag{4.1}$$

donde

$$p_\alpha(t) = \begin{cases} \frac{1}{\alpha}, & |t| \leq \frac{\alpha}{2}; \\ 0, & |t| > \frac{\alpha}{2}. \end{cases} \tag{4.2}$$

Con el fin de obtener una expresión cerrada de esta base, se parte de su transformada de Fourier

$$\mathrm{TF}\left\{\beta_\alpha^{[p\to p+q]}(t)\right\} = B_\alpha^{[p\to p+q]}(\Omega) = \left(\frac{\sin(\Omega/2)}{\Omega/2}\right)^{p+1}\left(\frac{\sin(\Omega\alpha/2)}{\Omega\alpha/2}\right)^{q}, \tag{4.3}$$

y se define la siguiente función causal

$$(t)_+^p = \begin{cases} t^p, & t \geq 0; \\ 0, & t < 0; \end{cases} \tag{4.4}$$

cuya transformada de Fourier, denotada como $X_+^p(\Omega)$, cumple la igualdad [Unser99]

$$(j\Omega)^{p+1} X_+^p(\Omega) = p!.$$

Haciendo uso de esta identidad, es posible reescribir (4.3)

$$\mathrm{TF}\left\{\beta_\alpha^{[p\to p+q]}(t)\right\} = \frac{X_+^{p+1+q}(\Omega)}{(p+q)!\ \alpha^q}\left[e^{j\Omega/2} - e^{-j\Omega/2}\right]^{p+1}\left[e^{j\Omega\alpha/2} - e^{-j\Omega\alpha/2}\right]^{q},$$

y, utilizando la fórmula del binomio de Newton, expandir los términos exponenciales

$$\begin{aligned}
&\mathrm{TF}\left\{\beta_\alpha^{[p\to p+q]}(t)\right\} \\
&= \frac{X_+^{p+1+q}(\Omega)}{(p+q)!\ \alpha^q}\sum_{k=0}^{p+1}\binom{p+1}{k}(-1)^k e^{-j\Omega(k-(p+1)/2)}\sum_{l=0}^{q}\binom{q}{l}(-1)^l e^{-j\Omega\alpha(l-q/2)} \\
&= \frac{1}{(p+q)!\ \alpha^q}\sum_{k=0}^{p+1}\sum_{l=0}^{q}\binom{p+1}{k}\binom{q}{l}(-1)^{k+l} e^{-j\Omega[k-(p+1)/2+\alpha(l-q/2)]} X_+^{p+1+q}.
\end{aligned}$$

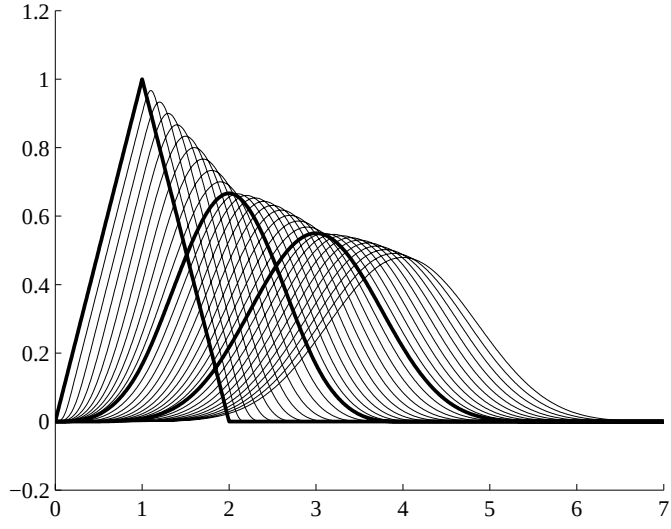


Figura 4.1: Base α-spline, $\beta_\alpha^{[p\to p+q]}(t)$, para $p = 1, 3, 5$, $q = 2$ y α variando entre 0 y 1.

Finalmente, considerando las exponenciales complejas en el dominio frecuencial como desplazamientos temporales, se obtiene la expresión cerrada de la base α-spline de orden $[p \to p+q]$

$$\beta_\alpha^{[p\to p+q]}(t) \tag{4.5}$$
$$= \frac{1}{(p+q)!\ \alpha^q} \sum_{k=0}^{p+1} \sum_{l=0}^{q} \binom{p+1}{k} \binom{q}{l} (-1)^{k+l} \left(t - k + \frac{p+1}{2} - \alpha + \alpha \frac{q}{2} \right)_+^{p+q},$$

que aparece representada en Figura 4.1 para distintos valores de p, q y α.

4.2.1. Propiedades

Cualquier función arbitraria (incluida la base α-spline) es susceptible de ser considerada como función base, ya sea para interpolar, aproximar o modelar. Sin embargo, es conveniente que el esquema de muestreo que se deriva de dicha base posea cierto paralelismo con la teoría clásica de muestreo (cuya función base es la función sinc) y, a la vez, resulte práctico. A partir de la base α-spline de orden $[p \to p+q]$, se define una señal α-spline como

$$s_\alpha(t) = \sum_{n=-\infty}^{+\infty} c[n] \beta_\alpha^{[p\to p+q]}(t-n), \tag{4.6}$$

y se define el espacio de señales α-spline

$$V(\beta_\alpha^{[p\to p+q]}) = \left\{ s_\alpha(t) = \sum_{n=-\infty}^{+\infty} c[n]\beta_\alpha^{[p\to p+q]}(t-n) \ : \ c[n] \in l_2 \right\}. \tag{4.7}$$

Este espacio está constituido por señales (continuas en el tiempo) caracterizadas por una secuencia de coeficientes $c[n]$. Dichos coeficientes α-spline, que constituyen la representación discreta de la señal $s_\alpha(t)$, no tienen que coincidir necesariamente con las muestras de la señal (como ocurre en el caso sinc).

A continuación, se demuestra que la familia de funciones base α-spline $\left\{\beta_\alpha^{[p\to p+q]}(t-n)\right\}_{n\in\mathbb{Z}}$ forma una base de Riesz del espacio $V(\beta_\alpha)$.

Proposición 4.1 (Base de Riesz α-spline) *El conjunto de funciones base α-spline $\{\beta_\alpha(t-n)\}_{n\in\mathbb{Z}}$ es una base de Riesz, siendo*

$$A = \left(\frac{2}{\pi}\right)^{2(p+1)} \operatorname{sinc}^{2q}\left(\frac{\alpha}{2}\right),$$
$$B = 1 + \frac{2^{2(p+1)}\left(2^{2(p+1)}-1\right)}{(2(p+1))!} B_{p+1},$$

dos valores que verifican

$$A||c||_{l_2}^2 \leq \left\| \sum_{n\in\mathbb{Z}} c[n]\beta_\alpha^{[p\to p+q]}(t-n) \right\|^2 \leq B||c||_{l_2}^2, \quad \forall\ c[n] \in l_2;$$

siendo B_n el n-ésimo número de Bernoulli.

Demostración 4.1 *La demostración se encuentra en el Apéndice B.1.*

Como se comentó en el Apartado 2.6.2, es preciso imponer una condición adicional a las bases α-spline, consistente en que (4.6) sea capaz de aproximar cualquier señal de forma tan precisa como sea necesario sin más que seleccionar un periodo de muestreo suficientemente pequeño. Tal y como se demuestra en [Unser00], esta condición equivale a que las funciones base formen una *partición de la unidad.*

Proposición 4.2 (Partición de la Unidad α-spline) *El conjunto de funciones base α-spline $\{\beta_\alpha(t-n)\}_{n\in\mathbb{Z}}$ forman una partición de la unidad.*

Demostración 4.2 *Haciendo uso de la fórmula estándar del sumatorio de Poisson*

$$\sum_{n\in\mathbb{Z}} x(n) = \sum_{k\in\mathbb{Z}} X(2\pi k),$$

donde $X(\Omega) = \mathrm{TF}\{x(t)\}$, *y teniendo en cuenta (4.3), se obtiene*

$$\begin{aligned}\sum_{n\in\mathbb{Z}} \beta_\alpha^{[p\to p+q]}(t-n) &= \sum_{k\in\mathbb{Z}} B_\alpha^{[p\to p+q]}(2\pi k)\, e^{-j2\pi kt} \\ &= \sum_{k\in\mathbb{Z}} \left[\frac{\sin(\pi k)}{\pi k}\right]^{p+1} \left[\frac{\sin(\alpha\pi k)}{\alpha\pi k}\right]^{q} e^{-j2\pi kt} = 1,\end{aligned}$$

puesto que $\sin(\pi k)/(\pi k) = \delta[k]$.

Estos resultados aseguran que la representación (4.6) es estable y que el espacio (4.7) definido por las bases α-spline es un subespacio cerrado de L_2, y establecen la base teórica sobre la que se fundamenta el resto del capítulo.

4.3. Base α-spline

Sin duda, los splines lineales y los splines cúbicos son, de todos los splines, los más empleados. Por esta razón, en esta Tesis se prestará especial atención a los α-splines que transicionan entre el spline lineal y el cúbico. Así, de forma general, se considerará la base α-spline, denotada $\beta_\alpha(t)$, como la convolución de dos pulsos rectangulares de duración unitaria y dos pulsos rectangulares de duración $0 \le \alpha \le 1$, es decir, se trata de una particularización de los α-splines presentados en la anterior sección para $p = 1$ y $q = 2$

$$\beta_\alpha(t) = \beta_\alpha^{[1\to 3]}(t) = \beta_0(t) * p_\alpha(t) * \beta_0(t) * p_\alpha(t). \tag{4.8}$$

Particularizando (4.5) se tiene

$$\beta_\alpha(t) = \frac{1}{6\alpha^2}\sum_{k=0}^{2}\sum_{l=0}^{2}\binom{2}{k}\binom{2}{l}(-1)^{k+l}\left(t-k+1-\alpha(l-1)\right)_+^3, \tag{4.9}$$

y desarrollando se obtienen la siguientes expresiones de la base α-spline,

$$\beta_\alpha(t) = \begin{cases} 1-|t|-\frac{1}{3\alpha^2}(\alpha-|t|)^3, & t\in T_1; \\ 1-|t|, & t\in T_2; \\ \frac{(\alpha+1)\left(3-(\alpha-2)^2-3t^2\right)}{6\alpha^2}+\frac{(\alpha-1)^2+t^2}{2\alpha^2}|t|, & t\in T_3; \\ 1-|t|+\frac{1}{6\alpha^2}(\alpha-1+|t|)^3, & t\in T_4; \\ \frac{1}{6\alpha^2}(\alpha+1-|t|)^3, & t\in T_5; \\ 0, & t\in T_6; \end{cases} \tag{4.10}$$

donde $T_1 = \{|t| < \text{mín}(\alpha, 1-\alpha)\}$, $T_2 = \{\alpha \le |t| < 1-\alpha\}$, $T_3 = \{1-\alpha \le |t| < \alpha\}$, $T_4 = \{\text{máx}(\alpha, 1-\alpha) \le |t| < 1\}$, $T_5 = \{1 \le |t| < 1+\alpha\}$ y $T_6 = \{|t| \ge 1+\alpha\}$.

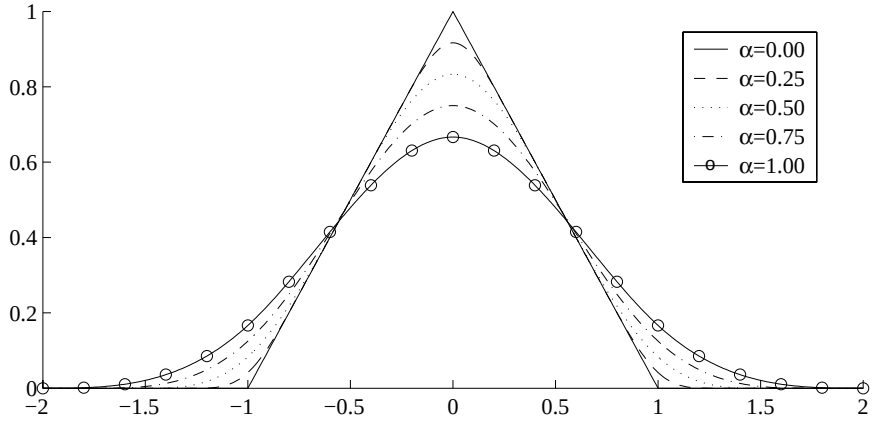


Figura 4.2: Base α-spline, $\beta_\alpha(t)$, para distintos valores del parámetro α.

Por otro lado, su transformada de Fourier es

$$B_\alpha(\Omega) = \mathrm{TF}\{\beta_\alpha(t)\} = \left[\frac{\sin(\Omega/2)}{\Omega/2}\right]^2 \left[\frac{\sin(\Omega\alpha/2)}{\Omega\alpha/2}\right]^2. \tag{4.11}$$

Se puede comprobar fácilmente, tanto en las expresiones (4.8) y (4.10) como en la Figura 4.2, que la base α-spline propuesta es idéntica a la base lineal, $\beta_1(t)$, cuando $\alpha = 0$; e igual al B-spline cúbico, $\beta_3(t)$, cuando $\alpha = 1$. Para valores intermedios del parámetro α se obtienen funciones base que se encuentran entre ambas soluciones, intercambiando la curvatura óptima de la interpolación mediante splines cúbicos por una solución más proxima a la interpolación lineal a tramos; pero siempre, excepto para el valor límite de $\alpha = 0$, manteniendo la continuidad de la primera y segunda derivadas.

En concreto, la expresión de la primera derivada de la base α-spline es

$$\beta_\alpha^{(1)}(t) = \begin{cases} \operatorname{sgn}(t)\frac{(\alpha-|t|)^2-\alpha^2}{\alpha^2}, & t \in T_1; \\ -\operatorname{sgn}(t), & t \in T_2; \\ -\operatorname{sgn}(t)\frac{4\alpha-(\alpha-|t|+1)^2+2x^2}{2\alpha^2}, & t \in T_3; \\ \operatorname{sgn}(t)\frac{(\alpha+|t|-1)^2-2\alpha^2}{2\alpha^2}, & t \in T_4; \\ -\operatorname{sgn}(t)\frac{(\alpha+|t|-1)^2}{2\alpha^2}, & t \in T_5; \\ 0, & t \in T_6; \end{cases} \tag{4.12}$$

mientras que la segunda derivada se escribe como,

$$\beta_\alpha^{(2)}(t) = \begin{cases} \frac{-2(\alpha-|t|)}{\alpha^2}, & t \in T_1; \\ 0 & t \in T_2; \\ \frac{-\alpha+3|t|-1}{\alpha^2}, & t \in T_3; \\ \frac{\alpha+|t|-1}{\alpha^2}, & t \in T_4; \\ \frac{\alpha-|t|+1}{\alpha^2}, & t \in T_5; \\ 0, & t \in T_6. \end{cases} \tag{4.13}$$

Se puede apreciar, tanto en (4.13) como en la Figura 4.3b, que, como en el caso del spline cúbico, toda la familia α-spline posee una derivada segunda continua y lineal a tramos. Este hecho resulta evidente al expresar dicha función como suma de pulsos triangulares, es decir,

$$\beta_\alpha^{(2)}(t) = \frac{1}{\alpha}\left[\Delta\left(\frac{t+1}{\alpha}\right) - 2\Delta\left(\frac{t}{\alpha}\right) + \Delta\left(\frac{t-1}{\alpha}\right)\right].$$

Por último, la tercera derivada de la función base α-spline viene dada por,

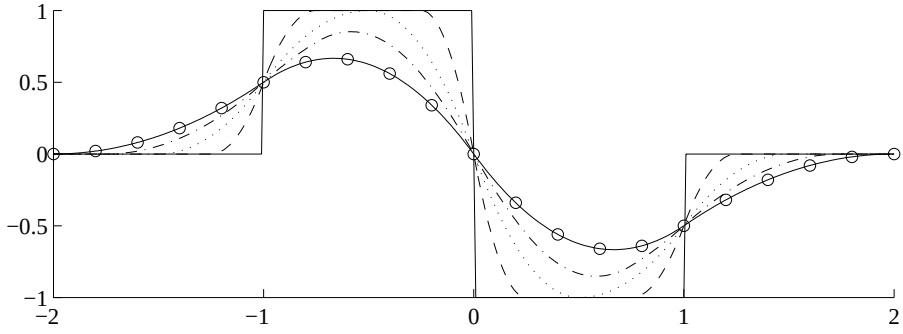
$$\beta_\alpha^{(3)}(t) = \begin{cases} 2\,\mathrm{sgn}\,(t)/\alpha^2, & t \in T_1; \\ 0, & t \in T_2; \\ 3\,\mathrm{sgn}\,(t)/\alpha^2, & t \in T_3; \\ \mathrm{sgn}\,(t)/\alpha^2, & t \in T_4; \\ -\,\mathrm{sgn}\,(t)/\alpha^2, & t \in T_5; \\ 0, & t \in T_6. \end{cases} \tag{4.14}$$

Ésta es una función escalonada (ver Figura 4.3.c), es decir, la derivada tercera de la interpolación α-spline es constante a tramos. Por ello, puede resultar clarificador expresarla en términos de funciones pulso rectangulares,

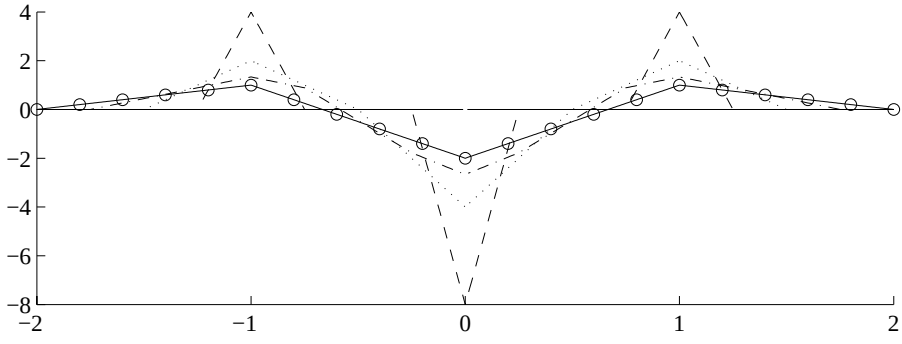
$$\begin{aligned} \beta_\alpha^{(3)}(t) = \frac{1}{\alpha}\,[&p_\alpha(t+1+\alpha/2) - p_\alpha(t+1-\alpha/2) - 2p_\alpha(t+\alpha/2) \\ &+2p_\alpha(t-\alpha/2) + p_\alpha(t-1+\alpha/2) - p_\alpha(t-1-\alpha/2)]\,. \end{aligned} \tag{4.15}$$

Conviene recalcar que la base α-spline hereda las propiedades de la familia α-spline, $\beta_\alpha^{[p\to p+q]}(t)$, puesto que no es más que un caso particular de la misma. En concreto, el conjunto de bases α-spline, $\{\beta_\alpha(t-n)\}_{n\in\mathbb{Z}}$ forma una base de Riesz del espacio

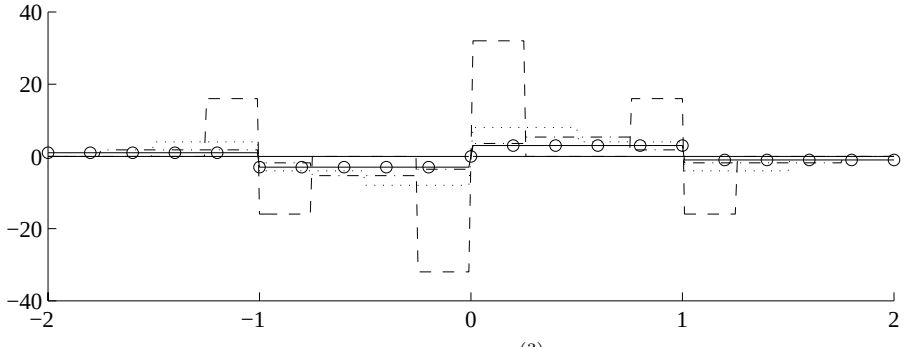
$$V(\beta_\alpha) = \left\{ s_\alpha(t) = \sum_{n=-\infty}^{+\infty} c[n]\beta_\alpha(t-n) \;:\; c[n] \in l_2 \right\},$$



(a) Derivada 1ª, $\beta_\alpha^{(1)}(t)$.



(b) Derivada 2ª, $\beta_\alpha^{(2)}(t)$.



(c) Derivada 3ª, $\beta_\alpha^{(3)}(t)$.

Figura 4.3: Derivadas de la función base α-spline, para distintos valores del parámetro α.

pues se cumple que

$$A||c||_{l_2}^2 \leq \left\| \sum_{n\in\mathbb{Z}} c[n]\beta_\alpha\left(t-n\right) \right\|^2 \leq B||c||_{l_2}^2, \quad \forall\, c[n] \in l_2;$$

con

$$A = \left(\frac{2}{\pi}\right)^4 \operatorname{sinc}^4\left(\frac{\alpha}{2}\right),$$
$$B = \frac{4}{3}.$$

Así mismo, las bases α-spline constituyen una partición de la unidad, lo que, unido a lo anterior, asegura un adecuado comportamiento de la interpolación α-spline.

4.4. Interpolación α-spline

La interpolación α-spline, $s_\alpha(t)$, de una señal $x(t)$ se puede escribir como

$$s_\alpha(t) = \sum_{n\in\mathbb{Z}} c[n]\beta_\alpha\left(t-n\right), \tag{4.16}$$

donde $\beta_\alpha\left(t-n\right)$ es la base α-spline centrada en $t=n$, y $c[n]$ son los coeficientes α-spline que cumplen la condición de interpolación

$$s_\alpha\left(n\right) = x(n), \quad \forall\, n \in \mathbb{Z}. \tag{4.17}$$

Esta expresión hace referencia a señales, o funciones, de duración infinita. Dado que en la práctica las señales manejadas (piénsese, por ejemplo, en imágenes) tienen duración finita $(x(n), n = 1, \ldots, N)$ se hace necesario imponer un conjunto de condiciones de contorno. La resolución de este problema puede afrontarse de una forma similar a la tradicional interpolación mediante splines cúbicos [Boor78], es decir, planteando un sistema de ecuaciones tridiagonal que se resuelve usando técnicas numéricas estándar, con el fin de obtener los coeficientes $c[n]$ de (4.16).

Así mismo, puede utilizarse la metodología (más eficiente) planteada en [Unser93a] y [Unser93b] basada en filtrado digital, aunque, evidentemente, con ciertas modificaciones que se presentarán posteriormente.

4.4.1. Solución matricial

Habitualmente, los problemas de interpolación y aproximación mediante splines se resuelven utilizando formulaciones matriciales [Boor78]. De forma

totalmente equivalente se puede afrontar el problema de interpolación mediante α-splines.

Se considera la interpolación de N puntos, es decir,

$$s_\alpha(n) = x(n), \quad \forall\, n = 1, 2, \ldots, N. \tag{4.18}$$

Para simplificar la notación, los valores de la base α-spline en los instantes de muestreo (valores enteros de t) se denotan como

$$\begin{aligned} a &= \beta_\alpha(t)|_{t=0} = \frac{3-\alpha}{3}, \\ b &= \beta_\alpha(t)|_{t=+1} = \beta_\alpha(t)|_{t=-1} = \frac{\alpha}{6}; \end{aligned}$$

siendo el resto de valores ($\beta_\alpha(n)$, $\forall\, |n| \geq 2$) nulos. El soporte compacto de las funciones base α-spline provoca que en cada una de las condiciones de interpolación de (4.18) intervengan únicamente tres funciones base: las correspondientes al punto en cuestión, al anterior y al posterior; resultando

$$\begin{aligned} x(n) = s_\alpha(n) &= \sum_{k=-\infty}^{+\infty} c[k]\beta_\alpha(n-k) \\ &= bc[n-1] + ac[n] + bc[n+1], \quad 2 \leq n \leq N-1. \end{aligned} \tag{4.19}$$

En los extremos del intervalo se deben considerar las condiciones de frontera. Asumiendo que la señal original es periódica de periodo N

$$x(t) = x(t + kN), \quad \forall\, k; \tag{4.20a}$$

y, por tanto, también los coeficientes α-spline

$$c[n] = c[n + kN], \quad \forall\, k; \tag{4.20b}$$

entonces, en los extremos de la interpolación deben cumplirse las condiciones

$$\begin{aligned} x(1) &= bc[0] + ac[1] + bc[2] = bc[N] + ac[1] + bc[2], \\ x(N) &= bc[N-1] + ac[N] + bc[N+1] = bc[N-1] + ac[N] + bc[1]. \end{aligned}$$

Expresado en forma matricial, el sistema a resolver cuando se consideran condiciones de frontera periódicas es

$$\begin{bmatrix} a & b & 0 & 0 & \ldots & 0 & b \\ b & a & b & 0 & \ldots & 0 & 0 \\ 0 & b & a & b & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ b & 0 & 0 & 0 & \ldots & b & a \end{bmatrix} \begin{bmatrix} c[1] \\ c[2] \\ c[3] \\ \vdots \\ c[N] \end{bmatrix} = \begin{bmatrix} x(1) \\ x(2) \\ x(3) \\ \vdots \\ x(N) \end{bmatrix}. \tag{4.21}$$

La simetría especular en los extremos (situación ésta habitual en procesado de imagen) fuerza que

$$\begin{aligned} x(t) &= x(2N-t), \quad N \leq t \leq 2N-1; \\ c[n] &= c[2N-n], \quad N \leq n \leq 2N-1; \end{aligned}$$

y que tanto la señal original como la secuencia de coeficientes de la interpolación sean periódicas de periodo $2N-1$, es decir,

$$\begin{aligned} x(t) &= x(t+k(2N-1)), \quad \forall\, k; \\ c[n] &= c[n+k(2N-1)], \quad \forall\, k; \end{aligned}$$

resultando que, además de las condiciones dadas por (4.19), en los extremos debe cumplirse

$$\begin{aligned} x(1) &= bc[0] + ac[1] + bc[2] = bc[1] + ac[1] + bc[2], \\ x(N) &= bc[N-1] + ac[N] + bc[N+1] = bc[N-1] + ac[N] + bc[N-1]. \end{aligned}$$

Por lo tanto, el sistema matricial correspondiente a la interpolación α-spline con condiciones de frontera especulares es

$$\begin{bmatrix} a+b & b & 0 & 0 & \dots & 0 & 0 \\ b & a & b & 0 & \dots & 0 & 0 \\ 0 & b & a & b & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 2b & a \end{bmatrix} \begin{bmatrix} c[1] \\ c[2] \\ c[3] \\ \vdots \\ c[N] \end{bmatrix} = \begin{bmatrix} x(1) \\ x(2) \\ x(3) \\ \vdots \\ x(N) \end{bmatrix}. \tag{4.22}$$

Es posible imponer otras condiciones de frontera, tales como fijar el valor de la derivada (o de la derivada segunda) en los puntos extremos del intervalo. En cualquier caso el sistema de ecuaciones puede ser resuelto muy eficientemente mediante eliminación Gaussiana o factorización LU [Golub89].

4.4.2. Solución mediante filtrado digital

A partir de la expresión (4.16), es posible considerar la interpolación como una convolución (filtrado) de los coeficientes de interpolación $c[n]$ con la función base α-spline

$$s_\alpha(t) = \sum_{n=\in\mathbb{Z}} c[n]\beta_\alpha(t-n) = c[n] * \beta_\alpha(t). \tag{4.23}$$

Este proceso de interpolación, desde el punto de vista de filtrado digital, se muestra en la Figura 4.4: resulta evidente que, para cumplir las condiciones de interpolación (4.17), es preciso que el denominado filtro directo α-spline, denotado por $d_\alpha[n]$, sea el inverso del filtro indirecto, $\beta_\alpha(t)$, discretizado.

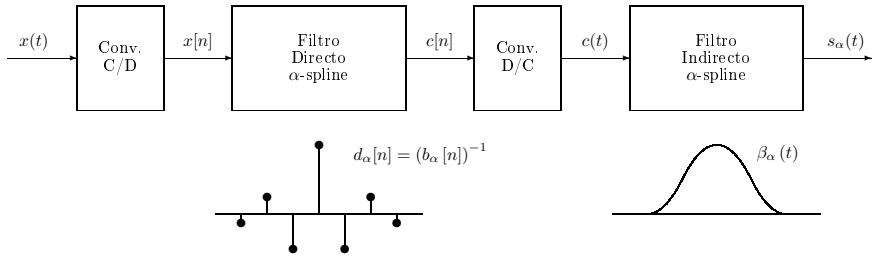


Figura 4.4: Esquema de interpolación α-spline.

Muestreando la base α-spline en los instantes $t = n \in \mathbb{Z}$, se obtiene

$$b_\alpha[n] = (\alpha\delta[n+1] + (6-2\alpha)\delta[n] + \alpha\delta[n-1])/6, \tag{4.24}$$

cuya transformada z es

$$B_\alpha(z) = \left(\alpha z + (6-2\alpha) + \alpha z^{-1}\right)/6.$$

Por lo tanto, para obtener los coeficientes $c[n]$ es necesario filtrar la secuencia $x[n]$ mediante el filtro inverso de $b_\alpha[n]$, es decir, un filtro digital con la siguiente función de transferencia

$$D_\alpha(z) = (B_\alpha)^{-1}(z) = \frac{6}{\alpha z + (6-2\alpha) + \alpha z^{-1}} = \frac{6}{\alpha}\frac{1}{1-z_1 z^{-1}}\frac{-z_1}{1-z_1 z}, \tag{4.25}$$

que posee un par de polos reales recíprocos $\{(z_1, z_1^{-1}),\ |z_1| < 1\}$, es decir, z_1 denota el polo situado dentro del círculo unidad y cuyo valor es

$$z_1 = \frac{\alpha - 3 + \sqrt{9-6\alpha}}{\alpha} \in \mathbb{R}. \tag{4.26}$$

Dado que $0 \leq \alpha \leq 1$, entonces z_1 toma valores negativos en el rango

$$-2+\sqrt{3} = -0{,}26795 \leq z_1 \leq 0.$$

Por otro lado, es posible obtener el valor de α a partir de z_1, sin más que despejar de (4.26),

$$\alpha = -\frac{6z_1}{(1-z_1)^2}. \tag{4.27}$$

Considerando esta expresión y realizando la transformada z inversa de (4.25), se obtiene la respuesta al impulso del filtro directo α-spline

$$d_\alpha[n] = \frac{-6z_1}{\alpha}\left(z_1^n u[n]\right) * \left(z_1^{-n} u[-n]\right) = \frac{(1-z_1)}{(1+z_1)} z_1^{|n|}, \tag{4.28}$$

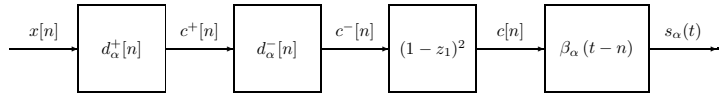


Figura 4.5: Interpolación α-spline mediante filtrado digital.

que representa un filtro digital simétrico que decae exponencialmente y que puede ser implementado mediante un algoritmo recursivo [Unser93b] más eficiente, en cuanto a número de operaciones requeridas, que la resolución del sistema matricial equivalente. En concreto, esta solución alternativa se basa en la descomposición del filtro directo spline en dos filtros: uno causal y otro anticausal. Como muestra la Figura 4.5, los coeficientes, $c[n]$, de la interpolación de una secuencia, $x[n]$, se pueden obtener mediante el filtrado de dicha secuencia con el filtro digital $d_\alpha[n]$ de (4.28), es decir,

$$c[n] = x[n] * d_\alpha[n] = x[n] * d_\alpha^+[n] * d_\alpha^-[n],$$

donde el filtro causal, $d_\alpha^+[n]$, y el anticausal, $d_\alpha^-[n]$, toman la forma

$$d_\alpha^\pm[n] = z_1^{\pm n} u[\pm n],$$

con transformada z

$$D_\alpha^\pm(z) = \frac{1}{1 - z_1 z^{\mp 1}}. \tag{4.29}$$

Por lo tanto, (4.28) puede ser reescrita como

$$d_\alpha[n] = \frac{-6z_1}{\alpha}\, d_\alpha^+[n] * d_\alpha^-[n] = (1-z_1)^2 d_\alpha^+[n] * d_\alpha^-[n]. \tag{4.30}$$

Esta metodología fue desarrollada por M. Unser y otros autores [Unser93b] para realizar la interpolación spline de cualquier orden imponiendo condiciones de frontera especulares (típicamente utilizadas en el procesado de imágenes). En esta Tesis, según se ha descrito en el presente apartado, se adapta dicho método recursivo a la interpolación α-spline y se extiende de manera que permita utilizar condiciones de frontera periódicas (más adecuadas para el procesado de señales unidimensionales), tal y como se presenta a continuación.

4.4.2.1. Condiciones de frontera periódicas

La interpolación α-spline de una secuencia $x[n]$ de N puntos con condiciones de frontera periódicas, es decir, de una señal que cumple

$$x(t) = x(t + kN), \quad \forall\, k;$$

puede resolverse recursivamente de forma muy eficiente, pues requiere únicamente $(5N-2)$ operaciones.

En primer, lugar es preciso realizar el filtrado de las muestras $x[n]$ mediante el filtro causal IIR $d_\alpha^+[n]$ (ver la Figura 4.5) con el fin de obtener los N valores de la secuencia $c^+[n]$. Según (4.29), es evidente que $c^+[n] = x[n] + z_1 c^+[n-1]$; sin embargo es necesario obtener un valor inicial para comenzar la recursión. Haciendo uso de la condición de periodicidad se obtiene

$$c^+[1] = x[1] + z_1 c^+[0] = x[1] + z_1 c^+[N],$$

y, siguiendo la recursión,

$$\begin{aligned} c^+[1] &= x[1] + z_1 c^+[N] = x[1] + z_1(x[N] + z_1 c^+[N-1]) = \cdots \\ &= x[1] + \sum_{n=1}^{N-1} z_1^n x[N-n+1] + z_1^{N+1} c^+[1], \end{aligned}$$

resultando un valor

$$c^+[1] = \frac{1}{1-z_1^N}\left(x[1] + \sum_{n=1}^{N-1} z_1^n x[N-n+1]\right), \tag{4.31a}$$

que permite iniciar el filtrado recursivo

$$c^+[n] = x[n] + z_1 c^+[n-1], \quad n = 2, 3, \ldots, N. \tag{4.31b}$$

En segundo lugar, se realiza el filtrado de la secuencia $c^-[n]$ con el filtro anticausal IIR $d_\alpha^-[n]$. En esta ocasión es preciso calcular el valor $c^-[N]$ que posibilite iniciar la recursión $c^-[n] = c^+[n] + z_1 c^-[n+1]$ dada por (4.29). Así, puesto que $c^-[N+1] = c^-[1]$, se tiene

$$\begin{aligned} c^-[N] &= c^+[N] + z_1 c^-[N+1] = c^+[N] + z_1 c^-[1] \\ &= c^+[N] + z_1(c^+[N-1] + z_1 c^-[2]) = \cdots \\ &= c^+[N] + \sum_{n=1}^{N-1} z_1^n c^+[n] + z_1^N c^-[N], \end{aligned}$$

obteniéndose el valor inicial

$$c^-[N] = \frac{1}{1-z_1^N}\left(c^+[N] + \sum_{n=1}^{N-1} z_1^n c^+[n]\right), \tag{4.31c}$$

de la recursión

$$c^-[n] = c^+[n] + z_1 c^-[n+1], \quad n = N-1, N-2, \ldots, 1; \tag{4.31d}$$

Finalmente, se aplica la constante multiplicativa a la secuencia $c^-[n]$ para obtener los coeficientes α-spline

$$c[n] = (1 - z_1)^2 c^-[k], \quad n = 1, 2, \ldots, N; \tag{4.31e}$$

y evaluar (4.16) para obtener el valor de la interpolación α-spline en los puntos deseados.

4.4.2.2. Condiciones de frontera especulares

En aplicaciones de procesado de imagen, resulta habitual imponer condiciones de frontera especulares que eviten discontinuidades en los bordes de la imagen interpolada, para lo cual se extiende la función $x[n]$ por ambos lados utilizando su imagen especular, es decir,

$$x(t) = x(2N - t), \quad N \leq t \leq 2N - 1;$$

resultando una señal periódica de periodo $2N - 1$, es decir,

$$x(t) = x(t + k(2N - 1)), \quad \forall\, k.$$

En estas circunstancias, [Unser93b] desarrolla el algoritmo recursivo para obtener los coeficientes de la interpolación spline. Resulta inmediata su modificación para adaptarlo a la interpolación α-spline con simetría especular

$$\begin{aligned}
c^+[1] &= \frac{1}{1 - z_1^{2N-2}} \sum_{n=0}^{2N-3} z_1^n x[n]; \\
c^+[n] &= x[n] + z_1 c^+[n-1], \; n = 2, 3, \ldots, N; \\
c^-[N] &= \frac{1}{1 - z_1^2} \left(c^+[N-1] + z_1 c^+[N-2] \right); \\
c^-[n] &= c^+[n] + z_1 c^-[n+1], \; n = N-1, \ldots, 1; \\
c[n] &= (1 - z_1)^2 c^-[n], \; n = 1, 2, \ldots, N.
\end{aligned}$$

El algoritmo es idéntico al presentado para el caso de condiciones periódicas salvo por los valores iniciales de las recursiones ($c^+[1]$ y $c^-[N]$).

4.4.3. Interpolación rápida mediante α-splines

Los splines cúbicos se utilizan ampliamente en la interpolación de datos multidimensionales (principalmente de imágenes). Sin embargo, en aplicaciones en las que la capacidad de cómputo sea un recurso escaso, la interpolación lineal (splines de orden uno) se convierte en una opción habitual, debido a su simplicidad. Como ya se ha comentado, el uso de splines requiere solucionar dos

problemas: un problema directo para obtener los coeficientes de las funciones base; y un problema indirecto que genera la señal interpolada a partir de dichos coeficientes. La interpolación lineal presenta la doble ventaja de no precisar la solución del problema directo (los datos son en sí mismos los coeficientes de las bases triangulares) y de la simplicidad y eficiencia a la hora de generar las muestras de salida.

Cuando se precisa un alto grado de suavidad, la solución más habitual son los splines cúbicos: por ejemplo, en redimensionamiento y transformaciones geométricas de imágenes, una interpolación lineal genera imágenes compuestas por planos y aristas, mientras que la interpolación spline cúbica produce imágenes con curvas y transiciones suaves. En la interpolación spline cúbica es preciso solucionar el problema directo (bien sea mediante inversión matricial [Boor78] o filtrado digital [Unser93b]), para lo cual son necesarias multiplicaciones y sumas en aritmética de punto flotante, mientras que el problema indirecto se puede implementar en punto fijo.

En [Ferrari99] se presenta una técnica que permite calcular de manera eficiente los coeficientes de la interpolación spline mediante sumas y desplazamientos en aritmética de punto fijo, lo que simplifica (y en algunos casos posibilita) la implementación hardware, sobre todo en aplicaciones en tiempo real o similares.

En concreto, si el filtro digital interpolatorio posee una raíz del tipo

$$z_j = -\left(\frac{1}{2}\right)^m, \tag{4.32}$$

entonces los coeficientes de dicho filtro son de la forma

$$d[n] = \left(\frac{1}{2}\right)^{mn}, \tag{4.33}$$

pudiéndose implementar mediante sumas, diferencias y desplazamientos (equivalentes a divisiones por potencias de 2) en punto fijo. Con este fin, [Ferrari99] propone un conjunto de bases splines, que denomina $(1/2)^m$-splines, que cumplen con el requisito (4.32), que poseen características próximas a los splines cúbicos y que posibilitan una implementación computacionalmente más eficiente. Dentro de la familia de α-splines presentada en esta Tesis, esta implementación surge de forma natural, pues existen infinitos valores del parámetro de suavidad α que proporcionan interpolaciones cuyo filtro directo posee raíces como en (4.32).

Así, utilizando las expresiones (4.26) y (4.27) que relacionan α con la raíz z_1 del filtro α-spline se tiene

$$z_1 = -\left(\frac{1}{2}\right)^m \;\Leftrightarrow\; \alpha = \frac{3}{1 + 2^{m-1} + 2^{-m-1}}, \quad m = 2, 3, 4, \ldots \tag{4.34}$$

m	**polo** z_1	**parámetro** α	**filtro** $d_\alpha[n]$
2	$-\frac{1}{4}$	0,9600	$\frac{5}{3}\left(-\frac{1}{4}\right)^{\lvert n\rvert}$
3	$-\frac{1}{8}$	0,5926	$\frac{9}{7}\left(-\frac{1}{8}\right)^{\lvert n\rvert}$
4	$-\frac{1}{16}$	0,3945	$\frac{17}{15}\left(-\frac{1}{16}\right)^{\lvert n\rvert}$
5	$-\frac{1}{32}$	0,1763	$\frac{33}{31}\left(-\frac{1}{32}\right)^{\lvert n\rvert}$

Tabla 4.1: Filtros directos α-spline con coeficientes potencia de dos.

y haciendo uso de (4.28) se obtienen los coeficientes del filtro digital directo, $d_\alpha[n]$, necesario para calcular los coeficientes de interpolación α-spline, con la particularidad de que sus valores son siempre de la forma (4.33). La Tabla 4.1 muestra algunos casos de α-splines con esta propiedad.

Es posible, por tanto, implementar la interpolación α-spline mediante aritmética de punto fijo manteniendo la flexibilidad aportada por el parámetro de suavidad α, que permite obtener desde un resultado muy próximo al spline cúbico ($m = 2$) hasta soluciones menos suaves, próximas a la interpolación lineal (cuando m se hace grande).

La aplicación de esta técnica a la interpolación de imágenes permitirá la utilización de circuitos integrados más simples (con aritmética de punto fijo) con el fin de obtener imágenes remuestreadas (desplazamientos, ampliaciones, ...) con una calidad, tanto desde el punto de vista subjetivo como de relación señal a error, prácticamente idéntica a la interpolación spline cúbica si se utilizan α-splines con $\alpha = 0{,}96$.

4.5. α-spline cardinal

El α-spline cardinal se introduce con el fin de establecer paralelismos y similitudes entre el esquema clásico de interpolación (basado en la función $\operatorname{sinc}(x)$) y la interpolación α-spline; lo que permitirá, entre otras cosas, extender estas técnicas de interpolación a las comunicaciones digitales, tal y como se verá en el Capítulo 6. Se continúa considerando, sin pérdida de generalidad, una frecuencia de muestreo unitaria ($T = 1$). Recordando que la interpolación de señales limitadas en banda se puede expresar como

$$x(t) = \sum_{n\in\mathbb{Z}} x(n)\operatorname{sinc}(t-n),$$

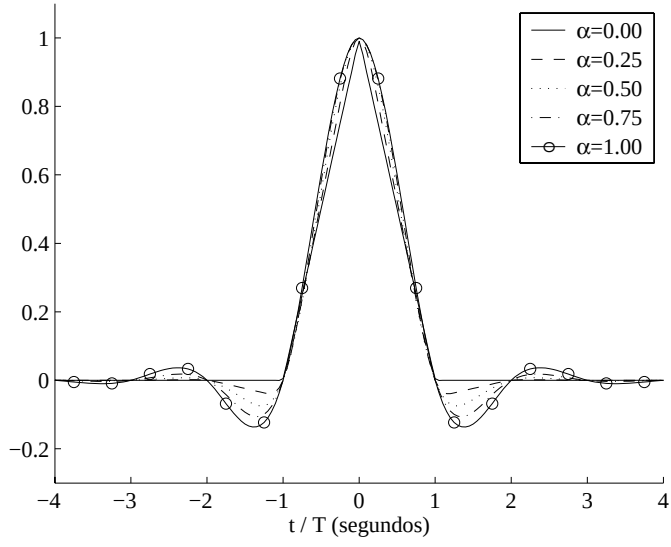


Figura 4.6: α-spline cardinal: respuesta al impulso.

el objetivo consiste en encontrar una formulación equivalente para el caso de la interpolación α-spline, del tipo

$$x(t) = \sum_{n \in \mathbb{Z}} x(n) \eta_\alpha(t - n),$$

donde $\eta_\alpha(t)$ representa el α-spline cardinal. Claramente, este filtro viene dado por la convolución del filtro directo con el filtro indirecto α-spline, por lo que su expresión matemática se puede representar como

$$\eta_\alpha(t) = \sum_{n \in \mathbb{Z}} d_\alpha[n] \beta_\alpha\left(t - n\right). \tag{4.35}$$

Su respuesta al impulso para distintos valores de α, calculada mediante las expresiones (4.10) y (4.28), se representa en la Figura 4.6. En dicha gráfica se comprueba que el α-spline cardinal cumple la propiedad de interpolación: valor unitario en el origen y valores nulos en los múltiplos enteros del tiempo entre muestras. Así mismo, cuando se varía el parámetro α entre 0 y 1, el α-spline cardinal transiciona entre la forma triangular con soporte localizado en $[-1, 1]$ y una forma muy próxima a la función $\operatorname{sinc}(t)$.

La respuesta en frecuencia del proceso interpolatorio nos puede aportar alguna información adicional. Para su cálculo es preciso obtener la transformada de Fourier de (4.35), que equivale a calcular el producto de las transformadas

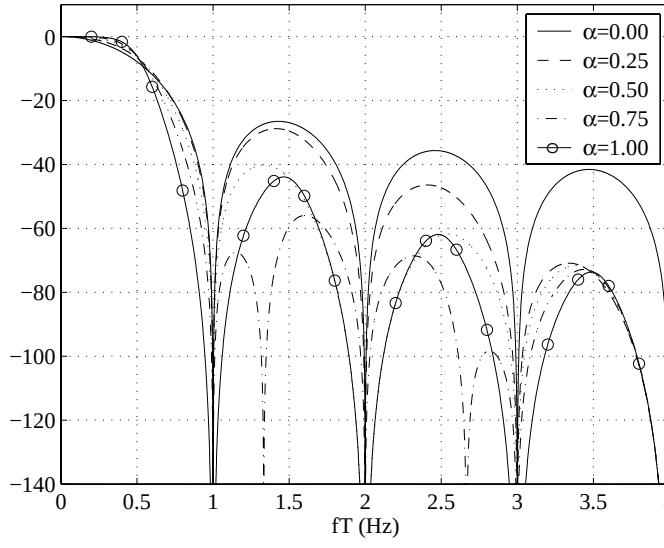


Figura 4.7: α-spline cardinal: respuesta en frecuencia.

de Fourier del filtro α-spline directo, $d_\alpha[n]$, y del filtro α-spline indirecto, $\beta_\alpha(t)$. Por un lado, y a partir de (4.25), se obtiene

$$\mathrm{TF}\left\{d_\alpha[n]\right\} = D_\alpha(z)|_{z=e^{j\omega}} = \frac{3}{3-\alpha+\alpha\cos(\omega)}. \tag{4.36}$$

Por otro lado, el hecho de que el núcleo α-spline en tiempo continuo proceda de la convolución de cuatro pulsos permite obtener su transformada de Fourier (4.11), la cual, combinada con (4.36),[2] proporciona la respuesta en frecuencia de la interpolación α-spline

$$H_\alpha(\Omega) = \mathrm{TF}\left\{\eta_\alpha(t)\right\} = \frac{3\ \mathrm{sinc}^2(\Omega/2\pi)\,\mathrm{sinc}^2(\alpha\Omega/2\pi)}{3-\alpha+\alpha\cos(\Omega)}. \tag{4.37}$$

La Figura 4.7 representa dicha respuesta en frecuencia para distintos valores de α. Se puede observar el comportamiento paso bajo de la interpolación, siendo el caso $\alpha = 1$ el más próximo al filtro paso bajo ideal.

4.6. Suavidad de la interpolación α-spline

La suavidad es una característica que, inicialmente, resulta deseable en muchas aplicaciones de interpolación y que, por tanto, es conveniente evaluar. En

[2] Téngase en cuenta que, a lo largo de todo el capítulo, se está considerando $T = 1$, por lo tanto $\Omega = \omega$.

esta sección se estimará el grado de suavidad de la interpolación α-spline y su dependencia con el parámetro α. En primer lugar, se definirá cuantitativamente la suavidad de una señal para, posteriormente, calcular el valor correspondiente a la interpolación α-spline. Conviene resaltar que la evaluación a priori de la suavidad (o, como se verá más adelante, de la varianza o del área residual) de la interpolación α-spline permite proponer una serie de técnicas que buscan el valor del parámetro α que genera una solución con la suavidad (u otra característica) requerida, lo cual abre un amplio abanico de aplicaciones prácticas de los α-splines.

4.6.1. Concepto de suavidad

Como se comentó en la introducción de esta Tesis, los splines cúbicos surgieron originalmente como un modelo matemático que simulara el trazado de curvas a mano alzada, ya que los splines cúbicos imitan la posición de una tira flexible forzada a pasar por un conjunto de puntos dados. La energía de tensión[3] (*strain energy*) de la tira, que es inversamente proporcional a la suavidad de la misma, viene dada por la integral de la curvatura

$$E = \int_b^a \frac{[x^{(2)}(t)]^2}{[1 + (x^{(1)}(t))^2]^{\frac{5}{2}}}\, dt. \tag{4.38}$$

La *elástica* es el spline ideal, es decir, la función $x(t)$ que minimiza (4.38). Debido a la inherente dificultad de trabajar con esta formulación, es habitual utilizar una versión linealizada, y por lo tanto más simple, de la energía de tensión

$$E_L = \int_b^a [x^{(2)}(t)]^2 dt. \tag{4.39}$$

Esta expresión es válida únicamente si se cumple $\left(x^{(1)}(t)\right)^2 \ll 1$. A pesar de que dicha suposición es violada en muchas ocasiones prácticas, su uso está muy extendido, dado que facilita la búsqueda de funciones que minimicen (4.38). En realidad, es la energía linealizada, E_L, dada por (4.39) la que resulta minimizada por la interpolación mediante splines cúbicos (con condiciones de frontera libres)[Boor68].

A continuación, se evalúa la energía de tensión de la interpolación α-spline, demostrando que la suavidad es inversamente proporcional al parámetro α. Se afrontará el problema desde dos puntos de vista: el primero de ellos considerará señales deterministas periódicas, mientras que el segundo tendrá en cuenta señales aleatorias.

[3] Se define la energía de tensión como la energía almacenada en un cuerpo en virtud de una deformación elástica o, de forma equivalente, como el trabajo que debe realizarse para producir cierta deformación elástica.

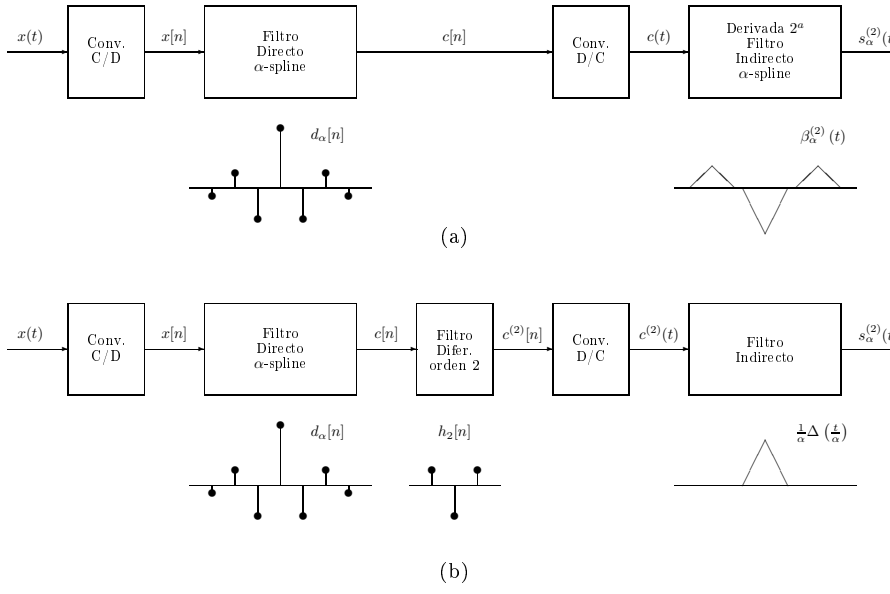


Figura 4.8: Derivada segunda de la interpolación α-spline determinista: a) derivación continua, b) derivación discreta.

4.6.2. Suavidad de la interpolación α-spline de señales deterministas periódicas

Se considera la interpolación de una secuencia de N muestras que representan M periodos de una señal o función senoidal, $x[n]$, con frecuencia discreta $\omega_0 = 2\pi/T_0 = 2\pi M/N$,

$$x[n] = A\sin(\omega_0 n + \Phi_0) = A\sin(2\pi\frac{M}{N}n + \Phi_0), \tag{4.40}$$

siendo $0 \leq M \leq N-1$ y $0 \leq \omega_0 \leq 2\pi$.

La interpolación α-spline de esta secuencia proporciona una señal continua, $s_\alpha(t)$, cuya suavidad es inversamente proporcional a la potencia de tensión, definida en este caso como

$$\mathrm{P}_S\left(s_\alpha(t)\right) = \frac{1}{MT_0}\int_{MT_0}\left(s_\alpha^{(2)}(t)\right)^2 dt. \tag{4.41}$$

Debido a la linealidad e invarianza del procedimiento interpolatorio (véase la Figura 4.8) y considerando condiciones de contorno periódicas (4.20), la

derivada segunda de la interpolación α-spline es

$$s_\alpha^{(2)}(t) = \sum_{n=-\infty}^{+\infty} c[n]\beta_\alpha^{(2)}(t-n) = \sum_{n=-\infty}^{+\infty} c^{(2)}[n]\frac{1}{\alpha}\Delta\left(\frac{t-n}{\alpha}\right), \tag{4.42}$$

donde $\Delta(t)$ representa la función pulso triangular y $c^{(2)}[n]$ es la diferencia segunda de la secuencia discreta de los coeficientes α-spline

$$c^{(2)}[n] = c[n] * h_2[n] = x[n] * d_\alpha[n] * h_2[n],$$

siendo $h_2[n]$ el filtro discreto derivador de segundo orden, cuya respuesta en frecuencia es

$$H_2(\omega) = \mathrm{TF}\left\{\delta[n+1] - 2\delta[n] + \delta[n-1]\right\} = 2\cos(\omega) - 2. \tag{4.43}$$

Puesto que la señal de entrada, $x[n]$, es senoidal, también lo serán la secuencia de coeficientes, $c[n]$, su diferencia segunda, $c^{(2)}[n]$, la señal interpolada, $s_\alpha(t)$, y su segunda derivada, $s_\alpha^{(2)}(t)$. Por ello, se puede escribir

$$\begin{aligned} c^{(2)}[n] &= x[n]\ D_\alpha(z)H_2(z)|_{z=e^{j\omega_0}} \\ &= A\sin(\omega_0 n + \Phi_0)\frac{3}{3+\alpha(\cos\omega_0 - 1)}(2\cos\omega_0 - 2). \end{aligned}$$

Introduciendo ahora (4.42) en (4.41), la potencia de tensión se puede reescribir como

$$\mathrm{P}_S(s_\alpha(t)) = \frac{1}{N}\int_N \left[\sum_{n=1}^{N} c^{(2)}[n]\frac{1}{\alpha}\Delta\left(\frac{t-n}{\alpha}\right)\right]^2 dt.$$

Si el parámetro de suavidad está comprendido en $0 \le \alpha \le 1/2$, entonces los pulsos $\Delta((t-n)/\alpha)$ no se solapan, es decir, toman valores nulos en $t \le n - 1/2$ y $t \ge n + 1/2$. Por lo tanto, el orden del sumatorio y de la integral pueden intercambiarse, resultando

$$\mathrm{P}_S(s_\alpha(t)) = \frac{1}{N}\sum_{n=1}^{N}\left[c^{(2)}[n]\right]^2 \int_{n-1/2}^{n+1/2}\left[\frac{1}{\alpha}\Delta\left(\frac{t-n}{\alpha}\right)\right]^2 dt. \tag{4.44a}$$

Por el contrario, si $1/2 \le \alpha \le 1$ entonces el pulso triangular n-ésimo solapa

con el anterior y el posterior, por lo que

$$\begin{aligned}
\mathrm{P}_S\left(s_\alpha(t)\right) &= \frac{1}{N}\sum_{n=1}^{N}\int_n^{n+1}\left[\frac{c^{(2)}[n]}{\alpha}\Delta\left(\frac{t-n}{\alpha}\right)\frac{c^{(2)}[n+1]}{\alpha}\Delta\left(\frac{t-n-1}{\alpha}\right)\right]^2 dt \\
&= \frac{1}{N}\sum_{n=1}^{N}\int_n^{n+1}\left[\frac{1}{\alpha}c^{(2)}[n]\Delta\left(\frac{t-n}{\alpha}\right)\right]^2 dt \\
&+ \frac{1}{N}\sum_{n=1}^{N}\int_n^{n+1}\left[\frac{1}{\alpha}c^{(2)}[n+1]\Delta\left(\frac{t-n-1}{\alpha}\right)\right]^2 dt \\
&+ \frac{2}{N}\sum_{n=1}^{N}\int_n^{n+1}\frac{1}{\alpha^2}c^{(2)}[n]c^{(2)}[n+1]\Delta\left(\frac{t-n}{\alpha}\right)\Delta\left(\frac{t-n-1}{\alpha}\right) dt.
\end{aligned}$$

Puesto que se consideran condiciones periódicas (en concreto $c^{(2)}[N+1] = c^{(2)}[1]$), y dada la simetría de la función pulso triangular, es posible reescribir la expresión anterior como

$$\begin{aligned}
\mathrm{P}_S\left(s_\alpha(t)\right) &= \frac{1}{N}\sum_{n=1}^{N}\left[c^{(2)}[n]\right]^2\int_n^{n+1}\left[\frac{1}{\alpha}\Delta\left(\frac{t-n}{\alpha}\right)\right]^2 dt \\
&+ \frac{1}{N}\sum_{n=1}^{N}c^{(2)}[n]c^{(2)}[n+1]\int_n^{n+1}\frac{1}{\alpha^2}\Delta\left(\frac{t-n}{\alpha}\right)\Delta\left(\frac{t-n-1}{\alpha}\right) dt.
\end{aligned} \tag{4.44b}$$

Sabiendo que la autocorrelación determinista de la diferencia segunda de los coeficientes α-spline vale

$$\begin{aligned}
\sum_{n=1}^{N}\left(c^{(2)}[n]\right)^2 &= \left(\frac{6A(\cos\omega_0-1)}{3+\alpha(\cos\omega_0-1))}\right)^2\sum_{n=1}^{N}\sin^2\left(\omega_0 n+\Phi_0\right) \\
&= N\frac{A^2}{2}\left(\frac{6(\cos\omega_0-1)}{3+\alpha(\cos\omega_0-1)}\right)^2,
\end{aligned}$$

y

$$\sum_{n=1}^{N}c^{(2)}[n]c^{(2)}[n+1] = N\frac{A^2}{2}\left(\frac{6(\cos\omega_0-1}{3+\alpha(\cos\omega_0-1)}\right)^2\cos\omega_0;$$

y que las integrales que intervienen en (4.44) son

$$I_{\Delta 0} = \int_{-\infty}^{\infty} \left[\Delta\left(\frac{t}{\alpha}\right)\right]^2 dt = \frac{2\alpha}{3}; \tag{4.45a}$$

$$I_{\Delta 1} = \int_{-\infty}^{\infty} \Delta\left(\frac{x-1}{\alpha}\right)\Delta\left(\frac{x}{\alpha}\right) dx = \begin{cases} 0, & 0 \leq \alpha \leq 1/2, \\ \dfrac{(2\alpha-1)^3}{6\alpha^2}, & 1/2 \leq \alpha \leq 1; \end{cases} \tag{4.45b}$$

entonces, la potencia de tensión de la interpolación α-spline (considerando condiciones de frontera periódicas) de una secuencia senoidal de amplitud A y frecuencia discreta ω_0 resulta

$$\mathrm{P}_S\left(s_\alpha(t)\right) = \frac{12A^2}{\alpha}\left(\frac{\cos\omega_0 - 1}{3+\alpha(\cos\omega_0 - 1)}\right)^2 G(\alpha,\omega_0), \tag{4.46a}$$

donde

$$G(\alpha,\omega) = \begin{cases} 1, & 0 \leq \alpha \leq 1/2; \\ 1 + \dfrac{(2\alpha-1)^3}{2\alpha^3}\cos\omega, & 1/2 < \alpha \leq 1. \end{cases} \tag{4.46b}$$

Mediante (4.46), la potencia de tensión de la interpolación de una señal sinusoidal de potencia unidad puede ser evaluada de manera exacta en función de su frecuencia, ω, y del parámetro de suavidad, α, utilizando la expresión

$$\mathrm{P}_S\left(\alpha,\omega\right) = \frac{24}{\alpha}\left(\frac{\cos\omega - 1}{3+\alpha(\cos\omega - 1)}\right)^2 G(\alpha,\omega). \tag{4.47}$$

Esta expresión decrece monótonamente con α, salvo una pequeña región de valores próximos a la frecuencia de Nyquist ($\omega_0 = 2\pi M/N = \pi$) y con el parámetro α comprendido entre 0.5 y 0.67, tal y como muestra la Figura 4.9. Conviene resaltar que, en dicha región, la potencia de tensión crece de forma muy leve, manteniéndose prácticamente constante con respecto a α.

El resultado anterior puede generalizarse para una secuencia $x[n]$ cualquiera. Dado que una secuencia finita siempre puede ser expresada como combinación lineal de exponenciales complejas de distintas frecuencias (mediante su Transformada Discreta de Fourier), y que la potencia de la derivada segunda de la interpolación resulta igual a la suma de las potencias resultantes de interpolar cada una de las exponenciales por separado (debido a la ortogonalidad de las funciones base de dicho desarrollo), entonces es posible evaluar la potencia de la derivada segunda de la interpolación α-spline de cualquier secuencia arbitraria, $x[n]$, de longitud N con condiciones de contorno periódicas. La Transformada Discreta de Fourier (DFT) de dicha secuencia arbitraria viene dada por

$$X[k] = \sum_{n=0}^{N-1} x[n]e^{-j2\pi kn/N},$$

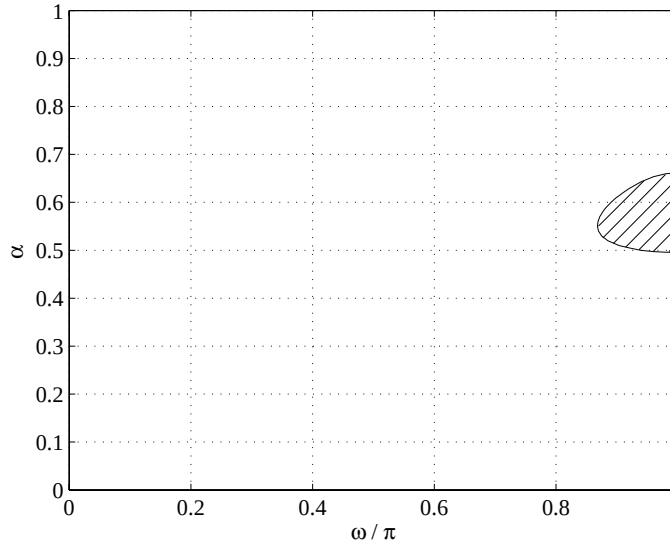


Figura 4.9: Región (área sombreada) donde la suavidad de la interpolación α-spline decrece con α.

donde $\omega = 2\pi k/N$ representa la frecuencia discreta; por lo que, debido a la ortogonalidad de las exponenciales complejas, la potencia de tensión puede ser evaluada a partir de la DFT de la secuencia y de (4.47)

$$\mathrm{P}_S\left(s_\alpha(t)\right) = \frac{1}{N^2}\sum_{k=0}^{N-1}\left|X[k]\right|^2 \mathrm{P}_S\left(\alpha, \omega = 2\pi k/N\right). \tag{4.48}$$

En términos generales, esta potencia de tensión disminuye con α para cualquier señal considerada (salvo para valores extremos de ω_0 próximos a π), es decir, la suavidad de la interpolación aumenta conforme crece α. En la práctica, aunque la secuencia a interpolar posea componentes frecuenciales pertenecientes a la región de no decrecimiento, el pequeño incremento de la potencia de tensión provocado por dichas componentes cuando aumenta α se verá compensado con el mayor decremento provocado por el resto de las componentes frecuenciales de la señal que no pertenecen a dicha región.

4.6.3. Suavidad de la interpolación α-spline de señales aleatorias

En este apartado se obtiene una expresión de la energía de tensión linealizada de la interpolación α-spline considerando procesos estocásticos, $\mathbf{X}(t)$, en

lugar de señales deterministas, $x(t)$. De esta manera, mediante el operador esperanza matemática, la expresión de la energía de tensión (4.39) se transforma en potencia media de tensión

$$\mathrm{P}_S\left(\mathbf{S}_\alpha\right) = E\left[\left(\mathbf{S}_\alpha^{(2)}(t)\right)^2\right] = \sigma^2_{\mathbf{S}_\alpha^{(2)}} + \eta^2_{\mathbf{S}_\alpha^{(2)}} = R_{\mathbf{S}_\alpha^{(2)}}(0),$$

donde $\eta_{\mathbf{S}_\alpha^{(2)}}$, $\sigma^2_{\mathbf{S}_\alpha^{(2)}}$ y $R_{\mathbf{S}_\alpha^{(2)}}(\tau)$, representan, respectivamente, la media, la varianza y la autocorrelación de la derivada segunda de la interpolación α-spline. El interés se centra, por tanto, en el valor de la potencia de la derivada segunda de la interpolación α-spline y en demostrar que, nuevamente, decrece de forma monótona conforme aumenta α.

El esquema de interpolación mostrado en la Figura 4.8 sirve para ilustrar la interpolación de procesos estocásticos sin más que sustituir las señales deterministas (por ejemplo $x(t)$ y $c[n]$) por señales aleatorias ($\mathbf{X}(t)$ y $\mathbf{C}[n]$).

El proceso estocástico constituido por la interpolación α-spline, $\mathbf{S}_\alpha^{(2)}(t)$, no es estacionario en sentido amplio pero, introduciendo una variable aleatoria $\mathbf{\Theta}$ uniformemente distribuida en $[0, T]$[4] e independiente de $\mathbf{S}(t)$, es posible formar el proceso estocástico, $\mathbf{Z}_\alpha(t)$, estacionario en sentido amplio

$$\mathbf{Z}_\alpha^{(2)}(t) = \mathbf{S}_\alpha^{(2)}(t + \mathbf{\Theta}). \tag{4.49}$$

Dado que, tanto la derivación como las operaciones involucradas en el proceso de interpolación (filtrado directo y filtrado indirecto) son lineales e invariantes en el tiempo, es posible expresar el proceso estocástico derivada segunda de la interpolación, $\mathbf{Z}_\alpha^{(2)}(t)$, en forma de convolución

$$\mathbf{Z}_\alpha^{(2)}(t) = \sum_{n=-\infty}^{\infty} \mathbf{C}^{(2)}[n] \frac{1}{\alpha} \Delta\left(\frac{t}{\alpha} - \mathbf{\Theta}\right),$$

donde $\mathbf{C}^{(2)}[n]$ es la secuencia aleatoria formada por la diferencia segunda de los coeficientes de la representación α-spline de $\mathbf{X}(t)$, y $\Delta\,(t)$ representa la función pulso triangular.

Es necesario, en primer lugar, calcular la autocorrelación de la derivada segunda de la interpolación α-spline, $R_{\mathbf{Z}_\alpha^{(2)}}(\tau)$, mediante convolución de la autocorrelación de $\mathbf{C}^{(2)}[n]$ y la autocorrelación determinista de $\Delta\,(t/\alpha)$,

$$R_{\mathbf{Z}_\alpha^{(2)}}(\tau) = \sum_{n=-\infty}^{\infty} R_{\mathbf{C}^{(2)}}[n] \int_{-\infty}^{\infty} \frac{1}{\alpha} \Delta\left(\frac{t - n + \tau}{\alpha}\right) \frac{1}{\alpha} \Delta\left(\frac{t}{\alpha}\right) dt,$$

[4]Como ya se comentó, se considera, sin pérdida de generalidad, un periodo de muestreo $T = 1$.

donde la autocorrelación de los coeficientes α-spline, $R_{\mathbf{C}^{(2)}}[n]$, resulta de filtrar el proceso discreto de entrada, $\mathbf{X}[n]$, con el filtro directo, $d^{\alpha}[n]$, y aplicar el operador diferencia de orden dos (ver Figura 4.8b). Haciendo uso de la autocorrelación determinista de la diferencia segunda del filtro directo α-spline, $\rho_{\alpha}^{(2)}[n]$, dada en (C.2), se tiene

$$R_{\mathbf{C}^{(2)}}[n] = R_{\mathbf{X}}[n] * \rho_{\alpha}^{(2)}[n]. \tag{4.50}$$

Por consiguiente, la potencia de la derivada segunda de la interpolación se corresponderá con su autocorrelación particularizada en $\tau = 0$

$$\mathrm{P}_S\left(\mathbf{Z}_{\alpha}\right) = R_{\mathbf{Z}_{\alpha}^{(2)}}(0) = \frac{1}{\alpha^2} \sum_{n=-\infty}^{\infty} R_{\mathbf{C}^{(2)}}[n] \int_{-\infty}^{\infty} \Delta\left(\frac{t-n}{\alpha}\right) \Delta\left(\frac{t}{\alpha}\right) dt,$$

y puesto que la integral únicamente toma valores no nulos para $n = -1, 0, 1$, entonces se puede escribir

$$\begin{aligned}
\mathrm{P}_S\left(\mathbf{Z}_{\alpha}\right) &= \frac{1}{\alpha^2} R_{\mathbf{C}^{(2)}}[-1] \int_{-\infty}^{\infty} \Delta\left(\frac{t+1}{\alpha}\right) \Delta\left(\frac{t}{\alpha}\right) dt \\
&+ \frac{1}{\alpha^2} R_{\mathbf{C}^{(2)}}[0] \int_{-\infty}^{\infty} \Delta^2\left(\frac{t}{\alpha}\right) dt \\
&+ \frac{1}{\alpha^2} R_{\mathbf{C}^{(2)}}[+1] \int_{-\infty}^{\infty} \Delta\left(\frac{t-1}{\alpha}\right) \Delta\left(\frac{t}{\alpha}\right) dt \\
&= \frac{1}{\alpha^2} \left[R_{\mathbf{C}^{(2)}}[0] I_{\Delta 0} + \left(R_{\mathbf{C}^{(2)}}[-1] + R_{\mathbf{C}^{(2)}}[+1]\right) I_{\Delta 1}\right].
\end{aligned}$$

Finalmente, se obtiene

$$\mathrm{P}_S\left(\mathbf{Z}_{\alpha}\right) = \frac{1}{\alpha^2} \left(R_{\mathbf{C}^{(2)}}[0] I_{\Delta 0} + 2 R_{\mathbf{C}^{(2)}}[1] I_{\Delta 1}\right), \tag{4.51}$$

donde los términos no nulos de las integrales toman el valor dado en (4.45).

En definitiva, es posible calcular la potencia de la derivada segunda de la interpolación α-spline a partir de la autocorrelación de la secuencia aleatoria a interpolar, $R_{\mathbf{X}}[n]$, sin más que hacer uso de la expresion (4.51) junto con (4.50). En posteriores secciones se llevará a cabo dicho cálculo particularizado para señales con una autocorrelación determinada. Sería posible realizar estos cálculos en el dominio de la frecuencia, aunque, salvo casos triviales, resulta más complicado que el procedimiento descrito.

4.6.3.1. Señales blancas

Si el proceso a interpolar, $\mathbf{X}[n]$, es blanco de media nula, es decir, si su autocorrelación es de la forma

$$R_{\mathbf{X}}[n] = \sigma_{\mathbf{X}}^2 \delta[n],$$

entonces, usando (4.50), la autocorrelación de la diferencia segunda de los coeficientes α-spline es

$$R_{\mathbf{C}^{(2)}}[n] = \sigma_{\mathbf{X}}^2 \rho_\alpha^{(2)}[n],$$

con lo que (4.51) resulta

$$\mathrm{P}_S\left(\mathbf{Z}_\alpha\right) = \frac{\sigma_{\mathbf{X}}^2}{\alpha^2}\left(\rho_\alpha^{(2)}[0] I_{\Delta 0} + 2\rho_\alpha^{(2)}[1] I_{\Delta 1}\right). \tag{4.52}$$

Este valor depende exclusivamente de la varianza de la entrada, $\sigma_{\mathbf{X}}^2$, y del parámetro de suavidad α a través de las expresiones de $I_{\Delta n}$ y de $\rho_\alpha^{(2)}[n]$ dadas en (4.45) y (C.2) respectivamente. Desarrollando dichas expresiones, se obtiene un resultado más simple expresando (4.52) en función del polo z_1 en lugar de en función del parámetro α. Así resulta que la potencia de la interpolación α-spline de una secuencia aleatoria blanca toma el valor

- si $0 \leq \alpha \leq 1/2$, entonces

$$\mathrm{P}_S\left(\mathbf{Z}_\alpha\right) = -\sigma_{\mathbf{X}}^2 \frac{2(z_1+3)(z_1-1)^6}{9z_1(z_1+1)^3}; \tag{4.53a}$$

- si $1/2 \leq \alpha \leq 1$, entonces

$$\begin{aligned}\mathrm{P}_S\left(\mathbf{Z}_\alpha\right) = -\sigma_{\mathbf{X}}^2 &\frac{(z_1-1)^6}{1944 z_1^4 (z_1+1)^3} \\ &\cdot(z_1^8 + 31z_1^7 + 331z_1^6 + 1303z_1^5 + 1189z_1^4 - 491z_1^3 - 575z_1^2 - 59z_1 - 2).\end{aligned} \tag{4.53b}$$

Se comprueba fácilmente que la derivada de la potencia de tensión es positiva para todos los posibles valores de z_1, es decir,

$$\frac{d\mathrm{P}_S\left(\mathbf{Z}_\alpha\right)}{dz_1} \geq 0, \quad \forall z_1 \in [-2+\sqrt{3}, 0];$$

y puesto que existe una relación monótona decreciente entre el parámetro α y el polo z_1 (4.26)

$$\frac{d\alpha}{dz_1} = 6\frac{z_1+1}{(z_1-1)^3} \leq 0, \quad \forall z_1 \in [-2+\sqrt{3}, 0], \tag{4.54}$$

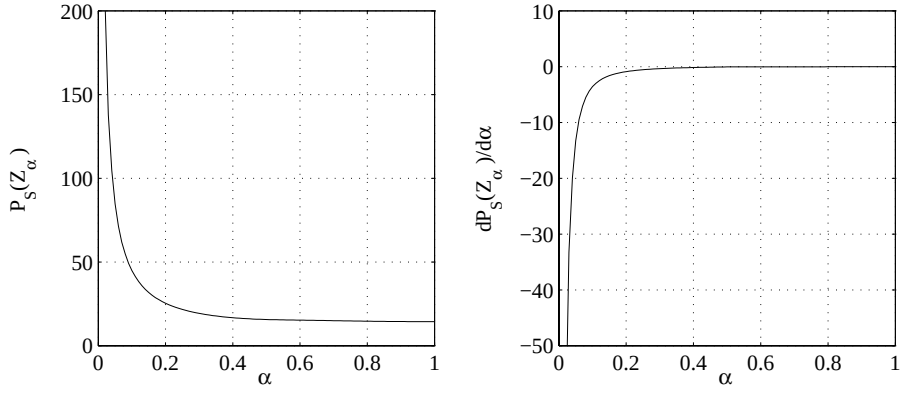


Figura 4.10: Potencia de tensión de la interpolación α-spline de señales blancas de potencia unidad y su derivada frente a α.

entonces es posible afirmar que la potencia de tensión decrece conforme crece el valor de α, puesto que

$$\frac{d\mathrm{P}_S\left(\mathbf{Z}_\alpha\right)}{d\alpha} =\leq 0, \quad \forall\alpha \in [0,1].$$

En la Figura 4.10 se muestra la dependencia de la potencia de la derivada segunda de la interpolación α-spline de una función blanca. Se concluye, por tanto, que la suavidad de la interpolación α-spline de una secuencia aleatoria blanca decrece conforme disminuye el valor de z_1, es decir, crece conforme aumenta el parámetro de suavidad α.

4.6.3.2. Señales aleatorias senoidales

Si el proceso a interpolar es senoidal de frecuencia ω_0 y fase aleatoria, su autocorrelación se puede escribir como

$$R_{\mathbf{X}}[n] = \sigma_{\mathbf{X}}^2 \cos\left(\omega_0 n\right), \quad \forall k \in \mathbb{Z},\ 0 \leq \omega \leq \pi.$$

Dado que la respuesta en frecuencia del filtro directo α-spline es (4.36)

$$\left|D_\alpha(\omega)\right|^2 = \frac{9}{[3+\alpha(\cos\omega - 1)]^2},$$

y que la respuesta en frecuencia del filtro diferencia segunda es

$$\left|H_2(\omega)\right|^2 = \left|TF\left\{\delta[n+1] - 2\delta[n] + \delta[n-1]\right\}\right|^2 = (2\cos\omega - 2)^2,$$

entonces

$$\begin{aligned} R_{\mathbf{C}^{(2)}}[k] &= \sigma_{\mathbf{X}}^2 \cos(\omega_0 n) \left|D_\alpha(\omega_0)\right|^2 \left|H_2(\omega_0)\right|^2 \\ &= \sigma_{\mathbf{X}}^2 \cos(\omega_0 n) \left(\frac{3}{3+\alpha(\cos\omega_0 - 1)}\right)^2 (2\cos\omega_0 - 2)^2 ; \end{aligned}$$

con lo que, aplicando (4.51), la potencia de la derivada segunda de la interpolación α-spline de una señal senoidal se puede expresar como

$$\begin{aligned} \mathrm{P}_S(\mathbf{Z}_\alpha) &= \frac{1}{\alpha^2}\left(R_{C^{(2)}}[0] I_{\Delta 0} + 2R_{C^{(2)}}[1] I_{\Delta 1}\right) \\ &= \frac{\sigma_{\mathbf{X}}^2}{\alpha^2}\left(\frac{6(\cos\omega_0 - 1)}{3+\alpha(\cos\omega_0 - 1)}\right)^2 \left(I_{\Delta 0} + 2\cos\omega_0 I_{\Delta 1}\right). \end{aligned}$$

Finalmente, considerando (4.45), la anterior expresión resulta

$$\mathrm{P}_S(\mathbf{Z}_\alpha) = \sigma_{\mathbf{X}}^2 \frac{24}{\alpha^2}\left(\frac{\cos\omega_0 - 1}{3+\alpha(\cos\omega_0 - 1)}\right)^2 G(\alpha, \omega_0), \tag{4.55}$$

expresión idéntica a (4.46) referente a la potencia de tensión de señales deterministas senoidales. Por consiguiente, al igual que ocurría en el caso determinista, es posible afirmar que la suavidad de la interpolación α-spline de procesos estocásticos senoidales aumenta conforme crece el valor de α salvo en la pequeña región del plano $\alpha - \omega$, representada en Figura 4.9, en la que la suavidad es prácticamente constante con respecto a α.

4.7. Varianza de la interpolación α-spline

En determinadas ocasiones, las técnicas de interpolación generan oscilaciones y sobredisparos en la solución. Existen diversas medidas que se pueden obtener con el fin de cuantificar numéricamente estos efectos, siendo la varianza y el área residual dos de las más significativas. En esta sección se calculará la varianza de la interpolación α-spline y su dependencia con el parámetro α considerando como entrada secuencias de duración finita y procesos estocásticos, particularizando, en este último caso, para señales aleatorias blancas y sinusoidales.

Proposición 4.3 *Sea $x[n]$ una secuencia finita de N muestras $1 \leq n \leq N$, con transformada discreta de Fourier (DFT) $X[k]$, la interpolación α-spline con condiciones de frontera periódicas de dicha secuencia proporciona una señal*

$$s_\alpha(t) = \sum_{n=1}^{N} c[n]\beta_\alpha(t-n),$$

cuya varianza es

$$Var\left(s_\alpha(t)\right) = \frac{1}{N^2}\sum_{k=1}^{N-1}\left|X[k]\right|^2 Var\left(\alpha, \omega = 2\pi k/N\right), \tag{4.56}$$

siendo

$$Var\left(\alpha,\omega\right) = \frac{9\left(I_0 + 2\cos\left(\omega\right)I_1 + 2\cos\left(2\omega\right)I_2 + 2\cos\left(3\omega\right)I_3\right)}{\left[3+\alpha(\cos\omega - 1)\right]^2}, \tag{4.57}$$

y donde la integrales

$$I_n = \int_{-\infty}^{\infty}\beta_\alpha\left(x\right)\beta_\alpha\left(x-n\right)dx, \tag{4.58}$$

son polinomios del parámetro de suavidad α

$$\begin{aligned}
I_0 &= \begin{cases} \frac{31\alpha^3-70\alpha^2+140}{210}, & 0\le\alpha\le\frac{1}{2};\\ \frac{-35\alpha^7+238\alpha^6-672\alpha^5+980\alpha^4-280\alpha^3+84\alpha^2-14\alpha+1}{630\alpha^4}, & \frac{1}{2}\le\alpha\le 1;\end{cases}\\
I_1 &= \begin{cases} \frac{-62\alpha^3+105\alpha^2+105}{630}, & 0\le\alpha\le\frac{1}{2};\\ \frac{400\alpha^7-2296\alpha^6+4704\alpha^5-3080\alpha^4+1960\alpha^3-588\alpha^2+98\alpha-7}{5040\alpha^4}, & \frac{1}{2}\le\alpha\le 1;\end{cases}\\
I_2 &= \begin{cases} \frac{31\alpha^3}{1260}, & 0\le\alpha\le\frac{1}{2};\\ \frac{-97\alpha^7+448\alpha^6-672\alpha^5+560\alpha^4-280\alpha^3+84\alpha^2-14\alpha+1}{1260\alpha^4}, & \frac{1}{2}\le\alpha\le 1;\end{cases}\\
I_3 &= \begin{cases} 0, & 0\le\alpha\le\frac{1}{2};\\ \frac{(2\alpha-1)^7}{5040\alpha^4}, & \frac{1}{2}\le\alpha\le 1.\end{cases}
\end{aligned}$$

Demostración 4.3 *La demostración se encuentra en el Apéndice C.2.*

Proposición 4.4 *Sea el proceso estocástico discreto,* $\mathbf{X}[n]$*, estacionario y con autocorrelación* $R_{\mathbf{X}}[n]$*, su interpolación* α*-spline con condiciones de frontera periódicas proporciona una señal*

$$\mathbf{S}_\alpha(t) = \sum_{n=-\infty}^{+\infty}\mathbf{C}[n]\beta_\alpha\left(t-n\right),$$

cuya varianza es

$$Var\left(\mathbf{S}_\alpha\right) = R_{\mathbf{C}}[0]I_0 + 2R_{\mathbf{C}}[1]I_1 + 2R_{\mathbf{C}}[2]I_2 + 2R_{\mathbf{C}}[3]I_3, \tag{4.59}$$

donde el valor de las integrales I_n *viene dado por (4.58) y la autocorrelación de los coeficientes* α*-spline es*

$$R_{\mathbf{C}}[n] = R_{\mathbf{X}}[n] * \rho_\alpha[n], \tag{4.60}$$

siendo $\rho_\alpha[n]$ la autocorrelación determinista del filtro $d_\alpha[n]$,

$$\rho_\alpha[n] = \frac{(1-z_1)z_1^{|n|}(|n|+1+(1-|n|)z_1^2)}{(z_1+1)^3}.$$

Demostración 4.4 *La demostración se encuentra en el Apéndice C.3.*

Proposición 4.5 *Sea el proceso estocástico discreto, $\mathbf{X}[n]$, estacionario y blanco con autocorrelación $R_{\mathbf{X}}[n] = \sigma_{\mathbf{X}}^2\delta[n]$, la interpolación α-spline con condiciones de frontera periódicas de dicha secuencia aleatoria proporciona una señal cuya varianza es*

- *si $0 \leq \alpha \leq \frac{1}{2}$, entonces*

$$Var(\mathbf{S}_\alpha) = -2\sigma_{\mathbf{X}}^2 \frac{35z_1^5 + 488z_1^4 + 1079z_1^3 + 595z_1^2 + 70z_1 - 35}{105(z_1-1)^2(z_1+1)^3}; \quad (4.61a)$$

- *si $\frac{1}{2} \leq \alpha \leq 1$, entonces*

$$\begin{aligned} Var(\mathbf{S}_\alpha) = \frac{-\sigma_{\mathbf{X}}^2}{1632960z_1^4(z_1-1)^2(z_1+1)^3} \big[& z_1^{16} + 71z_1^{15} + 2175z_1^{14} \\ & + 37387z_1^{13} + 391727z_1^{12} + 2530731z_1^{11} + 9591043z_1^{10} \\ & + 19088975z_1^9 + 21519117z_1^8 + 15050501z_1^7 + 4966381z_1^6 \\ & -1908879z_1^5 - 1772155z_1^4 - 68663z_1^3 - 4143z_1^2 - 139z_1 - 2 \big]. \end{aligned} \quad (4.61b)$$

donde z_1 (4.26) es el polo del filtro directo α-spline.

Demostración 4.5 *La demostración se encuentra en el Apéndice C.3.1.*

Proposición 4.6 *Sea el proceso estocástico discreto, $\mathbf{X}[n]$, estacionario senoidal de frecuencia ω_0 y fase aleatoria con autocorrelación*

$$R_X[n] = \sigma_{\mathbf{X}}^2 \cos(\omega_0 n) \quad \forall n \in \mathbb{Z},\ 0 \leq \omega_0 \leq \pi,$$

la interpolación α-spline con condiciones de frontera periódicas de dicha secuencia aleatoria proporciona una señal cuya varianza es

$$Var(\mathbf{S}_\alpha) = \sigma_{\mathbf{X}}^2 \frac{9\left(I_0 + 2\cos(\omega_0)I_1 + 2\cos(2\omega_0)I_2 + 2\cos(3\omega_0)I_3\right)}{[3+\alpha(\cos\omega_0 - 1)]^2}, \quad (4.62)$$

donde el valor de I_n viene dado por (4.58) .

Demostración 4.6 *La demostración se encuentra en el Apéndice C.3.2.*

4.8. Área residual de la interpolación α-spline

El área residual es una medida de cuánto se aparta la función interpolatoria de la solución lineal a tramos. En esta sección se obtendrá una expresión de la dependencia de dicha medida con el parámetro α.

El área residual, $\mathrm{A}_r\left(x(t)\right)$, de una señal $x(t)$ se calcula como la integral del cuadrado de la diferencia entre la propia función y la interpolación lineal a tramos correspondiente, es decir,

$$\mathrm{A}_r\left(x(t)\right) = \sum_n \int_n^{n+1} \left[x(t) - L_n(t)\right]^2 dt, \tag{4.63}$$

donde $L_n(t)$ es la expresión de la recta que une los puntos $x(n)$ y $x(n+1)$.

La utilización de un criterio basado en el área residual fue presentado en [Kashyap96] con el fin de evitar el problema del sobredisparo provocado por la interpolación spline cúbica. Básicamente, el procedimiento propuesto en dicho artículo consiste en utilizar splines de orden mayor que el cúbico con el fin de disponer de parámetros libres que son utilizados para minimizar el área residual.

Dado que la familia α-spline propuesta en esta Tesis permite obtener soluciones intermedias entre la solución interpolatoria con área residual nula (interpolación lineal a tramos: α-spline con $\alpha = 0$) y la solución más suave (interpolación cúbica: α-spline con $\alpha = 1$) resulta de interés conocer cuál es la dependencia del área residual de la interpolación con el parámetro de control α para distintos tipos de señales de entrada, tanto deterministas como aleatorias.

Proposición 4.7 *Sea $x[n]$ una secuencia finita de N muestras $1 \leq n \leq N$, con transformada discreta de Fourier (DFT), $X[k]$ para $k = 0, \ldots, N-1$, la interpolación α-spline con condiciones de frontera periódicas de dicha secuencia proporciona una señal*

$$s_\alpha(t) = \sum_{n=1}^{N} c[n]\beta_\alpha\left(t - n\right),$$

cuya área residual es

$$A_r\left(s_\alpha(t)\right) = \frac{1}{N^2} \sum_{k=1}^{N-1} \left|X[k]\right|^2 A_r\left(\alpha, \omega = 2\pi k/N\right), \tag{4.64}$$

siendo

$$\begin{aligned} A_r\left(\alpha, \omega\right) = & \frac{9\left(I_0 + 2\cos\left(\omega\right)I_1 + 2\cos\left(2\omega\right)I_2 + 2\cos\left(3\omega\right)I_3\right)}{\left[3 + \alpha(\cos\omega - 1)\right]^2} \\ & - \frac{6\left(I_{A0} + 2\cos\left(\omega\right)I_{A1} + 2\cos\left(2\omega\right)I_{A2}\right)}{3 + \alpha(\cos\omega - 1)} + \frac{2 + \cos\omega_0}{3}, \end{aligned}$$

donde I_n viene dado por (4.58) e I_{An} representa

$$I_{An} = \int_{-\infty}^{\infty} \beta_\alpha(x)\,\Delta(x-n)\,dx = I_{A-n}, \tag{4.65}$$

cuyos valores no nulos son

$$\begin{aligned} I_{A0} &= \frac{3\alpha^3 - 10\alpha^2 + 40}{60}, \\ I_{A1} &= \frac{-2\alpha^3 + 5\alpha^2 + 10}{60}, \\ I_{A2} &= \frac{\alpha^3}{120}. \end{aligned}$$

Demostración 4.7 *La demostración se encuentra en el Apéndice C.4.*

Proposición 4.8 *Sea el proceso estocástico discreto, $\mathbf{X}[n]$, estacionario y con autocorrelación $R_{\mathbf{X}}[n]$, su interpolación α-spline con condiciones de frontera periódicas proporciona una señal*

$$\mathbf{S}_\alpha(t) = \sum_{n=-\infty}^{+\infty} \mathbf{C}[n]\beta_\alpha(t-n),$$

cuya área residual es

$$A_r(\mathbf{S}_\alpha) = Var(\mathbf{S}_\alpha) + Var(\mathbf{S}_\alpha)|_{\alpha=0} - 2\,R_{\mathbf{X}}[n] * \eta_\alpha(\tau) * \Delta(\tau)|_{\tau=0}, \tag{4.66}$$

donde $\eta_\alpha(\tau)$ es el α-spline cardinal (4.35) y las varianzas de la interpolación α-spline se calculan según la Proposición 4.4.

Demostración 4.8 *La demostración se encuentra en el Apéndice C.5.*

Proposición 4.9 *Sea el proceso estocástico discreto, $\mathbf{X}[n]$, estacionario y blanco con autocorrelación $R_{\mathbf{X}}[n] = \sigma^2_{\mathbf{X}}\delta[n]$, la interpolación α-spline con condiciones de frontera periódicas de dicha secuencia aleatoria proporciona una señal cuya área residual es*

- *si $0 \le \alpha \le \frac{1}{2}$, entonces*

$$A_r(\mathbf{S}_\alpha) = 2\sigma^2_{\mathbf{X}} z_1^2 \frac{35z_1^6 - 35z_1^5 - 376z_1^4 + 380z_1^3 + 1985z_1^2 + 895z_1 + 140}{105(z_1-1)^4(z_1+1)^3}; \tag{4.67a}$$

- *si* $\frac{1}{2} \leq \alpha \leq 1$, *entonces*

$$\begin{aligned} A_r(\mathbf{S}_\alpha) = & \frac{-\sigma_{\mathbf{X}}^2}{1632960 z_1^4 (z_1-1)^4 (z_1+1)^3} \left[z_1^{18} + 69 z_1^{17} + 2034 z_1^{16} \right. \\ & + 33108 z_1^{15} + 319128 z_1^{14} + 1784664 z_1^{13} + 3832668 z_1^{12} \\ & + 2437620 z_1^{11} - 8374158 z_1^{10} - 25010630 z_1^9 - 31920144 z_1^8 \\ & - 23353956 z_1^7 - 10406256 z_1^6 - 2787792 z_1^5 - 550332 z_1^4 \\ & \left. -60516 z_1^3 - 3867 z_1^2 - 135 z_1 - 2 \right] ; \end{aligned} \tag{4.67b}$$

donde z_1 *(4.26) es el polo del filtro directo* α*-spline.*

Demostración 4.9 *La demostración se encuentra en el Apéndice C.5.1.*

Proposición 4.10 *Sea el proceso estocástico discreto,* $\mathbf{X}[n]$*, estacionario senoidal de frecuencia* ω_0 *y fase aleatoria con autocorrelación*

$$R_X[n] = \sigma_{\mathbf{X}}^2 \cos(\omega_0 n) \quad \forall n \in \mathbb{Z},\ 0 \leq \omega_0 \leq \pi,$$

la interpolación α*-spline con condiciones de frontera periódicas de dicha secuencia aleatoria proporciona una señal cuya área residual es*

$$\begin{aligned} A_r(\mathbf{S}_\alpha) = & \sigma_{\mathbf{X}}^2 \frac{9\left(I_0 + 2\cos(\omega_0) I_1 + 2\cos(2\omega_0) I_2 + 2\cos(3\omega_0) I_3\right)}{[3+\alpha(\cos\omega_0 - 1)]^2} \\ & - \sigma_{\mathbf{X}}^2 \frac{6\left(I_{A0} + 2\cos(\omega_0) I_{A1} + 2\cos(2\omega_0) I_{A2}\right)}{3+\alpha(\cos\omega_0 - 1)} + \sigma_{\mathbf{X}}^2 \frac{2+\cos\omega_0}{3}, \end{aligned}$$

donde las integrales I_n *e* I_{An} *son polinomios en* α *dados por (4.58) y (4.65) respectivamente.*

Demostración 4.10 *La demostración se encuentra en el Apéndice C.5.2.*

4.9. Comportamiento de las características de la interpolación

Las expresiones presentadas en esta Tesis relativas a la suavidad, varianza y área residual de la interpolación α-spline permiten estudiar su comportamiento respecto al parámetro α. Así, cuando la señal a interpolar es blanca, la potencia de tensión (4.53) (el inverso de la suavidad) decrece monótonamente con α, mientras que la varianza (4.61) y el área residual (4.67) crecen monótonamente. Este comportamiento monótono se muestra en la Figura 4.11, en la que

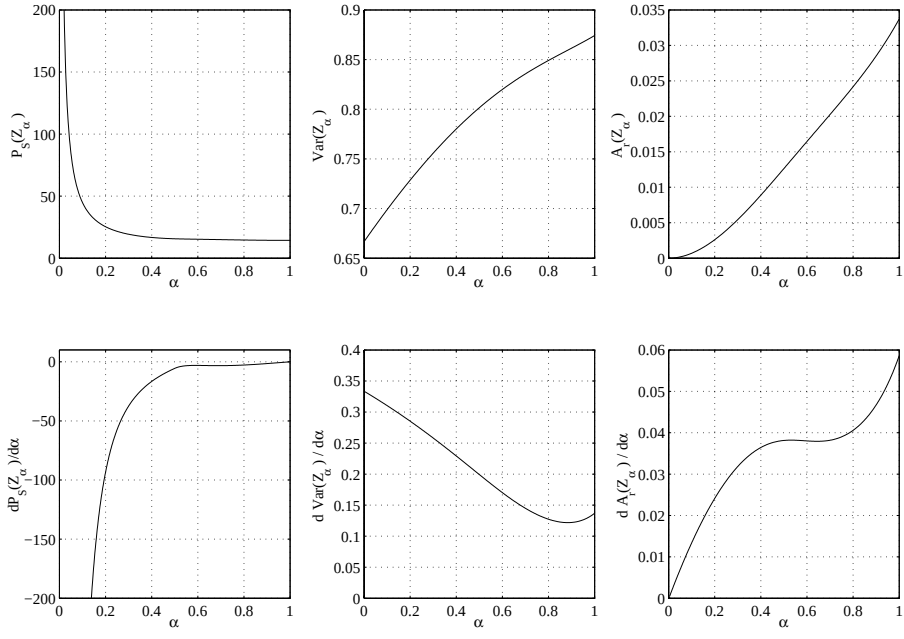


Figura 4.11: Comportamiento respecto a α de la potencia de tensión (izda.), varianza (centro), área residual (dcha.) y las correspondientes derivadas (abajo) de la interpolación α-spline de señales blancas de varianza unidad.

simplemente se dibuja el valor de las tres expresiones para una señal de entrada de varianza unitaria.

En lo que respecta a señales deterministas, el comportamiento de suavidad, varianza y área residual respecto a α dependerá de cada señal concreta. Con el fin de considerar un caso lo más genérico y representativo posible, en esta Tesis se ha estudiado el comportamiento de secuencias senoidales, pues posibilita extender los resultados en base a la transformada de Fourier de la secuencia a interpolar, tal y como se describe en el Apartado 4.6.2, en la Proposición 4.3 y en la Proposición 4.7. Particularizando estos resultados, si se considera una secuencia senoidal de frecuencia $\omega \in [-\pi, \pi]$ y fase arbitraria se comprueba que tanto la suavidad, como la varianza y el área residual de la interpolación α-spline crecen monótonamente respecto a α para cualquier valor de frecuencia considerado. Únicamente para valores de la frecuencia discreta muy próximos π y ciertos valores de α comprendidos en $[0{,}5, 1]$ (representados en la Figura 4.12) las características estudiadas se tornan decrecientes, aunque prácticamente constantes.

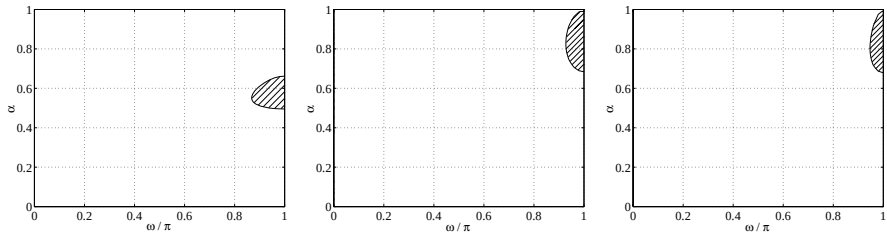


Figura 4.12: Regiones (áreas sombreadas) donde la suavidad (izquierda), la varianza (centro) y el área residual (derecha) de la interpolación α-spline decrecen con α.

4.10. Generalización para muestreo no uniforme

Hasta el momento, en esta Tesis se ha presentado una nueva familia de splines, los α-splines, considerando el muestreo uniforme de la señal o función a interpolar. En esta sección se generalizará la familia α-spline para el caso en el que las muestras no se encuentren equiespaciadas.

4.10.1. Base α-spline no uniforme

Sea un vector de cinco nodos (*knots*) no decrecientes $[t_0, t_1, t_2, t_3, t_4]$; el objetivo de este apartado consiste en encontrar la correspondiente base α-spline. Para ello se impondrán una serie de condiciones que permitan obtener su expresión analítica. La base así obtenida deberá poseer una derivada segunda continua y lineal a tramos, por lo que la derivada tercera será de tipo escalonada.

En el caso no uniforme, la derivada tercera de la base α-spline toma la forma

$$\begin{aligned}\beta_\alpha^{(3)}(t) = {} & Ap_\alpha\left(\frac{t+\alpha/2}{t_1-t_0}-t_1\right) - Bp_\alpha\left(\frac{t-\alpha/2}{t_2-t_1}-t_1\right) \\ & - Cp_\alpha\left(\frac{t+\alpha/2}{t_2-t_1}-t_2\right) + Dp_\alpha\left(\frac{t-\alpha/2}{t_3-t_2}-t_2\right) \\ & + Ep_\alpha\left(\frac{t+\alpha/2}{t_3-t_2}-t_3\right) - Fp_\alpha\left(\frac{t-\alpha/2}{t_4-t_3}-t_3\right),\end{aligned} \tag{4.68}$$

donde los seis valores A, B, C, D, E y F son dependientes de los nodos t_0, t_1, t_2, t_3, t_4 y del parámetro α. Para determinar el valor de dichas constantes es preciso imponer ciertas condiciones, que se derivan directamente de las características propias de la base α-spline. En la Figura 4.13 se representan, a modo de ejemplo, la base α-spline, $\beta_\alpha(t)$, y su derivada tercera, $\beta_\alpha^{(3)}(t)$, correspondientes al vector de nodos $[-2, -0{,}4, 0{,}8, 1{,}4, 2]$ para $\alpha = 0{,}2$ y $\alpha = 0{,}7$.

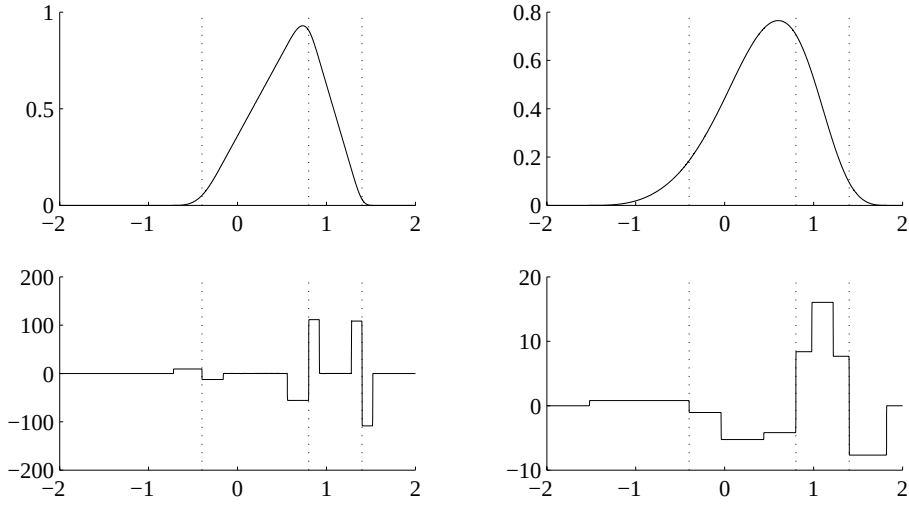


Figura 4.13: Base α-spline (arriba) y su tercera derivada (abajo) correspondientes al vector de nodos $[-2, -0{,}4, 0{,}8, 1{,}4, 2]$ para $\alpha = 0{,}2$ (izda.) y $\alpha = 0{,}7$ (dcha.).

En concreto, de la continuidad de la segunda derivada se extraen tres condiciones:

$$A(t_1 - t_0) - B(t_2 - t_1) = 0, \tag{4.69a}$$

$$C(t_2 - t_1) - D(t_3 - t_2) = 0, \tag{4.69b}$$

$$E(t_3 - t_2) - F(t_4 - t_3) = 0; \tag{4.69c}$$

de la continuidad de la derivada primera

$$A(t_1-t_0)^2+B(t_2-t_1)^2-C(t_2-t_1)^2-D(t_3-t_2)^2+E(t_3-t_2)^2+F(t_4-t_3)^2 = 0; \tag{4.69d}$$

mientras que la continuidad de la base α-spline implica

$$\begin{aligned} &A\left[\alpha/3(t_1 - t_0) + (t_2 - t_1)\right](t_1 - t_0)^2 \\ &+B(1 - \alpha/3)(t_2 - t_1)^3 - C\alpha/3(t_2 - t_1)^3 \\ &+D\alpha/3(t_3 - t_2)^3 - E(1 - \alpha/3)(t_3 - t_2)^3 \\ &-F\left[\alpha/3(t_4 - t_3) + (t_3 - t_2)\right](t_4 - t_3)^2 = 0. \end{aligned} \tag{4.69e}$$

Por último, se hace necesaria una sexta condición que posibilite la obtención de los seis valores buscados. Dicha condición viene impuesta por la necesidad de

que las bases α-spline constituyan partición de unidad para cualquier conjunto de nodos considerado

$$\sum_n \beta_{n,\alpha}(t - t_n) = 1. \tag{4.69f}$$

Dado el sistema de seis ecuaciones y seis incógnitas definido por (4.69) y considerando la distancia entre nodos

$$h_n = t_n - t_{n-1}, \quad n \in \{1, 2, 3, 4\};$$

se obtienen los siguientes valores:

$$A = \frac{6}{\alpha^2 h_1(h_1 + h_2)(\alpha h_1 + (3 - 2\alpha)h_2 + \alpha h_3)}, \tag{4.70a}$$

$$B = \frac{6}{\alpha^2 h_2(h_1 + h_2)(\alpha h_1 + (3 - 2\alpha)h_2 + \alpha h_3)}, \tag{4.70b}$$

$$C = \frac{6[\alpha(h_1 + h_4) + (3 - \alpha)(h_2 + h_3)]}{\alpha^2 h_2(h_2 + h_3)(\alpha h_1 + (3 - 2\alpha)h_2 + \alpha h_3)(\alpha h_2 + (3 - 2\alpha)h_3 + \alpha h_4)}, \tag{4.70c}$$

$$D = \frac{6[\alpha(h_1 + h_4) + (3 - \alpha)(h_2 + h_3)]}{\alpha^2 h_3(h_2 + h_3), (\alpha h_2 + (3 - 2\alpha)h_2 + \alpha h_3)(\alpha h_2 + (3 - 2\alpha)h_3 + \alpha h_4)}, \tag{4.70d}$$

$$E = \frac{6}{\alpha^2 h_3(h_3 + h_4)(\alpha h_2 + (3 - 2\alpha)h_3 + \alpha h_4)}, \tag{4.70e}$$

$$F = \frac{6}{\alpha^2 h_4(h_3 + h_4)(\alpha h_2 + (3 - 2\alpha)h_3 + \alpha h_4)}. \tag{4.70f}$$

Sustituyendo los valores de (4.70) en (4.68) e integrando tres veces se obtienen las expresiones de la base α-spline, $\beta_\alpha(x)$, que aparecen en el Apéndice D.1.

4.10.2. Interpolación α-spline no uniforme

Para obtener la representación α-spline, $s_\alpha(t)$, de una señal $x(t)$ en los nodos $t_1, t_2, \ldots, t_N$ son necesarias N bases α-spline (cada una de ellas centrada en cada uno de los N nodos), para lo cual es preciso definir dos nodos adicionales a ambos lados del intervalo de interpolación: t_{-1}, t_0 y t_{N+1}, t_{N+2}; de manera que dicha representación α-spline se puede escribir como

$$s_\alpha(t) = \sum_{n=1}^{N} c[n]\beta_{n,\alpha}(t),$$

donde $\beta_{n,\alpha}(t)$ es la base correspondiente al vector de nodos $[t_{n-2}, \ldots, t_{n+2}]$ y $c[n]$ es el coeficiente α-spline que posibilita el cumplimiento de la condición de

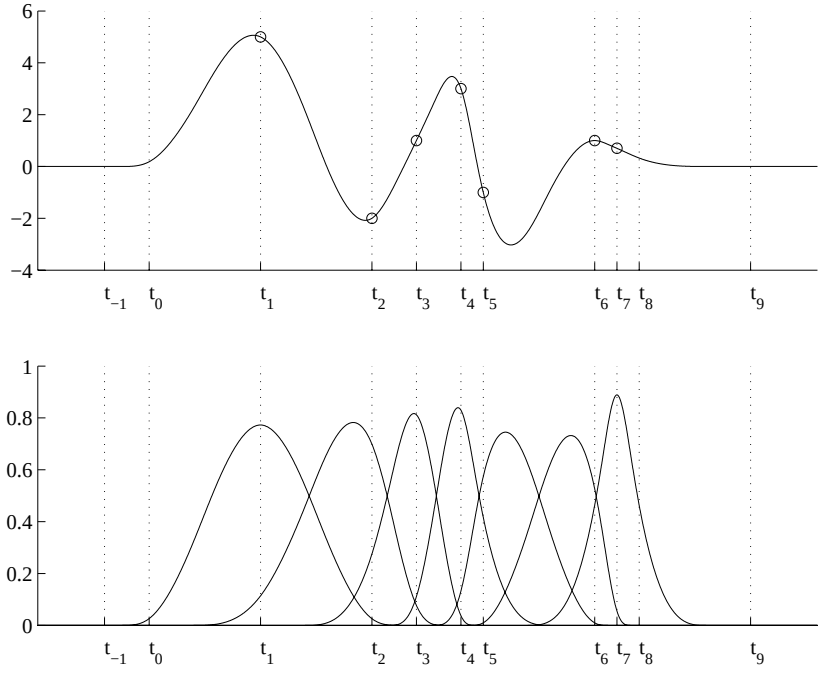


Figura 4.14: Interpolación α-spline (arriba) correspondiente al vector de nodos $t_1, \ldots, t_N$ y bases α-spline (abajo) correspondientes al conjunto de nodos $t_{-1}, \ldots t_{N+2}$.

interpolación en los nodos

$$s_\alpha(t_n) = x(t_n), \quad \forall\, n = 1, 2, \ldots, N. \tag{4.71}$$

En la Figura 4.14 se representa la interpolación α-spline de $N = 7$ muestras en los instantes $t_1, t_2, \ldots, t_7$ y las siete bases α-spline centradas en dichos nodos. Se puede observar cómo es preciso introducir los nodos t_{-1} y t_0 para poder definir la primera función base, $\beta_{1,\alpha}(t)$, cuyo soporte es $t_{-1} \geq t \geq t_3$. De igual manera, $\beta_{N,\alpha}(t)$ precisa de los nodos adicionales t_8 y t_9.

El valor de la base α-spline en los nodos se obtiene directamente de las

expresiones del Apéndice D.1; resultando

$$\beta_{n,\alpha}(t_{n-1}) = \frac{\alpha h_{n-1}^2}{(h_{n-1}+h_n)(\alpha h_{n-1}+(3-2\alpha)h_n+\alpha h_{n+1})}, \tag{4.72a}$$

$$\beta_{n,\alpha}(t_n) = \frac{\alpha h_{n-1}+(3-\alpha)h_n}{(\alpha h_{n-1}+(3-2\alpha)h_n+\alpha h_{n+1})} \tag{4.72b}$$

$$-\frac{\alpha h_n^2(\alpha h_{n-1}+(3-\alpha)h_n+(3-\alpha)h_{n+1}+\alpha h_{n+2})}{(h_n+h_{n+1})(\alpha h_{n-1}+(3-2\alpha)h_n+\alpha h_{n+1})(\alpha h_n+(3-2\alpha)h_{n+1}+\alpha h_{n+2})}$$

$$\beta_{n,\alpha}(t_{n+1}) = \frac{\alpha h_{n+2}^2}{(h_{n+2}+h_{n+1})(\alpha h_{n+2}+(3-2\alpha)h_{n+1}+\alpha h_n)}. \tag{4.72c}$$

En el caso de que se desee obtener una solución, $s_\alpha(t)$, que posea periodicidad es necesario fijar el valor del periodo de la señal

$$H = (t_N - t_1) = \sum_{n=2}^{N} h_n,$$

lo que implica que los nodos adicionales necesarios para resolver el problema interpolatorio toman los valores

$$\begin{aligned} t_{N+1} &= t_1 + H, \\ t_{N+2} &= t_2 + H, \\ t_0 &= t_N - H, \\ t_{-1} &= t_{N-1} - H. \end{aligned}$$

Finalmente, haciendo uso de las expresiones de (4.72), sólo resta resolver el sistema de ecuaciones en forma matricial que se deriva de (4.71)

$$\mathbf{B}\cdot\mathbf{c} = \mathbf{x}, \tag{4.74}$$

donde

$$\begin{aligned} \mathbf{c} &= [c[1], c[2], \ldots, c[N]]^T, \\ \mathbf{x} &= [x(t_1), x(t_2), \ldots, x(t_N)]^T; \end{aligned}$$

y

$$\mathbf{B} = \begin{bmatrix} \beta_{1,\alpha}(t_1) & \beta_{2,\alpha}(t_1) & 0 & \ldots & 0 & \beta_{N,\alpha}(t_N) \\ \beta_{1,\alpha}(t_2) & \beta_{2,\alpha}(t_2) & \beta_{3,\alpha}(t_2) & \ldots & 0 & 0 \\ 0 & \beta_{2,\alpha}(t_3) & \beta_{3,\alpha}(t_3) & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & \beta_{N-1,\alpha}(t_{N-1}) & \beta_{N,\alpha}(t_{N-1}) \\ \beta_{1,\alpha}(t_1) & 0 & 0 & \ldots & \beta_{N-1,\alpha}(t_N) & \beta_{N,\alpha}(t_N) \end{bmatrix}.$$

Se puede observar el total paralelismo entre este sistema y el presentado en (4.21) correspondiente al caso de muestreo uniforme.

Si se desea obtener una solución $s_\alpha(t)$ que posea simetría especular en los extremos del intervalo, es preciso forzar los nodos adicionales de la siguiente manera,

$$\begin{aligned} t_{N+1} &= t_N + (t_N - t_{N-1}), \\ t_{N+2} &= t_N + (t_N - t_{N-2}), \\ t_0 &= t_1 - (t_2 - t_1), \\ t_{-1} &= t_1 - (t_3 - t_1); \end{aligned}$$

y resolver el sistema de ecuaciones dado por (4.74) utilizando, en esta ocasión, una matriz, $\mathbf{B}$, de la forma

$$\mathbf{B} = \begin{bmatrix} \beta_{1,\alpha}(t_1) & 2\beta_{2,\alpha}(t_1) & 0 & \dots & 0 & 0 \\ \beta_{1,\alpha}(t_2) & \beta_{2,\alpha}(t_2) & \beta_{3,\alpha}(t_2) & \dots & 0 & 0 \\ 0 & \beta_{2,\alpha}(t_3) & \beta_{3,\alpha}(t_3) & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \beta_{N-1,\alpha}(t_{N-1}) & \beta_{N,\alpha}(t_{N-1}) \\ 0 & 0 & 0 & \dots & 2\beta_{N-1,\alpha}(t_N) & \beta_{N,\alpha}(t_N) \end{bmatrix};$$

resultando un sistema de ecuaciones equivalente al descrito en (4.22) para resolver el caso uniforme con simetría especular.

Conviene resaltar que no es posible obtener expresiones cerradas de la suavidad, la varianza y el área residual de la interpolación α-spline para el caso de muestreo no uniforme. Sin embargo, es lógico pensar que el comportamiento de dichas características será similar al caso uniforme, es decir, que la varianza, el área residual y la suavidad crecen conforme crece el parámetro α.

4.11. Simulaciones

En este apartado se realizan simulaciones con el fin de validar los resultados y conclusiones presentados en anteriores secciones.

4.11.1. Interpolación α-spline de tonos

La verificación de las expresiones de suavidad (4.48), varianza (4.56) y área residual (4.64) de la interpolación α-spline desarrolladas en este capítulo se ha llevado a cabo mediante simulaciones. A modo de ejemplo, se ha generado una secuencia, $x[n]$, de $N = 512$ muestras de una señal multitono y se ha realizado la interpolación α-spline[5] para distintos valores del parámetro de suavidad α.

[5]Se han utlizado, indistintamente, el método interpolatorio matricial dado por (4.21) y el método de filtrado dado por el algoritmo (4.31).

Con el fin de estimar la suavidad, la varianza y el área residual de la señal de tiempo continuo, se utiliza un factor de remuestreo muy elevado, en este caso $R = 1000$, es decir, se obtiene una secuencia interpolada de $L = N \cdot R = 512000$ muestras. En la Figura 4.15 se observa un segmento de la señal interpolada y de su segunda derivada para dos valores del parámetro de suavidad (en concreto $\alpha = 0{,}3$ y $\alpha = 0{,}8$), resultando evidente la mayor suavidad de la interpolación cuando $\alpha = 0{,}8$ frente a una segunda derivada más elevada para $\alpha = 0{,}3$.

Desde un punto de vista cuantitativo, los valores de la varianza, la potencia de tensión, y el área residual de la señal interpolada (evaluados, respectivamente, como la media del cuadrado de las muestras de la señal, de su segunda derivada y de la diferencia entre la señal y la interpolación lineal) coinciden de forma exacta con los obtenidos por medio de las expresiones analíticas, tal y como muestra la Tabla 4.2.

		$\boldsymbol{\alpha = 0{,}3}$	$\boldsymbol{\alpha = 0{,}8}$
potencia de tensión	usando (4.46)	17,66708	13,29942
	simulada	17,66698	13,29940
varianza	usando (4.56)	0,8131709	0,9070321
	simulada	0,8131709	0,9070321
área residual	usando (4.64)	0,005138431	0,02285829
	simulada	0,005138431	0,02285829

Tabla 4.2: Valores simulados y analíticos de la potencia de tensión, varianza y área residual de la interpolación α-spline de una señal multitono para dos valores de α.

Con el fin de comprobar el comportamiento de los tres parámetros de interés (suavidad, varianza y área residual) con respecto a α se ha llevado a cabo una batería de simulaciones similares a las realizadas anteriormente (señal multitono de longitud $N = 512$, factor de remuestreo $R = 1000$) variando, para cada señal, el parámetro de suavidad entre $\alpha = 0$ y $\alpha = 1$ con paso de 0,01. Para cada valor de α se calculan los tres valores, tanto analíticamente como por simulación, comprobándose que, además de la perfecta coincidencia entre los valores calculados teóricamente y los obtenidos por simulación, su comportamiento es monótono con respecto a α.

Como ejemplo, la Figura 4.16 muestra los resultados de la interpolación de una de esas señales (cuya varianza es de 1,0972), representándose los valores de potencia de tensión, varianza y área residual en función de α, tanto analíticos como simulados, aunque, debido a la concordancia entre ambos resultados, sólo se aprecia una curva en cada gráfica. Se puede observar como, para esta señal, la potencia de tensión decrece monótonamente con α (la suavidad crece monótonamente), mientras que la varianza y el área residual crecen monótonamente con el parámetro de suavidad.

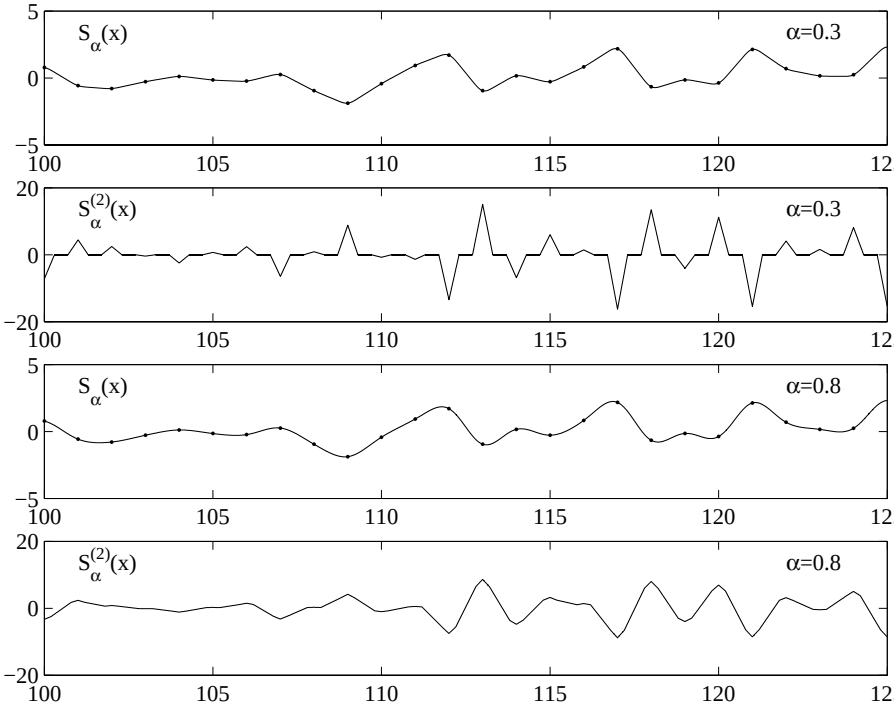


Figura 4.15: Interpolación α-spline de una señal multitono y su segunda derivada para $\alpha = 0{,}3$ y $\alpha = 0{,}8$.

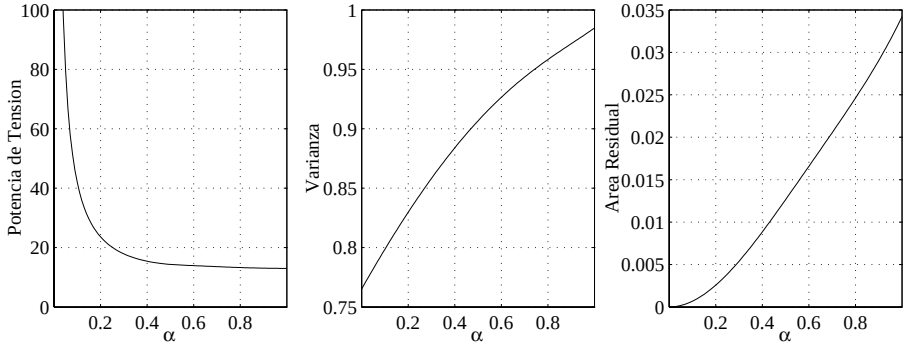


Figura 4.16: Interpolación α-spline de una señal multitono para distintos valores de α. De izquierda a derecha: potencia de tensión, varianza y área residual.

4.11.2. Interpolación α-spline de ruido blanco

Partiendo de M secuencias de N muestras cada una de ruido Gaussiano blanco de media nula y varianza unidad, se lleva a cabo la interpolación α-spline, obteniéndose una nueva secuencia sobremuestreada ($R = 100$ muestras por cada muestra original) con el fin de permitir evaluar de forma precisa la suavidad, varianza y área residual de la señal resultante. Para $M = 20$, $N = 512$, $R = 100$ y el parámetro α variando entre 0 y 1, los resultados de las simulaciones muestran una total concordancia con los resultados teóricos obtenidos. En la Figura 4.17 se observan los resultados de las simulaciones (potencia de tensión, varianza y área residual), obteniéndose unos resultados idénticos a los teóricos presentados en Figura 4.10.a, Figura C.1.a y Figura C.3.a, lo que confirma la validez de las expresiones obtenidas en esta Tesis.

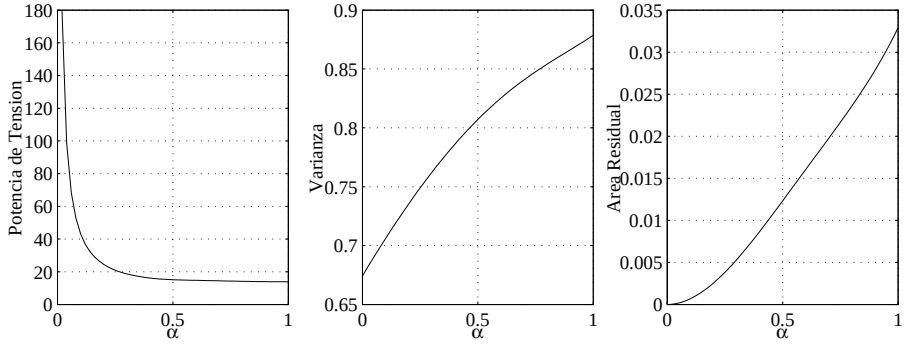


Figura 4.17: Interpolación α-spline de señales blancas de varianza unidad. De izquierda a derecha: potencia de tensión, varianza y área residual.

4.12. Aplicaciones de los α-splines

4.12.1. Control de suavidad, varianza o área residual en la interpolación α-spline

El hecho de que los valores de suavidad, varianza y área residual de la interpolación α-spline sean conocidos (gracias a las expresiones (4.48),(4.56) y (4.64) respectivamente) con anterioridad a la realización de la interpolación (pues dependen únicamente de las muestras de la función a interpolar y del parámetro α) posibilita la búsqueda *a priori* del valor de $\alpha = \alpha_0$ que producirá una señal interpolada con el valor deseado de suavidad, varianza o área residual. Dicha búsqueda, dado el comportamiento monótono de las funciones consideradas, se puede resolver utilizando métodos numéricos estándar (y simples).

4.12.2. Interpolación α-spline de señales no estacionarias

En determinadas aplicaciones se trabaja con señales no estacionarias, es decir, con señales cuyas características estadísticas varían dependiendo del segmento de señal que se considere. Bajo estas circunstancias, es posible utilizar α-splines para interpolar dichas señales y controlar la varianza, la suavidad o el área residual de la solución usando distintos valores de $\alpha = \alpha_k$ en distintas regiones de la señal; lo que se consigue aplicando

$$s_\alpha(t) = \sum_{n=-\infty}^{+\infty} c[n]\beta_{\alpha_n}(t-n),$$

y sustituyendo cada elemento $\mathbf{B}(i,n)$ de la matriz de la ecuación (4.21) con el valor $\beta_{\alpha_n}(i-n)$; lo que permite escoger un valor α_n distinto en torno a cada muestra $x[n]$. La Figura 4.18 muestra un ejemplo de esta aplicación, donde se utilizan tres valores diferentes de α para controlar el valor máximo de la derivada segunda de la interpolación a pesar de la no estacionariedad de la varianza de las muestras originales.

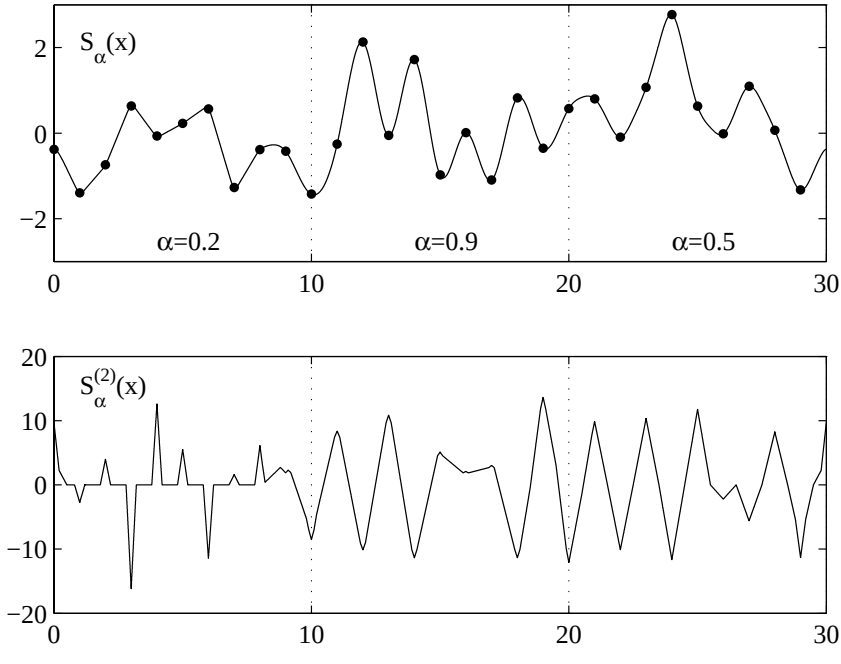


Figura 4.18: Interpolación α-spline (arriba) y su derivada segunda (abajo) utilizando múltiples valores de α.

En el caso de trabajar con muestreo no uniforme, se deberá aplicar la ecuación

$$s_\alpha(t) = \sum_{n=-\infty}^{+\infty} c[n]\beta_{n,\alpha_n}(t),$$

y sustituir los elementos $\mathbf{B}(i,n)$ de (4.74) con el valor $\beta_{n,\alpha_n}(t)$ correspondiente.

4.12.3. Interpolación no uniforme α-spline

En este apartado se lleva a cabo una comparativa de la interpolación no uniforme α-spline frente a la interpolación $\operatorname{sinc}(t)$ ideal, la interpolación lineal y la interpolación spline cúbica. Los resultados demuestran una gran sensibilidad al ruido de la interpolación sinc (pequeños errores en muestras con un tiempo entre muestras muy pequeño son magnificados por los kernels correspondientes a esas muestras), proporcionando señales interpoladas con enormes sobredisparos y con una relación señal a error intolerable, incluso para relaciones señal a ruido relativamente altas de los datos. Por su parte, la interpolación lineal se comporta de forma más robusta, aunque a costa de un resultado no derivable y de subestimar la varianza de la señal original. Por contra, la solución spline cúbica proporciona un resultado suave, pero con sobredisparos relativamente grandes y con una varianza superior a la de las muestras. Entre estas dos últimas soluciones, la interpolación α-spline proporciona un resultado con una varianza idéntica a las muestras originales, derivable y con sobredisparos contenidos.

Un ejemplo consiste en una señal muestreada no uniformemente, cuyos valores de amplitud siguen una distribución Gaussiana de media nula y varianza unidad, y donde el tiempo entre muestras es una variable aleatoria exponencial de media uno (es decir, el periodo de muestreo medio es $T = 1$s). Como se observa en la Figura 4.19 y en la Tabla 4.3, la interpolación α-spline con $\alpha = 0{,}328$ proporciona un resultado suave y continuo hasta la segunda derivada y su varianza coincide con la de las muestras originales; por contra, la solución lineal presenta discontinuidades en la derivada y subestima la varianza; mientras que la interpolación spline resulta suave y continua pero sobreestima enormemente la varianza y presenta sobredisparos inaceptables. También se muestran las soluciones S/H (*Sample and Hold*) y NNR (*Nearest Neighbor Resampling*) que representan métodos de interpolación sencillos y que preservan la varianza pero en los que las discontinuidades de la función resultante pueden ser inadmisibles en determinadas aplicaciones (como sería el caso de señales de velocidad).

Dado que la varianza de la interpolación α-spline crece monótonamente con el parámetro α, en las simulaciones se aprovecha esta circustancia para realizar la búsqueda del α óptimo. Dada una secuencia $x[n]$ compuesta por muestras no equiespaciadas, el algoritmo es simple, y consiste en los siguientes pasos:

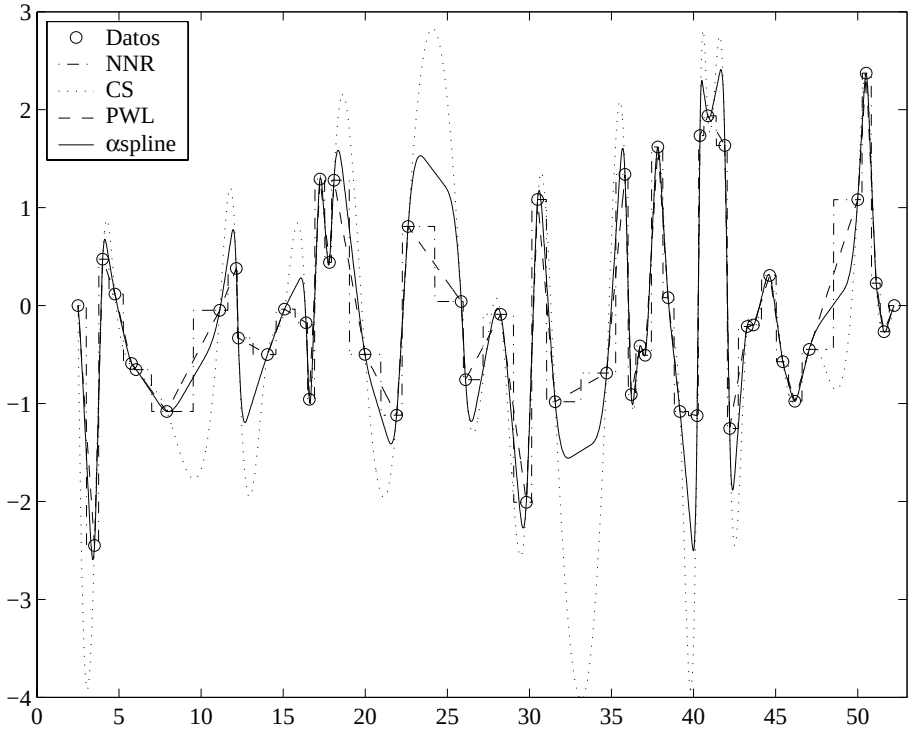


Figura 4.19: Interpolación no uniforme α-spline ($\alpha = 0{,}328$) frente a técnicas NNR, PWL y spline cúbica.

	Datos	**S/H**	**NNR**	**PWL**	**Spline**	**α-spline**
Varianza	1,022	0,846	0,857	0,595	2,305	1,026
Continuidad		no	no	sí	2ª deriv.	2ª deriv.

Tabla 4.3: Interpolación no uniforme α-spline ($\alpha = 0{,}328$) frente a técnicas S/H, NNR, PWL y spline cúbica.

1. Inicializar $\alpha = 0$ y $\lambda = 0{,}5$.

2. Realizar la interpolación α-spline, $s_\alpha(t)$, evaluar su varianza, $\mathrm{Var}\,(s_\alpha(t))$, y el error cometido, $\mathrm{error} = \mathrm{Var}\,(x[n]) - \mathrm{Var}\,(s_\alpha(t))$.

3. Si $|\mathrm{error}| \leq \mathrm{error}_{\mathrm{admisible}}$ finaliza.

4. Hacer $\alpha' = \alpha + \lambda\,\mathrm{sgn}\,(\mathrm{error})$, $\lambda' = \lambda/2$ y volver al paso 2.

Evidentemente, este método puede emplearse igualmente para buscar el valor de α que proporciona una suavidad o un área residual determinadas.

En [Waele00] se lleva a cabo un análisis teórico de los métodos de interpolación no uniforme más simples (NNR y SH). En dicho trabajo se concluye que estos métodos de interpolación son robustos y proporcionan una estimación no sesgada de la varianza. Los métodos complejos que considera (interpolación lineal y spline cúbica) presentan la desventaja de proporcionar una estimación incorrecta de la varianza. Adicionalmente, cuando se aplican estas técnicas a datos reales, como por ejemplo a señales procedentes de anemometría Laser-Doppler (LDA: *Laser-Doppler Anemometry*), el autor observa picos espúreos en la interpolación spline cúbica [Yuou96, Rolain98]. Por contra, la interpolación lineal es robusta frente a dichos picos, pero subestima la varianza. Por todo ello, concluye que el método del vecino más cercano es la técnica preferida para remuestrear datos adquiridos irregularmente.

En este contexto, es evidente que la familia de α-splines propuesta en esta Tesis puede aportar una solución de compromiso que resulte útil. Con este fin, se lleva a cabo la interpolación de muestras irregulares de la velocidad de un flujo turbulento, aunque su aplicación se podría extender al ámbito de la geofísica [Yuou96] y a los osciloscopios digitales [Rolain98]. Dichas muestras han sido obtenidas mediante anemometría Laser-Doppler, que permite medir la velocidad de un fluido o un gas de forma no intrusiva. El registro de datos de partida se corresponde con los utilizados en [Waele00] y la interpolación con los distintos métodos proporcionan los resultados presentados en la Tabla 4.4 y en la Figura 4.20. En concreto, al considerar un segmento de 100 muestras de la señal, se observa que la interpolación lineal subestima la varianza de las muestras, mientras que la interpolación cúbica la sobresestima y genera graves sobredisparos (como los observados en la Figura 4.20 en torno a $t = 0{,}6$s y en $t = 0{,}68$s). Por lo tanto, existe margen de maniobra para encontrar una interpolación α-spline que ajuste el valor de la varianza, que proporcione una solución derivable y que no presente excesivo sobredisparo. Tras aplicar el algoritmo de búsqueda del parámetro de suavidad óptimo, resulta una interpolación α-spline con $\alpha = 0{,}289$.

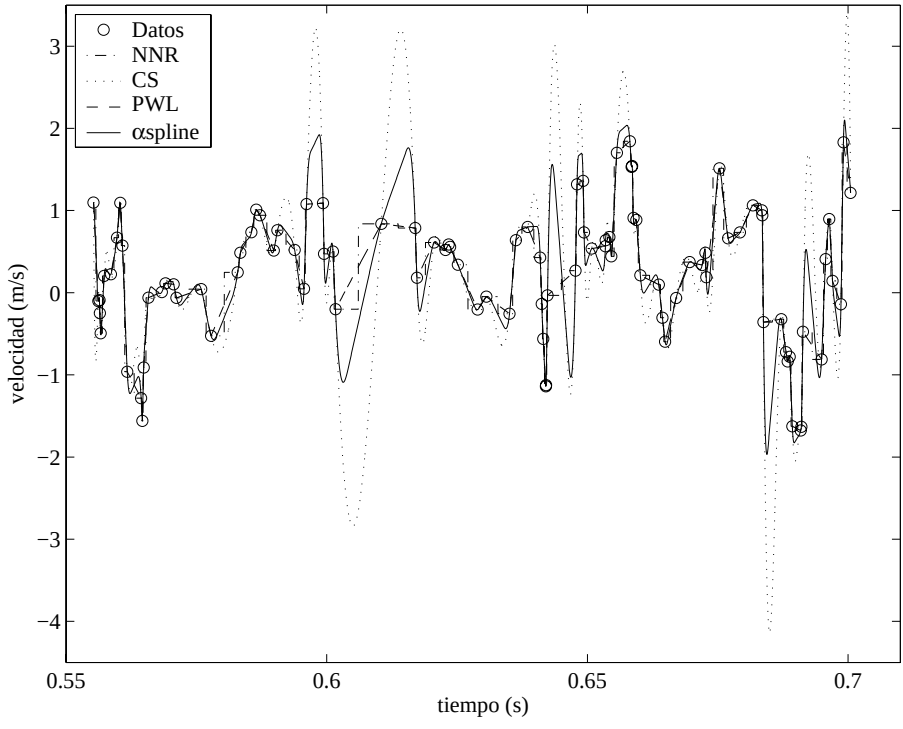


Figura 4.20: Comparativa de técnicas de interpolación no uniforme aplicadas a la velocidad de un flujo turbulento.

	Datos	**S/H**	**NNR**	**PWL**	**Spline**	**α-spline**
Varianza	0,603	0,377	0,445	0,3843	1,412	0,602
Continuidad		no	no	sí	2ª deriv.	2ª deriv.

Tabla 4.4: Interpolación no uniforme de la velocidad de un flujo turbulento ($\alpha = 0{,}289$).

4.13. Conclusiones

En este capítulo se ha presentado una nueva familia de splines, los α-splines, que permite obtener una función interpolatoria comprendida entre las soluciones proporcionadas por dos splines de orden arbitrario, controlando mediante el parámetro α la preponderancia de las características de una u otra. Se ha demostrado en este capítulo que las bases α-spline forman una base de Riesz y una partición de la unidad, todo lo cual asegura que el espacio de señal generado es un subespacio cerrado de L_2 y que la representación α-spline es estable, no ambigua y capaz de aproximar cualquier señal de forma tan precisa como sea necesario sin más que seleccionar un periodo de muestreo suficientemente pequeño. Es posible afirmar, por lo tanto, que la interpolación α-spline se comporta de forma adecuada y puede ser ampliamente utilizada para aprovechar sus especiales características. Puesto que, de entre todos los splines, los lineales y los cúbicos son los más utilizados, la base α-spline comprendida entre orden uno y orden tres adquiere una relevancia adicional. Por ello, este capítulo se ha dedicado en profundidad a este caso particular, que, evidentemente, hereda todas las características de la familia α-spline genérica.

Un aspecto fundamental, común a todos los splines, es la posibilidad de resolver el problema de interpolación α-spline mediante filtrado digital (siempre que las muestras se encuentren equiespaciadas), lo que posibilita una implementación computacionalmente muy eficiente. En este capítulo se han presentado las expresiones que permiten dicha implementación y se propone el empleo de ciertas bases α-spline que, por las especiales características de los filtros digitales que intervienen, evitan la necesidad de operaciones de punto flotante, resultando un sistema de interpolación mucho más simple y rápido.

En el presente capítulo también se han presentado expresiones cerradas que relacionan la suavidad, la varianza y el área residual de la interpolación α-spline con las muestras de la señal, de tal manera, que es posible calcular el valor de dichas características sin llevar a cabo la interpolación. Dado el comportamiento monótono con respecto a α de la suavidad, la varianza y el área residual, es posible (y sencillo) llevar a cabo la búsqueda del parámetro α que proporciona una interpolación con las características deseadas. Esta metodología se ha mostrado de gran utilidad en diversos ejemplos, principalmente en casos de muestreo no uniforme, donde el ruido en las muestras provoca sobredisparos intolerables en la interpolación spline cúbica (y más aún en la interpolación sinc). Con este fin, en las últimas secciones del capítulo se ha presentado la generalización de las bases α-spline considerando muestreo no uniforme, así como el algoritmo de interpolación correspondiente, el cual, forzosamente, es de tipo matricial.

Capítulo 5

Modelado α-spline

5.1. Introducción

Existen aplicaciones de modelado, como por ejemplo las relativas a dispositivos de microondas o a amplificadores de potencia, en las que las técnicas tradicionales como las ecuaciones cerradas, las series de Volterra o las tablas de búsqueda resultan difíciles de implementar debido a sus altos requerimientos de memoria o a su coste computacional. Es más, ciertos requisitos adicionales, como la derivabilidad del modelo o cierto grado de suavidad, son fundamentales en dichas aplicaciones (por ejemplo, para evaluar productos de intermodulación, recrecimiento espectral, etc.) y deben ser tenidos en cuenta.

En esta línea, las redes neuronales han sido propuestas en numerosos trabajos [Zaabab95, Rousset96, Shirakawa97] con el fin de modelar el comportamiento no lineal de circuitos y dispositivos activos, basándose en la capacidad que poseen las redes neuronales de aproximar cualquier función no lineal y su habilidad para aprender a partir de datos experimentales. Sin embargo, en la mayoría de los casos se utiliza el perceptrón multicapa (MLP: *Multilayer Perceptron*), cuyos requisitos de memoria y de cómputo son elevados. En otros trabajos, [Santamaría99, García99] se ha comprobado que una versión diferente de redes neuronales, la red de funciones de base radial generalizadas (GRBF: *Generalized Radial Basis Function*), proporciona mejores prestaciones que el MLP, sobre todo cuando interesa modelar, por ejemplo, el comportamiento de las derivadas en torno a un punto de polarización.

Otra solución a considerar ante este tipo de problemas de modelado son los splines. El modelado spline, entendido como la búsqueda de una función spline que minimice el error cuadrático medio entre su valor y los datos, es un problema lineal en los coeficientes si se fijan de antemano los puntos de frontera o ruptura (también denominados nodos o *knots*) de los distintos tramos.

En el caso multidimensional, las bases spline unidimensionales se transforman en bases M-dimensionales por medio de productos tensoriales, y la segmentación del espacio de los datos de entrada se realiza por medio de una rejilla M-dimensional. Sin embargo, hay que resaltar el hecho de que la utilización de fronteras prefijadas (por ejemplo, equiespaciadas) proporciona resultados subóptimos (incluso muy pobres) [Hlavachova97] y la dependencia no lineal del error respecto a la posición de dichos puntos de ruptura dificulta y limita la aplicabilidad del modelado spline.

Si se dispone de un conjunto de datos a modelar compuesto por L valores y se pretende buscar un modelo de P tramos, existen

$$\binom{L-2}{P-2} = \frac{(L-2)!}{(L-P)!(P-2)!}$$

posibles combinaciones de fronteras, número que crece muy rápidamente y que imposibilita una búsqueda exhaustiva incluso para valores pequeños de L. Para solventar este contratiempo ciertos autores han planteado técnicas iterativas de colocación de los nodos [Boor68, Jupp78, Dierckx95] o técnicas basadas en particionamiento recursivo [Friedman91]. Así mismo, las técnicas de segmentación de señales polinómicas a tramos, basadas en distintos criterios, como el AIC (*Akaike's Information Criterion*)[Akaike74], el MDL (*Minimum Description Length*)[Rissanen78] o el MAP (*Maximum A Posteriori*) [Djuric94, Giurcaneanu00], son aplicables al problema de la localización de los nodos. Recientemente, se ha aplicado con relativo éxito la técnica del calentamiento y progresivo enfriado (*simulated annealing*) a la optimización de los puntos de ruptura en el modelado spline de sistemas no lineales [Yiu01]. En cualquier caso, todas estas técnicas poseen una complejidad alta, motivo suficiente para buscar soluciones más simples que sean capaces de aprovechar las ventajas de los splines sin pagar un precio excesivo en cuanto a complejidad.

Una solución alternativa es el modelo canónico lineal a tramos (CPWL: *Canonical Piecewise-Linear*) [Chua86, Chua88], que proporciona soluciones con un alto grado de exactitud, con un número de parámetros bajo y con unos requerimientos de cómputo en general menores que las redes neuronales. Sin embargo, debido a su naturaleza lineal a tramos, el modelo resultante presenta una derivada segunda nula y discontinuidades en la primera derivada, lo que imposibilita su uso en modelado de dispositivos o sistemas no lineales con comportamiento suave o en los que interese el valor de sus derivadas. En [Lin93] se propone una generalización del modelo CPWL que sustituye la función lineal de activación por una función polinómica, resultando un modelo suave a tramos que, lamentablemente, posee cierta tendencia a la oscilación heredada de los polinomios que utiliza.

En esta Tesis se propone una extensión diferente del modelo canónico lineal a tramos, manteniendo las ventajas del modelo CPWL (simplicidad y precisión)

y proporcionando suavidad y derivabilidad. Para ello se ha tenido en cuenta el hecho de que la base α-spline con $\alpha = 0$, al igual que el modelo CPWL, proporciona funciones lineales a tramos y que el parámetro α posibilita controlar la suavidad de una solución que, en todos lo casos, posee continuidad hasta la segunda derivada. Por lo tanto, se propone sustituir la función de activación valor absoluto del modelo CPWL por una función basada en el nucleo α-spline, generándose así modelos derivables, con transiciones suaves entre las distintas regiones del espacio de entrada y con una suavidad controlable gracias al parámetro α.

En los siguientes apartados se presenta, en primer lugar, el modelo canónico lineal a tramos y el procedimiento propuesto para obtener un modelo lineal a tramos suavizado basado en el nucleo α-spline, así como la metodología de entrenamiento de dicho modelo. Posteriormente se presentarán dos aplicaciones: el modelado α-spline de amplificadores de potencia basado en medidas temporales de entrada-salida obtenidas mediante técnicas de submuestreo, y el modelado α-spline de gran señal de transistores HEMT sometidos a iluminación óptica.

5.2. Modelo α-spline

5.2.1. El modelo lineal a tramos canónico

Aunque es bien conocido que la representación lineal a tramos canónica propuesta por Chua [Chua86, Chua88] es un subconjunto de la clase más general de las funciones lineales a tramos, es la única que admite una formulación compacta y explícita [Leenaerts98]. Dicha representación realiza un mapeo $f : \mathbb{R}^M \to \mathbb{R}^N$ de la forma

$$f(\mathbf{x}) = \mathbf{a} + \mathbf{B}\mathbf{x} + \sum_{i=1}^{P} \mathbf{c}_i |\langle \boldsymbol{\nu}_i, \mathbf{x}\rangle - \beta_i| \tag{5.1}$$

donde $\mathbf{x}$ y $\boldsymbol{\nu}_i$ son vectores de la misma dimensión, M, que el espacio de entrada; $\mathbf{a}$ y $\mathbf{c}_i$ son vectores de la misma dimensión, N, que el espacio de salida; $\mathbf{B}$ es una matriz $N \times M$; β_i es un escalar; y $\langle \cdot, \cdot \rangle$ denota el producto interno. La inexistencia de información redundante en (5.1) reduce al máximo la memoria requerida para almacenar los parámetros del modelo, y su especial estructura permite desarrollar algoritmos muy eficientes para su cálculo.

El modelo canónico lineal a tramos particiona el espacio de entrada en diferentes regiones por medio de P fronteras impuestas por hiperplanos de dimensión $M-1$, y construye el modelo por medio de la combinación de hiperplanos con forma de bisagra de dimensión $M-1$. En cada región del espacio de entrada, la función se compone de una combinación lineal de hiperplanos lineales,

lo que genera un modelo lineal a tramos y continuo. En (5.1) se observa que la expresión dentro de la función valor absoluto define las fronteras de la partición del espacio de entrada, resultando, además, que en esas fronteras, la transición de la función se realiza siguiendo las características de los hiperplanos bisagra, es decir, según la forma de la función valor absoluto, que es continua, pero no derivable en los puntos frontera. El modelado lineal a tramos es capaz de proporcionar un resultado muy preciso con un número bajo de parámetros, pero su discontinuidad en la derivada limita su uso en determinadas aplicaciones, como, por ejemplo, en el modelado de dispositivos.

5.2.2. Suavizado de la función valor absoluto

Con el fin de solventar la limitación de la no derivabilidad, se pretende sustituir la abrupta función valor absoluto de (5.1) por una función derivable que suavice las uniones de los hiperplanos en las fronteras sin incrementar sustancialmente el número de parámetros de la solución. Existen diversas posibilidades que permiten una transición suave entre tramos lineales, siendo de utilidad considerar el caso unidimensional en el que la función, $f : \mathbb{R} \to \mathbb{R}$, toma la forma

$$f(x) = a + Bx + \sum_{i=1}^{P} c_i |x - \beta_i|, \tag{5.2}$$

donde los β_i representan los puntos de ruptura (*knots*) que determinan los intervalos en los cuales la función modeladora es una recta. De forma equivalente, (5.2) puede escribirse en términos de la función signo de la siguiente manera:

$$f(x) = a + Bx + \sum_{i=1}^{P} c_i (x - \beta_i) \operatorname{sgn}(x - \beta_i). \tag{5.3}$$

Sería deseable disponer de una familia de funciones que proporcionara una transición suave y permitiera un control paramétrico de la brusquedad de cambio en las transiciones. Existen en la literatura numerosas funciones con estas características; por ejemplo, en [Bacon71] la función $\operatorname{sgn}(x)$ es reemplazada por una familia de funciones aproximadoras, denotadas por $\operatorname{trn}(x)$, que cumplen las siguientes restricciones:

$$\lim_{x \to \pm\infty} (x \operatorname{trn}(x) - |x|) = 0,$$

$$\lim_{\alpha \to \infty} \operatorname{trn}(x/\alpha) = \operatorname{sgn}(x),$$

$$\operatorname{trn}(0) = \operatorname{sgn}(0) = 0.$$

La primera condición asegura que $x \operatorname{trn}(x)$ se aproxima a $x \operatorname{sgn}(x) = |x|$ conforme $x \to \infty$. La segunda implica que el modelo lineal a tramos (5.3) es

un caso límite particular de un modelo con transiciones suaves. Por último, la tercera condición asegura que el modelo suavizado pasa por los puntos resultantes de la intersección de los tramos lineales de (5.3). Esta familia posee un parámetro, α, que controla la suavidad de las transiciones: modificando este parámetro se varía la suavidad en torno a los puntos de ruptura. Alguna de las funciones pertenecientes a esta familia son [Seber89]:

$$\mathrm{ftan}(x,\alpha) = x\tanh(\alpha x), \tag{5.5}$$

y

$$\mathrm{fpol}(x,\alpha) = \begin{cases} \frac{1}{8\alpha}\left[15(\alpha x)^2 - (\alpha x)^4 + 3(\alpha x)^6\right], & |x| \leq \frac{1}{\alpha}; \\ |x|, & |x| > \frac{1}{\alpha}. \end{cases} \tag{5.6}$$

La curvatura, K, es un parámetro geométrico que proporciona cierto conocimiento acerca del comportamiento de estas funciones, y cuyo valor es

$$K = \frac{f^{(2)}(x)}{[1 + (f^{(1)}(x))^2]^{3/2}}.$$

En el punto de ruptura, es decir, cuando $x = 0$, la curvatura de (5.5) y (5.6) es 2α y $15\alpha/4$, respectivamente. Se puede comprobar cómo, cuanto mayor sea el valor de K, las funciones tienden más rápidamente a la función valor absoluto. Así mismo, se observa que ambas funciones alcanzan valores mayores que la unidad en su primera derivada y que su derivada segunda presenta valores negativos, lo que implica sobredisparos que dificultan la obtención de soluciones con derivadas suaves. Es posible utilizar otras funciones que sustituyan a $|x|$ y que solventen esta situación, como por ejemplo la propuesta por [Griffiths73],

$$\mathrm{hyp}(x,\alpha) = \sqrt{x^2 + \alpha},$$

cuya curvatura toma el valor $K = 1/\sqrt{\alpha}$.

Recientemente, se ha propuesto como base la función [Lázaro01b]

$$\mathrm{lch}(x,\alpha) = \frac{1}{\alpha}\ln(\cosh(\alpha x)), \tag{5.7}$$

cuya curvatura es $K = \alpha$, para desarrollar un modelo lineal a tramos suavizado denominado modelo lineal a tramos suavizado (SPWL: *Smoothed Piecewise-Linear*).

En esta Tesis se propone una función suavizadora relacionada íntimamente con el núcleo α-spline presentado en anteriores capítulos. Dicha función, que denotaremos por $\varphi(x,\alpha)$, sigue la expresión

$$\varphi(x,\alpha) = \begin{cases} \frac{x^2}{\alpha} - \frac{|x|^3}{3\alpha^2}, & |x| \leq \alpha; \\ |x| - \frac{\alpha}{3}, & |x| > \alpha; \end{cases} \tag{5.8}$$

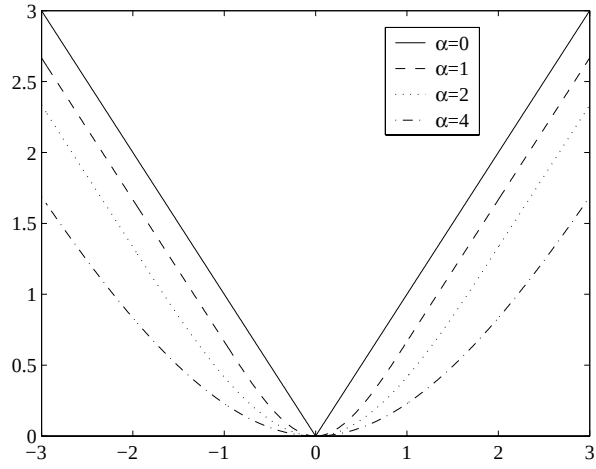


Figura 5.1: Función de suavizado α-spline, $\varphi(x,\alpha)$, para diversos valores del parámetro de suavidad α.

y posee una curvatura $K = 2/\alpha$. En la Figura 5.1 puede apreciarse su forma para distintos valores del parámetro de suavidad.

Las dos primeras derivadas de la función (5.8) vienen dadas por

$$\varphi^{(1)}(x,\alpha) = \begin{cases} \frac{2x}{\alpha} - \frac{x|x|}{\alpha^2}, & |x| \leq \alpha; \\ \operatorname{sgn}(x), & |x| > \alpha; \end{cases} \tag{5.9}$$

y

$$\varphi^{(2)}(x,\alpha) = \begin{cases} \frac{2}{\alpha} - \frac{2|x|}{\alpha^2}, & |x| \leq \alpha; \\ 0 & |x|, > \alpha. \end{cases} \tag{5.10}$$

La Figura 5.2 muestra las cinco funciones de suavizado así como sus dos primeras derivadas para un valor del grado de curvatura de $K = 4$. Se observa que las derivadas de $\mathrm{hyp}(x,\alpha)$, $\mathrm{lch}(x,\alpha)$ y $\varphi(x,\alpha)$ no presentan sobredisparo, es más, para la misma curvatura estas tres funciones son más suaves que $\mathrm{ftan}(x,\alpha)$ y $\mathrm{fpol}(x,\alpha)$. Así mismo, se ha observado que $\mathrm{hyp}(x,\alpha)$, $\mathrm{lch}(x,\alpha)$ y $\varphi(x,\alpha)$ proporcionan resultados muy similares (y superiores a $\mathrm{ftan}(x,\alpha)$ y $\mathrm{fpol}(x,\alpha)$) a la hora de modelar dispositivos, por lo que la elección entre dichas funciones se toma en función de sus posibles ventajas de implementación. Así, la función $\varphi(x,\alpha)$ y sus derivadas, además de poseer unas expresiones sencillas (polinómicas), pueden ser evaluadas de forma mucho más rápida que aquellas en las que intervienen expresiones basadas en $\tanh(x)$ y $\operatorname{sech}(x)$, lo que redunda en menores tiempos de estimación (entrenamiento) del modelo y de evaluación de la función aproximadora.

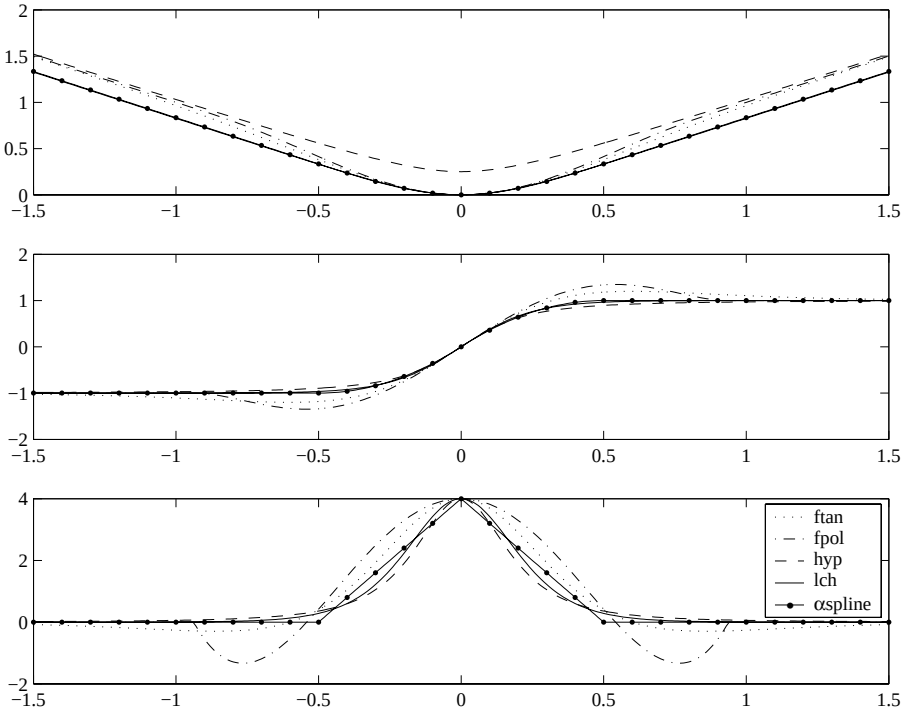


Figura 5.2: Funciones de suavizado (arriba) y sus dos primeras derivadas para una curvatura $K = 4$ ($\alpha = 0{,}5$).

5.2.3. El modelo αPWL

Aunque en la discusión previa se han considerado únicamente funciones unidimensionales, la extensión a un espacio de entrada multidimensional es inmediata. En concreto, el modelo lineal a tramos suavizado basado en el núcleo α-spline que realiza un mapeo $f:\ \mathbb{R}^M \to \mathbb{R}^N$ se denota como $\alpha\mathrm{PWL}(M, P, N)$ y adopta la forma

$$f(\mathbf{x}) = \mathbf{a} + \mathbf{B}\mathbf{x} + \sum_{i=1}^{P} \mathbf{c}_i \varphi\left(\langle \boldsymbol{\nu}_i, \mathbf{x}\rangle - \beta_i, \alpha_i\right). \tag{5.11}$$

En este caso, la frontera i-ésima está gobernada por $\varphi\left(\langle \boldsymbol{\nu}_i, \mathbf{x}\rangle - \beta_i, \alpha_i\right)$, una función suave, continua y derivable, por lo que, en consecuencia, el modelo global compartirá esas mismas propiedades. En la Figura 5.3 se muestra un ejemplo de dicha función base para un espacio de entrada bidimensional y para $\boldsymbol{\nu} = [1\ 0{,}25]$, $\beta = 0$ y $\alpha = 2/K = 0{,}5$; pudiéndose apreciar la forma y suavidad de la misma.

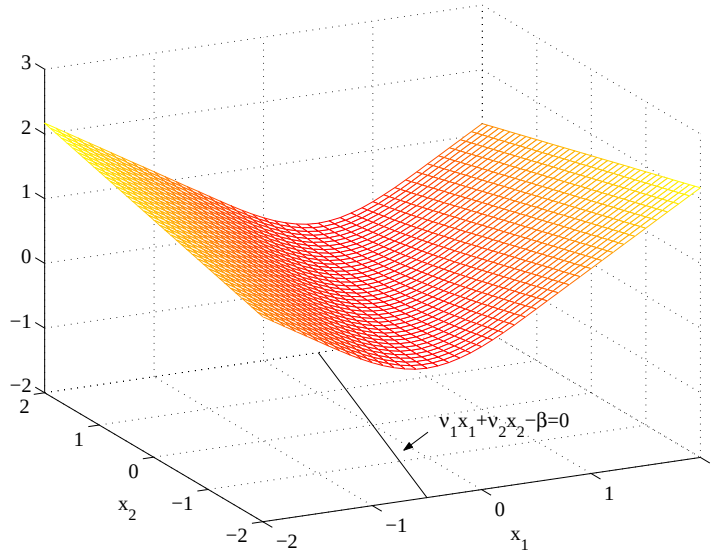


Figura 5.3: Función base $\varphi\left(\langle \boldsymbol{\nu}, \mathbf{x}\rangle - \beta, \alpha\right)$ en un espacio bidimensional y una frontera definida por la recta $x_1 - 0{,}25x_2 = 0$.

Para el caso particular unidimensional en el que las fronteras se encuentren equiespaciadas, el resultado de utilizar la aproximación lineal a tramos canónica suavizada con la función $\varphi\left(x, \alpha\right)$ es equivalente a la aproximación α-spline. De hecho, existe una relación entre la función base α-spline, $\beta_\alpha\left(x\right)$, y la función de suavizado, $\varphi\left(x, \alpha\right)$, dada por la expresión

$$\beta_\alpha\left(x\right) = -\varphi\left(x, \alpha\right) + \frac{1}{2}\varphi\left(x-1, \alpha\right) + \frac{1}{2}\varphi\left(x+1, \alpha\right). \tag{5.12}$$

Ésta es una formulación similar al suavizado de la función lineal a tramos propuesta para el modelado MARS (*Multivariate Adaptive Regression Splines*) en [Friedman91].

5.2.4. Suavidad del modelo αPWL

Como se ha comprobado en el Capítulo 4, la suavidad de una función interpolatoria α-spline crece conforme aumenta el parámetro de suavidad α. Dada la relación (5.12) existente entre la base interpolatoria α-spline y la base aproximadora α-spline, es lógico pensar que la suavidad del modelo αPWL (5.11) presentará una dependencia monótona creciente con el parámetro α.

En esta sección se demuestra que la suavidad del modelo global tiene una dependencia monótona respecto al parámetro α (se considerará una versión simplificada del modelo en la que se utiliza el mismo valor de α en todas las regiones o tramos). Por tanto, α se comporta como un parámetro de regularización, permitiendo ponderar entre la fidelidad del ajuste a los datos y la suavidad del modelo resultante.

Las técnicas habituales de regularización minimizan una función de coste compuesta por dos términos: el primero mide la proximidad a los datos y el segundo pondera el coste asociado con un funcional que mide la suavidad de la solución, es decir,

$$E = \sum_i (f(x_i) - y_i)^2 + \lambda ||P_s f(x)||^2, \tag{5.13}$$

donde

- y_i representa las medidas;
- λ es el parámetro de regularización, que controla el compromiso entre el grado de suavidad de la solución y su ajuste a los datos;
- P_s es el funcional estabilizador.

La suavidad puede ser medida de formas muy variadas, aunque, generalmente, el estabilizador P acostumbra a incorporar alguna derivada de la función. Una clase de estabilizadores ampliamente utilizados son aquellos que utilizan el funcional [Poggio90]

$$||P_s^m f||^2 = \sum_{i_1,\ldots,i_m}^{n} \int_{Rn} (\partial_{i_1,\ldots,i_m} f(x))^2 dx$$

donde $\partial_{i_1,\ldots,i_m} = \partial^m / \partial x_{i_1} \ldots \partial x_{i_m}$ y $m \geq 1$. Este estabilizador, invariante frente a rotaciones o desplazamientos, será el que se utilice para controlar la suavidad de nuestro modelo.

Con estas premisas, se pretende obtener un punto de encuentro entre el modelo lineal a tramos suavizado y la teoría de la regularización. En concreto, se demostrará que la relación entre la suavidad (evaluada mediante la integral de las derivadas de segundo orden) y el parámetro α es monótona y que, por tanto, es posible controlar la suavidad del modelo mediante una selección apropiada del valor de α. Dada la estrecha relación entre la función base $\varphi(x, \alpha)$ y la base α-spline, podrían extrapolarse los resultados de la interpolación α-spline al modelo αPWL. Sin embargo, se pretende alcanzar dicho resultado siguiendo un camino diferente.

Se comienza considerando un ejemplo unidimensional, para el cual la expresión del estabilizador tiene en cuenta la integral del cuadrado de la derivada

segunda de la función. Dado que la derivada segunda de la función base propuesta es localizada ($\varphi^{(2)}(x,\alpha) = 0$ para $|x| \geq \alpha$), entonces la derivada segunda de cada componente únicamente se solapa con la de las componentes más próximas. Sin pérdida de generalidad, se supone que el modelo está compuesto por dos componentes separadas por una distancia $b \geq 0$, es decir,

$$f(x) = c_1\varphi(x,\alpha) + c_2\varphi(x-b,\alpha).$$

En este caso, el término de regularización viene dado por la expresión

$$||P_s^2 f||^2 = \begin{cases} (c_1^2+c_2^2)\frac{8}{3\alpha}, & \alpha \leq b/2; \\ (c_1^2+c_2^2)\frac{8}{3\alpha} + c_1c_2\frac{4(2\alpha-b)^3}{3\alpha^4}, & b/2 \leq \alpha \leq b; \\ (c_1^2+c_2^2)\frac{8}{3\alpha} + c_1c_2\frac{4(3b^3+4\alpha^3-6b^2\alpha)}{3\alpha^4}, & \alpha \geq b; \end{cases} \tag{5.14}$$

que resulta monótona decreciente respecto a α, independientemente del valor que tome b, es decir, la separación entre las dos componentes, o de los pesos c_1 y c_2; pues su derivada respecto a α es negativa, como se demuestra en el Apéndice E.1.

Al extender el anterior análisis a un espacio de entrada bidimensional, surge el problema de que la segunda derivada no tiene soporte finito en la dirección de las fronteras.[1] Sin embargo, es posible evitar este inconveniente integrando en una región finita apropiada. Para simplificar la evaluación de las distintas integrales, se escoge una región de integración, R' delimitada por líneas ortogonales a las fronteras, tal y como muestra la Figura 5.4. Con esta consideración, el estabilizador adopta la siguiente forma:

$$||P_s^2 f||^2 = \int_{R'} \left[\left(\frac{\partial^2 f}{\partial x_1^2}\right)^2 + 2\left(\frac{\partial^2 f}{\partial x_1 \partial x_2}\right)^2 + \left(\frac{\partial^2 f}{\partial x_2^2}\right)^2 \right] dx_1 dx_2.$$

De nuevo, se considera un modelo compuesto únicamente por dos componentes, es decir,

$$f(x) = \sum_{i=1}^{2} c_i\varphi(m_{1i}x_1 + m_{2i}x_2 + t_i, \alpha).$$

Sin pérdida de generalidad, es posible suponer $t_1 = t_2 = 0$, lo que implica que el punto correspondiente a la intersección de las dos fronteras se encuentra en el origen de coordenadas, y que $m_{11} = 0$, $m_{12} = m_{22} = 1$ y $m_{21} = m$. Bajo

[1] Recuérdese que, en un modelo lineal a tramos canónico bidimensional, las regiones están delimitadas por rectas del tipo $m_1x_1 + m_2x_2 = t$ que constituyen las líneas de ruptura de los planos que forman el modelo.

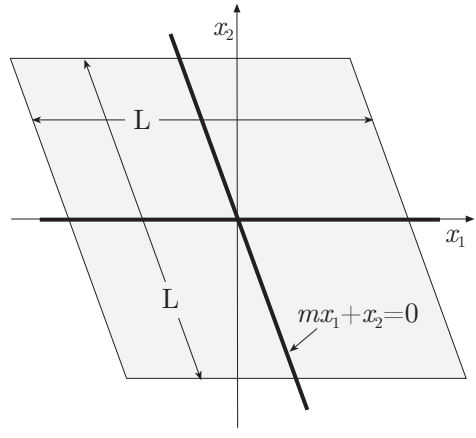


Figura 5.4: Región de integración para un espacio de entrada bidimensional.

estas condiciones, el término de regularización toma el valor

$$\begin{aligned}||P_s^2 f||^2 = &\int_{R'} c_1^2 \left[\varphi^{(2)}\left(x_2,\alpha\right)\right]^2 dx_1 dx_2 \\ &+ \int_{R'} c_2^2(m^4+2m^2+1)\left[\varphi^{(2)}\left(mx_1+x_2,\alpha\right)\right]^2 dx_1 dx_2 \\ &+ \int_{R'} 2c_1c_2\varphi^{(2)}\left(x_2,\alpha\right)\varphi^{(2)}\left(mx_1+x_2,\alpha\right) dx_1 dx_2,\end{aligned}$$

donde los dos primeros términos corresponden a la contribución de cada componente dentro de la región de integración R' mientras que el tercer término considera la intersección de ambas componentes, que será finita dado el soporte compacto de dicha intersección. Finalmente, se obtiene

$$||P_s^2 f||^2 = \frac{8L}{3\alpha}(c_1^2 + c_2^2(m^2+1)^2) + \frac{8}{m}c_1c_2,$$

siendo su derivada con respecto a α

$$\frac{d||P_s^2 f||^2}{d\alpha} = \frac{-8L}{3\alpha^2}(c_1^2 + c_2^2(m^2+1)^2) \leq 0,$$

lo que demuestra la dependencia monótona decreciente de la suavidad del modelo con respecto a α. En definitiva, incrementar α provoca que las transiciones entre las secciones lineales del modelo sean más suaves, por lo que la suavidad del modelo global aumenta.

5.2.5. Entrenamiento del modelo

El modelo α-spline posee tres tipos distintos de parámetros que deben ser estimados: los que definen las fronteras que particionan el espacio de entrada ($\boldsymbol{\nu}_i$ y β_i); los que determinan la combinación lineal de las distintas componentes del modelo ($\mathbf{a}$, $\mathbf{B}$ y $\mathbf{c_i}$); y el parámetro de suavidad α.

El algoritmo de aprendizaje del modelo es de tipo iterativo, y se basa en sucesivas adaptaciones de la partición del espacio de entrada y en la estima de los coeficientes óptimos del modelo para dicha partición. La adaptación de los parámetros que definen las fronteras de la partición se realiza utilizando un método basado en el gradiente de segundo orden, en concreto, se emplea el gradiente y el Hessiano de la función error con respecto a los parámetros. Una vez que son fijadas las fronteras del espacio de entrada, el error de aproximación del modelo es una función cuadrática de los parámetros que definen la combinación lineal de las componentes, por lo que su mínimo se encuentra de forma muy simple resolviendo un sistema lineal de ecuaciones. Posteriormente, se vuelven a adaptar las fronteras y el proceso se repite iterativamente.

Éste es, básicamente, el método propuesto por L. Chua para optimizar los parámetros del modelo canónico lineal a tramos [Chua86], salvo por el hecho de que la función de activación es, en nuestro caso, $\varphi(x, \alpha)$ (5.8) en lugar de la función valor absoluto. Por ello, es preciso obtener las expresiones que posibilitan el entrenamiento del modelo α-spline.

Se considera el caso general de la aproximación de un mapeo ($f : \mathbb{R}^M \to \mathbb{R}^N$) usando un conjunto de L datos de entrada-salida del tipo $(\mathbf{x}^{[l]}, \mathbf{y}^{[l]})$ con $l = 1, \ldots, L$, siendo $\mathbf{x}^{[l]} = [x_1^{[l]}, x_2^{[l]}, \ldots, x_M^{[l]}]^T$ el vector l-ésimo de los datos de entrada (de dimensión M); e $\mathbf{y}^{[l]} = [y_1^{[l]}, y_2^{[l]}, \ldots, y_N^{[l]}]^T$ el vector l-ésimo de los datos de salida (de dimensión N).

Es posible eliminar un coeficiente de cada frontera reescribiendo $\langle \boldsymbol{\nu}_i, \mathbf{x} \rangle - \beta_i$ en (5.11) como

$$d_i(\mathbf{x}) = m_{i,1}x_1 + m_{i,2}x_2 + \cdots + m_{i,M-1}x_{M-1} - x_M + t_i,$$

donde $d_i(\mathbf{x})$ denota la i-ésima frontera evaluada en $\mathbf{x}$. Con esta consideración, el modelo α-spline (5.11) con P fronteras puede escribirse

$$\mathbf{y} = f(\mathbf{x}) = \mathbf{a} + \mathbf{B}\mathbf{x} + \sum_{i=1}^{P} \mathbf{c}_i \varphi\left(d_i(\mathbf{x}), \alpha\right).$$

Los parámetros de este modelo pueden agruparse en un vector, $\mathbf{z}_f$, de longitud PM que define las P fronteras que particionan el espacio de entrada

$$\mathbf{z}_f = [m_{1,1}, \ldots, m_{1,M-1}, \ldots, m_{P,1}, \ldots, m_{P,M-1}, t_1, \ldots, t_P]^T;$$

y una matriz, $\mathbf{Z}_c$, de dimensión $N \times (M+1+P)$ que reúne los coeficientes asociados a la combinación lineal de las funciones base del modelo

$$\mathbf{Z}_c = \begin{bmatrix} a_1 & b_{1,1} & \dots & b_{1,M} & c_{1,1} & \dots & c_{1,P} \\ a_2 & b_{2,1} & \dots & b_{2,M} & c_{2,1} & \dots & c_{2,P} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ a_N & b_{N,1} & \dots & b_{N,M} & c_{N,1} & \dots & c_{N,P} \end{bmatrix}.$$

El modelo debe minimizar la función de coste dada por el error cuadrático entre la salida real y la proporcionada por el modelo

$$E(\mathbf{Z}_c, \mathbf{z}_f) = \sum_{l=1}^{L} \left\| \mathbf{y}^{[l]} - \left(\mathbf{a} + \mathbf{B}\mathbf{x}^{[l]} + \sum_{i=1}^{P} \mathbf{c}_i \varphi \left(d_i(\mathbf{x}^{[l]}), \alpha \right) \right) \right\|^2. \tag{5.15}$$

Para llevar a cabo la adaptación de los distintos parámetros, el algoritmo se inicia fijando unas fronteras iniciales, es decir, el vector $\mathbf{z}_f$. Habitualmente, esta partición inicial se realiza de forma aleatoria con la restricción de que las fronteras pertenezcan al dominio de los datos de entrada. Una vez fijado $\mathbf{z}_f$, la función de coste (5.15) es una función cuadrática en $\mathbf{Z}_c$, por lo que su mínimo se obtiene mediante pseudoinversa

$$\mathbf{Z_c} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{Y}, \tag{5.16}$$

donde $\mathbf{Y} = [\mathbf{y}^{[1]}, \dots, \mathbf{y}^{[L]}]^T$ es la matriz $L \times N$ de los datos de salida, y $\mathbf{A}$ es una matriz de dimensión $(M+P+1) \times L$ de la forma

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ x_1^{[1]} & x_1^{[2]} & \dots & x_1^{[L]} \\ \vdots & \vdots & & \vdots \\ x_M^{[1]} & x_M^{[2]} & \dots & x_M^{[L]} \\ u_1^{[1]} & u_1^{[2]} & \dots & u_1^{[L]} \\ \vdots & \vdots & & \vdots \\ u_P^{[1]} & u_P^{[2]} & \dots & u_P^{[L]} \end{bmatrix},$$

siendo $u_i^{[l]}$ el valor de la i-ésima función base para el vector de entrada l-ésimo, es decir, $u_i^{[l]} = \varphi\left(d_i(\mathbf{x}^{[l]}), \alpha\right)$.

Una vez obtenidos los parámetros $\mathbf{Z}_c$ óptimos para la partición inicial, el algoritmo debe estimar una nueva partición, para lo cual se evalúa el gradiente, $\mathbf{g}$, y el Hessiano, $\mathbf{H}$, de (5.15) con respecto a $\mathbf{z}_f$, lo que permite determinar la dirección óptima, $\mathbf{s}$, en que se debe modificar la partición $\mathbf{z}_f$, y que viene dada por la expresión

$$\mathbf{s} = -\mathbf{H}^{-1}\mathbf{g}. \tag{5.17}$$

El gradiente se calcula como la derivada parcial de la expresión de error con respecto a cada uno de los parámentros $\mathbf{z_f}$, y puesto que la función de salida (y, por tanto, el error) es N-dimensional, se evalúa como un sumatorio en las N dimensiones

$$\mathbf{g} = \nabla_{\mathbf{z}_f} E(\mathbf{Z}_c, \mathbf{z}_f) = \left[\frac{\partial E(\mathbf{Z}_c, \mathbf{z}_f)}{\partial z_{f1}}, \ldots, \frac{\partial E(\mathbf{Z}_c, \mathbf{z}_f)}{\partial z_{fPM}}\right]^T = 2\sum_{n=1}^{N} \mathbf{K}_n \mathbf{G} \mathbf{e}_n,$$

mientras que el Hessiano viene dado por

$$\mathbf{H} = \nabla^2_{\mathbf{z}_f} E(\mathbf{Z_c}, \mathbf{z}_f) = \nabla_{\mathbf{z}_f} \mathbf{g} = 2\sum_{n=1}^{N} \left(\mathbf{K}_n \nabla_{\mathbf{z}_f} \mathbf{G}\ \mathbf{e}_n + \mathbf{K}_n \mathbf{G}\mathbf{G}^T \mathbf{K}_n^T\right), \quad (5.18)$$

donde $\mathbf{e}_n$ es el vector de error del modelo según la dimensión n-ésima de salida, es decir, $\mathbf{e}_n = \left[e_n^{[1]}, \ldots, e_n^{[L]}\right]^T$; $\mathbf{K}_n$ es la matriz $MP \times MP$

$$\mathbf{K}_n = \mathrm{diag}(\underbrace{c_{n,1}, \ldots, c_{n,1}}_{M-1 \text{ veces}}, \underbrace{c_{n,2}, \ldots, c_{n,2}}_{M-1 \text{ veces}}, \ldots, \underbrace{c_{n,P}, \ldots, c_{n,P}}_{M-1 \text{ veces}}, c_{n,1}, c_{n,2}, \ldots, c_{n,P});$$

y $\mathbf{G}$ la matriz $MP \times L$

$$\mathbf{G} = \begin{bmatrix} x_1^{[1]} p_1^{[1]} & x_1^{[2]} p_1^{[2]} & \ldots & x_1^{[L]} p_1^{[L]} \\ \vdots & \vdots & & \vdots \\ x_{M-1}^{[1]} p_1^{[1]} & x_{M-1}^{[2]} p_1^{[2]} & \ldots & x_{M-1}^{[L]} p_1^{[L]} \\ \vdots & \vdots & & \vdots \\ x_1^{[1]} p_P^{[1]} & x_1^{[2]} p_P^{[2]} & \ldots & x_1^{[L]} p_P^{[L]} \\ \vdots & \vdots & & \vdots \\ x_{M-1}^{[1]} p_P^{[1]} & x_{M-1}^{[2]} p_P^{[2]} & \ldots & x_{M-1}^{[L]} p_P^{[L]} \\ p_P^{[1]} & p_P^{[2]} & \ldots & p_P^{[L]} \\ \vdots & \vdots & & \vdots \\ p_P^{[1]} & p_P^{[2]} & \ldots & p_P^{[L]} \end{bmatrix},$$

siendo $p_i^{[l]}$ el valor de la primera derivada de la función base i-ésima para el vector de entrada l-ésimo (5.9), es decir, $p_i^{[l]} = \varphi^{(1)}\left(d_i(\mathbf{x}^{[l]}), \alpha\right)$.

Es necesario mencionar que el primer término del Hessiano en (5.18) incorpora la segunda derivada de las funciones base del modelo αPWL, cuyo valor viene dado por (5.10). Se trata de una función localizada en las fronteras del dominio, por lo que únicamente aquellos datos de entrada próximos a las fronteras contribuyen a dicho término. En la práctica se ha observado que no

se aprecia degradación en las prestaciones del algoritmo si se prescinde de este factor, obteniéndose, por contra, un importante ahorro computacional. Por ello, la evaluación del Hessiano se llevará a cabo utilizando la expresión

$$\mathbf{H} = 2\sum_{n=1}^{N} \mathbf{K}_n \mathbf{G}\mathbf{G}^T \mathbf{K}_n^T.$$

Una vez calculada la dirección de búsqueda (5.17), las fronteras de la siguiente iteración se obtienen aplicando

$$\mathbf{z}_f' = \mathbf{z}_f + \mu \mathbf{s},$$

donde $\mu = \operatorname{argmin}\left(E(\mathbf{Z}_c, \mathbf{z}_f + \mu \mathbf{s})\right)$.

De esta forma se obtiene una nueva partición, $\mathbf{z}_f$, del dominio de entrada para la cual se deben calcular los coeficientes óptimos, $\mathbf{Z}_c$, utilizando (5.16). El proceso completo se repite de forma iterativa, adaptando sucesivamente la partición y los coeficientes hasta que se alcanza un determinado nivel de error o hasta que se sobrepasa un número máximo de iteraciones.

5.2.6. Estimación del parámetro de suavidad

El parámetro α incide de manera crucial en las características del modelo αPWL, principalmente controlando la suavidad del modelo. La estimación de dicho parámetro se puede llevar a cabo de diversas maneras, dependiendo en gran medida de la aplicación concreta.

La primera opción consiste en fijar un valor α de antemano, de forma que se asegure una determinada suavidad de la solución, lo que puede entenderse como una técnica de regularización como en (5.13).

Otra posibilidad consiste en seleccionar el valor de α que minimice el error cuadrático medio de la solución (5.15) utilizando, por ejemplo, el algoritmo de descenso del gradiente dado por la expresión

$$\alpha' = \alpha - \mu \frac{\partial E(\mathbf{Z_c}, \mathbf{z}_f)}{\partial \alpha}.$$

Si bien en la expresión general del modelo αPWL (5.11) se contempla la posibilidad de utilizar un parámetro de suavidad distinto para cada función base, el ligero incremento en las prestaciones que se obtiene no compensa, en la mayoría de las aplicaciones prácticas, el incremento en la complejidad (tanto en el entrenamiento como en la evaluación del modelo) y el aumento del número de parámetros que provoca.

Finalmente, en aquellas aplicaciones en las que se disponga de información acerca del valor de la(s) derivada(s) de la función a modelar, es posible la

búsqueda del valor óptimo de α que produzca resultados acordes con dicha información. En ese caso, se puede proceder a ajustar los parámetros que definen las fronteras, $\mathbf{z}_f$, y la combinación lineal, $\mathbf{Z}_c$, del modelo mediante el algoritmo que minimiza (5.15) descrito previamente y adaptar el valor de α para minimizar, por ejemplo, el error cuadrático medio entre la derivada del modelo y los datos. De esta forma, si se dispone de los $L \times N$ valores de la derivada de la función de salida con respecto a la variable j-ésima de entrada

$$\frac{d\mathbf{y}^{[l]}}{dx_j} = \mathbf{y}^{(1_j)[l]} = [y_1^{(1_j)[l]}, y_2^{(1_j)[l]}, \ldots, y_N^{(1_j)[l]}]^T,$$

entonces se puede evaluar la norma l_2 del error entre el valor de dichas medidas y la derivada con respecto a x_j proporcionada por el modelo αPWL

$$\begin{aligned} E^{(1_j)}(\mathbf{Z}_c, \mathbf{z}_f) &= \sum_{l=1}^{L} \left\| \mathbf{y}^{(1_j)[l]} - \left(\mathbf{b}_j + \sum_{i=1}^{P} \mathbf{c}_i m_{i,j} \varphi^{(1)} \left(d_i(\mathbf{x}^{[l]}), \alpha \right) \right) \right\|^2 \\ &= \sum_{l=1}^{L} \sum_{n=1}^{N} \left(e_n^{(1_j)[l]} \right)^2, \end{aligned} \tag{5.19}$$

donde $\mathbf{b}_j$ representa la j-ésima columna de la matriz de coeficientes $\mathbf{B}$, y donde $e_n^{(1_j)[l]}$ representa la diferencia entre la verdadera derivada con respecto a x_j y la estimada por el modelo αPWL para el conjunto de datos l-ésimo y según la dirección n-ésima del espacio de salida

$$e_n^{(1_j)[l]} = y_n^{(1_j)[l]} - \left(b_{n,j} + \sum_{i=1}^{P} c_{n,i} m_{i,j} \varphi^{(1)} \left(d_i(\mathbf{x}^{[l]}), \alpha \right) \right).$$

La derivada del error (5.19) con respecto a α es

$$\begin{aligned} \frac{\partial E^{(1_j)}(\mathbf{Z}_c, \mathbf{z}_f)}{\partial \alpha} &= 2 \sum_{l=1}^{L} \sum_{n=1}^{N} e_n^{(1_j)[l]} \frac{\partial e_n^{(1_j)[l]}}{\partial \alpha} \\ &= 2 \sum_{l=1}^{L} \sum_{n=1}^{N} e_n^{(1_j)[l]} \left(\sum_{i=1}^{P} \mathbf{c}_i m_{i,j} \frac{d(\mathbf{x})}{\alpha} \varphi^{(2)} \left(d(\mathbf{x}), \alpha \right) \right); \end{aligned} \tag{5.20}$$

donde, haciendo uso de (5.9) y (5.9), se ha considerado

$$\frac{d\varphi^{(1)}\left(d(\mathbf{x}), \alpha\right)}{d\alpha} = -\frac{d(\mathbf{x})}{\alpha} \varphi^{(2)}\left(d(\mathbf{x}), \alpha\right).$$

Por lo tanto, se puede llevar a cabo la búsqueda del parámetro de suavidad α mediante un algoritmo iterativo de descenso del gradiente que haga uso de (5.20)

$$\alpha' = \alpha - \mu \frac{\partial E^{(1_j)}(\mathbf{Z}_c, \mathbf{z}_f)}{\partial \alpha},$$

donde μ es el parámetro de paso del algoritmo.

5.2.7. Prestaciones

Dada la gran similitud de las funciones base (y sus derivadas) del modelo αPWL y del modelo SPWL propuesto en [Lázaro01b], así como de los respectivos algoritmos de entrenamiento, se puede afirmar (y se ha comprobado mediante simulaciones, así como en diversas aplicaciones prácticas) que la capacidad de aproximación del modelo propuesto en esta Tesis es similar a la del modelo SPWL. La única diferencia reseñable, de gran importancia en aplicaciones prácticas, radica en los menores requerimientos computacionales del modelo αPWL, motivado por la mayor simplicidad de las funciones bases y sus derivadas (polinomios de grado tres) comparadas con las funciones que intervienen en el modelo SPWL: $\ln(\cosh(x))$ para la función base y $\tanh(x)$ y $\text{sech}^2(x)$ para la primera y segunda derivada, respectivamente.

Es necesario mencionar que, al igual que ocurre con el modelo SPWL (y en general, al igual que ocurre con todos los modelos derivados del modelo canónico lineal a tramos), el modelo propuesto en esta Tesis presenta el problema de la presencia de mínimos locales en la superficie de error definida por (5.15), lo que puede conducir a la obtención de soluciones subóptimas, dependiendo de la inicialización de las fronteras que particionan el espacio de entrada. Con el fin de paliar este problema, se llevarán a cabo varios entrenamientos del modelo (partiendo de inicializaciones distintas) y se seleccionará la mejor solución obtenida, es decir, aquella que proporciona un menor error cuadrático entre la estimación y las medidas.

5.3. Modelado de amplificadores de potencia

En esta sección se presentará una metodología novedosa de caracterización de amplificadores de potencia de radiofrecuencia mediante técnicas de submuestreo [Ibáñez00, Santamaría03]. Las señales temporales así obtenidas posibilitarán la obtención de diversos modelos no lineales, entre los que se incluye el modelo αPWL propuesto en esta Tesis; esto permitirá comparar sus prestaciones a la hora de modelar este tipo de dispositivos cuando se excitan con señales de entrada de banda ancha. Previamente se llevará a cabo un repaso del estado del arte en este ámbito.

5.3.1. Estado del arte

Los amplificadores de potencia (HPA: *High Power Amplifier*) presentan un comportamiento claramente no lineal, salvo que se utilicen en regiones muy

alejadas de la saturación,[2] lo cual resulta muy negativo desde el punto de vista económico y de la eficiencia energética. Por esta razón, disponer de modelos precisos y representativos de los amplificadores de potencia constituye un factor fundamental a la hora de simular los sistemas completos de comunicaciones y resulta indispensable para desarrollar técnicas de compensación que linealicen su comportamiento (por ejemplo, por medio de predistorsión).

Esta Tesis se ocupa de la distorsión introducida por amplificadores de potencia cuando la señal de entrada es de banda ancha. Sin embargo, tradicionalmente las medidas, la caracterización y los modelos se basan en excitar los dispositivos mediante señales de onda continua (tonos puros en el centro de la banda de interés) y en realizar un barrido de la amplitud de las mismas. Sin duda, las curvas AM-AM y AM-PM son las medidas más habituales de este tipo de caracterización, en la que se supone un comportamiento del dispositivo independiente de la frecuencia (o, en otras palabras, se supone que el amplificador no posee memoria) en todo el ancho de banda de la señal de comunicaciones, lo que, en principio, restringe su uso a aplicaciones de banda estrecha.

Como ejemplo, el modelo de Saleh [Saleh81] (sin duda, el utilizado más extensamente) ajusta las curvas AM-AM y AM-PM mediante fórmulas analíticas dependientes de dos parámetros. Con el fin de extender su aplicación a señales de banda ancha, diversos trabajos [Saleh81, Abuelma'atti84, Jeruchim00] proponen considerar los parámetros del modelo como funciones de la frecuencia, pero, en cualquier caso, el ajuste de los mismos se lleva a cabo a partir de medidas del comportamiento del amplificador ante señales de excitación de onda continua, no utilizando señales de banda ancha. Este hecho cuestiona la validez de los modelos en el contexto de los modernos sistemas de comunicaciones digitales (OFDM, CDMA, QAM, ...) en los que las señales poseen un considerable ancho de banda y una envolvente no constante.[3] Por otro lado, las medidas de onda continua no permiten caracterizar de forma precisa fenómenos tales como el recrecimiento espectral (*spectral regrowth*) y la distorsión de intermodulación, por lo que los modelos derivados de ellas serán incapaces de representar el comportamiento real del amplificador.

A pesar de todo, el modelo de Saleh se sigue utilizando aún en la actualidad para proponer y analizar distintas técnicas de linealización en sistemas de banda ancha [González01, D'Andrea01]. Es más, el rendimiento de dichas técnicas se evalúa por medio de simulaciones, por lo que es de esperar que los desajustes entre el modelo y el amplificador real provoquen cierta degradación en las prestaciones del sistema de linealización empleado.

[2] Una medida habitual es el denominado OBO (*Output Back-Off*) obtenido como el cociente (expresado en decibelios) entre la potencia de saturación del amplificador (máxima potencia posible a la salida) y la potencia de salida actual.

[3] Evaluado en términos del PAPR (*Peak-to-Average Power Ratio*): relación entre la potencia pico instantánea y la potencia media.

Una de las contribuciones de esta Tesis consiste en obtener modelos no lineales de HPA ajustando las señales temporales de entrada-salida (en concreto la envolvente compleja) para señales de entrada de banda ancha. En los últimos años se han propuesto diversos métodos de caracterización de amplificadores de potencia de radiofrecuencia en el dominio temporal [Heutmaker97, Borich98]. En general, estas técnicas muestrean una versión demodulada de la señal de salida del amplificador, por lo que precisan etapas de conversión de frecuencia (moduladores y demoduladores) y preamplificadores. Dichos dispositivos deben ser extremadamente lineales, con el fin de evitar una distorsión no lineal añadida que desvirtúe la caracterización. En [Clark88] se presenta una solución que cancela los errores derivados de la conversión de frecuencia, aunque a costa de una calibración precisa de los conversores y de un montaje realmente complicado.

En esta Tesis se propone utilizar técnicas de submuestreo[4] para muestrear directamente la entrada y la salida (atenuadas si es preciso) de radiofrecuencia del HPA. Utilizando este esquema de adquisición es posible desarrollar modelos a partir de muestras temporales submuestreadas.

En las secciones siguientes se describirá el sistema de adquisición y caracterización basado en submuestreo, se propondrá la aplicación del modelo α-spline y se compararán sus prestaciones con otros modelos no lineales.

5.3.2. Sistema de caracterización basado en submuestreo

En los últimos años, los conversores analógico-digital han experimentado significativos avances tanto en lo que respecta a la velocidad como a la resolución, a la vez que los precios han disminuido progresivamente. De forma paralela, los avances en el procesado digital de señal han sido evidentes y se ha consolidado la tendencia de sustituir el procesado analógico de señal por su equivalente digital. Sin embargo, el conocido límite de Nyquist representa un serio obstáculo que frena la aplicación de las técnicas de procesado digital de señal a frecuencias por encima de unos cientos de MHz.

Pero si la señal es de tipo paso banda, es posible preservar toda su información muestreando a frecuencias inferiores a la frecuencia de Nyquist. Es la denominada técnica de submuestreo (o muestreo paso banda) [Vaughan91, Hill94, Wepman00], que permite una perfecta reconstrucción de la señal paso banda original si la frecuencia de muestreo es superior al doble del ancho de banda de la señal. Una limitación práctica importante es que el ADC debe ser capaz de operar a la frecuencia más alta de la señal paso banda. El submuestreo es una técnica utilizada en diversas aplicaciones, como por ejemplo, en los receptores de radiocomunicación, en los que se digitaliza la señal recibida

[4] Por submuestreo se entiende el muestreo de una señal modulada utilizando una frecuencia de muestreo inferior a la de Nyquist, con el fin de llevar a cabo una traslación frecuencial por medio de un solapamiento espectral intencionado.

en frecuencia intermedia o, incluso, en radiofrecuencia (lo que se conoce en la actualidad como *software radio*). En esta Tesis se utiliza esta técnica para caracterizar dispositivos y desarrollar modelos no lineales a partir de los datos obtenidos.

5.3.2.1. Submuestreo

El submuestreo, o muestreo paso banda, consiste en muestrear una señal paso banda (típicamente una señal modulada) utilizando una frecuencia de muestreo inferior a la frecuencia de Nyquist, lo que permite realizar una traslación frecuencial mediante un solapamiento o *aliasing* intencionado y controlado. Una señal paso banda ideal no posee componentes frecuenciales por encima de cierta frecuencia f_h ni por debajo de una cierta frecuencia f_l. Para posibilitar una reconstrucción exacta de este tipo de señales, es preciso que la frecuencia de muestreo sea, al menos, igual al doble de su ancho de banda, $f_h - f_l$, es decir

$$f_s \geq 2(f_h - f_l).$$

Adicionalmente, para evitar que se produzca solapamiento espectral de las componentes de la señal, la frecuencia de muestreo, f_s, debe cumplir

$$\frac{2f_h}{k} \leq f_s \leq \frac{2f_l}{k-1},$$

donde k es un valor entero que satisface

$$2 \leq k \leq \frac{f_h}{f_h - f_l},$$

y

$$f_h \leq 2f_l.$$

El submuestreo reduce drásticamente la velocidad de muestreo (k veces), pero mantiene la limitación de que el conversor analógico-digital debe ser capaz de operar a la frecuencia más alta que posea la señal: el circuito de muestreo y retención (*sample-and-hold*), que constituye la etapa de entrada de todo ADC, debe ser capaz de seguir señales de frecuencia f_h, capturarlas y mantenerlas el tiempo necesario para que se lleve a cabo la cuantificación, que se realizará a una tasa de $f_s \ll 2f_h$ muestras por segundo. Así mismo, puede ser necesario utilizar filtros analógicos paso banda que eviten la distorsión producida por ruido y señales interferentes fuera de banda.

5.3.2.2. Esquema de adquisición

En los últimos años se han desarrollado diversos métodos para la caracterización temporal de amplificadores de potencia de RF [Heutmaker97, Borich98].

Todos ellos utilizan un esquema basado en el muestreo en banda base de la señal demodulada, tal y como muestra la Figura 5.5a. En este tipo de montajes, la señal de test se genera digitalmente y se realiza una conversión de frecuencia que la sitúa en la banda de radiofrecuencia de interés. La señal debe mantenerse en unos niveles de potencia suficientemente bajos para evitar la distorsión del conversor de frecuencia, por lo que se precisa introducir un preamplificador muy lineal y capaz de proporcionar unos niveles de señal adecuados para excitar al amplificador de potencia de forma que se pueda observar su respuesta en torno a la región de saturación. Tras el amplificador se sitúa un filtro paso banda que elimina toda señal fuera de la llamada *primera zona*[5] seguido por un atenuador que minimiza la distorsión de la conversión a banda base. Finalmente, la señal banda base se muestrea a la frecuencia de Nyquist.

La desventaja principal de este esquema reside en los altos requerimientos de linealidad que se exigen al preamplificador y a los conversores (tanto hacia arriba como hacia abajo) de frecuencia. En cualquier caso, este método es incapaz de aislar el comportamiento del amplificador de potencia, puesto que las medidas incluyen los efectos de la totalidad de la cadena (modulación, amplificación y demodulación). En [Clark88] se propone una solución que elimina los errores derivados de las conversiones frecuenciales por medio de una cuidadosa calibración de los conversores, a costa de un montaje final notablemente complejo.

Como alternativa, en esta Tesis [Ibáñez01] se propone un montaje basado en la técnica de submuestreo (Figura 5.5b). Un generador digital de señal proporciona la señal de test, que es subida en frecuencia y preamplificada con el fin de excitar el amplificador de potencia. Una réplica exacta de la señal de entrada del HPA, proporcionada por un divisor, es filtrada, atenuada (cuando el nivel de señal es excesivamente alto) y submuestreada por uno de los canales del conversor analógico-digital. De la misma forma, el otro canal del ADC submuestrea la señal de salida del HPA tras los correspondientes filtrado y atenuación.

La principal ventaja de este esquema de medida es que caracteriza únicamente el comportamiento del amplificador de potencia, pues se adquiere tanto la señal de salida como la de entrada, ambas en radiofrecuencia. Además de la ausencia de la etapa de conversión a banda base, otra ventaja es que este montaje hace innecesario corregir los errores de las etapas de conversión de frecuencia, puesto que se adquieren las verdaderas señales de entrada y salida. Finalmente, también se evitan los problemas relacionados con los errores de temporización y retardos introducidos por los conversores de frecuencia.

[5] Primera zona o zona principal se refiere al rango de frecuencias en torno a la frecuencia central de una señal o sistema pasobanda. En la salida de sistemas no lineales aparecen componentes frecuenciales fuera de la primera zona, que suelen ser filtrados por filtros o circuitos de sintonía, por lo que habitualmente se descartan a la hora de caracterizar el comportamiento de este tipo de dispositivos.

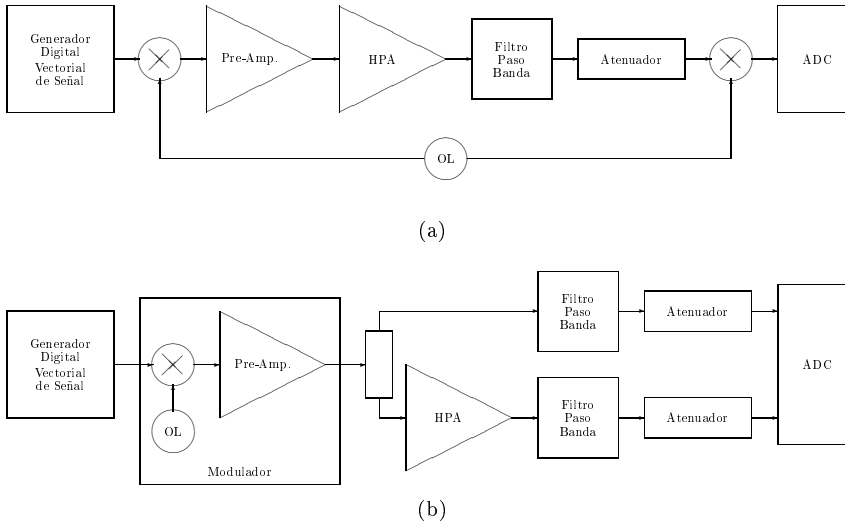


Figura 5.5: Esquema de caracterización temporal de HPA de radiofrecuencia: a) tradicional, b) propuesto.

El conversor analógico-digital constituye un elemento crucial en este esquema de submuestreo, pues debe poseer un ancho de banda de entrada suficientemente grande como para ser capaz de seguir las componentes de mayor frecuencia de la señal (en la banda de la portadora). De igual forma, debe proporcionar una frecuencia de muestreo adecuada para capturar el ancho de banda de interés de la señal paso banda. En la actualidad, este último requisito no es muy difícil de cumplir, pues existen conversores comerciales que alcanzan cientos de megamuestras por segundo e, incluso, varios gigamuestras por segundo. Estas prestaciones permiten estudiar el comportamiento de los amplificadores cuando se excitan con señales de gran ancho de banda y/o relajar las especificaciones de los filtros paso banda antisolapamiento y/o incrementar la resolución del ADC por medio de técnicas de procesado de señal (sobremuestreo de la señal submuestreada y posterior diezmado [Candy92]) .

5.3.3. Medidas experimentales

El amplificador de potencia utilizado en esta Tesis está basado en un transistor MESFET de arseniuro de galio de Motorola (en concreto, el MRFC181) que proporciona una potencia de salida de 33 dBm con una ganancia de 30 dB utilizando una alimentación de $4,8$ V y sintonizado para proporcionar máxima potencia a 1,45 GHz.

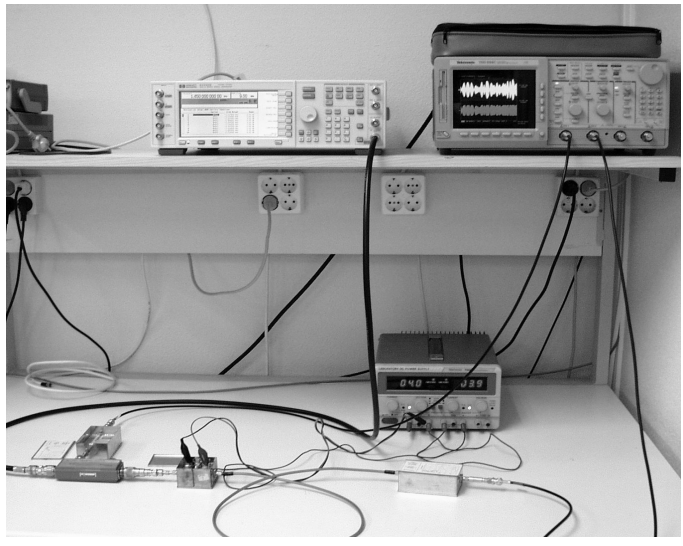


Figura 5.6: Banco de caracterización de amplificadores mediante submuestreo.

La Figura 5.6 muestra el montaje experimental descrito previamente y correspondiente al esquema de la Figura 5.5b. Un generador digital vectorial de señal (el Agilent E4432B ESG-D) proporciona la señal multiportadora centrada en la frecuencia de interés y con el nivel de potencia adecuado; seguidamente la señal atraviesa un filtro paso banda sintonizado a 1,45 GHz y con un ancho de banda de 80 MHz; y, finalmente, la señal es adquirida utilizando un osciloscopio digital (Tektronix TDS694C) capaz de muestrear hasta 10 gigamuestras por segundo, con una capacidad de almacenamiento de 120000 muestras y un ancho de banda de entrada de 3 GHz. Una réplica exacta de esta señal, proporcionada por el divisor, es amplificada por el amplificador de potencia en estudio, filtrada paso banda, atenuada y adquirida por medio del segundo canal del osciloscopio.

En este estudio, el generador E4432B produce señales OFDM con 64 subportadoras cuya separación es controlada (Δf = 45, 60, 75, 90, 105, 120, 135 y 150 KHz) con el fin de obtener anchos de banda, (BW_i) entre 3 y 10 MHz, aproximadamente. De la misma forma, la potencia de entrada al amplificador se controla por medio del propio generador (P_i = 0, 3, 6 y 9 dBm) cubriendo todo el comportamiento del amplificador, desde un punto de trabajo prácticamente lineal hasta una operación muy saturada. Por último se consideran distintos formatos de modulación para cada subportadora (por ejemplo, BPSK, QPSK y 64QAM). Sin embargo, las principales conclusiones que se extraen de este estudio no dependen del tipo de modulación empleado en las subportadoras.

El procedimiento seguido para adquirir la envolvente compleja de las señales en el dominio temporal en cada uno de los experimentos se describe a continuación. En primer lugar, el osciloscopio digital adquiere las señales de entrada y de salida utilizando una frecuencia de muestreo de 1,25 GHz, por lo que, debido al submuestreo, las señales resultan centradas en 1,45 GHz − 1,25 GHz = 200 MHz. Seguidamente, los registros son transferidos a un ordenador personal vía GPIB (*General Purpose Interface Bus*, también conocido como bus IEEE488.1).

Dado que los filtros paso banda de la Figura 5.5b no son idénticos, existe un retardo (adicional al introducido por el propio amplificador de potencia) entre la señal de entrada y la de salida que debe ser corregido antes de llevar a cabo posteriores procesamientos. Este retardo se estima mediante la búsqueda del máximo de la correlación cruzada entre las envolventes complejas de entrada y salida. Conviene resaltar que este retardo se estima a la máxima frecuencia de muestreo (en este caso 1,25 GHz), por lo que el retardo que se pueda atribuir erróneamente al HPA siempre será inferior al periodo de muestreo ($T = 0{,}8$ ns). Utilizando esta estimación lineal del retardo, se dispone de las señales envolvente compleja de entrada y de salida adecuadamente alineadas en el tiempo.

El siguiente paso en el preprocesamiento consiste en demodular las señales utilizando una secuencia exponencial compleja ($g[n] = e^{-j2\pi 0{,}16n}$) que desplace la parte positiva del espectro de la señal OFDM hasta la frecuencia cero (en definitiva, se demodula la señal digitalmente). Las señales complejas resultantes se filtran paso bajo utilizando un filtro FIR de 175 coeficientes cuyas especificaciones son: frecuencia de corte de 15 MHz, banda de transición de 7 MHz, rechazo fuera de banda de 60 dB y rizado en banda de 1dB. Finalmente las señales son diezmadas por un factor de cuarenta, por lo que la frecuencia de muestreo final se sitúa en 31,25 MHz. De esta forma, la envolvente compleja de la señal OFDM de mayor ancho de banda ocupa una banda de $0-5$ MHz, lo que supone un factor de sobremuestreo en torno a tres, que puede ser considerado como suficiente para caracterizar el recrecimiento espectral del amplificador. Con estos parámetros, la relación señal a ruido de los registros adquiridos se sitúa ligeramente por encima de 35 dB, valor que impondrá una cota superior al rendimiento que podrá proporcionar un modelo perfecto del amplificador de potencia derivado de dichos registros.

La longitud de los registros almacenados tras el diezmado es de 3000 muestras, por lo que al repetir cada experimento tres veces se dispone de 9000 muestras de cada una de las señales envolvente compleja de entrada y de salida para cada pareja (BW_i, P_i) considerada.

A modo de ilustración de las posibilidades que el esquema de medida propuesto brinda a la hora de capturar las características no lineales del amplificador, en la Figura 5.7 se muestra la densidad espectral de potencia de las señales de entrada y salida capturadas cuando se excita el amplificador con una señal

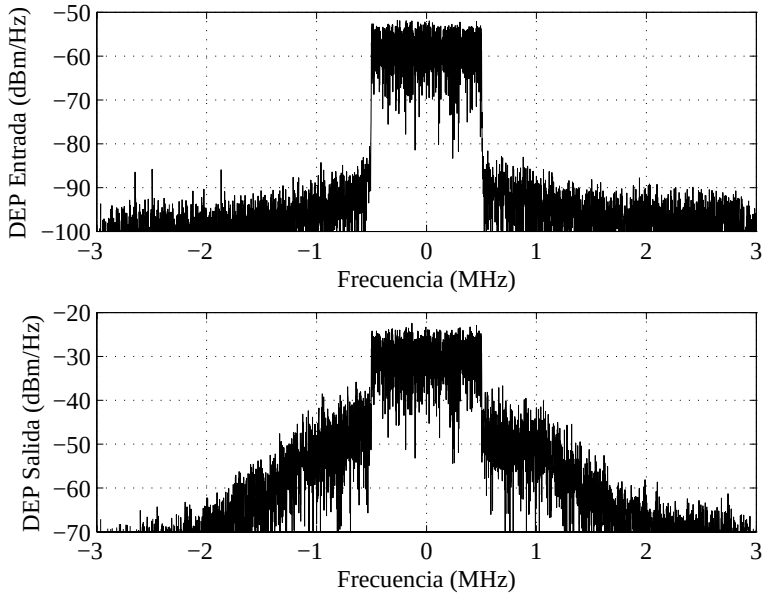


Figura 5.7: Densidad espectral de potencia de entrada (arriba) y salida (abajo) para una señal OFDM ($BW_i = 1$ MHz, $P_i = 0{,}3$ dBm).

OFDM de 1 MHz de ancho de banda y una potencia de 0,3 dBm (que produce una salida de 28,7 dBm); observándose claramente el recrecimiento espectral que sufre la señal de salida. Para esa misma señal de excitación, la Figura 5.8 muestra las secuencias de la amplitud de la envolvente compleja de entrada y de salida, apreciándose claramente el efecto de saturación del amplificador. Por último, la Figura 5.9 representa la relación entrada-salida de la amplitud de la envolvente compleja, permitiendo apreciar el comportamiento prácticamente lineal cuando la amplitud de entrada es baja y una severa saturación cuando la entrada alcanza valores elevados.

Es posible apreciar otros efectos, aparte de la saturación, cuando se dispone de medidas temporales de entrada-salida: tal y como muestra la Figura 5.9, el amplificador presenta histéresis, incluso en la región lineal, lo que sugiere cierta memoria. Éste es un aspecto crucial a la hora de modelar un sistema no lineal como el amplificador de potencia, pues si éste presenta memoria, el modelo deberá igualmente incorporarla. En la Figura 5.10 se aprecia aún más claramente la histéresis del amplificador basado en el MRFC1818 cuando la entrada es una señal modulada en amplitud con doble banda lateral (AM-DSB con índice de modulación del 70 % y señal moduladora sinusoidal de 20 KHz) y una potencia de entrada $P_i = -1{,}6$ dBm.

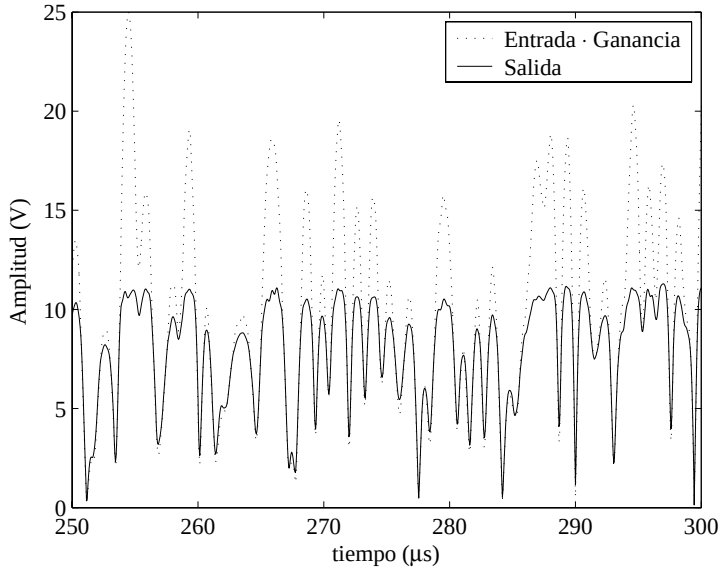


Figura 5.8: Amplitud de la envolvente compleja para una señal OFDM ($BW_i = 1$ MHz, $P_i = 0{,}3$ dBm).

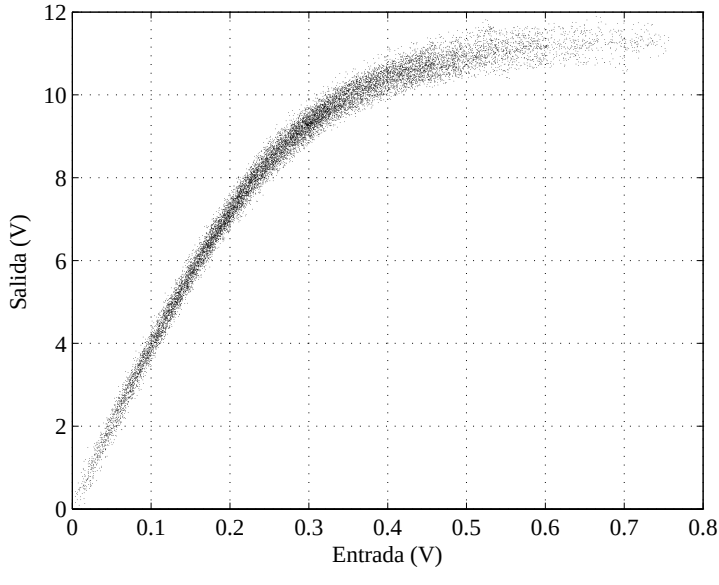


Figura 5.9: Gráfica entrada-salida de la amplitud de la envolvente compleja de una señal OFDM ($BW_i = 1$ MHz, $P_i = 0{,}3$ dBm).

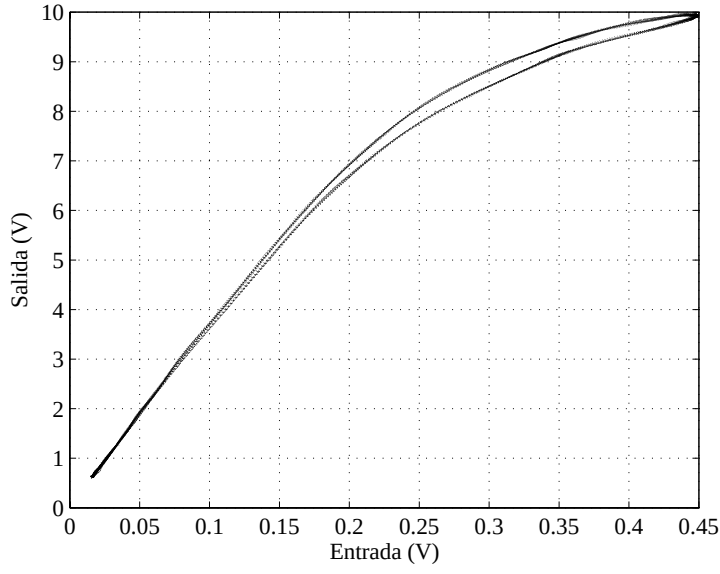


Figura 5.10: Gráfica entrada-salida de la amplitud de la envolvente compleja de una señal DSB-AM ($BW_i = 40$ KHz, $P_i = -1{,}6$ dBm).

La Figura 5.11 presenta otro ejemplo que muestra la severidad del comportamiento no lineal del amplificador. En este caso, se representa la constelación a la salida del amplificador cuando se excita con una señal OFDM-64QAM de 6 MHz de ancho de banda y 3 dBm de potencia. Contrariamente a lo que ocurre con señales de portadora única (para las que los efectos de compresión y deformación se aprecian nítidamente en la constelación), en sistemas multiportadora la distorsión no lineal provoca tres efectos: una rotación de fase, una ligera deformación de la constelación y, fundamentalmente, una distorsión que puede ser modelada como ruido aditivo. Teniendo en cuenta que una señal OFDM con un número suficientemente elevado de subportadoras puede modelarse por un proceso aleatorio Gaussiano complejo (cuyos módulo y fase siguen distribuciones Rayleigh y uniforme, respectivamente), este tipo de distorsión no lineal puede caracterizarse teóricamente como se muestra en [Dardari00, Banelli00].

5.3.4. Resultados y comparación de modelos

En esta sección se lleva a cabo la particularización del modelo αPWL y una breve descripción de otros modelos no lineales aplicados al modelado de amplificadores basado en datos temporales adquiridos mediate submuestreo. Para cada modelo se han estudiado configuraciones polares (modulo/fase) y

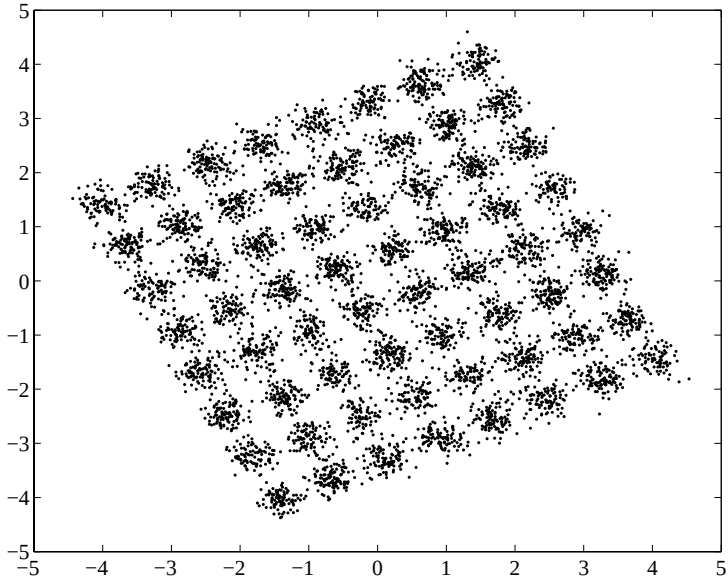


Figura 5.11: Constelación a la salida del amplificador para una señal de entrada OFDM-64QAM ($BW_i = 10$ MHz, $P_i = 3$ dBm).

en cuadratura (I/Q), si bien la estructura en cuadratura proporciona mejores resultados, por lo que únicamente consideraremos este tipo de modelos. Tras discutir la metodología empleada para estimar la memoria del sistema a modelar, se presentarán los distintos resultados.

5.3.4.1. Modelo αPWL de amplificadores de potencia

Los datos experimentales del amplificador son utilizados para entrenar el modelo αPWL propuesto en esta Tesis, con la particularidad de que, en esta aplicación, el espacio de entrada posee $M = 4$ dimensiones, al considerar las variables correspondientes a la señales de entrada en fase y cuadratura ($x_I[n]$ y $x_Q[n]$) y una versión retardada de las mismas ($x_I[n-d]$ y $x_Q[n-d]$). Como se demostrará posteriormente, el modelo de amplificador más adecuado para esta aplicación posee únicamente un coeficiente de memoria, motivo por el cual se precisan cuatro variables de entrada. Respecto al espacio de salida, evidentemente se trata de un espacio bidimensional, $N = 2$, pues la envolvente compleja de salida del amplificador posee componentes en fase y cuadratura ($y_I[n], y_Q[n]$). Al considerar una partición del espacio de entrada por medio de P fronteras, resulta un modelo que denotaremos como αPWL(4,P,2) y que

representa un mapeo ($f : \mathbb{R}^4 \rightarrow \mathbb{R}^2$) dado por la expresión

$$\mathbf{y}_n = f(\mathbf{x}_n) = \mathbf{a} + \mathbf{B}\mathbf{x}_n + \sum_{i=1}^{P} \mathbf{c}_i \varphi\left(\langle \boldsymbol{\nu}_i, \mathbf{x}_n \rangle - \beta_i, \alpha_i\right), \tag{5.21}$$

donde $\mathbf{a}$ y $\mathbf{c}_i$ son vectores de dimensión 2×1, $\boldsymbol{\nu}_i$ es un vector 4×1, $\mathbf{B}$ una matriz 2×4, β_i un escalar y $\langle \cdot, \cdot \rangle$ representa el producto interno. La adaptación de los parámetros de este modelo (las fronteras de partición del espacio de entrada: $\boldsymbol{\nu}_i$ y β_i; los valores de la combinación lineal de las componentes: $\mathbf{a}$, $\mathbf{B}$ y $\mathbf{c}_i$; y el parámetro de suavidad α_i) se realiza de forma iterativa, siguiendo los planteamientos descritos en el Apartado 5.2.5.

5.3.4.2. Otros modelos no lineales

Probablemente, el modelo no lineal sin memoria más ampliamente usado para modelar amplificadores de potencia es el modelo de Saleh [Saleh81], que considera que si la señal paso banda de entrada muestreada es

$$x[n] = a[n] \cos\left(\omega_0 n + \phi[n]\right),$$

entonces la salida se escribe como

$$\begin{aligned} y[n] &= A(a[n]) \cos\left(\omega_0 n + \phi[n] + \Phi(a[n])\right) \\ &= I(a[n]) \cos\left(\omega_0 n + \phi[n]\right) - Q(a[n]) \sin\left(\omega_0 n + \phi[n]\right), \end{aligned}$$

donde

$$A(a[n]) = \frac{\alpha_a a[n]}{1+\beta_a a^2[n]}, \tag{5.22a}$$

$$\Phi(a[n]) = \frac{\alpha_\Phi a^2[n]}{1+\beta_\Phi a^2[n]}, \tag{5.22b}$$

representan, respectivamente, las curvas AM/AM y AM/PM del amplificador; y donde

$$I(a[n]) = \frac{\alpha_i a[n]}{1+\beta_i a^2[n]}, \tag{5.22c}$$

$$Q(a[n]) = \frac{\alpha_q a^3[n]}{(1+\beta_q a^2[n])^2}, \tag{5.22d}$$

son las funciones correspondientes a la rama en fase y en cuadratura. Típicamente, los cuatro parámetros del modelo (bien del modelo polar o bien del modelo en cuadratura) se obtienen mediante el ajuste de las ecuaciones (5.22) a las medidas de la relación entre la amplitud de entrada y la amplitud y el desfase a la salida para una señal de entrada de onda continua (es decir, un tono de cierta frecuencia). Sin embargo, dado que se pretende desarrollar modelos específicos para señales de banda ancha (en particular, señales OFDM),

en esta Tesis se ajustarán los parámetros de este modelo en base a las muestras adquiridas de la señal envolvente compleja de entrada y de salida.

El modelo de Saleh asume que las características del HPA son independientes de la frecuencia, es decir, se trata de un modelo sin memoria. Sin embargo, cuando se consideran señales de entrada de gran ancho de banda (o, más concretamente, cuando el ancho de banda de la señal de entrada es comparable al ancho de banda del dispositivo[6]) se hace preciso un modelo dependiente de la frecuencia. Para tener en cuenta estos efectos de memoria, en esta Tesis se considera la utilización de muestras pasadas de las señales envolvente compleja a la hora de estimar el modelo adecuado del amplificador. Así, denotando por $x[n]$ e $y[n]$ las señales envolvente compleja de entrada y salida, los modelos no lineales que se consideran pueden ser expresados por medio del siguiente mapeo no lineal:

$$(y_I[n], y_Q[n]) = f\left(x_I[n], x_Q[n], \ldots, x_I[n-d], x_Q[n-d]\right). \tag{5.23}$$

La elección del máximo retardo, d, desempeña un papel crucial en el rendimiento de este tipo de modelos. Su valor depende de las características concretas del amplificador y de otros factores, como el ancho de banda de la señal y el factor de sobremuestreo de las medidas. En esta Tesis, se utiliza la información mutua entre la secuencia $y[n]$ y la secuencia retardada $x[n-k]$ como criterio para estimar el valor óptimo del retardo d. La información mutua retardada fue propuesta por A. M. Fraser y H. L. Swinney [Fraser86] como una herramienta para estimar el retardo en este tipo de situaciones, en las que es preciso tener en cuenta relaciones no lineales, aspecto éste que no es considerado por métodos más simples basados en la función de correlación cruzada. En particular, un estudio detallado, que se describirá posteriormente, de las series temporales de entrada-salida del amplificador concluye que su memoria está constituida únicamente por una muestra, es decir, cualquiera de los modelos con memoria del HPA deberá utilizar los valores de la muestra actual y de la pasada de la señal envolvente compleja de entrada, no obteniéndose ninguna mejora significativa de rendimiento utilizando un número mayor de muestras de memoria.

El primer modelo con memoria utilizado es el modelo polinómico de envolvente (EPM: *Envelope Polynomial Model*) [Jeruchim00, Ibáñez01] que aparece representado en la Figura 5.12. Los submodelos de fase y cuadratura de orden

[6] Por encima de 40 MHz en el caso de amplificadores basados en tubos de onda progresiva (TWT: *Traveling-Wave Tube*) y a partir unos pocos megahercios para los amplificadores de potencia de estado sólido (SSPA: *Solid State Power Amplifier*).

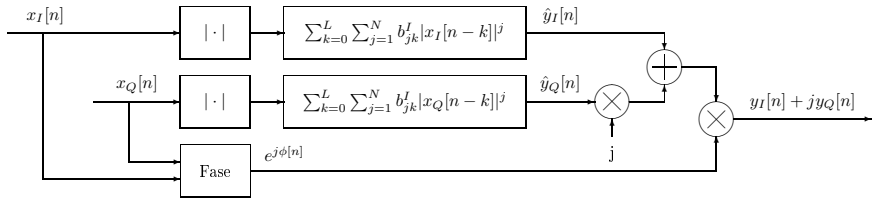


Figura 5.12: Modelo polinómico de envolvente.

(L, N) poseen las siguientes relaciones entrada-salida:

$$\hat{y}_I[n] = \sum_{k=0}^{L}\sum_{j=1}^{P} b_{jk}^{I}|x_I[n-k]|^j,$$

$$\hat{y}_Q[n] = \sum_{k=0}^{L}\sum_{j=1}^{P} b_{jk}^{I}|x_Q[n-k]|^j,$$

donde L denota la memoria del sistema y P el máximo grado de los polinomios (nótese que no se consideran sus términos constantes). En el modelo, los polinomios operan sobre el módulo de las componentes I y Q, mientras que la fase de la envolvente compleja de entrada se suma en la salida.

Un modelo polinómico con memoria más genérico consiste en una representación del amplificador de potencia mediante series de Volterra. En particular, se considera una forma de las series de Volterra adecuada para representar sistemas paso banda

$$y[n] = \sum_{k=0}^{P}\left(\frac{\binom{2k+1}{k}}{2^{2k}}\sum_{l_1=0}^{L}\cdots\sum_{l_{2k+1}=0}^{L} h_{2k+1}[l_1,\ldots,l_{2k+1}]\prod_{r=1}^{k} x^*[n-l_r]\prod_{s=k+1}^{2k+1} x[n-l_s]\right),$$

donde $x[n]$ e $y[n]$ son, respectivamente, las envolventes complejas de entrada y salida, y $h_{2k+1}[l_1,\ldots,l_{2k+1}]$ representan los núcleos (*kernel*) equivalentes paso bajo de la serie de Volterra. La ecuación anterior representa una expansión en serie de Volterra de un sistema causal paso banda en el que los términos que no generan señal en torno a la frecuencia central (en la *primera zona*) han sido eliminados de la serie. La complejidad de la serie de Volterra depende del número de términos impares de la expansión (1, 3,..., $2P+1$) así como de su memoria L: el modelo lo denotaremos como Volterra$(2P+1, L)$.

El cuarto modelo considerado en este estudio es el perceptrón multicapa, cuyo mapeo entrada-salida viene dado por

$$\mathbf{y}_n = \mathbf{W}_2^T \tanh\left(\mathbf{W}_1\mathbf{x}_n + \mathbf{b}_1\right) + \mathbf{b}_2,$$

donde $\mathbf{x}_n = (x_I[n], x_Q[n], x_I[n-1], x_Q[n-1])^T$ es el vector de entrada (correspondiente a la muestra actual y a la anterior), $\mathbf{y}_n = (y_I[n], y_Q[n])^T$ es el vector de salida, $\mathbf{W}_1$ es una matriz $P \times 4$ que conecta la capa de entrada con la capa oculta, $\mathbf{b}_1$ es el vector $P \times 1$ de los términos de sesgo de las neuronas ocultas, $\mathbf{W}_2$ es la matriz $P \times 2$ de pesos que conectan las neuronas de la capa oculta con las de salida, y $\mathbf{b}_2$ es el vector 2×1 de sesgo de las neuronas de salida. Por consiguiente, se trata de una estructura MLP$(4, P, 2)$, donde P denota el número de neuronas en la capa oculta, que se entrena utilizando el algoritmo de retropropagación [Haykin98] con el fin de minimizar el error cuadrático medio entre la salida obtenida y la deseada.

Por último, además del modelo αPWL propuesto en esta Tesis y dada su gran similitud con el modelo lineal a tramos suavizado [Lázaro01b], se utilizará, con fines comparativos, el modelo SPWL$(4, P, 2)$, cuya expresión de mapeo viene dada por

$$\mathbf{y} = \mathbf{a} + \mathbf{B}\mathbf{x}_n + \sum_{i=1}^{P} \mathbf{c}_i \operatorname{lch}\left(\langle \boldsymbol{\nu}_i, \mathbf{x}_n \rangle - \beta_i, \alpha_i\right),$$

donde los parámetros que intervienen en el modelo (y el entrenamiento del mismo) son equivalentes a los de (5.21).[7]

5.3.4.3. Estimación de la memoria del sistema

Antes de iniciar el modelado de un amplificador de potencia, es necesario precisar cúal es la memoria que posee el dispositivo, y, por lo tanto, cúal es el máximo retardo de tiempo que deben considerar los modelos no lineales, pues éste es un factor que, como ya se comentó, desempeña un papel crucial en el rendimiento que se puede obtener. Suponiendo que el número de portadoras es suficientemente elevado, la señal OFDM puede modelarse como un proceso estocástico Gaussiano complejo cuyas componentes en fase y cuadratura son independientes. Por esta razón, se considerará el problema más simple de estimar el valor óptimo de retardo, d, correspondiente a la componente en fase, $y_I[n] = f(x_I[n], \ldots, x_I[n-d])$; pudiéndose extender los resultados obtenidos al modelo global dado por (5.23). Para resolver esta cuestión se recurre a un criterio de Teoría de Información. En concreto, se estimará la información mutua entre la serie temporal de salida, $y_I[n]$, y una versión retardada de la serie temporal de entrada, $x_I[n-k]$: un valor próximo a cero de la información mutua indica que no existe relación estadística (ni lineal ni no lineal) entre ambas series temporales. Este criterio ha sido aplicado en diversos trabajos con el fin de estimar la dimensión de sistemas dinámicos a partir de señales temporales experimentales [Fraser86, Pineda93].

[7] Se puede obtener una descripción más detallada de la metodología de entrenamiento en [Lázaro01b, Lázaro01a, Chua86].

Dadas dos variables aleatorias, $\mathbf{X}$ e $\mathbf{Y}$, su información mutua puede ser estimada usando la divergencia de Kullback-Leibler (KL) entre la función densidad de probabilidad (fdp) conjunta y las marginales factorizadas, es decir,

$$I_{\mathrm{KL}}(\mathbf{YX}) = \iint f_{\mathbf{YX}}(y,x) \log \frac{f_{\mathbf{YX}}(y,x)}{f_{\mathbf{Y}}(y) f_{\mathbf{X}}(x)} dydx. \tag{5.24}$$

La información mutua es una medida apropiada de la dependencia entre dos variables aleatorias. Siempre es no negativa y su valor es cero si y sólo si las variables son estadísticamente independientes. El problema radica en la dificultad de su estimación a partir de datos experimentales, por lo que, en esta Tesis, se utiliza una medida alternativa dada por el criterio

$$I_{\mathrm{QMI}}(\mathbf{YX}) = \log \frac{\left(\iint f^2_{\mathbf{YX}}(y,x)dydx\right)\left(\iint f^2_{\mathbf{Y}}(y) f^2_{\mathbf{X}}(x)dydx\right)}{\left(\iint f_{\mathbf{YX}}(y,x) f_{\mathbf{Y}}(y) f_{\mathbf{X}}(x)dydx\right)^2}, \tag{5.25}$$

conocida como información mutua cuadrática (QMI: *Quadratic Mutual Information*) y propuesta en [Principe99, Principe00]. Esta medida puede interpretarse como un coeficiente de correlación generalizado que estima el ángulo entre la fdp conjunta y el producto de las marginales. Si se estima la fdp conjunta y ambas marginales utilizando el método de enventanado de Parzen con funciones núcleo Gaussianas, entonces el QMI puede ser evaluado de forma muy simple, lo que supone la principal ventaja de (5.25) frente a (5.24).

En definitiva, el objetivo consiste en cuantificar la cantidad de información "nueva" (innovación) que proporciona $x_I[n-k]$ acerca de $y_I[n]$; sin embargo, con anterioridad a la estimación de la información mutua entre $x_I[n-k]$ e $y_I[n]$, es necesario sustraer, de alguna manera, la información que ya proporcionan las entradas previas $x_I[n], \ldots, x_I[n-k+1]$. En concreto, el procedimiento consiste en evaluar la información mutua entre la serie temporal retardada $x_I[n-k]$ y el residuo resultante tras llevar a cabo la predicción lineal $e[n] = y_I[n] - \sum_{l=0}^{k-1} a_l x_I[n-l]$ para $k \geq 1$. De esta forma, se elimina toda dependencia estadística lineal entre $y_I[n]$ y las entradas previas $x_I[n], \ldots, x_I[n-k+1]$.

5.3.4.4. Resultados

La Figura 5.13 muestra los resultados obtenidos al aplicar el criterio QMI para señales OFDM-QPSK de potencia $P_i = 9$ dBm y diferentes anchos de banda. Para $k \geq 2$ la información mutua entre $y_I[n]$ y $x_I[n-k]$ es prácticamente nula, por lo que ambas series temporales pueden considerarse estadísticamente independientes. La conclusión de este análisis es que toda la información sobre $y[n]$ puede ser extraída de $x[n]$ y $x[n-1]$, es decir, la memoria del modelo no lineal debe ser de una muestra.

Teniendo en cuenta la memoria necesaria para modelar el amplificador y las prestaciones proporcionadas por los distintos modelos no lineales, a partir de la

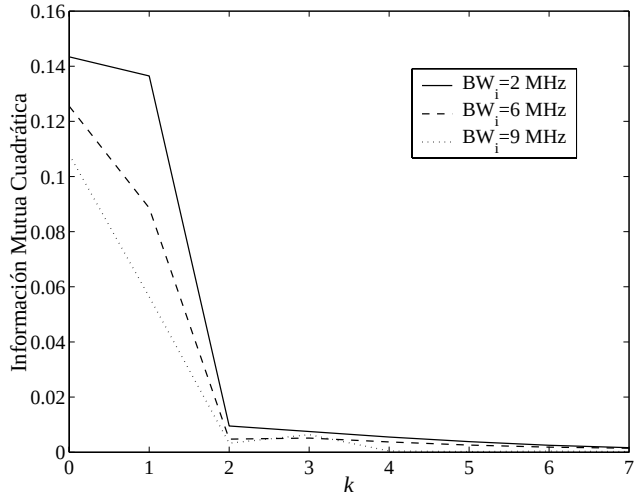


Figura 5.13: Información mutua cuadrática entre $y_I[n]$ y $x_I[n-k]$ en función del retardo k para una señal de entrada OFDM-QPSK con $P_i = 9$ dBm y diferentes anchos de banda.

envolvente compleja de entrada y de salida del amplificador obtenidas mediante submuestreo, se entrenarán y evaluarán los siguientes modelos no lineales:

- Saleh de 4 parámetros;
- EPM(1,3) de grado 3 y memoria 1, es decir, de 12 parámetros;
- Volterra(3,1) con una muestra de memoria y en el que únicamente se consideran los términos lineales y de tercer orden, lo que supone 20 parámetros;
- MLP(4,10,2), es decir, con 10 neuronas en la capa oculta, resultando 72 parámetros;
- SPWL(4,10,2) con 10 fronteras y 71 parámetros;
- αPWL(4,10,2) con 10 fronteras y 71 parámetros.

Para llevar a cabo un estudio comparativo detallado de dichos modelos no lineales (incluido el modelo αPWL) se ha considerado una señal de excitación OFDM-BPSK. En concreto, los conjuntos de datos de entrenamiento y verificación obtenidos mediante técnicas de submuestreo (siguiendo el esquema de la Figura 5.5 y el montaje de la Figura 5.6) están constituidos por un registro de 9000 muestras de las señales envolvente compleja de entrada y de salida

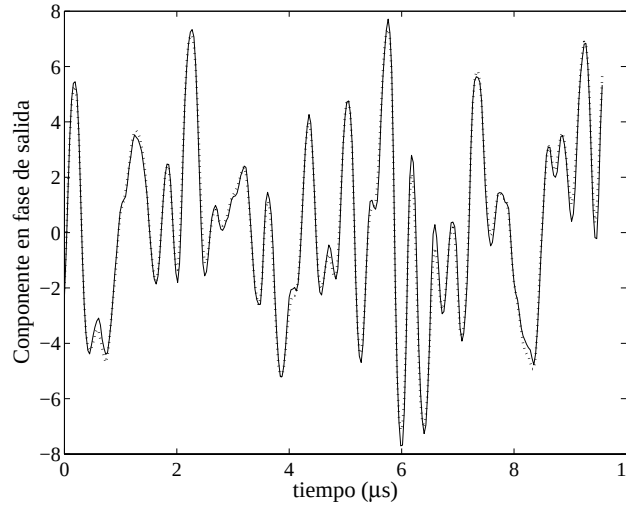


Figura 5.14: Señal de salida medida (línea de puntos) y estimada por el modelo αPWL para una señal de entrada OFDM-QPSK de $BW_i = 6$ MHz y $P_i = 3$ dBm.

para cada potencia de entrada y ancho de banda considerados, reservándose $L = 3000$ muestras para el entrenamiento y 6000 muestras para verificación de las prestaciones (test). Se han llevado a cabo adquisiciones para ocho anchos de banda distintos ($BW_i = 3$, 4, 5, 6, 7, 8, 9 y 10 MHz) y para cuatro potencias de entrada diferentes ($P_i = 0$, 3, 6 y 9 dBm). Por tanto, el conjunto global de medidas de entrenamiento y test para cada potencia concreta está formado, respectivamente, por 24000 y 48000 muestras complejas. El objetivo consistirá en obtener un modelo diferente para cada potencia de entrada e independiente del ancho de banda de la señal.

Con el fin de proporcionar una idea cualitativa de la capacidad del modelo αPWL a la hora de modelar el amplificador de potencia bajo estudio, la Figura 5.14 presenta la señal de salida medida frente a la señal de salida estimada por el modelo (únicamente se representa la señal en cuadratura) para una señal de entrada OFDM-QPSK de 6 MHz de ancho de banda y 3 dBm de potencia (lo que representa un comportamiento no lineal moderado). Se aprecia en la figura una buena concordancia entre las señales medidas y las estimadas, aunque se debe mencionar que la misma se degrada para anchos de banda y potencias superiores, tal y como se mostrará, con más detalle, posteriormente.

Una figura de mérito, que proporciona una medida cuantitativa del comportamiento del modelo en la banda de interés, es la relación señal a error (SER),

definida como

$$\mathrm{SER} = 10\log\left(\frac{\sum_n |y[n]|^2}{\sum_n |\hat{y}[n] - y[n]|^2}\right),$$

donde $\hat{y}[n]$ es la salida del modelo e $y[n]$ es la verdadera salida del amplificador de potencia.

Otro aspecto que debe ser tenido en cuenta es el comportamiento del modelo fuera de banda, para lo cual se recurre a la relación de potencia del canal adyacente (ACPR: *Adjacent Channel Power Ratio*), definida como el cociente entre la potencia de salida en la banda de la señal de entrada y la potencia de salida en los canales adyacentes superior o inferior. En esta Tesis se utilizará la media de la potencia de ambos canales laterales, es decir,

$$\mathrm{ACPR}(S(f)) = 10\log\left(\frac{2\int_{B/2}^{B/2} S(f)df}{\int_{3B/2}^{-B/2} S(f)df + \int_{B/2}^{3B/2} S(f)df}\right),$$

donde B representa el ancho de banda de la señal de entrada y $S(f)$ la densidad espectral de potencia de la señal de salida. Con el fin de evaluar la capacidad de los modelos no lineales para reproducir el ACPR del amplificador, se establece la relación

$$\Delta\mathrm{ACPR} = \mathrm{ACPR}(\hat{S}(f)) - \mathrm{ACPR}(S(f)),$$

siendo $\hat{S}(f)$ la densidad de potencia de la salida estimada por el modelo y $S(f)$ la verdadera densidad espectral de potencia de la salida del amplificador, calculada a partir de los registros adquiridos.

La Figura 5.15 compara la SER proporcionada por los seis modelos no lineales estudiados para una potencia de entrada de $P_i = 0$ dBm (lo que sitúa al amplificador en una zona ligeramente no lineal) y para $P_i = 9$ dBm (comportamiento extremadamente no lineal, en plena región de saturación). Se puede apreciar que el rendimiento de todos los modelos disminuye a medida que el ancho de banda de la señal de entrada crece. Lógicamente, si el número de parámetros de los modelos se mantiene fijo, su capacidad para ajustar señales de mayor ancho de banda se verá reducida. A esta circunstancia hay que añadir el hecho de que, a mayor ancho de banda de entrada, se produce mayor distorsión en las señales capturadas debido al solapamiento espectral.

Por otro lado, los modelos de Saleh y EPM (con 4 y 12 parámetros, respectivamente) proporcionan, obviamente, peores resultados que los modelos MLP, Volterra, SPWL y αPWL, pues estos últimos poseen un mayor número de parámetros (79, 20, 71 y 71, respectivamente). A la vista de los resultados, se puede concluir que, cuando el amplificador se encuentra alejado del punto de saturación, los modelos de Volterra, SPWL y αPWL proporcionan los mejores resultados; mientras que, cuando la potencia de entrada se incrementa hasta valores próximos a la potencia de saturación, entonces los tres modelos

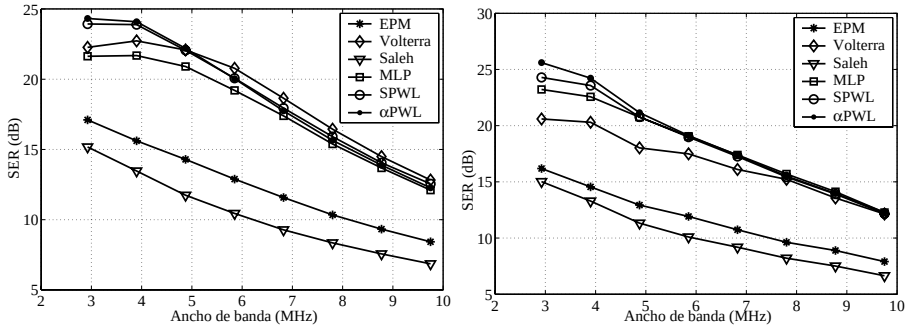


Figura 5.15: SER para una potencia de entrada de $P_i = 0$ dBm (izquierda) y $P_i = 9$ dBm (derecha).

con estructura de red neuronal (MLP, SPWL y αPWL) proporcionan la mejor aproximación de la señal de salida. Cabe destacar que el rendimiento del modelo de Volterra se degrada cuando se consideran anchos de banda pequeños, especialmente para potencias altas, lo cual se debe al hecho de que la salida de un modelo polinómico (como el Volterra o el EPM) diverge cuando el nivel de la señal de entrada se hace muy grande. Se trata de la conocida dificultad de las funciones polinómicas para extrapolar, lo que limita la aplicación de este tipo de modelos para modelar de forma precisa efectos de fuerte saturación.

Para evaluar el rendimiento fuera de banda, es decir, la capacidad para modelar el recrecimiento espectral, la Figura 5.16 muestra el ΔACPR obtenido con cada modelo para $P_i = 0$ y 9 dBm. En dichas gráficas un valor de ΔACPR $= 0$ dB significa un ajuste perfecto entre la potencia fuera de banda generada por el amplificador y la estimada por el modelo. Conviene resaltar varios aspectos: en primer lugar, a pesar del número de parámetros relativamente alto y del buen comportamiento en banda, los modelos SPWL y αPWL tienden a sobreestimar el ACPR, principalmente para anchos de banda grandes, lo que significa que el recrecimiento espectral producido por el modelo no lineal es superior al real. De forma similar, este comportamiento se aprecia en el modelo de Saleh. Si consideramos los resultados relativos al ACPR para todas las potencias estudiadas, el MLP es el modelo que proporciona resultados más precisos.

Finalmente, considerando tanto el rendimiento en términos de SER como en ACPR, el modelo de Volterra, con únicamente 20 parámetros, proporciona un buen compromiso entre prestaciones y complejidad, al menos para no linealidades no excesivamente severas. Por otro lado, los modelos SPWL y αPWL proporcionan los mejores resultados (y prácticamente idénticos entre sí) en términos de SER para cualquier rango de frecuencias y potencias considerados,

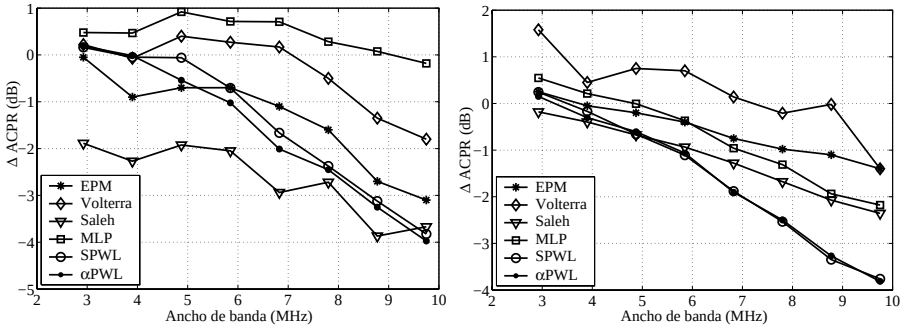


Figura 5.16: ΔACPR para una potencia de entrada de $P_i = 0$ dBm (izquierda) y $P_i = 9$ dBm (derecha).

aunque a costa de un número relativamente alto de parámetros. Entre estos dos últimos, la balanza se decanta en favor del modelo αPWL debido a que requiere unos tiempos de entrenamiento y evaluación un 45 % inferiores a los del modelo SPWL, fruto de la mayor simplicidad de las funciones base que utiliza.

5.4. Modelo gran señal de un transistor HEMT

En esta sección, se propone utilizar el modelo α-spline con el fin de aproximar la característica (no lineal) I/V de un transistor HEMT (*High Electron Mobility Transistor*) de microondas, incluyendo el efecto que sobre el mismo tiene la aplicación de cierta potencia lumínica.

En los últimos años se ha incrementado considerablemente el interés por desarrollar sistemas optoelectrónicos de microondas, motivado por las ventajas intrínsecas que poseen las comunicaciones ópticas, por el creciente uso de frecuencias de microondas en los distintos sistemas de comunicaciones y por la posibilidad que brinda la tecnología para integrar componentes ópticos y de microondas en un único substrato de arseniuro de galio (GaAs). Los circuitos ópticos son ventajosos debido a que pueden integrarse en los circuitos de microondas sin interferir con éstos, y a que poseen bajas pérdidas, pequeñas dimensiones, tiempos de respuesta bajos y gran ancho de banda.

La iluminación directa de un circuito monolítico de microondas (MMIC: *Microwave Monolithic Integrated Circuit*) constituye una técnica muy atractiva dada la versatilidad que proporcionan las aplicaciones asociadas con los sistemas de comunicaciones y control basados en fibra óptica. El transistor de efecto de campo (FET: *Field Effect Transistor*) de GaAs constituye el elemento básico de los MMIC's y puede ser utilizado a modo de fotodetector integrado en

el propio chip monolítico, lo que constituye un auténtico puerto óptico. Por lo tanto, resulta de interés conocer la interacción entre la óptica y las microondas en un FET y cómo varían los parámetros y características del transistor bajo iluminación.

Es bien conocido el hecho de que cuando se ilumina un dispositivo GaAs se produce un efecto de absorción en el espacio comprendido entre Puerta-Drenador y Puerta-Fuente y una fotoexcitación en el área activa. De hecho, estos dispositivos exhiben efectos tanto fotoconductivos como fotovoltaicos, que pueden ser convenientemente amplificados por medio de resistencias externas. Esto significa que, tanto las curvas estáticas de continua como los parámetros del circuito equivalente de pequeña señal, varían cuando se aplica energía óptica al dispositivo [Calandra94, Madjar92].

Sin embargo, el comportamiento real en gran señal está gobernado por la característica dinámica pulsada I/V, que depende del punto de polarización en reposo. Hasta la aplicación en [Lázaro01a] del modelo lineal a tramos suavizado, el único trabajo relativo a los efectos de la iluminación óptica sobre el comportamiento dinámico (responsable de la eficiencia y de la potencia de salida) de este tipo de dispositivos se basaba en la aproximación mediante funciones analíticas [Navarro98]. Esta métodología se fundamenta en una investigación intensiva del comportamiento dinámico en gran señal del transistor, estudiando las distintas dependencias y ajustando las mismas por medio de un conjunto adecuado de funciones analíticas. Por ejemplo, en el caso particular de un HEMT, se ha identificado una dependencia logarítmica con la potencia óptica y una relación tangente hiperbólica con la tensión de puerta. Aunque este método proporciona resultados aceptables, es claro que los modelos desarrollados son válidos únicamente para cada dispositivo concreto. Si se necesita modelar un nuevo dispositivo es preciso llevar a cabo un detallado estudio del mismo y buscar las funciones analíticas que mejor se ajusten a los datos experimentales. El proceso resultante es enormemente laborioso, justificando la búsqueda de herramientas automáticas que proporcionen modelos igual de precisos a partir de las medidas experimentales.

5.4.1. Resultados del modelo αPWL

Se han utilizado las medidas experimentales de un HEMT de microondas Phillips D02AH(4×30 μm) con una longitud de puerta de 0,2 μm para entrenar un modelo αPWL que ajuste la dependencia no lineal de la corriente de drenador (I_{ds}) con respecto a las tensiones de polarización (V_{gso}, V_{dso}), a las tensiones instantáneas (v_{gs}, v_{ds}), y a la potencia óptica incidente (P_o). Se trata, por tanto, de un problema de aproximación funcional en el que el espacio de entrada es de dimensión $M = 5$ y el de salida de dimensión $N = 1$, es decir, ($f : \mathbb{R}^5 \rightarrow \mathbb{R}$). El comportamiento del transistor ha sido medido en los siguientes puntos de polarización: $V_{dso} = -0{,}75$, 0 y 0,75 V; $V_{gso} = 0$, 2 y

4 V. Respecto a las tensiones instantáneas, v_{ds} varía entre 0 y 4 V con pasos de 0,25 V; mientras que v_{gs} varía entre -1 y 0,75 V con pasos de 0,25 V. Por último, todas las medidas se repiten para distintas potencias ópticas: $P_o = 0$, 0,25, 0,5, 0,75, 1, 2, 5 y 10 mW. Por lo tanto, se dispone de un conjunto de 9792 muestras de entrada-salida, que se emplean para entrenar tanto el modelo αPWL como otros modelos utilizados con fines comparativos.

Concretamente, se han ajustado y comparado seis modelos distintos que llevan a cabo un mapeo de $M = 5$ a $N = 1$: un MLP$(5, P, 1)$, una red GRBF$(5, P, 1)$ propuesta en [Santamaría99], el modelo analítico de [Navarro98], un modelo convencional CPWL$(5, P, 1)$, el modelo SPWL$(5, P, 1)$ propuesto en [Lázaro01b] y el modelo αPWL$(5, P, 1)$ presentado en esta Tesis. Con el fin de evaluar la precisión de cada uno de estos métodos se utiliza la relación señal a error (SER: *Signal-to-Error Ratio*) de la I_{ds} estimada para distintos valores del parámetro P (que representa el número de neuronas en la capa oculta para los modelos MLP y GRBF, y el número de fronteras en los modelos lineales a tramos).

El entrenamiento de los modelos GRBF, CPWL, SPWL y αPWL se realiza utilizando un subconjunto de $L = 1000$ muestras elegidas aleatoriamente entre todo el conjunto de medidas. Los modelos se validan utilizando el resto de muestras y se evalúa la SER considerando la totalidad de las muestras disponibles. Al utilizar un conjunto de datos de entrenamiento reducido se consigue mantener la complejidad de cómputo en unos niveles aceptables sin comprometer la capacidad de generalización de los distintos modelos. Para el entrenamiento del MLP se obtienen mejores resultados utilizando un subconjunto de entrenamiento de 2000 muestras. Dicho MLP se entrena empleando el algoritmo convencional de retropropagación con un parámetro de aprendizaje adaptativo, mientras que la red GRBF se entrena añadiendo un centroide en cada iteración y reentrenando la totalidad de la red siguiendo el algoritmo descrito en [Santamaría99].

Los resultados proporcionados por los distintos modelos se muestran en la Tabla 5.1, en la que aparecen los valores de SER correspondientes a la estima de I_{ds} y de su derivada con respecto a los voltajes instantáneos (v_{ds} y v_{gs}). En este ejemplo, los valores de las derivadas no han sido medidos experimentalmente, sino que se ha ajustado un modelo spline cúbico a todo el conjunto de datos (resultando un modelo impráctico dado su elevado número de parámetros) y se han obtenido las derivadas a partir de dicho modelo. Así mismo, tanto en el modelo SPWL como en el modelo αPWL se ha adaptado su parámetro de suavidad gracias a la información de la primera derivada.

Se puede observar en la Tabla 5.1 que, dado un número de parámetros, los modelos SPWL y αPWL proporcionan los mejores resultados, tanto en la función como en sus derivadas. Como era de esperar, ambos modelos poseen una capacidad de aproximación muy similar dado el gran parecido existente

Modelo	Número de parámetros	I_{ds} SER(dB)	$\frac{dI_{ds}}{dV_{ds}}$ SER(dB)	$\frac{dI_{ds}}{dV_{gs}}$ SER(dB)
MLP(5,5,1)	36	25,50	14,00	14,40
CPWL(5,5,1)	36	30,62	9,52	13,04
SPWL(5,5,1)	37	31,41	18,94	16,28
αPWL(5,5,1)	37	30,32	15,76	16,41
MLP(5,9,1)	64	27,17	12,62	14,75
CPWL(5,10,1)	66	32,44	10,44	12,96
SPWL(5,10,1)	67	33,14	18,69	17,67
αPWL(5,10,1)	67	32,02	18,20	17,53
MLP(5,11,1)	78	28,09	13,71	14,66
CPWL(5,12,1)	78	33,04	11,51	14,34
SPWL(5,12,1)	79	33,71	18,87	17,42
αPWL(5,12,1)	79	32,66	17,44	18,19
GRBF(5,8,1)	88	28,84	10,09	13,79
Analítico	98	32,92	−0,8674	16,53

Tabla 5.1: Comparativa de los resultados de modelar el HEMT Phillips D02AH (4×30 μm).

entre las funciones de activación que utilizan (ver Figura 5.2). La Figura 5.17 muestra la función original y sus derivadas, además de la aproximación dada por el modelo αPWL. En concreto se representa la corriente de drenador, I_{ds}, para distintos valores de las tensiones instantáneas de puerta y drenador, v_{gs} y v_{ds}, considerando una potencia óptica incidente de $P_o = 1$ mW y unas tensiones de polarización $V_{gs0} = 0$ V y $V_{ds0} = 2$ V.

Otro aspecto de gran importancia es el coste computacional requerido para estimar los parámetros de los distintos modelos. En la Tabla 5.2 se presentan los tiempos de entrenamiento de los distintos modelos programados en MATLAB sobre un ordenador personal Pentium IV 2,8 GHz con 1 GByte de memoria RAM. Se aprecia que el tiempo de entrenamiento es significativamente menor para los modelos CPWL, SPWL y αPWL que para el resto de modelos: mientras que los primeros requieren no más de 4 minutos para llevar a cabo el ajuste, las redes MLP y GRBF pueden precisar tiempos superiores a una hora. Esto representa una gran ventaja para los modelos lineales a tramos. Es interesante resaltar que, a pesar de poseer una capacidad de aproximación prácticamente idéntica, el tiempo de cómputo del modelo αPWL es siempre inferior al del modelo SPWL propuesto en [Lázaro01b] (lo que supone un ahorro computacional en torno al 20 %), tanto para entrenar el modelo como para evaluar la salida dadas las entradas. Esto se debe a que las funciones de activación del modelo

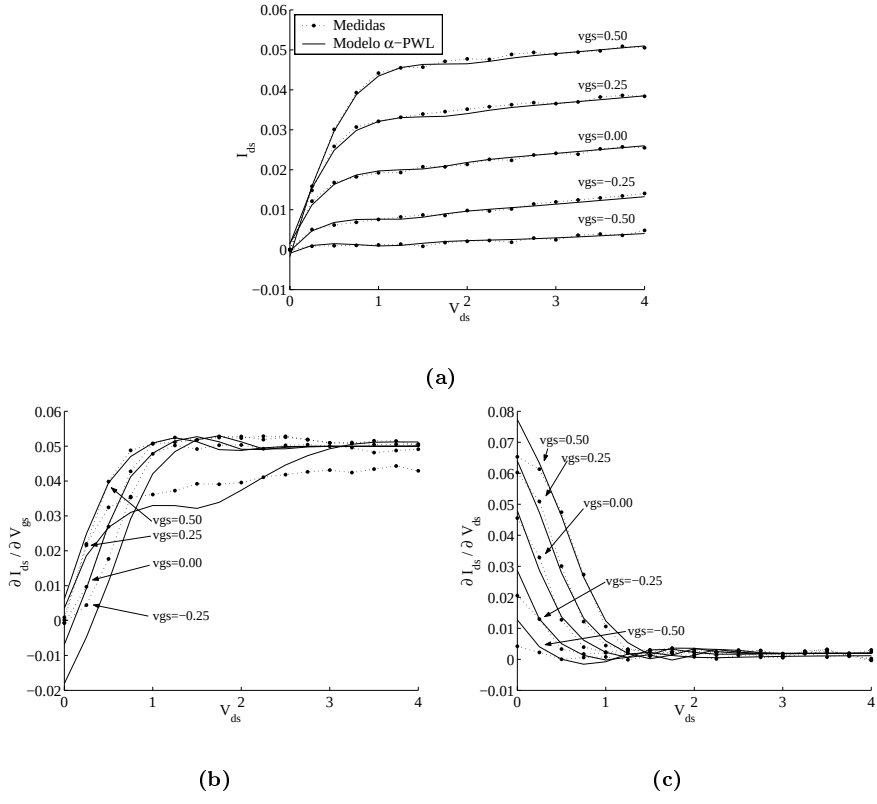


Figura 5.17: Curvas experimentales y generadas por el modelo αPWL para el transistor HEMT (polarización: $V_{gs0} = 0$ V, $V_{ds0} = 2$ V y $P_o = 1$ mW): a) corriente de drenador, I_{ds}, b) derivada con respecto a v_{gs}, c) derivada con respecto a v_{ds}.

αPWL (y sus derivadas) son funciones polinómicas de tercer grado (5.8), por lo que su evaluación es sensiblemente menos costosa que en el caso del modelo SPWL, en el que intervienen funciones de activación del tipo logaritmo de coseno hiperbólico (5.7).

5.5. Conclusiones

Las resultados obtenidos en el Capítulo 4 relativos a la interpolación mediante la base α-spline, justifican su empleo en el campo del modelado. Por ello, en este capítulo, se ha propuesto su uso como función núcleo del modelo canónico lineal a tramos, obteniéndose un modelo simple, con gran capacidad

Modelo	Tiempo de entrenamiento (minutos)
MLP(5,5,1)	43
CPWL(5,5,1)	1,5
SPWL(5,5,1)	2,4
αPWL(5,5,1)	2,2
MLP(5,9,1)	90
CPWL(5,10,1)	2,3
SPWL(5,10,1)	3,6
αPWL(5,10,1)	3,1
MLP(5,11,1)	110
CPWL(5,12,1)	2,8
SPWL(5,12,1)	4,3
αPWL(5,12,1)	3,6
GRBF(5,8,1)	30

Tabla 5.2: Coste computacional del entrenamiento de los distintos modelos del HEMT Phillips D02AH (4×30 μm).

de aproximación, derivable y tan suave como sea preciso, sin más que ajustar el parámetro de suavidad, α. Así, se ha propuesto el modelo αPWL, cuya suavidad se ha demostrado que posee una dependencia monótona respecto a α. También se ha presentado el algoritmo de entrenamiento del modelo para el caso general de M variables de entrada y N variables de salida, incluyendo la posibilidad de ajuste del parámetro de suavidad para minimizar el error del modelo o para ajustar ciertas derivadas de la función a modelar.

Existen dos ventajas del modelo αPWL frente a otros modelos que suavizan el modelo CPWL. En primer lugar, para una curvatura máxima dada, el núcleo α-spline no genera sobredisparos y es el que proporciona una mayor suavidad. Ésta es una característica de gran importancia en un modelo parsimonioso. En segundo lugar, el modelo αPWL es computacionalmente más eficiente, por la naturaleza polinómica de las funciones base (y sus derivadas) que intervienen.

Como aplicación práctica del modelo αPWL, se ha propuesto su utilización para modelar amplificadores de potencia de radiofrecuencia. Para ello, se hace uso de una técnica de caracterización basada en el submuestreo, lo que posibilita disponer de muestras temporales de la entrada y de la salida del amplificador mediante un esquema de adquisición más simple que los empleados habitualmente. En este ámbito, se ha demostrado la utilidad del modelo αPWL, pues proporciona el mejor ajuste en banda de todos los modelos considerados. Sin embargo, a la hora de modelar el comportamiento fuera de banda (más con-

cretamente, el recrecimiento espectral) el modelo MLP proporciona los mejores resultados, aunque a costa de un gasto computacional mucho mayor.

Respecto al modelado de transistores HEMT bajo iluminación óptica, el modelo αPWL ha proporcionado un ajuste similar que el modelo CPWL con un coste computacional ligeramente superior. Sin embargo, su derivabilidad y la capacidad de ajuste de las derivadas lo convierten en el modelo idóneo, puesto que, además, computacionalmente es menos costoso que el modelo SPWL (que lo iguala en prestaciones).

Capítulo 6

Técnicas Interpolatorias en Comunicaciones

6.1. Introducción

Entre los muchos aspectos referentes a la familia α-spline que se presentaron en el Capítulo 4 se encuentra el α-spline cardinal, que permite una representación de la interpolación α-spline similar a la proporcionada por el Teorema de Muestreo para señales limitadas en banda y, adicionalmente, equivalente a la representación de una señal de comunicaciones en función de los símbolos transmitidos. En este capítulo se aborda el problema de las comunicaciones digitales aprovechando el paralelismo existente entre las mismas y la interpolación funcional; paralelismo que, a pesar de resultar evidente, no ha sido, a nuestro juicio, explotado suficientemente. Es posible considerar el *criterio de Nyquist* de comunicaciones sin interferencia intersimbólica como un caso particular de interpolación funcional en el que la expansión en serie de funciones base se restringe a muestras equiespaciadas y funciones limitadas en banda. En aquellas aplicaciones en las que la restricción en banda sea prescindible, la consideración de las comunicaciones desde el punto de vista de la interpolación abre las puertas a nuevos diseños que hagan uso de las técnicas consideradas con anterioridad en esta Tesis. Con este fin se presenta dicho enfoque en relación a dos aspectos fundamentales de las comunicaciones digitales: la interferencia entre símbolos (ISI: *Intersymbol Interference*) y el filtro adaptado.

En concreto se introducirá una nueva metodología [Ibáñez02] que emplea la interpolación spline y sus técnicas con el fin de diseñar un sistema de comunicaciones digitales libre de ISI y adaptado, resultando una estructura de comunicaciones que posibilita una implementación eficiente, produce un rendimiento óptimo (en términos de probabilidad de error) y muestra una mejor

característica espectral que otros esquemas habituales de codificación de línea, como, por ejemplo, aquellos que hacen uso de conformaciones rectangulares (NRZ: *Non-Return to Zero*). De forma paralela, se presentará la metodología de diseño de sistemas de comunicaciones adaptados y libres de ISI que utilizan como filtro conformador la base α-spline presentada en el Capítulo 4. Posteriormente, se generalizará la metodología propuesta con el fin de permitir su aplicación a cualquier tipo de pulso conformador, no sólo pulsos de tipo spline.

Cuando el receptor de comunicaciones digitales muestrea al periodo de símbolo, el sistema que se obtiene de la metodología propuesta incluye filtros IIR que, en la práctica, deben ser truncados. En la última parte del capítulo se demuestra que, cuando existe sobremuestreo, es posible conseguir una implementación que solamente incluya filtros FIR. Esta técnica [Santamaría04] hace uso de la teoría de parejas de filtros biortogonales (BP: *Biortogonal Partners* [Vaidyanathan01b]) y permite obtener sistemas adaptados y libres de ISI empleando exclusivamente filtros FIR.

Previamente, se realizará una breve introducción a los conceptos de interferencia intersimbólica y filtro adaptado. En la Figura 6.1 se representa un esquema que servirá para ilustrar un sistema de comunicaciones digitales banda base típico en el que aparecen y se denotan las principales componentes y señales que lo forman.

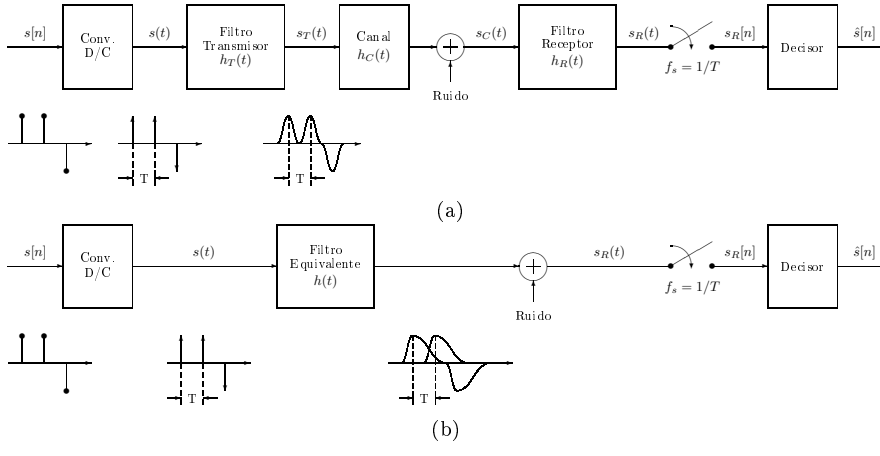


Figura 6.1: Sistema de comunicaciones digitales banda base: a) esquema general, b) modelo equivalente.

6.2. Conceptos básicos

6.2.1. Interferencia entre símbolos

En la Figura 6.1, que representa un modelo equivalente del sistema típico de comunicaciones digitales banda base, se observa cómo el efecto combinado de filtro transmisor, canal limitado en banda y filtro receptor, provoca que los pulsos recibidos se solapen unos con otros, interfiriendo en el proceso de detección y degradando la probabilidad de error. Es la denominada interferencia entre símbolos. Habitualmente, dado el canal $H_C(f)$, el problema radica en determinar los filtros transmisor, $H_T(f)$, y receptor, $H_R(f)$, que minimizan la ISI a la entrada del decisor.

Las condiciones para una transmisión de datos libre de interferencia entre símbolos sobre un canal limitado en banda fueron establecidas por H. Nyquist en [Nyquist28]. En dicho artículo, se demuestra que el ancho de banda mínimo necesario para detectar R_S símbolos por segundo sin ISI es $R_S/2$ Hz. Esta situación se produce cuando la función de transferencia del filtro equivalente del sistema de comunicación, $H(f)$, es rectangular, tal y como muestra la Figura 6.2b (línea continua). En ese caso, el filtro se denomina filtro ideal de Nyquist. En sistemas paso bajo, la respuesta al impulso del sistema es $h(t) = \text{sinc}\,(t/T)$ (línea continua en Figura 6.2a), resultando evidente que, a pesar de la (infinita) longitud de las colas de los pulsos, las pertenecientes a todos los símbolos toman un valor nulo en el instante de muestreo, a excepción hecha, evidentemente, del pulso perteneciente al símbolo que corresponde detectar. Por tanto, bajo estas condiciones y suponiendo sincronización perfecta, no existirá degradación debido a la ISI.

Aunque, en teoría, es posible transmitir sin ISI por un canal de anchura $W = 1/2T$ Hz una tasa de $R_S = 1/T = 2W$ símbolos/s, en la práctica, y dado que el filtro ideal de Nyquist es irrealizable, dicha tasa binaria constituye un límite teórico.

En aplicaciones reales, es preciso admitir un cierto exceso de banda (respecto al mínimo establecido por Nyquist) con el fin de trabajar con filtros implementables. Una función de transferencia, $H(f)$, utilizada muy frecuentemente y que cumple con los criterios anteriormente expuestos, es el filtro conformador en coseno alzado (*raised-cosine filter*), que toma la forma

$$H(f) = \begin{cases} 1, & |f| < 2W_0 - W; \\ \cos^2\left(\frac{\pi}{4}\frac{|f|+W-2W_0}{W-W_0}\right), & 2W_0 - W < |f| < W; \\ 0, & |f| > W; \end{cases} \tag{6.1}$$

donde W es el ancho de banda total y $W_0 = 1/2T = R_S/2$ representa el ancho de banda de Nyquist para el caso del filtro ideal y el ancho de banda a -6 dB para el filtro coseno alzado.

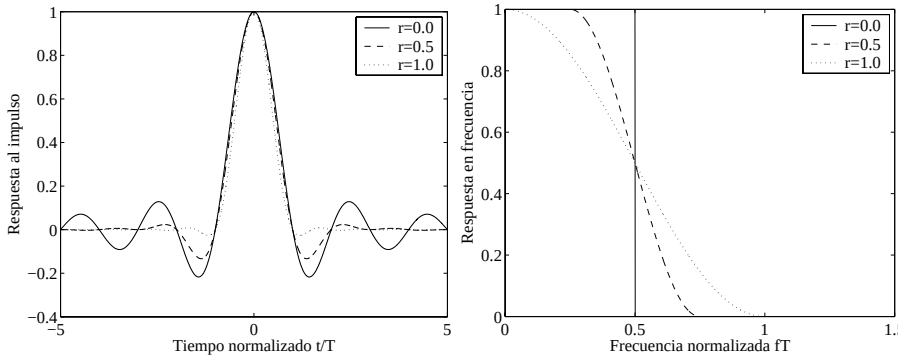


Figura 6.2: Filtro de Nyquist con factor de *roll-off* $r = 0$, 0,5 y 1: a) respuesta al impulso, b) respuesta en frecuencia.

La diferencia $W - W_0$ se denomina exceso de ancho de banda, mientras que el término $r = (W - W_0)/W_0$ es el denominado factor de *roll-off*, siendo $0 \leq r \leq 1$. La respuesta al impulso correspondiente a la función de transferencia (6.1) es

$$h(t) = 2W_0 \operatorname{sinc}(2W_0 t) \frac{\cos[2\pi(W - W_0)t]}{1 - [4(W - W_0)t]^2},$$

que, como se observa en la Figura 6.2a, cumple con el primer criterio de Nyquist, tomando un valor unidad en $t = 0$ y anulándose en $t = n/W_0 = nT$, para todo $n = \pm 1, \pm 2, \ldots$

6.2.2. Filtro adaptado

Los filtros a la entrada de un receptor de comunicaciones tienen por objetivo rechazar componentes espectrales no deseadas (ruido e interferencias), permitiendo a la vez que la señal, contenida en un cierto rango frecuencial, los atraviese con la mínima distorsión. Para ello, dichos filtros, típicamente paso banda, poseen ganancia constante y fase lineal en la banda de interés y un rechazo mínimo determinado en el resto del espectro.

Por contra, el llamado filtro "adaptado" se diseña de manera distinta, pues su función es maximizar a su salida la relación señal a ruido (SNR: *Signal to Noise Ratio*) cuando a su entrada se aplica una señal conocida en presencia de ruido aditivo blanco Gaussiano (AWGN: *Additive White Gaussian Noise*). Los filtros convencionales se aplican a señales definidas únicamente por su ancho de banda, mientras que el filtro adaptado se aplica a señales cuya forma de onda es conocida.

Dado el receptor del esquema de la Figura 6.3 en el que a la señal (conocida) de entrada de duración T, $s_i(t)$, se suma AWGN, $n(t)$, el diseño del

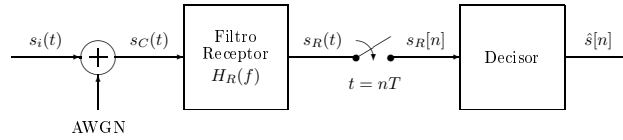


Figura 6.3: Esquema de receptor de comunicaciones digiales banda base.

filtro adaptado es bien conocido [Sklar01]. En el instante $t = nT$, la salida del muestreador, $s_R[n] = s_R(nT)$, estará formada por una componente de señal, $a[n] = a(nT)$, y una componente de ruido, $n_0[n]$, con lo que el filtro adaptado, que deberá maximizar la relación señal a ruido, deberá cumplir

$$H_R(f) = H_0(f) = kS_i^*(f)e^{2\pi fT},$$

o, en el dominio temporal,

$$h_R(t) = h_0(t) = F\left\{kS_i^*(f)e^{2\pi fT}\right\}^{-1}.$$

Puesto que se supone que $s_i(t)$ es una señal real, y haciendo uso de las propiedades de la transformada de Fourier, es posible escribir

$$h_0(t) = \begin{cases} ks_i(T-t), & 0 \leq t \leq T, \\ 0, & \text{en otro caso;} \end{cases} \tag{6.2}$$

es decir, la respuesta al impulso del filtro que maximiza la relación señal a ruido a la entrada del decisor es una versión especular de la señal recibida $s_i(t)$ y retardada el tiempo de símbolo, T, condición esta última necesaria con el fin de hacer (6.2) causal.

Conviene hacer notar dos particularidades de este resultado: por un lado, es válido únicamente para ruido aditivo blanco Gaussiano; y, por otro lado, debe tenerse en cuenta que el filtro adaptado es el espejo de la señal recibida, $s_i(t)$, no de la señal transmitida, $s(t)$, lo que implica tener en cuenta la posible distorsión del canal.

6.3. Comunicaciones digitales vistas como problema de interpolación

El teorema de Nyquist establece las condiciones para una transmisión sin ISI sobre un canal limitado en banda. Sin embargo, en ciertos supuestos, esta limitación no es aplicable; bien porque no sea de interés (por ejemplo, en sistemas UWB: *Ultrawide Band* o en sistemas de espectro ensanchado) o bien

porque, estrictamente, los pulsos conformadores (y las señales en general) de duración finita no están limitados en banda [Slepian76].

Por otro lado, el enunciado del primer criterio de Nyquist en el dominio del tiempo indica que en un sistema de comunicaciones digitales no existirá ISI si la combinación de filtro transmisor, $h_T(t)$, canal, $h_C(t)$, y filtro receptor, $h_R(t)$, posee una respuesta al impulso equivalente, $h(t)$, que se anule en múltiplos del periodo de símbolo, T, es decir,

$$h(nT + t_0) = 0, \quad n = \pm 1, \pm 2, \pm 3, \ldots,$$

siendo t_0 un cierto retardo introducido por los sistemas que intervienen. Evidentemente, este filtro equivalente provoca que la señal recibida, $s_R(t)$, tome en los instantes de muestreo los mismos valores que los símbolos originales, $s[n]$, salvo por el retardo t_0 y una posible constante multiplicativa, que podremos obviar en adelante; es decir,

$$s_R(nT + t_0) = s_R[n] = s[n].$$

Ésta es una expresión equivalente, salvo por el retardo t_0, a la definición de interpolación presentada en (2.1). Este paralelismo, representado en la Figura 6.4, se podría expresar de la siguiente manera: en comunicaciones (Figura 6.4a), el interés se centra en transmitir una serie de símbolos, $s[n]$, de tal manera que en recepción, tras atravesar una serie de sistemas lineales (filtro transmisor, canal y filtro receptor), el valor de la señal recibida, $s_R(t)$, coincida exactamente (salvo por un retardo y una posible constante multiplicativa) con los símbolos originales en los instantes de muestreo. En interpolación (Figura 6.4b), a partir de unos valores conocidos de la función, $f(x_i)$, (los símbolos en comunicaciones) se pretende construir (como expansión en serie de funciones base) una nueva función, $P(x)$, que tome los valores de la función original en los puntos considerados.

Aunque, enunciado de esta manera, el paralelismo entre el problema de interpolación funcional y el diseño de sistemas de comunicaciones resulta evidente, algunas de sus implicaciones no han sido, a nuestro juicio, aprovechadas en el mundo de las comunicaciones.

Se puede decir que el criterio de Nyquist de comunicaciones sin interferencia intersimbólica es un caso particular del problema de interpolación funcional, estableciéndose dicha particularización por medio de la consideración de puntos (muestras, símbolos) equiespaciados y funciones interpolatorias limitadas en banda. Este segundo condicionante, como ya se ha puesto de manifiesto en el Capítulo 2, puede resultar prescindible en ciertas aplicaciones, en las cuales, considerar las comunicaciones desde el punto de vista de la interpolación funcional abre puertas a nuevos diseños.

Bajo esta nueva perspectiva, en las siguientes secciones se abordará el diseño de sistemas de comunicaciones sin ISI y adaptados basados en las técnicas interpolatorias con splines.

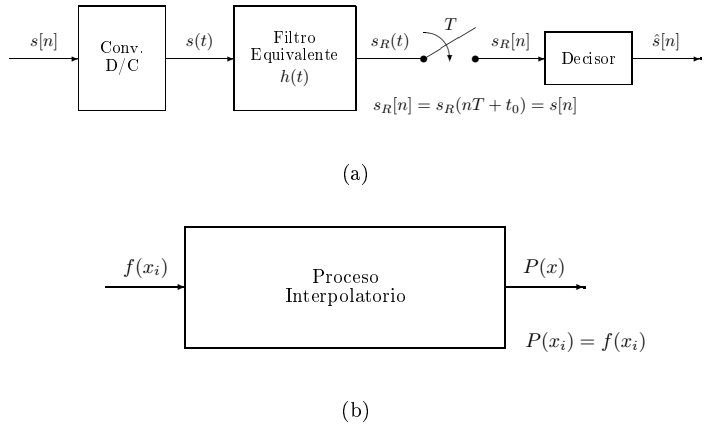


Figura 6.4: Paralelismo interpolación-comunicaciones: a) sistema equivalente de comunicaciones digitales banda base, b) sistema interpolatorio.

6.4. Sistemas de comunicación basados en splines

A pesar del éxito y la enorme difusión del filtro conformador en coseno alzado (6.1), existe una amplia y activa área de investigación relativa al diseño de filtros que satisfagan el criterio de Nyquist y que a la vez cumplan una serie de objetivos, como pueden ser la maximización de la energía en un cierto intervalo [Panayirci99] o la posesión de la propiedad de ausencia de ISI tanto en versiones adaptadas como no adaptadas [Xia97, Kisel00]. Es más, debido al coste computacional asociado al filtrado en coseno alzado, determinadas aplicaciones prácticas se basan en filtros conformadores mucho más simples (como, por ejemplo, pulsos rectangulares o cosenoidales) con características espectrales muy alejadas del coseno alzado, incluso no limitadas en banda.

Siguiendo la idea presentada en la anterior sección, en esta Tesis se propone utilizar splines para interpolar la secuencia discreta correspondiente a los símbolos de una transmisión. De esta manera, se obtiene una señal libre de ISI cuando se muestrea a la velocidad de símbolo. Aunque cualquier técnica interpolatoria (entendida como la generación, a partir de los símbolos, de una señal que pase exactamente por dichos símbolos) genera una señal de comunicaciones libre de ISI, la interpolación basada en splines posee un número de ventajas que la convierten en una opción interesante cuando la complejidad de implementación y el contenido espectral son factores de importancia.

Como se demostrará posteriormente, cualquier interpolación empleando splines de orden impar puede ser llevada a cabo mediante el filtrado de la

secuencia original (en este caso los símbolos) usando dos filtros adaptados. De esta forma resulta sencillo diseñar un receptor óptimo (desde el punto de vista de minimización de la BER). Si comparamos dicho esquema de comunicaciones con una comunicación basada en pulsos rectangulares, se observa una mejor característica espectral a costa de un moderado (y, posiblemente, asumible en la práctica) incremento computacional.

6.4.1. Interpolación spline

Con el fin de facilitar ulteriores desarrollos de este capítulo, se vuelven a presentar aquí algunas de las principales propiedades de la interpolación spline que ya fueron estudiadas en mayor profundidad en el Capítulo 2 y en [Unser93a, Unser93b, Unser99, Boor78]. Como es bien sabido, una base spline de orden m, $\beta_m(t)$, es una función simétrica con forma de campana que puede ser construida mediante la convolución de $m+1$ pulsos rectangulares idénticos del tipo

$$\beta_0(t) = \begin{cases} 1, & |t| < 1/2; \\ 0, & |t| \geq 1/2. \end{cases}$$

La interpolación spline de una determinada secuencia de datos, $s[n]$, consiste en la determinación de un conjunto de coeficientes $c[n]$ de forma que la señal de tiempo continuo

$$s(t) = \sum_{n\in\mathbb{Z}} c[n]\beta_m\left(\frac{t}{T} - n\right) \tag{6.3}$$

pase exactamente por los datos, es decir,

$$s[n] = s(nT), \quad n \in \mathbb{Z}. \tag{6.4}$$

Una de las ventajas de la expansión B-spline (6.3) es que los coeficientes $c[n]$ pueden ser obtenidos de forma sencilla mediante técnicas de filtrado digital. Con el fin de aclarar esta idea, es conveniente introducir el núcleo discreto B-spline $b_m[n]$, que se obtiene muestreando la base spline en los instantes enteros, resultando

$$b_m[n] = \beta_m(n) = \beta_m(t)|_{t=n\in\mathbb{Z}},$$

con lo que (6.4) puede reescribirse en forma de convolución

$$s[n] = \sum_{k\in\mathbb{Z}} c[k]b_m[n-k] = c[n] * b_m[n].$$

De acuerdo con esta expresión, $s[n]$ (los símbolos en nuestra aplicación de comunicaciones) se obtiene filtrando $c[n]$ con $b_m[n]$. Por tanto, y de forma

inversa, para obtener $c[n]$ a partir de $s[n]$, es suficiente con aplicar el filtro que verifique

$$b_m[n] * d_m[n] = \delta[n],$$

o, de otra forma,

$$d_m[n] = (b_m)^{-1}[n] \overset{z}{\longleftrightarrow} D_m(z) = \frac{1}{B_m(z)},$$

donde $d_m[n]$, conocido como filtro directo spline, es un filtro todo polos con ciertas propiedades de simetría que permiten su implementación mediante la conexión en serie de un filtro recursivo causal, $d_m^+[n]$, y otro recursivo anticausal, $d_m^-[n]$. Como se expondrá en el siguiente apartado, esta propiedad resulta crucial a la hora de obtener filtros adaptados y libres de ISI mediante splines.

El proceso completo de interpolación puede expresarse en forma de convolución,

$$s(t) = \sum_{n\in\mathbb{Z}} s[n]\eta_m\left(\frac{t}{T} - n\right),$$

en la cual se usa la secuencia discreta $s[n]$ como coeficientes, y donde $\eta_m(t)$ representa el spline cardinal de grado m, que se obtiene mediante la convolución del filtro directo y del núcleo B-spline

$$\eta_m(t) = \sum_{n\in\mathbb{Z}} d_m[n]\beta_m(t-n).$$

Si, para simplificar la notación posterior, se redefine el filtro conformador libre de ISI como

$$\varphi_{m,T}(t) = \eta_m\left(\frac{t}{T}\right), \tag{6.5}$$

entonces la señal de comunicaciones basada en interpolación spline adopta la forma

$$s(t) = \sum_{n\in\mathbb{Z}} s[n]\varphi_{m,T}(t-nT).$$

La Figura 6.5 muestra la respuesta al impulso y la respuesta en frecuencia de los filtros conformadores spline $\varphi_{m,T}(t)$ de grado 3, 5, 9 y ∞. Conforme m se incrementa, $\varphi_{m,T}(t)$ tiende a la función $\operatorname{sinc}(t/T)$, es decir, al filtro paso bajo ideal, que se corresponde con el filtro ideal de Nyquist. Así mismo, observando los cruces por cero en múltiplos de T de la Figura 6.5a, es evidente que la técnica de interpolación spline propuesta produce señales libres de ISI.

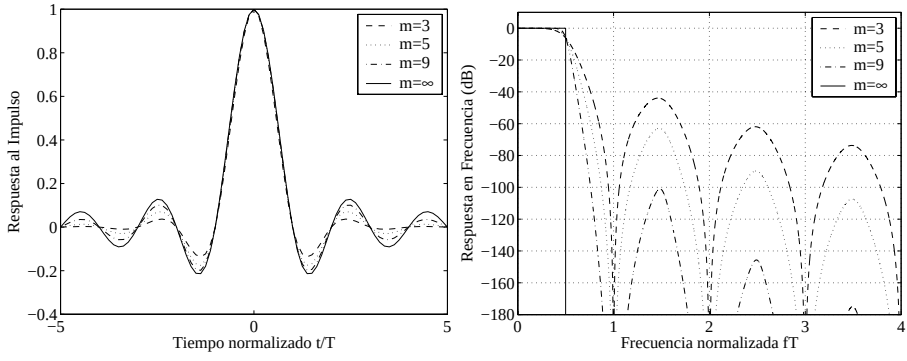


Figura 6.5: Filtro conformador $\varphi_m(t/T)$ de grado $m = 3, 5, 9$ y ∞: a) respuesta al impulso, b) respuesta en frecuencia.

6.4.2. Filtro spline cúbico para comunicaciones libres de ISI y adaptadas

En este apartado se discute cómo factorizar, entre el transmisor y el receptor, el filtro conformador sin ISI presentado en (6.5), de manera que se pueda obtener un esquema de comunicaciones adaptado. En adelante, y con el fin de facilitar la exposición, nos restringiremos a la interpolación mediante splines cúbicos, aunque los resultados obtenidos pueden ser extendidos a splines de cualquier orden impar.

Se proponen dos alternativas para descomponer (6.5) entre el transmisor y el receptor. La primera y más simple se basa en la estructura presentada en la Figura 6.6a, que aprovecha el hecho de que el núcleo B-spline cúbico consiste en la convolución de cuatro pulsos rectangulares o, de forma equivalente, de dos pulsos triangulares. Por lo tanto, es posible construir un sistema de comunicaciones libre de ISI utilizando como filtro transmisor un filtro triangular de duración $[-T, T]$ y como filtro receptor ese mismo filtro triangular combinado con el filtro directo B-spline cúbico, $d_3[n]$. Este esquema de comunicaciones no es óptimo (desde el punto de vista de BER) pues el receptor y el transmisor no están adaptados. Sin embargo, se ha comprobado mediante simulaciones que la degradación respecto a la estructura óptima adaptada es únicamente de 0,6 dB.

La segunda alternativa para la descomposición del filtro conformador busca adaptar transmisor y receptor. El filtro directo B-spline, $d_3[n]$, viene dado por

$$d_3[n] = \frac{1 - z_1}{1 + z_1} {z_1}^{|n|}, \tag{6.6}$$

donde $z_1 = -2 + \sqrt{3}$ es el polo del filtro $|z_1| < 1$. Es sencillo comprobar que

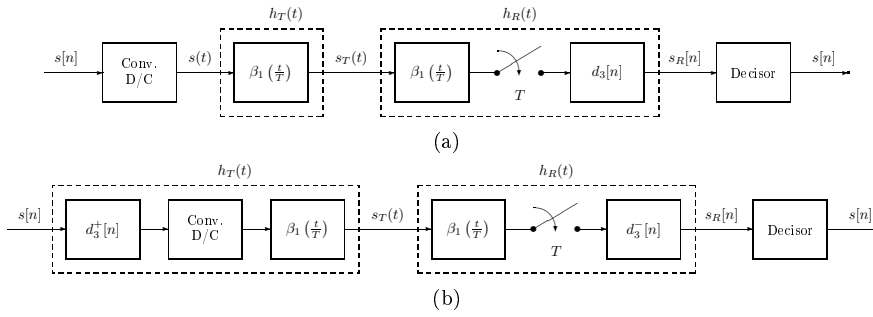


Figura 6.6: Sistema de comunicaciones basado en splines: a) esquema no adaptado, b) esquema adaptado.

este filtro se puede factorizar en dos filtros: uno causal

$$d_3^+[n] = (1 - z_1){z_1}^k u[n],$$

y otro anticausal

$$d_3^-[n] = (1 - z_1){z_1}^{-k} u[-n].$$

Puesto que $d_3^+[n] = d_3^-[-n]$, entonces $D_3^+(\omega) = D_3^-(\omega)^*$ y, por tanto, ambos filtros resultan estar adaptados. Por otro lado, y siguiendo el razonamiento de la anterior alternativa, el B-spline cúbico puede escribirse como la convolución de dos pulsos triangulares. En resumen, es posible descomponer el filtro conformador spline cúbico en un par de filtros adaptados siguiendo el esquema presentado en la Figura 6.6b. Los filtros que intervienen, $h_T(t)$ y $h_R(t)$, pueden denominarse filtros SQRCS (*Square Root Cubic Spline*)

$$h_T(t) = \sum_{n\in\mathbb{Z}} d_3^+[n]\beta^1\left(\frac{t}{T} - n\right),$$

$$h_R(t) = h_T(-t) = \sum_{n\in\mathbb{Z}} d_3^-[n]\beta^1\left(-\frac{t}{T} + n\right).$$

Teniendo en cuenta que la respuesta en frecuencia del filtro transmisor SQRCS es

$$H_T(\Omega) = \mathrm{TF}\left\{h_T(t)\right\} = (1 - z_1)\,\frac{\mathrm{sinc}^2(\Omega T/2\pi)}{1 - z_1 e^{-j\Omega T}},$$

y suponiendo que la secuencia de símbolos, $s[n]$, es blanca con varianza σ_S^2, entonces la densidad espectral de potencia de la señal transmitida, $s_T(t)$, vendrá dada por la expresión

$$S_T(f) = \frac{\sigma_S^2}{T}\left|H_T(f)\right|^2 = \frac{3\sigma_S^2}{T}\left(\frac{\mathrm{sinc}^4(fT)}{2 + \cos(2\pi fT)}\right).$$

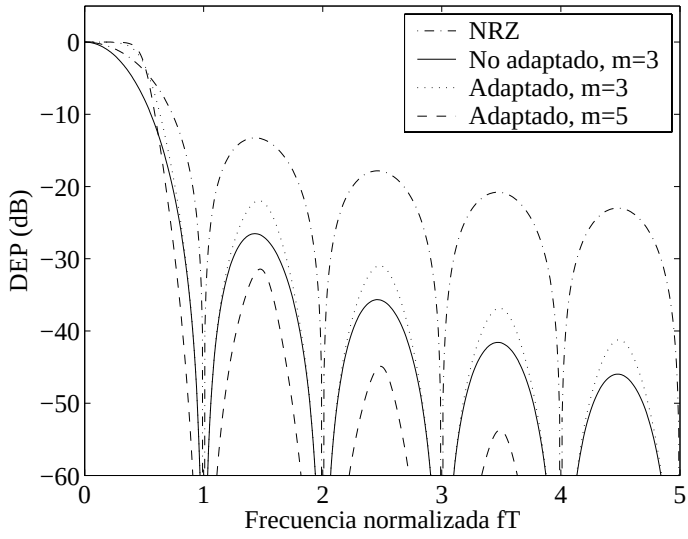


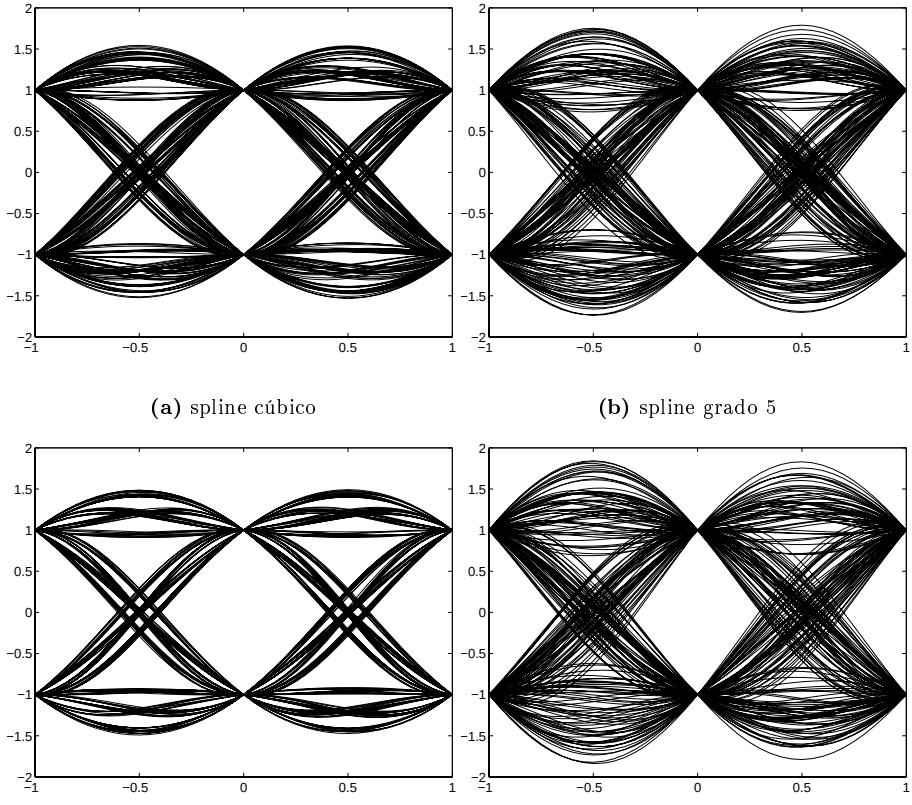
Figura 6.7: Densidad espectral de potencia de los esquemas de modulación basados en splines frente al sistema NRZ.

En la Figura 6.7 se muestra la DEP de $s_T(t)$ para $m = 3$ (spline cúbico) y $m = 5$, junto con la conformación triangular correspondiente a la alternativa no adaptada y el espectro de la conformación NRZ. Se observa que los lóbulos laterales de la conformación con spline cúbico adaptada se encuentran aproximadamente 7 dB por debajo de los correspondientes a la NRZ, mientras que, para una conformación spline de quinto orden, los lóbulos laterales se reducen aún más, situándose 20 dB por debajo de la NRZ. Por otro lado, el espectro de la conformación spline decae como $(fT)^4$, mientras que el NRZ decae de manera proporcional a $(fT)^2$.

Es conveniente recordar que, para ambas alternativas (conformación spline adaptada y no adaptada), la combinación de filtro transmisor y receptor es el spline cardinal (6.5), cuya respuesta en frecuencia (presentada en la Figura 6.5b) es

$$H_3(f) = \frac{3\,\mathrm{sinc}^4(fT)}{2 + \cos(2\pi fT)} = |H_T(f)|^2 .$$

Finalmente, como se muestra en la Figura 6.8a, el diagrama de ojo de la conformación mediante spline cúbico es muy similar al correspondiente a la raíz cuadrada de coseno alzado con factor de *roll-off* de 0,5. La Figura 6.8b muestra el diagrama de ojo tras el filtro receptor para una conformación basada en splines de orden 5.



(a) spline cúbico

(b) spline grado 5

(c) coseno alzado $r = 0{,}5$

(d) coseno alzado $r = 0{,}25$

Figura 6.8: Diagrama de ojo de conformación spline frente a coseno alzado.

6.4.3. Aspectos relativos a la implementación

En este apartado se discuten, brevemente, algunos aspectos relativos a la implementación de sistemas de comunicaciones mediante pulsos conformadores de tipo spline propuestos. A pesar de que la discusión se centrará en la conformación spline cúbica, los resultados son trasladables de forma inmediata a splines de orden (impar) superior. La principal dificultad de la metodología propuesta reside en la implementación de los filtros IIR digitales $d_3^+[n]$ y $d_3^-[n]$. El causal, $d_3^+[n]$, puede ser realizado mediante una única multiplicación-acumulación por cada símbolo de entrada

$$s_{d^+}[n] = s[n] + z_1 s_{d^+}[n-1].$$

Sin embargo, el filtro anticausal debe ser truncado, aunque, dado que sus

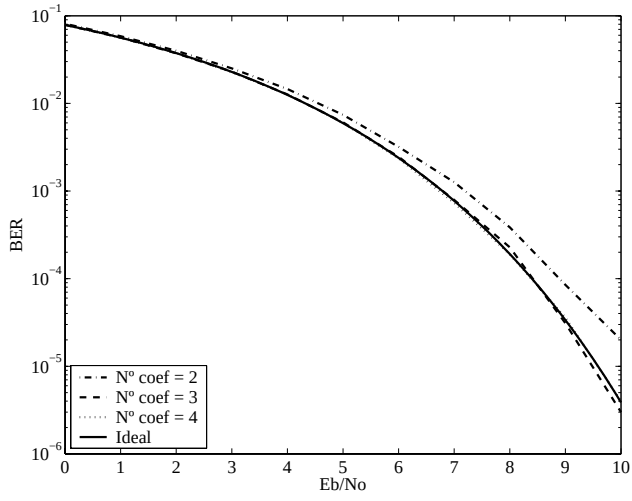


Figura 6.9: Degradación de BER debido al truncamiento de los filtros digitales transmisor y receptor.

coeficientes decrecen muy rápidamente, se ha comprobado mediante simulaciones que con únicamente tres coeficientes, e incluso con dos, se alcanza un rendimiento próximo al óptimo. En ese caso, y con el fin de mantener un sistema de transmisión adaptado, es necesario implementar también el filtro causal en transmisión mediante un filtro FIR truncado con el mismo número de coeficientes, alcanzándose un rendimiento en BER indistinguible del óptimo, tal como muestra la Figura 6.9, en la que aparecen las curvas de probabilidad de error frente a la energía de bit dividido por la densidad espectral de ruido para distintos niveles de truncamiento de los filtros. Dichas curvas se han obtenido al simular la transmisión y recepción de 10^7 bits. En el transmisor, los bits son filtrados utilizando el filtro digital $d_3^+[n]$ truncado a 2, 3 ó 4 coeficientes. Tras un sobremuestreo suficientemente elevado (40 muestras por bit), la señal es conformada por el pulso triangular. A la señal de comunicaciones resultante se le suma ruido aditivo blanco Gaussiano y en el receptor se realiza el filtrado con el pulso triangular, el correspondiente diezmado, el filtrado mediante el filtro $d_3^-[n]$ truncado con el mismo número de coeficientes que en el transmisor y la detección final de los bits.

Conviene señalar que, cuando se realiza el filtro anticausal FIR, no es necesario realizar multiplicaciones en el transmisor si los símbolos de entrada son binarios ($s[n] = \pm 1$), pues únicamente será preciso llevar a cabo operaciones de suma o resta, incrementando la simplicidad del transmisor. En cualquier caso, el filtro FIR del transmisor puede ser fácilmente realizado en forma de tabla

de búsqueda, basado en una memoria de tipo ROM (*Read Only Memory*) que contenga los valores de la señal de salida, $s_{d^+}[n]$, y que sea indexada con los datos de entrada, $s[n]$.

La conformación triangular necesaria en el transmisor puede ser implementada de forma sencilla si todo el filtrado digital se lleva a cabo en el receptor (es decir, con el esquema de comunicación no adaptado, Figura 6.6a). En ese caso, es posible obtener la señal transmitida $s_T(t)$, que no es más que la interpolación lineal a tramos de los símbolos, integrando la señal NRZ previamente derivada. En la práctica es posible utilizar un diferenciador analógico y un integrador basados en amplificadores operacionales, como aparece en la Figura 6.10a. Utilizando este montaje en el transmisor, el conversor digital analógico (DAC: *Digital to Analog Converter*) y su correspondiente filtro reconstructor analógico no son necesarios. En sistemas en los que la señal de comunicaciones sea de tipo multinivel (PAM: *Pulse Amplitude Modulation*), resulta imperativo el uso de un DAC a la frecuencia de símbolo, aunque el filtro reconstructor sigue siendo innecesario. En ese caso, el transmisor debe realizar digitalmente la derivación por medio de un filtro de dos coeficientes (Figura 6.10b).

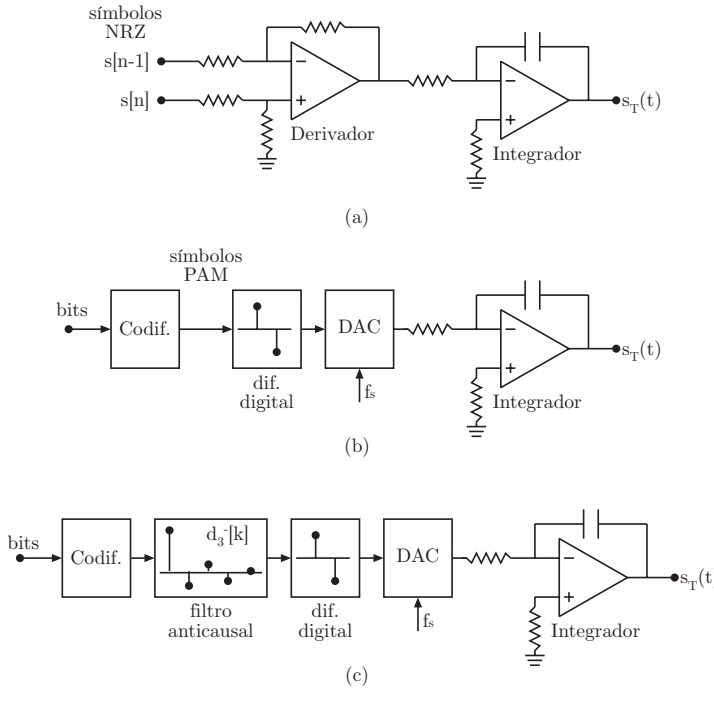


Figura 6.10: Implementación del transmisor spline cúbico: a) binario no adaptado, b) multinivel no adaptado, c) adaptado.

Cuando se desee una alternativa adaptada (Figura 6.6b), el filtro digital del transmisor estará formado por la combinación del derivador digital y el filtro FIR causal (Figura 6.10c), resultando en la práctica un filtro FIR de tan solo tres o cuatro coeficientes debido a su rápida velocidad de decaimiento.

Resumiendo, en el caso de utilizar un esquema de transmisión basado en conformación spline que requiera de un DAC, dicho conversor deberá trabajar a la frecuencia de símbolo, es decir, no sería necesario el factor de sobremuestreo típico de los sistemas habituales, lo que simplifica y abarata su realización. Así mismo, estos montajes no precisan de filtro reconstructor, sino únicamente de un circuito integrador de primer orden. Por otro lado, dada la simplicidad del filtro digital a realizar en el transmisor, éste puede ser incorporado a la tabla de búsqueda del codificador, en cuyo caso no sería necesario realizar operaciones aritméticas.

Respecto al receptor, si se opta por una implementación totalmente digital, será preciso muestrear la señal de entrada a una frecuencia de muestreo múltiplo de la tasa binaria $R_S = 1/T$. Para que el filtro adaptado digital, una versión discretizada del pulso triangular $\beta_1(t/T)$, sea suficientemente preciso, es necesario que el factor de sobremuestreo sea relativamente alto. Posteriormente, la señal debe ser diezmada y más tarde filtrada usando el filtro anticausal a la velocidad de símbolo. Este esquema posee un coste computacional mucho menor que los diseños basados en el coseno alzado. Así mismo, el receptor spline requiere un filtro antisolapamiento mucho menos restrictivo que, por ejemplo, el correspondiente a una transmisión NRZ, dado que sus lóbulos laterales son menores. Si se dispone de un correlador triangular no es necesario sobremuestreo, y el ADC puede trabajar a la velocidad de símbolo.

6.5. Sistemas de comunicación α-spline

En la sección anterior se ha presentado una técnica que permite implementar sistemas de comunicaciones digitales adaptados y libres de ISI a partir de la interpolación spline, gracias a la posibilidad de factorización de la base y del filtro directo spline (de orden impar). Dicha propiedad es compartida por la interpolación α-spline propuesta en el Capítulo 2. En efecto, la base α-spline, $\beta_\alpha(t)$, está constituida por la convolución de dos pulsos rectangulares de duración unitaria y dos pulsos rectangulares de duración $0 \leq \alpha \leq 1$ (4.8); mientras que el filtro directo α-spline, $d_\alpha[n]$, puede factorizarse en dos filtros adaptados (4.30).[1]

Por lo tanto, y como se expone en el esquema de la Figura 6.11, el transmisor de un sistema adaptado y libre de ISI que haga uso de los α-splines estará

[1] De hecho, la expresión del filtro directo α-spline (4.30) es idéntica a la del filtro directo spline cúbico (6.6), salvo por el valor del polo z_1.

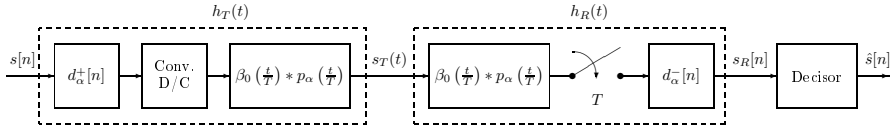


Figura 6.11: Sistema adaptado de comunicaciones basado en α-splines.

formado por un filtro digital, trabajando a la velocidad de símbolo, dado por

$$d_\alpha^+[n] = (1 - z_1){z_1}^n u[n],$$

y por el filtro conformador en tiempo contínuo, $p(t)$, resultado de la convolución

$$p(t) = \beta_0\left(\frac{t}{T}\right) * p_\alpha\left(\frac{t}{T}\right),$$

por lo que la respuesta al impulso del filtro transmisor adopta la expresión

$$h_T(t) = \sum_{n\in\mathbb{Z}} d_\alpha^+[n]p(t - nT).$$

De forma equivalente, y dada la simetría de $p(t)$, la respuesta al impulso del filtro receptor es

$$h_R(t) = \sum_{k\in\mathbb{Z}} d_\alpha^-[n]p(t - NT),$$

siendo

$$d_\alpha^-[n] = (1 - z_1){z_1}^{-n} u[-n],$$

y (4.26)

$$z_1 = \frac{\alpha - 3 + \sqrt{9 - 6\alpha}}{\alpha}.$$

Finalmente, dado que la respuesta frecuencial del filtro transmisor α-spline es

$$H_T(\Omega) = \mathrm{TF}\left\{h_T(t)\right\} = \frac{1 - z_1}{1 - z_1 e^{-j\Omega T}} \operatorname{sinc}\left(\frac{\Omega T}{2\pi}\right) \operatorname{sinc}\left(\frac{\alpha\Omega T}{2\pi}\right),$$

y, considerando una secuencia de símbolos, $s[n]$, blanca con varianza σ_S^2, entonces la densidad espectral de potencia de la señal transmitida es

$$S_T(f) = \frac{3\operatorname{sinc}^2(fT)\operatorname{sinc}^2(\alpha fT)}{3 - \alpha + \alpha\cos(2\pi fT)}.$$

Esta densidad espectral de potencia transmitida se representa en la Figura 6.12 para distintos valores de α. Por otro lado, en la Figura 6.13 se observan los

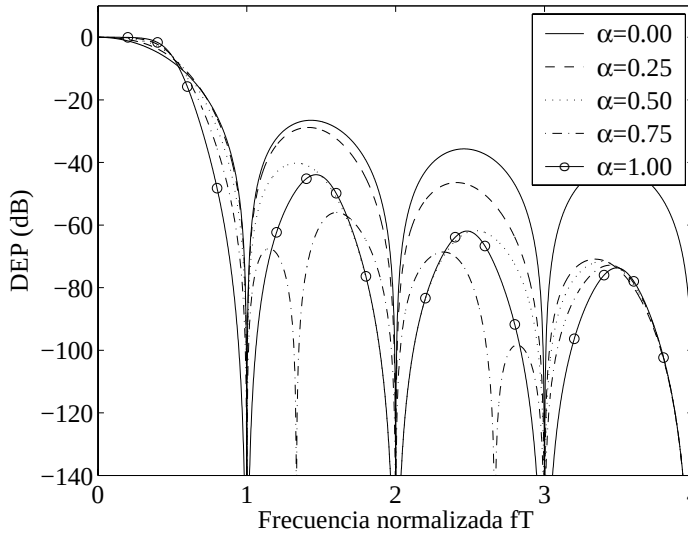


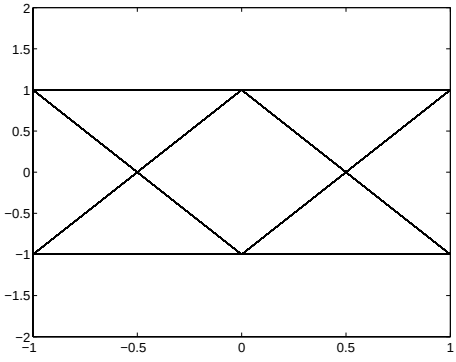
Figura 6.12: Densidad espectral de potencia de la señal transmitida usando conformación α-spline.

diagramas de ojo de una transmisión basada en α-splines para distintos valores del parámetro de suavidad.

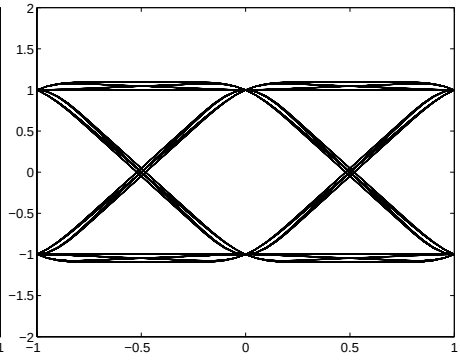
En ambas figuras se observa cómo la conformación α-spline posibilita comunicaciones libres de ISI y adaptadas, permitiendo, además, transicionar suavemente y de forma parametrizada entre un esquema NRZ, cuando $\alpha = 0$, y un esquema spline cúbico, cuando $\alpha = 1$.

6.6. Diseño de sistemas de comunicación adaptados y sin ISI

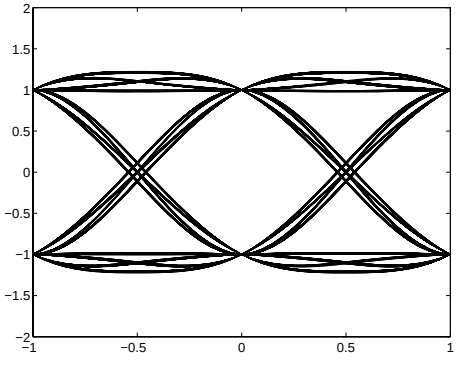
El paralelismo entre interpolación y comunicaciones digitales ha posibilitado el diseño de sistemas adaptados y libres de ISI basados en splines y α-splines, como se ha discutido en las anteriores secciones. Es evidente que cualquier tipo de pulso es susceptible de ser empleado en transmisión como pulso conformador, y que el receptor adaptado deberá filtrar la señal recibida utilizando un filtro cuya respuesta al impulso sea el pulso original invertido temporalmente. Sin embargo, en general, dicho sistema puede introducir ISI (además de la introducida por el propio canal), por lo que habitualmente se recurren a técnicas de igualación o de precodificación a costa, en muchas ocasiones, de perder la adaptación. Siguiendo el planteamiento de las anteriores secciones, es decir,



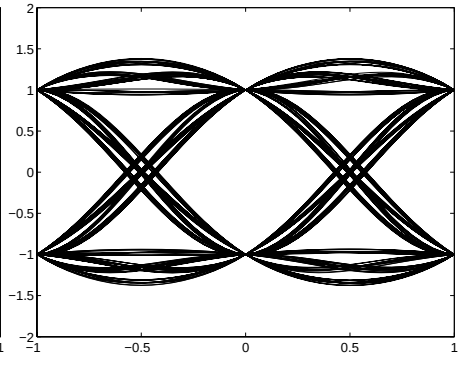
(a) α-spline $\alpha = 0$ (NRZ)



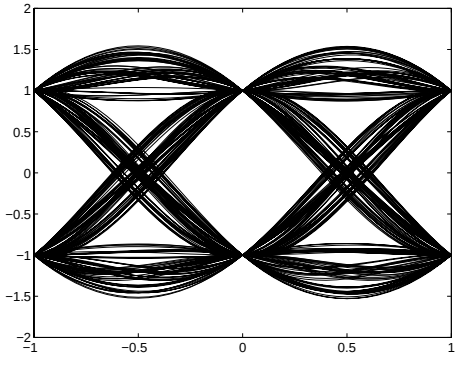
(b) α-spline $\alpha = 0{,}25$



(c) α-spline $\alpha = 0{,}5$



(d) α-spline $\alpha = 0{,}75$



(e) α-spline $\alpha = 1$ (Spline cúbico)

Figura 6.13: Diagramas de ojo de comunicaciones con conformación α-spline.

buscando la factorización de los filtros que intervienen en el proceso de la comunicación, es posible ampliar dicha metodología para utilizar pulsos distintos del spline (o del α-spline).

Si se considera un pulso conformador (filtro transmisor), $p(t)$, con una forma arbitraria y duración finita

$$p(t) = 0, \quad \forall |t| \geq \frac{N+1}{2}T, \tag{6.7}$$

donde T es el periodo de símbolo; la señal transmitida sigue la expresión

$$s_T(t) = \sum_{n \in \mathbb{Z}} s[n] p(t - nT).$$

El objetivo es encontrar un receptor adaptado para dicha señal $s_T(t)$, libre de interferencia entre símbolos. Evidentemente, como se comentó con anterioridad, la solución adaptada puede alcanzarse con el filtro receptor $p(-t)$, aunque esto provocará en general ISI.

6.6.1. Solución libre de ISI

Por simplicidad en posteriores desarrollos, se define el filtro $\varphi_p(t)$ como la convolución de los pulsos transmisor y receptor

$$\varphi_p(t) = p(t) * p(-t). \tag{6.8}$$

Éste es un filtro de fase cero, que provocará, en general, interferencia entre símbolos a la entrada del decisor. Sin embargo, y como se expondrá a continuación, mediante el prefiltrado de la señal de datos es posible evitar la ISI.

Si, tal como muestra la Figura 6.14, introducimos el prefiltro digital $d[n]$ y denominamos $s_d[n]$ a la señal prefiltrada, entonces, la señal recibida, $s_R(t)$, puede expresarse como

$$s_R(t) = \sum_{k \in \mathbb{Z}} s_d[n] \varphi_p(t - kT), \tag{6.9}$$

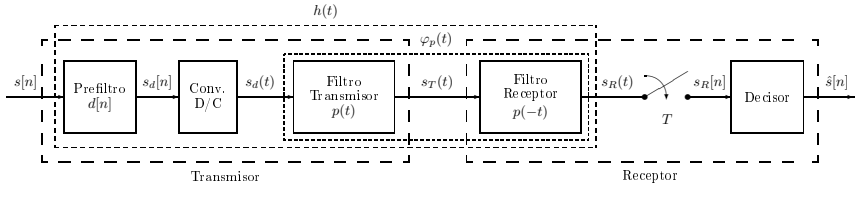


Figura 6.14: Esquema de comunicaciones libre de ISI.

o, de forma equivalente, como

$$s_R(t) = \sum_{n\in\mathbb{Z}} \left(\sum_{k\in\mathbb{Z}} s[k]d[n-k] \right) \varphi_p(t-nT) =$$
$$= \sum_{n\in\mathbb{Z}} s[n] \left(\sum_{k\in\mathbb{Z}} d[k]\varphi_p(t-kT-nT) \right) = \sum_{n\in\mathbb{Z}} s[n]h(t-nT),$$

donde la respuesta equivalente total del sistema de comunicaciones, $h(t)$, está constituida por la convolución del prefiltro, el filtro transmisor y el filtro receptor

$$h(t) = \sum_{n\in\mathbb{Z}} d[n]\varphi_p(t-nT).$$

La condición de ausencia de ISI en el receptor se puede escribir como

$$s_R(nT) = s[n], \tag{6.10}$$

es decir, la señal recibida, $s_R(t)$, toma exactamente los valores de la señal de datos, $s[n]$, en los instantes de muestreo, lo cual es equivalente a la condición de interpolación. Si se parte de (6.9) y se introduce el filtro discretizado $\varphi_p[n]$, el cual se obtiene muestreando $\varphi_p(t)$

$$\varphi_p[n] = \varphi_p(nT) = \varphi_p(t)|_{t=nT},$$

entonces la condición de interpolación (6.10) puede expresarse en forma de convolución discreta

$$s[n] = \sum_{k\in\mathbb{Z}} s_d[k]\varphi_p[n-k] = s_d[n] * \varphi_p[n].$$

Teniendo en cuenta que $s_d[n] = s[n] * d[n]$, resulta evidente que, con el fin de evitar la ISI, el prefiltro debe ser el filtro inverso de la versión discreta de $\varphi_p(t)$, es decir,

$$d[n] = \varphi_p^{-1}[n] \overset{z}{\longleftrightarrow} D(z) = \frac{1}{\Phi_p(z)}.$$

Esta metodología fue propuesta en [Unser93a] para aplicar técnicas de filtrado digital a la interpolación mediante splines. En esta ocasión, se recorre un camino, en cierta forma inverso, con el fin de aplicar técnicas interpolatorias en un área, las comunicaciones, que tradicionalmente ha sido contemplado de forma exclusiva desde el punto de vista del procesado de señal. Como resultado, el prefiltro digital, $d[n]$, el cual posibilita una transmisión libre de ISI, puede ser obtenido de una forma similar a la descrita en [Unser93b]. Sin embargo, el sistema de comunicaciones así obtenido no estará adaptado, por lo que su

rendimiento en BER no será óptimo. En el siguiente apartado se estudia una propiedad característica del prefiltro que posibilita la obtención de una solución que cumpla simultáneamente con los dos objetivos buscados: adaptación y ausencia de ISI.

6.6.2. Solución adaptada

Para la clase de filtro transmisor y receptor dados por (6.7), un símbolo concreto interfiere con N símbolos anteriores y N posteriores; es decir, el filtro discreto $\varphi_p[n]$ posee $2N+1$ coeficientes no nulos. Por tanto, el prefiltro digital, $d[n]$, será un filtro todo polos de orden $2N$ y con una respuesta al impulso infinita y simétrica, pudiéndose expresar su transformada z de la siguiente manera:

$$D(z) = \frac{1}{\Phi_p(z)} = \frac{1}{\varphi_p[0] + \sum_{n=1}^{N} \varphi_p[n]\left(z^n + z^{-n}\right)}. \tag{6.11}$$

Al igual que en [Unser93b], esta expresión puede ser factorizada como

$$D(z) = c_0 \prod_{i=1}^{N} \frac{-z_i}{(1 - z_i z^{-1})(1 - z_i z)} = c_0 \prod_{i=1}^{N} Q_{z_i}(z), \tag{6.12}$$

donde $c_0 = 1/\varphi_p[N]$ y (z_i, z_i^{-1}) son las N parejas de raíces del polinomio del denominador de (6.11), es decir,

$$\left\{(z_i, z_i^{-1}) \ / \ Q_{z_i}(z_i) = \infty, \ |z_i| < 1\right\}.$$

Es necesario comentar que, por simplicidad, en este desarrollo se supone que los polos, z_i, son simples. No obstante, la extensión para polos con multiplicidad superior a uno es inmediata. Dada la estructura del prefiltro digital (6.12), es posible su descomposición en un término causal, $D^+(z)$, y un término anticausal, $D^-(z)$, según

$$D(z) = c_0 \left(\prod_{i=1}^{N} -z_i\right) D^+(z) D^-(z), \tag{6.13a}$$

siendo

$$d^+[n] = z_1^n u[-n] * \cdots * z_N^n u[-n] \overset{z}{\longleftrightarrow} D^+(z) = \frac{1}{1 - z_i z^{-1}}, \tag{6.13b}$$

$$d^-[n] = z_1^{-n} u[n] * \cdots * z_N^{-n} u[n] \overset{z}{\longleftrightarrow} D^-(z) = \frac{1}{1 - z_i z^{-1}} = D^+(1/z). \tag{6.13c}$$

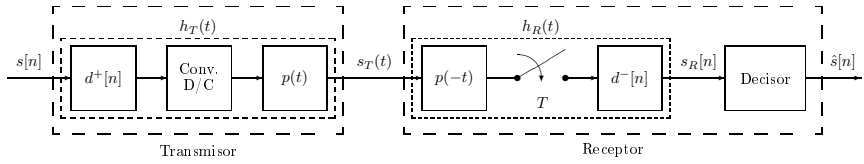


Figura 6.15: Esquema de comunicaciones adaptado y libre de ISI.

dos filtros de coeficientes reales.[2]

En resumen, la característica de fase cero del filtro $\varphi_p[n]$ posibilita una descomposición del prefiltro digital en dos filtros que, como se observa en (6.13c), están adaptados, es decir

$$d^+[n] = d^-[-n].$$

El esquema de comunicaciones presentado en la anterior sección y representado en Figura 6.14 junto con la etapa de filtrado dada por (6.11) está libre de interferencia entre símbolos, pero no permite un receptor adaptado. Sin embargo, debido a la simetría del prefiltro, éste puede ser descompuesto en dos términos adaptados, uno causal y el otro anticausal según (6.13), de forma que es posible implementar uno en el transmisor y el otro en el receptor. El esquema resultante se representa en la Figura 6.15, donde se obvia el término constante $c_0 \prod_{i=1}^{N} -z_i$.

Por lo tanto, los filtros transmisores y receptores definidos por

$$\begin{aligned} h_T(t) &= \sum_{n\in\mathbb{Z}} d^+[n]p(t-nT), \\ h_R(t) &= \sum_{n\in\mathbb{Z}} d^-[n]p(-t-nT), \end{aligned}$$

están adaptados y posibilitan una transmisión libre de ISI para cualquier pulso conformador, $p(t)$, considerado.

6.6.3. Aspectos relativos a la implementación

Al igual que ocurría al emplear splines como pulsos conformadores, desde un punto de vista práctico, la principal dificultad de la metodología propuesta reside en la implementación de los filtros IIR $d^+[n]$ y $d^-[n]$. El causal, $d^+[n]$, podría ser realizado mediante una única multiplicación-acumulación por cada símbolo de entrada. No obstante, el filtro anticausal, $d^-[n]$, debe ser de nuevo

[2] Dado que $d[n]$ es un filtro real, o bien sus polos son reales, o bien aparecen por parejas de complejos conjugados, por lo que tanto $d^+[n]$ como $d^-[n]$ serán reales.

truncado para mantener la causalidad, y el número de coeficientes a utilizar dependerá, evidentemente, del pulso $p(t)$ considerado en cada caso. Este hecho motiva que, si se desea preservar la adaptación, el prefiltro transmisor también deberá ser implementado en forma de FIR truncado con el mismo número de coeficientes.

Conviene señalar además que, cuando se aproxime el filtro anticausal mediante un FIR, no es necesario realizar multiplicaciones en el transmisor si los datos de entrada son binarios ($s[n] = \pm 1$), pues únicamente será preciso llevar a cabo operaciones de suma o resta, incrementando así la simplicidad del transmisor. En cualquier caso, el filtro discreto FIR del transmisor puede ser fácilmente realizado en forma de tabla de búsqueda, basado en una memoria de tipo ROM que contiene los valores de la señal de salida y que se indexa con los datos de entrada.

6.7. Implementación FIR en sistemas con sobremuestreo

6.7.1. Introducción

En los apartados anteriores de este capítulo se ha descrito una metodología general, basada en la teoría de interpolación funcional, para el diseño de sistemas de comunicaciones sin ISI y que permitan, simultáneamente, una implementación adaptada. Una limitación de la metodología propuesta (aunque poco importante desde el punto de vista práctico) es el hecho de que el prefiltro digital para una determinada forma de onda finita resulta ser IIR, lo que obliga a truncar los filtros adaptados obtenidos a partir de dicho prefiltro.

En este apartado se demostrará que esta limitación puede evitarse cuando se trabaja con un cierto factor de sobremuestreo en el receptor. En concreto, se presentará un procedimiento de diseño de sistemas de comunicaciones digitales con los condicionantes ya considerados en anteriores secciones:

- Sistema basado en pulso o filtro conformador arbitrario, $p(t)$, del tipo (6.7).
- Recepción libre de ISI.
- Receptor adaptado.

Adicionalmente, se considerará la siguiente limitación:

- Filtros FIR en transmisíón y recepción.

La metodología desarrollada en esta Tesis para cumplir simultáneamente las condiciones anteriores se basa en la teoría de las parejas de filtros biortogonales (*Biorthogonal Partners*), en adelante BP, recientemente desarrollada

por P. P. Vaidyanathan y B. Vrcelj [Vaidyanathan01c, Vrcelj02b, Vrcelj02a]. Un filtro digital $F(z)$ se dice que es pareja biortogonal de otro filtro digital $H(z)$ con respecto a un factor de sobremuestreo M si su conexión en serie $G(z) = H(z)F(z)$ satisface la propiedad Nyquist(M), es decir, si $g[n]$ posee cruces por cero cada M muestras excepto en el origen de tiempos, donde $g[0] = 1$.[3] En [Vaidyanathan01c] se presentan las condiciones necesarias y suficientes que debe cumplir un filtro $H(z)$ para que exista un filtro FIR $F(z)$ BP del anterior. De forma general, se puede decir que la existencia de un BP de tipo FIR requiere que las componentes polifase del filtro inicial $H(z)$ no posean ningún cero común. En el problema de diseño de filtros conformadores y receptores que nos ocupa, es necesario imponer otra serie de restricciones que aseguren que la respuesta global, $g[n]$, sea una función de autocorrelación, de manera que pueda ser descompuesta en un par de filtros adaptados. Básicamente, es preciso obtener una pareja de filtros FIR de forma que su respuesta conjunta sea un filtro FIR de fase lineal y Tipo I,[4] y, además, con espectro semidefinido positivo. A estos filtros los denominaremos pareja biortogonal semidefinida positiva (PSD-BP: *Positive Semidefinite Biorthogonal Partners*)

En este apartado se discutirán, en primer lugar, las condiciones para la existencia de parejas biortogonales PSD. Posteriormente, se formulará el diseño de dichas parejas de filtros biortogonales PSD como un problema de programación semidefinida convexa (SDP: *Semidefinite Programming*) para, finalmente, aplicar estas ideas a la conformación de pulsos en comunicaciones digitales.

6.7.2. Planteamiento del problema

La Figura 6.16 muestra un esquema básico de un sistema de comunicaciones digitales en tiempo discreto operando a una frecuencia de muestreo de T/M, siendo T el periodo de símbolo. En este modelo simplificado no consideraremos el canal, por lo que $G(z) = H(z)F(z)$ representa la respuesta combinada de los filtros transmisor y receptor sobremuestreados por un factor M.

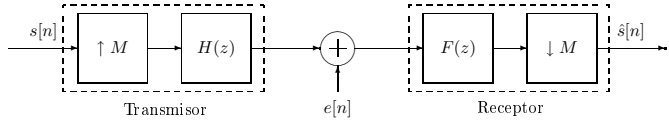


Figura 6.16: Modelo de un sistema de comunicaciones digitales paso bajo trabajando con un factor de sobremuestreo M.

[3] Ésta es la condición de cero ISI de un sistema con filtros transmisor, $H(z)$, y receptor, $F(z)$, trabajando con un factor de sobremuestreo M.

[4] $h[n] = h[M-n]$ con $M+1$ par y $h[n] \neq 0$ para $0 \leq n \leq M$.

Una posible formulación del problema consiste en tomar como punto de partida un filtro transmisor FIR, $H(z)$, previamente diseñado. Dicho filtro puede consistir, por ejemplo, en un filtro raíz cuadrada de coseno alzado truncado que provoca, por lo tanto, cierta ISI residual. Otro posible candidato como filtro inicial podría ser el núcleo discreto spline o α-spline sobremuestreado por un factor M. El objetivo es encontrar, si existe, otro filtro FIR, $F(z)$, tal que la respuesta conjunta $G(z) = H(z)F(z)$ sea una función autocorrelación y satisfaga la propiedad de Nyquist(M). De esta manera, será posible descomponer dicha función autocorrelación en un par de filtros adaptados, obteniéndose una implementación adaptada, libre de ISI y compuesta únicamente por filtros FIR.

Como ya se ha comentado, en ciertas situaciones se parte de un filtro FIR $H(z)$ y se busca otro filtro $F(z)$ de tal forma que la respuesta conjunta satisfaga ciertas restricciones. La primera de ellas es que $H(z)F(z)$ debe ser un filtro Nyquist(M),

$$[H(z)F(z)]_{\downarrow M} = 1, \tag{6.14}$$

es decir, el diezmado de orden M de $H(z)F(z)$ produce como resultado la función impulso $\delta[n]$. Cualquier filtro conformador que cumpla este requisito posibilitará una recepción libre de ISI.

Dos filtros que cumplan (6.14) se denominan parejas biortogonales (BP) con respecto a M y son tratados de forma extensa en [Vaidyanathan01c, Vrcelj02b]. En concreto, en [Vaidyanathan01c] se presentan las condiciones necesarias y suficientes que debe cumplir un filtro $H(z)$ para que exista una pareja biortogonal. Además, en dichos trabajos se describen aplicaciones en igualadores y en interpolación de mínimos cuadrados. A continuación, se resumen ciertos resultados necesarios para el desarrollo del presente trabajo. Si $M = 1$ (es decir, cuando el receptor toma una sola muestra por símbolo) y el filtro $H(z)$ no tiene ceros en el círculo unidad, la única pareja biortogonal existente es el filtro inverso $F(z) = 1/H(z)$, que es un filtro estable IIR. Éste es el resultado que ya se había obtenido con anterioridad para el prefiltro digital sin sobremuestreo. Una conclusión importante es que una implementación basada únicamente en filtros FIR sólo es posible si el receptor sobremuestrea la señal ($M > 1$). Éste es un resultado bien conocido en el ámbito de la igualación [Treichler96].

Si se supone que $M > 1$, el siguiente teorema en [Vaidyanathan01c] establece bajo que condiciones existe una pareja biortogonal de tipo FIR.

Teorema 6.1 (BP FIR) *Suponiendo un filtro FIR $H(z)$, entonces existe una pareja biortogonal de tipo FIR si y sólo si $H(z)$ no posee factores del tipo $(1 - z^{-M}e^{j\omega_1})$, que representan un conjunto de M ceros uniformemente espaciados en el círculo unidad.*

En la práctica, dado $H(z)$ es posible encontrar su BP resolviendo un sistema de ecuaciones lineales imponiendo la propiedad Nyquist(M). Esta idea, que será

descrita posteriormente con más detalle, es similar a la propuesta en [Vetterli86] para encontrar un banco de filtros de reconstrucción perfecta.

Resumiendo, los resultados previos nos permiten obtener una pareja biortogonal de tipo FIR (siempre que consideremos cierto sobremuestreo), de forma que la función de transferencia global esté libre de ISI. Sin embargo, para la conformación de pulsos en comunicaciones digitales es preciso descomponer la respuesta global en un par de filtros adaptados, pues ello posibilita la detección óptima, siempre que el ruido aditivo sea blanco. Esto introduce una restricción adicional consistente en que la respuesta global $G(z) = H(z)F(z)$ debe tener un espectro positivo, es decir,

$$G(\omega) \geq 0, \quad |\omega| \leq \pi. \tag{6.15}$$

Bajo esta condición, se pueden obtener un par de filtros adaptados mediante factorización espectral $G(z) = R(z)R(1/z)$.[5] Los factores de fase máxima y fase mínima pueden ser asignados arbitrariamente al filtro transmisor o al filtro receptor.

A una pareja biortogonal tipo FIR que cumple (6.15) la denominaremos en esta Tesis pareja biortogonal semidefinida positivamente (BP-PSD). Como se demostrará más tarde, el diseño de BP-PSD puede ser formulado como un problema de programación semidefinida convexa (SDP) que puede ser resuelto eficientemente utilizando métodos de punto interior [Ye97], disponibles en librerías de Matlab como *SeDuMi* [Sturm98].

La idea de formular el diseño de filtros conformadores como un problema SDP no es nueva, pues ha sido extensamente utilizada por T. N. Davidson y otros autores [Davidson00, Davidson01, Davidson02, Wu96]. En dichos trabajos, ciertas restricciones (tales como la propiedad Nyquist(M), máscaras espectrales o funciones de coste impuestas para incrementar la inmunidad frente al *jitter*) son transformadas en restricciones lineales y en restricciones de desigualdades lineales matriciales (LMI: *Linear Matrix Inequality*) que pueden ser resueltas mediante SDP. El problema en [Davidson00, Davidson01] también es parametrizado en términos de la función autocorrelación que resulta de la convolución de dos filtros adaptados. La principal diferencia entre estos trabajos y el que nos ocupa es que el procedimiento de diseño se inicia desde un filtro $H(z)$ de partida, mientras que Davidson y coautores arrancan "desde cero". Especificando por adelantado $H(z)$ se dispone de ciertas ventajas como, por ejemplo, el hecho de que es posible aplicar la teoría de parejas biortogonales y explotar ciertos resultados útiles relativos a la existencia de una implementación FIR y libre de ISI. Es más, puesto que la solución final es forzada a cumpir $G(z) = H(z)F(z)$, el resultado heredará del filtro inicial de diseño, $H(z)$, ciertas características (relativas al ancho de banda, por ejemplo). Es decir, con la

[5] Sólo se consideran aquí filtros con coeficientes reales.

metodología propuesta se intercambian ciertas características del diseño inicial (por ejemplo fase lineal de los filtros transmisores y receptores) por otras nuevas (por ejemplo ISI nula) manteniendo, al mismo tiempo, algunas de las características del diseño original (como el ancho de banda, implementación FIR, etc.).

En resumen, el problema planteado en esta Tesis puede formularse de la siguiente manera: dado un filtro FIR arbitrario, $H(z)$, y un factor de sobremuestreo, M, encontrar un filtro FIR, $F(z)$ tal que la respuesta global $G(z) = H(z)F(z)$ cumpla las condiciones

$$G(z)_{\downarrow M} = 1,$$

y

$$G(z) = R(z)R(1/z).$$

Desde el punto de vista de la conformación de pulsos, la primera condición implica encontrar una implementación FIR y libre de ISI, mientras que la segunda posibilita una descomposición adaptada de los filtros transmisor y receptor. Básicamente la idea consiste en que la respuesta global $g[n]$ debe tener forma de función de autocorrelación, es decir, filtro FIR de Tipo I con espectro semidefinido positivo. En el siguiente apartado, se discutirá cómo obtener un pareja de filtros biortogonales tales que $g[n]$ sea un filtro FIR tipo I. Posteriormente, se introducirá la restricción PSD en el procedimiento de diseño.

6.7.3. Parejas biortogonales de fase lineal

En este apartado se considera el siguiente problema: dado un filtro FIR arbitrario $H(z)$ y un factor de sobremuestreo M, encontrar una pareja biortogonal FIR, $F(z)$, tal que la respuesta global sea un filtro de fase lineal de Tipo I.[6]

Suponiendo que $H(z)$ y $F(z)$ son filtros FIR de longitud P y L, respectivamente, entonces $G(z)$ es un filtro FIR de longitud $P + L - 1$. Una primera cuestión a considerar es que para conseguir un filtro de tipo I y fase lineal, $P + L - 1$ debe ser un número impar, lo que implica que P y L tienen que ser ambos números pares o bien ambos impares.

Para resolver el problema se procederá de forma similar a [Vetterli86] con el fin de obtener un banco de filtros de reconstrucción perfecta de tipo FIR. En concreto, la convolución de $H(z)$ y $F(z)$ genera un conjunto de $P + L - 1$ ecuaciones lineales de la forma

[6] Es necesario destacar que el método descrito en [Vaidyanathan01c] no produce parejas biortogonales de fase lineal.

$$\begin{bmatrix} h[0] & 0 & \cdots & 0 & 0 \\ h[1] & h[0] & \cdots & 0 & 0 \\ h[2] & h[1] & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & h[1] & h[0] \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & h[P-1] \end{bmatrix} \begin{bmatrix} f[0] \\ f[1] \\ \vdots \\ f[L-1] \end{bmatrix} = \begin{bmatrix} g[0] \\ g[1] \\ \vdots \\ g[L+P-2] \end{bmatrix}. \tag{6.16}$$

Para obtener una solución válida, los coeficientes del filtro $g[n]$ deben cumplir

1. Condición Nyquist(M): $g[n]$ debe poseer cruces por cero periódicos separados por M muestras excepto[7] $g[(L+P-2)/2] = 1$.

2. Condición de simetría: $g[n] = g[L+P-2-n]$ para $n = 0, 1, \ldots, L+P-2$.

La condición Nyquist(M) fuerza $2\left\lfloor \frac{L+P-2}{2M} \right\rfloor + 1$ ecuaciones en (6.16), donde $\lfloor \cdot \rfloor$ denota la parte entera del argumento; mientras que la condición de simetría fuerza $P-1$ ecuaciones. Por lo tanto, disponemos de un total de $2\left\lfloor \frac{L+P-2}{2M} \right\rfloor + P$ ecuaciones para las L incógnitas $f[0], \ldots, f[L-1]$. Es fácil comprobar que, dados P y M, siempre es posible encontrar un valor de L que satisfaga

$$L = 2\left\lfloor \frac{L+P-2}{2M} \right\rfloor + P, \tag{6.17}$$

obteniendo, entonces, un sistema lineal de L ecuaciones y L incógnitas que es, en general, resoluble (si no existiera una pareja biortogonal de tipo FIR, la matriz sería singular y el sistema no tendría solución). El valor de L proporciona el BP más corto que cumple las condiciones de simetría y Nyquist(M). Valores más grandes de L nos permitirán imponer condiciones adicionales, como veremos más tarde.

Un ejemplo muy simple sirve para ilustrar el procedimiento de diseño de parejas biortogonales de fase lineal: considérese el filtro FIR $H(z) = -1 + 2z^{-1} + 3z^{-2}$ y un factor de sobremuestreo $M = 2$. En este caso, de acuerdo con (6.17) la longitud del BP más corto es $L = 5$, por lo que el sistema (6.16) resulta

$$\begin{bmatrix} -1 & 0 & 0 & 0 & 0 \\ 2 & -1 & 0 & 0 & 0 \\ 3 & 2 & -1 & 0 & 0 \\ 0 & 3 & 2 & -1 & 0 \\ 0 & 0 & 3 & 2 & -1 \\ 0 & 0 & 0 & 3 & 2 \\ 0 & 0 & 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} f[0] \\ f[1] \\ f[2] \\ f[3] \\ f[4] \end{bmatrix} = \begin{bmatrix} g[0] \\ g[1] \\ g[2] \\ g[3] \\ g[4] \\ g[5] \\ g[6] \end{bmatrix}.$$

[7] Puesto que se supone causalidad de $H(z)$ y $F(z)$, el filtro de Nyquist introducirá un retraso de $n_0 = (L+P-2)/2$ muestras, es decir, $(G(z)z^{n_o})_{\downarrow M} = 1$.

La condición Nyquist(2) nos proporciona las siguientes $2\left\lfloor\frac{L+P-2}{2M}\right\rfloor+1=3$ ecuaciones

$$\begin{bmatrix} 2 & -1 & 0 & 0 & 0 \\ 0 & 3 & 2 & -1 & 0 \\ 0 & 0 & 0 & 3 & 2 \end{bmatrix} \begin{bmatrix} f[0] \\ f[1] \\ f[2] \\ f[3] \\ f[4] \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix},$$

que pueden ser reescritas como

$$\mathbf{H}_1\mathbf{f} = \mathbf{d}_1.$$

Por otro lado, las condiciones de simetría proporcionan $P-1=2$ ecuaciones adicionales $g[0]=g[6]$ y $g[2]=g[4]$, que pueden reescribirse como

$$\begin{bmatrix} -1 & 0 & 0 & 0 & -3 \\ 3 & 2 & -4 & -2 & 1 \end{bmatrix} \begin{bmatrix} f[0] \\ f[1] \\ f[2] \\ f[3] \\ f[4] \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

o, en notación matricial,

$$\mathbf{H}_2\mathbf{f} = \mathbf{d}_2.$$

Finalmente, la solución $\mathbf{f}$ se obtiene resolviendo el problema lineal

$$\mathbf{H}_t\mathbf{f} = \mathbf{d}_t,$$

donde $\mathbf{H}_t = \begin{bmatrix} \mathbf{H}_1 \\ \mathbf{H}_2 \end{bmatrix}$ y $\mathbf{d}_t = \begin{bmatrix} \mathbf{d}_1 \\ \mathbf{d}_2 \end{bmatrix}$.

En este ejemplo, $\mathbf{H}_t$ es una matriz 5×5 no singular, la pareja biortogonal resultante es $F(z)=\frac{1}{80}\left(9+18z^{-1}+14z^{-2}+2z^{-3}-3z^{-4}\right)$, y la respuesta global viene dada por $G(z)=\frac{1}{80}\left(-9+49z^{-2}+80z^{-3}+49z^{-4}-9z^{-6}\right)$, que es un filtro de fase lineal Tipo I y Nyquist(2).

Este ejemplo ilustra el procedimiento general a seguir cuando el filtro de partida $H(z)$ no tiene fase lineal. Sin embargo, si $H(z)$ posee fase lineal, algunas de las condiciones previas resultan redundantes y pueden ser eliminadas. Para describir el procedimiento de diseño en este caso, es importante considerar la siguiente proposición:

Lema 6.1 *Para que un sistema $G(z)=H(z)F(z)$ sea de Tipo I y fase lineal, $H(z)$ y $F(z)$ deben ser filtros FIR de fase lineal del mismo tipo (ambos tipo I, II, III o IV).*

Teniendo presente este lema, cuya demostración es trivial y no se incluye aquí, la principal diferencia con respecto al procedimiento general radica en que

ahora sólo es necesario imponer a la solución $G(z)$ la condición Nyquist(M), tal y como muestra el siguiente ejemplo. En él se considera el filtro (de tipo I) $H(z) = -1+2z^{-1}+3z^{-2}+2z^{-3}-z^{-4}$ y un sobremuestreo $M = 2$. De acuerdo con el anterior razonamiento, $F(z)$ también debe ser un filtro tipo I con L coeficientes (el número de incógnitas se reduce a $(L+1)/2$). Por otro lado, el número de condiciones impuestas por la propiedad Nyquist(2) es $\left\lfloor \frac{L+P-2}{2M} \right\rfloor + 1$, por lo que el mínimo valor de L que proporciona un sistema determinado es $L = 2$.

$$\begin{bmatrix} -1 & 0 & 0 \\ 2 & -1 & 0 \\ 3 & 2 & -1 \\ 2 & 3 & 2 \\ -1 & 2 & 3 \\ 0 & -1 & 2 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} f[0] \\ f[1] \\ f[2] \end{bmatrix} = \begin{bmatrix} g[0] \\ g[1] \\ g[2] \\ g[3] \\ g[4] \\ g[5] \\ g[6] \end{bmatrix}.$$

Las restricciones Nyquist(2) son, en este caso, $g[3] = 1$ y $g[5] = 0$,

$$\begin{bmatrix} 2 & 3 & 2 \\ 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} f[0] \\ f[1] \\ f[2] \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix};$$

y la simetría de tipo I fuerza $f[0] = f[2]$, por lo que el sistema a resolver es

$$\begin{bmatrix} 3 & 4 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} f[1] \\ f[2] \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

La pareja biortogonal resultante es

$$F(z) = \frac{1}{10}\left(1 + 2z^{-1} + z^{-2}\right),$$

y la respuesta global del sistema viene dada por

$$G(z) = \frac{1}{10}\left(-1 + 6z^{-2} + 10z^{-3} + 6z^{-4} - z^{-6}\right).$$

6.7.4. Parejas biortogonales semidefinidas positivas

La respuesta conjunta, $G(z)$, de la pareja biortogonal obtenida mediante la técnica descrita en el anterior apartado no posee necesariamente un espectro positivo. Por ejemplo, el espectro de $G(z)z^3 = \frac{1}{10}\left(-z^3 + 6z + 10 + 6z^{-1} - z^{-3}\right)$ no es positivo para todo $\omega \in [0, \pi]$. Por lo tanto, es sabido por el teorema de Féjer-Riesz [Riesz55] que no es posible obtener un par de factores espectrales adaptados $G(z) = R(z)R(1/z)$. Para alcanzar este objetivo debemos imponer la restricción adicional $G\left(e^{j\omega}\right) \geq 0$ para todo $\omega \in [0, \pi]$, que

impone un número infinito de condiciones en $g[n]$. Como alternativa, puede utilizarse un lema que permite caracterizar las matrices semidefinidas positivas [Anderson74] para transformar estas restricciones en una restricción lineal en la que intervienen los elementos de una matriz semidefinida positiva [Davidson01, Davidson02]. En particular, una versión del lema descrita en estos trabajos establece que $g[-(L+P-2)/2], \ldots, g[0], \ldots, g[(L+P-2)/2]$ tiene espectro positivo si y sólo si existe una matriz $\mathbf{X}$ semidefinida positiva de dimensión $(L+P-2)/2 \times (L+P-2)/2$ tal que

$$\sum_{i=1}^{\frac{L+P}{2}-n} x_{i,i+n} = g[n], \quad n = 0, \ldots, \frac{L+P}{2} - 1; \tag{6.18}$$

donde $x_{i,j}$ denota el elemento en la fila i y columna j de la matriz $\mathbf{X}$.

La ecuación (6.18) puede ser reescrita en términos de la (desconocida) pareja biortogonal $f[n]$, lo que permite reformular el problema de la siguiente manera: dado $h[n]$ de longitud P y un factor de sobremuestreo M, encontrar $f[n]$ de longitud L sujeto a las siguientes restricciones lineales:

i)

$$\mathbf{H}_t \mathbf{f} = \mathbf{d}_t, \tag{6.19}$$

que fuerza la condición Nyquist(M) y la propiedad de fase lineal;

ii)

$$\sum_{i=1}^{\frac{L+P}{2}-n} x_{i,i+n} = \sum_{k=0}^{L-1} f[k]h[n-k], \quad n = 0, \ldots, \frac{L+P}{2} - 1; \tag{6.20}$$

iii) y la desigualdad lineal matricial (LMI)

$$\mathbf{X} \succ 0, \tag{6.21}$$

que denota que $\mathbf{X}$ ha de ser semidefinida positiva.

De esta forma se formula el diseño de una pareja biortogonal semidefinida positivamente como un conjunto de restricciones lineales dadas por (6.19) y (6.20), y una restricción dada por la desigualdad lineal matricial en (6.21). Por tanto, nos enfrentamos a un problema de programación semidefinida (SDP) que puede ser resuelto eficientemente utilizando métodos de punto interior.

Resolviendo este problema SDP se obtendrá, o bien una pareja biortogonal $f[n]$ que produce (al convolucionarlo con $h[n]$) una función autocorrelación apropiada, o bien un certificado de que no existe solución que cumpla las restricciones. En este último caso, se puede incrementar la longitud de $f[n]$ a

$L' = L + 2$ y repetir el procedimiento hasta que se obtenga un resultado apropiado, momento en el cual se podrá disponer de una pareja de filtros adaptados mediante factorización espectral de $g[n] = r[n] * r[-n]$. En el siguiente apartado se presentan, a título de ejemplo, dos aplicaciones que ilustran el procedimiento completo y su aplicación práctica.

6.7.5. Ejemplos de diseño

6.7.5.1. Sistema Spline

En el Apartado 6.4 se ha propuesto un sistema de comunicaciones digitales que utiliza filtros conformadores basados en splines. Dicho esquema está libre de ISI, alcanza el rendimiento óptimo en términos de BER (pues está adaptado), posee buenas características espectrales y su implementación es muy simple, tanto desde un punto de vista hardware como software. Sin embargo, únicamente se considera el caso $M = 1$; es decir, el receptor trabaja a una frecuencia de muestreo de una muestra por símbolo. En estas condiciones, el filtro transmisor de partida, $H(z)$, es el filtro B-spline, cuya pareja biortogonal es un filtro IIR que debe ser truncado. El procedimiento descrito en el apartado que nos ocupa permite considerar el caso en el que $M > 1$, posibilitando una implementación sin ISI, adaptada y que utiliza únicamente filtros FIR.

Se considera el ejemplo en el que $H(z)$ es el núcleo spline cúbico sobremuestreado por un factor $M = 2$:

$$H(z) = \frac{1}{48}\left(1 + 8z^{-1} + 23z^{-2} + 32z^{-3} + 23z^{-4} + 8z^{-5} + z^{-6}\right);$$

es decir, se trata de un filtro tipo I de fase lineal y longitud $P = 7$. Utilizando el procedimiento descrito en esta sección, es posible obtener una pareja biortogonal semidefinida positiva. Para solucionar este problema de programación semidefinida se utiliza un programa MATLAB denominado Sedumi [Sturm98] desarrollado por J. F. Sturm para la resolución de problemas SDP. Tras ciertas manipulaciones iniciales, que reformulan el diseño de parejas biortogonales PSD en el formato requerido por Sedumi, se obtiene la siguiente solución para $L = 5$: $F(z) = \frac{1}{6}\left(1 - 8z^{-1} + 20z^{-2} - 8z^{-3} + z^{-4}\right)$. La respuesta global es $G(z) = \frac{1}{288}\left(1 - 21z^{-2} + 164z^{-4} + 288z^{-5} + 164z^{-6} - 21z^{-8} + z^{-10}\right)$, que es un filtro Nyquist(2) con respuesta frecuencial positiva. Es interesante hacer notar que, en este ejemplo, el diseño se inicia con un filtro de longitud $L = 5$ (que es la mínima longitud que proporciona un filtro tipo I de fase lineal y Nyquist(2)), alcanzándose la solución en la primera iteración, lo cual significa que, en este caso, el filtro $g[n]$ Nyquist(2) de fase lineal más corto ya posee espectro positivo.

Finalmente, los filtros transmisores y receptores se obtienen como factores espectrales del tipo $G(z) = R(z)R(1/z)$. Para ello se utiliza un algoritmo de

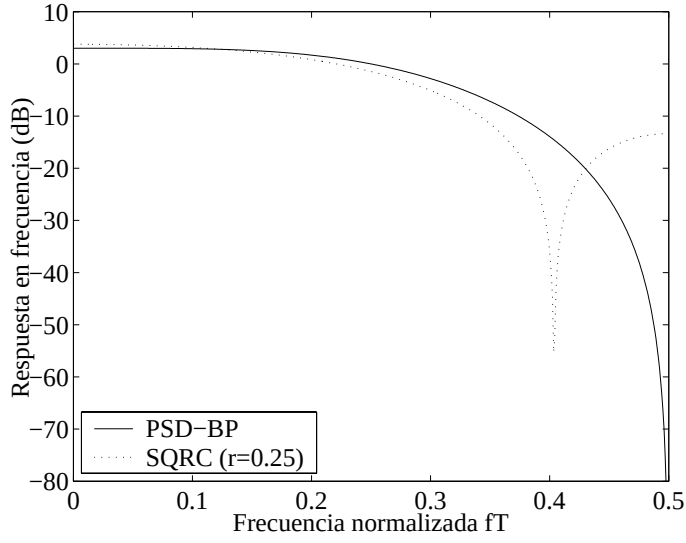


Figura 6.17: Respuesta en frecuencia de un filtro conformador PSD-BP basado en el filtro B-spline frente a un filtro SQRC de la misma longitud. Sobremuestreo $M = 2$.

factorización espectral descrito en [Mian82] que computa el cepstrum complejo de $g[n]$ y la DFT inversa de $\ln(G(2\pi k/N_{\mathrm{FFT}}))$ usando la fase desenrrollada y considerando FFT's de longitud $N_{\mathrm{FFT}} = 2^{14}$. Aplicando este algoritmo, el filtro transmisor resultante es $R(z) = 0{,}4623 + 0{,}8369z^{-1} + 0{,}2584z^{-2} - 0{,}1373z^{-3} - 0{,}0136z^{-4} + 0{,}0075z^{-5}$, mientras que el filtro receptor toma la forma $R(1/z)$.

Con fines comparativos, se ha diseñado con Matlab[8] un filtro raíz cuadrada de coseno alzado (SQRC: *Square Root Raised-Cosine*) de 5 coeficientes, factor de sobremuestreo $M = 2$ y exceso de ancho de banda $r = 0{,}25$. El filtro resultante es $R_{\mathrm{SQRC}}(z) = -0{,}0454 + 0{,}4397z^{-1} + 0{,}7554z^{-2} + 0{,}4397z^{-3} - 0{,}0454z^{-4}$ y su respuesta en frecuencia se representa en la Figura 6.17 junto con la del filtro propuesto $R(z)$.

Una figura de mérito del rendimiento final del sistema en términos de ISI viene dada por la distorsión ISI de pico, definida como [Alagha00, Tuncer02]

$$P_{\mathrm{ISI}} = \frac{\sum_{n\neq 0} |g[nM]|}{|g[0]|},$$

donde $g[n]$ es la respuesta conjunta de los filtros transmisor y receptor. En este ejemplo, $g[n]$ se obtiene como la convolución del factor espectral de fase

[8] El filtro se diseñó utilizando la instrucción `h=rcosfir(0.25,[-1,1],M,1,'sqrt')`.

mínima $r[n]$ con su versión reflejada $r[-n]$. La distorsión ISI de pico en este caso es $P_{\text{ISI}} = 9{,}6 \cdot 10^{-9}$, valor no exactamente nulo debido a inexactitudes numéricas del algoritmo de factorización espectral. Por otro lado, la distorsión ISI de pico del filtro SQRC de 5 coeficientes es $P_{\text{ISI}} = 0{,}2637$, un valor que puede afectar muy negativamente a la detección de la señal de comunicaciones.

6.7.5.2. Sistema SQRC

Se contempla en este ejemplo un sistema de comunicaciones digitales que utiliza filtros conformadores SQRC. En ciertas implementaciones, cuando el coste computacional debido al filtrado digital deba mantenerse en niveles bajos, se hace necesario utilizar filtros SQRC cortos, por ejemplo, utilizando Matlab es posible diseñar[9] un filtro SQRC de $P = 19$ coeficientes, exceso de ancho de banda $r = 0{,}35$ y factor de sobremuestreo $M = 3$.

El procedimiento descrito en este capítulo puede ser utilizado para intercambiar la propiedad de fase lineal del filtro SQRC por una ISI prácticamente nula. A partir de los resultados presentados en anteriores apartados, la mínima longitud de la pareja biortogonal es[10] $L_{\min} = 9$. La solución obtenida para esta longitud no es válida (no posee espectro positivo), por lo que, incrementando la longitud del filtro e iterando hasta obtener una solución válida, se obtiene una pareja biortogonal semidefinida positiva de $L = 31$ coeficientes, que genera un factor espectral de transmisión, $R(z)$, de 25 coeficientes de longitud.

Para establecer una comparación justa, en términos de longitud, se diseña un filtro SQRC de 25 coeficientes.[11] Los coeficientes y la respuesta en frecuencia de los tres filtros (SQRC inicial, BP-PSD y SQRC de 25 coeficientes) se muestran en la Figura 6.18. Se puede observar que el nuevo filtro conformador ha heredado el ancho de banda del filtro SQRC inicial a la vez que los lóbulos laterales se han incrementado levemente. Sin embargo, el BP-PSD proporciona una ISI de pico residual de $P_{\text{ISI}} = 2{,}1 \cdot 10^{-7}$, mientras que el SQRC de la misma longitud posee una ISI de $P_{\text{ISI}} = 0{,}0341$.

6.8. Conclusiones

En este capítulo se ha hecho uso de la estrecha relación existente entre las comunicaciones digitales y la interpolación funcional para proponer sistemas de comunicaciones adaptados y libres de ISI que emplean splines y α-splines. Dichos sistemas pueden resultar de interés en aplicaciones no extrictamente

[9] Se utiliza la instrucción `h=rcosfir(0.35,[-3,3],M,1,'sqrt')`.

[10] Dado que el filtro de partida $H(z)$ es de fase lineal, únicamente se aplican las restricciones Nyquist(M), por lo que se dispone de $L = 2\left\lfloor \frac{L+P-2}{2M} \right\rfloor + 1$ condiciones.

[11] `h=rcosfir(0.35,[-4,4],M,1,'sqrt')`.

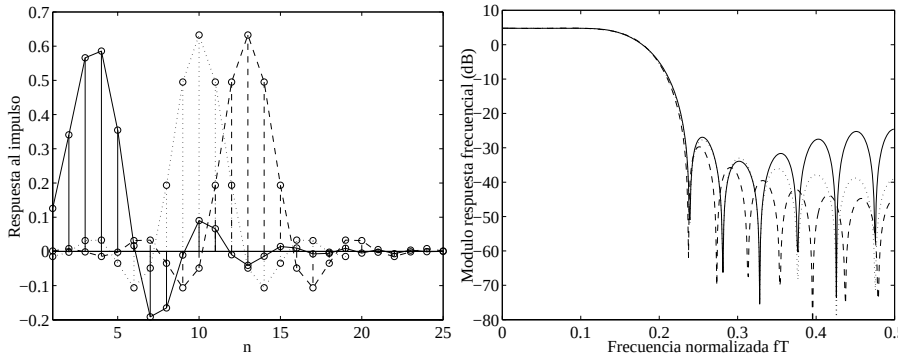


Figura 6.18: Respuesta al impulso (izquierda) y respuesta en frecuencia (derecha) del filtro BP-PSD de 25 coeficientes (línea continua), del filtro SQRC inicial de 19 coeficientes (puntos) y del filtro SQRC de 25 coeficientes (discontinua). Sobremuestreo $M = 3$.

limitadas en banda y en las que se precisen transmisores y receptores simples, pues permiten alcanzar el rendimiento óptimo de BER con unos gastos hardware y computacional mínimos. A modo de ejemplo, los sistemas de comunicaciones de alta velocidad que emplean códigos de línea NRZ (como muchos de los sistemas WLAN: *Wireless Local Aerea Network*) emplean habitualmente, por cuestiones de coste, receptores no adaptados y con cierta ISI. La técnica propuesta permitiría incrementar sustancialmente las prestaciones a un coste asumible.

En esa misma línea, la técnica descrita en el Apartado 6.6 permite implementar un sistema de comunicaciones que utilice filtros conformadores de forma arbitraria $p(t)$. A pesar de que la combinación de la respuesta del filtro transmisor y del filtro receptor introduce, en general, ISI, empleando la técnica propuesta es posible eliminarla manteniendo la condición de adaptación, realizando operaciones en transmisión y recepción a la frecuencia de símbolo y utilizando conversores ADC y DAC igualmente a la frecuencia de símbolo. Dichas operaciones son las requeridas por los filtros causal y anticausal, los cuales, típicamente, poseen longitudes muy cortas. Este esquema puede resultar de interés en el ámbito del espectro ensanchado por secuencia directa donde, en la actualidad, se utilizan conformaciones rectangulares y receptores no adaptados y con cierta ISI residual. La metodología propuesta posibilita cierto control espectral, pues la densidad espectral de potencia de la señal transmitida depende básicamente de la forma del pulso original, $p(t)$.

Cualquiera de las técnicas propuestas presenta la limitación de la necesidad de implementar filtros IIR tanto en el transmisor como en el receptor.

Una solución propuesta en este capítulo consiste en truncar ambos filtros, lo que proporciona resultados muy próximos a los óptimos en términos de BER, incluso para longitudes de filtros muy cortas, tal y como se ha demostrado mediante simulaciones.

En la misma línea, se ha propuesto una técnica, basada en parejas biortogonales semidefinidas positivas, que permite llevar a cabo una implementación FIR del receptor manteniendo las condiciones de ISI nula y adaptación. La metodología propuesta parte de un filtro inicial, del cual se mantienen ciertas características de interés (como el ancho de banda), para obtener una solución FIR, adaptada y libre de ISI a cambio de cierto sobremuestreo en el receptor.

Capítulo 7

Conclusiones, aportaciones y líneas futuras

En esta Tesis se han presentado diversas técnicas polinómicas en áreas de interpolación-muestreo-reconstrucción, modelado no lineal y comunicaciones digitales. En el presente capítulo se resumen las conclusiones y principales aportaciones de la Tesis, así como ciertas líneas de investigación que, a nuestro juicio, pueden ser de interés en un futuro.

7.1. Conclusiones y aportaciones

En el Capítulo 3 se ha demostrado la utilidad de las técnicas de interpolación polinómicas (splines y filtros de Lagrange) en aplicaciones de procesado de señal sujetas a condicionantes prácticos reales, como el sobremuestreo, la precisión finita y el coste computacional. Dicho capítulo ha pretendido servir para motivar y justificar el estudio y desarrollo de dichas técnicas que se ha llevado a cabo en el resto de la Tesis, no estando, sin embargo, exento de conclusiones de interés. En primer lugar, la interpolación spline cúbica es menos costosa computacionalmente que la interpolación sinc, y resulta, en muchas ocasiones, más precisa que interpolaciones que hacen uso de versiones truncadas de la función sinc (incluso cuando la cuadriplican en coste computacional). Es evidente que los resultados dependen fundamentalmente del contenido espectral de las señales a procesar, pero se ha demostrado que, en casos tan comunes como el de la interpolación de voz, la interpolación spline proporciona una solución equilibrada en términos de precisión y coste computacional. En segundo lugar, se ha obtenido una cota del error de interpolación de los filtros de Lagrange de retardo fraccionario que ha permitido desarrollar ecuaciones de diseño que relacionan el número de bits de resolución, el factor de sobremuestreo, el re-

tardo fraccionario y el orden del filtro. Así mismo, se ha comprobado que, a igualdad de precisión, el coste computacional de este tipo de filtros de Lagrange es comparable al de implementaciones polifásicas de técnicas multitasa, siendo inferior en el caso de factores de sobremuestreo superiores a diez.

Las técnicas de interpolación basadas en splines permiten un compromiso entre prestaciones y gasto computacional variando, para ello, el orden de la base spline utilizada (lineal, cuadrática, cúbica, etc.). Con objeto de incrementar la flexibilidad ofrecida por la interpolación spline, en el Capítulo 4 se ha propuesto una nueva familia de splines que ha sido denominada α-splines. En concreto, las principales aportaciones de dicho capítulo son:

- Una nueva familia de bases, denominadas bases α-spline, que constituyen una transición parametrizada entre dos bases spline de órdenes arbitrarios. Se ha demostrado que dichas funciones forman una base de Riesz y una partición de la unidad, lo que asegura que el espacio de señal α-spline es un subespacio cerrado de L_2 y que la representación α-spline es estable, no ambigua y capaz de aproximar cualquier señal con precisión arbitraria.

- Como caso particular, y de especial importancia, se ha considerado la base α-spline comprendida entre el spline lineal y el cúbico. Se propone el algoritmo de interpolación α-spline, tanto matricial como en forma de filtrado digital.

- La implementación de la interpolación α-spline mediante operaciones en punto fijo dadas las especiales características de los filtros interpoladores α-spline para ciertos valores del parámetro de suavidad α.

- Las expresiones cerradas de suavidad, varianza y área residual de la interpolación α-spline y la demostración de su comportamiento monótono con respecto al parámetro de suavidad. Dichas expresiones son función de α y de las muestras de partida (para el caso determinista) o de la autocorrelación de la secuencia aleatoria (en el caso estocástico).

- La extensión a muestreo no equiespaciado de la base α-spline y del correspondiente algoritmo matricial de interpolación.

Como conclusiones más importantes de este capítulo se pueden señalar:

- Dada la importancia y difusión de la interpolación lineal (debido a su sencillez y robustez) y de la interpolación spline cúbica (por su suavidad, derivabilidad y limitada complejidad), la interpolación α-spline se muestra como una solución de compromiso entre ambas soluciones y de gran utilidad, como se ha comprobado en diversos ejemplos prácticos.

- La dependencia monótona de la suavidad, la varianza y el área residual con respecto al parámetro α y la posibilidad de evaluar estos parámetros

a priori (es decir, antes de llevar a cabo la propia interpolación) permite implementar estrategias de búsqueda del valor de α óptimo que genere una solución que se adecúe a los requisitos particulares de cada problema de interpolación.

- En aplicaciones (principalmente de muestreo no uniforme) donde la interpolación lineal subestima la varianza de la señal original y, por contra, donde la interpolación cúbica la sobreestima (debido a los sobredisparos que provoca), la interpolación α-spline es capaz de aportar una solución continua, derivable, suave y que preserva el valor de la varianza para cierto valor de α. El mismo razonamiento se puede aplicar cuando se considere la suavidad, el área residual o cierto compromiso entre alguna de las tres características. En cualquier caso, la metodología de búsqueda del valor de α resulta simple, dado el comportamiento monótono de dichas características con respecto a α.

El Capítulo 5, referido al modelado α-spline, presenta las siguientes aportaciones:

- La utilización de la función base α-spline como núcleo del modelo canónico lineal a tramos para obtener el denominado modelo αPWL. Se trata de un modelo simple, derivable, con una capacidad de aproximación similar a la del modelo CPWL, y con una suavidad dependiente del parámetro α.

- La dependencia monótona creciente de la suavidad del modelo αPWL con respecto a α.

- El algoritmo de entrenamiento del modelo αPWL considerando dimensiones arbitrarias de los espacios de entrada y de salida. Se incluye el desarrollo necesario para llevar a cabo el ajuste del parámetro de suavidad en función del valor de la(s) derivada(s) de la función a modelar.

- La aplicación del modelo αPWL al modelado de amplificadores de potencia de radiofrecuencia, y la comparación de sus prestaciones con otros modelos no lineales como los modelos de Volterra, el perceptrón multicapa y el modelo de Shale.

- El empleo de técnicas de submuestreo para llevar a cabo la caracterización de amplificadores de potencia en RF. Las medidas temporales obtenidas permiten la obtención de modelos basados en la envolvente compleja de entrada-salida.

- El modelo αPWL del comportamiento en gran señal de un transistor HEMT bajo iluminación óptica.

Como conclusiones fundamentales extraídas del Capítulo 5, se pueden destacar las siguientes:

- El modelo αPWL proporciona resultados superiores a los obtenidos con redes neuronales como el MLP o la GRBF a la hora de modelar una función y sus derivadas. Su rendimiento es prácticamente idéntico al obtenido por el modelo SPWL basado en el núcleo logaritmo de coseno hiperbólico. Sin embargo, el menor gasto computacional necesario para entrenar y evaluar el modelo αPWL lo convierte en la mejor opción. De hecho, su coste computacional es sólo ligeramente superior al modelo CPWL, que constituye el caso límite del αPWL cuando $\alpha = 0$.

- El modelo αPWL, al igual que ocurre con el CPWL y SPWL, es sensible a los mínimos locales, por lo que los resultados obtenidos dependen, en gran medida, de la inicialización de las fronteras. A falta de un algoritmo de inicialización adecuado, la realización de varios entrenamientos a partir de fronteras iniciales aleatorias permite obtener resultados satisfactorios.

- El esquema de caracterización de amplificadores basado en técnicas de submuestreo permite aislar el comportamiento del amplificador, evitando la distorsión que introducen las etapas de conversión de frecuencia. Las medidas obtenidas permiten apreciar, de forma simple, las características de saturación, histéresis, memoria y recrecimiento espectral. Además posibilitan la obtención de modelos que representan fielmente el comportamiento del amplificador, tanto en banda como fuera de ella.

- En lo referente al modelado de amplificadores de potencia, el modelo αPWL se ha mostrado como el más preciso a la hora de describir el comportamiento en banda para todos los puntos de operación del amplificador (desde los prácticamente lineales hasta la región de saturación). Sin embargo, si únicamente se considera el comportamiento fuera de banda, el modelo MLP proporciona mejores resultados, a costa de un gasto computacional mucho mayor.

- Respecto al modelado de transistores HEMT sometidos a iluminación óptica, el modelo αPWL proporciona resultados de ajuste de la corriente de drenador similares a los del modelo CPWL y SPWL. Sin embargo, la capacidad de ajuste de las derivadas a costa de una baja carga computacional adicional convierte al modelo αPWL en el modelo idóneo.

En el Capítulo 6 se han abordado algunos aspectos del diseño de sistemas de comunicaciones digitales aprovechando ideas y conceptos de la teoría de la interpolación, siendo las principales aportaciones de dicho capítulo las siguientes:

- Aprovechando el conocido paralelismo entre comunicaciones digitales e interpolación funcional, se ha propuesto la utilización de bases spline como filtros conformadores. Sus especiales características y la descomposición de los filtros directos spline permiten implementar sistemas adaptados y sin ISI, con una destacable simplicidad tanto computacional como de hardware.
- De forma equivalente, se ha propuesto la utilización de bases α-spline para la realización de sistemas adaptados y sin ISI, extendiéndolo posteriormente a cualquier pulso de forma arbitraria.
- Se ha comprobado mediante simulaciones que es posible truncar los prefiltros IIR sin provocar por ello una pérdida apreciable en BER. En concreto, en los esquemas basados en el spline cúbico y en el α-spline, únicamente son necesarios tres coeficientes para mantener una BER prácticamente óptima.
- Una generalización de la teoría de parejas biortogonales que permite diseñar dos parejas biortogonales FIR tales que su respuesta global en frecuencia sea positiva y que puede, por tanto, factorizarse en dos filtros adaptados (filtros PSD-BP). Esta extensión permite, a partir de un pulso conformador inicial, una implementación FIR en transmisión y recepción libre de ISI y adaptada.

A la vista de lo presentado en el Capítulo 6 y de los resultados obtenidos, se pueden extraer las siguientes conclusiones:

- Las técnicas spline y α-spline propuestas resultan de interés en aplicaciones no estrictamente limitadas en banda y en las que se precisen transmisores y receptores simples, pues permiten alcanzar el rendimiento óptimo de BER con una complejidad hardware y un gasto computacional mínimo.
- La técnica de diseño de filtros PSD-BP proporciona un sistema final sin ISI, adaptado y FIR a cambio de cierto sobremuestreo. Adicionalmente, se ha comprobado en los ejemplos considerados que el sistema resultante hereda, en gran medida, las características espectrales del filtro conformador de partida.
- En aplicaciones que utilizan filtros raíz cuadrada de coseno alzado en las que surge cierta ISI debido a la corta longitud de los filtros utilizados, la técnica PSD-BP posibilita la obtención de filtros de longitud similar y que proporcionan una ISI virtualmente nula sin modificar apenas la respuesta frecuencial del sistema.

Por último, en el Apéndice F se hace referencia a las publicaciones a las que ha dado lugar la presente Tesis, tanto en lo que se refiere a revistas como a congresos nacionales e internacionales.

7.2. Líneas futuras

Considerando los resultados obtenidos en la presente Tesis, en este apartado se presentan posibles aplicaciones y ciertos aspectos de interés en los que podrían profundizar futuras investigaciones.

7.2.1. Implementación FIR causal de la interpolación spline y α-spline

Contrariamente a lo que ocurre en el procesado de imagen (señales bidimensionales), en el procesado en tiempo real de señales unidimensionales es necesario que los filtros que intervienen sean causales y finitos. A pesar de que las versiones IIR truncadas (como las propuestas en las comunicaciones basadas en splines y α-splines descritas en el Capítulo 6) se ha comprobado que proporcionan soluciones próximas a la óptima y que la técnica de diseños PSD-BP permite la implementación de sistemas FIR en sistemas que emplean sobremuestreo, existe margen de investigación para la búsqueda de esquemas de interpolación-aproximación basados en splines o α-splines que utilicen únicamente filtros FIR. Existen trabajos muy recientes en los que se persigue este mismo objetivo para el caso de la interpolación MOMS (*Maximal Order Interpolation of Minimal Support*) [Blu04], y es evidente que la resolución de dicho problema contribuiría a generalizar el uso de splines y α-splines en el procesado de series temporales.

7.2.2. Interpolación rápida α-spline en la generación de gráficos 3D

En el Capítulo 4 se ha demostrado la utilidad de la interpolación α-spline en la reconstrucción de la señal de velocidad de un flujo turbulento. Una línea de investigación abierta consiste en la búsqueda de nuevas aplicaciones que aprovechen las características de la familia α-spline. Un ejemplo a considerar es la generación de gráficos 3D y su manipulación (desplazamientos, rotaciones, deformaciones, etc.) en un ordenador, disciplina que ha experimentado un enorme desarrollo en la última década, principalmente gracias a los splines y a los avances en la tecnología microelectrónica. Sin duda, los splines cúbicos son los más utilizados en este ámbito, y es la obtención de los coeficientes spline (el problema directo) el proceso más costoso, pues se debe llevar a cabo el filtrado causal y anticausal (en dos o tres dimensiones, dependiendo de la aplicación) realizando operaciones en punto flotante.

En este tipo de aplicaciones, la interpolación rápida α-spline puede resultar de interés. En concreto, para un valor de $\alpha = 0{,}96$ se obtiene un esquema de interpolación prácticamente idéntico a la interpolación spline cúbica, con la ventaja de que los filtros digitales únicamente precisan sumas y desplazamientos

(dado que sus coeficientes son del tipo $d_\alpha[n] = 2^{-mn}$). Es evidente que esta interpolación α-spline se podría implementar de forma muy simple y eficiente en circuitos digitales integrados. Sin embargo, sería preciso llevar a cabo estudios perceptuales que compararan los resultados de la interpolación α-spline con los proporcionados por la interpolación spline cúbica.

7.2.3. Interpolación adaptativa con control de suavidad

En sistemas de procesado de señal *on line*, puede resultar de interés el control adaptativo de la suavidad de la señal interpolada. Por ello, una línea de investigación a considerar consistiría en el desarrollo de esquemas de interpolación α-spline con parámetro de suavidad adaptativo. Dado el comportamiento monótono de la suavidad con respecto a α, es de esperar que los algoritmos de adaptación sean simples y eficientes desde el punto de vista computacional. De forma similar, además de la suavidad se podría considerar el control adaptativo de la varianza, del área residual o de cierta combinación de cualquiera de las tres características.

7.2.4. Sistemas de comunicaciones

En aquellas situaciones prácticas en las que la tecnología permite generar pulsos con una forma determinada, la metodología descrita en el Capítulo 6 posibilita diseñar un sistema de comunicaciones adaptado y libre de ISI (y, por lo tanto, óptimo en el sentido de mínima probabilidad de error) a cambio de un incremento computacional ciertamente bajo debido a que los filtros digitales operan a la velocidad de símbolo y son, habitualmente, de longitud muy reducida. Por ejemplo, los pulsos de duración extremadamente pequeña (monociclo Gaussiano o similares [Lee01a, Lee01b]) que se han propuesto para su uso en sistemas de banda ultra-ancha (UWB: *Ultra Wideband* o también *impulse radio*) [Win98, Kusuma02, Oppermann04] podrían ser candidatos para implementar sistemas de comunicaciones de elevadísima velocidad, adaptados y sin ISI, a cambio de un coste computacional bajo y, posiblemente, asumible en el futuro.

7.2.5. Máquina de Vectores Soporte con núcleo α-spline

Un aspecto crucial en el entrenamiento de modelos formados como combinación lineal de ciertas funciones base consiste en la selección de los puntos sobre los que centrar las funciones base (nodos en el modelado spline en 1D, centroides en el caso de la Red de Funciones de Base Radial o hiperplanos de ruptura en el caso del modelo αPWL presentado en el Capítulo 5 de esta Tesis). Se trata de un problema no lineal para el que existen distintas soluciones subóptimas, por lo general específicas para cada modelo desarrollado.

Una solución a este problema la proporciona la Máquina de Vectores Soporte (SVM: *Support Vector Machine*). La SVM da una solución global al problema de modelado/regresión obteniendo simultáneamente tanto los puntos del conjunto de entrenamiento sobre los que centrar las funciones base como los parámetros de la combinación lineal óptima de dichas funciones. Empleando criterios derivados de la Teoría Estadística del Aprendizaje (*Statistical Learning Theory*) [Vapnik00], la SVM fue inicialmente propuesta para resolver problemas de clasificación binaria [Vapnik00, Cortés95], aunque posteriormente ha sido extendida para su uso en problemas de regresión/aproximación funcional [Smola98, Vapnik97, Muller98].

La idea fundamental del método consiste en realizar una transformación no lineal del espacio de entrada en otro espacio de características, *features*, de dimensión mayor (posiblemente infinita), en el que es posible resolver el problema de modelado/regresión mediante un hiperplano óptimo. El vector de pesos que define el hiperplano óptimo en el espacio transformado admite una expansión en términos de una serie de patrones representativos, denominados vectores soporte, de entre el conjunto de patrones de entrenamiento: éstos son los centroides o nodos del modelo.

Un aspecto muy importante para la aplicación de las SVM's es que, habitualmente, no es necesario conocer la transformación no lineal entre el espacio de entrada y el espacio de características; basta con conocer el producto escalar entre los vectores transformados que, bajo ciertas condiciones [Aizerman64], puede obtenerse de forma muy sencilla evaluando una función núcleo *simétrica*. Por ejemplo, la función de activación Gaussiana, habitualmente empleada en la Red de Funciones de Base Radial, es un kernel de estas características. Así mismo, el núcleo B-spline o su generalización α-spline propuesta en esta Tesis, también pueden interpretarse como producto escalar en un espacio transformado, por lo que ambos son kernels válidos para la formulación del problema mediante SVM's.

Una línea futura de interés, por tanto, consiste en la formulación de los modelos α-spline propuestos en esta Tesis dentro del marco de las SVM's. De esta manera, será posible obtener SVM's que, por ejemplo, transicionen de manera suave entre modelos lineales a tramos y modelos polinómicos; solucionando a la vez el problema de la selección de los centroides o nodos del modelo.

Apéndice A

Filtro de retardo fraccionario de Lagrange

A.1. Diseño de filtro con respuesta frecuencial máximamente plana

En este apéndice se demuestra la equivalencia entre el filtro interpolatorio de Lagrange y el obtenido mediante la técnica de diseño basada en el criterio de respuesta frecuencial máximamente plana [Oetken79]. Se puede considerar (3.20)

$$y[n] = x_i[n-D] = \sum_{k=0}^{N} h_L[k]x[n-k]$$

como una expresión general para la obtención del valor de una secuencia $x[n]$ en un instante $n-D$ a partir del conjunto de muestras situadas en torno a dicho instante y utilizando $N+1$ coeficientes.

El error cometido entre la señal así interpolada y el valor real correspondiente a la señal continua $x(t)$ dependerá de la diferencia entre el filtro ideal (3.18), $h_{\text{ideal}}[n]$, y el filtro utilizado, $h[n]$. Es posible expresar dicho error en el dominio frecuencial como

$$H_e\left(e^{j\omega}\right) = e^{-j\omega D} - \sum_{n=0}^{N} h[n]e^{-j\omega n}. \tag{A.1}$$

Si se opta por forzar esta función de error a comportarse de forma máximamente plana en torno a una cierta frecuencia, ω_0, entonces la interpolación será exacta en dicha frecuencia y, parece razonable pensar, muy próxima al ideal en

un entorno frecuencial más o menos amplio. Esto significa que las derivadas de (A.1) deben anularse en dicho punto, es decir,

$$H_e^{(k)}\left(e^{j\omega_0}\right) = \left.\frac{\partial^k H_e\left(e^{j\omega}\right)}{\partial\omega^k}\right|_{\omega=\omega_0} = 0, \quad k = 0, 1, \ldots, N. \tag{A.2}$$

Éste es el conocido como criterio de respuesta frecuencial máximamente plana, ya que el filtro obtenido presentará una banda de paso máximamente plana en torno a $\omega = \omega_0$ y una caída suave fuera de ella (filtros de tipo Butterworth). Sustituyendo (A.1) en (A.2), derivando y particularizando para $\omega_0 = 0$ se obtiene

$$D^k = \sum_{n=0}^{N} h[n]n^k, \quad k = 0, 1, \ldots, N;$$

es decir, N ecuaciones con N incógnitas (los N coeficientes del filtro $h[n]$) que se pueden expresar en forma matricial como

$$\mathbf{Vh} = \mathbf{v},$$

donde el vector $\mathbf{v}$ toma los valores $\mathbf{v} = [1, D, D^2, \ldots, D^N]^T$, $\mathbf{h}$ es el vector de coeficientes del filtro y $\mathbf{V}$ es una matriz de Vandermonde $N \times N$ y, por tanto, no singular e invertible; por lo que el sistema tendrá solución y ésta será única, y de valor [Kootsookos96]

$$h[n] = (-1)^{N-n}\binom{D}{n}\binom{D-n-1}{N-n}.$$

Como se puede observar, este resultado es idéntico a (3.21), por lo que se puede afirmar que el filtro de retardo fraccionario de Lagrange posee una respuesta máximamente plana alrededor de $\omega = 0$.

Apéndice B

Base de Riesz α-spline

En este apéndice se demuestra que el conjunto de bases α-spline forman una base de Riesz, con cotas

$$A = \left(\frac{2}{\pi}\right)^{2(p+1)} \left[\operatorname{sinc}\left(\frac{\alpha}{2}\right)\right]^{2q},$$
$$B = 1 + \frac{2^{2(p+1)}\left(2^{2(p+1)}-1\right)}{(2(p+1))!} B_{p+1}.$$

B.1. Cotas Riesz de la base α-spline

Dada la definición de base de Riesz (Definición 2.15) y la expresión de la transformada de Fourier de la base α-spline (4.3), se busca una cota superior e inferior de la función

$$S(\Omega) = \sum_{k\in\mathbb{Z}} \left[\operatorname{sinc}\left(\frac{\Omega}{2\pi} - k\right)\right]^{2(p+1)} \left[\operatorname{sinc}\left(\frac{\alpha\Omega}{2\pi} - \alpha k\right)\right]^{2q}. \tag{B.1}$$

Como consideración previa, es obvio que esta función es periódica de periodo 2π, es decir, $S(\Omega) = S(\Omega + 2k\pi)$, por lo que es posible restringir el estudio de las cotas al intervalo $[-\pi, \pi]$.

Respecto a la cota inferior, todos los términos de (B.1) son positivos, por lo que el sumatorio siempre será mayor o igual a cada uno de sus términos. En concreto se cumple

$$\begin{aligned} S(\Omega) &\geq \left[\operatorname{sinc}\left(\frac{\Omega}{2\pi}\right)\right]^{2(p+1)} \left[\operatorname{sinc}\left(\frac{\alpha\Omega}{2\pi}\right)\right]^{2q} \\ &\geq \left[\operatorname{sinc}\left(\frac{1}{2}\right)\right]^{2(p+1)} \left[\operatorname{sinc}\left(\frac{\alpha}{2}\right)\right]^{2q} = \left(\frac{2}{\pi}\right)^{2(p+1)} \left[\operatorname{sinc}\left(\frac{\alpha}{2}\right)\right]^{2q}, \end{aligned}$$

donde se ha tenido en cuenta el sumando correspondiente a $k = 0$ y el hecho de que en $[-\pi, \pi]$ las funciones $\mathrm{sinc}^2(\Omega/2\pi)$ y $\mathrm{sinc}^2(\alpha\Omega/2\pi)$ están acotadas inferiormente por el valor en $\Omega = \pi$.

Para la obtención de una cota superior se tiene en cuenta el hecho de que la función $\mathrm{sinc}^2(\Omega) \leq 1$, lo que se utiliza para acotar la función $\mathrm{sinc}^2(\alpha\Omega/2\pi - \alpha k)$ de (B.1), resultando

$$\begin{aligned} S(\Omega) &\leq \sum_{k\in\mathbb{Z}} \left[\mathrm{sinc}\left(\frac{\Omega}{2\pi} - k\right)\right]^{2(p+1)} \\ &\leq \sum_{k<0} \left[\mathrm{sinc}\left(\frac{-1}{2} - k\right)\right]^{2(p+1)} + 1 + \sum_{k>0} \left[\mathrm{sinc}\left(\frac{1}{2} - k\right)\right]^{2(p+1)} \\ &= 1 + \frac{2^{2(p+1)}\left(2^{2(p+1)} - 1\right)}{(2(p+1))!} B_{p+1}, \end{aligned}$$

puesto que se cumple

$$\begin{aligned} \mathrm{sinc}^2\left(\frac{\Omega}{2\pi} - k\right) &\leq \mathrm{sinc}^2\left(\frac{-1}{2} - k\right), \quad \forall\, k < 0,\ \Omega \in [-\pi, \pi]; \\ \mathrm{sinc}^2\left(\frac{\Omega}{2\pi}\right) &\leq \mathrm{sinc}^2(0) = 1, \quad \forall\, \Omega \in [-\pi, \pi]; \\ \mathrm{sinc}^2\left(\frac{\Omega}{2\pi} - k\right) &\leq \mathrm{sinc}^2\left(\frac{1}{2} - k\right), \quad \forall\, k > 0,\ \Omega \in [-\pi, \pi]; \end{aligned}$$

y

$$\begin{aligned} \sum_{k>0} \left[\mathrm{sinc}\left(\frac{1}{2} - k\right)\right]^{2(p+1)} &= \left(\frac{2}{\pi}\right)^{2(p+1)} \sum_{k>0} (1 - 2k)^{-2(p+1)} \\ &= \frac{2^{2p+1}\left(2^{2(p+1)} - 1\right)}{(2p+2)!} B_{p+1}, \end{aligned}$$

donde el valor de la serie y del n-ésimo número de Bernoulli, B_n, se encuentran respectivamente en [Spiegel70, páginas 109 y 115].

Apéndice C

Varianza y suavidad de la interpolación α-spline

En este apéndice se obtienen las expresiones de la varianza y el área residual de la interpolación α-spline de diversos tipos de señales, tanto deterministas como aleatorias. Previamente, se presenta una serie de expresiones asociadas al filtro directo α-spline necesarias para realizar posteriores cálculos.

C.1. Autocorrelaciones del filtro directo α-spline

Dada la expresión presentada en (4.28) del filtro directo α-spline:

$$d_\alpha[n] = \frac{(1-z_1)}{(z_1+1)} z_1^{|n|},$$

su autocorrelación determinista, que denotaremos como $\rho_\alpha[n]$, se escribe como

$$\rho_\alpha[n] = d_\alpha[n] * d_\alpha[-n] = d_\alpha[n] * d_\alpha[n],$$

siendo su valor

$$\rho_\alpha[n] = \frac{(1-z_1) z_1^{|n|} (|n| + 1 + (1-|n|) z_1^2)}{(z_1+1)^3}. \tag{C.1}$$

Así mismo, resulta de interés presentar la diferencia segunda del filtro directo α-spline $d_\alpha^{(2)}$,

$$d_\alpha^{(2)}[n] = d_\alpha[n] * (\delta[n+1] - 2\delta[n] + \delta[n-1]) = \begin{cases} -\dfrac{(z_1-1)^3}{(z_1+1)} z_1^{|n|-1}, & n \neq 0; \\ -\dfrac{2(z_1-1)^2}{(z_1+1)}, & n = 0; \end{cases}$$

y su autocorrelación determinista, $\rho_\alpha^{(2)}$, dada por la expresión

$$\rho_\alpha^{(2)}[n] = d_\alpha^{(2)}[n] * d_\alpha^{(2)}[-n] = d_\alpha^{(2)}[n] * d_\alpha^{(2)}[n], \tag{C.2a}$$

siendo alguno de sus valores los siguientes

$$\rho_\alpha^{(2)}[0] = \sum_{n=-\infty}^{\infty} (d_\alpha^{(2)}[n])^2 = \frac{2(z_1+3)(z_1-1)^4}{(z_1+1)^3}, \tag{C.2b}$$

$$\rho_\alpha^{(2)}[1] = \rho_\alpha^{(2)}[-1] = \sum_{n=-\infty}^{\infty} d_\alpha^{(2)}[n] d_\alpha^{(2)}[1-n] = \frac{2(z_1+2)(z_1-1)^5}{(z_1+1)^3}. \tag{C.2c}$$

C.2. Varianza de la interpolación α-spline de señales deterministas periódicas

Se considera la interpolación de N muestras pertenecientes a M periodos de una señal senoidal, $x[n]$, de frecuencia discreta $\omega_0 = 2\pi/T_0 = 2\pi M/N$:

$$x[n] = A\sin(\omega_0 n + \Phi_0) = A\sin\left(2\pi\frac{M}{N}n + \Phi_0\right),$$

con $0 \leq M \leq N-1$ y $0 \leq \omega_0 \leq 2\pi$.

La interpolación α-spline de esta secuencia proporciona una señal continua $s_\alpha(t)$ cuya potencia (que equivale a la varianza, pues posee media nula) se puede calcular, recordando (4.23), como

$$\mathrm{Var}\left(s_\alpha(t)\right) = \frac{1}{MT_0}\int_{MT_0} \left(s_\alpha(t)\right)^2 dt = \frac{1}{N}\int_N \left(\sum_{n=1}^{N} c[n]\beta_\alpha\left(t-n\right)\right)^2 dt.$$

Dado que en el intervalo $[n, n+1]$ solamente toman valores no nulos las bases $\beta_\alpha\left(t-n+1\right)$, $\beta_\alpha\left(t-n\right)$, $\beta_\alpha\left(t-n-1\right)$ y $\beta_\alpha\left(t-n-2\right)$, entonces

$$\begin{aligned}\mathrm{Var}\left(s_\alpha(t)\right) = \frac{1}{N}\sum_{n=1}^{N}\int_n^{n+1} &\left(c[n-1]\beta_\alpha\left(t-n+1\right) + c[n]\beta_\alpha\left(t-n\right)\right.\\ &\left.+c[n+1]\beta_\alpha\left(t-n-1\right) + c[n+2]\beta_\alpha\left(t-n-2\right)\right)^2 dt.\end{aligned}$$

Suponiendo condiciones de contorno periódicas, es decir, $c[n] = c[n+kN]$,

su potencia se puede escribir como

$$\begin{aligned}\mathrm{Var}\left(s_\alpha(t)\right) = & \frac{1}{N}\sum_{n=1}^{N} c^2[n]\int_{-\infty}^{\infty}\left(\beta_\alpha\left(t\right)\right)^2 dt \\ & + \frac{2}{N}\sum_{n=1}^{N} c[n]c[n+1]\int_{-\infty}^{\infty}\beta_\alpha\left(t\right)\beta_\alpha\left(t-1\right)dt \\ & + \frac{2}{N}\sum_{n=1}^{N} c[n]c[n+2]\int_{-\infty}^{\infty}\beta_\alpha\left(t\right)\beta_\alpha\left(t-2\right)dt \\ & + \frac{2}{N}\sum_{n=1}^{N} c[n]c[n+3]\int_{-\infty}^{\infty}\beta_\alpha\left(t\right)\beta_\alpha\left(t-3\right)dt.\end{aligned} \tag{C.3}$$

Al considerar una sinusoide periódica, sus coeficientes α-spline se pueden calcular como

$$c[n] = f[n]\cdot D_\alpha(z)|_{z=e^{j\omega_0}} = A\sin\left(\omega_0 n + \Phi_0\right)\frac{3}{3+\alpha(\cos\omega_0 - 1)},$$

por lo que los sumatorios de (C.3) toman el valor

$$\begin{aligned}\sum_{n=1}^{N} c[n]c[n+k] &= \frac{9A^2}{[3+\alpha(\cos\omega_0-1)]^2}\sum_{n=1}^{N}\sin\left(\omega_0 n+\phi_0\right)\sin\left(\omega_0(n+k)+\phi_0\right) \\ &= \frac{A^2}{2}\frac{9N}{[3+\alpha(\cos\omega_0-1)]^2}\cos\left(\omega_0 k\right).\end{aligned}$$

Finalmente, la varianza de la interpolación α-spline de N muestras correspondientes a M periodos de una secuencia senoidal de amplitud A y frecuencia ω_0 resulta

$$\mathrm{Var}\left(s_\alpha(t)\right) = \frac{A^2}{2}\frac{9\left(I_0 + 2\cos\left(\omega_0\right)I_1 + 2\cos\left(2\omega_0\right)I_2 + 2\cos\left(3\omega_0\right)I_3\right)}{[3+\alpha(\cos\omega_0-1)]^2}, \tag{C.4}$$

donde la integrales

$$I_n = \int_{-\infty}^{\infty}\beta_\alpha\left(t\right)\beta_\alpha\left(t-n\right)dt, \tag{C.5}$$

son polinomios del parámetro de suavidad α

$$
\begin{aligned}
I_0 &= \begin{cases} \frac{31\alpha^3-70\alpha^2+140}{210}, & 0 \leq \alpha \leq \frac{1}{2}; \\ \frac{-35\alpha^7+238\alpha^6-672\alpha^5+980\alpha^4-280\alpha^3+84\alpha^2-14\alpha+1}{630\alpha^4}, & \frac{1}{2} \leq \alpha \leq 1; \end{cases} \\
I_1 &= \begin{cases} \frac{-62\alpha^3+105\alpha^2+105}{630}, & 0 \leq \alpha \leq \frac{1}{2}; \\ \frac{400\alpha^7-2296\alpha^6+4704\alpha^5-3080\alpha^4+1960\alpha^3-588\alpha^2+98\alpha-7}{5040\alpha^4}, & \frac{1}{2} \leq \alpha \leq 1; \end{cases} \\
I_2 &= \begin{cases} \frac{31\alpha^3}{1260}, & 0 \leq \alpha \leq \frac{1}{2}; \\ \frac{-97\alpha^7+448\alpha^6-672\alpha^5+560\alpha^4-280\alpha^3+84\alpha^2-14\alpha+1}{1260\alpha^4}, & \frac{1}{2} \leq \alpha \leq 1; \end{cases} \\
I_3 &= \begin{cases} 0, & 0 \leq \alpha \leq \frac{1}{2}; \\ \frac{(2\alpha-1)^7}{5040\alpha^4}, & \frac{1}{2} \leq \alpha \leq 1; \end{cases}
\end{aligned}
$$

Mediante (C.4) es posible calcular la varianza de la interpolación α-spline de una secuencia senoidal de amplitud unidad en función de su frecuencia, ω, y del parámetro de suavidad, α, como

$$
\mathrm{Var}\,(\alpha, \omega) = \frac{9\,(I_0 + 2\cos(\omega) I_1 + 2\cos(2\omega) I_2 + 2\cos(3\omega) I_3)}{[3 + \alpha(\cos\omega - 1)]^2}
$$

y los valores de las integrales de (C.5).

Evaluando la anterior expresión, se comprueba que la varianza de interpolación α-spline crece monótonamente conforme aumenta el valor del parámetro α para cualquier valor de la frecuencia ω, excepto en una pequeña región ($0{,}7 < \alpha < 1$ y $\omega > 0{,}9\pi$ representada en la Figura 4.12), donde la varianza es prácticamente constante. En la práctica, cuando la señal a interpolar está formada por múltiples componentes frecuenciales, la varianza de la interpolación crecerá conforme crece el parámetro α. En este caso, es posible evaluar la varianza de la interpolación α-spline *a priori* (es decir, antes de realizar la interpolación), a partir de la DFT, $X[k]$, de las muestras a interpolar, $x[n]$, utilizando la expresión

$$
\mathrm{Var}\,(s_\alpha(t)) = \frac{1}{N^2} \sum_{k=1}^{N-1} |X[k]|^2\, \mathrm{Var}\,(\alpha, \omega = 2\pi k/N)\,,
$$

en la que no se incluye el término de continua de la señal ($k = 0$ en la DFT) pues no contribuye a la varianza de la interpolación, ya que la interpolación α-spline es capaz de sintetizar exactamente dicha señal constante, la cual posee una varianza nula.

C.3. Varianza de la interpolación α-spline de señales aleatorias

Para el cálculo de la varianza de la interpolación α-spline, $\mathrm{Var}\left(\mathbf{S}_\alpha\right)$, se procede de forma similar al Apartado 4.6.3, obteniendo previamente la autocorrelación del proceso estocástico interpolado

$$R_{\mathbf{S}_\alpha}(\tau) = \sum_{n=-\infty}^{\infty} R_{\mathbf{C}}[n] \int_{-\infty}^{\infty} \beta_\alpha\left(t-n+\tau\right)\beta_\alpha\left(t\right)dt.$$

La varianza media de la interpolación α-spline se obtiene particularizando la anterior expresión en $\tau = 0$

$$\mathrm{Var}\left(\mathbf{S}_\alpha\right) = \sum_{n=-\infty}^{\infty} R_{\mathbf{C}}[n] \int_{-\infty}^{\infty} \beta_\alpha\left(t-n\right)\beta_\alpha\left(t\right)dt.$$

Teniendo en cuenta el soporte compacto de las funciones base α-spline, las integrales que intervienen, cuyo valor viene dado por (C.5), son nulas para $|n| > 3$

$$\begin{aligned}\mathrm{Var}\left(\mathbf{S}_\alpha\right) = R_{\mathbf{C}}[0]I_0 &+ \left(R_{\mathbf{C}}[1] + R_{\mathbf{C}}[-1]\right)I_1 \\ &+ \left(R_{\mathbf{C}}[2] + R_{\mathbf{C}}[-2]\right)I_2 + \left(R_{\mathbf{C}}[3] + R_{\mathbf{C}}[-3]\right)I_3,\end{aligned}$$

y, dado que la autocorrelación $R_{\mathbf{C}}[n]$ es simétrica (se está suponiendo procesos estocásticos reales), entonces la expresión

$$\mathrm{Var}\left(\mathbf{S}_\alpha\right) = R_{\mathbf{C}}[0]I_0 + 2R_{\mathbf{C}}[1]I_1 + 2R_{\mathbf{C}}[2]I_2 + 2R_{\mathbf{C}}[3]I_3, \tag{C.6}$$

permite obtener el valor de la varianza de la interpolación α-spline evaluando las integrales (C.5) y conociendo la autocorrelación del proceso de entrada, pues $R_{\mathbf{C}}[n]$ se calcula mediante

$$R_{\mathbf{C}}[n] = R_{\mathbf{X}}[n] * \rho_\alpha[n], \tag{C.7}$$

donde los valores de $\rho_\alpha[n]$ vienen dados por (C.1).

En el dominio frecuencial, la varianza de la interpolación se puede calcular en función de la densidad espectral de potencia, $S_{\mathbf{X}}(\omega)$, del proceso de entrada y de la respuesta en frecuencia de la interpolación α-spline (4.37)

$$\begin{aligned}\mathrm{Var}\left(\mathbf{S}_\alpha\right) &= \frac{1}{2\pi}\int_{-\infty}^{+\infty} S_{\mathbf{S}_\alpha}(\Omega)d\Omega = \frac{1}{2\pi}\int_{-\infty}^{+\infty} S_{\mathbf{X}}(\Omega)\left|H_\alpha(\Omega)\right|^2 d\Omega \\ &= \frac{1}{2\pi}\int_{-\infty}^{+\infty} S_{\mathbf{X}}(\Omega)\frac{9\ \mathrm{sinc}^4(\Omega/2\pi)\ \mathrm{sinc}^4(\alpha\Omega/2\pi)}{\left[3+\alpha(\cos\Omega-1)\right]^2}\ d\Omega,\end{aligned}$$

aunque evaluar esta integral resulta complicado en la mayoría de los casos.

C.3.1. Varianza de la interpolación α-spline de señales blancas

Si el proceso $\mathbf{X}[n]$ es blanco de media nula, entonces la autocorrelación de los coeficientes α-spline es $R_{\mathbf{C}}[n] = \sigma^2_{\mathbf{X}}\rho[n]$, con lo que la varianza de la interpolación, haciendo uso de (C.6) y (C.1), resulta

$$\mathrm{Var}\,(\mathbf{S}_\alpha) = \sigma^2_{\mathbf{X}}\frac{(1-z_1)}{(z_1+1)^3}\left[(1+z_1^2)I_0 + 4z_1I_1 + 2z_1^2(3-z_1^2)I_2 + 4z_1^3(2-z_1^2)I_3\right].$$

Desarrollando esta expresión se obtiene:

- Si $0 \leq \alpha \leq \frac{1}{2}$, entonces

$$\mathrm{Var}\,(\mathbf{S}_\alpha) = -2\sigma^2_{\mathbf{X}}\frac{35z_1^5 + 488z_1^4 + 1079z_1^3 + 595z_1^2 + 70z_1 - 35}{105(z_1-1)^2(z_1+1)^3}. \quad \text{(C.8a)}$$

- Si $\frac{1}{2} \leq \alpha \leq 1$, entonces

$$\begin{aligned}\mathrm{Var}\,(\mathbf{S}_\alpha) = \frac{-\sigma^2_{\mathbf{X}}}{1632960z_1^4(z_1-1)^2(z_1+1)^3}\big[&z_1^{16} + 71z_1^{15} + 2175z_1^{14}\\ &+ 37387z_1^{13} + 391727z_1^{12} + 2530731z_1^{11} + 9591043z_1^{10}\\ &+ 19088975z_1^9 + 21519117z_1^8 + 15050501z_1^7 + 4966381z_1^6\\ &-1908879z_1^5 - 1772155z_1^4 - 68663z_1^3 - 4143z_1^2 - 139z_1 - 2\big].\end{aligned} \quad \text{(C.8b)}$$

En la Figura C.1 se observa la dependencia de la varianza de la interpolación α-spline de señales aleatorias blancas con el parámetro α. Bien sea a partir de dichas gráficas, o de las expresiones anteriores, se comprueba que

$$\frac{d\mathrm{Var}\,(\mathbf{S}_\alpha)}{dz_1} \leq 0, \quad \forall z_1 \in [-2+\sqrt{3}, 0],$$

por lo que

$$\frac{d\mathrm{Var}\,(\mathbf{S}_\alpha)}{d\alpha} \geq 0, \quad \forall \alpha \in [0,1].$$

Es decir, la varianza de la interpolación α-spline de un proceso estocástico blanco crece monótonamente conforme se incrementa el parámetro α.

C.3.2. Varianza de la interpolación α-spline de señales aleatorias senoidales

Si el proceso estocástico a interpolar es senoidal de frecuencia ω_0, su autocorrelación se puede escribir como

$$R_{\mathbf{X}}[n] = \sigma^2_{\mathbf{X}}\cos(\omega_0 n), \quad \forall n \in \mathbb{Z},\ 0 \leq \omega_0 \leq \pi,$$

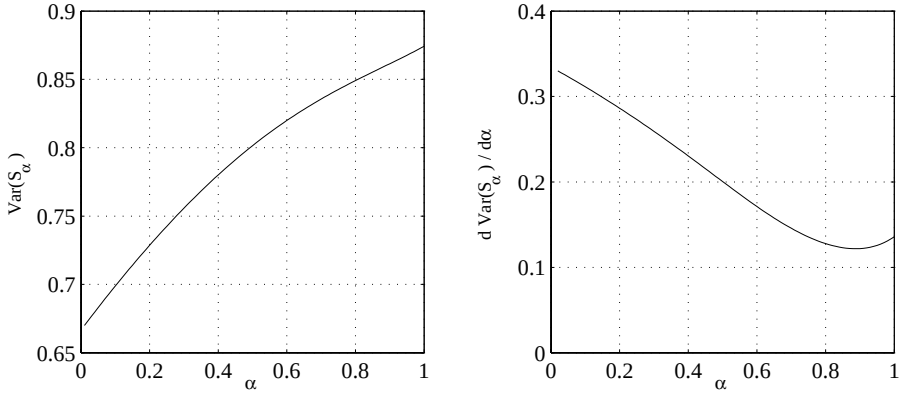


Figura C.1: Varianza de la interpolación α-spline de señales blancas (izquierda) y su derivada respecto a α (derecha).

y, considerando (C.7) y (4.36), la de los coeficientes α-spline como

$$R_{\mathbf{C}}[n] = \sigma_{\mathbf{X}}^2 \cos(\omega_0 n) \left|H_{d_\alpha}(\omega)\right|^2,$$

con lo que la varianza media de la interpolación, dada por (C.6), toma el valor

$$\mathrm{Var}\left(\mathbf{S}_\alpha\right) = \sigma_{\mathbf{X}}^2 \frac{9\left(I_0 + 2\cos\left(\omega_0\right)I_1 + 2\cos\left(2\omega_0\right)I_2 + 2\cos\left(3\omega_0\right)I_3\right)}{\left[3 + \alpha(\cos\omega_0 - 1)\right]^2}.$$

De nuevo, esta expresión es equivalente a la correspondiente a la varianza de señales deterministas senoidales, por lo que su comportamiento respecto al parámetro α resulta idéntico, es decir, la varianza crece conforme aumenta α para cualquier valor de α y de frecuencia, ω, considerados, salvo en la región representada en Figura 4.12, donde decrece levemente.

C.4. Área residual de la interpolación α-spline de señales deterministas periódicas

Nuevamente, se considera la interpolación α-spline de N muestras pertenecientes a M periodos de una señal senoidal, $x[n]$, con frecuencia discreta $\omega_0 = 2\pi/T_0 = 2\pi M/N$:

$$x[n] = A\sin\left(\omega_0 n + \Phi_0\right) = A\sin\left(2\pi\frac{N}{M}n + \Phi_0\right),$$

siendo $0 \leq M \leq N-1$ y $0 \leq \omega_0 \leq 2\pi$.

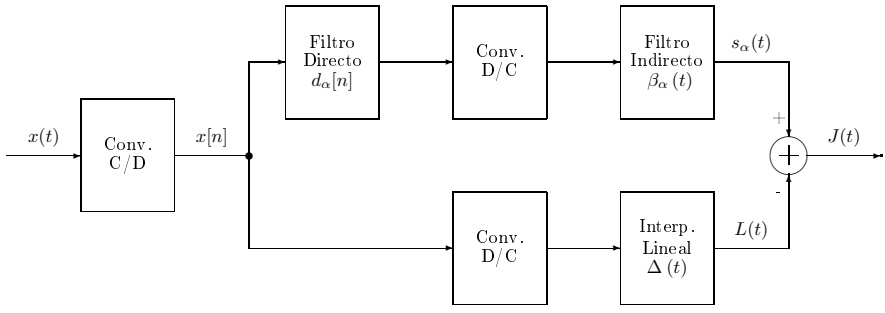


Figura C.2: Esquema de cálculo del área residual α-spline.

Para el cálculo del área residual de una función se define la señal $J(t)$ (ver Figura C.2) como la diferencia entre la señal interpolada α-spline, $s_\alpha(t)$, y la interpolación lineal, $L(t)$, es decir,

$$J(t) = s_\alpha(t) - L(t),$$

de forma que el área residual de la interpolación α-spline será igual a la potencia de la señal $J(t)$, puesto que consideramos señales periódicas, es decir,

$$\begin{aligned} \mathrm{A}_r\left(s_\alpha(t)\right) &= \frac{1}{MT_0}\int_{MT_0}\left[s_\alpha(t) - L(t)\right]^2 dt \\ &= \mathrm{Var}\left(s_\alpha(t)\right) + \mathrm{Var}\left(L(t)\right) - \frac{2}{MT_0}\int_{MT_0} s_\alpha(t)L(t)dt. \end{aligned} \tag{C.9}$$

En la anterior expresión, los dos primeros términos pueden ser evaluados utilizando (C.4).[1] Respecto al último término, es posible su cálculo utilizando (4.23) y el hecho de que la interpolación lineal no es más que la convolución de las muestras de la señal con la función $\Delta(t) = \beta_\alpha(t)|_{\alpha=0}$. De esta forma, se obtiene

$$\begin{aligned} I_A &= \frac{1}{MT_0}\int_{MT_0} s_\alpha(t)L(t)dt \\ &= \frac{1}{N}\int_N \left(\sum_{n=1}^{N} c[n]\beta_\alpha(t-n)\right)\left(\sum_{n=1}^{N} x[n]\Delta(t-n)\right) dt. \end{aligned}$$

Haciendo uso del soporte compacto de las funciones $\Delta(t)$ y $\beta_\alpha(t)$, y de la periodicidad tanto de la secuencia $x[n]$ como de los coeficientes α-spline $c[n]$,

[1] Dado que la interpolación lineal a tramos es un caso particular de la interpolación α-spline (cuando $\alpha = 0$), su varianza, usando (C.4), resulta $\mathrm{Var}\left(L(t)\right) = A^2(2+\cos\omega_0)/6$.

resulta

$$\begin{aligned} I_A &= \frac{1}{N}\sum_{k=1}^{N} c[k]\left(\sum_{j=-2}^{2} f[k+j]\int_{-\infty}^{\infty}\beta_\alpha\left(x-k\right)\Delta\left(x-k+j\right)dx\right) \\ &= \frac{I_{A0}}{N}\sum_{k=1}^{N} c[k]f[k] + \frac{2I_{A1}}{N}\sum_{k=1}^{N} c[k]f[k+1] + \frac{2I_{A2}}{N}\sum_{k=1}^{N} c[k]f[k+2], \end{aligned} \tag{C.10}$$

donde

$$I_{An} = \int_{-\infty}^{\infty}\beta_\alpha\left(x\right)\Delta\left(x-n\right)dx = I_{A-n}, \tag{C.11}$$

cuyos valores no nulos son

$$\begin{aligned} I_{A0} &= \frac{3\alpha^3 - 10\alpha^2 + 40}{60}, \\ I_{A1} &= \frac{-2\alpha^3 + 5\alpha^2 + 10}{60}, \\ I_{A2} &= \frac{\alpha^3}{120}. \end{aligned}$$

Los coeficientes α-spline de una sinusoide periódica vienen dados por

$$c[n] = x[n]\cdot D_\alpha(z)|_{z=e^{j\omega_0}} = A\sin\left(\omega_0 n + \Phi_0\right)\frac{3}{3+\alpha(\cos\omega_0 - 1)},$$

por lo que los sumatorios de (C.10) toman el valor

$$\begin{aligned} \sum_{n=1}^{N} c[n]x[n+k] &= \frac{3A^2}{3+\alpha(\cos\omega_0-1)}\sum_{n=1}^{N}\sin\left(\omega_0 n+\phi_0\right)\sin\left(\omega_0(n+k)+\phi_0\right) \\ &= \frac{A^2 N}{2}\frac{3\cos\left(\omega_0 k\right)}{3+\alpha(\cos\omega_0-1)}. \end{aligned}$$

Finalmente, (C.10) se reescribe como

$$I_A = \frac{A^2}{2}\frac{3}{3+\alpha(\cos\omega_0-1)}\left(I_{A0} + 2\cos\left(\omega_0\right)I_{A1} + 2\cos\left(2\omega_0\right)I_{A2}\right). \tag{C.12}$$

Por lo tanto, sustituyendo (C.12) y (C.4) en (C.9), el área residual de la interpolación α-spline de una señal senoidal de amplitud A y frecuencia ω_0 toma el valor

$$\begin{aligned} \mathrm{A}_r\left(s_\alpha(t)\right) &= \frac{A^2}{2}\frac{9\left(I_0 + 2\cos\left(\omega_0\right)I_1 + 2\cos\left(2\omega_0\right)I_2 + 2\cos\left(3\omega_0\right)I_3\right)}{\left[3+\alpha(\cos\omega_0-1)\right]^2} \\ &+ \frac{A^2}{2}\frac{2+\cos\omega_0}{3} \\ &- \frac{A^2}{2}\frac{6\left(I_{A0} + 2\cos\left(\omega_0\right)I_{A1} + 2\cos\left(2\omega_0\right)I_{A2}\right)}{3+\alpha(\cos\omega_0-1)}, \end{aligned} \tag{C.13}$$

donde las integrales denotadas por I_n e I_{An} son polinomios en α con valor dado por (C.5) y (C.11), respectivamente.

En definitiva, el área residual de la interpolación α-spline de un señal senoidal de amplitud unitaria depende, exclusivamente, de su frecuencia discreta ω y del valor del parámetro de suavidad α, según la expresión

$$\begin{aligned} \mathrm{A}_r\left(\alpha,\omega\right) = & \frac{9\left(I_0+2\cos\left(\omega\right)I_1+2\cos\left(2\omega\right)I_2+2\cos\left(3\omega\right)I_3\right)}{\left[3+\alpha(\cos\omega-1)\right]^2} \\ & -\frac{6\left(I_{A0}+2\cos\left(\omega\right)I_{A1}+2\cos\left(2\omega\right)I_{A2}\right)}{3+\alpha(\cos\omega-1)}+\frac{2+\cos\omega_0}{3}, \end{aligned} \tag{C.14}$$

Al igual que ocurría con la varianza y la suavidad de la interpolación α-spline, el área residual dada por la anterior expresión es una función monótona creciente del parámetro de suavidad α para cualquier frecuencia, ω, considerada. Únicamente existe una pequeña región próxima a la frecuencia de Nyquist y para valores de α próximos a uno (ver la Figura 4.12) en donde el área residual es prácticamente constante, aunque decreciente con respecto a α.

Si la señal a interpolar fuera una secuencia arbitraria, $x[n]$, el área residual de la interpolación α-spline se podría calcular a partir de su DFT, $X[k]$, y la función (C.14):

$$\mathrm{A}_r\left(s_\alpha(t)\right) = \frac{1}{N^2}\sum_{k=1}^{N-1}\left|X[k]\right|^2\mathrm{A}_r\left(\alpha,\omega=2\pi k/N\right),$$

C.5. Área residual de la interpolación α-spline de señales aleatorias

En el caso de considerar la interpolación α-spline de una señal aleatoria, $\mathbf{X}[n]$, con autocorrelación $R_{\mathbf{X}}[n]$ y densidad espectral de potencia $|S_{\mathbf{X}}(\omega)|^2$, el criterio del área residual se puede establecer como la potencia de la señal aleatoria $\mathbf{J}(t)$ representada en la Figura C.2. De dicha figura podemos concluir que el sistema que calcula el área residual de la interpolación α-spline posee una respuesta al impulso

$$h_J(t) = \eta_\alpha(t) - \Delta\left(t\right),$$

y una respuesta en frecuencia

$$H_J(\Omega) = H_\alpha(\Omega) - H_\Delta\left(\Omega\right) = \mathrm{sinc}^2(\Omega/2\pi)\left(\frac{3\ \mathrm{sinc}^2(\alpha\Omega/2\pi)}{3-\alpha(1-\cos\Omega)}-1\right).$$

Por lo tanto, la autocorrelación de la salida, $R_{\mathbf{J}}(\tau)$, será

$$R_{\mathbf{J}}(\tau) = R_{\mathbf{X}}[n] * \left(h_J(\tau) * h_J(-\tau)\right),$$

y, considerando (C.5) y la simetría de las funciones $\eta_\alpha(x)$ y $\Delta\,(x)$ que intervienen, es posible escribir

$$R_{\mathbf{J}}(\tau) = R_{\mathbf{X}}[n] * \left[\eta_\alpha(\tau) * \eta_\alpha(\tau) + \Delta\,(\tau) * \Delta\,(\tau) - 2\eta_\alpha(\tau) * \Delta\,(\tau)\right]. \qquad \text{(C.15)}$$

Finalmente, particularizando esta última expresión para $\tau = 0$, se obtiene el área residual de la interpolación α-spline.

C.5.1. Señales blancas

Si el proceso a interpolar es blanco y de media nula, es decir, $R_{\mathbf{X}}[n] = \sigma^2_{\mathbf{X}}\delta[n]$, entonces, haciendo uso de (C.15), la potencia del área residual de la interpolación α-spline se puede escribir como

$$\mathrm{A}_r\left(\mathbf{S}_\alpha\right) = \sigma^2_{\mathbf{X}} R_{\mathbf{J}}(0) = \mathrm{Var}\left(\mathbf{S}_\alpha\right) + \mathrm{Var}\left(\mathbf{S}_\alpha\right)\big|_{\alpha=0} - 2\sigma^2_{\mathbf{X}} \int_{-\infty}^{\infty} \eta_\alpha(\tau)\Delta\,(\tau)\, d\tau,$$

donde $\mathrm{Var}\left(\mathbf{X}_\alpha\right)$ es la varianza de la interpolación α-spline y $\mathrm{Var}\left(\mathbf{X}_\alpha\right)\big|_{\alpha=0}$ la varianza de la interpolación lineal; cuyos valores viene dados por (C.8). En cuanto a la integral, ésta se puede calcular como

$$\begin{aligned}\int_{-\infty}^{\infty} \eta_\alpha(t)\Delta\,(t)\, dt &= 2\int_0^1 \left(\sum_{n\in\mathbb{Z}} d_\alpha[n]\beta_\alpha\,(t-n)\right)(1-n)dx \\ &= 2\sum_{n\in\mathbb{Z}} d_\alpha[n]\int_0^1 (1-t)\beta_\alpha\,(t-n)\, dx,\end{aligned}$$

donde, debido al soporte local de la función $\beta_\alpha\,(t)$, la integral únicamente toma valores no nulos para $n = -1, 0, 1, 2$. Tras ciertas operaciones algebraicas se obtiene la potencia del área residual de la interpolación α-spline en función del polo z_1 del filtro directo α-spline:

- Si $0 \le \alpha \le \frac{1}{2}$, entonces

$$\mathrm{A}_r\left(\mathbf{S}_\alpha\right) = 2\sigma^2_{\mathbf{X}} z_1^2 \frac{35z_1^6 - 35z_1^5 - 376z_1^4 + 380z_1^3 + 1985z_1^2 + 895z_1 + 140}{105(z_1-1)^4(z_1+1)^3}.$$

- Si $\frac{1}{2} \le \alpha \le 1$, entonces

$$\begin{aligned}\mathrm{A}_r\left(\mathbf{S}_\alpha\right) = \frac{-\sigma^2_{\mathbf{X}}}{1632960z_1^4(z_1-1)^4(z_1+1)^3}\big[&z_1^{18} + 69z_1^{17} + 2034z_1^{16} \\ &+ 33108z_1^{15} + 319128z_1^{14} + 1784664z_1^{13} + 3832668z_1^{12} \\ &+ 2437620z_1^{11} - 8374158z_1^{10} - 25010630z_1^9 - 31920144z_1^8 \\ &- 23353956z_1^7 - 10406256z_1^6 - 2787792z_1^5 - 550332z_1^4 \\ &-60516z_1^3 - 3867z_1^2 - 135z_1 - 2\big];\end{aligned}$$

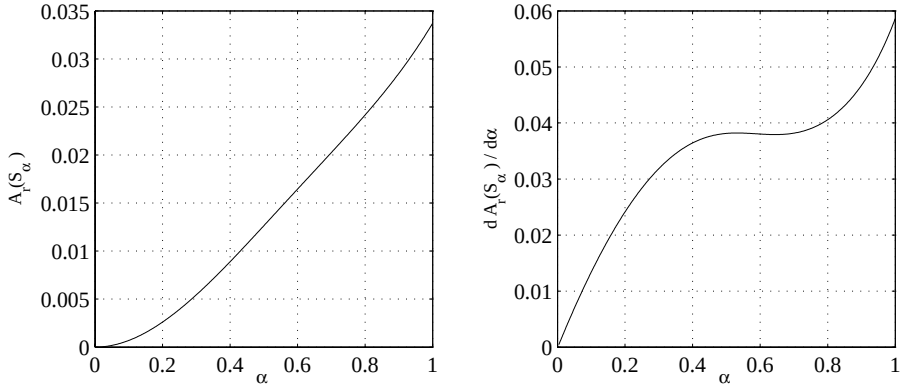


Figura C.3: Área residual de la interpolación α-spline de señales blancas (izquierda) y su derivada respecto a α (derecha).

función que, como se esperaba y se observa en la Figura C.3, decrece monótonamente con z_1, por lo que crece monótonamente con respecto a α, es decir, para un proceso blanco, el área residual de la interpolación α-spline crece conforme aumentamos el parámetro de suavidad α.

C.5.2. Señales aleatorias senoidales

Siguiendo el razonamiento previamente desarrollado para el cálculo de la suavidad y la varianza de la interpolación α-spline de señales aleatorias senoidales, el área residual de un proceso estocástico de frecuencia ω_0 y varianza $\sigma^2_{\mathbf{X}}$ tendrá una expresión similar a (C.13). En concreto:

$$\begin{aligned}
\mathrm{A}_r\left(\mathbf{S}_\alpha\right) = {} & \sigma^2_{\mathbf{X}} \frac{9\left(I_0 + 2\cos\left(\omega_0\right) I_1 + 2\cos\left(2\omega_0\right) I_2 + 2\cos\left(3\omega_0\right) I_3\right)}{\left[3 + \alpha(\cos\omega_0 - 1)\right]^2} \\
& + \sigma^2_F \frac{2 + \cos\omega_0}{3} \\
& - \sigma^2_F \frac{6\left(I_{A0} + 2\cos\left(\omega_0\right) I_{A1} + 2\cos\left(2\omega_0\right) I_{A2}\right)}{3 + \alpha(\cos\omega_0 - 1)}.
\end{aligned}$$

Apéndice D

Expresiones de la base α-spline no uniforme

D.1. Expresiones de la base α-spline no uniforme

Las expresiones de la base α-spline no uniforme, $\beta_\alpha(t)$, correspondiente al vector no decreciente de nodos $[t_0, t_1, t_2, t_3, t_4]$ y considerando el vector distancia entre nodos $[h_1, h_2, h_3, h_4]$, se obtienen integrando tres veces (4.68) e introduciendo los valores de (4.70).

En el caso de que $0 \leq \alpha \leq 0{,}5$:

- Si $t_1 - \alpha h_1 < t < t_1$,

$$\beta_\alpha(t) = \frac{(t - t_1 + \alpha h_1)^3}{\alpha^2 h_1 (h_1 + h_2)(\alpha h_1 + (3 - 2\alpha) h_2 + \alpha h_3)}.$$

- Si $t = t_1$,

$$\beta_\alpha(t) = \frac{\alpha h_1^2}{(h_1 + h_2)(\alpha h_1 + (3 - 2\alpha) h_2 + \alpha h_3)}.$$

- Si $t_1 < t < t_1 + \alpha h_2$,

$$\beta_\alpha(t) = \frac{\alpha^3 h_1^2 h_2 + 3\alpha^2 h_1 h_2 (t - t_1) + 3\alpha h_2 (t - t_1)^2 - (t - t_1)^3}{\alpha^2 h_2 (h_1 + h_2)(\alpha h_1 + (3 - 2\alpha) h_2 + \alpha h_3)}.$$

- Si $t = t_1 + \alpha h_2$,

$$\beta_\alpha(t) = \frac{\alpha h_1 + 2\alpha h_2}{(\alpha h_1 + (3-2\alpha)h_2 + \alpha h_3)}.$$

- Si $t_1 + \alpha h_2 < t < t_2 - \alpha h_2$,

$$\beta_\alpha(t) = \frac{\alpha h_1 + 2\alpha h_2 + 3(t - t_1 - \alpha h_2)}{(\alpha h_1 + (3-2\alpha)h_2 + \alpha h_3)}.$$

- Si $t = t_2 - \alpha h_2$,

$$\beta_\alpha(t) = \frac{\alpha h_1 + (3-4\alpha)h_2}{(\alpha h_1 + (3-2\alpha)h_2 + \alpha h_3)}.$$

- Si $t_2 - \alpha h_2 < t < t_2$,

$$\begin{aligned}\beta_\alpha(t) =& \frac{\alpha h_1 + 2\alpha h_2 + 3(t - t_1 - \alpha h_2)}{(\alpha h_1 + (3-2\alpha)h_2 + \alpha h_3)} \\ &- \frac{[\alpha h_1 + (3-\alpha)h_2 + (3-\alpha)h_3 + \alpha h_4](t - t_2 + \alpha h_2)^3}{\alpha^2 h_2(h_2+h_3)(\alpha h_1 + (3-2\alpha)h_2 + \alpha h_3)(\alpha h_2 + (3-2\alpha)h_3 + \alpha h_4)}.\end{aligned}$$

- Si $t = t_2$,

$$\begin{aligned}\beta_\alpha(t) =& \frac{\alpha h_1 + (3-\alpha)h_2}{(\alpha h_1 + (3-2\alpha)h_2 + \alpha h_3)} \\ &- \frac{\alpha h_1 + (3-\alpha)h_2 + (3-\alpha)h_3 + \alpha h_4}{(h_2+h_3)(\alpha h_1 + (3-2\alpha)h_2 + \alpha h_3)(\alpha h_2 + (3-2\alpha)h_3 + \alpha h_4)}\alpha h_2^2.\end{aligned}$$

- Si $t_2 < t < t_2 + \alpha h_3$,

$$\begin{aligned}\beta_\alpha(t) =& \frac{\alpha h_4 + 2\alpha h_3 + 3(-t + t_3 - \alpha h_3)}{(\alpha h_4 + (3-2\alpha)h_3 + \alpha h_2)} \\ &- \frac{[\alpha h_4 + (3-\alpha)h_3 + (3-\alpha)h_2 + \alpha h_1](-t + t_2 + \alpha h_3)^3}{\alpha^2 h_3(h_3+h_2)(\alpha h_4 + (3-2\alpha)h_3 + \alpha h_2)(\alpha h_3 + (3-2\alpha)h_2 + \alpha h_1)}.\end{aligned}$$

- Si $t = t_2 + \alpha h_3$,

$$\beta_\alpha(t) = \frac{\alpha h_4 + (3-4\alpha)h_3}{(\alpha h_4 + (3-2\alpha)h_3 + \alpha h_2)}.$$

- Si $t_2 + \alpha h_3 < t < t_3 - \alpha h_3$,

$$\beta_\alpha(t) = \frac{\alpha h_4 + 2\alpha h_3 + 3(-t + t_3 - \alpha h_3)}{(\alpha h_4 + (3-2\alpha)h_3 + \alpha h_2)}.$$

- Si $t = t_3 - \alpha h_3$,

$$\beta_\alpha(t) = \frac{\alpha h_4 + 2h_3}{(\alpha h_4 + (3-2\alpha)h_3 + \alpha h_2)}.$$

- Si $t_3 - \alpha h_3 < t < t_3$,

$$\beta_\alpha(t) = \frac{\alpha^3 h_4^2 h_3 + 3\alpha^2 h_4 h_3(-t+t_3) + 3\alpha(-t+t_3)^2 - (-t+t_3)^3}{\alpha^2 h_3(h_4+h_3)(\alpha h_4 + (3-2\alpha)h_3 + \alpha h_2)}.$$

- Si $t = t_3$,

$$\beta_\alpha(t) = \frac{\alpha h_4^2}{(h_4+h_3)(\alpha h_4 + (3-2\alpha)h_3 + \alpha h_2)}.$$

- Si $t_3 < t < t3 + \alpha h_4$,

$$\beta_\alpha(t) = \frac{(-t+t_3+\alpha h_4)^3}{\alpha^2 h_4(h_4+h_3)(\alpha h_4 + (3-2\alpha)h_3 + \alpha h_2)}.$$

- Si $t \leq t_1 - \alpha h_1$ ó $t \geq t_3 + \alpha h_4$,

$$\beta_\alpha(t) = 0.$$

Si se considera $0{,}5 \leq \alpha \leq 1$:

- Si $t_1 - \alpha h_1 < t < t_1$,

$$\beta_\alpha(t) = \frac{(t-t_1+\alpha h_1)^3}{\alpha^2 h_1(h_1+h_2)(\alpha h_1 + (3-2\alpha)h_2 + \alpha h_3)}.$$

- Si $t = t_1$,

$$\beta_\alpha(t) = \frac{\alpha h_1^2}{(h_1+h_2)(\alpha h_1 + (3-2\alpha)h_2 + \alpha h_3)}.$$

- Si $t_1 < t < t_2 - \alpha h_2$,

$$\beta_\alpha(t) = \frac{\alpha^3 h_1^2 h_2 + 3\alpha^2 h_1 h_2(t-t_1) + 3\alpha h_2(t-t_1)^2 - (t-t_1)^3}{\alpha^2 h_2(h_1+h_2)(\alpha h_1 + (3-2\alpha)h_2 + \alpha h_3)}.$$

- Si $t = t_2 - \alpha h_2$,

$$\beta_\alpha(t) = \frac{\alpha^3 h_1^2 + (1-\alpha)h_2(3\alpha^2 h_1 + (1-\alpha)(4\alpha-1)h_2)}{\alpha^2(h_1+h_2)(\alpha h_1 + (3-2\alpha)h_2 + \alpha h_3)}.$$

- Si $t_2 - \alpha h_2 < t < t_1 + \alpha h_2$,

$$\begin{aligned}\beta_\alpha(t) =& \frac{\alpha^3 h_1^2 h_2 + (1-\alpha)h_2^2(3\alpha^2 h_1 + (1-\alpha)(4\alpha-1)h_2)}{\alpha^2 h_2(h_1+h_2)(\alpha h_1 + (3-2\alpha)h_2 + \alpha h_3)} \\ &+ \frac{3h_2(\alpha^2 h_1 + (1-\alpha)(3\alpha-1)h_2)(t - t_2 + \alpha h_2)}{\alpha^2 h_2(h_1+h_2)(\alpha h_1 + (3-2\alpha)h_2 + \alpha h_3)} \\ &+ \frac{3h_2(2\alpha-1)(t-t_2+\alpha h_2)^2 - (t-t_2+\alpha h_2)^3}{\alpha^2 h_2(h_1+h_2)(\alpha h_1 + (3-2\alpha)h_2 + \alpha h_3)} \\ &- \frac{[\alpha h_1 + (3-\alpha)h_2 + (3-\alpha)h_3 + \alpha h_4](t-t_2+\alpha h_2)^3}{\alpha^2 h_2(h_2+h_3)(\alpha h_1 + (3-2\alpha)h_2 + \alpha h_3)(\alpha h_2 + (3-2\alpha)h_3 + \alpha h_4)}.\end{aligned}$$

- Si $t = t_1 + \alpha h_2$,

$$\begin{aligned}\beta_\alpha(t) =& \frac{\alpha h_1 + 2\alpha h_2}{(\alpha h_1 + (3-2\alpha)h_2 + \alpha h_3)} \\ &+ \frac{[\alpha h_1 + (3-\alpha)h_2 + (3-\alpha)h_3 + \alpha h_4](1-2\alpha)^3 h_2^2}{\alpha^2(h_2+h_3)(\alpha h_1 + (3-2\alpha)h_2 + \alpha h_3)(\alpha h_2 + (3-2\alpha)h_3 + \alpha h_4)}.\end{aligned}$$

- Si $t_1 + \alpha h_2 < t < t_2$,

$$\begin{aligned}\beta_\alpha(t) =& \frac{\alpha h_1 + 2\alpha h_2 + 3(t - t_1 - \alpha h_2)}{(\alpha h_1 + (3-2\alpha)h_2 + \alpha h_3)} \\ &- \frac{[\alpha h_1 + (3-\alpha)h_2 + (3-\alpha)h_3 + \alpha h_4](t-t_2+\alpha h_2)^3}{\alpha^2 h_2(h_2+h_3)(\alpha h_1 + (3-2\alpha)h_2 + \alpha h_3)(\alpha h_2 + (3-2\alpha)h_3 + \alpha h_4)}.\end{aligned}$$

- Si $t = t_2$,

$$\begin{aligned}\beta_\alpha(t) =& \frac{\alpha h_1 + (3-\alpha)h_2}{(\alpha h_1 + (3-2\alpha)h_2 + \alpha h_3)} \\ &- \frac{[\alpha h_1 + (3-\alpha)h_2 + (3-\alpha)h_3 + \alpha h_4]\alpha h_2^2}{(h_2+h_3)(\alpha h_1 + (3-2\alpha)h_2 + \alpha h_3)(\alpha h_2 + (3-2\alpha)h_3 + \alpha h_4)}.\end{aligned}$$

- Si $t_2 < t < t_3 - \alpha h_3$,

$$\begin{aligned}\beta_\alpha(t) =& \frac{\alpha h_4 + 2\alpha h_3 + 3(-t + t_3 - \alpha h_3)}{(\alpha h_4 + (3-2\alpha)h_3 + \alpha h_2)} \\ &- \frac{[\alpha h_4 + (3-\alpha)h_3 + (3-\alpha)h_2 + \alpha h_1](-t+t_2+\alpha h_3)^3}{\alpha^2 h_3(h_3+h_2)(\alpha h_4 + (3-2\alpha)h_3 + \alpha h_2)(\alpha h_3 + (3-2\alpha)h_2 + \alpha h_1)}.\end{aligned}$$

• Si $t = t_3 - \alpha h_3$,

$$\begin{aligned}\beta_\alpha(t) =& \frac{\alpha h_4 + 2\alpha h_3}{(\alpha h_4 + (3-2\alpha)h_3 + \alpha h_2)} \\ &+ \frac{[\alpha h_4 + (3-\alpha)h_3 + (3-\alpha)h_2 + \alpha h_1](1-2\alpha)^3 h_3^2}{\alpha^2(h_3+h_2)(\alpha h_4 + (3-2\alpha)h_3 + \alpha h_2)(\alpha h_3 + (3-2\alpha)h_2 + \alpha h_1)}.\end{aligned}$$

• Si $t_3 - \alpha h_3 < t < t_2 + \alpha h_3$,

$$\begin{aligned}\beta_\alpha(t) =& \frac{\alpha^3 h_4^2 h_3 + (1-\alpha)h_3^2(3\alpha^2 h_4 + (1-\alpha)(4\alpha-1)h_3)}{\alpha^2(h_4+h_3)(\alpha h_4 + (3-2\alpha)h_3 + \alpha h_4)} \\ &+ \frac{3h_3(\alpha^2 h_4 + (1-\alpha)(3\alpha-1)h_3)(t-t_2-\alpha h_3)}{\alpha^2(h_4+h_3)(\alpha h_4 + (3-2\alpha)h_3 + \alpha h_4)} \\ &+ \frac{3h_3(2\alpha-1)(t-t_2-\alpha h_3)^2 - (t-t_2-\alpha h_3)^3}{\alpha^2(h_4+h_3)(\alpha h_4 + (3-2\alpha)h_3 + \alpha h_4)} \\ &- \frac{[\alpha h_4 + (3-\alpha)h_3 + (3-\alpha)h_2 + \alpha h_1](t-t_2+\alpha h_3)^3}{\alpha^2 h_3(h_3+h_2)(\alpha h_4 + (3-2\alpha)h_3 + \alpha h_2)(\alpha h_3 + (3-2\alpha)h_2 + \alpha h_1)}.\end{aligned}$$

• Si $t = t_2 + \alpha h_3$,

$$\beta_\alpha(t) = \frac{\alpha^3 h_4^2 + (1-\alpha)h_3(3\alpha^2 h_4 + (1-\alpha)(4\alpha-1)h_3)}{\alpha^2(h_4+h_3)(\alpha h_4 + (3-2\alpha)h_3 + \alpha h_2)}.$$

• Si $t_2 + \alpha h_3 < t < t_3$,

$$\beta_\alpha(t) = \frac{\alpha^3 h_4^2 h_3 + 3\alpha^2 h_4 h_3(-t+t_3) + 3\alpha(-t+t_3)^2 - (-t+t_3)^3}{\alpha^2 h_3(h_4+h_3)(\alpha h_4 + (3-2\alpha)h_3 + \alpha h_2)}.$$

• Si $t = t_3$,

$$\beta_\alpha(t) = \frac{\alpha h_4^2}{(h_4+h_3)(\alpha h_4 + (3-2\alpha)h_3 + \alpha h_2)}.$$

• Si $t3 < t < t_3 + \alpha h_4$,

$$\beta_\alpha(t) = \frac{(-t+t_3+\alpha h_4)^3}{\alpha^2 h_4(h_4+h_3)(\alpha h_4 + (3-2\alpha)h_3 + \alpha h_2)}.$$

• Si $t \leq t_1 - \alpha h_1$ ó $t \geq t_3 + \alpha h_4$,

$$\beta_\alpha(t) = 0.$$

Apéndice E

Modelado α-spline

E.1. Comportamiento del término de regularización

En este apéndice se demuestra que el término de regularización (5.14), correspondiente a dos bases del modelo αPWL separadas una distancia $b \geq 0$, tiene un comportamiento monótono decreciente con respecto al parámetro de suavidad $\alpha \geq 0$ para cualquier valor de b y de las constantes c_1 y c_2. Derivando la expresión (5.14) con respecto a α se obtiene

$$\frac{d||P_s^2 f||^2}{d\alpha} = \begin{cases} \frac{-8}{3\alpha^2}(c_1^2 + c_2^2), & \alpha \leq b/2; \\ \frac{-8}{3\alpha^2}\left(c_1^2 + c_2^2 + 2c_1c_2\frac{(2\alpha-b)^2}{\alpha^2}\left(\frac{2\alpha-b}{\alpha} - \frac{3}{2}\right)\right), & b/2 \leq \alpha \leq b; \\ \frac{-8}{3\alpha^2}\left(c_1^2 + c_2^2 + 2c_1c_2\frac{2(3b^3+4\alpha^3-6b^2\alpha)}{3\alpha^4}\right), & \alpha \geq b. \end{cases} \tag{E.1}$$

Es evidente que la anterior expresión es negativa para $\alpha \leq b/2$. En el caso de que $b/2 \leq \alpha \leq b$, se puede reescribir (E.1) como

$$\frac{d||P_s^2 f||^2}{d\alpha} = \frac{-8}{3\alpha^2}\left[c_1^2 + c_2^2 + 2c_1c_2\theta_1\left(\frac{2\alpha - b}{\alpha}\right)\right], \tag{E.2}$$

siendo $\theta_1(x) = x^2(x - 3/2)$. Recordando que el parámetro de suavidad siempre es positivo y que, en este caso, $b/2 \leq \alpha \leq b$, entonces $0 \leq \frac{2\alpha-b}{\alpha} \leq 1$, por lo que

$$\left|\theta_1\left(\frac{2\alpha - b}{\alpha}\right)\right| \leq 1;$$

resultando que (E.2) es no positiva, dado que siempre se cumple que $c_1^2 + c_2^2 + 2kc_1c_2 \geq 0$ si $|k| \geq 0$.

De igual forma, cuando $\alpha \geq b$, es posible expresar (E.1) según

$$\frac{d||P_s^2 f||^2}{d\alpha} = \frac{-8}{3\alpha^2}\left[c_1^2 + c_2^2 + 2c_1c_2\theta_2\left(\frac{b}{\alpha}\right)\right], \tag{E.3}$$

siendo $\theta_2(x) = (6x^3 - 9x^2 + 2)/2$. Puesto que $\alpha \geq 0$, $b \geq 0$ y $\alpha \geq b$, entonces $0 \leq \frac{b}{\alpha} \leq 1$. Finalmente, se demuestra que

$$\left|\theta_2\left(\frac{b}{\alpha}\right)\right| \leq 1,$$

por lo que, haciendo uso de la desigualdad triangular, se obtiene que (E.3) es mayor o igual que cero.

En definitiva, (E.1) es una función no positiva para cualquier valor de α, b, c_1 y c_2 considerado, hecho éste que demuestra que el término de regularización propuesto en (5.14) decrece monótonamente con el parámetro de suavidad α.

Apéndice F

Publicaciones a las que ha dado lugar la realización de la Tesis

En este apéndice se enumeran las publicaciones a las que ha dado lugar la realización de la presente Tesis, tanto en revistas como en congresos nacionales e internacionales.

F.1. Revistas

- J. M. López-Higuera, I. Santamaría, P. Mottier, A. Cobo, C. Pantaleón, E. Ollier, M. Morante, J. Ibáñez, C. Chabrol y F. J. Madruga. "Optical sensors and their fusion in a quasi-smart structure for real time vibration monitoring and predictive maintenance of large power electric generators". *Journal of Intelligent Material Systems and Structures*, tomo 9, nº 11, páginas 938–946, noviembre 1999.

- D. Luengo, C. J. Pantaleón, J. Ibáñez y I. Santamaría. "Design of simultaneous sampling systems based on fractional delay Lagrange filters". *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, 2000.

- J. Ibáñez, C. Pantaleón, I. Santamaría, T. Fernández y D. Martínez. "Nonlinearity estimation in power amplifiers based on subsampled temporal data". *IEEE Transactions on Instrumentation and Measurement*, tomo 50, nº 4, páginas 882–887, agosto 2001.

- I. Santamaría, J. Ibañez, M. Lázaro, C. Pantaleón y L. Vielva. "Modeling nonlinear power amplifiers in OFDM systems from subsampled data: a comparative study using real measurements". *EURASIP Journal on Applied Signal Processing*, , nº 12, páginas 1219–1228, 2003.

- I. Santamaría, J. Ibañez, C. Pantaleon y F. Cruz-Roldan. "Positive semidefinite biorthogonal partners for pulse shaping". *Remitido para su publicación en Signal Processing*, 2004.

F.2. Congresos internacionales

- D. Luengo, C. Pantaleón, I. Santamaría, J. Ibáñez y E. Gómez-Cossío. "Simultaneous sampling by digital phase correction". En *Proc. Instrumentation and Measurement Technology Conference*, páginas 980–984. Otawa, mayo 1997.

- J. López-Higuera, I. Santamaría, A. Cobo, C. Pantaleón, M. Morante, J. Ibáñez y J. Echevarría. "Quasi-smart optical fibre sensor system for real time and predictive monitoring on large rotating machinery". En *Proc. CLEO/Europe*, páginas 858–862. 1998.

- A. Cobo, D. Luengo, J. Ibáñez, C. López, F. Madruga y C. Pantaleón. "A novel simultaneous sampling technique and its application to multipoint optical fiber sensor accelerometers". En *Proc. International Conference on Signal Processing Applications and Technology*, tomo 2, páginas 858–862. Toronto, septiembre 1998.

- J. Ibáñez, C. Pantaleón, I. Santamaría, D. Luengo y M. Lázaro. "Specific DSP based monitoring system for hydro-generator sets". En *Proc. International Conference on Signal Processing Applications and Technology*, tomo 2, páginas 932–936. Toronto, septiembre 1998.

- J. Ibáñez, C. Pantaleón, I. Santamaría, T. Fernández y D. Martínez. "Nonlinearity estimation in power amplifiers based on undersampled temporal data". En *Proc. Instrumentation and Measurement Technology Conference*, tomo 2, páginas 1266–1271. Baltimore, mayo 2000.

- J. Ibáñez, C. Pantaleón, J. Diez, I. Santamaría y L. Vielva. "Spline pulse shaping with ISI-Free and matched filter receiver". En *Proc. European Signal Processing Conference*, páginas 1–4. Toulouse, France, 2002.

- J. Ibáñez, I. Santamaría, C. Pantaleón y L. Vielva. "Parametric smoothing of spline interpolation". En *Proc. International Conference on Acoustics, Speech, and Signal Processing*. Montreal, mayo 2004.

F.3. Congresos nacionales

- D. Luengo, J. Ibáñez, C. Pantaleón y I. Santamaría. "Análisis y diseño de sistemas de retardo fraccional basados en sobremuestreo e interpoladores FIR de orden bajo". En *Actas del Simposium Nacional de la URSI*, tomo 1, páginas 721–724. Bilbao, septiembre 1997.

- M. Lázaro, J. A. García, J. Ibáñez, C. Pantaleón, I. Santamaría. "Modelado no lineal de dispositivos mediante funciones lineales a tramos suavizadas". En *Actas del Simposium Nacional de la URSI*, tomo 2, páginas 217–220. Bilbao, septiembre 1997.

- J. Diez, J. Ibáñez, C. Pantaleón, I. Santamaría. "Conformación óptima de señales de comunicaciones con splines cúbicos". En *Actas del Simposium Nacional de la URSI*. Madrid, septiembre 2001.

Apéndice G

Símbolos y abreviaturas

En este apéndice se describe la simbología utilizada en la Tesis y se enumeran las abreviaturas empleadas.

G.1. Simbología

A lo largo de esta Tesis se utiliza, en la medida de lo posible, la nomenclatura propia del procesado de señal. Así, las señales (o funciones) unidimensionales de variable independiente continua se denotan con dicha variable entre paréntesis, como por ejemplo

$$x(t), \text{ o bien } f(x);$$

mientras que su transformada de Fourier se escribe como

$$X(\Omega) = \mathrm{TF}\left\{x(t)\right\},$$

siendo $\Omega = 2\pi f$ la frecuencia continua en radianes por segundo.

Por el contratrio, las señales (o funciones) de variable independiente discreta se denotan utilizando corchetes, es decir,

$$x[n],$$

siendo su transformada de Fourier

$$X(\omega) = \mathrm{TF}\left\{x[n]\right\},$$

donde ω es la frecuencia discreta.

Respecto a las derivadas (diferencias en el caso discreto), y siempre que no provoque confusión, se emplea el siguiente formato

$$x^{(1)}(t), \text{ o, en el caso discreto, } x^{(1)}[n].$$

En el caso multidimensional, la derivada parcial con respecto a cierta variable utiliza la nomenclatura

$$\frac{\partial f(x)}{\partial x_1}.$$

Para denotar vectores se utilizan caracteres en minúscula y negrita (por ejemplo, $\mathbf{x}$), letras mayúsculas en negrita ($\mathbf{A}$) para las matrices y letras griegas minúsculas para los parámetros (como α o γ). Los espacios vectoriales se denotan utilizando letras mayúsculas en cursiva (X) y, en ocasiones, con la base generadora entre paréntesis, como en $V(\beta_\alpha)$. En dichos espacios, el producto interno se representa como $\langle x, y\rangle$ y la norma como $\|x\|$.

Las variables aleatorias se denotan empleando letras griegas mayúsculas en negrita (como $\mathbf{\Theta}$), y su esperanza matemática como $\mathrm{E}[\mathbf{\Theta}]$. Respecto a los procesos estocásticos, se utilizan letras mayúsculas en negrita seguidas de la variable independiente entre paréntesis o corchetes (por ejemplo, $\mathbf{X}(t)$ y $\mathbf{X}[n]$).

G.2. Abreviaturas

ACPR: Relación de potencia del canal adyacente (*Adjacent Channel Power Ratio*).

ADC: Conversor analógico-digital (*Analog to Digital Converter*).

AIC: Criterio de información de Akaike (*Akaike's Information Criterion*).

AWGN: Ruido aditivo blanco Gaussiano (*Additive White Gaussian Noise*).

BP: Pareja biortogonal (*Biortogonal Partner*).

CPWL: Canónico lineal a tramos (CPWL: *Canonical Piecewise-Linear*).

DAC: Conversor digital-analógico (*Digital to Analog Converter*).

EPM: Modelo polinómico de envolvente (*Envelope Polynomial Model*).

FDF: Filtro de retardo fraccionario (*Fractional Delay Filter*).

FET: Transistor de efecto de campo (*Field Effect Transistor*).

FIR: Respuesta al impulso finita (*Finite Impulse Response*).

GPIB: Bus de instrumentación IEEE488.1 (*General Purpose Interface Bus*).

GRBF: Función de base radial generalizada (*Generalized Radial Basis Function*).

HEMT: Transistor con gran movilidad electrónica (*High Electron Mobility Transistor*).

HPA: Amplificador de potencia (HPA: *High Power Amplifier*).

IIR: Respuesta al impulso infinita (*Infinite Impulse Response*).

ISI: Interferencia entre símbolos (*Intersymbol Interference*).

LDA: Anemometría Laser-Doppler (*Laser-Doppler Anemometry*).

LMI: Desigualdad lineal matricial (*Linear Matrix Inequality*).

LSB: Bit menos significativo (*Least Significant Bit*).

MAP: Máximo *a posteriori* (*Maximum A Posteriori*).

MARS: Regresión adaptativa multivariable mediante splines (*Multivariate Adaptative Regression Splines*).

MDL: Mínima longitud de descripción (*Minimum Description Length*).

MLP: Perceptrón multicapa (*Multilayer Perceptron*).

MMIC: Circuito monolítico integrado de microondas (*Microwave Monolithic Integrated Circuit*).

MOMS: Interpolación de máximo orden y mínimo soporte (*Maximal Order Interpolation of Minimal Support*).

NN: Red neuronal (*Neural Network*).

NNR: Regresión por vecino más cercano (*Nearest Neighbors Regression*).

NRZ: Código de línea sin retorno a cero (*Non-Return to Zero*).

OBO: Nivel de potencia de salida por debajo de la saturación (*Output Back-Off*).

PAM: Modulación por amplitud de pulso (*Pulse Amplitude Modulation*).

PAPR: Relación potencia de pico a potencia media (*Peak-to-Average Power Ratio*).

PPR: Regresión por búsqueda de proyección (*Projection Pursuit Regression*).

PSD-BP: Pareja biortogonal semidefinida positivamente (*Positive Semidefinite Biorthogonal Partner*).

PWL: Lineal a tramos (*Piecewise Linear*).

QMI: Información mutua cuadrática (*Quadratic Mutual Information*).

ROM: Memoria de sólo lectura (*Read Only Memory*).

S/H: Muestreo y retención (*Sample and Hold*).

SER: Relación señal a error (*Signal to Error Ratio*).

SNR: Relación señal a ruido (*Signal to Noise Ratio*).

SDP: Programación semidefinida convexa (*Semidefinite Programming*)

SPWL: Lineal a tramos suavizado (*Smoothed Piecewise-Linear*).

SQRC: Raíz cuadrada de coseno alzado (*Square root Raised-Cosine*).

SQRCS: Raíz cuadrada de spline cúbico (*Square Root Cubic Spline*).

SSPA: Amplificador de potencia de estado sólido (*Solid State Power Amplifier*).

SVM: Máquinas de vectores soporte (*Support Vector Machine*).

TWT: Tubo de onda progresiva (*Traveling-Wave Tube*).

UWB: Banda ultra ancha (*Ultrawide Band*).

WLAN: Red de área local inalámbrica (*Wireless Local Aerea Network*).

Bibliografía

[Abuelma'atti84] M. T. Abuelma'atti. "Frequency-dependent nonlinear quadrature model for TWT amplifiers". *IEEE Transactions on Communications*, tomo 32, nº 8, páginas 982–986, 1984.

[Ahlberg67] J. H. Ahlberg, E. N. Nilson y J. L. Walsh. *The theory of splines and their applications.* Academic Press, Orlando, 1967.

[Aizerman64] M. Aizerman, E. Braverman y L. Rozonoer. "Theoretical foundations of the potential function method in pattern recognition learning". *Automation and Remote Control*, tomo 25, páginas 821–837, 1964.

[Akaike74] H. Akaike. "A new look at the statistical model identification". *IEEE Trans. Automat. Contr.*, tomo 19, páginas 716–723, 1974.

[Alagha00] N. S. Alagha y P. Kabal. "Generalized raised-cosine filters". *IEEE Transactions on Communications*, tomo 47, páginas 989–997, julio 2000.

[Aldroubi94] A. Aldroubi y M. Unser. "Sampling procedures in function spaces and asymptotic equivalence with Shannon's sampling theory". *Numerical Functional Analysis and Optimization*, tomo 15, páginas 1–21, febrero 1994.

[Anderson74] B. D. O. Anderson, K. L. Hitz y N. D. Dien. "Recursive algorithm for spectral factorization". *IEEE Transactions on Circuits and Systems*, tomo 21, páginas 742–750, noviembre 1974.

[Bacon71] D. W. Bacon y D. G. Watts. "Estimating the transition between two intersecting straight lines". *Biometrika*, tomo 58, páginas 525–534, 1971.

[Banelli00] P. Banelli y S. Cocopardi. "Theoretical analysis and performance of OFDM signals in nonlinear AWGN channels". *IEEE Transactions on Communications*, tomo 48, nº 3, páginas 430–441, 2000.

[Barwicz90] A. Barwicz, D. Bellemere y B. Z. Morawski. "Digital correction of A/D conversion error due to multiplexing delay". *IEEE Transactions on Instrumentation and Measurement*, tomo 39, páginas 76–79, febrero 1990.

[Blu99a] T. Blu y M. Unser. "Quantitative Fourier analysis of approximation techniques: Part I interpolators and proyectors". *IEEE Transactions on Signal Processing*, tomo 47, páginas 2783–2795, octubre 1999.

[Blu99b] T. Blu y M. Unser. "Quantitative Fourier analysis of approximation techniques: Part I wavelets". *IEEE Transactions on Signal Processing*, tomo 47, páginas 2796–2806, octubre 1999.

[Blu04] T. Blu, P. Thévenaz y M. Unser. "High-quality causal interpolation for online unidimensional signal processing". En *Proc. European Signal Processing Conference*. Viena, Austria, septiembre 2004.

[Boor68] C. de Boor y J. R. Rice. "Least-squares cubic spline approximation II: variable knots". CSD TR21, Purdue University, Lafayette, 1968.

[Boor78] C. de Boor. *A practical guide to splines*. Springer-Verlag, New York, 1978.

[Borich98] V. Borich, J. H. Jong, J. East y W. E. Stark. "Nonlinear effects of power amplification on multicarrier spread spectrum systems". En *Proc. IEEE MTT-S International Microwave Symposium Digest*, páginas 323–326. Baltimore, Md, junio 1998.

[Calandra94] E. F. Calandra y G. Sirna. "CAD-oriented modelling of the optically-controled GaAs MESFET". En *Proc. GAAS'94 Symp.*, páginas 401–404. 1994.

[Candy92] J. Candy y G. Temes. "Oversampling methods for A/D and D/A conversion". En *Oversampling Delta-Sigma data converters*, páginas 1–25. IEEE Press, 1992.

[Cheney82] E. W. Cheney. *Introduction to approximation theory*. Chelsea Publishing Company, New York, segunda edición, 1982.

[Cherkassky96] V. Cherkassky, D. Gehring y F. Mulier. "Comparison of adaptative methods for function estimation from samples". *IEEE Transactions on Neural Networks*, tomo 7, nº 4, páginas 969–984, julio 1996.

[Chua86] L. O. Chua y A. C. Deng. "Canonical piecewise-linear modeling". *IEEE Transactions on Circuits and Systems*, tomo 33, nº 5, páginas 511–525, mayo 1986.

[Chua88] L. O. Chua y A. C. Deng. "Canonical piecewise-linear representation". *IEEE Transactions on Circuits and Systems*, tomo 35, páginas 101–111, 1988.

[Clark88] C. J. Clark, G. Chrisikos, M. S. Muha, A. A. Moulthrop y C. P. Silva. "Time-domain envelope measurement technique with application to wideband power amplifier modeling". *IEEE Transactions on Microwave Theory and Technology*, tomo 46, nº 12, páginas 1531–1540, diciembre 1988.

[Cobo98] A. Cobo, D. Luengo, J. Ibáñez, C. López, F. Madruga y C. Pantaleón. "A novel simultaneous sampling technique and its application to multipoint optical fiber sensor accelerometers". En *Proc. International Conference on Signal Processing Applications and Technology*, tomo 2, páginas 858–862. Toronto, septiembre 1998.

[Cortés95] C. Cortés y V. N. Vapnik. "Support-Vector networks". *Machine Learning*, tomo 20, páginas 273–297, 1995.

[Crochiere83] R. E. Crochiere y L. R. Rabiner. *Multirate digital signal processing*. Prentice Hall, New York, 1983.

[D'Andrea01] A. N. D'Andrea, V. Lottici y R. Reggiannini. "Nonlinear predistortion of OFDM signals over frequency-selective fading channels". *IEEE Transactions on Communications*, tomo 49, nº 5, páginas 837–843, 2001.

[Dardari00] D. Dardari, V. Tralli y A. Vaccari. "A theoretical characterization of nonlinear distortion effects in OFDM systems". *IEEE Transactions on Communications*, tomo 48, nº 10, páginas 1755–1764, 2000.

[Davidson00] T. N. Davidson, Z. Luo y J. M. Wong. "Design of orthogonal pulse shapes for communications via semidefinite programming". *IEEE Transactions on Signal Processing*, tomo 48, nº 5, páginas 1433–1444, mayo 2000.

[Davidson01] T. N. Davidson. "Efficient design of waveforms for robust pulse amplitude modulation". *IEEE Transactions on Signal Processing*, tomo 49, nº 12, páginas 3098–3111, diciembre 2001.

[Davidson02] T. N. Davidson, Z. Luo y J. Sturm. "Linear matrix inequality formulation of spectral mask constraints with application to FIR filter design". *IEEE Transactions on Signal Processing*, tomo 50, nº 11, páginas 2702–2715, noviembre 2002.

[Davis75] P. J. Davis. *Interpolation and approximation*. Dover Publications, New York, segunda edición, 1975.

[Dierckx95] P. Dierckx. *Curve and sourface fitting with splines*. Oxford University Press, Oxford, 1995.

[Djuric94] P. M. Djuric. "A MAP solution to off-line segmentation of signals". En *Proc. International Conference on Acoustics, Speech, and Signal Processing*, tomo 4, páginas 505–508. Adelaide, abril 1994.

[Erup93] L. Erup, F. M. Gardner y R. A. Garris. "Interpolation in digital modems, part II: Implementation and performance". *IEEE Transactions on Communications*, tomo 41, páginas 981–1008, 1993.

[Ferrar27] W. L. Ferrar. "On the consistency of cardinal function interpolation". *Proc. Roy. Soc. Edinburgh*, tomo 47, páginas 230–242, 1927.

[Ferrari99] L. A. Ferrari, J. H. Park, A. Healey y S. Leeman. "Interpolation using a fast spline transform (FST)". *IEEE Transactions on Circuits and Systems*, tomo 46, nº 8, páginas 891–906, agosto 1999.

[Fraser86] A. M. Fraser y H. L. Swinney. "Independent coordinates for strange attractors from mutual information". *Physical Review A*, tomo 33, nº 2, páginas 1134–1140, 1986.

[Friedman91] J. H. Friedman. "Multivariate adaptative regression splines". *The Annal of Statistics*, tomo 19, nº 1, páginas 1–141, 1991.

[Gabor46] D. Gabor. "Theory of communication". *J. Inst. Elect. Eng. London*, tomo 93–3, nº 26, página 429, 1946.

[García99] J. A. García, A. Tazón, A. Mediavilla, I. Santamaría, M. Lázaro, C. J. Pantaleón y J. C. Pedro. "Modeling MESFETs and HEMTs for intermodulation distortion behaviour using a generalized radial basis function network". *Int. Journal RF Microwave CAE*, tomo 9, páginas 261–276, 1999.

[Gardner93] F. M. Gardner. "Interpolation in digital modems, part I: Fundamentals". *IEEE Transactions on Communications*, tomo 41, páginas 501–507, 1993.

[Giurcaneanu00] C. D. Giurcaneanu, I. Tabus y J. Rissanen. "MDL based digital signal segmentation". En *Proc. European Signal Processing Conference*, páginas 10–21. Tampere, Finlandia, septiembre 2000.

[Golub89] G. H. Golub y V. F. V. Loan. *Matrix computations*. Johns Hopkins University Press, Baltimore, 1989.

[González01] F. J. González, J. J. Murillo y A. Artés. "GCMAC-based predistortion for digital modulations". *IEEE Transactions on Communications*, tomo 49, nº 9, páginas 1679–1689, 2001.

[Griffiths73] D. A. Griffiths y A. J. Miller. "Hyperbolic regression - a model based on two phase piecewise linear regression with a smooth transition between regimes". *Commun. Stat.*, tomo 2, páginas 561–569, 1973.

[Haykin94] S. Haykin. *Neural Networks, a comprehensive foundation*. IEEE Press, New York, 1994.

[Haykin98] S. Haykin. *Neural networks: a comprehensive foundation*. Macmillan Publishing, New York, 1998.

[Heutmaker97] M. S. Heutmaker. "The error vector and power amplifier distortion". En *Proc. IEEE Wireless Communication Conference*, páginas 100–104. Boulder, Colorado, USA, agosto 1997.

[Higgins96] J. R. Higgins. *Sampling theory in Fourier and signal analysis*. Osford University Press, New York, primera edición, 1996.

[Hill94] G. Hill. "The benefits of undersampling". *Electronic Design*, páginas 69–79, julio 1994.

[Hlavachova97] K. Hlavachova y M. Verleysen. "Placing spline knots in neural networks using splines as activation functions". *Neurocomputing*, tomo 17, páginas 159–166, 1997.

[Horbelt00] S. Horbelt, A. M. Barrutia, T. Blu y M. Unser. "Spline kernels for continuous-space image processing". En *Proc. International Conference on Acoustics, Speech, and Signal Processing*, tomo 4, páginas 2191–2194. Istanbul, Turkey, junio 2000.

[Ibáñez98] J. Ibáñez, C. Pantaleón, I. Santamaría, D. Luengo y M. Lázaro. "Specific DSP based monitoring system for hydrogenerator sets". En *Proc. International Conference on Signal Processing Applications and Technology*, tomo 2, páginas 932–936. Toronto, septiembre 1998.

[Ibáñez00] J. Ibáñez, C. Pantaleón, I. Santamaría, T. Fernández y D. Martínez. "Nonlinearity estimation in power amplifiers based on undersampled temporal data". En *Proc. Instrumentation and Measurement Technology Conference*, tomo 2, páginas 1266–1271. Baltimore, mayo 2000.

[Ibáñez01] J. Ibáñez, C. Pantaleón, I. Santamaría, T. Fernández y D. Martínez. "Nonlinearity estimation in power amplifiers based on subsampled temporal data". *IEEE Transactions on Instrumentation and Measurement*, tomo 50, nº 4, páginas 882–887, agosto 2001.

[Ibáñez02] J. Ibáñez, C. Pantaleón, J. Diez, I. Santamaría y L. Vielva. "Spline pulse shaping with ISI-Free and matched filter receiver". En *Proc. European Signal Processing Conference*, páginas 1–4. Toulouse, France, 2002.

[Ibáñez04] J. Ibáñez, I. Santamaría, C. Pantaleón y L. Vielva. "Parametric smoothing of spline interpolation". En *Proc. International Conference on Acoustics, Speech, and Signal Processing*. Montreal, mayo 2004.

[Isaacson66] E. Isaacson y H. B. Keller. *Analysis of numerical methods*. Wiley, New York, 1966.

[Jayant84] N. S. Jayant y P. Noll. *Digital coding of waveforms: principles and applications to speech and video*. Prentice Hall, New Jersey, 1984.

[Jerri77] A. J. Jerri. "The Shannon sampling theorem - its various extensions and applications: a tutorial review". *Proceedings of the IEEE*, tomo 65, nº 11, páginas 1565–1596, 1977.

[Jeruchim00] M. C. Jeruchim, P. Balaban y K. S. Shanmugan. *Simulation of Communication Systems: Modeling, Methodology and Techniques*. Kluwer Academic / Plenum Publishers, New York, segunda edición, 2000.

[Jupp78] D. L. B. Jupp. "Approximation to data by splines with free knots". *SIAM Journal on Numerical Analysis*, tomo 15, páginas 328–343, 1978.

[Kamada95] M. Kamada, K. Toraichi y R. E. Kalman. "A smooth signal generator based on quadratic B-spline functions". *IEEE Transactions on Signal Processing*, tomo 43, nº 5, páginas 1252–1255, mayo 1995.

[Kashyap96] R. L. Kashyap y S. Moni. "Using the residual area criterion for signal reconstruction from noiseless and noisy samples". *IEEE Transactions on Signal Processing*, tomo 44, nº 3, páginas 732–738, marzo 1996.

[Kisel00] A. V. Kisel. "Nyquist 1 universal filters". *IEEE Transactions on Communications*, tomo 48, nº 7, páginas 1095–1099, julio 2000.

[Kootsookos96] P. J. Kootsookos y R. C. Williamsons. "FIR approximation of fractional sample delay systems". *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, tomo 43, páginas 269–271, marzo 1996.

[Kotelnikov33] V. A. Kotelnikov. *On the transmission capacity of ether and wire in electrocommunications*. Izd. Red. Upr. Svyazzi RKKA, Moscú, 1933.

[Kreyszig78] E. Kreyszig. *Introduction to functional analysis and applications*. John Wiley, New York, 1978.

[Kroon91] P. Kroon y B. S. Atal. "On improving the performance of pitch predictors in speecho coding systems". En B. S. Atal, V. Cuperman y A. Gersho, editores, *Advances in Speech Coding*, tomo 114 de *The Kluwer international series in engeneering and computer science*. Kluwer, 1991.

[Kusuma02] J. Kusuma, A. Ridolfi y M. Vetterli. "Sampling of communication systems with bandwidth expansion". En *IEEE Conf. on Communication*. New York, mayo 2002.

[Laakso96] T. I. Laakso, V. Valimaki, M. Karjalainen y U. K. Laine. "Spliting the unit delay: tools for fractional filter design". *IEEE Signal Processing Magazine*, tomo 13, páginas 30–60, enero 1996.

[Lee01a] J. S. Lee y C. Nguyen. "Novel low-cost ultrawideband , ultra-short-pulse transmitter wiyh MESFET impulse-shaping circuitry for reduced distortion and improved pulse repetition rate". *IEEE Microwave and Wireless Components Letters*, tomo 11, nº 5, páginas 208–210, mayo 2001.

[Lee01b] J. S. Lee, C. Nguyen y T. Scullion. "New uniplanar subnanosecond monocycle pulse generator and transformer for time-domain microwave applications". *IEEE Transactions on Microwave Theory and Technology*, tomo 49, nº 6, páginas 1126–1129, junio 2001.

[Leenaerts98] D. M. W. Leenaerts y W. M. G. van Bokhoven. *Piecewise linear modelling and analysis*. Kluwer, Boston, MA, 1998.

[Lin93] J. N. Lin y R. Unbehauen. "Canonical representation: from piecewise-linear function to piecewise-smooth functions". *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, tomo 40, páginas 461–468, julio 1993.

[López-Higuera98] J. López-Higuera, I. Santamaría, A. Cobo, C. Pantaleón, M. Morante, J. Ibáñez y J. Echevarría. "Quasi-smart optical fibre sensor system for real time and predictive monitoring on large rotating machinery". En *Proc. CLEO/Europe*, páginas 858–862. 1998.

[López-Higuera99] J. M. López-Higuera, I. Santamaría, P. Mottier, A. Cobo, C. Pantaleón, E. Ollier, M. Morante, J. Ibáñez, C. Chabrol y F. J. Madruga. "Optical sensors and their fusion in

a quasi-smart structure for real time vibration monitoring and predictive maintenance of large power electric generators". *Journal of Intelligent Material Systems and Structures*, tomo 9, nº 11, páginas 938–946, noviembre 1999.

[Luengo97a] D. Luengo, J. Ibáñez, C. Pantaleón y I. Santamaría. "Análisis y diseño de sistemas de retardo fraccional basados en sobremuestreo e interpoladores FIR de orden bajo". En *Actas del Simposium Nacional de la URSI*, tomo 1, páginas 721–724. Bilbao, septiembre 1997.

[Luengo97b] D. Luengo, C. Pantaleón, I. Santamaría, J. Ibáñez y E. Gómez-Cossío. "Simultaneous sampling by digital phase correction". En *Proc. Instrumentation and Measurement Technology Conference*, páginas 980–984. Otawa, mayo 1997.

[Luengo00] D. Luengo, C. J. Pantaleón, J. Ibáñez y I. Santamaría. "Design of simultaneous sampling systems based on fractional delay Lagrange filters". *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, tomo 47, nº 5, páginas 482–485, mayo 2000.

[Lázaro01a] M. Lázaro. *Técnicas de modelado no lineal con incorporación de restricciones en las derivadas y su aplicación a comunicaciones*. Tesis Doctoral, Departamento de Ingeniería de Comunicaciones, Universidad de Cantabria, Santander, España, septiembre 2001.

[Lázaro01b] M. Lázaro, I. Santamaría, C. Pantaleón, A. Mediavilla, A. Tazón y C. Navarro. "Smoothing the canonical piecewise linear model: An efficient and derivable large-signal model for MESFET/HEMT transistors". *IEEE Transactions on Circuits and Systems*, tomo 48, nº 2, páginas 184–192, febrero 2001.

[Madjar92] A. Madjar, P. R. Herczfeld y A. Paolella. "Analytical model for optically generated currents in GaAs MESFETs". *IEEE Transactions on Microwave Theory and Technology*, tomo 40, páginas 1681–1691, 1992.

[Mathews91] V. J. Mathews. "Adaptative polynomial filters". *IEEE Signal Processing Magazine*, tomo 8, nº 3, páginas 10–26, julio 1991.

[Mathews00] V. J. Mathews y G. L. Sicuranza. *Polynomial Signal Processing*. Wiley Series in Telecomunications and Signal Processing. John Wiley and Sons, New York, 2000.

[Meijering02] E. Meijering. "A chronology of interpolation: from ancient astronomy to modern signal and image processing". *Proceedings of the IEEE*, tomo 90, nº 3, páginas 317–342, marzo 2002.

[Mian82] G. A. Mian y A. P. Nainer. "A fast procedure to design equiripple minimum-phase FIR filters". *IEEE Transactions on Circuits and Systems*, tomo 29, páginas 327–331, mayo 1982.

[Moon00] T. K. Moon y W. C. Stirling. *Mathematical methods and algorithms for signal processing*. Prentice Hall, New Jersey, primera edición, 2000.

[Muller98] K. R. Muller, A. J. Smola, G. Ratsch, B. Schölkopff, J. Kohlmorgen y V.Ñ. Vapnik. "Using Support Vector Machines for time series prediction". En B. Schölkopf, C. J. Burges y A. J. Smola, editores, *Advances in Kernel Methods: Support Vector Learning*. MIT Press, Cambridge, MA, USA, 1998.

[Murphy97] P. Murphy, A. Krukowsky y A. Tarczynski. "An efficient fractional sample delayer for digital beam steering". En *Proc. International Conference on Acoustics, Speech, and Signal Processing*, páginas 2245–2248. mayo 1997.

[Navarro98] C. Navarro. "Large signal dynamic properties of GaAs MESFET/HEMT devices under optical ilumination". En *Proc. GAAS'98 Symp.*, páginas 350–353. Amsterdam, 1998.

[Neubauer04] A. Neubauer. "Approximation error of shifted signals in spline spaces". *IEEE Transactions on Signal Processing*, tomo 52, nº 4, páginas 921–928, abril 2004.

[Nyquist28] H. Nyquist. "Certain topics in telegraph transmission theory". *Trans. AIEE*, tomo 47, páginas 617–644, abril 1928.

[Oetken79] G. Oetken. "A new approach for the design of digital interpolating filters". *IEEE Transactions on Acoustic and Speech Signal Processing*, tomo 23, nº 3, páginas 301–309, junio 1979.

[Oppermann04] I. Oppermann. *UWB Theory and Applications*. John Wiley and Sons, New York, 2004.

[Ouderaa88] E. van der Ouderaa y J. Renneboog. "Some formulas and applications of nonuniform sampling of bandwidth-limited signals". *IEEE Transactions on Instrumentation and Measurement*, tomo 37, nº 3, páginas 353–357, septiembre 1988.

[Owaga86] H. Owaga. "A unified approach to generalized sampling theorems". En *Proc. International Conference on Acoustics, Speech, and Signal Processing*, páginas 1657–1660. Tokio, 1986.

[Panayirci99] E. Panayirci, T. Ozugur y H. Caglar. "Design of optimum Nyquist signals based on generalized sampling theory for data communications". *IEEE Transactions on Signal Processing*, tomo 47, nº 6, páginas 1753–1759, junio 1999.

[Pang94] D. Pang, L. A. Ferrari y P. V. Sankar. "B-spline FIR filters". *Circuits, Systems and Signal Processing*, tomo 13, nº 1, páginas 31–64, 1994.

[Pantaleón95a] C. Pantaleón, F. Díaz, I. Santamaría y J. Ibáñez. "Diseño y realización de una unidad de adquisición y procesado digital de señales orientada a la monitorización de grandes máquinas eléctricas". GTASSETEL 051095, Universidad de Cantabria, Santander, noviembre 1995. Proyecto Brite 7289.

[Pantaleón95b] C. Pantaleón y A. R. Figueiras. "Competitive estimation of polynomial filter banks". En *Proc. IEEE Workshop on Nonlinear Signal and Image Processing*, tomo 2, páginas 935–938. Halkidiki, Grecia, junio 1995.

[Pineda93] F. J. Pineda y J. C. Sommerer. "Estimating generalized dimensions and choosing time delays: a fast algorithm". En A. S. Weigend y N. A. Gershenfeld, editores, *Time series prediction: forecasting the future and understanding the past*, tomo 15 de *SFI Studies in the Sciences of Complexity*, páginas 367–385. Addison-Wesley, Reading, Mass., USA, 1993.

[Poggio90] T. Poggio y F. Girosi. "Networks for approximation and learning". *Proceedings of the IEEE*, tomo 78, páginas 1481–1497, 1990.

[Press89] W. H. Press, B. P. Flannery, S. A. Teuklosky y W. T. Vetteriling. *Numerical recipes*. Cambridge University Press, New York, 1989.

[Principe99] J. C. Principe, D. Xu y J. W. Fisher III. "Information-theoretic learning". En S. Haykin, editor, *Unsupervised Adaptive Filtering*, páginas 265–319. John Wiley and Sons, New York, USA, 1999.

[Principe00] J. C. Principe, D. Xu, Q. Zhao y J. W. Fisher III. "Learning from samples with information theoretic criteria". *Journal of VLSI Signal Processing-Systems*, tomo 26, nº 1–2, páginas 61–77, 2000.

[Ralston78] A. Ralston y P. Rabinowitz. *A first course in numerical analysis*. McGraw-Hill, New York, segunda edición, 1978.

[Ramstad84] T. A. Ramstad. "Digital methods for conversion between arbitrary sampling frequencies". *IEEE Transactions on Acoustic and Speech Signal Processing*, tomo 32, nº 3, páginas 577–591, junio 1984.

[Riesz55] F. Riesz y B. S. Nagy. *Functional analysis*. Frederick Ungar Publ., New York, 1955.

[Rissanen78] J. Rissanen. "Modeling by shortest data description". *Automatica*, tomo 14, páginas 465–471, 1978.

[Rolain98] Y. Rolain, J. Schoukens y G. Vandersteen. "Signal reconstruction for non-equidistant finite length sample sets: a "KIS" approach". *IEEE Transactions on Instrumentation and Measurement*, tomo 47, páginas 1046–1052, 1998.

[Rousset96] J. Rousset, Y. Harkouss y J. M. Campovecchio. "An accurate neural netwoek model of FET for intermodulation and power analysis". En *Proc. 26th European Microwave Conference*. Praga, Checoslovaquia, 1996.

[Saleh81] A. A. M. Saleh. "Frequency-independent and frequency-dependent nonlinear models of TWT amplifiers". *IEEE Transactions on Communications*, tomo 29, nº 11, páginas 1715–1720, 1981.

[Santamaría99] I. Santamaría, M. Lázaro, C. J. Pantaleón, J. A. García, A. Tazón y A. Mediavilla. "A nonlinear MESFET model for intermodulation analysis using a generalized radial basis

function network". *Neurocomputing*, tomo 45, páginas 1–18, 1999.

[Santamaría03] I. Santamaría, J. Ibañez, M. Lázaro, C. Pantaleón y L. Vielva. "Modeling nonlinear power amplifiers in OFDM systems from subsampled data: a comparative study using real measurements". *EURASIP Journal on Applied Signal Processing*, tomo 12, páginas 1219–1228, 2003.

[Santamaría04] I. Santamaría, J. Ibañez, C. Pantaleón y F. Cruz-Roldán. "Positive semidefinite biorthogonal partners for pulse shaping". *Remitido para su publicación en Signal Processing*, 2004.

[Schetzen80] M. Schetzen. *The Volterra and Wiener theories of nonlinear systems*. John Wiley and Sons, New York, 1980.

[Schoenberg46] I. J. Schoenberg. "Contribution to the problem of approximation of equidistant data by analytic functions". *Quart. Appl. Math.*, tomo 4, páginas 45–99, 1946.

[Schumaker81] L. L. Schumaker. *Spline functions: basic theory*. John Willey and Sons, New York, 1981.

[Seber89] G. A. F. Seber y C. J. Wild. *Nonlinear regression*. John Willey and Sons, New York, 1989.

[Shannon49] C. E. Shannon. "Communication in the presence of noise". *Proceedings of the IRE*, tomo 37, páginas 10–21, 1949.

[Shannon98] C. E. Shannon. "Communication in the presence of noise". *Proceedings of the IEEE*, tomo 86, nº 2, páginas 447–457, 1998.

[Shirakawa97] K. Shirakawa, M. Shimiz, N. Okubo y Y. Daido. "A large signal characterization of a HEMT using a multilayer neural network". *IEEE Transactions on Microwave Theory and Technology*, tomo 45, páginas 1630–1633, 1997.

[Shoumekh88] M. Shoumekh. "Band-limited interpolation from unevenly spaced sampled data". *IEEE Transactions on Acoustic and Speech Signal Processing*, tomo 36, nº 1, páginas 110–122, enero 1988.

[Sklar01] B. Sklar. *Digital communications: fundamentals and applications*. Prentice Hall, New Jersey, segunda edición, 2001.

[Slepian76] D. Slepian. "On bandwidth". *Proceedings of the IEEE*, tomo 64, nº 3, páginas 292–300, marzo 1976.

[Smola98] A. J. Smola. *Learning with Kernels*. Tesis Doctoral, Technische Universität Berlin, Berlin, Germany, 1998.

[Spiegel70] M. R. Spiegel. *Manual de fórmulas y tablas matemáticas*. Schaum. McGraw-Hill, Méjico, 1970.

[Sturm98] J. F. Sturm. "Using SeDuMi 1.02, a MATLAB Toolbox for optimization over symmetric cones". Disponible en http://www.unimaas.nl/sturm/software/sedumi.html. Department of Econometrics, Tilburg University. The Netherlands, 1998.

[Treichler96] J. R. Treichler, I. Fijalkow y C. R. Johnson. "Fractionally spaced equalizers: how long should they really be?" *IEEE Signal Processing Magazine*, páginas 65–81, mayo 1996.

[Tuncer02] T. E. Tuncer. "ISI-free pulse shaping filters for receivers with or without a matched filter". En *Proc. International Conference on Acoustics, Speech, and Signal Processing*, tomo III, páginas 2269–2272. Orlando, FL, mayo 2002.

[Unser92] M. Unser, A. Aldroubi y M. Eden. "Polynomial spline signal approximation: filter design and asymptotic equivalence with Shannon's sampling theorem". *IEEE Transactions on Information Theory*, tomo 38, páginas 95–103, enero 1992.

[Unser93a] M. Unser, A. Aldroubi y M. Eden. "B-spline signal processing: part I - theory". *IEEE Transactions on Signal Processing*, tomo 41, nº 2, páginas 821–833, febrero 1993.

[Unser93b] M. Unser, A. Aldroubi y M. Eden. "B-spline signal processing: part II - efficient design and applications". *IEEE Transactions on Signal Processing*, tomo 41, nº 2, páginas 834–848, febrero 1993.

[Unser94] M. Unser y A. Aldroubi. "A general sampling theory for nonideal acquisition devices". *IEEE Transactions on Signal Processing*, tomo 42, nº 11, páginas 2915–2925, noviembre 1994.

[Unser97] M. Unser. "Ten good reasons for using spline wavelets". En *SPIE, Wavelets Applicat. Signal Image Processing*, tomo 3169, páginas 422–431. 1997.

[Unser98] M. Unser y J. Zerubia. "A generalized sampling theory without band-limiting constrains". *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, tomo 45, nº 8, páginas 959–969, agosto 1998.

[Unser99] M. Unser. "Splines: a perfect fit for signal and image processing". *IEEE Signal Processing Magazine*, noviembre 1999.

[Unser00] M. Unser. "Sampling: 50 years after Shannon". *Proceedings of the IEEE*, tomo 88, nº 4, páginas 569–587, abril 2000.

[Vaidyanathan01a] P. P. Vaidyanathan. "Generalizations of the sampling theorem: seven decades after Nyquist". *IEEE Transactions on Circuits and Systems*, tomo 48, nº 9, páginas 1094–1109, septiembre 2001.

[Vaidyanathan01b] P. P. Vaidyanathan. "Sampling theorems for non bandlimited signals: theoretical impact and practical applications". En *Proc. 4th Intl. Conf. on Sampling Theory and Appl.*. 2001.

[Vaidyanathan01c] P. P. Vaidyanathan y B. Vrcelj. "Biorthogonal partners and applications". *IEEE Transactions on Signal Processing*, tomo 49, nº 5, páginas 1013–1027, mayo 2001.

[Vapnik97] V. N. Vapnik, S. Golowich y A. J. Smola. "Support vector method for function approximation, regression estimation, and signal processing". En M. Mozer, M. Jordan y T. Petsche, editores, *Advances in Neural Information Processing Systems*, tomo 9, páginas 281–287. MIT Press, Cambridge, MA, USA, 1997.

[Vapnik00] V. N. Vapnik. *The nature of statistical learning thory*. Statistics for Engineering and Information Science. Springer Verlag, New York, 2000.

[Vaughan91] R. G. Vaughan, N. L. Scott y D. R. White. "The theory of bandpass sampling". *IEEE Transactions on Signal Processing*, tomo 39, nº 9, páginas 1973–1984, septiembre 1991.

[Vetterli86] M. Vetterli. "Filterbanks allowing perfect reconstruction". *Signal Processing*, tomo 10, páginas 219–244, 1986.

[Välimäki95] V. Välimäki y M. Karjalaine. "Implementation of fractional waveguide models using allpass filters". En *Proc. International Conference on Acoustics, Speech, and Signal Processing*, páginas 1524–1527. 1995.

[Vrcelj02a] B. Vrcelj y P. P. Vaidyanathan. "Fractional biorthogonal partners in channel equalization and signal interpolation". *IEEE Transactions on Signal Processing*, tomo 50, nº 3, páginas 528–542, marzo 2002.

[Vrcelj02b] B. Vrcelj y P. P. Vaidyanathan. "MIMO biorthogonal partners and applications". *IEEE Transactions on Signal Processing*, tomo 50, nº 3, páginas 528–542, marzo 2002.

[Waele00] S. de Waele y P. M. T. Broersen. "Error measurements for resampled irregular data". *IEEE Transactions on Instrumentation and Measurement*, tomo 49, nº 2, páginas 216–221, abril 2000.

[Wang00] Q. Wang y L. Wu. "Translation invariance and sampling theorem of wavelet". *IEEE Transactions on Signal Processing*, tomo 48, páginas 1471–1474, mayo 2000.

[Wepman00] J. A. Wepman. "Analog-to-digital converters and their applications in radio receivers". *IEEE Communications Magazine*, páginas 39–45, mayo 2000.

[Whittaker29] J. M. Whittaker. "The Fourier theory of the cardinal functions". *Proc. Math. Soc. Edinburgh*, tomo 1, páginas 169–176, 1929.

[Win98] M. Z. Win y R. A. Scholtz. "Impulse radio: how it works". *IEEE Communications Letters*, tomo 2, nº 1, páginas 10–12, enero 1998.

[Wu96] S. P. Wu, S. Boyd y L. Vandenberghe. "FIR filter design via semidefinite programming and spectral factorization". *Proc. of the 35th IEEE Conference on Decision and Control*, tomo 1/4, páginas 271–276, 1996.

[Xia97] X. G. Xia. "A family of pulse-shaping filters with ISI-free matched and unmatched filter properties". *IEEE Transactions on Communications*, tomo 45, nº 10, páginas 1157–1158, octubre 1997.

[Ye97] Y. Ye. *Interior point algorithms: theory and analysis*. Wiley, New York, 1997.

[Yen56] J. L. Yen. "On nonuniform sampling of bandwidth-limited signals". *IRE Trans. Circuit Theory*, tomo CT–33, nº 5, páginas 251–257, diciembre 1956.

[Yiu01] K. F. C. Yiu, S. Wang, K. L. Teo y A. C. Tsoi. "Nonlinear system modeling via knot-optimizing B-spline networks". *IEEE Transactions on Neural Networks*, tomo 12, nº 5, páginas 1013–1022, septiembre 2001.

[Yuou96] P. Yuou, E. Baert y M. F. Loutre. "Spectral analysis of climate data". *Surv. Geophysics*, tomo 17, páginas 619–663, 1996.

[Zaabab95] A. H. Zaabab, Q. J. Zhang y M. Nakhla. "A neural network modeling approach to circuit optimization and statistical design". *IEEE Transactions on Microwave Theory and Technology*, tomo 43, páginas 1349–1358, 1995.