

Trabajo Fin de Grado Grado en Ingeniería de las Tecnologías de Telecomunicación

Esquema de linealización con arquitectura de aprendizaje directo y regularización de Ridge

Autor: Elías Marqués Valderrama

Tutores: Juan Antonio Becerra González y María José Madero Ayora

Dpto. Teoría de la Señal y Comunicaciones
Escuela Técnica Superior de Ingeniería
Universidad de Sevilla

Sevilla, 2021



Trabajo Fin de Grado
Grado en Ingeniería de las Tecnologías de Telecomunicación

Esquema de linealización con arquitectura de aprendizaje directo y regularización de Ridge

Autor:

Elías Marqués Valderrama

Tutor:

Juan Antonio Becerra González

Profesor Ayudante Doctor de Universidad

María José Madero Ayora

Profesora Titular de Universidad

Dpto. Teoría de la Señal y Comunicaciones
Escuela Técnica Superior de Ingeniería
Universidad de Sevilla

Sevilla, 2021

Trabajo Fin de Grado: Esquema de linealización con arquitectura de aprendizaje directo y regularización de Ridge

Autor: Elías Marqués Valderrama

Tutores: Juan Antonio Becerra González y María José Madero Ayora

El tribunal nombrado para juzgar el trabajo arriba indicado, compuesto por los siguientes profesores:

Presidente:

Vocal/es:

Secretario:

acuerdan otorgarle la calificación de:

El Secretario del Tribunal

Fecha:

Agradecimientos

Este trabajo es el cúlmen de una etapa que ha estado llena de momentos que recordaré para siempre, donde he estado rodeado de personas que lo han hecho más especial aún, y es por esto que quiero dedicarles estas líneas.

En primer lugar, quiero agradecer a mis padres, Antonio y Rosa, toda la ilusión y dedicación que han puesto en mí; a pesar de los esfuerzos que cada día ha supuesto el estudio siempre me habéis apoyado por encima de todo. Gracias por haberme entregado la Fé por encima de cualquier cosa. Gracias a mis hermanos, Abraham, Israel y Samuel, sois la mejor familia que Dios me podía regalar. Gracias a mi gran familia y a mis hermanos de comunidad, que siempre han creído en mí.

Gracias a mis compañeros y amigos que me han acompañado en esta etapa, en especial a Miguel, Iván y Rubén, por todos los momentos que hemos compartido y que espero que sean muchos más. Quiero agradecer también con un cariño especial a Juan Antonio Becerra y a María José Madero, por vuestra ilusión, vuestras ganas de enseñar, la disponibilidad, el apoyo y la alegría que plasmáis en lo que hacéis. Ha sido un disfrute ser vuestro alumno.

Hago especial mención a Gloria, que ha sido mi apoyo en estos cuatro años. Contigo los momentos difíciles han sido más llevaderos y las alegrías más alegres.

*Elías Marqués Valderrama
Sevilla, 2021*

Resumen

En este trabajo se pone a prueba una de las herramientas que brinda la estadística para conseguir estabilidad en las soluciones de un sistema de ecuaciones. Los modelos de doble banda requeridos en múltiples ocasiones para linealizar los sistemas de comunicaciones móviles 5G, suelen ser un problema mal condicionado algebraicamente, presentando problemas en el cálculo de la matriz inversa. Así pues, se aplica regularización de Ridge como método para la obtención de soluciones fiables y alcanzar una arquitectura de aprendizaje estable. Los resultados experimentales se han logrado para un amplificador de potencia de clase AB, cuya tarea ha sido amplificar dos señales concurrentes 5G-NR de 10 MHz de ancho de banda. Este tipo de investigaciones toma importancia debido a la complejidad que va adquiriendo progresivamente la capa física, provocando que la linealización sea un reto cada vez mayor.

Abstract

In this dissertation, it is tested one of the tools that statistics provides in order to achieve stability in the solutions of a system of equations. The dual-band models, multiple times required to linearize 5G mobile communication systems, are usually a problem that is badly conditioned algebraically, introducing difficulties in the calculation of the inverse matrix. Thus, Ridge regularization is applied as a method to obtain reliable solutions and reach a stable learning architecture. Experimental results have been obtained for a class AB power amplifier, whose task has been to amplify two concurrent 5G-NR signals of 10 MHz bandwidth. This type of research becomes important due to the complexity that it is progressively acquiring in the physical layer, causing linearization to be an increasing challenge.

Índice Abreviado

<i>Resumen</i>	III
<i>Abstract</i>	V
<i>Índice Abreviado</i>	VII
<i>Notación</i>	XI
1 Introducción	1
2 Amplificador de potencia como sistema no lineal	3
2.1 Definición de sistemas lineales e invariantes en el tiempo	3
2.2 Amplificador de potencia como sistema no lineal con memoria	4
2.3 Caracterización de un sistema no lineal	6
2.4 Medida del efecto de la distorsión sobre una señal modulada	9
2.5 OFDM como señal modulada en 5G-NR	10
3 Modelado y linealización	13
3.1 Modelos basados en series de Volterra	13
3.2 Identificación de coeficientes del modelo: Regresión	19
3.3 Predistorsión digital	19
4 Técnicas de reducción de complejidad y estabilización del modelo	23
4.1 Identificación de problemas numéricos	23
4.2 Regularización de Ridge	24
4.3 Métodos de selección de componentes	26
4.4 Regularización de Ridge aplicada en el espacio de correlaciones	28
5 Diseño experimental y resultados	31
5.1 Descripción del banco de medidas	31
5.2 Caracterización inicial	33
5.3 Aplicación de DPD-DLA	34
5.4 Aplicación de regularización de Ridge a DPD-DLA	35
5.5 Aplicación de regularización de Ridge a DPD-DLA con selección de coeficientes RC-DOMP	38
6 Conclusiones y líneas futuras	41
Apéndice A Código de MATLAB® empleado en el trabajo	43
Apéndice B Artículo de investigación	53
<i>Índice de Figuras</i>	59
<i>Índice de Tablas</i>	61
<i>Índice de Códigos</i>	63

Bibliografía

65

Glosario

67

Índice

<i>Resumen</i>	III
<i>Abstract</i>	V
<i>Índice Abreviado</i>	VII
<i>Notación</i>	XI
1 Introducción	1
2 Amplificador de potencia como sistema no lineal	3
2.1 Definición de sistemas lineales e invariantes en el tiempo	3
2.2 Amplificador de potencia como sistema no lineal con memoria	4
2.3 Caracterización de un sistema no lineal	6
2.3.1 Prueba de un tono	6
2.3.2 Prueba de dos tonos	7
2.3.3 Caracterización AM/AM	8
2.4 Medida del efecto de la distorsión sobre una señal modulada	9
2.4.1 Relación de potencia con el canal adyacente (ACPR)	9
2.4.2 Magnitud del vector error (EVM)	9
2.4.3 Error cuadrático medio normalizado (NMSE)	10
2.5 OFDM como señal modulada en 5G-NR	10
3 Modelado y linealización	13
3.1 Modelos basados en series de Volterra	13
3.1.1 Memory Polynomial (MP)	15
3.1.2 Generalized Memory Polynomial (GMP)	16
3.1.3 Modelos bidimensionales	16
3.1.4 Expresión en forma matricial	17
3.2 Identificación de coeficientes del modelo: Regresión	19
3.3 Predistorsión digital	19
3.3.1 Arquitectura de aprendizaje indirecto	20
3.3.2 Arquitectura de aprendizaje directo	21
3.3.3 Arquitectura de aprendizaje directo para doble banda concurrente	21
4 Técnicas de reducción de complejidad y estabilización del modelo	23
4.1 Identificación de problemas numéricos	23
4.2 Regularización de Ridge	24
4.2.1 Matrices aumentadas	26
4.3 Métodos de selección de componentes	26
4.3.1 Orthogonal Matching Pursuit (OMP)	26
4.3.2 Doubly Orthogonal Matching Pursuit (DOMP)	27
4.3.3 Reduced Complexity Doubly Orthogonal Matching Pursuit (RC-DOMP)	28
4.4 Regularización de Ridge aplicada en el espacio de correlaciones	28

5	Diseño experimental y resultados	31
5.1	Descripción del banco de medidas	31
5.2	Caracterización inicial	33
5.3	Aplicación de DPD-DLA	34
5.4	Aplicación de regularización de Ridge a DPD-DLA	35
5.5	Aplicación de regularización de Ridge a DPD-DLA con selección de coeficientes RC-DOMP	38
6	Conclusiones y líneas futuras	41
	Apéndice A Código de MATLAB® empleado en el trabajo	43
	Apéndice B Artículo de investigación	53
	<i>Índice de Figuras</i>	59
	<i>Índice de Tablas</i>	61
	<i>Índice de Códigos</i>	63
	<i>Bibliografía</i>	65
	<i>Glosario</i>	67

Notación

\mathbb{R}	Cuerpo de los números reales
\mathbb{C}	Cuerpo de los números complejos
$\ \mathbf{v}\ $	Norma del vector \mathbf{v}
$\ \mathbf{A}\ $	Norma de orden r de la matriz \mathbf{A}
\mathbf{A}^{-1}	Inversa de la matriz \mathbf{A}
\mathbf{A}^\dagger	Matriz pseudoinversa de la matriz \mathbf{A}
\mathbf{A}^H	Transpuesto y conjugado de \mathbf{A}
sen	Función seno
cos	Función coseno
$\sin^n x$	Función seno de x elevado a n
$\cos^n x$	Función coseno de x elevado a n
$\delta(t)$	Función Delta de Dirac
$\xrightarrow{\mathcal{F}}$	Aplicación de la Transformada de Fourier
$\frac{dy}{dx}$	Derivada de y respecto de x
$n!$	Factorial de n
$\mathbf{I}_{(M \times M)}$	Matriz identidad de dimensión M
$\mathbf{0}_{(M \times L)}$	Matriz de ceros de dimensión $M \times L$
diag(\mathbf{A})	Vector diagonal de la matriz \mathbf{A}
$\ \mathbf{x}\ $	Norma-2 del vector \mathbf{x}
$\mathbf{x}_i, \quad i = 1, 2, \dots, n$	Elementos i , de 1 a n , del vector \mathbf{x}
\leq	Menor o igual
\geq	Mayor o igual
\gg	Mucho mayor que
\approx	Aproximadamente igual
$\mathbb{E}[x]$	Esperanza matemática de x
\mathbf{R}_x	Matriz de correlación de \mathbf{x}
\hat{x}	Valor estimado de x
$tr[\mathbf{A}]$	Traza de la matriz \mathbf{A}
$\mathbf{A} \leftarrow \mathbf{B}$	Asignación
$\mathbf{A} \otimes \mathbf{B}$	Producto tensor Kronecker de las matrices \mathbf{A} y \mathbf{B}
$Col(\mathbf{A})$	Espacio columna de la matriz \mathbf{A}
\mathbf{P}_X	Operación de proyección sobre el conjunto \mathbf{X}
\emptyset	Conjunto vacío
$i \notin S$	i no pertenece a S
$A \cup B$	Unión de los conjuntos A y B

1 Introducción

La vida es como montar en bicicleta. Para mantener el equilibrio debes seguir moviéndote.

ALBERT EINSTEIN, 1930

Desde el comienzo de los tiempos el ser humano ha tenido la necesidad de comunicarse. En cuanto a las comunicaciones a distancia se han ido reinventando distintos métodos para hacer llegar un mensaje al destino deseado. Así pues las telecomunicaciones constituyen el arte del modelado de las ondas electromagnéticas con fin de conseguir que las comunicaciones a distancia sean sencillas y universales. Desde las primeras generaciones de comunicaciones móviles hasta nuestros días esta necesidad ha ido transformándose, desde un servicio dedicado al envío de voz, hasta la irrupción de las *app*, video *in streaming*, etc, donde la tasa de datos es fundamental. Además, con el paso del tiempo, el servicio móvil ha ido instaurándose en la sociedad, convirtiéndose en una herramienta fundamental e indispensable.

Estas nuevas necesidades han supuesto un desafío tecnológico con una evolución sin precedentes. Desde el paso a las comunicaciones digitales con la segunda generación móvil se empieza a ofrecer servicios de mensajería móvil como el SMS, que poco a poco fue derivando en un mayor consumo debido a su popularidad. El aumento del mercado ha ido propiciando la evolución de las generaciones móviles, que por lo general han centrado sus esfuerzos en satisfacer la creciente demanda en velocidad de datos, conectividad, y cantidad de equipos conectados. La llegada de 5G supone un punto de inflexión en cuanto al enfoque. Si bien 4G con LTE-A pretende alcanzar una solución a largo plazo, el abanico de servicios demandados es cada vez más amplio gracias a la evolución tecnológica.

En este escenario se lanza 5G, que además de mejorar la tasa de datos, ofrece soluciones específicas de alta tasa, conectividad ultrafiable y conectividad en áreas de alta demanda. Destaca el empleo de la técnica de agregación de portadora (CA) para aumentar la tasa de transmisión proporcionada al usuario, que da lugar a bandas de transmisión concurrentes. Uno de los compromisos de 5G es la eficiencia energética, con el fin de ir evolucionando a unas comunicaciones más sostenibles. Es por esto que se dedican grandes esfuerzos en el diseño de los transceptores, y más concretamente en el amplificador de potencia (PA), que constituye el subsistema más crítico en cuanto a eficiencia en potencia. A simple vista parece un problema irremediable, ya que gran parte de la potencia que se suministra al PA se disipa en forma de calor, siendo este comportamiento inherente al dispositivo. Además se debe refrigerar convenientemente para que no se sobrecaliente, hecho que supone un aumento mayor en el consumo eléctrico de los transceptores. Pero no se trata del único factor que contribuye en la ineficiencia del amplificador, sino que existe otro fenómeno relacionado con su comportamiento. Este es debido al punto de funcionamiento al que se somete el amplificador, ya que, dado que es costoso hacer que funcione, se busca aprovechar al máximo su capacidad de generar una potencia elevada a la salida. Así pues, se trabaja con potencias de entrada muy cercanas al punto de saturación del PA, y debido a las características de la señal modulada, el amplificador termina por saturar. De manera que al conjunto de los problemas térmicos que degradan la eficiencia, se suman la saturación y los efectos no lineales, que convierten al PA en un sistema que introduce una fuerte componente de distorsión en la señal, además de provocar recrecimiento espectral.

Es por esto que se invierte tanto esfuerzo en desarrollar técnicas que permitan linealizar el sistema, para que además de conseguir una amplificación lineal, la señal enviada tenga las prestaciones adecuadas. En concreto la técnica de predistorsión digital consiste en diseñar un bloque no lineal cuyo comportamiento

sea el inverso al que produce el PA. Para ello se requiere modelar correctamente el comportamiento del amplificador, donde destacan las series de Volterra ampliamente utilizadas en la literatura. La exigencia en cuanto a la precisión de los modelos y el empleo de señales en doble banda, provoca que a menudo se empleen modelos de Volterra excesivamente grandes, que derivan en ocasiones en problemas numéricos. Es por esto que se hace uso de distintas técnicas para la reducción de la complejidad y la estabilización de las soluciones como la regularización de Ridge, que será ampliamente estudiada en este trabajo.

Así pues, el objetivo de este trabajo consiste en el estudio de la regularización de Ridge, como técnica principal, en combinación con otros métodos de reducción de complejidad, para así alcanzar la estabilidad, manteniendo al máximo las prestaciones del sistema. Dado que se trata de un problema ampliamente estudiado, presenta multitud de soluciones posibles, cada cual enfocada en un matiz distinto. Es por esto por lo que se profundizará en la naturaleza de la inestabilidad, para así asociar el método con la causa, pudiendo establecer comparaciones que ahonden en los veneficios de emplear dichas técnicas.

De manera que este documento comenzará con una introducción teórica a los sistemas no lineales en el Capítulo 2, con fin de motivar y caracterizar el problema. Seguidamente, en el Capítulo 3, se presentarán las técnicas de modelado, en las que se fundamenta la predistorsión digital y las arquitecturas de implementación para la identificación de sus coeficientes. En el Capítulo 4 se presentarán distintos métodos para abordar los problemas numéricos, y finalmente se pondrán a prueba experimentalmente en el Capítulo 5.

2 Amplificador de potencia como sistema no lineal

La visión sin acción es un sueño. Acción sin visión es simplemente pasar el tiempo. Acción con visión es hacer una diferencia positiva.

JACK WELCH

En telecomunicaciones, clásicamente se entiende un sistema como una caja negra que recibe una serie de entradas y devuelve un conjunto de salidas. El sistema estará definido en función de la relación que guarden las entradas con las salidas, por lo que este apartado tratará de definir el dispositivo a partir de dichas relaciones. Dado que el dispositivo con el que se va a trabajar es de una entrada y una salida la explicación se centrará en este caso. Así pues, sea $x(t)$ una señal continua, se define un sistema como aquel dispositivo capaz de transformar la señal $x(t)$ en otra señal $y(t)$, denominada señal transformada, siendo $T[\cdot]$ una operación de transformación a definir.

$$\begin{aligned} x(t) &\rightarrow \boxed{\text{Sistema}} \rightarrow y(t) \\ y(t) &= T[x(t)] \\ x(t) &\rightarrow y(t) \end{aligned}$$

Figura 2.1 Representación de un sistema.

De modo que los sistemas se clasificarán en función de las características de la transformación $T[\cdot]$.

2.1 Definición de sistemas lineales e invariantes en el tiempo

Debido a sus propiedades, los sistemas Lineales e Invariantes en el Tiempo (LTI) son los más interesantes en cuanto a la integridad de una señal, y es por esto que son preferibles siempre que se puedan emplear. Se define un sistema lineal como aquel que cumple el principio de superposición. Sean las salidas $y_1(t), \dots, y_N(t)$ correspondientes a las entradas $x_1(t), \dots, x_N(t)$ se cumple el principio de superposición si y solo si

$$\sum_{n=1}^N \alpha_n x_n(t) \rightarrow \sum_{n=1}^N \alpha_n y_n(t), \quad (2.1)$$

donde $\alpha_n \in \mathbb{C}$ e $y_n(t)$ representa la n -ésima salida correspondiente a la entrada $x_n(t)$. Cabe destacar que esta relación depende únicamente del tipo de transformación que defina el sistema, es decir, de $T[\cdot]$.

Del mismo modo, se define un sistema invariante en el tiempo como aquel que cumple la relación mostrada en la Figura 2.2. Esto quiere decir que el sistema responde de la misma forma independientemente del instante de tiempo en el que se introduzca la señal $x(t)$.

$$\begin{aligned}x(t) &\rightarrow y(t) \\x(t-t_0) &\rightarrow y(t-t_0)\end{aligned}$$

Figura 2.2 Sistema invariante en el tiempo.

En sistemas lineales e invariantes en el tiempo se trabaja con la respuesta impulsiva del sistema, que se define como la salida cuando $x(t) = \delta(t)$, pudiendo expresar la relación entre la entrada y la salida como una convolución. Se trata de una característica distintiva de los sistemas LTI, por ende, si se identifica que un sistema cumple esta operación puede decirse que es LTI.

$$\begin{aligned}\delta(t) &\rightarrow \boxed{\text{Sistema}} \rightarrow h(t) \\x(t) &\rightarrow \boxed{h(t)} \rightarrow y(t) \\y(t) &= h(t) * x(t) = \int_{-\infty}^{\infty} h(\tau)x(t-\tau) d\tau\end{aligned}$$

Figura 2.3 Respuesta impulsiva de un sistema.

Particularmente, el tipo de comportamiento que se desea que cumpla el PA es el de un escalado en magnitud, es decir, $h(t) = g_v \delta(t)$, siendo g_v la ganancia en amplitud. En otras palabras, la relación entre la potencia de entrada y la potencia de salida debe ser una dependencia lineal, cuya pendiente sea la ganancia del amplificador g .

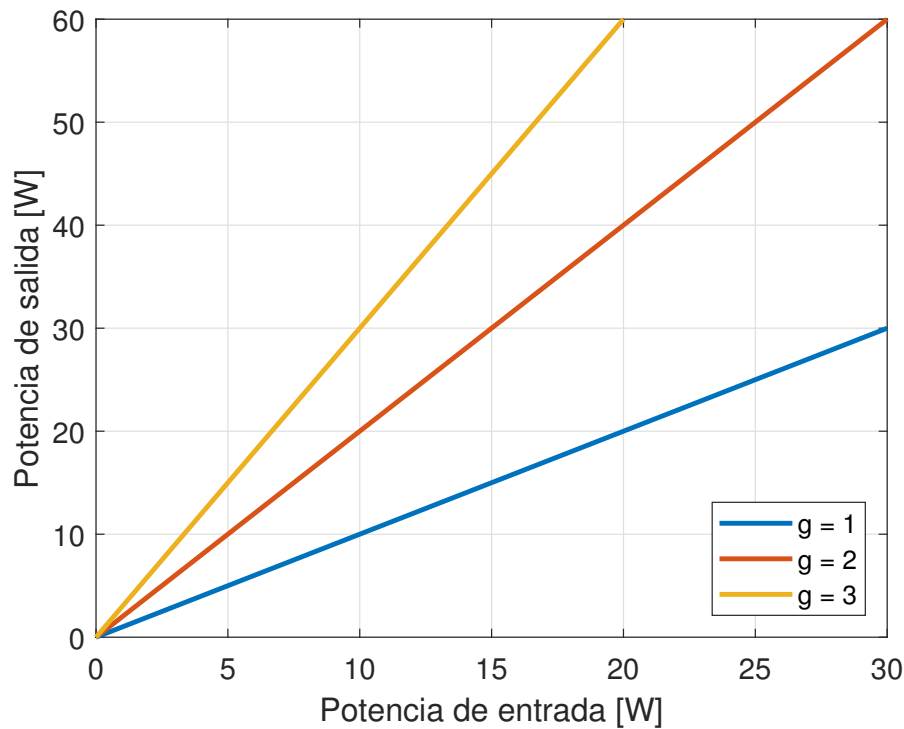


Figura 2.4 Potencia de salida frente a potencia de entrada con tres ganancias diferentes.

Para evaluar la invariabilidad en el tiempo sería necesaria una tercera dimensión en la Figura 2.4 con la que se observaría un plano.

2.2 Amplificador de potencia como sistema no lineal con memoria

Los sistemas comentados en la sección anterior son ideales y se alejan bastante de la realidad. El primer motivo es la conservación de la energía. Dado que el dispositivo tiene una capacidad de generación de potencia

limitada, existe una potencia de entrada a partir de la cual el PA no es capaz de mantener una dependencia lineal. En segundo motivo es el hecho de que se trata de un dispositivo electrónico, que cambiará sus prestaciones en función de la temperatura, y debido a que en general el punto de polarización es elevado, el dispositivo tiende a calentarse. De manera que a priori es posible predecir que el comportamiento del amplificador no va a ser el de un sistema LTI.

Comenzando por el dominio de la señal, según [1] se define un sistema no lineal como se muestra en la Figura 2.5.

$$x(t) \rightarrow \boxed{\text{SNL}} \rightarrow y_{NL}(t) = a_1x(t - \tau_1) + a_2x(t - \tau_2)^2 + \dots + a_nx(t - \tau_n)^n$$

Figura 2.5 Sistema no lineal.

En consecuencia, aparecen nuevos términos retrasados y no lineales. En los casos en los que $x(t)$ tuviese una magnitud pequeña podrían despreciarse los términos polinómicos. Esto sería deseable, pero no es compatible con la idea de conseguir una potencia de salida elevada, debido a que el amplificador no siempre tiene una ganancia alta. Esta circunstancia suele forzar a que la amplitud de la señal de entrada sea la mayor posible. Así pues, si se emplea la señal

$$x(t) = A(t) \cos[\omega_c t + \theta(t)], \quad (2.2)$$

donde se permite una modulación en amplitud o en fase indistintamente, y se cuenta únicamente con los tres primeros términos por simplicidad, la señal de salida del sistema no lineal será:

$$\begin{aligned} y_{NL}(t) = & a_1A(t - \tau_1) \cos[\omega_c t + \theta(t - \tau_1) - \phi_1] \\ & + a_2A(t - \tau_2)^2 \cos[\omega_c t + \theta(t - \tau_2) - \phi_2]^2 \\ & + a_3A(t - \tau_3)^3 \cos[\omega_c t + \theta(t - \tau_3) - \phi_3]^3, \end{aligned} \quad (2.3)$$

donde $\phi_j = \omega_c \tau_j$ es un desplazamiento en la fase. Desarrollando las potencias de los cosenos se obtiene finalmente:

$$\begin{aligned} y_{NL}(t) = & a_1A(t) \cos[\omega_c t + \theta(t) - \phi_1] \\ & + \frac{1}{2}a_2A(t)^2 + \frac{1}{2}a_2A(t)^2 \cos[2\omega_c t + 2\theta(t) - 2\phi_2] \\ & + \frac{3}{4}a_3A(t)^3 \cos[\omega_c t + \theta(t) - \phi_3] \\ & + \frac{1}{4}a_3A(t)^3 \cos[3\omega_c t + 3\theta(t) - 3\phi_3] \end{aligned} \quad (2.4)$$

donde se ha despreciado el retardo τ_k por la diferencia en la velocidad de variación entre la señal envolvente o la fase moduladora en relación con la portadora. Cada uno de los términos con frecuencia angular $k\omega_c$ se denomina armónico, donde la frecuencia fundamental se corresponde con el coseno con frecuencia angular ω_c .

Tal y como puede verse en la Figura 2.6, la forma de la señal a la entrada del sistema se ha visto fuertemente alterada. A la diferencia entre la señal de entrada escalada y la salida se le denomina distorsión. La primera implicación que puede extraerse de este análisis se intuye por conservación de la energía, es decir, la energía que entra al sistema tiene que salir de él. Teniendo esto presente puede interpretarse el fenómeno de la distorsión como un reparto de energías. Esto quiere decir que, idealmente se espera que el sistema asigne toda la energía al primer término de (2.4). Sin embargo, el amplificador invierte parte de su capacidad en generar términos no deseados y aplicando así menos potencia al término de señal.

Por otro lado se debe analizar el espectro. Para ello es conveniente aplicar la transformada de Fourier a la expresión de la Figura 2.5. Considerando $\phi_j \simeq 0$, modulación en amplitud, y agrupando las amplitudes de los distintos cosenos en los términos v_j se reescribe la ecuación (2.4) como:

$$y_{NL}(t) = v_0 + v_1 \cos[\omega_c t] + v_2 \cos[2\omega_c t] + \dots + v_n \cos[n\omega_c t], \quad (2.5)$$

cuya transformada sería:

$$y_{NL}(t) \xrightarrow{\mathcal{F}} Y_{NL}(f) = v_0\delta(f) + v_1\delta(f - f_c) + v_2\delta(f - 2f_c) + \dots + v_n\delta(f - nf_c). \quad (2.6)$$

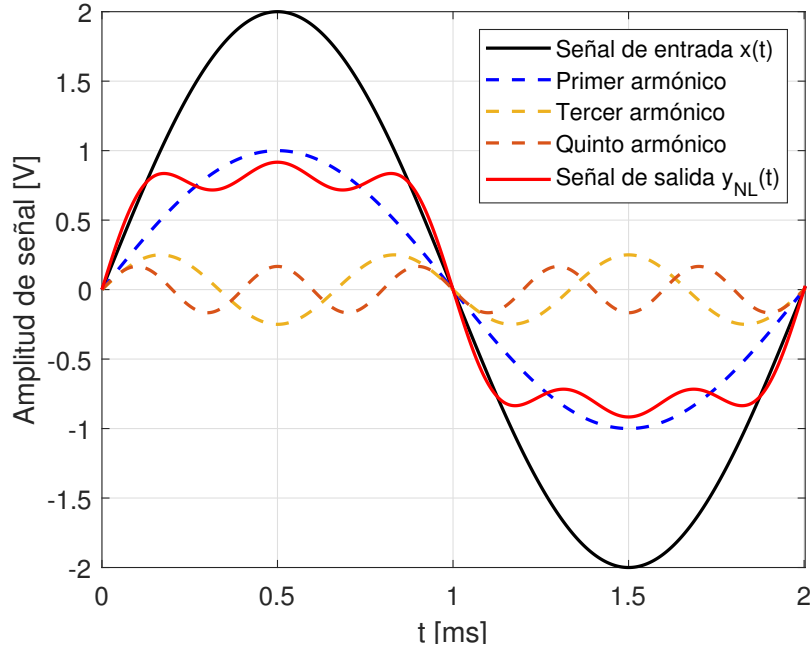


Figura 2.6 Descomposición de la salida y_{NL} en tres de sus armónicos.

Como la señal de entrada sólo tenía la frecuencia f_c , se deduce que el sistema ha generado frecuencias nuevas cada kf_c . A estas frecuencias se les da el nombre de frecuencias espurias, que son los armónicos ya mencionados. En definitiva, este fenómeno presenta la problemática de que distribuye la potencia de salida en armónicos que estropean la señal.

2.3 Caracterización de un sistema no lineal

Aunque los sistemas LTI presentan grandes ventajas, son sistemas ficticios, ya que lo habitual es que los dispositivos reales presenten un comportamiento no lineal. No obstante, sería posible marcar un umbral a partir del cual considerar que un dispositivo es lineal. La forma de comprobar cuán no lineal es un dispositivo es mediante una caracterización experimental, realizando medidas que sean capaces de cuantificar los efectos de la no linealidad. Por consiguiente, el sistema se puede caracterizar en base a las medidas que se exponen a continuación.

2.3.1 Prueba de un tono

Esta prueba consiste simplemente en poner en práctica la teoría del Apartado 2.2. Si un sistema lineal cumple que dada la entrada

$$x(t) = A_{in} \cos(2\pi f_c t), \quad (2.7)$$

su salida debe ser

$$y(t) = A_{out}(f) \cos[2\pi f_c t + \phi_{out}(f)]. \quad (2.8)$$

Por lo tanto, si el dispositivo bajo prueba (DUT) es no lineal, su salida será:

$$y(t) = \sum_{k=1}^{\infty} A_{out_k}(f, A_{in}) \cos[2\pi k f_c t + \phi_{out_k}(f, A_{in})]. \quad (2.9)$$

Este efecto puede verse fácilmente en un analizador de espectro. Como ya se ha mencionado anteriormente, la aparición de los armónicos se debe a un reparto en la potencia que suministra el amplificador. De tal forma que si se hace una comparativa entre la potencia asignada al armónico fundamental y la potencia entregada al

resto de armónicos se puede obtener un valor numérico relativo a la no linealidad del dispositivo

$$THD = \frac{\sqrt{\frac{1}{T} \int_0^T \left[\sum_{j=2}^{\infty} A_{out_k}(f, A_{in}) \cos[2\pi kft + \phi_{out_k}(f, A_{in})] \right]^2 dt}}{\sqrt{\frac{1}{T} \int_0^T A_{out_1}(f, A_{in}) \cos[2\pi kft + \phi_{out_1}(f, A_{in})]^2 dt}}, \quad (2.10)$$

donde THD significa distorsión armónica total. Por otro lado, la amplitud del armónico fundamental en la salida depende también de A_{in} , como puede verse en (2.9) y (2.10). Esto es lo que provoca el fenómeno de compresión, que será comentado en 2.3.3.

2.3.2 Prueba de dos tonos

Si bien la prueba de un tono es capaz de determinar si un sistema es lineal o no, aporta poca información. Concretamente la prueba de un tono no es capaz de determinar cómo afecta una señal en un canal adyacente al canal de interés. Es por esto que se repite el mismo experimento pero incluyendo otro tono a otra frecuencia distinta

$$x(t) = A_{1,in} \cos(2\pi f_1 t) + A_{2,in} \cos(2\pi f_2 t). \quad (2.11)$$

Por lo tanto, ahora el sistema distorsionará ambas señales, en la forma

$$y(t) = \sum_{n=1}^{\infty} A_{out_n}(f, A_{1,in}, A_{2,in}) \cos [2\pi f_n t + \phi_{out_n}(f, A_{1,in}, A_{2,in})] \quad \text{con } f_n = mf_1 + kf_2 \quad \text{para } m, k \in \mathbb{Z}, \quad (2.12)$$

donde además se define el orden de intermodulación como $n = |m| + |k|$. Es necesario prestar especial atención a la intermodulación de orden tres, ya que, como se muestra en la Figura 2.7, provoca productos de mezcla a frecuencias muy cercanas a la señal de interés. Destaca como figura de mérito de esta medida el punto de intercepto de tercer orden [2].

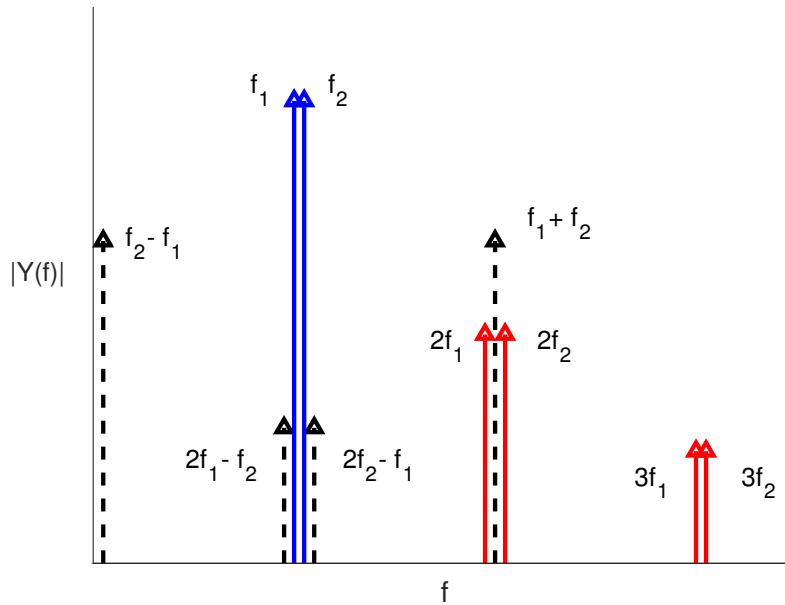


Figura 2.7 Resultado de la prueba de dos tonos. En azul se presentan las frecuencias de interés f_1 y f_2 , en rojo el segundo y tercer armónico y en discontinuo los productos de intermodulación de segundo y tercer orden.

2.3.3 Caracterización AM/AM

Esta caracterización es igual de intuitiva que la anterior y se puede realizar tanto para un tono como para una señal modulada. Si bien se desea que el amplificador sea lineal y guarde una relación proporcional entre la entrada y la salida, un dispositivo no lineal no la cumple. Así pues, un sistema no lineal tendrá una característica AM/AM similar a la mostrada en la Figura 2.8, en la que se considera que el sistema está en saturación a partir del punto de compresión de 1 dB, donde la respuesta del sistema y su respuesta lineal se separan 1 dB. El caso de la Figura 2.8 se trata de una no linealidad suave, siendo más habitual una transición más abrupta entre la zona lineal y la no lineal.

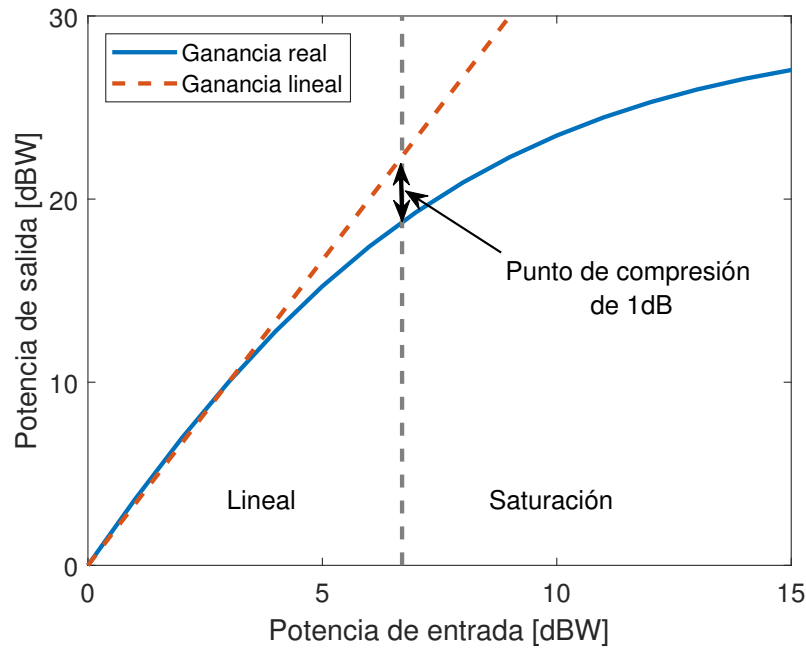


Figura 2.8 Ganancia de un amplificador no lineal.

Por último cabe destacar que, a pesar de que el amplificador sea un sistema no lineal y con memoria su parámetro más importante es la ganancia. Igualmente puede calcularse una ganancia promedio como

$$G_{avg}[dB] = 10 \log_{10} \left(\frac{P_{in}}{P_{out}} \right). \quad (2.13)$$

Nótese que al tratarse de un promedio, esta ganancia es un tanto mayor a la ganancia comprimida G_c , obtenida cuando se introduce la máxima potencia de entrada. La diferencia entre estos valores quedará determinada por el punto de operación, que será la potencia promedio de la señal a la entrada del amplificador, y las características de la señal, que harán que la potencia de entrada oscile entorno al punto de operación. La fluctuación máxima será cuantificada en la Sección 2.5. Así pues, cuanto mayor sea la zona de saturación empleada mayor desigualdad existirá entre G_{avg} y G_c .

Además, dado que se trata de un sistema real, es conveniente definir una medida de eficiencia. Así pues se define la eficiencia de potencia añadida (PAE) como

$$PAE[\%] = \frac{P_{out} - P_{in}}{P_{source}} \times 100, \quad (2.14)$$

siendo P_{source} la potencia que alimenta al amplificador. En general los amplificadores son poco eficientes siguiendo esta relación, ya que consumen gran parte de su potencia en disipación de calor y en la polarización del transistor que lo compone. Es por esto que se trata de un parámetro a tener en cuenta.

2.4 Medida del efecto de la distorsión sobre una señal modulada

Si bien la prueba de dos tonos pretende simular el efecto de un canal adyacente sobre el canal de interés, el resultado puede ser un tanto pobre en el sentido de que, en general, la señal modulada suele tener un ancho de banda mayor. Es por esto que es habitual emplear otras medidas que den un nivel mayor de profundidad en cuanto a la afectación de la distorsión sobre la señal, permitiendo así caracterizar la no linealidad en un entorno más realista y con un mayor nivel de detalle. El concepto es similar al de la prueba de dos tonos, pudiéndose entender una señal modulada como una multitono. Además, de ahora en adelante se supondrá que se están empleando modulaciones digitales. Estas pruebas son de tal relevancia que constituyen las figuras de mérito que evalúan la calidad del dispositivo.

2.4.1 Relación de potencia con el canal adyacente (ACPR)

Se trata de una medida que da una información que guarda relación con el THD pero con una señal que ocupa un cierto ancho de banda y analizando únicamente los canales adyacentes. De manera que se podría expresar el ACPR como

$$\text{ACPR}_T[\text{dBc}] = 10 \log_{10} \left(\frac{\int_{f_{L_1}}^{f_{L_2}} S_y(f) df + \int_{f_{U_1}}^{f_{U_2}} S_y(f) df}{\int_{f_{C_1}}^{f_{C_2}} S_y(f) df} \right), \quad (2.15)$$

donde los intervalos $[f_{L_1}, f_{L_2}]$ y $[f_{U_1}, f_{U_2}]$ representan los canales adyacentes inferior y superior respectivamente, y el intervalo $[f_{C_1}, f_{C_2}]$ es el ancho de banda de canalización. Es importante incidir en que $f_{L_2} \neq f_{C_1}$ y $f_{C_2} \neq f_{U_1}$, ya que se deja una pequeña banda de guarda entre los canales a evaluar. Además, se toma el mismo ancho de banda de la señal de canalización en los canales adyacentes, por lo que se pueden reescribir los intervalos como $[f_{-1} - B/2, f_{-1} + B/2]$ para el canal adyacente inferior y $[f_{+1} - B/2, f_{+1} + B/2]$ para el superior. De ahora en adelante se emplearán los subíndices -1 y $+1$ para denotar a los canales inferior y superior respectivamente. Por su parte, el ancho de canalización también puede denotarse como $[f_c - B/2, f_c + B/2]$. Por otro lado cabe esperar que, en general, el recrecimiento no sea simétrico debido a los efectos de memoria, de modo que se puede definir el ACPR en cada canal adyacente como

$$\begin{aligned} \text{ACPR}_{-1}[\text{dBc}] &= 10 \log_{10} \left(\frac{\int_{f_{-1}-B/2}^{f_{-1}+B/2} S_y(f) df}{\int_{f_c-B/2}^{f_c+B/2} S_y(f) df} \right) \\ \text{ACPR}_{+1}[\text{dBc}] &= 10 \log_{10} \left(\frac{\int_{f_{+1}-B/2}^{f_{+1}+B/2} S_y(f) df}{\int_{f_c-B/2}^{f_c+B/2} S_y(f) df} \right). \end{aligned} \quad (2.16)$$

Tanto en (2.15) como en (2.16) [dBc] representa la relación entre potencia de señal y potencia de portadora.

2.4.2 Magnitud del vector error (EVM)

Esta medida se centra exclusivamente en la recepción digital. Se trata de proyectar los símbolos recibidos en la representación vectorial del espacio de señal de la señal transmitida. La medida se efectúa como la diferencia promedio entre el símbolo recibido demodulado \vec{x}_i y el símbolo transmitido \vec{x}_i . Como se trata de una medida que involucra a la constelación al completo se computa para los N símbolos transmitidos como

$$\text{EVM}[\%] = \sqrt{\frac{\sum_{i=1}^N \|\vec{x}_i - \vec{x}_i\|^2}{\sum_{i=1}^N \|\vec{x}_i\|^2}} \times 100, \quad (2.17)$$

de manera que cuanto más cercano a cero sea el EVM mejor será la integridad de la señal recibida. Hay que incidir en que sería posible una detección sin probabilidad de error de símbolo aun teniendo $\text{EVM} > 0$, ya que

es posible que exista varianza en la constelación recibida que no sobrepase la frontera de decisión. Por otro lado, hay que tener en cuenta que si bien el EVM proporciona una medida de la varianza de la constelación recibida, no es capaz de distinguir entre la aportación del ruido y la de la distorsión sobre dicha varianza. De modo que, en general es necesario tener una idea del ruido que imprimen los dispositivos sobre la señal, y procurar que este sea lo más bajo posible para que así esta medida sirva como indicador de distorsión.

Por último puede añadirse un matiz a esta caracterización, y es el hecho de que la medida se centra exclusivamente en el canal de interés. Así pues, la distorsión evaluada tiene un aporte únicamente en el ancho de banda de canalización, por lo que puede denotarse como distorsión en banda.

2.4.3 Error cuadrático medio normalizado (NMSE)

Esta medida a diferencia del ACPR y del EVM no tiene como objetivo diferenciar entre distorsión en banda o fuera de banda. Más bien se intenta medir la calidad de la señal en términos absolutos, comparando la señal modelada y_{mod} con la señal real y_{med} de la forma

$$\text{NMSE}[\text{dB}] = 20 \log_{10} \left\{ \frac{\sum_{k=1}^{\infty} |y_{mod}(k) - y_{med}(k)|}{\sum_{k=1}^{\infty} |y_{med}(k)|} \right\}. \quad (2.18)$$

Esta comprobación permite evaluar cuán bueno es el modelo adoptado, aunque igualmente el NMSE sirve como medida de no linealidad en el caso en el que se elija como modelo un comportamiento lineal.

2.5 OFDM como señal modulada en 5G-NR

Conforme han ido evolucionando las generaciones móviles se han ido buscando formas de mejorar la experiencia del usuario. Entre todas las líneas de mejora destaca la capa física y las modulaciones empleadas, consiguiendo velocidades desde 271 kbps en GSM hasta 20 Gbps en 5G-NR. Si bien ya en 1966 Robert Chang [3] había planteado una multiplexación por división de frecuencias ortogonales (OFDM), no consigue implementarse en estándares móviles hasta la Release 8 de LTE. La motivación para emplear este tipo de modulación es la gran facilidad que presenta a la hora de ecualizar y multiplexar, siendo además más eficiente espectralmente.

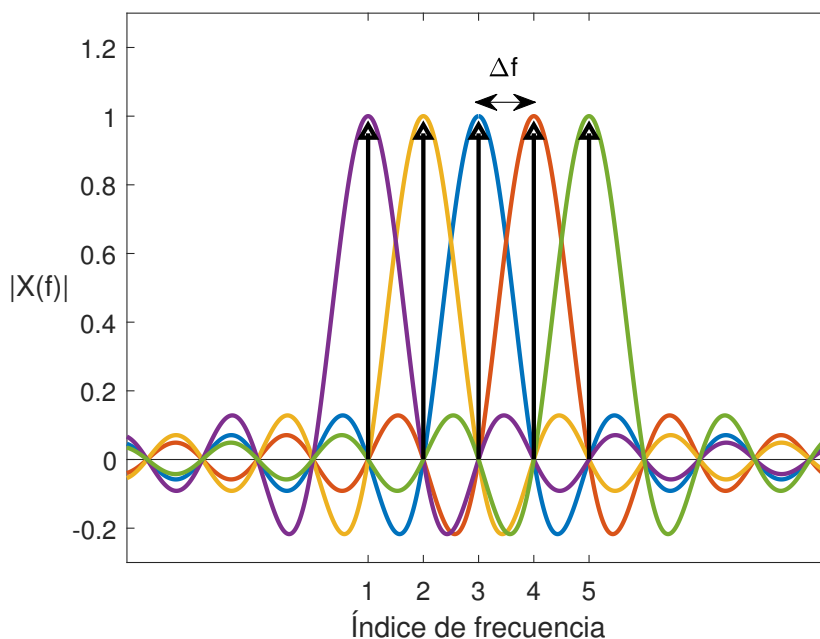


Figura 2.9 Espectro de señal OFDM con 5 subportadoras activas.

La modulación OFDM consiste en un conjunto de subportadoras espaciadas una cierta cantidad en frecuencia Δf , donde cada subportadora transporta un flujo de información, tal y como se muestra en la Figura 2.9.

Con fin de conseguir la ortogonalidad, la propuesta de Robert Chang fue evolucionando conforme la electrónica lo permitía. Los estudios posteriores consiguieron sustituir los osciladores locales por la operación de la Transformada rápida de Fourier Inversa (IFFT), que es más sencilla y barata de implementar. Así pues, se emplea el esquema mostrado en la Figura 2.10.

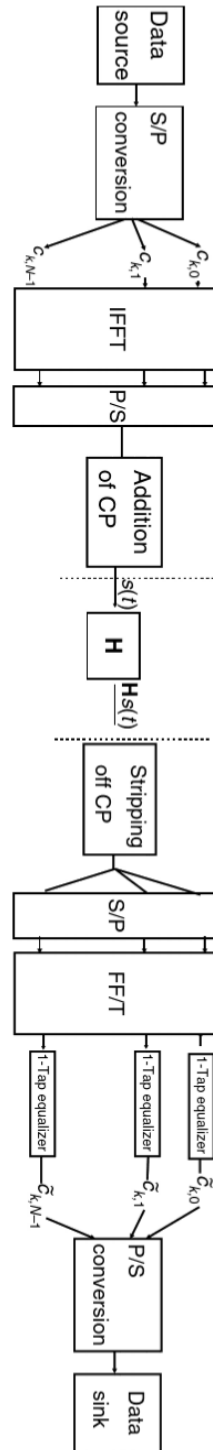


Figura 2.10 Implementación de una modulación OFDM [4].

Además, los estándares de comunicaciones móviles van fijando parámetros sobre la OFDM, como el concepto de bloque de recursos (RB), $\Delta f = 15$ kHz o la duración del prefijo cíclico (CP). En el paso de LTE a LTE-A se desbloquean modulaciones con mayores capacidades y se permite emplear CA. En la Release 15 (5G-NR) aparece el concepto de numerología, en la que se permite cambiar la separación entre portadoras según la expresión:

$$\Delta f = 2^\mu 15 \text{kHz} \quad \text{con } \mu = 0, 1, 2, 3, 4. \quad (2.19)$$

Si bien las ventajas de la OFDM son muchas, presenta un gran reto a la hora de pasar por la etapa amplificadora. Si se define la relación potencia pico a potencia promedio (PAPR) como

$$\text{PAPR}[\text{dB}] = 20 \log_{10} \left(\frac{\max |x[k]|}{\sqrt{\frac{1}{N} \sum_{n=1}^N |x[k]|^2}} \right), \quad (2.20)$$

se identifica que el problema que tienen en general las señales OFDM es que presentan una PAPR muy elevada. Esto quiere decir que hay valores que se alejan mucho de la media de potencia de señal. Como se desea aprovechar al máximo la potencia del amplificador es habitual trabajar con potencias muy cercanas a la saturación. De modo que, el hecho de que existan instantes en los que la señal toma un valor muy alto provocan que el amplificador trabaje en saturación, lo que en definitiva, distorsiona la señal.

3 Modelado y linealización

La ciencia nos enseña, en efecto, a someter nuestra razón a la verdad y a conocer y a juzgar las cosas como ellas son; es decir, como ellas quieren ser, y no como nosotros queremos que ellas sean.

MIGUEL DE UNAMUNO, 1913

Modelar un sistema real es un problema con grandes precedentes históricos. La idea consiste en generar una serie de expresiones matemáticas que se ajusten lo más fielmente posible a los fenómenos que se observan en la realidad. El modelo debe procurar ser generalista, en el sentido de que debe adaptarse correctamente a cualquier condición que pueda ser probada en la realidad sobre el sistema en cuestión. Además conocer el modelo es la mejor forma de entender cómo funciona la entidad modelada, pudiendo así predecir su comportamiento y aprovechar al máximo sus prestaciones.

Por ende, con el objetivo de conseguir la linealización de un sistema que a priori es no lineal, será necesario generar un modelo que se ajuste bien a su comportamiento para así poder tomar decisiones que contrarresten las componentes no deseadas.

3.1 Modelos basados en series de Volterra

En el Capítulo 2 ya se apuntaba un tipo de modelo de un sistema no lineal. La mejor forma de entender este tipo de aproximaciones es tener presente la filosofía que se aplica cuando se aproxima una función mediante las series de Taylor. El objetivo final es expresar una función complicada como una suma de polinomios, mucho más sencillos y manejables. De hecho, si se tratase de un sistema que únicamente es no lineal, la serie de Taylor sería una aproximación posible. Nótese que la ecuación de la Figura 2.5 se puede interpretar como

$$f(t) = \sum_{p=1}^{\infty} \frac{f^{(p)}(\tau)}{p!} (t - \tau)^p, \quad (3.1)$$

siendo $f^{(p)}(\tau)$ la derivada de orden p evaluada en τ . Haciendo la transformación $t \rightarrow s = x(t)$, agrupando el término constante, y acortando el sumatorio a P términos queda

$$f(x(t)) \approx \sum_{p=1}^P a_p x(t - \tau)^p, \quad (3.2)$$

pudiendo escribir la salida del sistema no lineal como

$$y_{NL}(t) \approx a_1 x(t - \tau) + a_2 x(t - \tau)^2 + \dots + a_P x(t - \tau)^P, \quad (3.3)$$

lo que es equivalente al modelo de Wiener con un filtro lineal igual a la unidad [5]. Este razonamiento es similar al adoptado en [6].

En este caso existe un matiz que hace que modelar el PA no sea abordable empleando esta aproximación, y es el hecho de que, si bien se trata de un sistema no lineal, también presenta memoria. Esto quiere decir que la respuesta que tenga el dispositivo ante una entrada en un instante t_1 depende de las entradas en instantes anteriores. Por otro lado, hay que tener en cuenta que se conoce parte del sistema que se va a modelar, ya que un amplificador al fin y al cabo sigue un diseño electrónico cuyos componentes son conocidos.

Ya en el siglo XIX Vito Volterra formuló una serie capaz de aproximar una función con efectos de memoria, y no fue hasta 1942 cuando Norbert Wiener comenzase a aplicar esta herramienta tan potente con un enfoque circuital. Esta estimación se conoce como serie de Volterra y fue capaz de revolucionar el análisis de sistemas, tratándose de un objeto de estudio que sigue vigente en la actualidad. Así pues, siguiendo el estilo de las series de Taylor se puede expresar la serie de Volterra como

$$y(t) = \sum_{p=1}^{\infty} y_p(t), \quad (3.4)$$

según [5], donde cada término denominado funcional sigue la forma

$$y_p(t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} h_p(\tau_1, \tau_2, \dots, \tau_p) x(t - \tau_1) x(t - \tau_2) \dots x(t - \tau_p) d\tau_1 d\tau_2 \dots d\tau_p. \quad (3.5)$$

De manera que los efectos de memoria están contenidos en los sucesivos retrasos τ_1, \dots, τ_p . La expresión (3.5) contiene infinitas combinaciones entre los retrasos de la señal $x(t)$, denominados regresores. No se puede perder de vista que la serie de Volterra se genera teniendo en cuenta el comportamiento de distintas entidades circuitales, de modo que cada regresor representa una aportación de la respuesta del circuito. Por otro lado, las ponderaciones $h(\tau_1, \tau_2, \dots, \tau_p)$ representan el *kernel* de la serie.

Para que (3.5) sea manejable y útil, requiere ser discretizada y truncada. Así pues, definiendo P como el orden máximo y Q como la profundidad de memoria, es decir, la cantidad de retrasos de la señal a tener en cuenta, se reescribe la serie de Volterra como

$$y[k] = \sum_{p=1}^{\infty} y_p[k] \approx \sum_{p=1}^P \sum_{q_1=0}^Q \dots \sum_{q_p=0}^Q h_p[q_1, \dots, q_p] \prod_{j=1}^p x[k - q_j]. \quad (3.6)$$

Nótese que la p en el productorio está en minúsculas, lo que significa que este límite variará en función del índice del funcional. También cabe destacar que a partir de ahora los retrasos q_p ya no indican tiempo, sino que son números enteros, ya que la señal ahora es discreta y la memoria se indica en muestras, no en tiempo. Además a partir de ahora se empleará = en lugar de \approx , por simplicidad en la notación, ya que se entiende que al haber truncado la serie se trata de una aproximación.

Debe tenerse en cuenta que la ecuación (3.6) está escrita para señales reales, por lo que se denomina comúnmente serie de Volterra en RF (radiofrecuencia). Teniendo en cuenta que puede ser deseable trabajar con muestras de la envolvente compleja de la señal en banda base, sería conveniente reescribir (3.6) de la siguiente forma según [7]:

$$y[k] = \sum_{m=0}^M \sum_{q_1=0}^Q \dots \sum_{q_{2m+1}=0}^Q h_{2m+1}[q_1, \dots, q_{2m+1}] \prod_{j=1}^{m+1} x[k - q_j] \prod_{j=m+2}^{2m+1} x^*[k - q_j], \quad (3.7)$$

donde $p = 2m + 1$ es el orden impar p -ésimo de no linealidad. Es importante no perder de vista que las señales ahora están en banda base y con el fin de no llegar confusión en cuanto al significado de $x[k]$ y $y[k]$ se hace la siguiente aclaración: siempre que se mencione que el modelo es en banda base, o se encuentren términos del estilo $|x[k - q_j]|^p$ se debe entender que se están empleando las muestras de la envolvente compleja, en lugar de la señal en RF. Por último, en [8] se justifica que es posible emplear también los órdenes p pares.

Llegados a este punto sería necesaria una modificación en la expresión (3.6) para acercar el modelo a una interpretación circuital. Con esto se quiere decir que, si bien el modelo surge por sí solo de la intención de modelar un sistema no lineal con memoria, la interpretación de cómo afecta cada término a la integridad de la señal es un tanto ardua. Es por esto que se propone una modificación en la realización de los sumatorios con fin de alcanzar una cierta semejanza con la teoría de sistemas. Así pues, siguiendo el razonamiento de [9], si se emplea una representación con coordenadas diagonales donde

$$n_l = q_{l+1} - q_l, \quad l = 1, 2, \dots, p - 1, \quad (3.8)$$

puede reescribirse cada funcional de serie de Volterra como:

$$y_p[k] = \sum_{n_1=-Q}^Q \cdots \sum_{n_{p-1}=-Q}^Q g_{n_1, \dots, n_{p-1}}^p[k] * v_{n_1, \dots, n_{p-1}}[k], \quad (3.9)$$

donde se define la nueva señal

$$v_{n_1, \dots, n_{p-1}}[k] = x[k] \prod_{l=1}^{p-1} x[n - n_l]. \quad (3.10)$$

Además * denota convolución unidimensional y

$$g_{n_1, \dots, n_{p-1}}^p[k] = h_p[k, k + n_1, \dots, k + n_{p-1}]. \quad (3.11)$$

es un filtro lineal determinado por h_p . De manera que el funcional $y_p[k]$ está compuesto por la superposición del filtrado lineal de las señales $v_{n_1, \dots, n_{p-1}}[k]$, tal y como se muestra en la Figura 3.1.

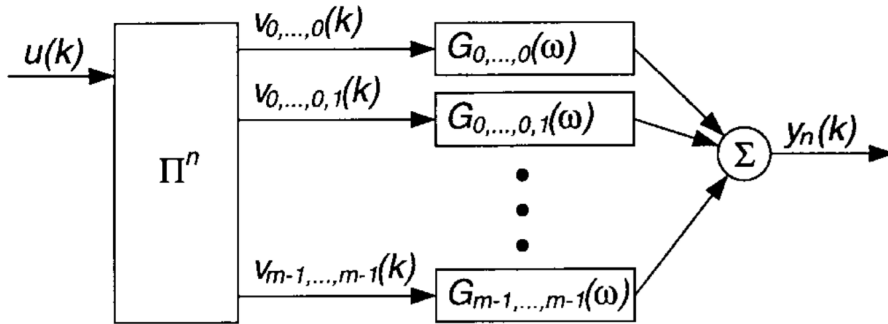


Figura 3.1 Descomposición del funcional $y_p[k]$ como superposición de filtrados lineales. En la figura $u(k)$ representa $x[k]$, y $m - 1$ es equivalente a Q . $G_{n_1, \dots, n_{p-1}}(\omega)$ es la transformada de Fourier en tiempo discreto de $g_{n_1, \dots, n_{p-1}}^p[k]$ [9].

Así pues, la expresión (3.9) permite entender el sistema como una no linealidad seguida de un conjunto de filtros lineales. Este enfoque, entre otras cosas, permite no perder de vista el hecho de que la señal de entrada se convoluciona con el sistema para obtener la señal de salida. De manera que tal y como muestra la Figura 3.1, la expresión (3.6) es equivalente a un conjunto de convoluciones unidimensionales, constituyendo por tanto, una convolución p -dimensional.

Por último se debe incidir en que se ha llegado a la expresión (3.6) por necesidad, para así poder trabajar con el modelo en software o implementarlo en placas programables como las *Field-programmable gate array* (FPGA). Este modelo se conoce como *Full Volterra* (FV), y aun tratándose de una operación digital, la serie de Volterra obtenida es demasiado extensa, por lo que la carga computacional sería excesiva. De manera que este modelo requiere de simplificaciones que lo hagan más funcional.

3.1.1 Memory Polynomial (MP)

Cabe destacar que, de la ecuación (3.3) (*memoryless*) a (3.6) (FV), se pasa de una aproximación pobre a una con demasiados términos. Así pues, una consideración que se efectúa sobre el modelo completo es plantear la hipótesis de que los regresores de mayor interés son aquellos que tienen el mismo retraso en todos sus factores. Este resultado es fruto de suponer banda estrecha sobre el modelo generalizado de Hammerstein [5]:

$$y_{GH}[k] = \sum_{p=1}^P \sum_{q=0}^Q h_p[q] x[k - q]^p, \quad (3.12)$$

donde el hecho de suponer banda estrecha permite afirmar que las combinaciones de interés son aquellas en las que los factores reales y complejos se diferencian en uno, es decir, tienen la forma $x[k]|x[k - q]|^{p-1}$ para

p impar, llegando así a la aproximación en banda base similar según [10] a la de Kim y Konstantinou [11]:

$$y_{MP}[k] = \sum_{p=1}^P \sum_{q=0}^Q h_p[q] x[k-q] |x[k-q]|^{p-1}, \quad (3.13)$$

donde el vector de memorias $[q_1, \dots, q_p]$ ha quedado reducido a una sola componente $[q]$. Obsérvese que según [8] puede usarse p par.

3.1.2 Generalized Memory Polynomial (GMP)

Aunque el modelo MP ha sido ampliamente utilizado a principios de siglo, se trata de un modelo demasiado restringido. El hecho de suponer que sólo existen componentes no lineales con un retraso común se aleja en cierta medida de la realidad. Así pues, en [5] se propone la inclusión de términos cruzados, definiendo el modelo GMP como

$$\begin{aligned} y_{GMP}[n] = & \sum_{k \in \mathcal{K}_a} \sum_{l \in \mathcal{L}_a} a_{kl} x[n-l] |x[n-l]|^k \\ & + \sum_{k \in \mathcal{K}_b} \sum_{l \in \mathcal{L}_b} \sum_{m \in \mathcal{M}_b} b_{klm} x[n-l] |x[n-l-m]|^k \\ & + \sum_{k \in \mathcal{K}_c} \sum_{l \in \mathcal{L}_c} \sum_{m \in \mathcal{M}_c} c_{klm} x[n-l] |x[n-l+m]|^k, \end{aligned} \quad (3.14)$$

donde se ha cambiado la notación, siendo en este momento n el índice de muestra. Las letras \mathcal{K} , \mathcal{L} y \mathcal{M} representan conjuntos, dejando a libre elección la eliminación de aquellos términos de la serie que se consideren menos relevantes. Así pues, los términos que forman parte del conjunto $\mathcal{K}_a \cup \mathcal{L}_a$ representan justamente el modelo MP, el conjunto $\mathcal{K}_b \cup \mathcal{L}_b \cup \mathcal{M}_b$ son regresores compuestos por la combinación de la envolvente actual con la envolvente retrasada y el conjunto $\mathcal{K}_c \cup \mathcal{L}_c \cup \mathcal{M}_c$ la combina con su adelanto.

3.1.3 Modelos bidimensionales

Si bien los modelos anteriores están formulados para estudiar el comportamiento del amplificador ante una entrada con una frecuencia concreta, no siempre es el escenario de funcionamiento. Además, con la irrupción de CA es habitual tener modulación en más de una banda, de modo que en esta sección se plantea un modelado para una amplificación en doble banda concurrente. Este escenario se puede asemejar a una prueba de dos tonos en la que las señales presentan un ancho de banda y están más distanciadas en frecuencia, tal y como se presenta en la Figura 3.2.

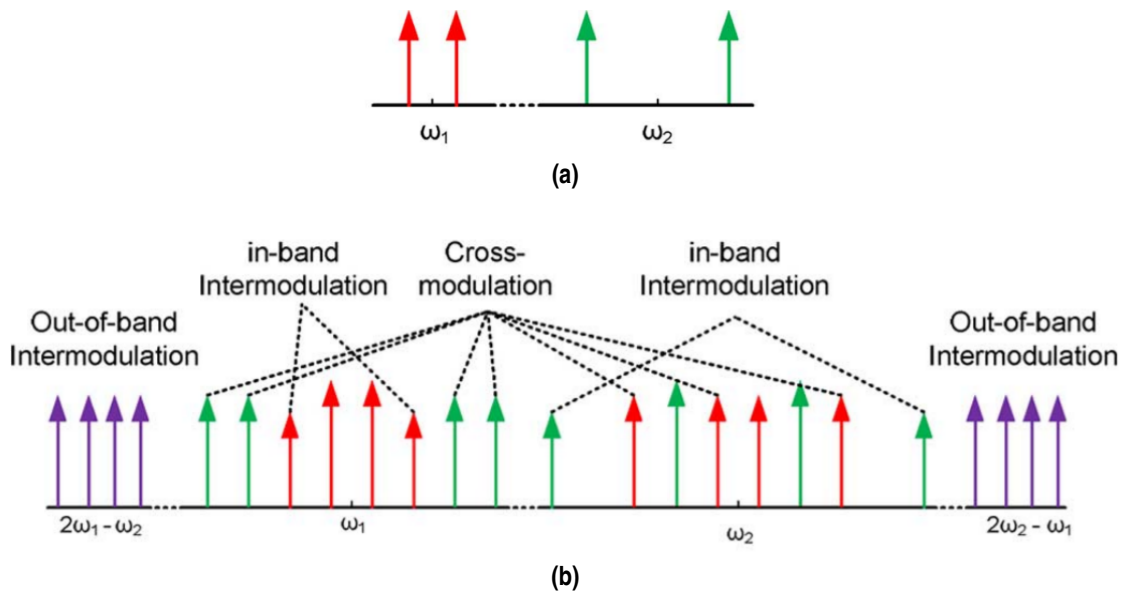


Figura 3.2 Espectro de la señal en: (a) la entrada y (b) la salida de un transmisor dual. [12].

Así pues, Bassam [12] desarrolla un modelo a partir del MP que tiene en cuenta dos señales de entrada separadas en frecuencia un cierto *offset*, llegando a la expresión

$$\begin{aligned} y_1[n] &= \sum_{p=0}^Q \sum_{r=0}^P \sum_{q=0}^R h_{prq}^{(1)} x_1[k-q] |x_1[k-q]|^{p-r} |x_2[k-q]|^r \\ y_2[n] &= \sum_{p=0}^Q \sum_{r=0}^P \sum_{q=0}^R h_{prq}^{(2)} x_2[k-q] |x_2[k-q]|^{p-r} |x_1[k-q]|^r. \end{aligned} \quad (3.15)$$

Este modelo se denominará 2D-MP o MP bidimensional. Como ya se ha comentado en la Sección 3.1.2, el modelo MP es un tanto pobre. Es por esto que se puede hacer la misma extensión de regresores sobre el modelo bidimensional, formulando así el 2D-GMP [13] como:

$$\begin{aligned} y_1[n] &= \sum_{k \in \mathcal{H}_a^{IB}} \sum_{l \in \mathcal{L}_a^{IB}} a_{kl}^{IB} x_1[n-l] |x_1[n-l]|^k \\ &+ \sum_{k \in \mathcal{H}_b^{IB}} \sum_{l \in \mathcal{L}_b^{IB}} \sum_{m \in \mathcal{M}_b^{IB}} b_{klm}^{IB} x_1[n-l] |x_1[n-l-m]|^k \\ &+ \sum_{k \in \mathcal{H}_c^{IB}} \sum_{l \in \mathcal{L}_c^{IB}} \sum_{m \in \mathcal{M}_c^{IB}} c_{klm}^{IB} x_1[n-l] |x_1[n-l+m]|^k \\ &+ \sum_{k \in \mathcal{H}_a^{CB}} \sum_{l \in \mathcal{L}_a^{CB}} a_{kl}^{CB} x_1[n-l] |x_2[n-l]|^k \\ &+ \sum_{k \in \mathcal{H}_b^{CB}} \sum_{l \in \mathcal{L}_b^{CB}} \sum_{m \in \mathcal{M}_b^{CB}} b_{klm}^{CB} x_1[n-l] |x_2[n-l-m]|^k \\ &+ \sum_{k \in \mathcal{H}_c^{CB}} \sum_{l \in \mathcal{L}_c^{CB}} \sum_{m \in \mathcal{M}_c^{CB}} c_{klm}^{CB} x_1[n-l] |x_2[n-l+m]|^k \\ &+ \sum_{k_1 \in \mathcal{H}_a^{MX1}} \sum_{k_2 \in \mathcal{H}_a^{MX2}} \sum_{l \in \mathcal{L}_a^{MX}} a_{k_1 k_2 l}^{MX} x_1[n-l] |x_1[n-l]|^{k_1} |x_2[n-l]|^{k_2} \end{aligned} \quad (3.16)$$

$$\begin{aligned} y_2[n] &= \sum_{k \in \mathcal{H}_a^{IB}} \sum_{l \in \mathcal{L}_a^{IB}} a_{kl}^{IB} x_2[n-l] |x_2[n-l]|^k \\ &+ \sum_{k \in \mathcal{H}_b^{IB}} \sum_{l \in \mathcal{L}_b^{IB}} \sum_{m \in \mathcal{M}_b^{IB}} b_{klm}^{IB} x_2[n-l] |x_2[n-l-m]|^k \\ &+ \sum_{k \in \mathcal{H}_c^{IB}} \sum_{l \in \mathcal{L}_c^{IB}} \sum_{m \in \mathcal{M}_c^{IB}} c_{klm}^{IB} x_2[n-l] |x_2[n-l+m]|^k \\ &+ \sum_{k \in \mathcal{H}_a^{CB}} \sum_{l \in \mathcal{L}_a^{CB}} a_{kl}^{CB} x_2[n-l] |x_1[n-l]|^k \\ &+ \sum_{k \in \mathcal{H}_b^{CB}} \sum_{l \in \mathcal{L}_b^{CB}} \sum_{m \in \mathcal{M}_b^{CB}} b_{klm}^{CB} x_2[n-l] |x_1[n-l-m]|^k \\ &+ \sum_{k \in \mathcal{H}_c^{CB}} \sum_{l \in \mathcal{L}_c^{CB}} \sum_{m \in \mathcal{M}_c^{CB}} c_{klm}^{CB} x_2[n-l] |x_1[n-l+m]|^k \\ &+ \sum_{k_2 \in \mathcal{H}_a^{MX2}} \sum_{k_1 \in \mathcal{H}_a^{MX1}} \sum_{l \in \mathcal{L}_a^{MX}} a_{k_2 k_1 l}^{MX} x_2[n-l] |x_2[n-l]|^{k_2} |x_1[n-l]|^{k_1}, \end{aligned}$$

donde IB denota *intra-band* y recoge distorsión de la señal consigo misma, CB significa *cross-band* referido a los términos de una banda que afectan a la otra y MX *mixed-band* que es una mezcla entre el efecto IB y CB. Por lo tanto el modelo 2D-GMP sería la versión generalizada del 2D-MP, ya que tiene en cuenta todos los aspectos que recoge el 2D-MP y además contempla desfases en el índice de memoria. De manera que el modelo de cada banda queda resumido en 18 conjuntos que siguen el mismo razonamiento que en el GMP.

3.1.4 Expresión en forma matricial

Es de gran utilidad a nivel de software e implementación poder definir el modelo de forma compacta mediante una matriz, ya que las operaciones matriciales han sido ampliamente desarrolladas y la mayoría de ellas

tienen una electrónica optimizada. Además, tomar este tipo de notación para expresar las series anteriormente mencionadas permite aplicar sobre ellas toda la teoría del álgebra. Para no perder generalidad se comenzará reescribiendo el modelo FV, el cual se recupera a continuación:

$$y[k] = \sum_{p=1}^{\infty} y_p[k] \approx \sum_{p=1}^P \sum_{q_1=0}^Q \dots \sum_{q_p=0}^Q h_p[q_1, \dots, q_p] \prod_{j=1}^p x[k - q_j]. \quad (3.17)$$

De partida, el argumento del conjunto de sumatorios presenta una problemática, y es que su dimensión va aumentando conforme lo hace p . Esto puede apreciarse observando algunas realizaciones, por ejemplo suponiendo $P = 2$ y $Q = 2$:

$$\begin{aligned} \sum_{p=1}^2 y_p[k] &= \sum_{q_1=0}^2 h_p[q_1, \dots, q_p] \prod_{j=1}^p x[k - q_j] + \sum_{q_1=0}^2 \sum_{q_2=0}^2 h_p[q_1, \dots, q_p] \prod_{j=1}^p x[k - q_j] \\ &= h_1[0]x[k] + h_1[1]x[k-1] + h_1[2]x[k-2] \\ &\quad + h_2[0,0]x[k]^2 + h_2[0,1]x[k]x[k-1] + h_2[0,2]x[k]x[k-2] \\ &\quad + h_2[1,0]x[k-1]x[k] + h_2[1,1]x[k-1]^2 + h_2[1,2]x[k-1]x[k-2] \\ &\quad + h_2[2,0]x[k-2]x[k] + h_2[2,1]x[k-2]x[k-1] + h_2[2,2]x[k-2]^2, \end{aligned} \quad (3.18)$$

donde puede verse que existen regresores repetidos que en general podrían estar ponderados por coeficientes diferentes. Como ya se comentaba anteriormente, tanto $h_p[q_1, \dots, q_p]$ como $\prod_{j=1}^p x[k - q_j]$, aumentan su dimensión con p . Pero la ventaja que presenta la serie es que, como puede verse en el ejemplo, la operación tiene forma de producto escalar. De manera que puede definirse el vector \mathbf{h}_p como:

$$\mathbf{h}_p = [h_1[0], \dots, h_1[Q], h_2[0,0], h_2[0,1], \dots, h_2[Q, Q], \dots, h_p[0,0, \dots, 0], \dots, h_p[Q, Q, \dots, Q]] \quad (3.19)$$

definiendo la matriz de regresores en función de sus columnas como:

$$\begin{aligned} \mathbf{X} &= [x[k], x[k-1], \dots, x[k-Q], x[k]x[k], x[k]x[k-1], \dots, x[k-Q]x[k-Q], \dots, \\ &\quad x[k] \cdot^{(P)} x[k], \dots, x[k-Q] \cdot^{(P)} x[k-Q]]. \end{aligned} \quad (3.20)$$

Extrapolando este desarrollo al GMP el resultado sería:

$$\mathbf{h} = [a[\mathcal{H}_a, \mathcal{L}_a], b[\mathcal{H}_b, \mathcal{L}_b, \mathcal{M}_b], c[\mathcal{H}_c, \mathcal{L}_c, \mathcal{M}_c]], \quad (3.21)$$

donde se efectúa la indexación de los conjuntos en su respectivo orden. Por otro lado, \mathbf{X} será de la forma:

$$\mathbf{X} = \begin{bmatrix} x[0] & x[-1] & \dots & x[-Q] & x[0]|x[0]|^2 & \dots & x[-Q]|x[-Q]|^2 & \dots \\ x[1] & x[0] & \dots & \vdots & x[1]|x[1]|^2 & \dots & \vdots & \dots \\ x[2] & x[1] & \dots & x[-1] & x[2]|x[2]|^2 & \dots & x[-1]|x[-1]|^2 & \dots \\ x[3] & x[2] & \dots & x[0] & x[3]|x[3]|^2 & \dots & x[0]|x[0]|^2 & \dots \\ \vdots & \vdots & & \vdots & \vdots & & \vdots & \\ \vdots & \vdots & & \vdots & \vdots & & \vdots & \\ x[N] & x[N-1] & \dots & x[N-Q] & x[N]|x[N]|^2 & \dots & x[N-Q]|x[N-Q]|^2 & \dots \end{bmatrix}, \quad (3.22)$$

siendo este un extracto de la matriz. Es importante destacar que, las muestras indicadas con un índice negativo son aquellas tomadas en los instantes previos a la muestra que se ha tomado como referencia de inicio, para así poder contemplar los efectos de la memoria. Por último, si se denomina a la cantidad de muestras como N , y al número de regresores como M , se tiene que $\mathbf{y} \in \mathbb{C}^{N \times 1}$, $\mathbf{X} \in \mathbb{C}^{N \times M}$ y que $\mathbf{h} \in \mathbb{C}^{M \times 1}$, quedando dimensionadas las entidades de interés. Desde este momento, \mathbf{X} se denominará matriz de regresores.

3.2 Identificación de coeficientes del modelo: Regresión

La regresión es el mecanismo mediante el cual se puede estimar la relación entre distintas variables. Si bien los modelos propuestos se ciñen bastante bien al comportamiento del amplificador, se debe asumir con total certeza que éstos nunca serán capaces de imitarlos a la perfección. Esta presunción se traduce en expresar el resultado del modelo como una estimación, es decir, teniendo presente que se comete un cierto error:

$$\mathbf{y} = \mathbf{X}\mathbf{h} + \mathbf{e}, \quad (3.23)$$

donde $\mathbf{e} \in \mathbb{C}^{N \times 1}$ representa el vector de error muestral, y el resto de entidades han sido definidas en la sección anterior. Esta aproximación tiene unas implicaciones muy fuertes sobre el modelo elegido, que pueden ser comprendidas desde un punto de vista algebraico. Así pues, la ecuación (3.23) debe entenderse como la proyección del vector \mathbf{y} en una base del subespacio $\text{Col}(\mathbf{X})$ donde las coordenadas de las componentes de dicha base están representadas en \mathbf{h} . Este enfoque será desarrollado en profundidad en el siguiente capítulo. El motivo de que se asuma un error \mathbf{e} viene impuesto por la elección del modelo, ya que no es posible modelar todas las componentes del vector \mathbf{y} y la predicción sobre cada componente también puede albergar un cierto error. De modo que, como se trata de una estimación, se emplea la siguiente notación:

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\mathbf{h}}, \quad (3.24)$$

donde $\hat{\mathbf{y}} = \mathbf{y} + \mathbf{e}$. El vector \mathbf{e} será por tanto una combinación entre las componentes de error dentro del espacio $\mathbb{C}^{N \times M}$ y las proyecciones de las componentes de error que quedan fuera del modelo sobre dicho subespacio, por lo que es deseable incluir una gran cantidad de regresores.

Así pues, el objetivo principal de la regresión consiste en encontrar el vector de coeficientes $\hat{\mathbf{h}}$ que haga mínimo el error cuadrático medio. La aproximación más intuitiva para abordar este problema es aplicar mínimos cuadrados (LS), es decir, minimizar la suma residual de cuadrados RSS. El resultado sería,

$$\hat{\mathbf{h}}_{LS} = \mathbf{X}^\dagger \mathbf{y}, \quad (3.25)$$

donde el subíndice LS denota que se ha empleado el estimador de mínimos cuadrados y \mathbf{X}^\dagger representa la pseudo-inversa de Moore-Penrose, que se obtiene como:

$$\mathbf{X}^\dagger = (\mathbf{X}^H \mathbf{X})^{-1} \mathbf{X}^H = (\mathbf{R})^{-1} \mathbf{X}^H, \quad (3.26)$$

y donde \mathbf{R} es la matriz de autocorrelaciones muestrales.

3.3 Predistorsión digital

Una vez desarrollado todo el modelado se presenta esta sección como nexo de unión entre los conceptos de modelado y linealización. Concretamente se trata de uno de los métodos capaces de conseguir linealizar un sistema, que no es más que contrarrestar la no linealidad del sistema. Así pues, la predistorsión digital, por sus siglas en inglés DPD (*Digital Predistortion*), es la técnica mediante la cual se emplea el modelo para actuar sobre la señal de entrada de manera que la respuesta conjunta sea equivalente a la de un sistema lineal. Dicha actuación tiene una representación directa en forma de sistema, tal y como se muestra en la Figura 3.3. Así pues, cuanto más preciso sea el modelo, mejor será la compensación de la no linealidad.

Se puede incidir en que históricamente se han ido desarrollando técnicas capaces de compensar los efectos de la no linealidad en los sistemas. Concretamente la ecualización fue el más extendido para mitigar los efectos de la propagación multicamino en el canal móvil. Pero este escenario es distinto, ya que, si bien el amplificador hace las veces de canal en cuanto a la distorsión se refiere, esta etapa se encuentra en el propio transmisor. En consecuencia, si se plantea realizar la ecualización en el receptor, se estaría enviando al canal una señal afectada por el recrecimiento espectral, pudiendo provocar interferencias en los canales adyacentes. Además la ecualización es un concepto aplicado en general a una señal independientemente de su modulación. Es aquí donde se diferencia DPD, dando una solución específica y aprovechando al máximo las cualidades de la señal a transmitir, evitando además, que la señal transmitida perjudique a los canales adyacentes. Así pues, puede decirse que DPD es una de las posibles soluciones al problema de la linealización. De modo que existen distintas formas de implementar dicha solución en función del momento en el que se calculen los coeficientes del modelo.

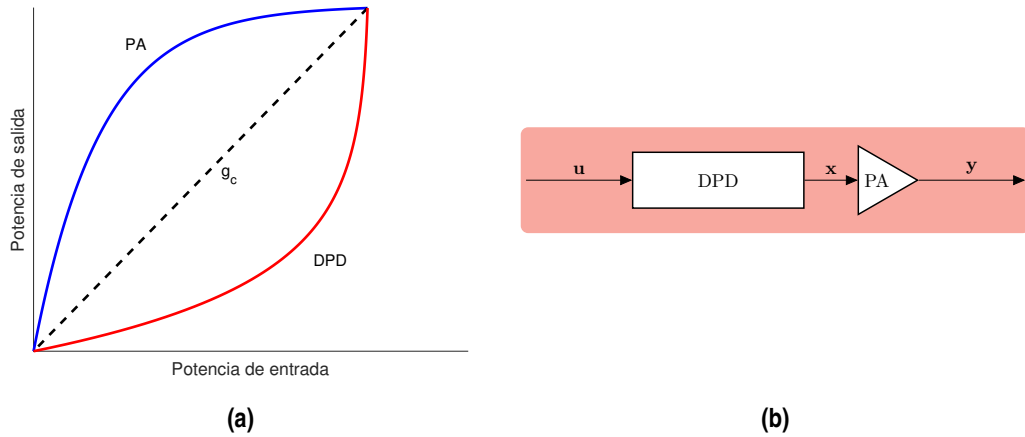


Figura 3.3 Esquema conceptual de aplicación del DPD.

Como nota aclaratoria el autor advierte de un cambio de notación en la explicación de las arquitecturas. El motivo es el hecho de que tradicionalmente se emplea la variable \mathbf{x} como entrada al amplificador, por lo que la señal generada pasará a llamarse \mathbf{u} . Así se puede decir que \mathbf{x} será la señal predistorsionada, es decir, $\mathbf{x} = \mathbf{u} - \mathbf{d}$, siendo \mathbf{d} la señal de distorsión. Por otro lado, la matriz de regresores se indicará con la letra que proceda en mayúsculas. Por último, el vector de coeficientes \mathbf{h} pasa a denotarse por \mathbf{w} porque en la literatura es común emplear la palabra *weights* (pesos) en lugar de coeficientes dentro de este ámbito.

3.3.1 Arquitectura de aprendizaje indirecto

Una de las formas posibles de realizar DPD es mediante una Arquitectura de Aprendizaje Indirecto (ILA), empleado en [14] y [10], entre otros. La idea consiste en emplear en la regresión las señales de entrada y salida del amplificador. La regresión simplemente establece la relación entre las variables \mathbf{x} e \mathbf{y} , siendo transparente en cuanto a la dirección de la relación. Es decir se puede estimar $\mathbf{y} \rightarrow f(\mathbf{x})$, esto es, el amplificador; o su inversa, es decir, $\mathbf{x} \rightarrow f^{-1}(\mathbf{y})$, empleándose esta última implicación en la etapa de entrenamiento. De manera que los coeficientes que se obtienen en la etapa de entrenamiento son ciertamente los de un sistema postdistorsionador, los cuales se copian en el bloque DPD, tal y como se muestra en la Figura 3.4. Se advierte de que g_o representa la ganancia objetivo, que puede ser cualquier ganancia comprendida entre la ganancia lineal y g_c . Se elige una u otra en función de las prestaciones deseadas, de modo que si se requiere de una fuerte amplificación a la salida se usará una g_o próxima a la ganancia lineal, mientras que si se requiere una respuesta muy lineal se empleará una g_o cercana a g_c .

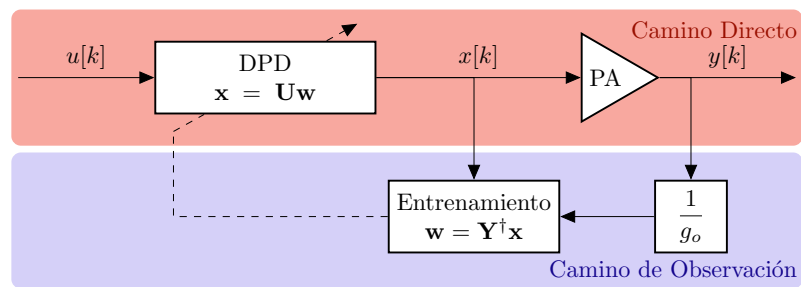


Figura 3.4 Implementación de DPD con arquitectura de aprendizaje indirecto.

Se debe tener en cuenta que en la primera etapa del aprendizaje el bloque DPD no tiene cargados los coeficientes del modelo, de modo que el primer entrenamiento se realiza con la señal $x[k] = u[k]$. Cabe apreciar que el hecho de que esta arquitectura sea indirecta reside en el orden a la hora de hacer las operaciones, es decir, primero se obtienen unos coeficientes diseñados para estimar la entrada del amplificador en función de su salida y después se dan como válidos para predistorsionar la señal de entrada. Así pues, si bien esta arquitectura funciona y ha sido ampliamente utilizada, no es óptima en cuanto a la obtención de un predistorsionador se refiere.

3.3.2 Arquitectura de aprendizaje directo

La Arquitectura de Aprendizaje Directo (DLA) vista en [15] nace en contraposición a ILA, de manera que si ILA centraba sus esfuerzos en conseguir una función que estimase $x[k]$ a partir de $y[k]$, DLA pretende obtener una relación en el camino directo. La forma que se plantea para conseguirlo es mediante la minimización del vector error, imponiendo que, si un sistema es puramente lineal, la única diferencia entre la señal de entrada y salida debe ser la ganancia del amplificador. La función de costes que rige la regresión consistirá justamente en eliminar el error mediante descenso del gradiente. Así pues, los pesos obtenidos en la regresión sirven para estimar el error. Dicha estimación, $\hat{\mathbf{e}} = \mathbf{U}\mathbf{w}_i$, se resta a la señal de entrada como se ve en la Figura 3.5.

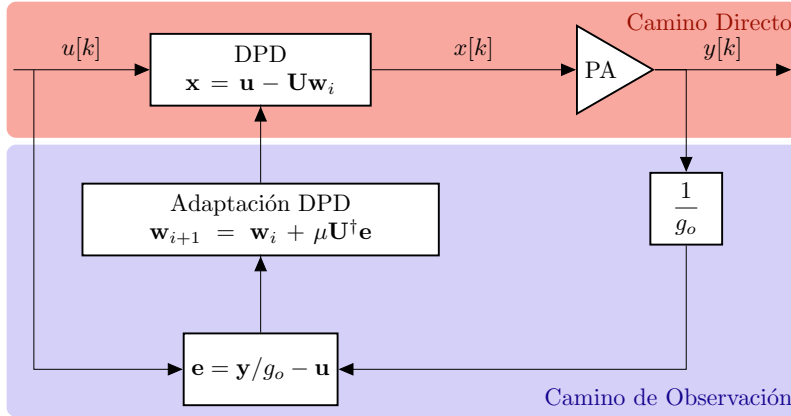


Figura 3.5 Implementación de DPD con arquitectura de aprendizaje directo.

En la figura puede apreciarse cómo el bucle de control va actualizando los pesos a cada iteración. La velocidad de aprendizaje viene regulada por la variable μ , cuyo valor mínimo es 0 y el máximo 1, siendo deseable que sea lo más cercano a 1 posible para obtener la convergencia más rápidamente. Dado que la corrección del error es progresiva, la actualización de los coeficientes se hace mediante la adición de los nuevos pesos a los antiguos, llamándose a esta etapa Adaptación DPD en la Figura 3.5. Por último cabe destacar que, en DLA, está estandarizado el uso de una ganancia objetivo igual a g_{avg} .

3.3.3 Arquitectura de aprendizaje directo para doble banda concurrente

Para conseguir que el esquema sea compatible con una transmisión en doble banda concurrente es necesario incluir un par de bloques. Hay que tener en cuenta que la regresión para este caso está planteada siguiendo el modelo de Enzinger [13], de modo que es conveniente adaptar la estructura para que admita dos entradas y sea capaz de evaluar dos salidas. Así pues, se muestra la arquitectura en la Figura 3.6.

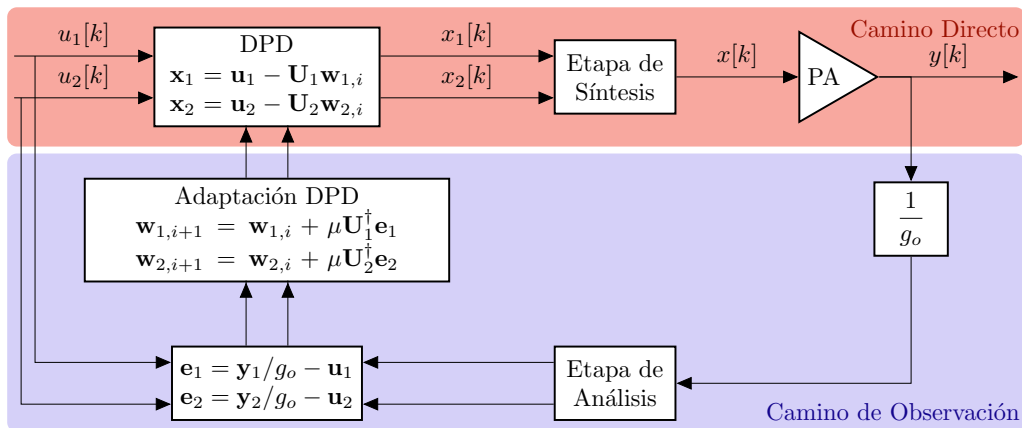


Figura 3.6 Implementación de DPD con arquitectura de aprendizaje directo para doble banda concurrente.

Véase que las señales $x_1[k]$ y $x_2[k]$ son muestras de la envolvente compleja de cada banda centradas en

banda base, y que la entrada al amplificador debe ser una señal $x[k]$ de doble banda y en RF. De modo que, en la etapa de síntesis, cada señal pasa por un modulador que convierte las muestras de la envolvente compleja de cada banda en dos señales en RF a frecuencias centrales diferentes y después las combina. Por otro lado, en la etapa de análisis se realiza un filtrado paso banda de cada banda, para después pasar estas señales que aún están en RF por un demodulador, cuya salida es la envolvente compleja de cada banda por separado. El orden de estas operaciones dependerá del hardware instalado.

4 Técnicas de reducción de complejidad y estabilización del modelo

Comprender que hay otros puntos de vista es el principio de la sabiduría.

THOMAS CAMPBELL, 1948

Como se ha comentado en el Capítulo 3, la regresión es el método a seguir para poder cuantificar la importancia de cada regresor. El gran número de componentes que se requieren para poder abordar este tipo de modelos, puede provocar que el estimador de LS no sea único. Esto es debido a la existencia de regresores muy correlados, o, lo que es lo mismo: columnas de \mathbf{X} muy parecidas entre sí, que provocan que la matriz de autocorrelaciones muestrales sea singular o cercana a serlo. Esta característica puede observarse en la Figura 4.1, donde los valores cercanos a 1 indican correlación alta. En otras palabras, el sistema presenta un número de condición muy elevado, de manera que ligeros cambios en \mathbf{X} desembocarían en grandes cambios en $(\mathbf{R})^{-1}$. Esto se traduce en que puede obtenerse una buena estimación de los coeficientes del modelo con buenas prestaciones en test pero malos para validación.

4.1 Identificación de problemas numéricos

Generalmente, es común el uso del número de condición para conocer si un sistema de ecuaciones es o no estable frente a las perturbaciones en el vector de entrada. Es posible relacionar el concepto de columnas correlacionadas con el número de condición. Se define el número de condición de una matriz \mathbf{A} invertible como [16]:

$$\text{cond}(\mathbf{A}) = \|\mathbf{A}^{-1}\| \|\mathbf{A}\|, \quad (4.1)$$

donde el operador $\|\cdot\|$ representa la norma de orden r . Realmente esta expresión no sirve para evaluar el condicionamiento de la matriz de regresores, ya que $\mathbf{X} \in \mathbb{C}^{N \times M}$, donde además en general $N \gg M$. En consecuencia es preciso abordar esta evaluación desde la matriz de autocorrelaciones muestrales \mathbf{R} . De modo que, como $\mathbf{R} = \mathbf{X}^H \mathbf{X}$ es cuadrada, simétrica, y semidefinida positiva se cumple que $\|\mathbf{R}\| \geq \rho(\mathbf{R}) = \max |\lambda_i(\mathbf{R})|$, $i = 1, \dots, r$, donde $\rho(\mathbf{R})$ es el radio espectral de la matriz \mathbf{R} . Así pues:

$$\text{cond}(\mathbf{R}) = \|\mathbf{R}\| \|\mathbf{R}^{-1}\| \geq \rho(\mathbf{R}) \rho(\mathbf{R}^{-1}) = \frac{\max_{1 \leq i \leq r} |\lambda_i(\mathbf{R})|}{\min_{1 \leq i \leq r} |\lambda_i(\mathbf{R})|}, \quad (4.2)$$

Este cálculo sólo sería cierto si \mathbf{R} es regular, ya que si no lo fuese contendría autovalores nulos, lo que provocaría que la expresión (4.2) quedase indeterminada, tendiendo a infinito.

Llegados a este punto, hay que tener presente que cuanto mayor es el número de condición, peor condicionado está el sistema. Existe una relación directa entre los autovalores de la matriz $\mathbf{X}^H \mathbf{X}$ y los valores singulares de la matriz \mathbf{X} que se indicará más adelante. Como análisis preliminar cabe preguntarse qué sucede al número de condición cuando existen en \mathbf{X} dos columnas muy parecidas entre sí o qué sucede si una de las columnas es cercana a ser nula. En el primero de estos casos es razonable pensar que si dos columnas

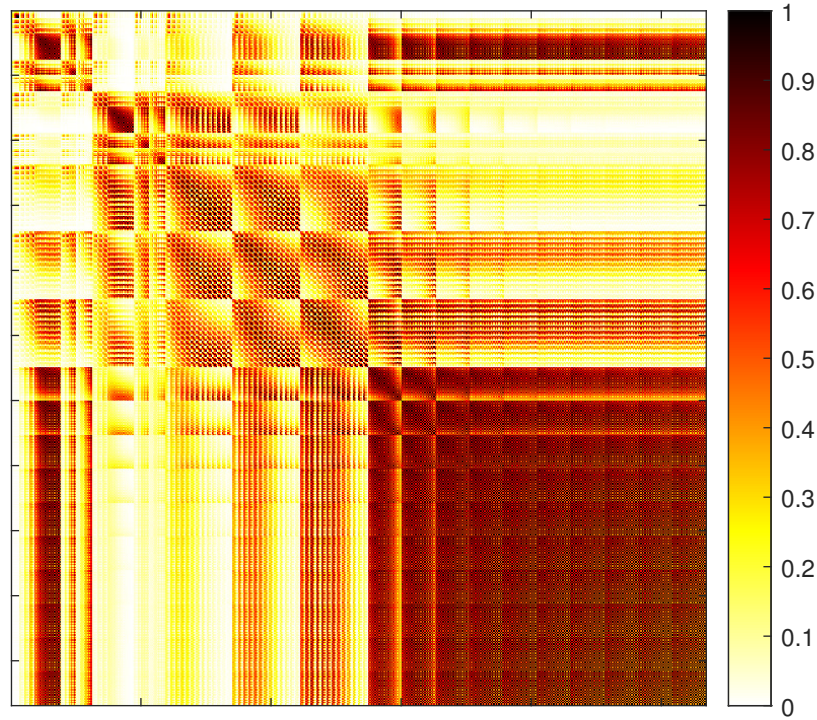


Figura 4.1 Matriz de autocorrelaciones muestrales para una de las bandas empleando un modelo 2D-GMP.

son muy parecidas entre sí, una de ellas es superflua y puede ser eliminada. En otras palabras puede ser sustituida por una columna de ceros. Esto provoca que aparezcan una fila y una columna nulas en \mathbf{R} , por lo que automáticamente \mathbf{R} tendrá un autovalor nulo. De modo que, en este caso el número de condición tiende a infinito y el sistema estaría muy mal condicionado. Lo mismo sucede con aquellas columnas de \mathbf{X} muy cercanas a 0.

4.2 Regularización de Ridge

Así pues, una de las soluciones que presenta la estadística ante problemas mal condicionados es la regularización de Ridge, centrada en solucionar el problema que introducen las columnas cercanas a 0. Esta operación abandona la idea de conseguir un estimador insesgado, siendo capaz de llegar a un compromiso entre el sesgo y la varianza en la estimación. Aunque a simple vista aumentar el sesgo parece perjudicial, se debe tener presente que se trata de un problema en el que se asume que no se puede superar una cierta exactitud en el resultado. Por lo tanto, desviar un poco la estimación para obtener mejor precisión a costa de la exactitud es una estrategia razonable.

El procedimiento a seguir consistiría en partir del estimador de mínimos cuadrados e incluir una penalización. Para evitar la inclusión de un término de sesgo se opera con las muestras de la matriz \mathbf{X} con media 0 [17] [18], y es por esto que desde este momento es lo mismo hablar de columnas casi nulas que de columnas poco variantes. Por su parte, el término constante se determinará si es preciso mediante la media de la estimación. Así pues, el objetivo ahora será minimizar la suma residual de cuadrados penalizada (PRSS) [17] cuyo desarrollo se expone a continuación:

$$\begin{aligned} \text{PRSS} &= \|\mathbf{y} - \hat{\mathbf{y}}\|_2^2 + \lambda \|\hat{\mathbf{h}}\|_2^2, = (\mathbf{y} - \hat{\mathbf{y}})^H (\mathbf{y} - \hat{\mathbf{y}}) + \lambda \hat{\mathbf{h}}^H \hat{\mathbf{h}} \\ &= \mathbf{y}^H \mathbf{y} - \hat{\mathbf{h}}^H \mathbf{X}^H \mathbf{y} - \mathbf{y}^H \mathbf{X} \hat{\mathbf{h}} + \hat{\mathbf{h}}^H \mathbf{X}^H \mathbf{X} \hat{\mathbf{h}} + \lambda \hat{\mathbf{h}}^H \hat{\mathbf{h}}, \end{aligned} \quad (4.3)$$

donde λ es el coeficiente de regularización. Como el objetivo es encontrar el mínimo de la función PRSS, a continuación, se deriva respecto de \mathbf{h} y se iguala a 0,

$$\frac{d\text{PRSS}}{d\mathbf{h}} = -2\mathbf{X}^H\mathbf{y} + 2\mathbf{X}^H\mathbf{X}\hat{\mathbf{h}} + 2\lambda\hat{\mathbf{h}} = 0, \quad (4.4)$$

y finalmente se despeja el modelo regularizado $\hat{\mathbf{h}}$:

$$\hat{\mathbf{h}}_{\text{Ridge}} = (\mathbf{X}^H\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}^H\mathbf{y}, \quad (4.5)$$

donde \mathbf{I} es la matriz identidad, y el subíndice *Ridge* indica que los coeficientes se han obtenido mediante la regresión penalizada.

En consecuencia, observando la ecuación (4.3), se extraen varios análisis. El primero de ellos es que la minimización del PRSS ya no consiste únicamente en aumentar los coeficientes $\hat{\mathbf{h}}_{\text{Ridge}}$ para disminuir $\|\mathbf{y} - \hat{\mathbf{y}}\|_2^2$, ya que ahora aparece un término de penalización regulado por λ que aumenta con $\|\hat{\mathbf{h}}_{\text{Ridge}}\|$. Esto puede interpretarse como una restricción en la norma ℓ_2 del vector de coeficientes. El segundo análisis se induce del primero, y es que el hecho de restringir $\|\hat{\mathbf{h}}_{\text{Ridge}}\|$ está sesgando la estimación, y esto asegura la invertibilidad de la matriz de regresores regularizada.

Para entender qué ha sucedido al introducir la penalización, es necesario realizar antes un pequeño desarrollo y plantear la descomposición en valores singulares (SVD). Esta descomposición es una aplicación del Teorema espectral para matrices simétricas reales [19], aplicable a cualquier matriz independientemente de si es cuadrada o no. De modo que, sea \mathbf{U} una matriz cuyas columnas ortogonales son los autovectores asociados a los autovalores no nulos de \mathbf{X} . Sea \mathbf{D} una matriz diagonal cuyas componentes son los valores singulares $[\sigma_1, \dots, \sigma_p]$ de la matriz de regresores. Se puede expresar \mathbf{X} como [17]:

$$\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^H, \quad (4.6)$$

donde $\mathbf{V} = \mathbf{X}^H\mathbf{U}\mathbf{D}^{-1}$, y $\mathbf{D} = [\sqrt{\lambda_1}, \dots, \sqrt{\lambda_p}]$. Así pues, aplicando (4.6) en (3.25) y multiplicando en ambas partes por \mathbf{X} por la izquierda:

$$\mathbf{X}\hat{\mathbf{h}}_{LS} = \mathbf{X}\mathbf{X}^\dagger\mathbf{y} = \mathbf{X}(\mathbf{X}^H\mathbf{X})^{-1}\mathbf{X}^H\mathbf{y} = \mathbf{U}\mathbf{D}\mathbf{V}^H(\underbrace{\mathbf{V}\mathbf{D}^H\mathbf{U}^H\mathbf{U}\mathbf{D}\mathbf{V}^H}_{\mathbf{I}})^{-1}\mathbf{V}\mathbf{D}^H\mathbf{U}^H\mathbf{y}, \quad (4.7)$$

dado que \mathbf{U} y \mathbf{V} son ortogonales, se cumple que $\mathbf{U}^H\mathbf{U} = \mathbf{V}^H\mathbf{V} = \mathbf{I}$. Además no se debe olvidar que \mathbf{U} y \mathbf{V} no son cuadradas, de modo que su inversa debe realizarse según (3.26). Teniendo esto en cuenta y aplicando que $(\mathbf{A}\mathbf{B})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$ se obtiene:

$$\mathbf{X}\hat{\mathbf{h}}_{LS} = \mathbf{U}\underbrace{(\mathbf{V}\mathbf{V}^H)^{-1}\mathbf{V}\mathbf{V}^H}_{\mathbf{I}}(\mathbf{D}^H\mathbf{D})^{-1}\underbrace{(\mathbf{V}^H\mathbf{V})^{-1}\mathbf{V}^H\mathbf{V}}_{\mathbf{I}}\mathbf{D}^H\mathbf{U}^H\mathbf{y} = \mathbf{U}\underbrace{\mathbf{D}\mathbf{D}^{-1}}_{\mathbf{I}}\underbrace{(\mathbf{D}^H)^{-1}\mathbf{D}^H}_{\mathbf{I}}\mathbf{U}^H\mathbf{y}, \quad (4.8)$$

finalmente:

$$\mathbf{X}\hat{\mathbf{h}}_{LS} = \mathbf{U}\mathbf{U}^H\mathbf{y}. \quad (4.9)$$

Si se compara (4.9) con (3.24) puede deducirse fácilmente que $\hat{\mathbf{y}} = \mathbf{U}\mathbf{U}^H\mathbf{y}$, tratándose de la aplicación de la proyección $\mathbf{P}_{\mathbf{X}}$ al vector \mathbf{y} [20]. Al comienzo de la sección se comentó que el objetivo de la regresión era proyectar cierto vector \mathbf{y} sobre una base del subespacio $\text{Col}(\mathbf{X})$ contenido en $\mathbb{C}^{N \times M}$. Por simplicidad será denominado espacio $\text{Col}(\mathbf{X})$, dando por hecho que está contenido en $\mathbb{C}^{N \times M}$. De modo que, identificando términos en (4.9), puede deducirse que $\mathbf{U}^H\mathbf{y}$ representa las coordenadas de la proyección sobre la base \mathbf{U} . Concretamente, \mathbf{U} es una de las posibles bases ortonormales del espacio columna de \mathbf{X} , llamada comúnmente base de componentes principales (PC) de \mathbf{X} [19]. Esta aproximación sirve para comprender qué sucede cuando se regulariza la regresión. Así pues, se puede deducir una expresión similar a (4.9) haciendo el mismo desarrollo partiendo de la expresión (4.5):

$$\mathbf{X}\hat{\mathbf{h}}_{\text{Ridge}} = \mathbf{X}(\mathbf{X}^H\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}^H\mathbf{y} = \mathbf{U}\mathbf{D}(\mathbf{D}^H\mathbf{D} + \lambda\mathbf{I})^{-1}\mathbf{D}\mathbf{U}^H\mathbf{y} = \sum_{j=1}^M u_j \frac{d_j^2}{d_j^2 + \lambda} u_j^H \mathbf{y}, \quad (4.10)$$

donde u_j es la columna j -ésima de la base \mathbf{U} y d_j el valor singular j -ésimo. Dado que $\hat{\mathbf{h}}_{\text{Ridge}}$ representa las coordenadas de la proyección en dicha base, el hecho de introducir la misma penalización en todas

las componentes de la diagonal de \mathbf{R} provoca una restricción mayor en aquellas componentes con menor variabilidad en $\hat{\mathbf{y}}$. Esto se deduce fácilmente observando que la penalización en la coordenada de la base \mathbf{U} no depende únicamente de λ , sino que también depende de d_j^2 , el autovalor λ_j asociado al autovector u_j . Un autovalor λ_j muy pequeño se traduce directamente en poca variabilidad en la base PC, de forma que, cuanto menor sea d_j^2 menor resulta ser $d_j^2/(d_j^2 + \lambda)$, por lo que se deduce finalmente que las componentes con menor variabilidad en la base PC son las más restringidas.

Esto último puede interpretarse como un método para la simplificación de la complejidad del sistema. Para cuantificarlo pueden definirse los grados de libertad efectivos como [17], [18]:

$$df(\lambda) = tr[\mathbf{X}(\mathbf{X}^H\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}^H] = tr[\mathbf{U}\mathbf{D}(\mathbf{D}^H\mathbf{D} + \lambda\mathbf{I})^{-1}\mathbf{D}\mathbf{U}^H], \quad (4.11)$$

donde tr significa la traza de la matriz, es decir, la suma de los elementos diagonales. Como la traza es conmutativa:

$$df(\lambda) = tr[(\mathbf{D}^H\mathbf{D} + \lambda\mathbf{I})^{-1}\underbrace{\mathbf{D}\mathbf{U}^H\mathbf{U}\mathbf{D}}_{\mathbf{I}}] = \sum_{j=1}^M \frac{d_j^2}{d_j^2 + \lambda}. \quad (4.12)$$

De modo que la ecuación (4.12) tiene una doble lectura, ya que no solo corrobora el hecho de que al incluir regularización se disminuye la complejidad del sistema, sino que también es posible calcular el λ necesario para conseguir que el sistema tenga df grados de libertad efectivos. Evidentemente, esta reducción de la complejidad está basada en la relevancia que los regresores \mathbf{X} con respecto a la base de componentes principales. El hecho de que el criterio de selección para la eliminación de componentes sea en este caso la variabilidad de la proyección de \mathbf{y} en la base PC puede ser un criterio poco acertado en lo que a este problema se refiere, ya que aparentemente, el motivo de la singularidad de \mathbf{X} reside con más relevancia en la correlación de sus regresores. Esto será discutido en secciones posteriores.

4.2.1 Matrices aumentadas

Aunque la ecuación (4.5) ha resultado de gran ayuda para entender cómo afecta la regularización a la regresión, es mejorable computacionalmente hablando. Dicha ecuación puede reescribirse como:

$$\hat{\mathbf{h}}_{Ridge} = (\tilde{\mathbf{X}}^H\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}^H\tilde{\mathbf{y}}, \quad (4.13)$$

donde $\tilde{\mathbf{X}}$ e $\tilde{\mathbf{y}}$ representan la matriz de coeficientes aumentada y la envolvente compleja de la salida respectivamente, esto es:

$$\tilde{\mathbf{X}} = \begin{bmatrix} \mathbf{X} \\ \sqrt{\lambda}\mathbf{I}_{(M \times M)} \end{bmatrix}, \tilde{\mathbf{y}} = \begin{bmatrix} \mathbf{y} \\ \mathbf{0}_{(M \times 1)} \end{bmatrix}, \quad (4.14)$$

donde $\mathbf{I}_{(M \times M)}$ es la matriz identidad con tamaño M y $\mathbf{0}_{(M \times 1)}$ es un vector de ceros de longitud M . Es sencillo demostrar que (4.13) es idéntica a (4.5). Así pues, la expresión (4.13) es preferida por ser más compacta y simple computacionalmente.

4.3 Métodos de selección de componentes

Otro tipo de técnica para reducir la complejidad más comúnmente usada es la selección de componentes. Si bien es cierto que cuantos más regresores se generen mejor se ajusta el modelo al identificar, no todos son adecuados. Seleccionar qué columnas de la matriz de regresores ofrecen los mejores resultados es una estrategia a seguir. Concretamente, para aplacar el problema de condicionamiento argumentado en la Sección 4.1 es conveniente que el criterio de selección esté enfocado en la correlación entre los regresores. Es por esto que a continuación se muestra el razonamiento para discernir qué columnas son mejores en base a sus correlaciones.

4.3.1 Orthogonal Matching Pursuit (OMP)

La idea principal recogida en el algoritmo OMP [21] consiste en ordenar la matriz de regresores mediante un método iterativo, el cual se detiene cuando se ha cumplido un requisito, gobernado por un algoritmo de parada [22] como el *Bayesian Information Criterion* (BIC) [23]. La selección de la mejor columna consiste en una correlación entre la matriz de regresores y un vector residuo, que es la diferencia entre la envolvente compleja de la salida y la envolvente compleja modelada por las componentes seleccionadas en la iteración.

Algoritmo 1 Orthogonal Matching Pursuit (OMP)**Input:** \mathbf{X}, \mathbf{y} **Output:** $S^{(\text{end})}, \hat{\mathbf{h}}^{(\text{end})}$

- 1: *Inicialización* : $\mathbf{r}^{(0)} \leftarrow \mathbf{y}, S^{(0)} \leftarrow \emptyset$
- 2: **for** $t = 1$ hasta que el criterio de parada se cumpla **do**
- 3: $\mathbf{g}_{\{i\}}^{(t)} \leftarrow \frac{\langle \mathbf{X}_{\{i\}}^H, \mathbf{r}^{(t-1)} \rangle}{\|\mathbf{X}_{\{i\}}\|_2} \mathbf{r}^{(t-1)}$
- 4: $s^{(t)} \leftarrow \arg \max_{i \notin S^{(t-1)}} |\mathbf{g}_{\{i\}}^{(t)}|$
- 5: $S^{(t)} \leftarrow S^{(t-1)} \cup s^{(t)}$
- 6: $\hat{\mathbf{h}}_{S^{(t)}} \leftarrow \mathbf{X}_{S^{(t)}}^\dagger \mathbf{y}$
- 7: $\hat{\mathbf{y}}^{(t)} \leftarrow \mathbf{X}_{S^{(t)}} \hat{\mathbf{h}}_{S^{(t)}}$
- 8: $\mathbf{r}^{(t)} \leftarrow \mathbf{y} - \hat{\mathbf{y}}^{(t)}$
- 9: **end for**

El hecho de que el producto $\mathbf{X}^H \mathbf{r}$ represente una correlación es debido a la estructura en forma de muestras de la señal $x[n]$ que posee la matriz de regresores y que ya ha sido comentada. De modo que OMP consigue ordenar la matriz de coeficientes de mayor a menor correlación con \mathbf{y} , ya que el residuo representa el resto de componentes que no están generadas por los regresores que ya han sido seleccionados en cada iteración. Sin embargo, en este método no se evalúa la correlación entre regresores.

4.3.2 Doubly Orthogonal Matching Pursuit (DOMP)

Para conseguir que la correlación entre regresores sea un factor a tener en cuenta a la hora de seleccionar una columna de \mathbf{X} , es necesario incluir una operación más en el algoritmo OMP. De esto se encarga el algoritmo DOMP [24] que incluye una aplicación del método de Gram-Schmidt tras haber seleccionado el regresor más importante en la iteración en base al OMP. Gram-Schmidt garantiza una base ortonormal del espacio columna de \mathbf{X} , cuyas componentes no están correladas. Además, tras la comparativa realizada en [25], destaca por mantener un rendimiento en linealización equivalente al de técnicas como OMP, *Subspace Pursuit* (SP) o *Compressive Sampling Matching Pursuit* (CoSaMP) con un menor número de coeficientes.

Algoritmo 2 Doubly Orthogonal Matching Pursuit (DOMP)**Input:** \mathbf{X}, \mathbf{y} **Output:** $S^{(\text{end})}, \hat{\mathbf{h}}^{(\text{end})}$

- 1: *Inicialización* : $\mathbf{r}^{(0)} \leftarrow \mathbf{y}, S^{(0)} \leftarrow \emptyset, \mathbf{Z}^{(0)} \leftarrow \mathbf{X}$
- 2: **for** $t = 1$ hasta que el criterio de parada se cumpla **do**
- 3: $\mathbf{g}_{\{i\}}^{(t)} \leftarrow \frac{\langle \mathbf{Z}_{\{i\}}^H, \mathbf{r}^{(t-1)} \rangle}{\|\mathbf{Z}_{\{i\}}\|_2} \mathbf{r}^{(t-1)}$
- 4: $s^{(t)} \leftarrow \arg \max_{i \notin S^{(t-1)}} |\mathbf{g}_{\{i\}}^{(t)}|$
- 5: $S^{(t)} \leftarrow S^{(t-1)} \cup s^{(t)}$
- 6: $\mathbf{p}^{(t)} \leftarrow \mathbf{Z}_{\{i^{(t)}\}}^{(t-1)H} \mathbf{Z}^{(t-1)}$
- 7: $\mathbf{Z}^{(t)} \leftarrow \mathbf{Z}^{(t-1)} - \mathbf{p}^{(t)} \otimes \mathbf{Z}_{\{i^{(t)}\}}^{(t-1)}$
- 8: $\hat{\mathbf{h}}_{S^{(t)}} \leftarrow \mathbf{X}_{S^{(t)}}^\dagger \mathbf{y}$
- 9: $\hat{\mathbf{y}}^{(t)} \leftarrow \mathbf{X}_{S^{(t)}} \hat{\mathbf{h}}_{S^{(t)}}$
- 10: $\mathbf{r}^{(t)} \leftarrow \mathbf{y} - \hat{\mathbf{y}}^{(t)}$
- 11: **end for**

Como puede intuirse del Algoritmo 2, \mathbf{Z} es una base ortogonal que proviene de ortogonalizar \mathbf{X} en orden decreciente de correlación con la señal a modelar. Es importante resaltar que en el paso 8 de este algoritmo se emplea \mathbf{X} y no \mathbf{Z} . Esto es debido a que \mathbf{Z} sólo se emplea como apoyo para encontrar la ortogonalidad entre los regresores. No obstante, sería posible emplear \mathbf{Z} en lugar de \mathbf{X} , de manera que los coeficientes $\hat{\mathbf{h}}_{LS}$ no representen pesos de las componentes de la serie de Volterra, sino que estén referidos a otra base enfocada más bien en la eliminación de las correlaciones de las columnas de \mathbf{X} . Esta base será denotada por base DOMP de \mathbf{X} , y será de suma importancia en las secciones posteriores.

4.3.3 Reduced Complexity Doubly Orthogonal Matching Pursuit (RC-DOMP)

Aunque el algoritmo DOMP es muy eficaz y consigue buenas prestaciones en cuanto a reducción de complejidad y estabilidad de las soluciones se refiere, tiene el problema de que es muy costoso computacionalmente. Concretamente el problema reside en la realización de la pseudo-inversa. Con fin de reducir el cómputo se desarrolla el *Reduced Complexity DOMP* (RC-DOMP) [26] que profundiza aún más en el significado del DOMP, encontrando una alternativa a la pseudo-inversa. El Algoritmo 3 está recogido en [27].

Algoritmo 3 Reduced Complexity Doubly Orthogonal Matching Pursuit (RC-DOMP).

Input: \mathbf{X}, \mathbf{y}

Output: $S^{(\text{end})}, \hat{\mathbf{h}}^{(\text{end})}$

- 1: *Inicialización* : $\mathbf{r}^{(0)} \leftarrow \mathbf{y}, S^{(0)} \leftarrow \emptyset, \mathbf{Z}^{(0)} = \mathbf{X}, \hat{\mathbf{y}}^{(0)} \leftarrow \mathbf{0}$,
 $\mathbf{R}_{\mathbf{Z}^{(0)}} = \mathbf{X}^H \mathbf{X}, \mathbf{R}_{\mathbf{Z}^{(0)}\mathbf{y}} = \mathbf{X}^H \mathbf{y}$
 - 2: **for** $k = 1$ hasta que el criterio de parada se cumpla **do**
 - 3: $s^{(k)} = \arg \max_{i \notin S^{(k)}} \frac{|\gamma_i|}{\sqrt{\rho_{ii}}}$
 - 4: $S^{(k)} = S^{(k-1)} \cup \{s^{(k)}\}$
 - 5: $\mathbf{T}^{(k)} = \left(t_{ij}^{(k)} \right) = \begin{cases} -\frac{\rho_{ij}}{\rho_{ii}} & i = s^{(k)}, j \notin S^{(k)} \\ \frac{1}{\sqrt{\rho_{ii}}} & i = j = s^{(k)} \\ 1 & i = j \neq s^{(k)} \\ 0 & \text{elsewhere} \end{cases}$
 - 6: $\mathbf{Z}^{(k)} = \mathbf{Z}^{(k-1)} \mathbf{T}^{(k)}$
 - 7: $\gamma_i^{(k)} = \gamma_i^{(k-1)} - \left(\frac{\rho_{s^{(k)}i}}{\rho_{s^{(k)}s^{(k)}}} \right)^* \gamma_{s^{(k)}}^{(k-1)}, i \notin S^{(k)}$
 - 8: $\gamma_i^{(k)} = \frac{\gamma_i^{(k-1)}}{\sqrt{\rho_{ii}}}, i = s^{(k)}$
 - 9: $\mathbf{R}_{\mathbf{Z}^{(k)}} = \mathbf{T}^{(k)H} \mathbf{R}_{\mathbf{Z}^{(k-1)}} \mathbf{T}^{(k)}$
 - 10: $\hat{\mathbf{h}}^{(k)} = \gamma_{S^{(k)}}$
 - 11: $\mathbf{r}^{(k)} = \mathbf{r}^{(k-1)} - \mathbf{Z}_{s^{(k)}}^{(k)} \gamma_{s^{(k)}}^{(k)}$
 - 12: $\hat{\mathbf{y}}^{(k)} = \hat{\mathbf{y}}^{(k-1)} + \mathbf{Z}_{s^{(k)}}^{(k)} \gamma_{s^{(k)}}^{(k)}$
 - 13: **end for**
-

Como ya se ha comentado, la ortogonalización respecto a la correlación se basa en aplicar Gram-Schmidt a las columnas de \mathbf{X} . Ahora bien, esto puede verse como un sistema de entrada-salida, donde la entrada es \mathbf{X} y la salida es la base ortonormal \mathbf{Z} . De modo que la idea consiste en condensar la operación de ortogonalización en una matriz de transformación $\mathcal{T} = \mathbf{T}^{(1)} \mathbf{T}^{(2)} \dots \mathbf{T}^{(n)}$, denominada también matriz de ortogonalización. Por otro lado, se puede entender que \mathcal{T} representa a su vez una base ortonormal del espacio columna de $\mathbf{X}^H \mathbf{X}$, análoga a la llamada base DOMP de \mathbf{X} , referida al espacio $\text{Col}(\mathbf{R})$.

Es conveniente hacer una apreciación sobre la matriz de correlaciones de \mathbf{X} , incidiendo en que se define como $\mathbf{R}_{\mathbf{X}} = \mathbb{E}[\mathbf{X}^H \mathbf{X}]$, cuyo estimador de máxima verosimilitud es la matriz de autocorrelaciones muestrales, que se ha denotado por \mathbf{R} . Esta implicación incumbe en la misma medida a \mathbf{Z} , donde directamente se ha representado $\mathbf{R}_{\mathbf{Z}}$ como la estimación de la matriz de correlación de \mathbf{Z} .

Por último, se comentaba que en DOMP sería posible plantear la pseudo-inversa con \mathbf{Z} en lugar de \mathbf{X} . Esto justamente se consigue con la correlación residual γ_i , que representa la correlación muestral cruzada de cada regresor con \mathbf{y} . Como la correlación de la columna k -ésima de \mathbf{X} con \mathbf{y} tiene distintas componentes en la nueva base \mathbf{Z} se debe eliminar esta correlación por separado aplicando de nuevo Gram-Schmidt, para que así cada componente del vector γ represente únicamente la ponderación en una dirección de la base DOMP.

4.4 Regularización de Ridge aplicada en el espacio de correlaciones

Hasta este punto sólo se han presentado las técnicas OMP, DOMP y RC-DOMP como métodos para seleccionar los regresores en función de las correlaciones. El hecho de conservar sólo aquellos regresores que produzcan un mejor comportamiento ya es una herramienta muy potente. Igualmente, en esta sección se plantea una combinación de la selección con la regularización. Si se tiene en cuenta que $\mathbf{Z} = \mathbf{X} \mathcal{T}$ es una

matriz ortonormal, se puede volver a escribir la ecuación de la proyección (4.9) como:

$$\mathbf{X}\hat{\mathbf{h}}_{LS}^{\mathcal{C}} = \mathbf{Z}\mathbf{Z}^H \mathbf{y}, \tag{4.15}$$

donde se ha querido denotar con \mathcal{C} que al tratarse de una base distinta a la de PC, la nueva base descompone las columnas de \mathbf{X} con componentes rotadas en el espacio. De modo que las coordenadas de la proyección de \mathbf{y} en la base \mathbf{Z} serán, en general, diferentes. Dado que las bases PC y \mathbf{Z} pertenecen al espacio generado por $Col(\mathbf{X})$ puede plantearse una proyección con \mathbf{Z} . Además, anteriormente se empleaba \mathbf{X} como el conjunto de vectores generador de la solución \mathbf{y} , que estaba referenciado a la base PC pero no era ortogonal. Es por esto que ahora se plantea emplear la propia matriz \mathbf{Z} como subconjunto sobre el que proyectar \mathbf{y} , ya que además de connotar una base incorrelada, es ortogonal. En otras palabras, se pretende cambiar el conjunto de proyección $Col(\mathbf{X})$ por $Col(\mathbf{Z})$, y es por esto que las coordenadas ya no tienen el significado de la serie de Volterra. De modo que se puede expresar (4.15) como:

$$\mathbf{Z}\hat{\mathbf{h}}_{LS}^{DOMP} = \mathbf{Z}\mathbf{Z}^H \mathbf{y}, \tag{4.16}$$

donde el superíndice *DOMP* denota que las coordenadas pertenecen a la base DOMP, \mathbf{Z} . Este movimiento es lógico, ya que se pretende penalizar la correlación, y las componentes menos variantes en \mathbf{Z} representan justamente las menos incorreladas. En la Figura 4.2 se ilustra graficamente el significado de las ecuaciones (4.9) y (4.16) tomando como ejemplo un hipotético espacio \mathbb{R}^3 para poder visualizarlo.

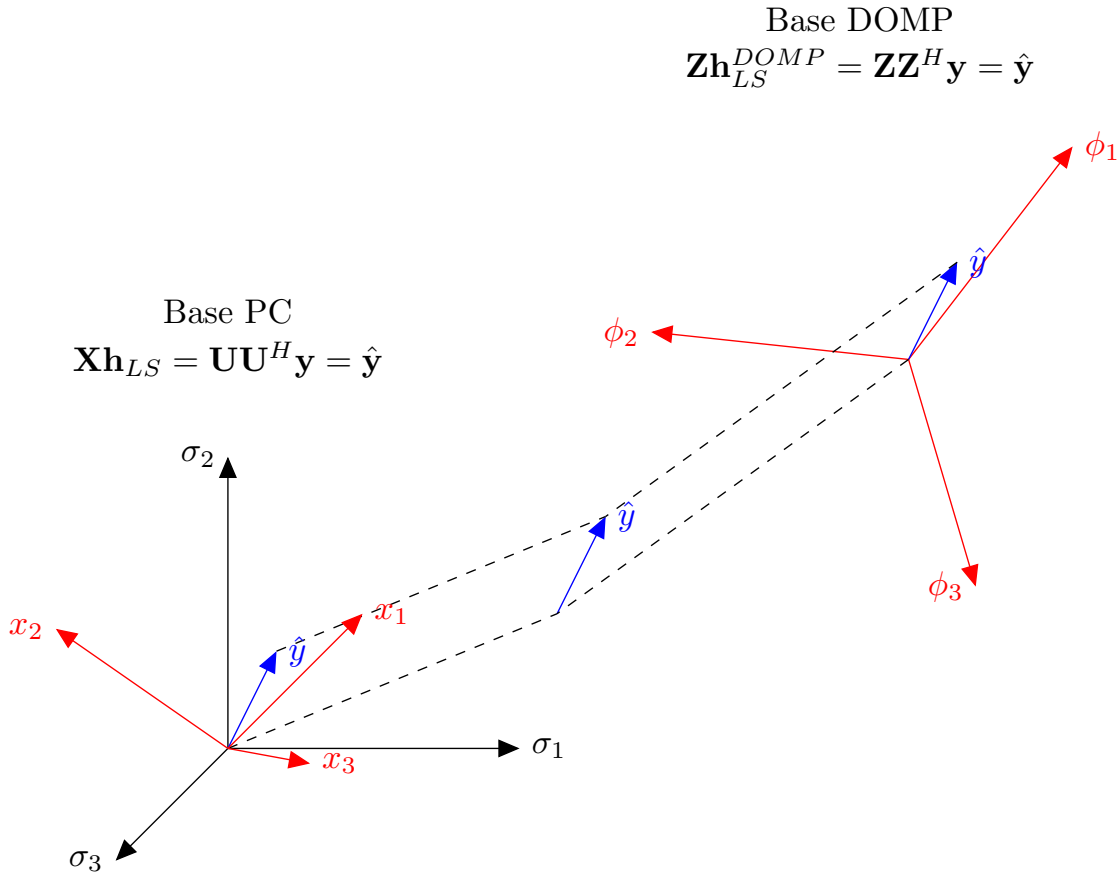


Figura 4.2 Representación geométrica de las ecuaciones de proyección. En rojo se indica la base sobre la que se proyecta $\hat{\mathbf{y}}$ en cada caso, σ_i representa la columna *i*-ésima de la matriz \mathbf{U} y x_i la columna *i*-ésima de la matriz \mathbf{X} .

Así pues, aplicando la regresión sobre la matriz \mathbf{Z} , los coeficientes estimados con LS son:

$$\hat{\mathbf{h}}_{LS}^{DOMP} = \mathbf{Z}^\dagger \mathbf{y} = (\mathbf{Z}^H \mathbf{Z})^{-1} \mathbf{Z}^H \mathbf{y}, \tag{4.17}$$

De modo que, si se regulariza y se multiplican ambos lados de la igualdad por \mathbf{Z} se obtiene:

$$\mathbf{Z}\hat{\mathbf{h}}_{Ridge}^{DOMP} = \mathbf{Z}(\mathbf{Z}^H\mathbf{Z} + \lambda\mathbf{I})^{-1}\mathbf{Z}^H\mathbf{y}, \quad (4.18)$$

y teniendo en cuenta que por ortogonalidad $\mathbf{Z}^H\mathbf{Z} = \mathbf{I}$ se llega finalmente a:

$$\mathbf{Z}\hat{\mathbf{h}}_{Ridge}^{DOMP} = \mathbf{Z}(\mathbf{I} + \lambda\mathbf{I})^{-1}\mathbf{Z}^H\mathbf{y} = \sum_{j=1}^M \phi_j \frac{1}{1+\lambda} \phi_j^H \mathbf{y}, \quad (4.19)$$

donde ϕ_j representa la columna j -ésima de \mathbf{Z} . La ecuación (4.19) se ha representado para comprender cómo se procede con las coordenadas en la base \mathbf{Z} , ya que no se puede olvidar que el algoritmo RC-DOMP evita la realización de la pseudo-inversa. La ecuación (4.19) se traduce en:

$$\hat{\mathbf{h}}_{Ridge}^{DOMP} = \frac{1}{1+\lambda} \hat{\mathbf{h}}_{LS}^{DOMP}. \quad (4.20)$$

Por lo tanto, la mejor forma de aprovechar los beneficios computacionales de RC-DOMP sin perder la referencia del espacio de Volterra es empleando la matriz de ortogonalización \mathcal{T} , tal y como se presenta en el Algoritmo 4.

Algoritmo 4 Aplicación de Ridge en el espacio ortogonal de las correlaciones

Input: \mathbf{X} , \mathcal{T} , \mathbf{y}

Output: $\hat{\mathbf{h}}_{Ridge}^{Volterra}$

- 1: *Ortogonalización* : $\mathbf{Z} = \mathbf{X}\mathcal{T}$
 - 2: *Cálculo de coeficientes en el espacio ortogonal de correlaciones* : $\hat{\mathbf{h}}_{LS}^{DOMP} = \mathbf{Z}^H\mathbf{y}$
 - 3: *Aplicación de Ridge como penalización en los coeficientes* : $\hat{\mathbf{h}}_{Ridge}^{DOMP} = \frac{1}{1+\lambda} \hat{\mathbf{h}}_{LS}^{DOMP}$
 - 4: *Traducción de los coeficientes al espacio de Volterra* : $\hat{\mathbf{h}}_{Ridge}^{Volterra} = \mathcal{T}\hat{\mathbf{h}}_{Ridge}^{DOMP}$
-

Por último es importante incidir en que la base DOMP es útil a la hora de actuar contra el efecto de las columnas correladas de \mathbf{X} , pero es preferible seguir entendiendo la matriz de regresores como funcionales de la serie de Volterra. Es por esto que en el último paso del algoritmo se vuelve a emplear la matriz de ortogonalización para traducir los coeficientes a la base PC.

5 Diseño experimental y resultados

Hay una gran diferencia entre conocer el camino y andar el camino.

MORFEO, EN MATRIX, 1999

Para poner a prueba la potencia de la teoría desarrollada es oportuno integrarla en un escenario real. Así pues, este capítulo comenzará con una descripción del entorno de trabajo, seguido de las especificaciones que definen el marco de calidad de la señal transmitida en el estándar 5G, que servirán de referencia para saber si los resultados son adecuados. Por último, se irán aplicando progresivamente las distintas técnicas presentadas, midiendo sus prestaciones.

5.1 Descripción del banco de medidas

Se ha empleado el laboratorio de Teoría de la Señal y Comunicaciones situado en la primera planta del bloque L-3 de los laboratorios de la Escuela Técnica Superior de Ingeniería (ETSI). El banco de medidas está compuesto por un generador vectorial de señal (VSG) SMU200A de Rohde & Schwarz, dos pre-amplificadores TVA-4W-422A+ de Minicircuits en cascada, un analizador vectorial de señal (VSA) PXA-N9030A de Keysight Technologies y una fuente de alimentación. El amplificador es de clase AB, se apoya en una placa de evaluación, y está basado en el HEMT de GaN CGH40010 de Cree, con una frecuencia de trabajo de 3.6 GHz. El generador dispone de una serie de instrucciones que pueden ser ejecutadas a través de MATLAB®, desde donde se configura la señal que pone a prueba al amplificador.

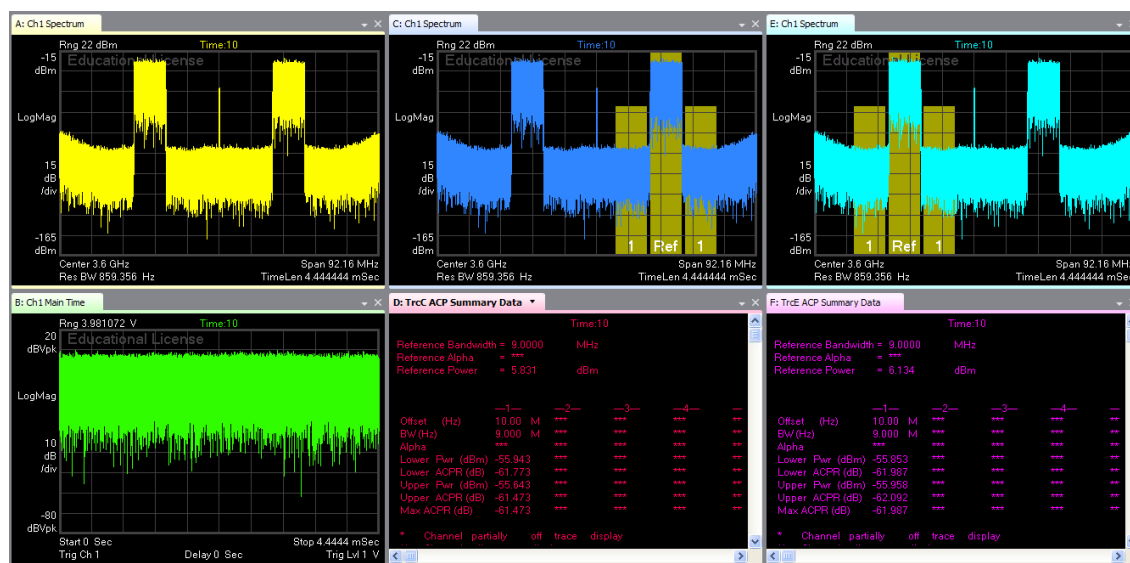
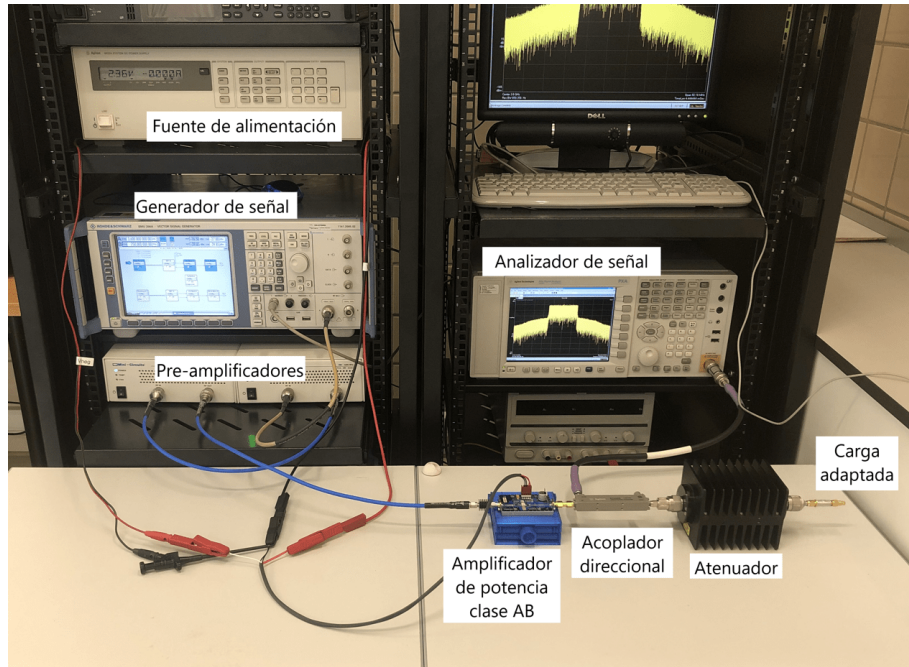
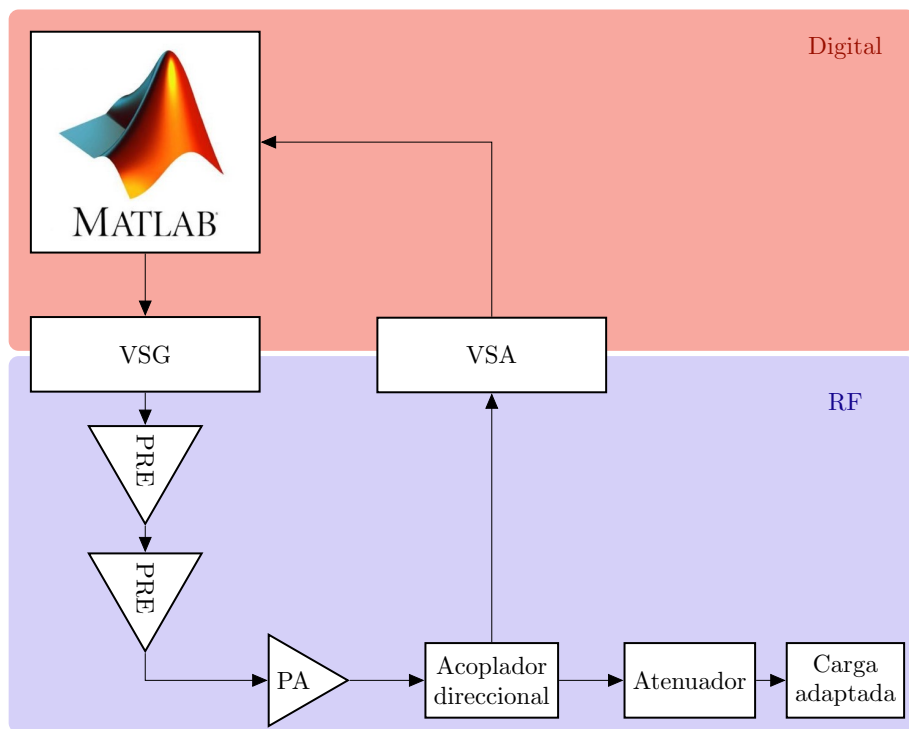


Figura 5.1 Monitor del analizador de espectros.

La combinación de las bandas de la señal se hace en MATLAB[®], y es por esto que solo se emplea un canal de VSG. Estas señales se han visto pre-amplificadas antes del DUT, ya que el generador no es capaz de suministrar la potencia suficiente como para llevar al amplificador a la zona de saturación. Téngase en cuenta que los pre-amplificadores trabajan en zona muy lineal para no distorsionar la señal. Por último es importante hacer la extracción de la señal convenientemente, ya que no deben introducirse en el VSA señales demasiado potentes. Es por esto que el acoplador direccional disgrega en dos la salida del PA, empleándose el puerto adaptado para el analizador consumiendo la potencia del puerto de salida en el atenuador. A su vez



(a)



(b)

Figura 5.2 Banco de pruebas (a) foto del montaje, (b) esquemático por bloques.

el atenuador debe terminarse con una carga adaptada, de lo contrario se producirían reflexiones indeseadas, falseando así la medida del VSA. Finalmente, el VSA aplica un filtrado paso banda centrado en 3.6 GHz en el que entran las dos bandas de la señal. Tras la demodulación, una vez se han obtenido las muestras de la envolvente compleja paso de baja de la señal de doble banda, se separan en software empleando MATLAB®.

La señal de doble banda ha sido generada de acuerdo con el estándar 5G-NR, empleando modulación OFDM en cada banda. Se ha empleado un ancho de banda de 10 MHz en sendas bandas y una numerología de 1, esto es, 30 kHz de separación entre subportadoras. Así pues primero se generó cada banda en banda base, empleando una secuencia de bits aleatoria y una frecuencia de muestreo $f_s = 96.16$ MHz. Para conseguir una separación entre las bandas de 40 MHz se aplicó un desplazamiento en frecuencia de -20 MHz a la banda inferior y $+20$ MHz a la banda superior cuando estas estaban en banda base. La semilla empleada para la secuencia de bits provocó una PAPR en torno a 11 dB. Con fin de proteger al amplificador se aplicó un *hard clipping*, que impidió que la PAPR superase los 12 dB.

Por último, en el analizador se adquirieron muestras de la envolvente compleja de la señal en el puerto acoplado, considerando las portadoras de cada una de las bandas y aplicando una tasa de muestreo $f_s/2$, esto es, un factor de sobremuestreo de 3. Además se redujo el ruido de medida promediando 10 realizaciones de la señal adquirida, habiendo tenido que sincronizar posteriormente la entrada y la salida. En la Figura 5.1 se muestra la pantalla de control del VSA, donde se establecen las bandas para la medida del ACPR. Una vez obtenidas las muestras son cargadas de nuevo en MATLAB® para poder medir así los parámetros de interés. El montaje junto con su esquemático se muestran en la Figura 5.2.

5.2 Caracterización inicial

Esta sección brindará ciertos resultados que servirán como punto de partida. Así pues, como se trata de una caracterización, la idea consiste en poner a prueba el PA sin realizar ningún cambio en la señal de entrada. De modo que la comparación entre el espectro a la entrada y a la salida se muestra en la Figura 5.3

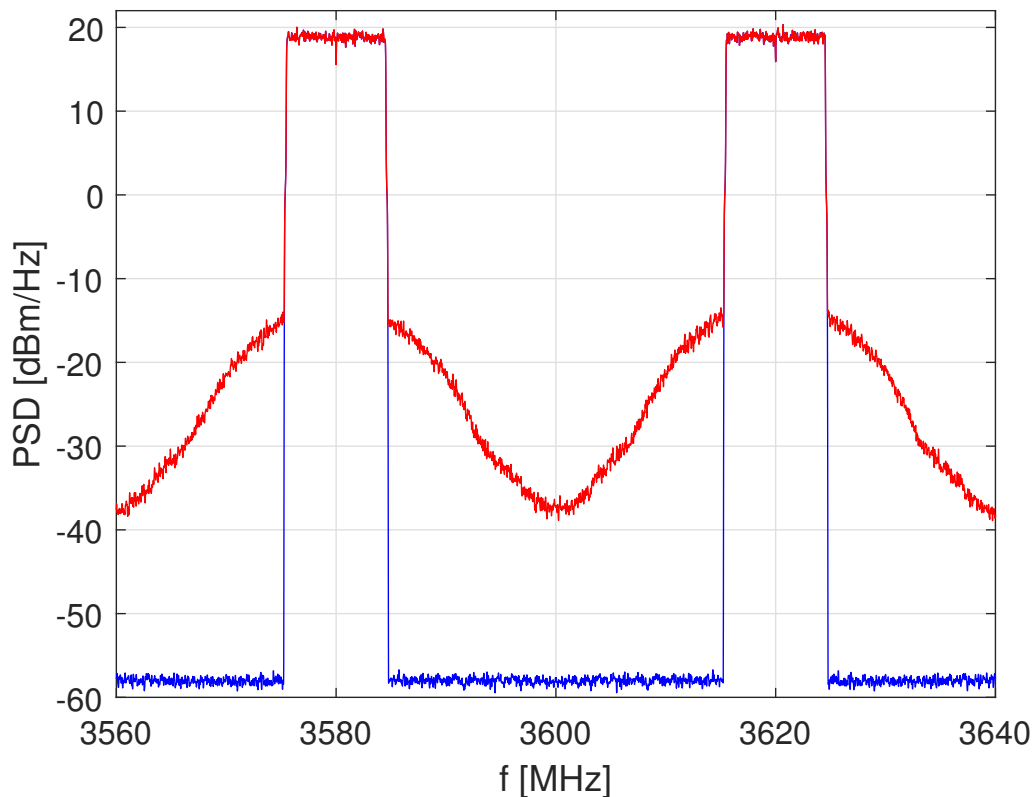


Figura 5.3 Densidad espectral de potencia de la señal de entrada multiplicada por la ganancia lineal y la salida sin aplicar DPD.

Puede verse claramente cómo la distorsión provocada por el amplificador deriva en un recrecimiento espectral de los canales adyacentes. Los efectos de esta distorsión quedan cuantificados en la Tabla 5.1.

Tabla 5.1 Parámetros medidos sin DPD.

NMSE ₁	NMSE ₂	ACPR _{1,-1/1,+1}	ACPR _{2,-1/2,+1}	EVM ₁	EVM ₂	PAE
(dB)		(dBc)		(%)		(%)
-30.3	-29.8	-39.1/-39.0	-38.5/-38.6	2.8	2.8	16.9

5.3 Aplicación de DPD-DLA

Con el fin de eliminar el recrecimiento espectral se acude a la técnica de DPD, ya desarrollada en el Capítulo 3. Concretamente se aplica el modelo 2D-GMP de Enzinger [13], cuyos conjuntos \mathcal{H} , \mathcal{L} y \mathcal{M} se muestran en la Tabla 5.2

Tabla 5.2 Configuración del modelo 2D-GMP.

Conjunto	Intra-banda	Banda cruzada	Banda mezclada
\mathcal{H}_a	[0 1 2 3 4 5 6 7 8 9 10 11 12 13]	[1 2 3 4 5 6 7 8 9 10 11 12 13]	[1 2 3 4 5 6 7 8 9 10 11 12 13]
\mathcal{L}_a	[5 3 3 3 1 1 1 1 1 1 1 1 1 1]	[3 3 3 1 1 1 1 1 1 1 1 1 1 1]	[3 3 3 1 1 1 1 1 1 1 1 1 1 1]
\mathcal{H}_b	[1 2 3 4 5]	[1 2 3 4 5]	[1 2 3 4 5 6 7 8 9]
\mathcal{L}_b	[2 2 1 1 1]	[2 2 1 1 1]	[2 2 2 2 2 2 2 2]
\mathcal{M}_b	[1 1 1 1 1]	[1 1 1 1 1]	[1 1 1 1 1 1 1 1 1]
\mathcal{H}_c	[1 2 3 4 5]	[1 2 3 4 5]	[2 3 4 5 6 7 8 9]
\mathcal{L}_c	[2 2 1 1 1]	[2 2 1 1 1]	[2 2 2 2 2 2 2 2]
\mathcal{M}_c	[1 1 1 1 1]	[1 1 1 1 1]	[1 1 1 1 1 1 1 1]

Este modelo produce en total 534 regresores, siendo tan elevado que resulta inestable en su regresión. En la Figura 5.4 se muestra cómo el bucle de control tiende a la divergencia provocando que los coeficientes calculados empeoren el ACPR.

Tabla 5.3 Parámetros medidos con DPD en 5 iteraciones. Caso inestable.

Iter	NMSE ₁	NMSE ₂	ACPR _{1,-1/1,+1}	ACPR _{2,-1/2,+1}	EVM ₁	EVM ₂	PAE
	(dB)		(dBc)		(%)		(%)
1	-30.4	-30.0	-39.3/-39.1	-38.6/-38.7	2.7	2.8	17.7
2	-33.6	-33.1	-42.2/-42.3	-41.9/-41.3	2.0	2.1	17.7
3	-36.3	-35.0	-44.8/-45.0	-44.2/-43.0	1.6	1.7	17.6
4	-38.4	-34.7	-47.1/-47.5	-43.9/-42.9	1.4	1.7	17.7
5	-34.2	-26.7	-48.6/-49.0	-34.1/-34.1	1.8	3.9	17.8

Este experimento fue realizado con una factor de aprendizaje $\mu = 1/3$, que en principio se podría afirmar que es relativamente bajo. Esta divergencia es debida a los problemas numéricos que presenta la matriz de regresores, tal y como se comentaba en el capítulo anterior. La inestabilidad además de empeorar el resto de parámetros como se recoge en la Tabla 5.3, se manifestó en un aumento desmesurado de la PAPR, que fue aplacado con el *hard clipping*. También merece la pena comentar el hecho de que la divergencia se alcanzó antes en la banda superior que en la inferior, aunque puede verse cómo la banda inferior había comenzado a diverger.

Por último, se puede comentar que el hecho de que la zona de separación entre los efectos de cada banda en la Figura 5.4 no sea la frecuencia de 3600 MHz se debe que la separación aplicada es inferior a la frecuencia de muestreo, produciéndose un pequeño efecto de *aliasing* únicamente en la zona central.

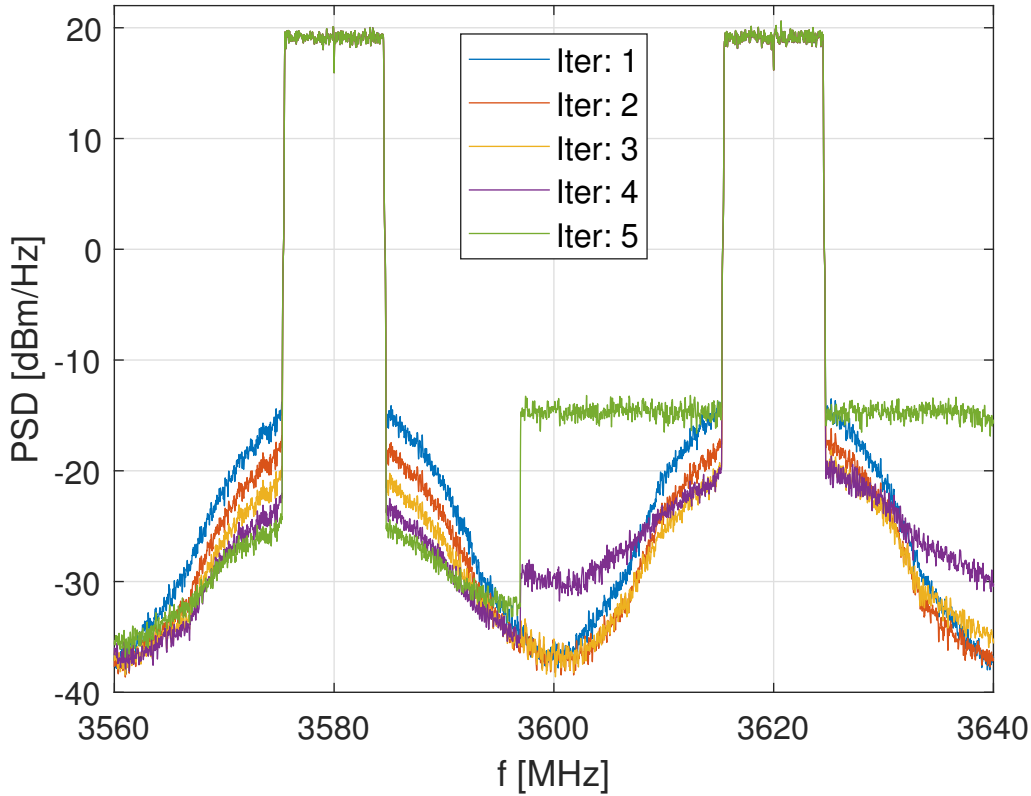


Figura 5.4 Densidad espectral de potencia de la señal de salida para cada iteración del DPD.

5.4 Aplicación de regularización de Ridge a DPD-DLA

Así pues, llegados a este punto se puede decir que este es el escenario idóneo para plantear la regularización de Ridge. La situación lo requiere, ya que no se desea bajar en número de coeficientes del modelo porque se empeorarían las prestaciones del resultado y tampoco puede transmitirse la señal sin hacer DPD. Por consiguiente, la aplicación de Ridge plantea un pequeño cambio en la arquitectura de DLA, ya que presenta dos etapas. La primera consistiría en optimizar la regularización con un λ_{opt} respecto a un criterio y la segunda aplicarlo en el DPD. El criterio de optimización será la minimización del NMSE de validación, empleando un set de muestras de $u[k]$ para identificar la afectación de λ sobre el NMSE en regresión y otro

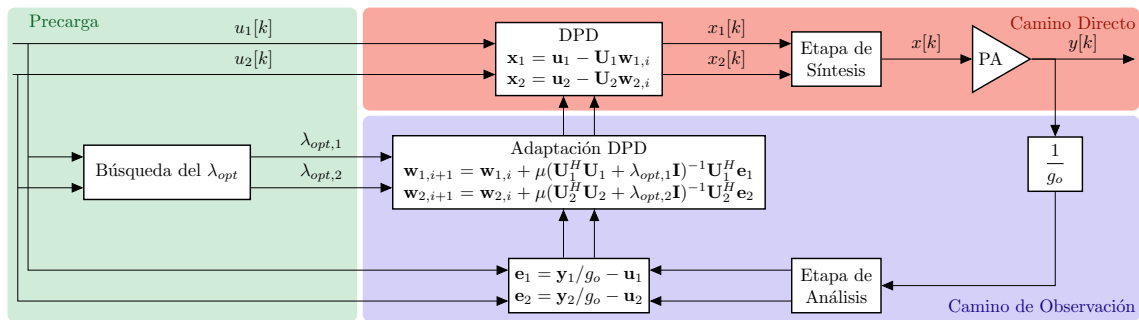


Figura 5.5 Esquema de implementación de la regularización de Ridge en un escenario de DPD en DLA con doble banda.

set para la validación. Concretamente se han empleado el 10% de las muestras para identificar, tomando como criterio de elección que sean máximas en amplitud. Existen otros criterios como el mostrado en [28] que podrían emplearse para la elección del set de muestras de identificación, pero se ha escogido este por simplicidad. De modo que toda esta operación se concentra en el bloque “Búsqueda del λ_{opt} ” que se presenta en el esquema de la Figura 5.5.

Si bien el λ_{opt} depende de la realización, está igualmente relacionado con el estadístico de la señal, por lo que a priori se espera que el óptimo no cambie demasiado. De todas formas, dado que el esquema DLA contempla recalcular sus coeficientes en función del comportamiento, volver a obtener λ_{opt} es recomendable. El bloque “Búsqueda del λ_{opt} ” calcula progresivamente modelos regularizados variando λ y evaluando el NMSE con sus respectivos sets de muestras, declarando λ_{opt} aquel λ que consigue un menor NMSE en validación. Cada banda se calcula por separado, mostrándose la realización del cálculo para la banda 1 en la Figura 5.6.

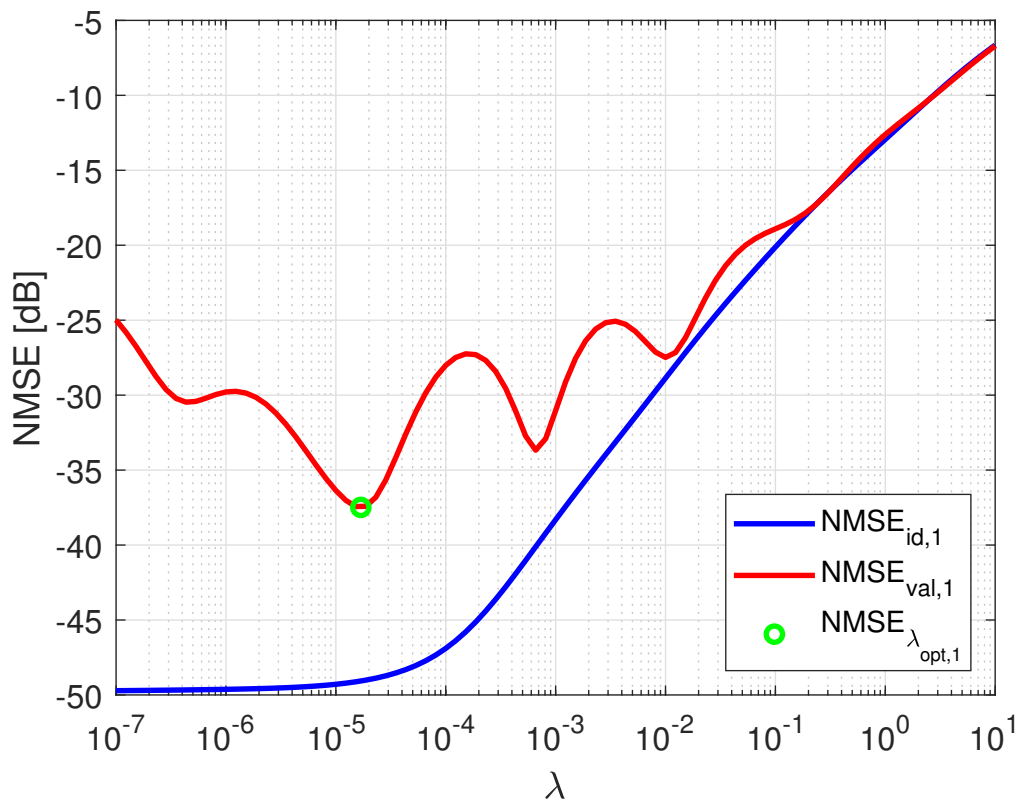


Figura 5.6 Cálculo del λ_{opt} de la banda inferior en función del NMSE de validación.

Así pues, se consigue asegurar la estabilidad del modelo perjudicando lo mínimo posible las prestaciones de la señal obtenida. Existen un par de matices a destacar en cuanto a la obtención del λ_{opt} . El primero de ellos es que no se debe confundir el NMSE que se obtiene en este paso previo con el que resulta de aplicar DLA, ya que en este caso se están evaluando los coeficientes como modelo de amplificador, mientras que en la arquitectura DLA se van a encargar de modelar el error. Es por esto que la Figura 5.6 no coincidirá con los resultados de NMSE que se obtendrán al validar con el sistema completo. El segundo matiz reside en la forma de la gráfica del NMSE de validación. El hecho de que haya mínimos locales se traduce en que, por alguna razón, al aumentar la regularización se induce a aumentar el error en ciertas zonas. La interpretación que esto puede tener es el hecho de que las componentes en la base PC no están ordenadas en función de la correlación, sino de su amplitud. Esto provocaría que, conforme la regularización va afectando a las componentes comienzan a penalizarse direcciones de la base PC que se correspondan con direcciones de importancia en la base DOMP. En otras palabras, parece que a primera vista el criterio a tener en cuenta es la correlación antes que la amplitud. En cualquier caso, aunque aplicar Ridge en PC no parezca óptimo, es efectivo, consiguiendo que el DPD sea estable. En la Figura 5.7 se muestra el espectro con una señal de

validación tras aplicar el modelo de la Tabla 5.2 regularizado y un factor de aprendizaje $\mu = 1/3$, quedando reflejadas sus prestaciones en la Tabla 5.4. Cabe destacar que, además de haber mitigado la aparición de recrecimiento espectral en las bandas adyacentes ha disminuido la distorsión en banda, que se traduce en un descenso del EVM. Además la PAE no se ha visto alterada, por lo que a efectos numéricos no empeora la eficiencia del amplificador. Es más, puede decirse que mejora, ya que gracias al DPD está empleando toda su potencia en la banda de interés.

Tabla 5.4 Parámetros medidos aplicando DPD y Ridge, con $\lambda_{opt,1} = 1.69 \cdot 10^{-5}$, $\lambda_{opt,2} = 2.28 \cdot 10^{-5}$ tras 15 iteraciones con $\mu = 1/3$.

Caso	NMSE ₁	NMSE ₂	ACPR _{1,-1/1,+1}	ACPR _{2,-1/2,+1}	EVM ₁	EVM ₂	PAE
	(dB)		(dBc)		(%)		(%)
Sin DPD	-30.3	-29.8	-39.1/-39.0	-38.5/-38.6	2.8	2.8	16.9
DPD-DLA	-49.3	-50.7	-63.0/-63.5	-63.0/-62.5	1.2	1.3	16.9

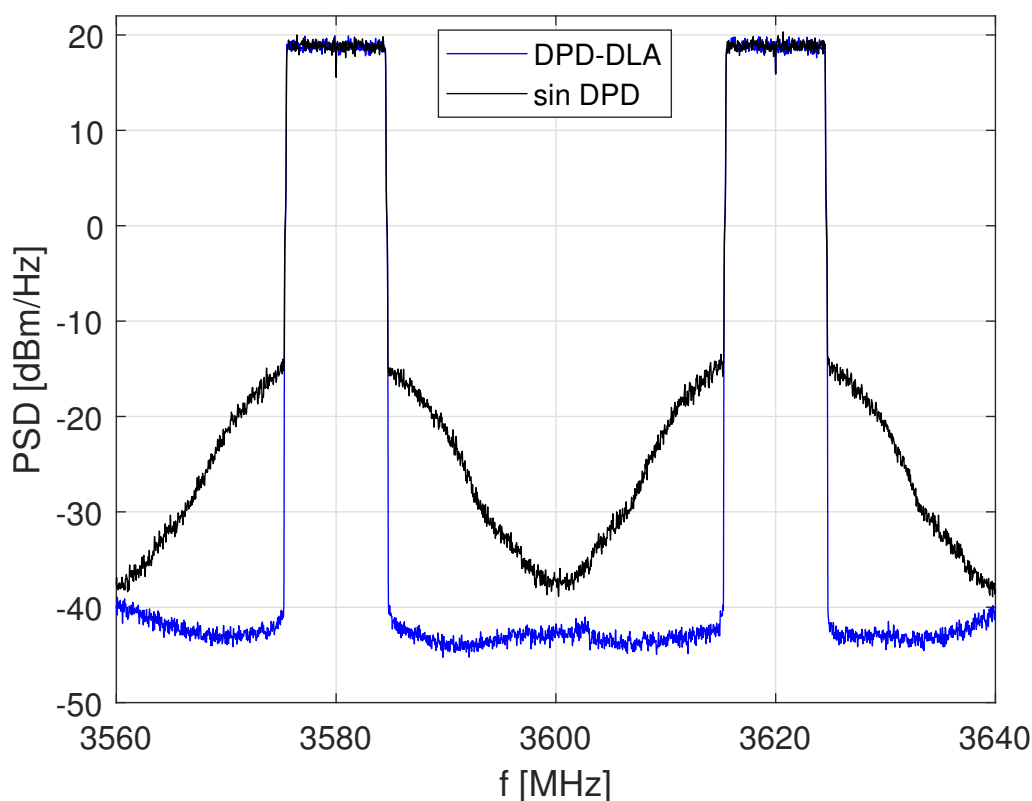


Figura 5.7 Comparativa entre el espectro a la salida del PA sin aplicar DPD y habiendo aplicado DPD.

Dado que el hecho de regularizar ha estabilizado el modelo, cabe plantear un aumento en el factor de aprendizaje con el fin de obtener la convergencia más rápidamente. Así pues, en la Figura 5.8 se muestra la evolución del NMSE y el ACPR para factores de aprendizaje $\mu = 1/3$ y $\mu = 1$, siendo llamativo el hecho de que, a pesar de que el modelo resultaba inestable por sí solo con un factor $\mu = 1/3$, regularizando es posible forzar al máximo la velocidad de aprendizaje. Además los resultados convergen entre la 5ª y la 6ª iteración, aproximadamente tres veces más rápido y obteniendo prácticamente el mismo resultado que en la última iteración de $\mu = 1/3$, por lo que es preferible fijar $\mu = 1$.

Por último existe un indicativo más de que el sistema ha sido linealizado, y es la relación de potencia de entrada frente a la ganancia, equivalente a la respuesta AM-AM. En este caso si el dispositivo es lineal debe observarse una constante que indica la ganancia del dispositivo. En cambio la ganancia de un amplificador

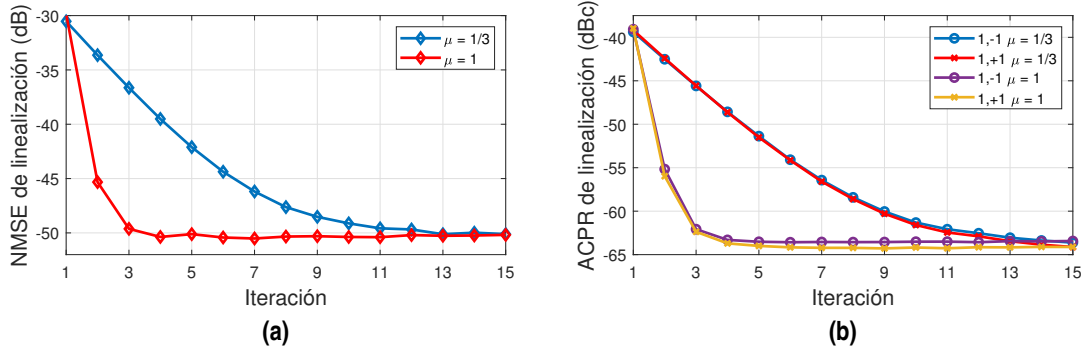


Figura 5.8 Evolución del (a) NMSE de linealización y (b) ACPR de la banda 1 para valores de $\mu = 1/3$ y 1 frente al número de iteraciones.

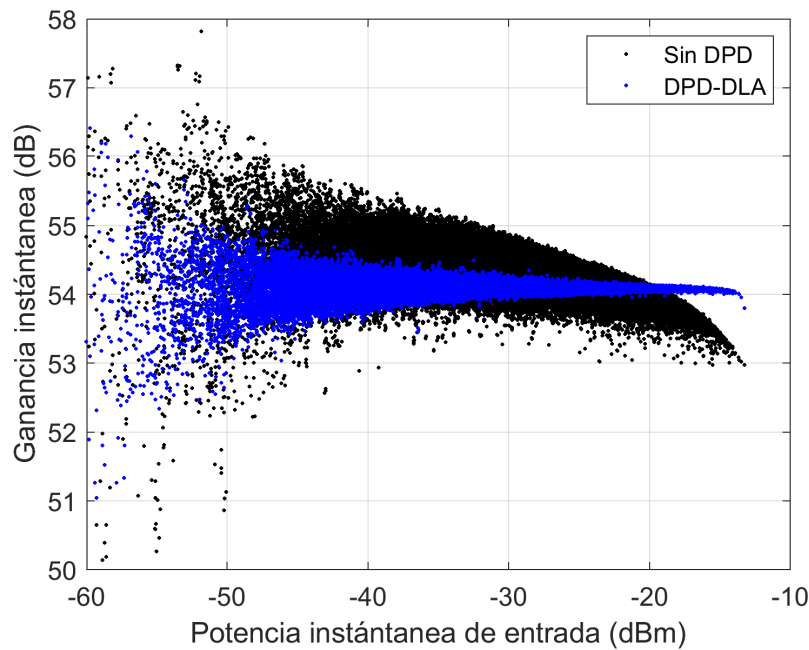


Figura 5.9 Linealización de la ganancia instantánea al emplear DPD.

no lineal disminuye cuando llega a la saturación. Además, los efectos de la memoria hacen que no se vea una sola línea, sino que constituyen una nube de puntos, de modo que, cuanto mejor sea la linealización, más se parecerá esta característica a una constante. Este resultado se muestra en la Figura 5.9, donde se ve cómo el DPD ha conseguido estabilizar la ganancia instantánea con respecto a la potencia de entrada. Nótese el pequeño descenso que se advierte en la ganancia cuando se alcanzan las potencias de entrada más elevadas. Esto es debido al empleo de la ganancia promedio en DLA.

5.5 Aplicación de regularización de Ridge a DPD-DLA con selección de coeficientes RC-DOMP

Esta sección trata de justificar el comportamiento observado en la Figura 5.6 mediante resultados en simulación. Así pues, como se comentó en el Capítulo 3, es posible trazar una estrategia para que la regularización se centre en penalizar las componentes correladas. Concretamente consistiría en aplicar primero el algoritmo de RC-DOMP (Algoritmo 3) en el bloque “Búsqueda del λ_{opt} ”, obteniendo así la matriz \mathcal{T} e integrar posteriormente la penalización en la base DOMP aplicando el Algoritmo 4, seleccionando las columnas óptimas de la matriz de regresores. Procediendo de esta forma se obtiene la Figura 5.10 de la que pueden hacerse un par de comentarios. En primer lugar, no presenta máximos locales, es decir, es siempre decreciente

hasta que cambia la tendencia en torno a $\lambda = 10^{-3}$. Relacionando este comportamiento con el concepto de los grados de libertad efectivos, si bien al aumentar λ la restricción es la misma en todas las direcciones de una base ortogonal, aquellas con menor varianza se extinguen antes. De manera que este resultado corrobora el hecho de que el descenso del NMSE va en consonancia con la penalización de las direcciones más correladas.

Es preciso incidir en que, dado que el modelo completo era inestable es necesario seleccionar componentes, ya que la ortogonalización de la matriz de regresores es también una operación con problemas numéricos debido a que las componentes muy correladas derivan en columnas muy cercanas a cero en la matriz de autocorrelaciones muestrales. Además, en la Figura 5.10 se ha calculado NMSE desde $\lambda = 10^{-16} \approx 0$, para así incluir en la gráfica la solución en mínimos cuadrados. De manera que esta diferencia se debe a la respuesta en validación del modelo con las columnas óptimas. Por otro lado, hay que tener en cuenta que la aplicación

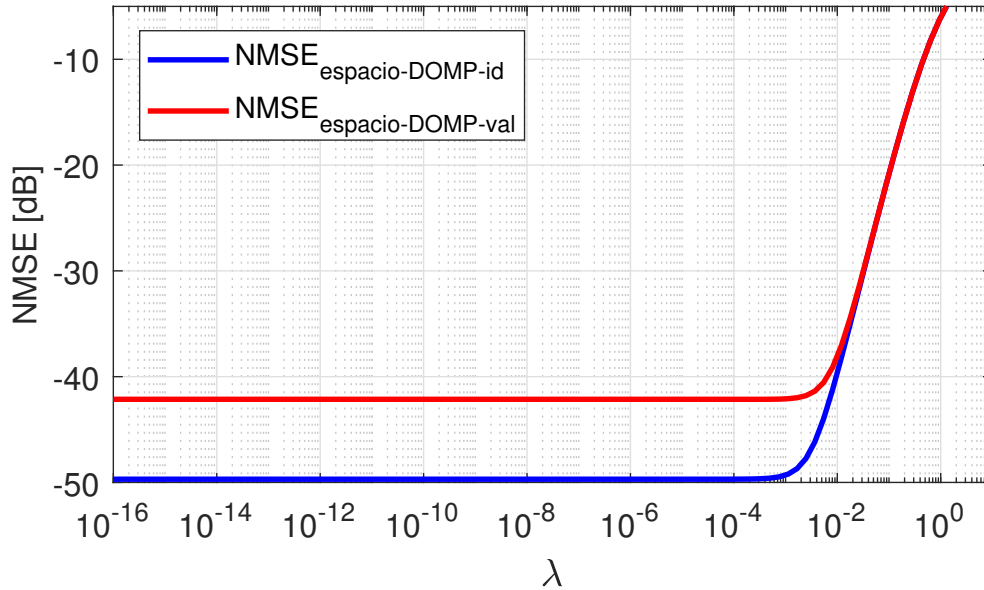


Figura 5.10 Curvas de NMSE para identificación y validación en función de λ .

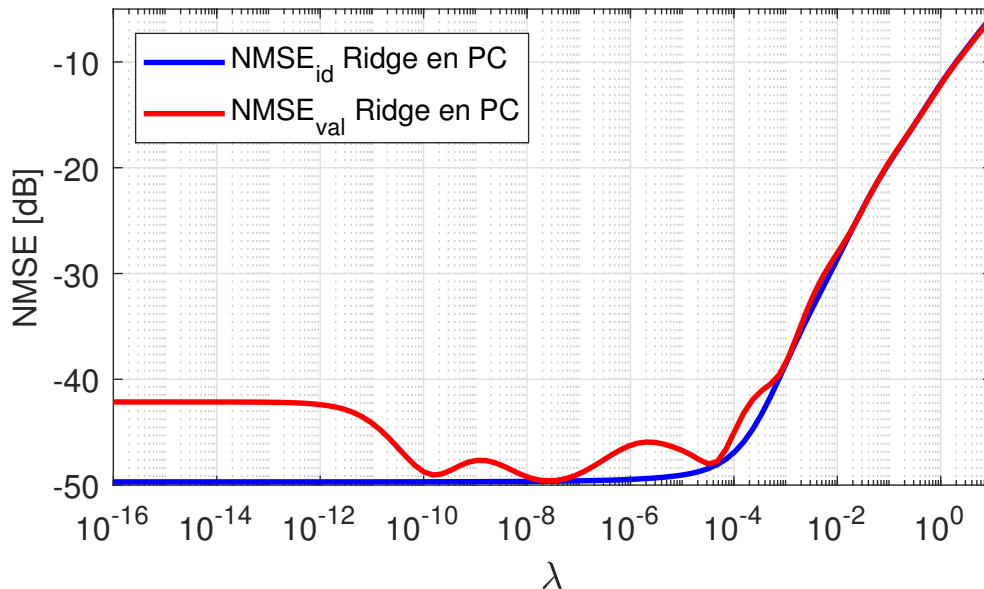


Figura 5.11 Curvas de NMSE para identificación y validación en función de λ .

del Algoritmo 4 está centrado exclusivamente en penalizar la correlación entre las columnas de la matriz de regresores. Dado que se ha seleccionado primero y regularizado después sobre el mismo criterio, la mejora al aplicar Ridge no es muy notable, y esto se manifiesta en una pendiente casi nula al aumentar λ . Por ello, se plantea a continuación un tercer enfoque, que consiste en procurar corregir ambos problemas. Así pues, si primero se aplica RC-DOMP y luego se regulariza en la base PC, se eliminan en primer lugar las columnas muy correladas y en segunda instancia se penalizan las poco variantes. De manera que se obtiene así la Figura 5.11.

Pueden realizarse dos apreciaciones mediante la comparativa entre la Figura 5.10 y 5.11. La primera de ellas es el hecho de que, como se comentaba en el Capítulo 4, el problema numérico reside tanto en la correlación entre columnas como en la existencia de columnas poco variantes. Así pues cabía esperar que planteando una solución que tuviese en cuenta ambos problemas se obtuviesen mejores resultados. La segunda lectura que ofrece la comparativa está en relación al momento en el que Ridge comienza a empeorar los resultados. El valor a partir del cual la gráfica comienza a ser creciente dependerá exclusivamente de la relación entre las varianzas de las distintas direcciones de cada base. Así pues, si las componentes importantes de una base se diferencian notablemente de las poco variantes λ deberá tomar un valor elevado para que se aprecie su efecto en las direcciones dominantes. Es por esto que se deduce de las gráficas que existe una diferencia mayor en la base DOMP que en la base de componentes principales en relación a este aspecto.

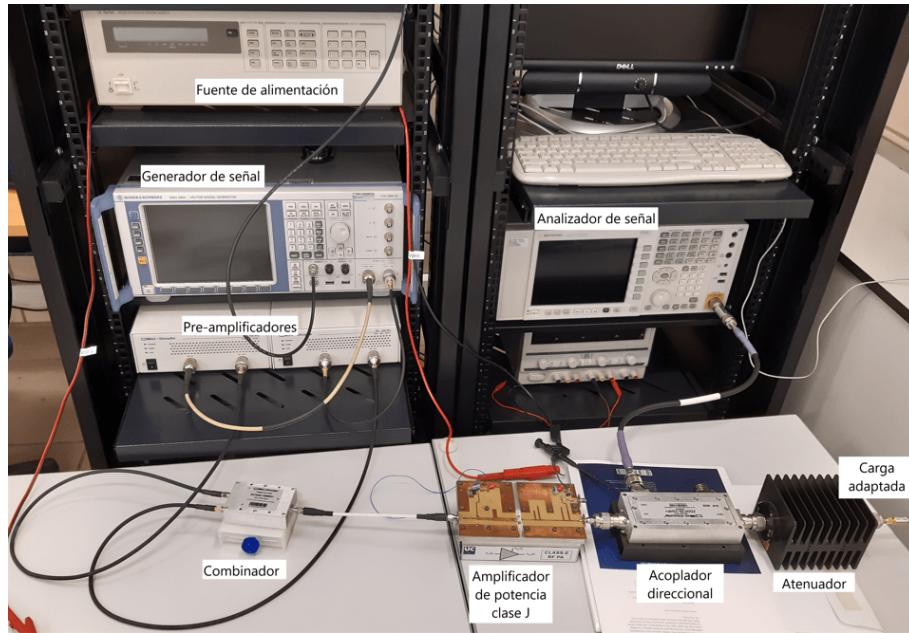
6 Conclusiones y líneas futuras

No nos toca a nosotros decidir qué tiempo vivir, sólo podemos elegir qué hacer con el tiempo que se nos ha dado.

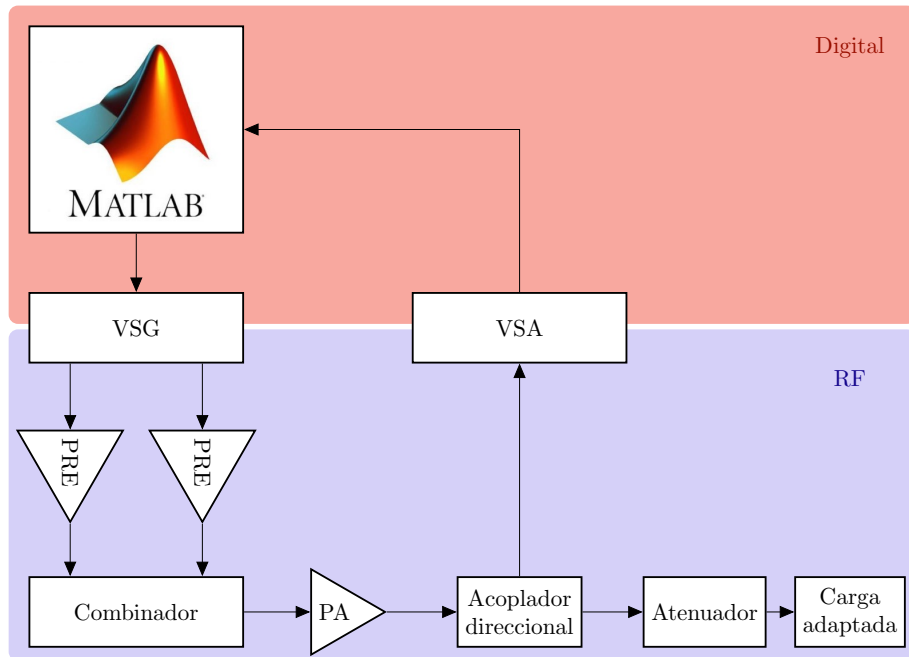
GANDALF, EN *EL SEÑOR DE LOS ANILLOS*, 1954

Este trabajo ha pasado por muchos aspectos dentro de la teoría de sistemas, buscando en cada apartado conducir estos conceptos hacia la predistorsión digital. Además, se ha buscado un enfoque específico hacia los sistemas multibanda, que se encuentran en auge y seguirán siendo de gran importancia en las futuras generaciones móviles. Dentro del reto que supone la no linealidad del PA, la necesidad de implementar modelos en doble banda concurrente aumenta aún más si cabe la dificultad para encontrar soluciones estables. Así pues, las pruebas experimentales realizadas han puesto en evidencia distintos aspectos. El primero, presentado en la Sección 5.2, ya se conocía; y es la evidencia de que es necesario realizar DPD para compensar los efectos de la distorsión en banda y fuera de banda. El segundo es que, al trabajar con modelos de doble banda, se añade un problema extra, y es el hecho de que la implementación de la solución en mínimos cuadrados dentro de una arquitectura de aprendizaje directo es inestable, tal y como se muestra en la Sección 5.3. Este escenario ha conducido a la necesidad de implementar técnicas que reduzcan la complejidad del modelo y hagan estable a la solución. Entre todas las técnicas posibles se ha optado por el desarrollo de la regularización de Ridge como método de estabilización y las técnicas de selección de componentes presentadas en el Capítulo 4 como método de reducción de complejidad. Con estas herramientas por delante, se han planteado tres implementaciones distintas, que han ido ahondando progresivamente en la búsqueda de los elementos que provocan la inestabilidad. A la vista de los resultados se puede afirmar que, el primero y más importante, es el mostrado en la Sección 5.4, que confirma a la regularización de Ridge como un método válido para la estabilización de las soluciones. Además queda de manifiesto el compromiso de la regularización con la validación, mostrado en las curvas de NMSE.

Por otro lado se han realizado dos experimentos más, que direccionan las líneas futuras de este trabajo. Éstos han consistido en la aplicación conjunta de las técnicas de selección de componentes y la regularización de Ridge. En la Sección 5.5 se ha puesto de manifiesto que existen distintas configuraciones de aplicación, donde la base empleada juega un papel fundamental. De modo que, a corto plazo, se podría continuar afinando aún más en la determinación de las componentes de inestabilidad, para así adoptar la técnica que más se ajuste a dichas componentes. Además, podrían emplearse otros amplificadores como el clase J, que requieran un modelo más complejo. De modo que, a largo plazo, se podría continuar afinando aún más en la determinación de las componentes de inestabilidad, para así adoptar la técnica que más se ajuste a dichas componentes. Además, se ha comprobado experimentalmente que el amplificador clase AB no presenta demasiada complejidad a la hora de linealizarlo, ya que, si bien presenta inestabilidad al aplicar DPD, tiene una transición relativamente suave a la zona de saturación. Es por esto que se propone como siguiente paso poner a prueba estas técnicas en un amplificador clase J, cuyo montaje ya se ha preparado y se muestra en la Figura 6.1.



(a)



(b)

Figura 6.1 Banco de pruebas (a) foto del montaje, (b) esquemático por bloques.

Apéndice A

Código de MATLAB® empleado en el trabajo

En este apéndice se recogen los códigos empleados para el desarrollo de las pruebas experimentales. Una gran parte de este apéndice se trata de modificaciones de los códigos desarrollados por el Grupo de Sistemas de Radiocomunicación, al que el autor expresa su gratitud por su dedicación en el desarrollo de este trabajo. Además, contienen llamadas a funciones que no se van a recoger en este apéndice dada su extensión.

Código A.1 Código principal, donde se realiza la generación de la señal 5G-NR, cálculo del modelo, empleo de la arquitectura DLA con las respectivas medidas sobre el amplificador y aplicación de la regularización de Ridge a la matriz de regresores.

```
clearvars; clc; close all;
addpath('toolbox'); addpath('toolbox_Chalmers');

nmse = @(ymod,y) 20*log10(norm(ymod-y)/norm(y));
dBm = @(x) 10*log10(rms(x).^2/100)+30;
scale_dBm = @(x,P) x*10^((P-dBm(x))/20);
PAPR = @(x) 20*log10(max(abs(x))/rms(x));
dBminst = @(x) 10*log10(abs(x).^2/100)+30;

% Load static configuration
cd('confset'); configurationB; cd ..; %534 regresores
UsarDomp = false;
RMSin = -24; % Power In-Put
mu = 1/3;
Nmeas = 1; % Number of measurements
Nave = 10;
Niter = 15; % Iteration
Nval = 3;

ovs = 3; % OverSampling
fsgen = 1024*30e3*ovs; % Frequency used to generate Signal
fsmed = fsgen; % Measurement Freq.
NRB = 25; % 20 MHz (number of resource blocks)

%% Generacion OFDM
Foff = [-20e6,20e6];
[u,us,~,datau] = genera_5GNR_multicarrier_v4([NRB,NRB], [30e3,30e3],...
    [10e6,10e6], Foff, fsgen, [8,8], [2,1]);
usL = us(:,1);
usR = us(:,2);
fsint = fsgen/2;

medida_out = [];
```

```

filenamedate = datestr(clock,30);
filename = ['measurements' filesep 'exp_' filenamedate];
figfilename = ['results' filesep 'exp_' filenamedate];

%% First measurement and initialization
u = scale_dBm(u,RMSin);
[y, x, data] = myPA(u, dBm(u), Nmeas, Nave);

[uLs, uRs] = separate_bands(u,fsmed,fsint,Foff);
[yLs, yRs] = separate_bands(y,fsmed,fsint,Foff);
Pout = dBm(y);
Pin = dBm(u);

% Ganancia Estática
G = norm(y)/norm(u);
GL = G;
GR = G;

analiza_medidas_OFDM_B_MIG(uLs,yLs,datau,0,fsint,fsgen,99);

%% Calculation of maximum memory buffers (Qpmax) and (Qnmax), 2D-GMP
Qa1 = max([max(x1_limits.intra.La), max(x1_limits.cross.La),...
           max(x1_limits.mixed.La)]);
Qa2 = max([max(x2_limits.intra.La), max(x2_limits.cross.La),...
           max(x2_limits.mixed.La)]);
Qb1 = max([max(x1_limits.intra.Lb) + max(x1_limits.intra.Mb),...
           max(x1_limits.cross.Lb) + max(x1_limits.cross.Mb),...
           max(x1_limits.mixed.Lb) + max(x1_limits.mixed.Mb)]);
Qb2 = max([max(x2_limits.intra.Lb) + max(x2_limits.intra.Mb),...
           max(x2_limits.cross.Lb) + max(x2_limits.cross.Mb),...
           max(x2_limits.mixed.Lb) + max(x2_limits.mixed.Mb)]);
Qc1 = max([max(x1_limits.intra.Lc) - min(x1_limits.intra.Mc),...
           max(x1_limits.cross.Lc) - min(x1_limits.cross.Mc),...
           max(x1_limits.mixed.Lc) - min(x1_limits.mixed.Mc)]);
Qc2 = max([max(x2_limits.intra.Lc) - min(x2_limits.intra.Mc),...
           max(x2_limits.cross.Lc) - min(x2_limits.cross.Mc),...
           max(x2_limits.mixed.Lc) - min(x2_limits.mixed.Mc)]);

Qpmax1 = max([Qa1, Qb1, Qc1]);
Qpmax2 = max([Qa2, Qb2, Qc2]);

Qnmax1 = max([x1_limits.intra.Mc, x1_limits.cross.Mc, 0]);
Qnmax2 = max([x2_limits.intra.Mc, x2_limits.cross.Mc, 0]);

N = length(yLs);

%Construcción de la matriz de regresores
[UL, RmatL] = buildW_2GMP(uLs, uRs, [N-Qpmax1+1:N , 1:N , 1:Qnmax1],...
                          Qpmax1, Qnmax1, x1_limits);
[UR, RmatR] = buildW_2GMP(uRs, uLs, [N-Qpmax2+1:N , 1:N , 1:Qnmax2],...
                          Qpmax2, Qnmax2, x2_limits);

[~,cL]=size(UL);
[~,cR]=size(UR);
sL = 1:cL;
sR = 1:cR;

```

```

if UsarDomp == true
    indices = sel_indices(yLs,yLs,.1);
    [~, sL, noptL, ~, ~] = RCDOMP(UL(indices,:), yLs(indices), RmatL, 51);
    indices = sel_indices(yRs,yRs,.1);
    [~, sR, noptR, ~, ~] = RCDOMP(UR(indices,:), yRs(indices), RmatR, 51);
    cL = 50;
    cR = 50;
    sL = sL(1:cL).';
    sR = sR(1:cR).';
    [nmse_val_L, nmse_id_L, ncond_L] = ridge_function(UL(:,sL),yLs,0);
    [nmse_val_R, nmse_id_R, ncond_R] = ridge_function(UR(:,sR),yRs,0);
end

%Inicialización del modelo:
wL = zeros(cL,1);
dwL = zeros(cL,1);
wR = zeros(cR,1);
dwR = zeros(cR,1);

%% Cálculo de lambda_opt:
options = optimset('Display','iter','PlotFcns',@optimplotfval);
[lambda_opt_L, NMSE_L_opt] = fminsearch(@(lambda) ridge_function(UL(:,sL),...
    yLs,lambda),10^-5,options);
[lambda_opt_R, NMSE_R_opt] = fminsearch(@(lambda) ridge_function(UR(:,sR),...
    yRs,lambda),10^-5,options);
% Seguridad por si el algoritmo devuelve un lambda_opt negativo
if lambda_opt_L < 0
    lambda_opt_L = 10^-8;
elseif lambda_opt_R < 0
    lambda_opt_R = 10^-8;
end

%% Cálculo de matriz inversa
fprintf('\nCálculo de la matriz inversa \n');
uL_pot_col = vecnorm(UL(:,sL)); %Potencia de cada columna
UL_norm = UL(:,sL)*diag(uL_pot_col.^(-1));
uR_pot_col = vecnorm(UR(:,sR)); %Potencia de cada columna
UR_norm = UR(:,sR)*diag(uR_pot_col.^(-1));

%Precarga de las matrices usadas en el bloque de Adaptación DPD:
UL_inv = (inv(UL_norm'*UL_norm+eye(cL)*lambda_opt_L)*UL_norm');
UR_inv = (inv(UR_norm'*UR_norm+eye(cR)*lambda_opt_R)*UR_norm');

%% DLA loop
figure(30),title('y');

nmsesL = [];
nmsesR = [];

for iter=1:Niter
    fprintf(['Iter: ' num2str(iter) '\n']);

    %DPD:
    xLs = uLs-UL(:,sL)*wL;
    xRs = uRs-UR(:,sR)*wR;

```

```

%Etapa de Síntesis
x = unite_bands(xLs,xRs,fsint,fsmed,Foff);

PAPR(x)
dBm(x)

PAPRx = 20*log10(max(abs(x))/rms(x));
disp(sprintf('PAPR antes de clipping: %4.2f dB\n', PAPRx));
PAPRd = 12;
clip = (10^((PAPRd-PAPRx)/20))*max(abs(x));
x(abs(x)>clip) = clip*exp(1i*angle(x(abs(x)>clip)));
disp(sprintf('PAPR después de clipping: %4.2f dB\n', PAPR(x)));

[y, x, data] = myPA(x, dBm(x), Nmeas, Nave);

espectro(y - mean(y),fsgen,30);

%Etapa de Análisis
[xLs, xRs] = separate_bands(x,fsmed,fsint,Foff);
[yLs, yRs] = separate_bands(y,fsmed,fsint,Foff);

eL = yLs/GL-uLs;
eR = yRs/GR-uRs;

%Adaptación DPD:
dwL = diag(uL_pot_col.^(-1))*UL_inv*eL;
wL = wL + mu * dwL;
dwR = diag(uR_pot_col.^(-1))*UR_inv*eR;
wR = wR + mu * dwR;

figure(10); plot(dBminst(u), dBminst(y)- dBminst(u),'.');
hold on;

[PerformanceL]=analiza_medidas_OFDM_B_MIG(uLs,yLs,datau,0,fsint,fsgen,99);
[PerformanceR]=analiza_medidas_OFDM_B_MIG(uRs,yRs,datau,0,fsint,fsgen,101);
medida.descr = ['Itera : ' num2str(iter)];
medida.PerformanceL = PerformanceL;
medida.PerformanceR = PerformanceR;
medida.x = x;
medida.y = y;
medida.data = data;
medida_out = [medida_out medida];
medida2figure_MIG(medida_out)
end

%% Validación
for iter = 1:Nval
    [uVal,~,~,datau] = genera_5GNR_multicarrier_v4([NRB,NRB], [30e3,30e3],...
        [10e6,10e6], Foff, fsgen, [8,8], [2*iter+101,2*iter+100]);
    uVal = scale_dBm(uVal,RMSin);

    [uLs, uRs] = separate_bands(uVal,fsmed,fsint,Foff);

    [UL, RmatL] = buildW_2GMP(uLs, uRs, [N-Qpmax1+1:N , 1:N , 1:Qnmax1],...
        Qpmax1, Qnmax1, x1_limits);
    [UR, RmatR] = buildW_2GMP(uRs, uLs, [N-Qpmax2+1:N , 1:N , 1:Qnmax2],...
        Qpmax2, Qnmax2, x2_limits);

```



```

xLs = uLs-UL(:,sL)*wL;
xRs = uRs-UR(:,sR)*wR;

x = unite_bands(xLs,xRs,fsint,fsmed,Foff);

dBm(x)

PAPRx = 20*log10(max(abs(x))/rms(x));
disp(sprintf('PAPR antes de clipping: %4.2f dB\n', PAPRx));
PAPRd = 12;
clip = (10^((PAPRd-PAPRx)/20))*max(abs(x));
x(abs(x)>clip) = clip*exp(1i*angle(x(abs(x)>clip)));
disp(sprintf('PAPR después de clipping: %4.2f dB\n', PAPR(x)));

[y, x, data] = myPA(x, dBm(x), Nmeas, Nave);

espectro(y - mean(y),fsgen,31);

[yLs, yRs] = separate_bands(y,fsmed,fsint,Foff);

figure(11); plot(dBminst(uVal), dBminst(y)- dBminst(uVal),'.');
hold on;

[PerformanceL]=analiza_medidas_OFDM_B_MIG(uLs,yLs,datau,0,fsint,fsgen,99);
[PerformanceR]=analiza_medidas_OFDM_B_MIG(uRs,yRs,datau,0,fsint,fsgen,101);
medida.descr = ['Itera_val : ' num2str(iter)];
medida.PerformanceL = PerformanceL;
medida.PerformanceR = PerformanceR;
medida.x = x;
medida.y = y;
medida.data = data;
medida_out = [medida_out medida];
medida2figure_MIG(medida_out)
end

save(filename,'medida_out','RMSin','model_id','sL','sR','mu',...
'lambda_opt_L','lambda_opt_R');

```

Código A.2 Construcción de la matriz de regresores para el modelo 2D-GMP.

```

function [Xmat, Rmat] = buildW_2GMP(x1, x2, n, Qpmax, Qnmax, x1_limits)
% This function builds the measurement matrix of the 2-D GMP-modified model.
% Only one side

%% Linear and intra-band block
Regr_a_intra = []; Regr_b_intra = []; Regr_c_intra = [];
Xa_intra = []; Xb_intra = []; Xc_intra = [];
% x1 Linear and intra-band block
indk = 0;
for k = 1:length(x1_limits.intra.Ka)
    for l = 0:x1_limits.intra.La(k)
        indk = indk + 1;
        Xa_intra(:,indk) = x1(n(1+Qpmax-l:end-Qnmax-l)).*(abs(x1(n(1+Qpmax-l:
            end-Qnmax-l)))^(x1_limits.intra.Ka(k)));
        if(x1_limits.intra.Ka(k)~=0)

```

```

        Reqr_a_intra{indk} = sprintf('x1(n-%d)*|x1(n-%d)|^{%d}',1,1,
            x1_limits.intra.Ka(k));
    else
        Reqr_a_intra{indk} = sprintf('x1(n-%d)',1);
    end
end
end

indk = 0;
for k = 1:length(x1_limits.intra.Kb)
    for l = 0:x1_limits.intra.Lb(k)
        for m = 1:x1_limits.intra.Mb(k)
            indk = indk + 1;
            Xb_intra(:,indk) = x1(n(1+Qpmax-1:end-Qnmax-1)).*(abs(x1(n(1+Qpmax-1
                -m:end-Qnmax-1-m))))^(x1_limits.intra.Kb(k)));
            Reqr_b_intra{indk} = sprintf('x1(n-%d)|x1(n-%d)|^{%d}',1,1+m,
                x1_limits.intra.Kb(k));
        end
    end
end

indk = 0;
for k = 1:length(x1_limits.intra.Kc)
    for l = 0:x1_limits.intra.Lc(k)
        for m = 1:x1_limits.intra.Mc(k)
            indk = indk + 1;
            Xc_intra(:,indk) = x1(n(1+Qpmax-1:end-Qnmax-1)).*(abs(x1(n(1+Qpmax-1
                +m:end-Qnmax-1+m))))^(x1_limits.intra.Kc(k)));
            Reqr_c_intra{indk} = sprintf('x1(n-%d)|x1(n-%d)|^{%d}',1,1-m,
                x1_limits.intra.Kc(k));
        end
    end
end

% Cross-band block
Reqr_a_cross = []; Reqr_b_cross = []; Reqr_c_cross = [];
Xa_cross = []; Xb_cross = []; Xc_cross = [];

% x1 cross-band block
indk = 0;
for k = 1:length(x1_limits.cross.Ka)
    for l = 0:x1_limits.cross.La(k)
        indk = indk + 1;
        Xa_cross(:,indk) = x1(n(1+Qpmax-1:end-Qnmax-1)).*(abs(x2(n(1+Qpmax-1:
            end-Qnmax-1))))^(x1_limits.cross.Ka(k)));
        Reqr_a_cross{indk} = sprintf('x1(n-%d)*|x2(n-%d)|^{%d}',1,1,x1_limits.
            cross.Ka(k));
    end
end

indk = 0;
for k = 1:length(x1_limits.cross.Kb)
    for l = 0:x1_limits.cross.Lb(k)
        for m = 1:x1_limits.cross.Mb(k)
            indk = indk + 1;
            Xb_cross(:,indk) = x1(n(1+Qpmax-1:end-Qnmax-1)).*(abs(x2(n(1+Qpmax-1
                -m:end-Qnmax-1-m))))^(x1_limits.cross.Kb(k)));

```

```

        Regr_b_cross{indk} = sprintf('x1(n-%d)|x2(n-%d)|^{%d}',1,1+m,
            x1_limits.cross.Kb(k));
    end
end
end

indk = 0;
for k = 1:length(x1_limits.cross.Kc)
    for l = 0:x1_limits.cross.Lc(k)
        for m = 1:x1_limits.cross.Mc(k)
            indk = indk + 1;
            Xc_cross(:,indk) = x1(n(1+Qpmax-1:end-Qnmax-1)).*(abs(x2(n(1+Qpmax-1:
                +m:end-Qnmax-1+m))))^(x1_limits.cross.Kc(k)));
            Regr_c_cross{indk} = sprintf('x1(n-%d)|x2(n-%d)|^{%d}',1,1-m,
                x1_limits.cross.Kc(k));
        end
    end
end

%% Mixed block
Regr_mixed = [];
X_mixed = [];

% x1 mixed block
indk = 0;
for k1 = 1:length(x1_limits.mixed.Ka)
    for k2 = 1:length(x1_limits.mixed.Ka)
        for l = 0:x1_limits.mixed.La(k1)
            indk = indk + 1;
            X_mixed(:,indk) = x1(n(1+Qpmax-1:end-Qnmax-1)).*(abs(x1(n(1+Qpmax-1:
                end-Qnmax-1))))^(x1_limits.mixed.Ka(k1)).*(abs(x2(n(1+Qpmax-1:
                end-Qnmax-1))))^(x1_limits.mixed.Ka(k2)));
            Regr_mixed{indk} = sprintf('x1(n-%d)*|x1(n-%d)|^{%d}*|x2(n-%d)|^{%d}
                }',1,1,x1_limits.mixed.Ka(k1),1,x1_limits.mixed.Ka(k2));
        end
    end
end

Xmat = [Xa_intra, Xb_intra, Xc_intra, Xa_cross, Xb_cross, Xc_cross, X_mixed];
Rmat = [Regr_a_intra, Regr_b_intra, Regr_c_intra, Regr_a_cross, Regr_b_cross,
    Regr_c_cross, Regr_mixed]';

end

```

Código A.3 Aplicación de la regularización de Ridge en la base PC.

```

function [nmse_val, nmse_id, ncond] = ridge_function(X,y,lambda)

indices_id = sel_indices(y,y,0.1);

%Normalización:
x_norm = vecnorm(X(indices_id,:)); %Potencia de cada colum
X_id = X(indices_id,:);
X_id_n = X_id*diag(x_norm.^(-1));

X_val = X;

```

```

%Aplicación de Ridge con las matrices aumentadas:
[~,p] = size(X_id);
X_id_gorro = [X_id_n;sqrt(lambda)*eye(p)];
R_ridge_gorro = X_id_gorro'*X_id_gorro;
y_gorro = [y(indices_id);zeros(p,1)];
h_vec = diag(x_norm.^(-1))*((inv(R_ridge_gorro)*X_id_gorro')*y_gorro);

% y en identificación y su nmse:
ymod_id = X_id* h_vec;
ncond = cond(R_ridge_gorro);
nmse_id = 20*log10(norm(ymod_id-y(indices_id))/norm(y(indices_id)));

% y en validación y su nmse:
ymod_val = X_val* h_vec;
nmse_val = 20*log10(norm(ymod_val-y)/norm(y));

end

```

Código A.4 Aplicación de la regularización de Ridge en la base DOMP.

```

function [nmse_Domp_Ridge_id, nmse_Domp_Ridge_val] = ridge_RCDOMP(X,T,y,lambda)

nmse = @(ymod,y) 20*log10(norm(ymod-y)/norm(y));

indices_id = sel_indices(y,y,0.1);

%Normalización:
x_norm = vecnorm(X(indices_id,:)); %Potencia de cada columna
X_id = X(indices_id,:);
X_id_n = X_id*diag(x_norm.^(-1));

X_val = X;

%Ortogonalización:
Z = X_id_n*T;

%Proyección en la base DOMP:
h_orth = Z'*y(indices_id);

%Regularización:
h_orth = h_orth/(1+lambda);

%Cambio de la base DOMP a la base PC:
h_volterra = diag(vecnorm(X_id).^(-1))*T*h_orth;

% y en identificación y su nmse:
yest = X_id*h_volterra;
nmse_Domp_Ridge_id = nmse(yest,y(indices_id));

% y en validación y su nmse:
yest_val = X_val*h_volterra;
nmse_Domp_Ridge_val = nmse(yest_val,y);

end

```

Código A.5 Generación de las gráficas de NMSE de identificación y validación.

```

modo = 'Ridge_en_PC'; %'Ridge_en_DOMP', 'SelectRCDOMP+Ridge'

lambda = logspace(-16, 1, 100);
switch modo
    case 'Ridge_en_PC'
        for i = 1:length(lambda)
            [nmse_val_L(i),nmse_id_L(i),~]=ridge_function(UL,yLs,0.1,lambda(i));
            [nmse_val_R(i),nmse_id_R(i),~]=ridge_function(UR,yRs,0.1,lambda(i));
        end
    case 'Ridge_en_DOMP'
        for i = 1:length(lambda)
            [nmse_Domp_Ridge_L_id(i),nmse_Domp_Ridge_L_val(i)] = ridge_RCDOMP(UL
                (:,sL(1:noptL)), TL(sL(1:noptL),sL(1:noptL)), yLs, lambda(i));
            [nmse_Domp_Ridge_R_id(i),nmse_Domp_Ridge_R_val(i)] = ridge_RCDOMP(UR
                (:,sR(1:noptR)), TR(sR(1:noptR),sR(1:noptR)), yRs, lambda(i));
        end
    case 'SelectRCDOMP+Ridge'
        for i = 1:length(lambda)
            [nmse_val_L(i),nmse_id_L(i),~]=ridge_function(UL(:,sL(1:noptL)),yLs
                ,0.1,lambda(i));
            [nmse_val_R(i),nmse_id_R(i),~]=ridge_function(UR(:,sR(1:noptR)),yRs
                ,0.1,lambda(i));
        end
end

figure(400)
title('Banda L')
switch modo
    case 'Ridge_en_PC'
        semilogx(lambda, nmse_id_L);
        hold on
        semilogx(lambda, nmse_val_L);
        legend('NMSE_{id}','NMSE_{val}');
    case 'Ridge_en_DOMP'
        semilogx(lambda, nmse_Domp_Ridge_L_id);
        hold on
        semilogx(lambda, nmse_Domp_Ridge_L_val);
        legend('NMSE_{espacio-DOMP-id}','NMSE_{espacio-DOMP-val}');
    case 'SelectRCDOMP+Ridge'
        semilogx(lambda, nmse_id_L);
        hold on
        semilogx(lambda, nmse_val_L);
        legend('NMSE_{id} Ridge en PC','NMSE_{val} Ridge en PC');
end

figure(401)
title('Banda R')
switch modo
    case 'Ridge_en_PC'
        semilogx(lambda, nmse_id_R);
        hold on
        semilogx(lambda, nmse_val_R);
        legend('NMSE_{id}','NMSE_{val}');
    case 'Ridge_en_DOMP'
        semilogx(lambda, nmse_Domp_Ridge_R_id);

```

```
    hold on
    semilogx(lambda, nmse_Domp_Ridge_R_val);
    legend('NMSE_{espacio-DOMP-id}', 'NMSE_{espacio-DOMP-val}');
case 'SelectRCDOMP+Ridge'
    semilogx(lambda, nmse_id_R);
    hold on
    semilogx(lambda, nmse_val_R);
    legend('NMSE_{id} Ridge en PC', 'NMSE_{val} Ridge en PC');
end
```

Apéndice B

Artículo de investigación

La realización de este trabajo me ha permitido ser coautor, junto con mi compañero Miguel Nogales, del artículo que se muestra a continuación. Todo ha sido gracias a la iniciativa del Grupo de Sistemas de Radiocomunicación. Este artículo ha sido aceptado en la XXXVI edición del Simposio Nacional de la URSI.

Arquitectura de aprendizaje directo en DPD de doble banda con regularización de Ridge

Miguel Nogales, Elías Marqués-Valderrama, Juan A. Becerra, María J. Madero-Ayora, Carlos Crespo-Cadenas
 mignoggon@alum.us.es, elimarval@alum.us.es, jabecerra@us.es, mjmadero@us.es, ccrespo@us.es
 Dpto. de Teoría de la Señal y Comunicaciones. Universidad de Sevilla. Camino de los descubrimientos s/n. 41092. Sevilla.

Resumen—In this paper, a direct learning architecture is employed for the digital predistortion of a power amplifier with a concurrent dual-band input signal. Conventional dual-band behavioral models need to cope with a large number of regressors, being prone to algebraic issues that arise in the computation of the inverse of the covariance matrix. Ridge regularization is applied to the modelling equation in order to improve the stability of the learning process and the precision of the estimated coefficients in validation. Experimental results for the linearization performance achieved with the proposed approach are provided for a class AB power amplifier driven by two concurrent 10-MHz 5G New Radio signals.

I. INTRODUCCIÓN

A medida que las generaciones de comunicaciones móviles han evolucionado se han establecido requerimientos de cada vez mayores tasas de datos. Con este fin, el uso de esquemas de modulación eficientes como OFDM permiten un aprovechamiento óptimo del espectro radioeléctrico. La señales basadas en este esquema de modulación se caracterizan por tener una alta relación potencia promedio a potencia de pico (PAPR), lo que obliga al amplificador de potencia a trabajar en régimen no lineal. Técnicas como la predistorsión digital (DPD) permiten mitigar la distorsión generada por la no linealidad, evitando así emisiones fuera de banda no deseadas [1].

Otra estrategia ampliamente utilizada en arquitectura de comunicaciones inalámbricas es la multiplexación en el dominio de la frecuencia. La combinación de esta técnica con el acceso múltiple por división en frecuencia FDMA permite una gran flexibilidad de asignación de recursos, como ocurre en la capa física de la quinta generación de comunicaciones móviles (5G-NR). El uso de DPD con este tipo de esquemas multibanda presenta un gran reto, ya que el número de coeficientes del modelo se dispara y aparecen problemas numéricos que traen inestabilidad a la regresión.

En el caso de modelos multibanda, los regresores de Volterra tienen la forma de productos de los equivalentes paso bajo de cada banda y sus envolventes con distintas combinaciones de retrasos [2], [3]. Dicha combinatoria es muy rica, y además de generar un número ingente de coeficientes crea un conjunto de regresores del modelo muy parecidos entre sí, lo que lleva a un mal condicionamiento de la matriz de medida en la regresión [4].

El enfoque principal a la hora de abordar una regresión con una matriz mal condicionada consiste en la regularización, que se basa en modificar ligeramente la función de coste que se pretende minimizar para que la dependencia de los coeficientes con las señales de entrada disminuya. La regularización de Ridge es una de las técnicas clásicas de este tipo [5], [6].

En este trabajo exploramos el uso de regularización de Ridge en un esquema de aprendizaje directo (DLA) multibanda. Para ello, en la Sección II se introducen los conceptos básicos de la DLA y cómo se extienden en un contexto multibanda. A continuación, en la Sección III se establece el marco de trabajo de regresión de Ridge. La Sección IV presenta resultados experimentales de predistorsión de un amplificador de potencia comercial. Finalmente, la Sección V cubre las conclusiones y finaliza el presente trabajo.

II. ARQUITECTURA DE APRENDIZAJE DIRECTO EN DOBLE BANDA

El esquema tradicional de DLA consta principalmente de un bucle cuya tarea es predistorsionar cierta señal de entrada \mathbf{u} ,

$$\mathbf{x} = \mathbf{u} - \mathbf{d}, \quad (1)$$

de modo que se le restará la señal llamada de distorsión, \mathbf{d} , y dará como resultado la envolvente compleja de la señal que se empleará como entrada al amplificador, \mathbf{x} . La señal \mathbf{d} se define como

$$\mathbf{d} = \mathbf{U}\mathbf{w}, \quad (2)$$

donde la información adquirida por el sistema será almacenada en el vector de pesos \mathbf{w} mientras que la matriz \mathbf{U} contendrá los regresores, que serán generados siguiendo la estructura del modelo adoptado, entre los que el modelo *generalized memory polynomial* (GMP) [7] es uno de los más ampliamente utilizados. En cuanto al cálculo de los coeficientes \mathbf{w} en cada iteración,

$$\mathbf{w}_{i+1} = \mathbf{w}_i + \mu \mathbf{U}^\dagger \mathbf{e}, \quad (3)$$

este será realizado mediante la pseudoinversa de Moore-Penrose

$$\mathbf{U}^\dagger = (\mathbf{U}^H \mathbf{U})^{-1} \mathbf{U}^H = \mathbf{R}^{-1} \mathbf{U}^H, \quad (4)$$

donde \mathbf{R} es la matriz de correlación y la actualización estará regulada por un coeficiente que determinará la velocidad del aprendizaje, μ , cuyo valor pertenecerá al intervalo (0, 1). El lazo se cierra obteniendo el error,

$$\mathbf{e} = \frac{\mathbf{y}}{g_0} - \mathbf{u}, \quad (5)$$

siendo \mathbf{y} la envolvente compleja a la salida del amplificador de potencia en esa iteración, mientras que g_0 es la ganancia estática, obtenida del cociente entre las normas de \mathbf{y} y \mathbf{u} .

Una vez definido el sistema DLA de una sola banda, el sistema dual surge del primero duplicando sus bloques de procesamiento de señal, dedicando uno a cada banda. Esta separación de los sistemas en bandas da lugar a varias implicaciones, como la necesidad de trabajar con el doble de variables

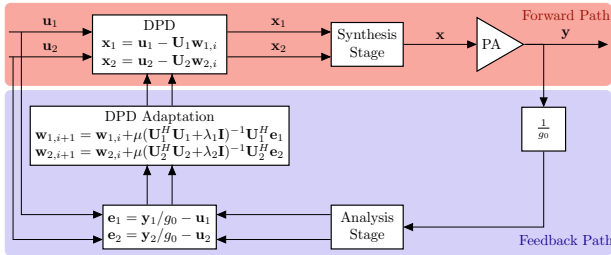


Fig. 1: Esquema de linealización con aprendizaje directo y regularización de Ridge para una señal de doble banda concurrente.

o generar los regresores con modelos de dos bandas, que de forma general contemplan términos mezclados y cruzados. Un ejemplo de este tipo de modelos es el GMP bidimensional (2D-GMP) [8]. Este cambio aportará importantes mejoras en rendimiento para el sistema debido a las interacciones entre las dos bandas.

Un ejemplo del esquema final se muestra en la Fig. 1, donde se puede apreciar la similitud con el sistema adaptado a una sola banda, pero con los añadidos propios del modelo dual. Así pues, siguiendo dicho esquema, se denota con el subíndice 1 a la banda con frecuencia central inferior y con el subíndice 2 a la banda con la frecuencia central superior.

III. REGULARIZACIÓN DE RIDGE

En esta sección se introduce la técnica de regularización con carácter general, para ser aplicada posteriormente en la arquitectura DLA. Un problema de modelado tiene la siguiente forma:

$$\mathbf{y} = \mathbf{X} \mathbf{h} + \mathbf{e}, \quad (6)$$

donde \mathbf{X} es la matriz de regresores, \mathbf{h} los coeficientes del modelo y \mathbf{e} el error. El objetivo de este problema será encontrar los coeficientes que hagan que el error sea nulo. Dado que se trata de una estimación es aconsejable emplear la notación $\hat{\mathbf{y}} = \mathbf{X} \hat{\mathbf{h}}$, donde $\hat{\mathbf{h}}$ denota estimación.

El gran número de regresores que se requieren para poder abordar un modelo de doble banda, puede provocar que el estimador de mínimos cuadrados (LS) no sea único. Además la existencia de regresores muy correlados conduce a que la matriz de correlación generalmente sea singular o cercana a serlo, es decir, presenta un número de condición muy elevado. Esto quiere decir que, en el caso en el que sea invertible, ligeros cambios en \mathbf{X} desembocarían en grandes cambios en $(\mathbf{R})^{-1}$. Esto se traduce en que puede obtenerse una buena estimación de los coeficientes del modelo en test pero malos para validación.

Así pues, la solución que se presenta es la regularización de Ridge, que si bien abandona la idea de conseguir un estimador insesgado, es capaz de llegar a un compromiso entre el sesgo y varianza en la estimación. A continuación se muestra la deducción de la expresión del nuevo modelo a partir de una penalización en la suma residual de cuadrados (PRSS),

$$\text{PRSS} = \|\mathbf{y} - \hat{\mathbf{y}}\|_2^2 + \lambda \|\hat{\mathbf{h}}\|_2^2, \quad (7)$$

donde λ es el coeficiente de regularización. Como el objetivo es encontrar el mínimo de la función PRSS, a continuación, se deriva respecto de \mathbf{h} y se iguala a 0,

$$\frac{d\text{PRSS}}{d\mathbf{h}} = -2\mathbf{X}^H \mathbf{y} + 2\mathbf{X}^H \mathbf{X} \hat{\mathbf{h}} + 2\lambda \hat{\mathbf{h}} = 0, \quad (8)$$

y finalmente se despeja el modelo regularizado $\hat{\mathbf{h}}$:

$$\hat{\mathbf{h}} = (\mathbf{X}^H \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^H \mathbf{y}, \quad (9)$$

donde \mathbf{I} es la matriz identidad.

De manera que, observando la ecuación (7), se extraen varios análisis. El primero de ellos es que la minimización ya no consiste en aumentar los coeficientes \hat{h}_j , ya que ahora aparece un término de penalización regulado por λ que aumenta con $\|\hat{\mathbf{h}}\|_2$. Esto puede interpretarse como una restricción en la norma ℓ_2 del vector de coeficientes. El segundo análisis se induce del primero, y es que el hecho de restringir $\|\hat{\mathbf{h}}\|_2$ está sesgando la estimación, y esto asegura la invertibilidad de la matriz de regresores regularizada.

La implicación que esto tiene en el modelado se comprende sobre el espacio de componentes principales (PC) de \mathbf{X} . La regresión consiste en proyectar y en el subespacio que contiene a \mathbf{X} , cuya base en este desarrollo se considera el espacio de PC de \mathbf{X} . Dado que $\hat{\mathbf{h}}$ representa las coordenadas de la proyección en dicha base, el hecho de introducir la misma penalización en todas las componentes de la diagonal de la matriz de correlación provoca una restricción mayor en aquellas componentes con menor variabilidad en $\hat{\mathbf{y}}$. De modo que además de conseguir mayor estabilidad y fiabilidad en la regresión sirve como método de simplificación en cuanto a su complejidad.

A. Aplicación de la regularización de Ridge en DLA

Generalmente la matriz de regresores se denota como \mathbf{X} , hecho que motiva la notación empleada en este apartado. En cambio la arquitectura DLA representada en la Fig. 1 entiende que la entrada al sistema es \mathbf{u} en lugar de \mathbf{x} . Esto es debido a que tradicionalmente se emplea \mathbf{x} para indicar la entrada en el amplificador. Para no llevar a confusión se ha decidido que \mathbf{u} represente la señal generada, cuya matriz de regresores es \mathbf{U} , de modo que \mathbf{x} es la señal predistorsionada según (1).

IV. RESULTADOS EXPERIMENTALES

En esta sección se pondrá a prueba experimentalmente la solución teórica propuesta en las Secciones II y III. En primera instancia se hace una breve explicación del experimento realizado para proceder posteriormente a la explicación de los resultados obtenidos.

A. Descripción del banco de medidas

Para confirmar la validez de la propuesta presentada en este trabajo, se evaluaron sus prestaciones para linealización mediante un banco de medidas compuesto por un generador de señal SMU200A de Rohde & Schwarz, dos pre-amplificadores TVA-4W-422A+ de Minicircuits en cascada, un analizador de señal PXA-N9030A de Keysight Technologies y una fuente de alimentación. Como amplificador bajo prueba se empleó la placa de evaluación de un amplificador de clase AB basado en el HEMT de GaN CGH40010 de Cree, con una frecuencia de trabajo de 3.6 GHz.

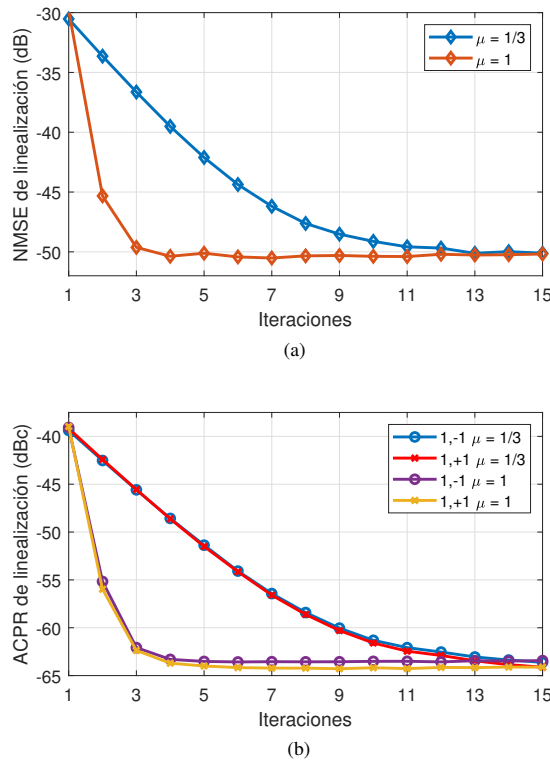


Fig. 2: Evolución del (a) NMSE de linealización y (b) ACPR de la banda 1 para valores de $\mu = 1/3$ y 1 frente al número de iteraciones.

La señal de sondeo de doble banda estaba compuesta por dos señales OFDM de acuerdo con el estándar 5G-NR, cada una con un ancho de banda de 10 MHz y una separación entre subportadoras de 30 kHz, presentando una PAPR en torno a 11 dB. Antes de combinar las señales de banda base, generadas con una tasa de muestreo $f_s = 92.16$ MHz, se aplicó un desplazamiento de frecuencia $-f_s/4$ a la señal de la banda 1 y un desplazamiento $+f_s/4$ a la señal de la banda 2. Estos valores se escogieron para optimizar el rango de frecuencias observables con el equipamiento disponible.

En el analizador, se adquirieron muestras de la envolvente compleja de la señal a la salida del amplificador bajo prueba considerando las portadoras de cada una de las bandas y una tasa de muestreo $f_s/2$, lo que supone un factor de sobremuestreo de 3 para cada banda. El ruido de medida se redujo promediando 10 realizaciones de la señal adquirida y posteriormente se sincronizaron las señales de entrada y salida. Cabe mencionar que, en la arquitectura de aprendizaje directo empleada para predistorsión digital, a las señales predistorsionadas se les aplicó una técnica de *clipping* que impedía que su PAPR superara los 12 dB como protección.

B. Comportamiento del modelo

Se planteó una configuración del modelo 2D-GMP [8], con 534 regresores en cada banda, lo que supone un número demasiado elevado y da lugar a problemas numéricos, de manera que, sin la aplicación de la regularización de Ridge, el funcionamiento de la arquitectura DLA tiende a la diver-

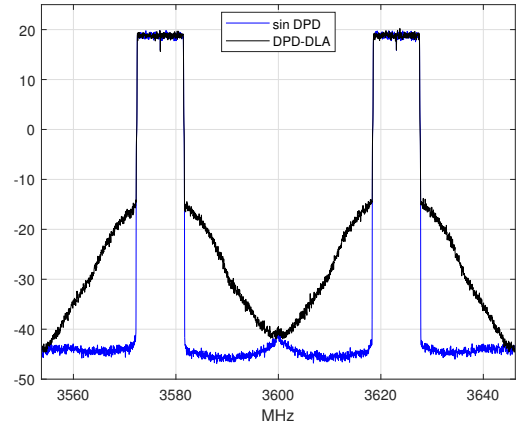


Fig. 3: Densidad espectral de potencia de la señal de salida del amplificador sin emplear DPD frente al uso de DPD.

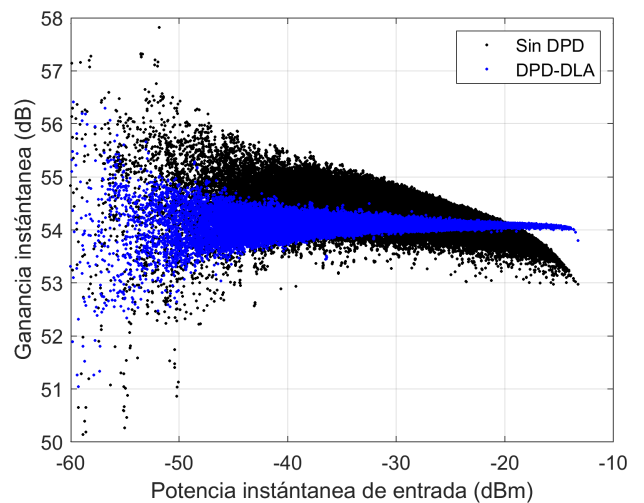


Fig. 4: AM-AM sin emplear DPD comparado con AM-AM al emplear DPD-DLA.

gencia. A este mismo modelo se le aplicó la regularización de Ridge, teniendo como resultado las prestaciones que se detallan a continuación.

En primer lugar, el error cuadrático medio normalizado (NMSE) frente al número de iteraciones DLA se muestra en la Fig. 2(a). Se observa un decrecimiento a medida que avanzan las iteraciones de aprendizaje, siendo más rápido a mayor valor del paso de aprendizaje μ . Por otro lado, en la Fig. 2(b) se muestra el comportamiento de la relación de potencia de canal adyacente (ACPR), que experimenta una evolución muy similar a la del NMSE. Esto quiere decir que se ha conseguido mitigar el recrecimiento espectral, tal y como se muestra en Fig. 3.

Por otro lado cabe destacar que el hecho de que la regularización de Ridge establezca el comportamiento del bucle de control en la arquitectura DLA permite aumentar μ a su valor máximo. Esto implica que es posible alcanzar el valor óptimo de los coeficientes w_i en un menor número de iteraciones, tal y como se observa en la Fig. 2.

TABLA I: Indicadores de Linealización para los Casos sin DPD y con DPD-DLA.

Caso	NMSE ₁	NMSE ₂	ACPR _{1,-1}	ACPR _{1,+1}	ACPR _{2,-1}	ACPR _{2,+1}	EVM ₁	EVM ₂	P _{out}	PAE
	(dB)		(dBc)				(%)		(dBm)	(%)
Sin DPD	-30.5	-30.0	-39.4	-39.2	-38.7	-38.7	2.7	2.8	30.1	17.9
DPD-DLA	-49.7	-49.8	-62.8	-63.3	-62.8	-62.1	1.2	1.2	30.1	18.0

La linealización también se puede observar en la Fig. 4, la cual relaciona potencia instantánea de entrada con la ganancia instantánea, y donde el caso óptimo se encuentra al conseguir que la nube de puntos sea horizontal, lo cual se logra con este método.

Por último, el rendimiento de la arquitectura empleada se muestra en la Tabla I. Es importante destacar que la eficiencia de potencia añadida (PAE) no se ha visto alterada. Esto quiere decir que el empleo de estas técnicas no empeoran la eficiencia del amplificador. También resalta la reducción de la magnitud del vector de error (EVM), que implica una mejora en la recepción de la señal, siendo sinónimo de una menor dispersión en la constelación de la señal recibida.

V. CONCLUSIONES

La aplicación de la arquitectura DLA junto con la regularización de Ridge ha resultado en una mejora en prestaciones de linealización, ya que la señal a la salida del amplificador cumple con el criterio de NMSE y las máscaras espectrales impuestas por 5G-NR sin degradar su eficiencia. Además, la combinación del 2D-GMP con la regularización de Ridge ha conseguido regresionar un modelo que a priori era inestable por la gran cantidad de regresores que contiene.

Por otro lado, en favor de la técnica de regularización de Ridge, cabe mencionar que la única modificación en el modelo ha sido la inserción de la constante en la diagonal de la matriz de correlación. Su implementación es muy sencilla y garantiza estabilidad en situaciones en las que el número de regresores es demasiado elevado.

REFERENCIAS

- [1] A. Katz, J. Wood, and D. Chokola, "The evolution of PA linearization: From classic feedforward and feedback through analog and digital predistortion," *IEEE Microw. Mag.*, vol. 17, no. 2, pp. 32–40, Feb. 2016.
- [2] P. Roblin, S. K. Myoung, D. Chaillot, Y. G. Kim, A. Fathimulla, J. Strahler, and S. Bibyk, "Frequency-selective predistortion linearization of RF power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 56, no. 1, pp. 65–76, Jan. 2008.
- [3] S. A. Bassam, W. Chen, M. Helaoui, F. M. Ghannouchi, and Z. Feng, "Linearization of concurrent dual-band power amplifier based on 2D-DPD technique," *IEEE Microw. Wireless Compon. Lett.*, vol. 21, no. 12, pp. 685–687, Dec. 2011.
- [4] P. L. Gilabert, R. N. Braithwaite, and G. Montoro, "Beyond the Moore-Penrose inverse: Strategies for the estimation of digital predistortion linearization parameters," *IEEE Microw. Mag.*, vol. 21, no. 12, pp. 34–46, Dec. 2020.
- [5] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, ser. Springer Series in Statistics. New York, NY, USA: Springer New York Inc., 2001.
- [6] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning – with Applications in R*, ser. Springer Texts in Statistics. New York: Springer, 2013, vol. 103.
- [7] D. Morgan, Z. Ma, J. Kim, M. Zierdt, and J. Pastalan, "A generalized memory polynomial model for digital predistortion of RF power amplifiers," *IEEE Transactions on Signal Processing*, vol. 54, no. 10, pp. 3852–3860, Oct. 2006.
- [8] S. A. Bassam, M. Helaoui, and F. M. Ghannouchi, "2-D digital predistortion (2-D-DPD) architecture for concurrent dual-band transmitters," *IEEE Trans. Microw. Theory Techn.*, vol. 59, no. 10, pp. 2547–2553, Oct. 2011.

Índice de Figuras

2.1	Representación de un sistema	3
2.2	Sistema invariante en el tiempo	4
2.3	Respuesta impulsiva de un sistema	4
2.4	Potencia de salida frente a potencia de entrada con tres ganancias diferentes	4
2.5	Sistema no lineal	5
2.6	Descomposición de la salida y_{NL} en tres de sus armónicos	6
2.7	Resultado de la prueba de dos tonos. En azul se presentan las frecuencias de interés f_1 y f_2 , en rojo el segundo y tercer armónico y en discontinuo los productos de intermodulación de segundo y tercer orden	7
2.8	Ganancia de un amplificador no lineal	8
2.9	Espectro de señal OFDM con 5 subportadoras activas	10
2.10	Implementación de una modulación OFDM [4]	11
3.1	Descomposición del funcional $y_p[k]$ como superposición de filtrados lineales. En la figura $u(k)$ representa $x[k]$, y $m - 1$ es equivalente a Q . $G_{n_1, \dots, n_{p-1}}(\omega)$ es la transformada de Fourier en tiempo discreto de $g_{n_1, \dots, n_{p-1}}^p[k]$ [9]	15
3.2	Espectro de la señal en: (a) la entrada y (b) la salida de un transmisor dual. [12]	16
3.3	Esquema conceptual de aplicación del DPD	20
3.4	Implementación de DPD con arquitectura de aprendizaje indirecto	20
3.5	Implementación de DPD con arquitectura de aprendizaje directo	21
3.6	Implementación de DPD con arquitectura de aprendizaje directo para doble banda concurrente	21
4.1	Matriz de autocorrelaciones muestrales para una de las bandas empleando un modelo 2D-GMP	24
4.2	Representación geométrica de las ecuaciones de proyección. En rojo se indica la base sobre la que se proyecta \hat{y} en cada caso, σ_i representa la columna i -ésima de la matriz \mathbf{U} y x_i la columna i -ésima de la matriz \mathbf{X}	29
5.1	Monitor del analizador de espectros	31
5.2	Banco de pruebas (a) foto del montaje, (b) esquemático por bloques	32
5.3	Densidad espectral de potencia de la señal de entrada multiplicada por la ganancia lineal y la salida sin aplicar DPD	33
5.4	Densidad espectral de potencia de la señal de salida para cada iteración del DPD	35
5.5	Esquema de implementación de la regularización de Ridge en un escenario de DPD en DLA con doble banda	35
5.6	Cálculo del λ_{opt} de la banda inferior en función del NMSE de validación	36
5.7	Comparativa entre el espectro a la salida del PA sin aplicar DPD y habiendo aplicado DPD	37
5.8	Evolución del (a) NMSE de linealización y (b) ACPR de la banda 1 para valores de $\mu = 1/3$ y 1 frente al número de iteraciones	38
5.9	Linealización de la ganancia instantánea al emplear DPD	38
5.10	Curvas de NMSE para identificación y validación en función de λ	39
5.11	Curvas de NMSE para identificación y validación en función de λ	39

6.1 Banco de pruebas (a) foto del montaje, (b) esquemático por bloques

42

Índice de Tablas

5.1	Parámetros medidos sin DPD	34
5.2	Configuración del modelo 2D-GMP	34
5.3	Parámetros medidos con DPD en 5 iteraciones. Caso inestable	34
5.4	Parámetros medidos aplicando DPD y Ridge, con $\lambda_{opt,1} = 1.69 \cdot 10^{-5}$, $\lambda_{opt,2} = 2.28 \cdot 10^{-5}$ tras 15 iteraciones con $\mu = 1/3$	37

Índice de Códigos

A.1	Código principal, donde se realiza la generación de la señal 5G-NR, cálculo del modelo, empleo de la arquitectura DLA con las respectivas medidas sobre el amplificador y aplicación de la regularización de Ridge a la matriz de regresores	43
A.2	Construcción de la matriz de regresores para el modelo 2D-GMP	47
A.3	Aplicación de la regularización de Ridge en la base PC	49
A.4	Aplicación de la regularización de Ridge en la base DOMP	50
A.5	Generación de las gráficas de NMSE de identificación y validación	51

Bibliografía

- [1] J. Pedro and N. Carvalho, *Intermodulation distortion in microwave and wireless circuits*. Boston: Artech House, Jun. 2003.
- [2] *Apuntes de la asignatura Sistemas de Infraestructura de Telecomunicación. Grado en Ingeniería de las Tecnologías de Telecomunicación (ETSI). Departamento de Teoría de la Señal y Comunicaciones. Universidad de Sevilla*, Jan. 2020.
- [3] R. W. Chang, "Synthesis of band-limited orthogonal signals for multichannel data transmission," *The Bell System Technical Journal*, vol. 45, no. 10, pp. 1775–1796, Dec. 1966.
- [4] A. Molisch, *Wireless communications*. Chichester: John Wiley, Dec. 2010.
- [5] D. Morgan, Z. Ma, J. Kim, M. Zierdt, and J. Pastalan, "A generalized memory polynomial model for digital predistortion of rf power amplifiers," *IEEE Transactions on Signal Processing*, vol. 54, no. 10, pp. 3852–3860, Oct. 2006.
- [6] P. Celka, N. Bershad, and J.-M. Vesin, "Stochastic gradient identification of polynomial wiener systems: analysis and application," *IEEE Transactions on Signal Processing*, vol. 49, no. 2, pp. 301–313, Feb. 2001.
- [7] C. Crespo-Cadenas, M. J. Madero-Ayora, J. Reina-Tosina, and J. Becerra-González, "Formal deduction of a volterra series model for complex-valued systems," *Signal Processing*, vol. 131, pp. 245–248, 2017. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0165168416302043>
- [8] C. Crespo-Cadenas, M. J. Madero-Ayora, and J. A. Becerra, "A bivariate volterra series approach to modeling and linearization of power amplifiers," in *2021 IEEE Topical Conference on RF/Microwave Power Amplifiers for Radio and Wireless Applications (PAWR)*, 2021, pp. 4–7.
- [9] G. Raz and B. van Veen, "Baseband volterra filters for implementing carrier based nonlinearities," *IEEE Transactions on Signal Processing*, vol. 46, no. 1, pp. 103–114, Jan. 1998.
- [10] L. Ding, G. Zhou, D. Morgan, Z. Ma, J. Kenney, J. Kim, and C. Giardina, "A robust digital baseband predistorter constructed using memory polynomials," *IEEE Transactions on Communications*, vol. 52, no. 1, pp. 159–165, Jan. 2004.
- [11] J. Kim and K. Konstantinou, "Digital predistortion of wideband signals based on power amplifier model with memory," *IEEE Transactions on Signal Processing*, vol. 37, no. 23, pp. 1417–1418, Nov. 2001.
- [12] S. A. Bassam, M. Helaoui, and F. M. Ghannouchi, "2-D digital predistortion (2-D-DPD) architecture for concurrent dual-band transmitters," *IEEE Trans. Microw. Theory Techn.*, vol. 59, no. 10, pp. 2547–2553, Oct. 2011.
- [13] H. Enzinger, K. Freiberger, and C. Vogel, "Competitive linearity for envelope tracking: Dual-band crest factor reduction and 2d-vector-switched digital predistortion," *IEEE Microwave Magazine*, vol. 19, no. 1, pp. 69–77, Feb. 2018.

- [14] C. Eun and E. Powers, "A new volterra predistorter based on the indirect learning architecture," *IEEE Transactions on Signal Processing*, vol. 45, no. 1, pp. 223–227, Jan. 1997.
- [15] D. Zhou and V. DeBrunner, "A novel adaptive nonlinear predistorter based on the direct learning algorithm," in *2004 IEEE International Conference on Communications (IEEE Cat. No.04CH37577)*, vol. 4, Jun. 2004, pp. 2362–2366 Vol.4.
- [16] *Apuntes de la asignatura Métodos Matemáticos. Grado en Ingeniería de las Tecnologías de Telecomunicación (ETSI). Departamento de Matemática Aplicada II. Universidad de Sevilla*, Sep. 2019.
- [17] R. T. T. Hastie and R. Friedman, *The elements of statistical learning: data mining, inference, and prediction*, 2nd ed. Springer, 2017.
- [18] T. H. G. James, D. Witten and R. Tibshirani, *An Introduction to Statistical Learning with Applications in R*, 1st ed., ser. New York. Springer, 2013.
- [19] J. M. Virués-Gavira, *Matemáticas I: Grado en Ingeniería de las Tecnologías de Telecomunicación*. Sevilla: Sección de Publicaciones, Escuela Técnica Superior de Ingeniería, 2018.
- [20] N. W. van Wieringen, "Lecture notes on ridge regression," 2015.
- [21] J. Reina-Tosina, M. Allegue-Martínez, C. Crespo-Cadenas, C. Yu, and S. Cruces, "Behavioral modeling and predistortion of power amplifiers under sparsity hypothesis," *IEEE Transactions on Microwave Theory and Techniques*, vol. 63, no. 2, pp. 745–753, Feb. 2015.
- [22] C.-L. T. Allan D R McQuarrie, *Regression and Times Series Model Selection*. Singapore: World Scientific, 1998.
- [23] J. A. Becerra, M. J. Madero-Ayora, R. G. Noguer, and C. Crespo-Cadenas, "On the optimum number of coefficients of sparse digital predistorters: A bayesian approach," *IEEE Microwave and Wireless Components Letters*, vol. 30, no. 12, pp. 1117–1120, Dec. 2020.
- [24] J. A. Becerra, M. J. Madero-Ayora, J. Reina-Tosina, C. Crespo-Cadenas, J. García-Frías, and G. Arce, "A doubly orthogonal matching pursuit algorithm for sparse predistortion of power amplifiers," *IEEE Microwave and Wireless Components Letters*, vol. 28, no. 8, pp. 726–728, Aug. 2018.
- [25] J. A. Becerra, M. J. Madero-Ayora, and C. Crespo-Cadenas, "Comparative analysis of greedy pursuits for the order reduction of wideband digital predistorters," *IEEE Transactions on Microwave Theory and Techniques*, vol. 67, no. 9, pp. 3575–3585, Sep. 2019.
- [26] J. A. Becerra, M. J. Madero-Ayora, J. Reina-Tosina, C. Crespo-Cadenas, J. Garcia-Frias, and G. Arce, "A reduced-complexity doubly orthogonal matching pursuit algorithm for power amplifier sparse behavioral modeling," *IEEE Topical Conference on RF/Microwave Power Amplifiers for Radio and Wireless Applications (PAWR)*, pp. 1–3, Jan. 2019.
- [27] J. A. Becerra, M. J. M. Ayora, J. Reina-Tosina, and C. Crespo-Cadenas, "Sparse identification of volterra models for power amplifiers without pseudoinverse computation," *IEEE Transactions on Microwave Theory and Techniques*, vol. 68, no. 11, pp. 4570–4578, Aug. 2020.
- [28] D. Byrne, R. Farrell, and J. Dooley, "Sample pruning based on leverage for digital pre-distortion," in *2020 International Workshop on Integrated Nonlinear Microwave and Millimetre-Wave Circuits (INMMiC)*, Jul. 2020, pp. 1–3.

Glosario

- 2D-GMP** *Bidimensional Generalized Memory Polynomial*. 17, 34
- 2D-MP** *Bidimensional Memory Polynomial*. 17
- ACPR** Relación de potencia con el canal adyacente. 9, 10, 33, 34, 37
- BIC** *Bayesian Information Criterion*. 26
- CA** Agregación de portadora. 1, 12, 16
- CoSaMP** *Compressive Sampling Matching Pursuit*. 27
- CP** Prefijo cíclico. 12
- DLA** Arquitectura de Aprendizaje Directo. 21, 35, 36, 38
- DOMP** *Doubly Orthogonal Matching Pursuit*. 27, 28, 29, 30, 36, 38, 40
- DPD** *Digital Predistortion*. 19, 20, 21, 34, 35, 36, 37, 38
- DUT** Dispositivo Bajo Prueba. 6, 32
- ETSI** Escuela Técnica Superior de Ingeniería. 31
- EVM** Magnitud del vector error. 9, 10, 37
- FPGA** *Field-programmable gate array*. 15
- FV** *Full Volterra*. 15, 18
- GMP** *Generalized Memory Polynomial*. 16, 17, 18
- IFFT** Transformada rápida de Fourier Inversa. 11
- ILA** Arquitectura de Aprendizaje Indirecto. 20, 21
- LS** Mínimos cuadrados. 19, 23, 29
- LTI** Lineal e Invariante en el Tiempo. 3, 4, 5, 6
- MP** *Memory Polynomial*. 16, 17
- NMSE** Error cuadrático medio normalizado. 10, 35, 36, 37, 39
- OFDM** Multiplexación por división de frecuencias ortogonales. 10, 11, 12, 33
- OMP** *Orthogonal Matching Pursuit*. 26, 27, 28

- PA** Amplificador de potencia. 1, 2, 4, 5, 14, 32, 33
- PAE** Eficiencia de potencia añadida. 8, 37
- PAPR** Relación potencia pico a potencia promedio. 12, 33, 34
- PC** Componentes principales. 25, 26, 29, 30, 36, 40
- PRSS** Suma residual de cuadrados penalizada. 24, 25
-
- RB** Bloque de recursos. 12
- RC-DOMP** *Reduced Complexity DOMP*. 28, 30, 38, 40
- RF** Radiofrecuencia. 14, 22
- RSS** Suma residual de cuadrados. 19
-
- SP** *Subspace Pursuit*. 27
- SVD** Descomposición en valores singulares. 25
-
- THD** Distorsión Armónica Total. 7, 9
-
- VSA** Analizador Vectorial de Señal. 31, 32, 33
- VSG** Generador Vectorial de Señal. 31