

# Trabajo Fin de Máster Máster Universitario en Ingeniería de Telecomunicación

## Predistorsión de amplificadores de potencia con técnicas de aprendizaje distribuido

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, 2023





Trabajo Fin de Máster  
Máster Universitario en Ingeniería de Telecomunicación

# **Predistorsión de amplificadores de potencia con técnicas de aprendizaje distribuido**

Autor:

Elías Marqués Valderrama

Tutor:

Juan Antonio Becerra González

Profesor Titular 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, 2023



Trabajo Fin de Máster: Predistorsión de amplificadores de potencia con técnicas de aprendizaje distribuido

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

---

Con este trabajo finaliza una etapa que ha estado llena de dificultades pero también de alegrías. Sé que he salido adelante porque he estado apoyado y rodeado de personas que lo han hecho posible. Han sido vitales y es por esto que quiero dedicarles unas palabras.

En primer lugar, quiero darle las gracias a mis padres, Antonio y Rosa. Gracias por apoyarme, y por haber creído en mí desde el principio. Cuando el curso se ha complicado os ha tocado observar con paciencia. La Fé, que me habéis transmitido siempre, me ha dado la esperanza y el consuelo de que no se sufre sin sentido y que esta historia también ha estado bien hecha. 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, porque habéis sido un apoyo imprescindible y siempre habéis creído en mí.

Gracias a mis compañeros y amigos que me han acompañado estos años. Iván, gracias de corazón, eres el mejor compañero de batalla que he podido tener y ha sido un disfrute vivir este último año a tu lado. Juan Antonio y María José, muchísimas gracias. Desempeñáis vuestra vocación con pasión y cariño, y lo habéis demostrado con la dedicación, paciencia y atención que me habéis mostrado durante este tiempo. Os he vivido como compañero y como alumno y sois una bendición. También quiero agradecer con un cariño especial a Irene Fondón. Gracias porque literalmente has contado conmigo para todo, desde el primer día, con una confianza en mi desde el principio que me ha impulsado muchísimo. Te mereces lo mejor. Gracias también a todo el cuerpo docente que una y otra vez os preocupáis por cómo perfeccionar el máster y cómo hacernos sentir realizados y motivados con el estudio.

Las palabras más especiales van para ti, Gloria. Han sido dos años duros donde menos mal que estabas tu. Como me sueles decir, tenemos nuestras cosas, pero somos el mejor equipo. Hoy toca alegrarnos por fin, y me siento muy feliz de que estés conmigo para cerrar juntos esta etapa.

*Elías Marqués Valderrama  
Sevilla, 2023*





# Resumen

---

Este trabajo introduce ciertos aspectos que son necesarios para trabajar con el modelado y la linealización de amplificadores de potencia en entornos distribuidos. La problemática de la distorsión en amplificadores de potencia se vuelve más compleja conforme aumentan características como el ancho de banda y que cada vez adquieren mayor demanda en los sistemas de comunicaciones móviles 5G. Además, surge especial interés en técnicas que permitan la computación distribuida, que no solo permite operar sobre dispositivos más simples, sino que también habilita arquitecturas en paralelo, acelerando este tipo de cómputos. Este trabajo trata de combinar técnicas como *genetic-based Volterra subspace generator* (GVG) para la búsqueda del modelo y el método de dirección alterna de multiplicadores (ADMM) para modelar y linealizar amplificadores de potencia, mediante una solución distribuida. Además, se profundiza en el estudio de la validación, mediante la comprensión del impacto de la relación potencia de pico a potencia promedio (PAPR) en métricas como la divergencia de Kullback-Leibler. Éstas, se pusieron a prueba con señales 5G-NR con un ancho de banda de 50 MHz, demostrando que ADMM es capaz de alcanzar prestaciones competitivas al comparar con la resolución por mínimos cuadrados.



# Abstract

---

This work introduces certain aspects that are necessary to work with the modelling and linearisation of power amplifiers in distributed environments. The problem of distortion in power amplifiers becomes more complex as features such as bandwidth grow and are increasingly in demand in 5G mobile communications systems. In addition, there is a special interest in techniques that enable distributed computing, which not only allows operating on simpler devices, but also enables parallel architectures, accelerating this type of computation. The combination of techniques such as genetic-based Volterra subspace generator (GVG) for model search and the alternating direction method of multipliers (ADMM) for modelling and linearising power amplifiers, by means of a distributed solution, is proposed. In addition, the study of model validation is further explored by understanding the impact of the peak-to-average power ratio (PAPR) on metrics such as Kullback-Leibler divergence. These were tested on 5G-NR signals with a bandwidth of 50 MHz, demonstrating that ADMM is able to achieve competitive performance when compared to least squares resolution.



# Índice Abreviado

---

<i>Resumen</i>	III
<i>Abstract</i>	V
<i>Índice Abreviado</i>	VII
<i>Notación</i>	XI
<b>1 Introducción</b>	<b>1</b>
<b>2 Modelado no lineal con series de Volterra</b>	<b>3</b>
2.1 Series de Volterra	3
2.2 Modelado estructurado	4
2.3 Modelado no estructurado	5
2.4 Expresión matricial	7
2.5 Planteamiento fundamental de la regresión	8
2.6 Capacidad de generalización de la regresión	9
<b>3 Resolución del problema mediante el método de dirección alterna de multiplicadores (ADMM)</b>	<b>15</b>
3.1 Ascenso dual: función Lagrangiana	15
3.2 Método de Lagrange aumentado	16
3.3 Método de dirección alterna de multiplicadores (ADMM)	17
3.4 Variaciones de la función a minimizar	18
3.5 Implementación de ADMM centralizado para el problema propuesto: Modelado	18
3.6 Implementación de ADMM centralizado para el problema propuesto: Linealización	22
3.7 Complejidad computacional	24
3.8 Consenso de los distintos subproblemas	25
<b>4 Diseño experimental y resultados</b>	<b>27</b>
4.1 Banco de medidas	27
4.2 Figuras de mérito	28
4.3 Implementación de ADMM	29
4.4 Modelado	30
4.5 Linealización	32
4.6 Ventajas de la reducción del coste computacional	39
<b>5 Conclusiones y líneas futuras</b>	<b>41</b>
<b>Apéndice A Código de MATLAB® empleado en el trabajo</b>	<b>43</b>
<b>Apéndice B Artículos de investigación</b>	<b>45</b>
<i>Índice de Figuras</i>	59
<i>Índice de Tablas</i>	61

*Bibliografía*

63

*Glosario*

67

# Índice

---

<i>Resumen</i>	III
<i>Abstract</i>	V
<i>Índice Abreviado</i>	VII
<i>Notación</i>	XI
<b>1 Introducción</b>	<b>1</b>
<b>2 Modelado no lineal con series de Volterra</b>	<b>3</b>
2.1 Series de Volterra	3
2.2 Modelado estructurado	4
2.3 Modelado no estructurado	5
2.4 Expresión matricial	7
2.5 Planteamiento fundamental de la regresión	8
2.6 Capacidad de generalización de la regresión	9
2.6.1 Divergencia de KL	10
2.6.2 Sensibilidad respecto a la señal de entrada	11
<b>3 Resolución del problema mediante el método de dirección alterna de multiplicadores (ADMM)</b>	<b>15</b>
3.1 Ascenso dual: función Lagrangiana	15
3.2 Método de Lagrange aumentado	16
3.3 Método de dirección alterna de multiplicadores (ADMM)	17
3.4 Variaciones de la función a minimizar	18
3.5 Implementación de ADMM centralizado para el problema propuesto: Modelado	18
3.6 Implementación de ADMM centralizado para el problema propuesto: Linealización	22
3.7 Complejidad computacional	24
3.8 Consenso de los distintos subproblemas	25
<b>4 Diseño experimental y resultados</b>	<b>27</b>
4.1 Banco de medidas	27
4.2 Figuras de mérito	28
4.3 Implementación de ADMM	29
4.4 Modelado	30
4.4.1 Ponderación de modelos	32
4.5 Linealización	32
4.5.1 Implementación directa de ADMM en DLA	33
4.5.2 Reajustes para alcanzar la linealización	35
4.6 Ventajas de la reducción del coste computacional	39
<b>5 Conclusiones y líneas futuras</b>	<b>41</b>
<b>Apéndice A Código de MATLAB® empleado en el trabajo</b>	<b>43</b>

<b>Apéndice B Artículos de investigación</b>	<b>45</b>
<i>Índice de Figuras</i>	59
<i>Índice de Tablas</i>	61
<i>Bibliografía</i>	63
<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 $\mathbf{A}$
$ x $	Módulo de $x$
$\mathbf{A}^{-1}$	Inversa de la matriz $\mathbf{A}$
$\mathbf{A}^\dagger$	Matriz pseudoinversa de la matriz $\mathbf{A}$
$\mathbf{A}^T$	Transpuesto de $\mathbf{A}$
$\mathbf{A}^H$	Transpuesto y conjugado de $\mathbf{A}$
$\text{cond}(\mathbf{A})$	Número de condición de la matriz $\mathbf{A}$
$x^*$	Conjugado de $x$
$\text{Re}\{\cdot\}$	Parte real
$\text{Im}\{\cdot\}$	Parte imaginaria
$\nabla f(x)$	Gradiente de $f(x)$
$\mathcal{O}$	Número de operaciones
$\mathbf{x} \rightarrow \mathbf{y}$	Asignación
$\frac{\delta \mathbf{y}}{\delta \mathbf{x}}$	Derivada de $\mathbf{y}$ respecto de $\mathbf{x}$
$\mathbf{x}_*$	$\mathbf{x}$ óptima
$\bar{X} = \{x[1], \dots, x[n]\}$	Secuencia de $n$ datos
$\mathbf{I}$	Matriz identidad
$\mathbf{0}$	Matriz de ceros
$\bar{\mathbf{u}}$	Media de los vectores $\mathbf{u}^1 \dots \mathbf{u}^M$
$\ \mathbf{x}\ _1$	Norma-1 del vector $\mathbf{x}$
$\mathbf{x}_i, \quad i = 1, 2, \dots, n$	Trozo $i$ -ésimo del vector $\mathbf{x}$
$\mathbf{x}^i, \quad i = 1, 2, \dots, n$	Réplica $i$ -ésima del vector $\mathbf{x}$
$\mathbf{x}^k$	$\mathbf{x}$ en la $k$ -ésima iteración
$\leq$	Menor o igual
$\geq$	Mayor o igual
$\gg$	Mucho mayor que
$\ll$	Mucho menor que
$\approx$	Aproximadamente igual
$t.q.$	Tal que
$s.t.$	Sujeto a
$\underset{\mathbf{x}}{\text{argmax}} f(\mathbf{x})$	$\mathbf{x}$ que hace máxima la función $f(\mathbf{x})$
$\underset{\mathbf{x}}{\text{argmin}} f(\mathbf{x})$	$\mathbf{x}$ que hace mínima la función $f(\mathbf{x})$
$\max_{\mathbf{x}} f(\mathbf{x})$	Máximo de $f(\mathbf{x})$ en la dirección de $\mathbf{x}$
$\min_{\mathbf{x}} f(\mathbf{x})$	Mínimo de $f(\mathbf{x})$ en la dirección de $\mathbf{x}$
$\text{sign}(a)$	Signo de $a$

---

$\mathbf{S}_\kappa(a)$	Función de umbral $\kappa$ suave de $a$
$\mathbb{E}[x]$	Esperanza matemática de $x$
$\mathbf{R}_x$	Matriz de correlación de $x$
$\hat{x}$	Valor estimado de $x$
$Col(\mathbf{A})$	Espacio columna de la matriz $\mathbf{A}$
$\forall$	Para todo
$i \in S$	$i$ pertenece a $S$
$i \notin S$	$i$ no pertenece a $S$
$(0,1]$	Conjunto entre 0 y 1 excluyendo el 0
$\{\cdot\}$	Conjunto
$x \sim \mathcal{N}(m, \sigma^2, \gamma)$	$x$ sigue una distribución normal de media $m$ , covarianza $\sigma^2 = \mathbb{E}[x^*x]$ y pseudocovarianza $\gamma = \mathbb{E}[xx]$

# 1 Introducción

---

*No es el conocimiento, sino el acto de aprendizaje; y no la posesión, sino el acto de llegar a ella, lo que concede el mayor disfrute*

CARL FRIEDRICH GAUSS

En las últimas décadas, las comunicaciones móviles se están convirtiendo en un pilar fundamental para el desarrollo de los servicios de la información, con un afianzamiento sólido en la sociedad que cada vez más demanda conectividad en escenarios de movilidad. Si se observa la evolución de las distintas generaciones móviles, no cabe duda de que aumentar las prestaciones en cuanto a velocidad de datos ha sido el gran desafío en las cuatro primeras generaciones móviles. Esto ha facilitado que se comience a pensar en nuevos servicios, cada vez más imaginativos e innovadores, que no solo son posibles con el aumento del ancho de banda. Es por esto que, en la Quinta Generación móvil (5G) se establecen la banda ancha móvil mejorada (eMBB), las comunicaciones ultra fiables de baja latencia (URLLC) y las comunicaciones tipo máquina masivas (mMTC) como los tres casos de uso sobre los que se basa esta nueva generación.

En este contexto, 5G coloca al esquema de multiplexación por división de frecuencias ortogonales (OFDM) como pilar fundamental para alcanzar sus objetivos, siendo de especial interés para el caso de uso eMBB, permitiéndose anchos de banda muy elevados. Este tipo de señales tienen la desventaja de que presentan una relación potencia de pico a potencia promedio (PAPR) demasiado elevada, haciendo que cuando se coloca el punto de operación del amplificador de potencia (PA) cercano a la zona de saturación, donde es más eficiente, proliferen la distorsión, que degrada la señal e imposibilita las comunicaciones en las bandas adyacentes. Esto ha provocado que en los últimos años surja un interés en la comunidad científica por atajar este problema, donde toman especial relevancia los trabajos de predistorsión digital (DPD), que se constituye como una de las técnicas por excelencia.

Hasta ahora se ha llevado a cabo un gran esfuerzo por desplegar la 5G *New Radio* (5G-NR), quedando en segundo plano la implantación de las mejoras en el *core* de la red. Si bien está comenzando el despliegue de soluciones, aún quedan unos años para poder comenzar a disfrutar de forma real de los casos de uso URLLC y mMTC. Esta arquitectura tiene como soporte el concepto del *Edge Computing*, que trata de funcionar como apoyo para que los dispositivos puedan hacer uso de modelos complejos, que se resuelven rápidamente en el borde de la red. Es por ello, que parece conveniente adaptar la resolución de los problemas de predistorsión desde un enfoque centralizado a una arquitectura distribuida, a fin de poder aprovechar las estructuras de cómputo que, a día de hoy, ya están diseñándose. Esta intención confluye con las complicaciones que presentan los escenarios complejos, donde la resolución de las ecuaciones de medida suelen conllevar problemas numéricos, causando inestabilidad en la regresión y devolviendo soluciones muy sensibles a variaciones en la entrada o que incluso no son capaces de converger.

Así pues, este trabajo pretende abordar distintos aspectos sobre el procesado distribuido, presentando el método de dirección alterna de multiplicadores (ADMM) como alternativa a la solución de mínimos cuadrados (LS), a fin de evaluar la viabilidad de este tipo de soluciones y las prestaciones que ofrece. Para ello, este documento comenzará con una introducción teórica sobre el modelado y la linealización de este tipo de dispositivos en el Capítulo 2, además de explorar distintos aspectos del set de entrenamiento que pueden ser determinantes a la hora de modelar los amplificadores de potencia. El Capítulo 3 se adentrará en el procesado distribuido, comentando el balance entre las implicaciones que este tiene sobre las prestaciones

de las soluciones alcanzadas y la mejora computacional. Finalmente, en el Capítulo 4 se muestran tanto el laboratorio como las pruebas experimentales que ponen a prueba los contenidos desarrollados, cerrando el trabajo en el Capítulo 5, que contendrá las conclusiones y las líneas futuras.

## 2 Modelado no lineal con series de Volterra

---

*La esencia de la verdad es la verdad de la esencia*

MARTÍN HEIDEGGER, 1952

El primer paso a la hora de trabajar con amplificadores de potencia es el modelado, ya que será empleado posteriormente para poder llevar a cabo la tarea de linealización. Desde el punto de vista de la ingeniería, un modelo o ley de funcionamiento suele tener la forma de expresión matemática. La formulación para la obtención de la ley en cuestión se conoce como regresión, que consiste en encontrar la relación entre variables no deterministas como una función matemática. El tipo de regresión puede ser lineal, logarítmica, exponencial, etc, y se elige una relación u otra en base al conocimiento previo del problema y a la capacidad de ajuste y generalización de la solución alcanzada. Si bien se suele prestar mucha atención al tipo de regresión, la estructura del modelo es fundamental. La estructura es justamente la cualidad descriptiva del proceso o elemento que se pretende modelar y define lo que se conoce como el espacio de características. En otras palabras, un modelo establece la relación entre las distintas características que se esperan expliquen el comportamiento que se desea modelar.

### 2.1 Series de Volterra

En el mundo de los amplificadores de potencia (PA), el paradigma de modelado por excelencia es el basado en series de Volterra. Esto es debido a que el PA constituye un sistema no lineal con memoria. Puede entenderse como una extensión de un sistema no lineal al caso con memoria, que se puede obtener de las series de Taylor siguiendo el modelo de Wiener como

$$y_{NL}(t) \approx a_1 x(t - \tau) + a_2 x^2(t - \tau) + \dots + a_p x^p(t - \tau), \quad (2.1)$$

donde el orden máximo es  $P$  y presenta diferentes respuestas en función del retraso  $\tau$ . De manera que, según [1], se define formalmente la serie de Volterra como

$$y(t) = \sum_{p=1}^{\infty} y_p(t), \quad (2.2)$$

siendo cada  $y_p(t)$  un funcional, que toma la siguiente forma

$$y_p(t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} h_p(\tau_1, \tau_2, \dots, \tau_p) x(t - \tau_1) x(t - \tau_2) \dots x(t - \tau_p) d\tau_1 d\tau_2 \dots d\tau_p. \quad (2.3)$$

Cada combinación de productos de  $x(t)$  con sus respectivos retrasos se denomina regresor, quedando contenido el *kernel* de la serie en los distintos coeficientes  $h_p(\tau_1, \tau_2, \dots, \tau_p)$ . Aunque conceptualmente la serie de Volterra ya queda bien definida, es necesario truncar la serie y discretizarla, para que sea computable y poder adaptar el modelo directamente a la predistorsión digital (DPD), del inglés *digital predistortion*. Por

tanto, la serie quedaría como

$$y[n] = \sum_{p=1}^{\infty} y_p[n] \approx \sum_{p=1}^P \sum_{q_1=0}^{Q_1} \dots \sum_{q_p=0}^{Q_p} h_p[q_1, \dots, q_p] \prod_{j=1}^p x[n - q_j], \quad (2.4)$$

donde  $Q$  representa la máxima profundidad de memoria y  $P$  el máximo orden.

En relación a lo comentado anteriormente sobre la regresión, la ecuación (2.4) representaría un modelo cuya entrada es la señal  $x[n]$  y la salida es  $y[n]$ . Las características a relacionar con los distintos coeficientes  $h_p[q_1, \dots, q_p]$  serían los regresores. Y aunque este modelo es perfectamente válido, es conveniente realizar un ajuste más para empezar a trabajar con él. Hasta ahora,  $x[n]$  e  $y[n]$  representarían muestras de señales reales. Por ello, la ecuación (2.4) se denomina serie de Volterra en RF (radiofrecuencia). En telecomunicaciones es preferible trabajar con muestras de la envolvente compleja de la señal en banda base, por lo que se reescribe (2.4) según [2] como:

$$y[n] = \sum_{p=0}^P \sum_{q_1=0}^{Q_1} \sum_{q_2=0}^{Q_2} \dots \sum_{q_{2p+1}=0}^{Q_{2p+1}} h_{2p+1}[q_1, q_2, \dots, q_{2p+1}] \prod_{r=1}^{p+1} x[n - q_r] \prod_{s=p+2}^{2p+1} x^*[n - q_s], \quad (2.5)$$

teniendo así un orden no lineal máximo  $2P + 1$  y una profundidad de memoria para cada orden de  $Q_{2p+1}$ .

Esta formulación se basa en la suposición de banda estrecha [3], que será válida siempre que la frecuencia de portadora  $f_c$  sea mucho mayor que el ancho de banda de la señal modulada  $BW$ , esto es,  $f_c \gg BW$ , tomando interés los regresores de la forma  $x[n - q]|x[n - q]|^{p-1}$  para  $p$  impar. Aunque el desarrollo teórico por el que se expresa la serie de Volterra en términos de su equivalente en banda base limita (2.5) al uso de estos regresores con  $p$  impar, en [4]-[5] se demuestra mediante un enfoque de serie de Volterra bivariable que es posible la aparición de términos con  $p$  par, lo que justifica que el uso de estos regresores es beneficioso. Es por esto que los siguientes modelos presentados tendrán incorporados estos términos. La ecuación (2.5) será referida como FV (*full Volterra*) y no debe perderse de vista que, a partir de ahora, la entrada  $x[n]$  y la salida  $y[n]$  serán en todo momento las muestras de las envolventes complejas de las respectivas señales en banda base.

Por otro lado, en ocasiones es recomendable incluir regresores dependientes de la señal imagen  $x^*[n]$  para tener en cuenta las posibles imperfecciones I/Q de la señal modulada. Estas se manifiestan por la existencia de correlación entre  $x[n]$  y  $x^*[n]$  [6], que se genera a lo largo de la cadena RF y que obliga a trabajar con señales impropias. Estos regresores están basados en procesamiento de señal ampliamente lineal [7] y en la propuesta del modelo de Volterra general para señales complejas de [2]. Esta serie de Volterra compleja (CVS) viene dada por la siguiente expresión:

$$\begin{aligned} y[k] = & h_{0,0} + \sum_{p=1}^P \left\{ \sum_{\mathbf{q}_p=0}^{\mathbf{Q}_{p,0}} h_{p,0}[\mathbf{q}_p] \prod_{r=1}^p x[n - q_r] + \right. \\ & \sum_{\mathbf{q}_p=0}^{\mathbf{Q}_{0,p}} h_{0,p}[\mathbf{q}_p] \prod_{s=1}^p x^*[n - q_s] + \\ & + \sum_{m=1}^{p-1} \sum_{\mathbf{q}_p=0}^{\mathbf{Q}_{p-m,m}} \sum_{\mathbf{p}_m=0}^{\mathbf{P}_{p-m,m}} h_{p-m,m}[\mathbf{q}_{p-m}, \mathbf{p}_m] \\ & \left. \prod_{r=1}^{p-m} x[n - q_r] \prod_{s=1}^m x^*[n - p_s] \right\}, \quad (2.6) \end{aligned}$$

donde puede verse como la salida depende tanto de la señal  $x[n]$  como de su imagen  $x^*[n]$ .

## 2.2 Modelado estructurado

En este trabajo, realizar un modelado estructurado va a consistir en partir del modelo FV o CVS e imponer cotas para obtener un subespacio de regresores más manejable y reducido. Dentro de las infinitas elecciones que pueden llevarse a cabo sobre el modelo FV existen aproximaciones reducidas que han sido ampliamente

empleadas en la literatura, debido principalmente a que los casos de uso en los que se han planteado, han dado unas prestaciones razonables en comparación con la complejidad de la estructura.

El modelo más sencillo conceptualmente hablando que recoge regresores no lineales con memoria es el *Memory Polynomial* (MP), adquiriendo la siguiente expresión en banda base

$$y_{MP}[n] = \sum_{p=1}^P \sum_{q=0}^Q h_p[q] x[n-q] |x[n-q]|^{p-1}, \quad (2.7)$$

presentada en [8] y equivalente a la de Kim y Konstantinou [9]. Aunque el modelo resulta bastante simple, alcanza unas prestaciones muy bajas siendo únicamente factible para amplificadores muy *sencillos*, es decir, con no linealidades débiles y pocos efectos en memoria.

Esto es lo que motiva el hecho de que Morgan en [1], proponga una extensión más completa donde se tengan en cuenta términos de memoria cruzada. Este modelo se le conoce como *Generalized Memory Polynomial* (GMP) y se define como

$$\begin{aligned} y_{GMP}[n] = & \sum_{k=0}^{K_a} \sum_{l=0}^{L_a} a_{kl} x[n-l] |x[n-l]|^k + \\ & + \sum_{k=1}^{K_b} \sum_{l=0}^{L_b} \sum_{m=1}^{M_b} b_{klm} x[n-l] |x[n-l-m]|^k + \\ & + \sum_{k=1}^{K_c} \sum_{l=0}^{L_c} \sum_{m=1}^{M_c} c_{klm} x[n-l] |x[n-l+m]|^k. \end{aligned} \quad (2.8)$$

La forma de trabajar con los límites de los sumatorios en este modelo es un tanto peculiar, y en ocasiones se emplean letras de conjuntos como en [10] para dejar constancia de que no se trata de un límite fijo. En concreto,  $\{L_a\} = f(K_a)$ ,  $\{L_b, M_b\} = f(K_b)$  y  $\{L_c, M_c\} = f(K_c)$ . Esta libertad a la hora de elegir los límites en función del orden que se trabaje permite tener en cuenta efectos como el *memory fading*, que presenta una dependencia con la memoria decreciente en ordenes superiores. Se recuerda que, en este modelo, se permiten regresores que incorporen productos  $|x|^k$  con  $k$  impar, que en principio están fuera del espacio FV univariable, pero quedan justificados desde el punto de vista de un modelo bivariable y, además, se ha probado que son beneficiosos [4]-[5].

El modelo GMP es ampliamente utilizado en problemas de predistorsión, debido a que, si se tiene algún conocimiento previo del comportamiento físico del PA, se puede seguir un razonamiento *ad hoc* a la hora de proponer una estructura inicial de regresores. En este tipo de enfoques, cuando el escenario se complica, se tiende a extender los límites de la serie, esperando poder incorporar así regresores que aporten más riqueza al modelo, buscando mejorar el ajuste. El criterio en cuanto a la profundidad de los límites del modelo suele decidirse en base a lo observado en la característica AM/AM del PA, ejemplificada en la Figura 2.1. Por tanto, si se observa una compresión muy abrupta, se espera que el modelo contenga regresores de alto orden, mientras que si la nube de puntos presenta una variación alta alrededor de la media, se entiende que el modelo debería tener mucha profundidad de memoria.

Aunque los modelos estructurados tienen la ventaja del razonamiento *ad hoc* anteriormente comentado, presentan el problema de que dejan de ser adecuados en escenarios más complicados. El aumento del ancho de banda es una de las características de la señal que puede añadir dificultad a los modelos. De manera que, puede darse el escenario en el que el modelo GMP deje de dar buenas prestaciones aún cuando se le permite una alta dimensión de representación [11]. Este fenómeno se ejemplifica en la Figura 2.2 donde trabajando con una señal 5G-NR de 50 MHz, por más que se aumente el espacio de regresores, no se consigue aportar precisión al modelo.

Esto sugiere que emplear regresores únicamente de la forma  $x[n-l] |x[n-l \mp m]|^k$  no es suficiente en el caso de ejemplo, siendo necesario el empleo de otro tipo de funciones base que no forman parte del subespacio GMP.

## 2.3 Modelado no estructurado

El estudio presentado en [11], además de ejemplificar un caso de saturación en prestaciones del modelo GMP, deja de manifiesto una desventaja en la metodología que hay detrás de los modelos estructurados.

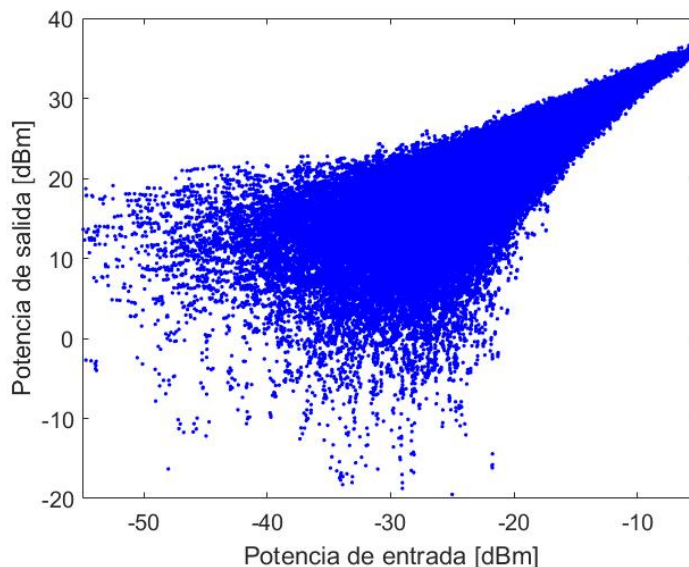


Figura 2.1 Ejemplo de característica AM/AM de un amplificador clase J.

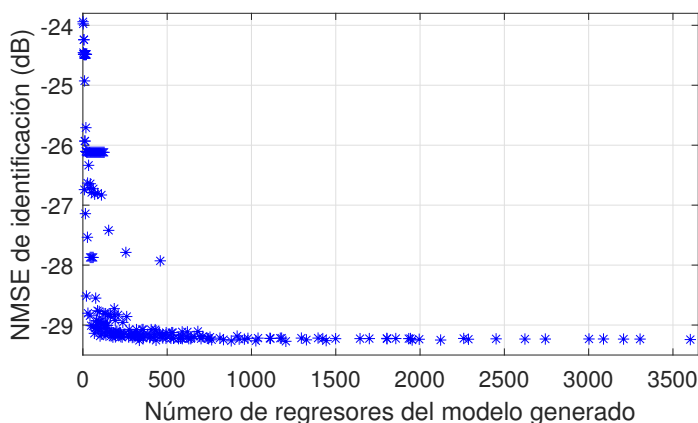


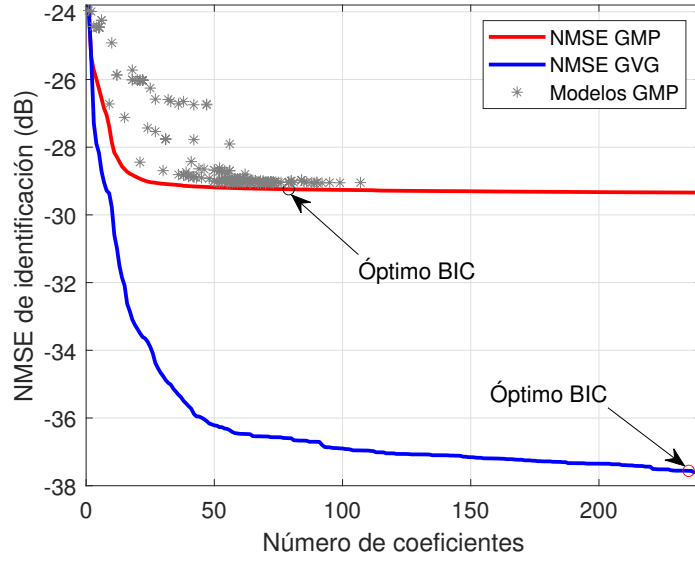
Figura 2.2 Saturación del NMSE de identificación tras realizar selección de coeficientes en un conjunto de 231 modelos GMP con diferentes parámetros, con un máximo de 3063 regresores [11].

Concretamente, cuando el escenario es muy complejo, se da por hecho que se va a tener que hacer uso de una técnica de podado de coeficientes, donde se van a eliminar gran parte de los regresores incluidos, manteniendo la mejor precisión posible. Esto unido a la saturación de prestaciones con respecto al número de coeficientes para escenarios complejos hace que en ocasiones sea necesario plantear una metodología distinta.

En esta tesitura, se propone un enfoque a la inversa, el cual recibirá el nombre de modelado no estructurado. La idea no consiste ahora en imponer límites al modelo FV, sino justo al contrario, explorar el espacio FV de forma inteligente sin un límite fijo más allá de la cantidad de regresores con los que se desea trabajar. Y dado que el escenario de este trabajo tiene el mismo contexto que el presentado en [11], será el modelo escogido. Esta filosofía va en la línea de los algoritmos genéticos, donde la referencia en este trabajo será el *genetic-based Volterra subspace generator* (GVG), presentado en [12] e implementado en [13], a fin de generar el subespacio de Volterra con aquellos regresores que mejor se ajusten al comportamiento particular del PA. Con esta adaptación hacia los algoritmos genéticos, cada regresor constituye un cromosoma que queda definido por la combinación de los constituyentes  $x$ ,  $x^*$  y  $|x|$ . Así, cada cromosoma queda definido por un orden no lineal y una profundidad de memoria.

La incorporación de nuevos cromosomas se hace aplicando *doubly orthogonal matching pursuit* (DOMP) [14], asegurando que la información añadida sea la máxima posible por cada nuevo regresor. Aunque la generación





**Figura 2.3** Evolución del NMSE de identificación respecto a la cantidad de regresores incluidos en el modelo, ordenados en base a DOMP. La nube de puntos representa los NMSE de identificación alcanzados por distintos modelos GMP, cuando se aplica selección de coeficientes mediante DOMP [11].

de un modelo en base a esta técnica presenta un fuerte coste computacional, tiene la ventaja de que se pueden explorar regresores poco convencionales sin necesidad de computar a priori una matriz que englobe infinidad de regresores. En el caso de escenarios complejos, no hay suficiente conocimiento previo como para poder realizar un modelado estructurado guiado por el aumento de los límites de la serie, que planteado sobre el modelo FV o CVS, resulta inviable. La necesidad del paso a este tipo de modelos queda ejemplificada en la Figura 2.3, donde se ve cómo GVG supera con creces las máximas prestaciones ofrecidas por modelos GMP muy extensos.

Este trabajo se realizará bajo el mismo escenario, es decir, una señal 5G-NR con un ancho de banda de 50 MHz, empleando el mismo laboratorio. Este es el principal motivo por el que se respalda el uso del GVG como generador de modelo de Volterra desde su enfoque no estructurado.

## 2.4 Expresión matricial

En esta sección se da paso a una interpretación algebraica, que además de ayudar conceptualmente a comprender qué implica la regresión, hace de soporte para la computación en máquinas. Este desarrollo ya se llevó a cabo en [10], por lo que se recupera aquí brevemente. La clave consiste en que el sumatorio de la serie tiene forma de producto de una matriz por un vector, donde la componente  $[q_1, \dots, q_{2p+1}]$  del vector  $\mathbf{h}$  se multiplica por el correspondiente regresor, que debe ser columna de  $\mathbf{X}$ . En forma algebraica, compactar estas operaciones resulta en las siguientes definiciones de vector de coeficientes y matriz de regresores:

$$\mathbf{h} = [h_1[0], \dots, h_1[Q], h_2[0,0], h_2[0,1], \dots, h_2[Q, Q], \dots, h_p[0,0, \dots, 0], \dots, h_p[Q, Q, \dots, Q]] \quad (2.9)$$

$$\mathbf{X} = \begin{bmatrix} x[0] & x[-1] & \dots & x[-Q] & x[0]|x[0]|^2 & \dots & x[-Q]|x[-Q]|^2 & \dots \\ x[1] & x[0] & \dots & \vdots & x[1]|x[1]|^2 & \dots & \vdots & \dots \\ x[2] & x[1] & \dots & x[-1] & x[2]|x[2]|^2 & \dots & x[-1]|x[-1]|^2 & \dots \\ x[3] & x[2] & \dots & x[0] & x[3]|x[3]|^2 & \dots & x[0]|x[0]|^2 & \dots \\ \vdots & \vdots & & \vdots & \vdots & & \vdots & \\ \vdots & \vdots & & \vdots & \vdots & & \vdots & \\ x[N-1] & x[N-2] & \dots & x[N-Q-1] & x[N-1]|x[N-1]|^2 & \dots & x[N-Q-1]|x[N-Q-1]|^2 & \dots \end{bmatrix}, \quad (2.10)$$

donde el índice negativo indica que se trata de instantes anteriores a la muestra que se ha tomado como referencia de inicio. Por tanto, se nota  $N$  como cantidad de muestras,  $M$  como el número de regresores,  $\mathbf{X} = [\phi_1, \phi_2, \dots, \phi_M] \in \mathbb{C}^{N \times M}$  como la matriz de regresores,  $\mathbf{h} \in \mathbb{C}^{1 \times M}$  como el vector de coeficientes y las muestras de la envolvente compleja de la salida en banda base como  $\mathbf{y} \in \mathbb{C}^{N \times 1}$ . De manera que, cualquiera de las series representadas anteriormente quedaría sintetizada en la siguiente expresión:

$$\mathbf{y} = \mathbf{X}\mathbf{h}, \quad (2.11)$$

en la que se indagará más adelante.

Durante el capítulo se han estado empleando los términos espacio y subespacio, referidos al resultado de las distintas series. Visto ahora el modelo en forma de matriz, resulta más sencillo relacionar estos conceptos. Concretamente, la ecuación (2.11) se entiende algebraicamente como una operación de proyección del vector  $\mathbf{y}$  sobre una base del subespacio  $\text{Col}(\mathbf{X})$ , siendo las coordenadas de dicha proyección las componentes del vector  $\mathbf{h}$ . Así pues, se entiende que el espacio completo de representación de la serie de Volterra se alcanza con el modelo CVS. Elegir un subconjunto dentro de un espacio es justamente definir un subespacio, y es lo que se hace cada vez que se elige uno de los modelos presentados anteriormente. Dentro del álgebra, el subespacio más intuitivo es el compuesto por la base cartesiana, que sería un conjunto de tres componentes dentro del espacio  $\mathbb{R}^n$ . Esto suele provocar que, por inercia, se piense que el subespacio de un modelo de Volterra tiene  $\text{Col}(\mathbf{X})$  como una base ortogonal, al estilo de la cartesiana. Nada más lejos de la realidad para el caso de las series de Volterra, donde la matriz de regresores presenta una fuerte correlación en sus columnas. Esto es lo que motiva el empleo de distintas técnicas de selección de coeficientes como *compressive sampling matching pursuit* (CoSaMP) [15], *subspace pursuit* (SP) [16] y *DOMP* [14] comparadas en [17]. Además, al ser algoritmos con un coste computacional alto, se buscan implementaciones más rápidas como *DOMP* en su versión de complejidad reducida [18] y [19]. También se aborda el problema con enfoques bayesianos como en [20] o [21]. Todas estas técnicas buscan en definitiva que el subespacio  $\text{Col}(\mathbf{X})$  sea óptimo, facilitando la tarea de la regresión.

## 2.5 Planteamiento fundamental de la regresión

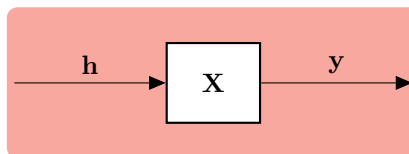
En la sección anterior se ha hecho especial hincapié en la idoneidad del modelo subyacente, teniendo como resultado la estructura del modelo. Esta estructura ya ha sido embebida en la matriz de regresores, pero queda por determinar la relación entre sus columnas que consigan un resultado lo más parecido posible al comportamiento real del PA. A continuación, se hará un estudio exhaustivo de la resolución de la regresión en este tipo de escenarios.

En la Sección 2.4 se ha llegado a la conclusión de que cualquier modelo escogido se puede representar en forma compacta según la expresión (2.11). Con esta reescritura de la serie de Volterra parece cambiar en cierta forma el orden lógico del sistema. Concretamente, el PA se entiende como sistema no lineal con memoria, al estilo de lo representado en la Figura 2.4. Lo natural es entender que la entrada es la envolvente compleja de la señal  $x[n]$  y su salida  $y[n]$ .

$$x[n] \rightarrow \boxed{\text{SNLM}} \rightarrow y[n] = \sum_{p=0}^P \sum_{q_1=0}^{Q_1} \sum_{q_2=0}^{Q_2} \cdots \sum_{q_{2p+1}=0}^{Q_{2p+1}} h_{2p+1}(q_1, q_2, \dots, q_{2p+1}) \prod_{r=1}^{p+1} x[n - q_r] \prod_{s=p+2}^{2p+1} x^*[n - q_s]$$

**Figura 2.4** Representación del PA. Las siglas SNLM simbolizan un sistema no lineal con memoria.

En cambio, con la adaptación del modelo de Volterra a la expresión matricial, resulta más sencillo trabajar según la Figura 2.5. Ahora, también puede interpretarse como entrada el vector de coeficientes. Así, aunque la estructura del modelo sigue siendo fija, el comportamiento del mismo estará influenciado por las características de  $\mathbf{x}$ . Este análisis se llevará a cabo más adelante.

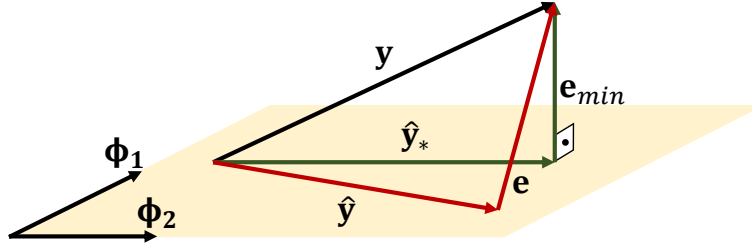


**Figura 2.5** Representación del PA según la ecuación (2.11).

Para ser rigurosos, es conveniente ajustar la notación de la ecuación (2.11):

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\mathbf{h}}, \quad (2.12)$$

donde  $\hat{\cdot}$  denota estimación. Por tanto, el objetivo de la regresión es encontrar  $\hat{\mathbf{h}}$  tal que el error  $\mathbf{e} = \mathbf{y} - \mathbf{X}\hat{\mathbf{h}}$ , sea el mínimo posible. El vector  $\mathbf{e}$  no solo recogerá las componentes de error dentro del subespacio  $\mathbb{C}^{N \times M}$  definido por el modelo de Volterra, sino que a su vez, incluye las proyecciones de las componentes de error que quedan fuera del modelo. El mínimo error se obtiene al proyectar  $\mathbf{y}$  en  $\text{Col}(\mathbf{X})$ , tal como se ve en la Figura 2.6.



**Figura 2.6** La estimación de mínimo error  $\hat{\mathbf{y}}_*$  se obtiene mediante una proyección ortogonal al espacio de representación.

Por tanto, el mínimo error se obtiene mediante la aplicación de mínimos cuadrados (LS), que resulta en la proyección ortogonal de la  $\mathbf{y}$  en el espacio del modelo, contenido en los regresores  $\phi_m$ . Para ello, se minimiza la suma residual de cuadrados (RSS):

$$\text{RSS} = \|\mathbf{y} - \hat{\mathbf{y}}\|_2^2 = (\mathbf{y} - \hat{\mathbf{y}})^H (\mathbf{y} - \hat{\mathbf{y}}) = \mathbf{y}^H \mathbf{y} - \hat{\mathbf{h}}^H \mathbf{X}^H \mathbf{y} - \mathbf{y}^H \mathbf{X} \hat{\mathbf{h}} + \hat{\mathbf{h}}^H \mathbf{X}^H \mathbf{X} \hat{\mathbf{h}}. \quad (2.13)$$

Minimizar la función RSS consiste en aplicar el procedimiento habitual de minimización de funciones, esto es, calcular la derivada e igualar a cero. En este caso esto es posible porque se trata de una función de variable compleja pero con resultado real. Por ende, si se tiene en cuenta el cálculo de Wirtinger [22], esto es, trabajar con  $\mathbf{h}$  y  $\mathbf{h}^*$  como variables independientes, y sabiendo que el gradiente tiene un mayor descenso en la variable conjugada, se resuelve:

$$\frac{d\text{RSS}}{d\mathbf{h}^*} = -2\mathbf{X}^H \mathbf{y} + 2\mathbf{X}^H \mathbf{X} \hat{\mathbf{h}} = 0, \quad (2.14)$$

despejando  $\hat{\mathbf{h}}$

$$\hat{\mathbf{h}} = (\mathbf{X}^H \mathbf{X})^{-1} \mathbf{X}^H \mathbf{y} = \mathbf{X}^\dagger \mathbf{y}. \quad (2.15)$$

Esta es la solución de LS y, como se ha comentado previamente, es la que consigue el mínimo error. Quedan por tanto de manifiesto dos características relacionadas con la regresión a tener en cuenta para poder obtener la solución. La primera de ellas es el cómputo de la matriz de autocorrelaciones muestrales  $\mathbf{R}_\mathbf{X} = \mathbf{X}^H \mathbf{X}$  cuya inversa es la matriz de precisión  $\mathbf{P}_\mathbf{X} = \mathbf{R}_\mathbf{X}^{-1}$ . El principal problema en el cómputo de esta matriz proviene de la fuerte correlación que presentan las columnas de  $\mathbf{X}$ , donde las técnicas presentadas en [23] ayudan a mitigar esos efectos, bajando el número de condición de  $\mathbf{R}_\mathbf{X}$ . La segunda característica a tener en cuenta es la propia función de minimización RSS. El método para encontrar el mínimo o la imposición de restricciones pueden ayudar a lidiar con la inestabilidad. Esta segunda característica será explorada en más detalle en el siguiente capítulo.

## 2.6 Capacidad de generalización de la regresión

El planteamiento fundamental presentado en la sección anterior realiza lo que se conoce como identificación del modelo. Concretamente, ya existe una estructura prefijada de regresores y la tarea de indentificar es justamente calcular el vector de coeficientes  $\hat{\mathbf{h}}$  óptimo que minimiza el error. Lo que se espera de un buen modelo es que tenga la capacidad de generalizar, es decir, que sepa adaptarse adecuadamente a distintas señales  $\mathbf{x}$ . Esto es fundamental, no solo por el hecho de que el amplificador puede cambiar su comportamiento en función de su entrada  $\mathbf{x}$ , sino que también hay que tener en cuenta que en la Figura 2.6 si cambia  $\mathbf{x}$

cambia también el espacio de representación. Por lo que a priori no se puede asegurar que los coeficientes  $\mathbf{h}$  calculados para una señal de entrada  $\mathbf{x}_1$  sean válidos para otra señal distinta  $\mathbf{x}_2$ . Es importante destacar que el sistema que se modela es un PA, que cambia físicamente de comportamiento cuando cambia el punto de operación, que viene dado por la potencia promedio y la PAPR de la señal de entrada. En este estudio se fija la potencia promedio, como se apuntará en la Sección 2.6.2, para fijar lo máximo posible el punto de operación y procurar que el sistema a modelar no cambie, ya que si es así el problema se complicaría demasiado.

### 2.6.1 Divergencia de KL

Una vez abierta esta discusión sobre la idoneidad de los modelos en función de la señal de entrada, resulta interesante ver de forma cuantitativa hasta qué punto el escenario puede ser cambiante en función de la señal de entrada  $x[n]$ . Para ello, puede ser útil emplear la divergencia de Kullback-Leibler (DKL) como métrica. Así, sea  $p_X^d(x)$  la función de distribución de probabilidad de una secuencia de datos  $\bar{X} = \{x[1], x[2], \dots, x[n]\}$  provenientes de una variable aleatoria  $X$  discreta, y sea  $q(x)$  un posible modelo de distribución que ha podido generar los datos, se puede tratar de codificar la secuencia de datos en base a ambas distribuciones como:

$$l_q(x_j) = \log_2 \frac{1}{q(x)}, \quad (2.16)$$

$$l_{p^*}(x_j) = \log_2 \frac{1}{p_X^d(x)}, \quad (2.17)$$

con  $j = 1, \dots, m$  siendo  $m$  la cantidad de elementos no repetidos en la secuencia. Se puede evaluar sus respectivas longitudes de código como:

$$L(C_q) = \sum_x p_X^d(x) \log_2 \frac{1}{q(x)} = H(p_X^d(x), q), \quad (2.18)$$

$$L(C_{p^*}) = \sum_x p_X^d(x) \log_2 \frac{1}{p_X^d(x)} = H(p_X^d(x)), \quad (2.19)$$

donde el código óptimo es justamente el que se obtiene de la estadística de los datos, y donde  $H(p_X^d(x), q)$  representa la entropía cruzada entre  $p_X^d(x)$  y  $q$ , y  $H(p_X^d(x))$  es la entropía de  $p_X^d(x)$ . Esto da pie a la definición de la divergencia de Kullback-Leibler:

$$D_{KL}(p_X^d||q) = L(C_q) - L(C_{p^*}) \geq 0, \quad (2.20)$$

que devuelve la ineficiencia de codificar la secuencia asumiendo que sigue el modelo  $q$ . Si el modelo  $q$  coincide con el de los datos  $p_X^d$ ,  $L(C_q) = L(C_{p^*})$ , quedando  $D_{KL}(p_X^d||q) = 0$ . Por último, a partir de  $D_{KL}$ , se puede plantear la probabilidad de que el modelo  $q$  genere una la secuencia  $\bar{x}$  con distribución  $p_X^d$  como:

$$P(\bar{x}_q \sim p_X^d) = 2^{-nD_{KL}(p_X^d||q)}. \quad (2.21)$$

**Demostración.** Sea la fuente de información  $\mathbf{x}$  una fuente discreta sin memoria, con alfabeto  $\mathcal{X} = \{x_1, \dots, x_N\}$ , donde cada símbolo tiene una probabilidad de ocurrencia  $\{p_1, \dots, p_N\}$ , el número de veces que ocurre el símbolo  $x_i$  es  $n_i$ . Si se observa una secuencia de longitud  $n \gg 1$  se puede aplicar la ley de los grandes números, teniendo una ocurrencia por símbolo aproximada de  $n_i \approx np_i$ . De forma que, se computa la probabilidad de ocurrencia de una secuencia típica en función de la entropía de  $X$  ( $H(X)$ ) como:

$$P(\bar{X} = \bar{x}) = \prod_{i=1}^N p_i^{n_i} \approx \prod_{i=1}^N p_i^{np_i} = \prod_{i=1}^N 2^{np_i \log_2 p_i} = 2^{\sum_{i=1}^N np_i \log_2 p_i} = 2^{-nH(X)}. \quad (2.22)$$

Por otro lado, el modelo  $q$  plantea la hipótesis de que la probabilidad de ocurrencia del símbolo  $x_i$  es  $\{q_1, \dots, q_N\}$ . Al estudiar una secuencia de  $n \gg 1$  símbolos, la ocurrencia del símbolo  $x_i$  seguirá siendo  $n_i \approx np_i$ . Así, al igual que con (2.22):

$$P(\bar{X}_q = \bar{x}) = \prod_{i=1}^N q_i^{n_i} \approx \prod_{i=1}^N q_i^{np_i} = \prod_{i=1}^N 2^{np_i \log_2 q_i} = 2^{\sum_{i=1}^N np_i \log_2 q_i} = 2^{-nH(X, X_q)}. \quad (2.23)$$

Siguiendo la notación anterior, proviniendo  $X$  del modelo generado por las muestras,  $p_X^d$ , y  $X_q$  del modelo  $q$ , se puede ver que la ecuación (2.22) devuelve la probabilidad de ocurrencia de la secuencia típica  $\bar{x}$  según el modelo generado por los datos  $p_X^d$ , mientras que (2.23) es la probabilidad de ocurrencia de la secuencia típica  $\bar{x}$  según el modelo  $q$ .

Con probabilidad prácticamente 1, las secuencias de longitud  $n$  que entregue la fuente  $\mathbf{x}$  serán secuencias típicas, si  $n \gg 1$ . De modo que, el número de secuencias típicas de longitud  $n$  de la fuente  $\mathbf{x}$  será  $N_{sec} = 2^{nH(X)}$ . Así, la probabilidad de que el modelo  $q$  genere una secuencia  $\mathbf{x}$  con distribución muestral  $p_X^d$ , será:

$$P(\bar{x}_q \sim p_X^d) = N_{sec} P(\bar{X}_q = \bar{x}) = 2^{nH(X)} 2^{-nH(X, X_q)} = 2^{-n(-H(X) + H(X, X_q))} = 2^{-nD_{KL}(p_X^d || q)}, \quad (2.24)$$

llegando así a la expresión (2.21).

En este punto conviene modificar ligeramente la ecuación (2.21) para conseguir una visualización adecuada. Esto es debido a que plantear la probabilidad de que el modelo  $q$  genere una secuencia de  $n$  muestras con distribución  $p_X^d$  es demasiado exigente, además de ser menos probable cuanto mayor es  $n$ . Realmente, la información en cuanto al parecido de cada distribución con  $p_X^d$  respecto del modelo  $q$  está contenida en  $D_{KL}(p_X^d || q)$ . Por tanto, la interpretación de la probabilidad es escalable en función de dónde se juzgue que está el umbral de  $D_{KL}(p_X^d || q)$  a partir del cual no se considera que exista parecido. Además, dado que  $2^x$  es una función monótona y creciente, no se altera el orden al escalar el exponente. Es por eso que se propone normalizar la probabilidad  $P(\bar{x}_q \sim p_X^d)$  por  $L/n$ :

$$P(\bar{x}_q \sim p_X^d)^{L/n} = 2^{-LD_{KL}(p_X^d || q)}, \quad (2.25)$$

pudiendo ajustar convenientemente la gráfica de probabilidad al dominio que se maneja. Véase cómo  $n$  es el número de muestras o longitud de secuencia, mientras que  $L$  representa un factor de escala a elegir.

### 2.6.2 Sensibilidad respecto a la señal de entrada

La señal modulada por excelencia en 5G es la OFDM. Desde que su planteamiento en [24] hasta su implementación típica a día de hoy [25], ha habido un fuerte desarrollo en la implementación de las operaciones *Inverse Fast Fourier Transform*/*Fast Fourier Transform* (IFFT/FFT), facilitando el soporte hardware. A pesar de ser una modulación con múltiples ventajas y ser la gran apuesta de la 5G-NR, tiene el problema de que presenta una potencia pico a potencia promedio PAPR muy elevada, que se define como

$$\text{PAPR}[\text{dB}] = 20 \log_{10} \left( \frac{\max |x[n]|}{\sqrt{\frac{1}{N} \sum_{n=1}^N |x[n]|^2}} \right). \quad (2.26)$$

Esto es debido a que en la etapa de la IFFT puede darse aleatoriamente una combinación de fases que dé lugar a una suma constructiva, haciendo que la amplitud de un instante determinado aumente mucho con respecto a la media de las señal.

A su vez, se define la *peak envelope power* (PEP) como

$$\text{PEP}[\text{dB}] = \text{PAPR}[\text{dB}] + P_{avg}[\text{dB}], \quad (2.27)$$

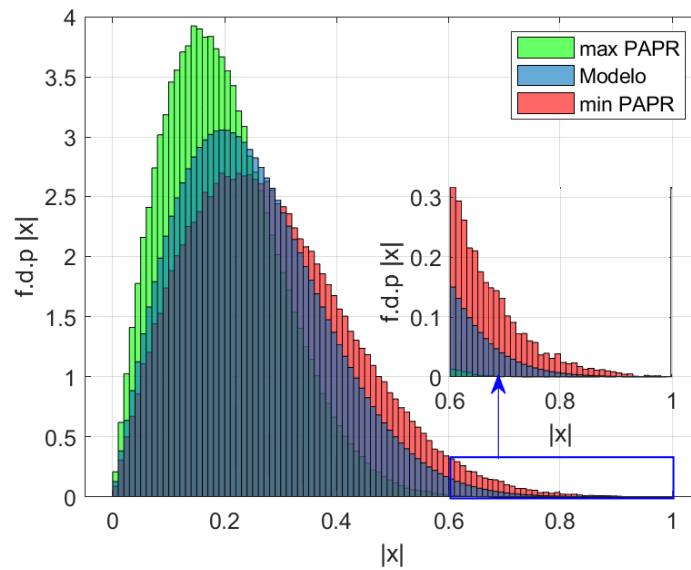
donde  $P_{avg}$  significa potencia promedio. Como generalmente se pretende situar la  $P_{avg}$  cercana al punto de saturación, puesto que el amplificador es más eficiente en esa zona, la PAPR suele ser el punto débil de la señal, que hace que el PA entre en saturación. Pero para el modelo la característica crítica debe ser la PEP. Esto es debido a que la naturaleza de las series de Volterra es polinómica, de manera que se espera que extrapolen mal los resultados en PEP que estén fuera del rango con el que se ha identificado. Por este motivo, tiene sentido plantear un estudio sobre cómo es la señal con la que tiene que trabajar el amplificador, para controlar hasta qué punto es esperable que el modelo se adapte bien a nuevas entradas, puesto que la serie de Volterra es sensible en este sentido.

Como se ha comentado en el párrafo anterior, gran parte de la distorsión ocasionada por el PA proviene de métricas que tienen que ver con la amplitud de la señal. Es por ello que, es plausible plantear la distribución

de probabilidad de la amplitud de la envolvente compleja de la señal de entrada  $\mathbf{x}$ . La metodología propuesta es la siguiente:

1. Extraer un modelo a partir de una población de señales mediante el método de Montecarlo. Éste será el modelo  $q$  mencionado en la Sección 2.6.1.
2. Realizar una comparativa entre la distribución  $p_X^d$  de cada señal OFDM de la población y el modelo  $q$ , computando la Divergencia de Kullback-Leibler.
3. Obtener una probabilidad de extrapolación del modelo mediante la expresión (2.25).

En primer lugar, se obtiene el modelo de Montecarlo, que consiste en realizar sucesivas realizaciones de la señal OFDM, cambiando la semilla de aleatorización. De esta forma se probaron 4000 secuencias de longitud  $n = 737820$  muestras, donde la señal OFDM mantuvo sus características. Es importante destacar que toda la teoría de la Sección 2.6.1 es relativa a fuentes discretas. Aunque el valor de la amplitud de la señal es continuo se ha discretizado a la hora de clasificar en los distintos *bins* para realizar el histograma. Es por eso que se considera que la teoría anteriormente presentada es extrapolable en este caso.

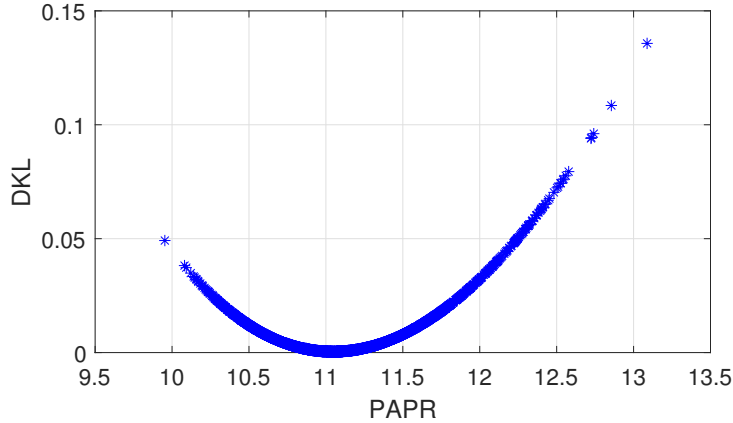


**Figura 2.7** Función de distribución de probabilidad (f.d.p) del modelo de Montecarlo junto con las señales de máxima y mínima PAPR. La f.d.p del modelo de Montecarlo presenta valores en todo el dominio de  $|\mathbf{x}|$ .

Además, en la figura se presenta una comparativa con los dos casos extremos, esto es, el mínimo y el máximo de la PAPR de cada señal de la población. Los detalles de la OFDM generada serán explicados en más profundidad en el Capítulo 4, ya que serán las mismas que se empleen durante la experimentación. Nótese cómo los valores de la Figura 2.7 están concentrados entre 0 y 1. Esto es debido a que al generar la señal se ha substraído su media y normalizado por el máximo módulo, esto es, la muestra de pico. Aunque la envolvente compleja generada esté normalizada para que su valor máximo sea 1, al aplicar la señal modulada al PA, se le fija una potencia promedio y se obtienen los valores de potencia deseados para el punto de operación.

Generalmente, en la OFDM,  $\mathbf{x} \sim \mathcal{N}(0, \sigma^2, 0)$ , es decir, las partes real e imaginaria se distribuyen como  $\text{Re}\{\mathbf{x}\} \sim \mathcal{N}(0, \sigma^2/2)$  e  $\text{Im}\{\mathbf{x}\} \sim \mathcal{N}(0, \sigma^2/2)$ , y son independientes e idénticamente distribuidas. Computar el módulo de  $\mathbf{x}$  consiste en obtener el marginal para el módulo de la distribución de  $\mathbf{x}$  cuando está representada en polares. En este caso, la señal es propia o circular, donde la correlación compleja es  $\rho = 0$ . Al calcular el marginal se genera justamente una función Rayleigh, como se ha obtenido en el resultado de la Figura 2.7 y como queda desarrollado en [22]. Se llama a la atención sobre el hecho de que la señal  $\mathbf{x}$  ciertamente es propia, ya que, como en cualquier sistema de comunicaciones, las partes real e imaginaria de la señal generada no están a priori correladas. El caso impropio, y con él la necesidad del empleo del modelo CVS, surge al tratar con el sistema completo, ya que los equipos de comunicaciones y el PA introducen las imperfecciones I/Q. Pero como se trata de un estudio previo al modelado se trabaja con la señal generada  $\mathbf{x}$ , por tanto con una

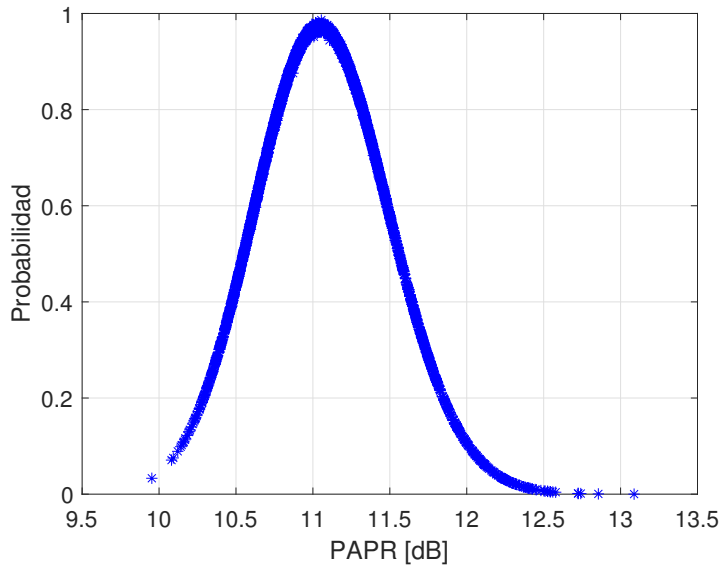
gaussiana circular, cuyo marginal para el módulo es la Rayleigh mencionada anteriormente que se observa en la Figura 2.7.



**Figura 2.8** Divergencia de Kullback-Leibler de la población de 4000 señales OFDM con el modelo de Montecarlo  $D_{KL}(p_{X_i}^d || q)$ .

Al generar la señal entre 0 y 1, el caso con mayor PAPR concentra sus valores en torno a una media más baja, mientras que la menor PAPR distribuye sus valores a lo largo de todo el rango de la función de distribución como se ve en la Figura 2.7. Esto ayuda a comparar en términos absolutos cómo se diferencia la distribución del módulo de la envolvente al cambiar la PAPR. Para dar un valor de cuánto difieren estas señales respecto del modelo de Montecarlo, se computa la divergencia de Kullback-Leibler, y como se ve que el principal indicador de discrepancia es la diferencia en la PAPR, se ordenan en base a esta métrica, obteniendo así la Figura 2.8. Téngase en cuenta que la PAPR del modelo de Montecarlo es  $\text{PAPR}_{\text{Montecarlo}} = 11.06$  dB.

Finalmente, se puede computar la probabilidad de generar una secuencia de  $n = 737820$  normalizada y escalada por  $L/n$ , siguiendo la ecuación (2.25), para facilitar así la visualización. Concretamente, se ha escogido  $L = 100$ , donde el resultado será interpretado a partir de ahora como una capacidad de extrapolación.



**Figura 2.9** Probabilidad de que  $q$  genere una secuencia  $p_{X_i}^d$  de longitud  $n = 737820$ , normalizada y escalada por  $L/n$  con  $L = 100$ . Este resultado se interpreta a partir de ahora como una probabilidad de extrapolación.

Véase en la Figura 2.9 cómo la métrica obtenida es bastante sensible al cambio de la PAPR, ya que, para una variación de 0.5 dB, la probabilidad cae a 0.5 desde la media. Esto quiere decir que si se obtiene un

modelo con una PAPR aproximada de 11 dB no se puede esperar que valide bien para una señal de 11.5 dB.

Por último, hay un matiz a tener en cuenta. Como se ha comentado anteriormente, los modelos de Volterra son sensibles a cambios en la PEP. Se vuelve a incidir en que se ha hecho el estudio frente a la PAPR debido a que cuando se vaya a trabajar con el PA, se va a mantener la  $P_{\text{avg}}$  fija.



# 3 Resolución del problema mediante el método de dirección alterna de multiplicadores (ADMM)

---

*Divide et impera.*

JULIO CÉSAR

Una vez planteada la estadística de la población a modelar es momento de centrarse en la función de minimización. Como se ha presentado en el capítulo anterior, la regresión por mínimos cuadrados constituye la solución analítica del problema de mínimos cuadrados de la función de variable compleja y salida real RSS. Pero también se ha visto que presenta ciertos problemas de estabilidad que han de ser atajados, bien mediante selección de coeficientes o bien planteando una forma alternativa de alcanzar el mínimo de la función RSS. Por otro lado, 5G ha empezado a promover el denominado *edge computing* (cómputo en el borde), para facilitar el cumplimiento de los requisitos que se plantean en el caso de comunicaciones fiables y de baja latencia, denominadas *ultra reliable and low latency communications* (URLLC). En este contexto comienzan a ser cuestionables los esquemas de aprendizaje automático con computación centralizada, ya que suponen un límite en cuanto a la latencia. El aumento de las exigencias provoca incluso que se hable de la *edge intelligence*, cuyo principal habilitador es la computación federada. Esto descansa en el planteamiento de sistemas distribuidos, más simples, que trabajando en conjunto consiguen realizar computaciones complejas [26]. Dado que tanto el modelado como la linealización son operaciones con un grado alto de complejidad, parece adecuado plantear el empleo de este tipo de técnicas para su resolución.

Debido a la variedad de implicaciones que hay detrás de la minimización distribuida, se hará una breve pasada por aquellos aspectos que facilitan la formulación de ADMM. Siguiendo el itinerario propuesto en [27], es adecuado pasar por ciertos aspectos previos a la formulación del problema de ADMM, para así entender que la solución que se alcanza mediante esta modificación es adecuada. Estos aspectos se tratarán con una función genérica  $f(\mathbf{x})$ , para adecuarla en secciones posteriores a la notación empleada anteriormente.

## 3.1 Ascenso dual: función Lagrangiana

Se plantea el siguiente problema de minimización con restricciones lineales, denominado problema primal:

$$\min_{\mathbf{x}} f(\mathbf{x}), \quad s.t. \quad \mathbf{Ax} = \mathbf{b}, \quad (3.1)$$

donde al menos  $f(\mathbf{x})$  es una función propia y cerrada, sin necesidad aún de asumir convexidad [28] y  $\mathbf{x} \in \mathbb{R}^{nx1}$ , y donde  $\mathbf{A} \in \mathbb{R}^{pxn}$  y  $\mathbf{b} \in \mathbb{R}^{px1}$  definen el conjunto de  $p$  restricciones. La función Lagrangiana simplemente modifica la  $f(\mathbf{x})$  para tener en cuenta las restricciones, incluyéndolas como una suma ponderada de penalizaciones:

$$L(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \boldsymbol{\lambda}^T (\mathbf{Ax} - \mathbf{b}), \quad (3.2)$$

donde  $\boldsymbol{\lambda} \in \mathbb{R}^{p \times 1}$  es el multiplicador de Lagrange cuyo tamaño dependerá del número de restricciones  $p$ . De esta forma, se define la función dual como

$$d(\boldsymbol{\lambda}) = \min_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}). \quad (3.3)$$

La función dual constituye un límite inferior de  $f(\mathbf{x})$ , de manera que en el óptimo se cumple que,  $d(\boldsymbol{\lambda}_*) \leq f(\mathbf{x}_*)$  [29]. La igualdad se podrá alcanzar en el caso en el que  $f(\mathbf{x})$  sea convexa, teniendo el caso de una dualidad fuerte, como se muestra en la Figura 3.1. Si no se cumple esta condición, se está ante un caso de dualidad débil. Desarrollando (3.3) se obtiene:

$$\begin{aligned} d(\boldsymbol{\lambda}) &= \min_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}) \\ &= \min_{\mathbf{x}} \left( f(\mathbf{x}) + \boldsymbol{\lambda}^T \mathbf{A}\mathbf{x} - \boldsymbol{\lambda}^T \mathbf{b} \right) \\ &= \min_{\mathbf{x}} \left( f(\mathbf{x}) + \boldsymbol{\lambda}^T \mathbf{A}\mathbf{x} \right) - \boldsymbol{\lambda}^T \mathbf{b} \\ &= - \max_{\mathbf{x}} \left( -f(\mathbf{x}) - \boldsymbol{\lambda}^T \mathbf{A}\mathbf{x} \right) - \boldsymbol{\lambda}^T \mathbf{b}. \end{aligned} \quad (3.4)$$

Así, sea  $\mathcal{D} = \{\boldsymbol{\lambda} \text{ t.q. } d(\boldsymbol{\lambda}) > -\infty\}$  el dominio de la función dual, se define el problema dual como:

$$\max_{\boldsymbol{\lambda} \in \mathcal{D}} d(\boldsymbol{\lambda}). \quad (3.5)$$

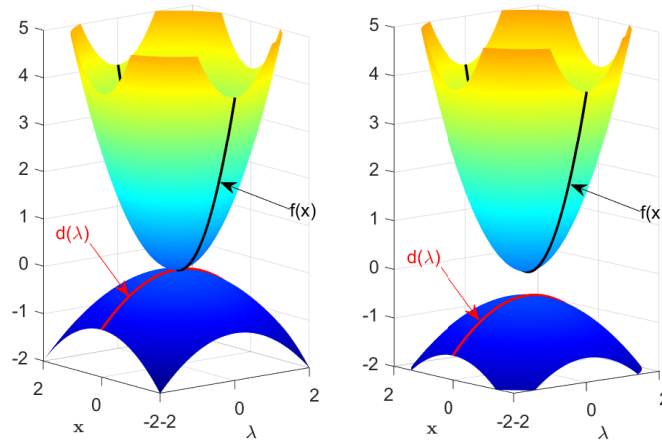


Figura 3.1 Ejemplo gráfico de dualidad fuerte (izquierda) y dualidad débil (derecha).

Las ecuaciones (3.4) y (3.5) permiten la interpretación ilustrada en la Figura 3.1. Véase que para alcanzar la solución se debe ir avanzando por partes. Esto sale a la luz al tratar de resolver el problema dual, cuyo gradiente será  $\nabla d(\boldsymbol{\lambda}^k) = \mathbf{A}\mathbf{x}^{k+1} - \mathbf{b}$ , donde  $\mathbf{x}^{k+1}$  minimiza  $L(\mathbf{x}, \boldsymbol{\lambda}^k)$ . Por tanto, se alcanza la solución iterativamente, ascendiendo con el gradiente en la función dual y minimizando el Lagrangiano:

$$\mathbf{x}^{k+1} = \operatorname{argmin}_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}^k), \quad (3.6)$$

$$\boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k + \alpha_k (\mathbf{A}\mathbf{x}^{k+1} - \mathbf{b}), \quad (3.7)$$

siendo  $\alpha_k \in (0,1]$  el tamaño del salto.

### 3.2 Método de Lagrange aumentado

Si bien, el método dual se presenta como una solución de minimización posible, puede no tener solución en el caso que atañe a este trabajo. Concretamente, aunque la función RSS es cuadrática, su convexidad puede estar en duda a causa de la correlación entre los regresores. Aunque este aspecto puede ser salvable por

aplicaciones como la distancia de Bregman [27], puede ser conveniente aplicar un término aumentado, para así también dar velocidad al avance del gradiente en (3.7). De manera que, se define la función de Lagrange aumentada como:

$$L_{\beta}(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \boldsymbol{\lambda}^T (\mathbf{Ax} - \mathbf{b}) + \frac{\beta}{2} \|\mathbf{Ax} - \mathbf{b}\|_2^2, \quad (3.8)$$

donde  $\beta$  es el parámetro aumentado. La ecuación (3.8) está asociada a un nuevo problema primal:

$$\min_{\mathbf{x}} \left( f(\mathbf{x}) + \frac{\beta}{2} \|\mathbf{Ax} - \mathbf{b}\|_2^2 \right), \quad s.t. \quad \mathbf{Ax} = \mathbf{b}. \quad (3.9)$$

Este nuevo enfoque asegura la diferenciabilidad de la función dual asociada y reduce la exigencia sobre la convexidad de  $f(\mathbf{x})$ . La solución al nuevo problema aumentado sería similar a (3.6) y (3.7):

$$\mathbf{x}^{k+1} = \underset{\mathbf{x}}{\operatorname{argmin}} L_{\beta}(\mathbf{x}, \boldsymbol{\lambda}^k), \quad (3.10)$$

$$\boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k + \beta (\mathbf{Ax}^{k+1} - \mathbf{b}). \quad (3.11)$$

En resumen, este pequeño cambio en la función de coste ayuda a la estabilidad y a la velocidad de convergencia en la resolución del problema.

### 3.3 Método de dirección alterna de multiplicadores (ADMM)

Hasta el momento, la función a minimizar ha sido única y dependiente únicamente de  $\mathbf{x}$ . ADMM, se plantea para el caso especial en el que la estructura del problema es separable:

$$\min_{\mathbf{x}, \mathbf{y}} (f(\mathbf{x}) + g(\mathbf{y})), \quad s.t. \quad \mathbf{Ax} + \mathbf{By} = \mathbf{b}. \quad (3.12)$$

La función de Lagrange aumentada para este problema sería:

$$L_{\beta}(\mathbf{x}, \mathbf{y}, \boldsymbol{\lambda}) = f(\mathbf{x}) + g(\mathbf{y}) + \boldsymbol{\lambda}^T (\mathbf{Ax} + \mathbf{By} - \mathbf{b}) + \frac{\beta}{2} \|\mathbf{Ax} + \mathbf{By} - \mathbf{b}\|_2^2, \quad (3.13)$$

donde la resolución, al igual que se procede con (3.8), consiste en resolver iterativamente:

$$(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}) = \underset{\mathbf{x}, \mathbf{y}}{\operatorname{argmin}} L_{\beta}(\mathbf{x}, \mathbf{y}, \boldsymbol{\lambda}^k), \quad (3.14)$$

$$\boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k + \beta (\mathbf{Ax}^{k+1} + \mathbf{By}^{k+1} - \mathbf{b}). \quad (3.15)$$

La alternancia en la dirección es relativa a la variable que se minimiza. En [30], se motiva la separación de la minimización de dos variables acopladas por las restricciones, y aunque este caso difiere ligeramente en forma, es posible separar (3.14) en dos, habilitando una actualización de  $\mathbf{x}^{k+1}$  e  $\mathbf{y}^{k+1}$  de forma separada [27]:

$$\mathbf{x}^{k+1} = \underset{\mathbf{x}}{\operatorname{argmin}} \left( f(\mathbf{x}) + g(\mathbf{y}^k) + \boldsymbol{\lambda}^{T,k} (\mathbf{Ax} + \mathbf{By}^k - \mathbf{b}) + \frac{\beta}{2} \|\mathbf{Ax} + \mathbf{By}^k - \mathbf{b}\|_2^2 \right), \quad (3.16)$$

$$\mathbf{y}^{k+1} = \underset{\mathbf{y}}{\operatorname{argmin}} \left( f(\mathbf{x}^{k+1}) + g(\mathbf{y}) + \boldsymbol{\lambda}^{T,k} (\mathbf{Ax}^{k+1} + \mathbf{By} - \mathbf{b}) + \frac{\beta}{2} \|\mathbf{Ax}^{k+1} + \mathbf{By} - \mathbf{b}\|_2^2 \right), \quad (3.17)$$

$$\boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k + \beta (\mathbf{Ax}^{k+1} + \mathbf{By}^{k+1} - \mathbf{b}). \quad (3.18)$$

La posibilidad de resolver el problema mediante estas ecuaciones es lo que da nombre al método. Puede verse cómo en (3.16) se minimiza  $L_{\beta}(\mathbf{x}, \mathbf{y}, \boldsymbol{\lambda})$  en la dirección de  $\mathbf{x}$ , mientras que en (3.17) se minimiza en la dirección de  $\mathbf{y}$ , actualizando el vector de penalizaciones  $\boldsymbol{\lambda}$  en (3.18).

### 3.4 Variaciones de la función a minimizar

Hasta ahora, la función a minimizar ha sido la ecuación (2.13), que aporta la solución de mínimo error, tal como se planteaba en la Sección 2.5. Ciertamente, el problema asociado a RSS tiene la forma de (3.1), pero es habitual que en este tipo de escenarios el número de condición de la matriz a invertir sea alto. En [31], se plantea una forma de atajar este problema mediante los denominados métodos de regularización (*shrinkage methods*), entre los que se encuentran la regularización de Ridge, la aplicación de *least absolute shrinkage and selection operator* (LASSO) o el estimador de Bayes que es una combinación de los anteriores. La regularización de Ridge, como se presentó en [32], consiste en una regularización de norma  $\ell_2$  y se propuso como un método adecuado para llevar a cabo la predistorsión digital cuando la inestabilidad es muy elevada. La principal diferencia entre estas dos regularizaciones es que Ridge, al ser de norma  $\ell_2$ , permite una solución suave, donde se ajustan las distintas componentes del resultado para satisfacer la función de coste. En cambio, LASSO al ser de norma  $\ell_1$  restringe la cantidad de componentes que aparecen en la solución, eliminando los regresores menos relevantes en cuanto a varianza. Además, como se indica en [33], LASSO se aplica con una umbralización. Esto implica que hay una diferencia respecto al mejor subset (*hard threshold*) de componentes presentado en [31], y es el hecho de que los regresores supervivientes también experimentan una reducción en el coeficiente (*soft threshold*), mejorando igualmente el número de condición. Se puede intuir cómo se trata de un compromiso entre la estabilidad de las soluciones y la exactitud de la solución, es decir, empeorando ligeramente la solución se consigue robustez.

Normalmente, se suele trabajar bajo la hipótesis de que la regresión admite solución escasa (*sparse solution*), ya que se da por hecho que LASSO va a eliminar un buen número de componentes [34]. Esto abre una discusión acerca de la condición *sparse*, que principalmente se fundamenta en los resultados que aportan los algoritmos de selección de coeficientes, y es necesario dedicarle algunas líneas puesto que se desea contrastar el rendimiento de la técnica en contraposición con el que se consigue aplicando DOMP. Técnicamente, en el paso final del algoritmo DOMP, se realiza una elección del mejor subset de componentes, donde el umbral de corte del *hard threshold* lo fija el *Bayesian Information Criterion* (BIC) [35], que busca el óptimo entre el número de regresores del modelo y la información aportada. La gran diferencia reside en que esta selección se realiza después de haber ordenado los coeficientes en base a la correlación, por lo que el criterio no está en la varianza de cada componente sino en cuánta información novedosa aporta el regresor. Así, también es de esperar que la exactitud de la solución aportada por DOMP sea mejor que la de hacer LASSO, puesto que la selección es más inteligente en el primer caso que en el segundo. Esto también pone de manifiesto que, dado que la regularización de norma  $\ell_1$  no tiene solución cerrada, existen diversidad de métodos alternativos para resolver el mismo problema.

Se pretende pues, hacer una combinación de la regularización de LASSO con la estructura de ADMM centralizado que se presenta a continuación.

### 3.5 Implementación de ADMM centralizado para el problema propuesto: Modelado

Como se ha comentado en la introducción de este trabajo, el objetivo fundamental es conseguir una computación distribuida para el modelado y predistorsión de los PA. El matiz de ADMM centralizado reside en la topología de la solución que se pretende alcanzar, donde distintos nodos independientes resuelven parte del problema, dependiendo de un nodo central que distribuye la información y hace converger la solución. Esto irá saliendo a la luz conforme se vaya planteando el problema.

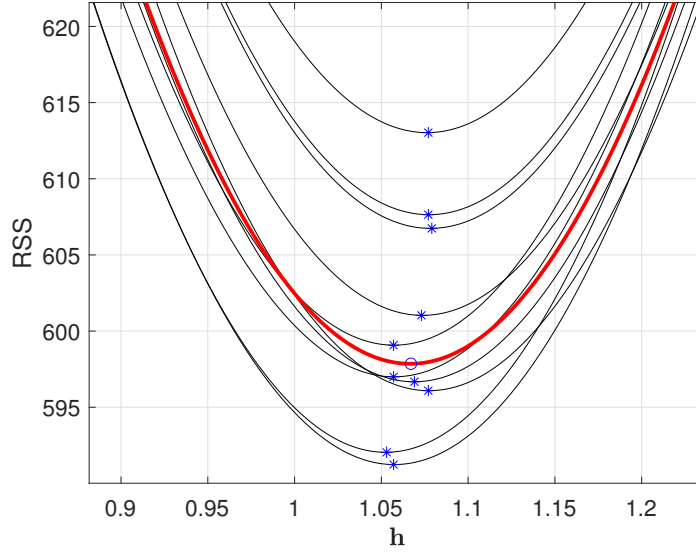
Aunque la introducción de las regularizaciones comentadas anteriormente puede formularse con una sola función, con el fin de poder dividir más fácilmente el problema en trozos. Empezando por la suma residual de cuadrados penalizada (PRSS), del inglés, *penalized residual sum of squares*:

$$\text{PRSS}_{\text{LASSO}} = \|\mathbf{y} - \mathbf{X}\mathbf{h}\|_2^2 + \lambda \|\mathbf{h}\|_1, \quad (3.19)$$

puede entenderse como el lagrangiano de un problema donde la restricción se realiza sobre la norma  $\ell_1$  del vector de coeficientes. Véase cómo ahora  $\lambda$  representa el coeficiente de regularización. Pero para aprovechar la potencia de lo presentado en la Sección 3.3, se plantea asociar la función PRSS al siguiente problema:

$$\min_{\mathbf{u}, \mathbf{h}} \frac{1}{2} \|\mathbf{y} - \mathbf{X}\mathbf{u}\|_2^2 + \frac{\lambda}{2} \|\mathbf{h}\|_1 \quad s.t. \quad \mathbf{A}\mathbf{u} + \mathbf{B}\mathbf{h} = \mathbf{c}, \quad (3.20)$$

donde en principio  $\mathbf{A}$ ,  $\mathbf{B}$  y  $\mathbf{c}$  están por determinar. Para facilitar el desarrollo posterior, se incluye un factor de  $1/2$  en ambos términos de la expresión, que no cambia la forma de la función. En el caso de uso de este trabajo, la función  $f_1(\cdot) = \dots = f_M(\cdot)$  será siempre la misma, ya que se desea que cada agente resuelva el mismo problema, donde  $\{1, \dots, M\}$  ahora representa el conjunto de  $M$  agentes que van a tratar de resolver el subproblema que le corresponda. El hecho de que cada agente resuelva el problema con una parte de los datos tendrá un efecto negativo en la exactitud de la solución. Este efecto no tiene por qué ser demasiado perjudicial, pero es ineludible, y ha de tenerse en cuenta. En la Figura 3.2 se muestra el efecto sobre la curva resultante de la función RSS, empleando datos de ejemplo.



**Figura 3.2** Ejemplo de función de minimización RSS. En rojo se representa la curva obtenida con todos los puntos mientras que en negro se presentan las curvas correspondientes a dividir el problema en 10 partes con el mismo número de muestras.

Realmente, se desea que las expresiones (3.19) y (3.20) alcancen la misma solución, por lo que se entiende que siguiendo [36]  $\mathbf{A} = \mathbf{I}$ ,  $\mathbf{B} = -\mathbf{I}$  y  $\mathbf{c} = \mathbf{0}$ . Se reescribe (3.20) como [37]:

$$\min_{\mathbf{u}, \mathbf{h}} \frac{1}{2} \|\mathbf{y} - \mathbf{X}\mathbf{u}\|_2^2 + \frac{\lambda}{2} \|\mathbf{h}\|_1 \quad s.t. \quad \mathbf{u} - \mathbf{h} = \mathbf{0}, \quad (3.21)$$

resultando en una forma que admite la solución por alternancia de dirección.

Ya se ha planteado la resolución de este problema anteriormente, pero con el fin de alcanzar la computación distribuida es conveniente reformularlo. Siguiendo [27], es interesante buscar una expresión semejante a

$$\min_{\{\mathbf{u}^i\}, \mathbf{h}} \sum_{i=1}^M f_i(\mathbf{u}^i) \quad s.t. \quad \mathbf{u}^i = \mathbf{h}, \quad \forall i \in \{1, \dots, M\}, \quad (3.22)$$

donde ahora,  $M$  agentes van a tratar de resolver el problema de forma distribuida. Este razonamiento es lo que se presenta en [36] como *row wise division*, es decir, división sabia de filas. Por tanto, si se divide la matriz inteligentemente por filas se replantea el problema (3.21), como se hace en [37]:

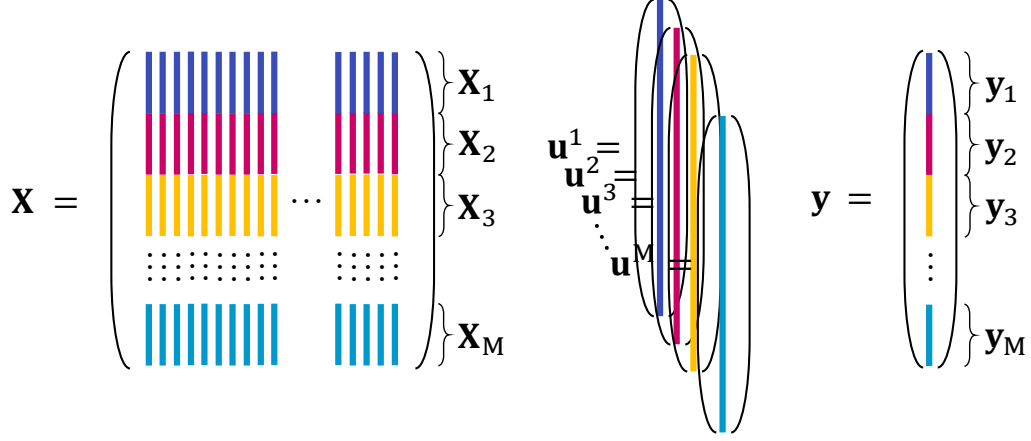
$$\min_{\{\mathbf{u}^i\}, \mathbf{h}} \frac{1}{2} \sum_{i=1}^M \|\mathbf{y}_i - \mathbf{X}_i \mathbf{u}^i\|_2^2 + \frac{\lambda}{2} \|\mathbf{h}\|_1 \quad (3.23)$$

$$s.t. \quad \mathbf{u}^i - \mathbf{h} = \mathbf{0}, \quad \forall i \in \{1, \dots, M\},$$

cuyo lagrangiano es

$$L_\rho(\mathbf{u}^1, \dots, \mathbf{u}^M, \mathbf{h}, \mathbf{d}^1, \dots, \mathbf{d}^M) = \frac{1}{2} \sum_{i=1}^M \|\mathbf{y}_i - \mathbf{X}_i \mathbf{u}^i\|_2^2 + \frac{\lambda}{2} \|\mathbf{h}\|_1 + \sum_{i=1}^M \mathbf{d}^T (\mathbf{u}^i - \mathbf{h}) + \frac{\rho}{2} \sum_{i=1}^M \|\mathbf{u}^i - \mathbf{h}\|_2^2, \quad (3.24)$$

donde el superíndice  $i$  significa réplica  $i$ -ésima mientras que el subíndice  $i$  representa la porción  $i$ -ésima. Se advierte en este punto un cambio en la notación. Véase cómo ahora  $\lambda$  es el coeficiente de regularización,  $\mathbf{d}$  ha pasado a ser el vector de multiplicadores de Lagrange y  $\rho$  el parámetro aumentado. La división en  $M$  problemas que representa la nueva función de minimización se ilustra en la Figura 3.3.



**Figura 3.3** Esquemático de división sabia de filas, donde se ha de minimizar el conjunto de  $\{f_1(\mathbf{u}_1), \dots, f_M(\mathbf{u}_M)\}$  funciones de minimización, según (3.22).

Por sencillez, se lleva a cabo la transformación planteada en [36]:

$$\mathbf{d}^T \mathbf{r} + \frac{\rho}{2} \|\mathbf{r}\|_2^2 = \frac{\rho}{2} \|\mathbf{r} + \mathbf{s}\|_2^2 - \frac{\rho}{2} \|\mathbf{s}\|_2^2, \quad (3.25)$$

donde  $\mathbf{r}$  representa la restricción  $\mathbf{r} = \mathbf{u} - \mathbf{h}$  y  $\mathbf{s} = \mathbf{d}/\rho$  es el nuevo vector de multiplicadores de Lagrange aumentados. Así, el lagrangiano, con el particionado, queda de la forma:

$$L_\rho(\mathbf{u}^1, \dots, \mathbf{u}^M, \mathbf{h}, \mathbf{s}^1, \dots, \mathbf{s}^M) = \frac{1}{2} \sum_{i=1}^M \|\mathbf{y}_i - \mathbf{X}_i \mathbf{u}^i\|_2^2 + \frac{\lambda}{2} \|\mathbf{h}\|_1 + \sum_{i=1}^M \frac{\rho}{2} \|\mathbf{u}^i - \mathbf{h} + \mathbf{s}^i\|_2^2 - \frac{\rho}{2} \sum_{i=1}^M \|\mathbf{s}^i\|_2^2. \quad (3.26)$$

Hasta ahora se ha planteado la estructura del problema, viendo que realmente tiene semejanza con la resolución por el método de dirección alterna de multiplicadores. Ahora bien, el caso planteado tiene una peculiaridad en cuanto a la forma de sus restricciones, y es el hecho de que  $\mathbf{u}^i = \mathbf{h}$  demanda que todas las soluciones  $\mathbf{u}^i$  converjan a un mismo resultado  $\mathbf{h}$ . Esto es lo que se va a denominar consenso entre los agentes. A priori, la mejor medida del consenso es la media y es la solución que se plantea en [36], [38] y [27], entre otros. A continuación, se desarrollan los pasos en ambas direcciones, empezando por el cómputo de los agentes, siguiendo por el consenso y terminando por la actualización de los multiplicadores aumentados. Desarrollando el avance de minimización en  $\mathbf{u}$  quedaría:

$$\begin{aligned} \mathbf{u}^{k+1} &= \underset{\mathbf{u}}{\operatorname{argmin}} \left( \frac{1}{2} \|\mathbf{y} - \mathbf{X}\mathbf{u}\|_2^2 + \frac{\lambda}{2} \|\mathbf{h}^k\|_1 + \frac{\rho}{2} \|\mathbf{u} - \mathbf{h}^k + \mathbf{s}^k\|_2^2 - \frac{\rho}{2} \|\mathbf{s}^k\|_2^2 \right) \\ &= \underset{\mathbf{u}}{\operatorname{argmin}} \left( \frac{1}{2} (\mathbf{y}^H \mathbf{y} - \mathbf{u}^H \mathbf{X}^H \mathbf{y} - \mathbf{y}^H \mathbf{X} \mathbf{u} + \mathbf{u}^H \mathbf{X}^H \mathbf{X} \mathbf{u}) + \frac{\lambda}{2} \|\mathbf{h}^k\|_1 \right. \\ &\quad \left. + \frac{\rho}{2} (\mathbf{u}^H \mathbf{u} - (\mathbf{h}^k - \mathbf{s}^k)^H \mathbf{u} - \mathbf{u}^H (\mathbf{h}^k - \mathbf{s}^k) + (\mathbf{h}^k - \mathbf{s}^k)^H (\mathbf{h}^k - \mathbf{s}^k)) - \frac{\rho}{2} \mathbf{s}^{H,k} \mathbf{s}^k \right). \end{aligned} \quad (3.27)$$

Para encontrar el argumento en  $\mathbf{u}$  que minimiza la expresión, se deriva

$$\begin{aligned} \frac{\delta L_\rho(\mathbf{u}, \mathbf{h}^k, \mathbf{s}^k)}{\delta \mathbf{u}^*} &= \frac{1}{2} (-2\mathbf{X}^H \mathbf{y} + 2\mathbf{X}^H \mathbf{X} \mathbf{u}) + \frac{\rho}{2} (2\mathbf{u} - 2(\mathbf{h}^k - \mathbf{s}^k)) \\ &= -\mathbf{X}^H \mathbf{y} + \mathbf{X}^H \mathbf{X} \mathbf{u} + \rho \mathbf{u} - \rho(\mathbf{h}^k - \mathbf{s}^k), \end{aligned} \quad (3.28)$$

se iguala a 0 y se despeja  $\mathbf{u}$ , obteniendo:

$$\begin{aligned} \mathbf{X}^H \mathbf{X} \mathbf{u} + \rho \mathbf{u} &= \mathbf{X}^H \mathbf{y} + \rho(\mathbf{h}^k - \mathbf{s}^k) \\ \mathbf{u}^{k+1} &= (\mathbf{X}^H \mathbf{X} + \rho \mathbf{I})^{-1} (\mathbf{X}^H \mathbf{y} + \rho(\mathbf{h}^k - \mathbf{s}^k)), \end{aligned} \quad (3.29)$$

que recuperando el particionado resulta:

$$\mathbf{u}^{i,(k+1)} = (\mathbf{X}_i^H \mathbf{X}_i + \rho \mathbf{I})^{-1} (\mathbf{X}_i^H \mathbf{y}_i + \rho(\mathbf{h}^k - \mathbf{s}^{i,(k)})). \quad (3.30)$$

De manera que, cada agente se encarga de resolver la ecuación (3.30) a partir de  $\mathbf{h}^k$  que proviene del nodo central. Desarrollando la ecuación del nodo central, manteniendo esta vez el laplaciano con la división:

$$\mathbf{h}^{k+1} = \underset{\mathbf{h}}{\operatorname{argmin}} \sum_{i=1}^M \left( \frac{1}{2} \|\mathbf{y}_i - \mathbf{X}_i \mathbf{u}^{i,(k+1)}\|_2^2 + \frac{\lambda}{2M} \|\mathbf{h}\|_1 + \frac{\rho}{2} \|\mathbf{u}^{i,(k+1)} - \mathbf{h} + \mathbf{s}^{i,(k)}\|_2^2 - \frac{\rho}{2} \|\mathbf{s}^{i,(k)}\|_2^2 \right). \quad (3.31)$$

Es importante destacar que  $(\lambda/2) \|\mathbf{h}\|_1$  se ha introducido en el sumatorio con el factor  $1/M$  puesto que la función se computa una única vez en el nodo central. Eliminando los términos que no dependen de  $\mathbf{h}$ :

$$\mathbf{h}^{k+1} = \underset{\mathbf{h}}{\operatorname{argmin}} \frac{\rho}{2} \sum_{i=1}^M \left( \frac{\lambda}{M\rho} \|\mathbf{h}\|_1 + \|\mathbf{u}^{i,(k+1)} - \mathbf{h} + \mathbf{s}^{i,(k)}\|_2^2 \right). \quad (3.32)$$

En la línea de lo que se comenta en [39], bastaría con minimizar el segundo término y aplicarle el umbral suave (*soft threshold*). De modo que, si se desarrolla la segunda parte:

$$\mathbf{h}^{k+1} = \underset{\mathbf{h}}{\operatorname{argmin}} \frac{\rho}{2} \sum_{i=1}^M \left( \mathbf{u}^{i,(k+1),H} \mathbf{u}^{i,(k+1)} - (\mathbf{h} - \mathbf{s}^{i,(k)})^H \mathbf{u}^{i,(k+1)} - \mathbf{u}^{i,(k+1),H} (\mathbf{h} - \mathbf{s}^{i,(k)}) + (\mathbf{h} - \mathbf{s}^{i,(k)})^H (\mathbf{h} - \mathbf{s}^{i,(k)}) \right), \quad (3.33)$$

operando y conservando de nuevo la parte dependiente de  $\mathbf{h}$ :

$$\mathbf{h}^{k+1} = \underset{\mathbf{h}}{\operatorname{argmin}} \frac{\rho}{2} \sum_{i=1}^M \left( -\mathbf{h}^H \mathbf{u}^{i,(k+1)} - \mathbf{u}^{i,(k+1),H} \mathbf{h} + \mathbf{h}^H \mathbf{h} - \mathbf{h}^H \mathbf{s}^{i,(k)} - \mathbf{s}^{i,(k),H} \mathbf{h} \right). \quad (3.34)$$

Al igual que antes, para encontrar el argumento en  $\mathbf{h}$  que minimiza la expresión, se deriva

$$\frac{\delta L_\rho(\mathbf{u}^{k+1}, \mathbf{h}, \mathbf{s}^k)}{\delta \mathbf{h}^*} = \frac{\rho}{2} \sum_{i=1}^M \left( -2\mathbf{u}^{i,(k+1)} + 2\mathbf{h} - 2\mathbf{s}^{i,(k)} \right), \quad (3.35)$$

se iguala a 0 y se despeja, teniendo como resultado:

$$\begin{aligned} M\mathbf{h} &= \sum_{i=1}^M \mathbf{u}^{i,(k+1)} + \mathbf{s}^{i,(k)} \\ \mathbf{h}^{k+1} &= \frac{1}{M} \sum_{i=1}^M \mathbf{u}^{i,(k+1)} + \mathbf{s}^{i,(k)}. \end{aligned} \quad (3.36)$$

Queda de manifiesto lo afirmado anteriormente, siendo el resultado del consenso justamente la media del resultado de los agentes a la que se le suma el vector de multiplicadores aumentado. Para terminar con la aplicación de la regularización de LASSO en el cálculo de  $\mathbf{h}$ , se aplica el *soft threshold*, que toma la siguiente forma:

$$\mathbf{S}_\kappa(a) = \begin{cases} a - \kappa \operatorname{sign}(a), & |a| > \kappa \\ 0, & |a| < \kappa. \end{cases} \quad (3.37)$$

Por último, se actualiza el vector de multiplicadores de Lagrange aumentados al igual que en (3.18):

$$\mathbf{s}^{i,(k+1)} = \mathbf{s}^{i,(k)} + \mathbf{u}^{i,(k+1)} - \mathbf{h}^{(k+1)}. \quad (3.38)$$

A continuación, se recogen las tres ecuaciones iterativas que conforman el método:

$$\mathbf{u}^{i,(k+1)} = (\mathbf{X}_i^H \mathbf{X}_i + \rho \mathbf{I})^{-1} (\mathbf{X}_i^H \mathbf{y}_i + \rho (\mathbf{h}^{(k)} - \mathbf{s}^{i,(k)})), \quad (3.39)$$

$$\mathbf{h}^{(k+1)} = \mathbf{S}_{\frac{\lambda}{M\rho}} (\bar{\mathbf{u}}^{(k+1)} + \bar{\mathbf{s}}^{(k)}), \quad (3.40)$$

$$\mathbf{s}^{i,(k+1)} = \mathbf{s}^{i,(k)} + \mathbf{u}^{i,(k+1)} - \mathbf{h}^{(k+1)}, \quad (3.41)$$

donde la barra en  $\bar{\mathbf{u}}^{(k+1)}$  y  $\bar{\mathbf{s}}^{(k)}$  representa la operación de hacer la media. En la Figura 3.4 se pretende ilustrar la topología distribuida de esta técnica, donde  $M$  nodos computan una solución individualizada para que el nodo central devuelva el consenso.

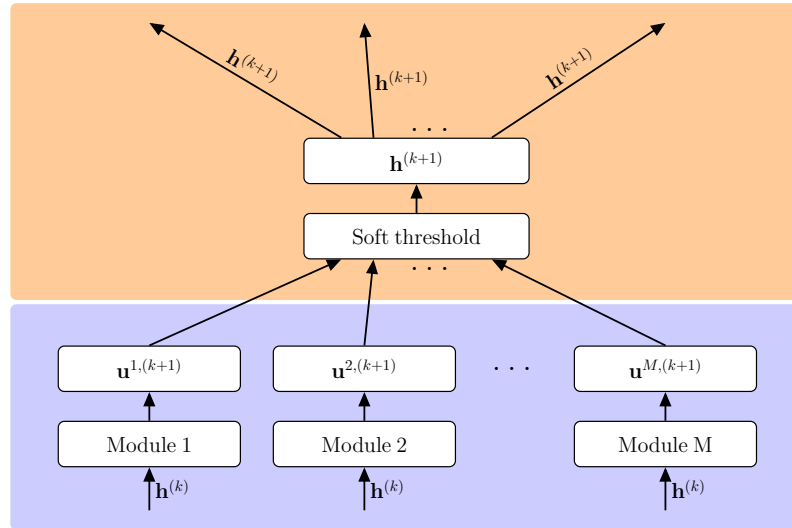


Figura 3.4 Topología de ADMM centralizado para la resolución del problema planteado [37].

### 3.6 Implementación de ADMM centralizado para el problema propuesto: Linealización

Antes de empezar, se advierte al lector que en esta sección va a encontrar distintos cambios de notación, debidos principalmente a que, en el contexto de linealización, se suele emplear  $\mathbf{x}$  para referirse a la entrada al amplificador y  $\mathbf{u}$  a la señal original generada. Por tanto,  $\mathbf{x}$  es la señal predistorcionada  $\mathbf{x} = \mathbf{u} - \mathbf{d}$ , donde  $\mathbf{d}$  es la señal de distorsión. Por último, al tratar con arquitecturas de aprendizaje máquina es habitual que el vector de coeficientes  $\mathbf{h}$  se denote por  $\mathbf{w}$ , refiriéndose así al concepto equivalente del vector de pesos (*weights*).

Una vez planteado ADMM como método posible para la resolución del problema, es el momento de introducirlo dentro de la disciplina de DPD. Hasta ahora, toda la resolución ha girado en torno al objetivo de encontrar un modelo de amplificador, es decir, una expresión que relacione la entrada y la salida, y que para una señal nueva, el modelo sea capaz de predecir con la mayor exactitud posible cuál va a ser la respuesta del PA. Pues bien, cuando se habla de linealización, se quiere ir un paso más allá, tratando de aprovechar la información que aporta el modelo para conseguir que el sistema de amplificación se comporte linealmente con la mayor eficiencia energética posible. Este concepto de DPD como sistema en cascada se muestra en la Figura 3.5.

Observando la Figura 3.5, se entiende que el objetivo del DPD consiste en compensar al PA, es decir, transformar la señal original  $\mathbf{u}$  en una versión predistorcionada  $\mathbf{x}$  que consiga que la salida y presente la





Figura 3.5 Esquema conceptual de DPD.

menor distorsión posible. Una de las implementaciones ampliamente utilizadas que permiten entender el concepto de DPD es la Arquitectura de Aprendizaje Indirecto (ILA). Principalmente consiste en determinar la relación  $\mathbf{x} \rightarrow f^{-1}(\mathbf{y})$  durante la etapa de entrenamiento, mediante el mismo procedimiento de regresión. Esta forma de proceder implica dar por hecho que los coeficientes que estiman la entrada en función de la salida son válidos para representar la inversa del PA, pudiéndolos emplear para predistorsionar la señal de entrada. La hipótesis es válida, demostrada por Schetzen en [40], y prueba de ello es el empleo de la técnica en multitud de trabajos, como por ejemplo en [41] y [8].

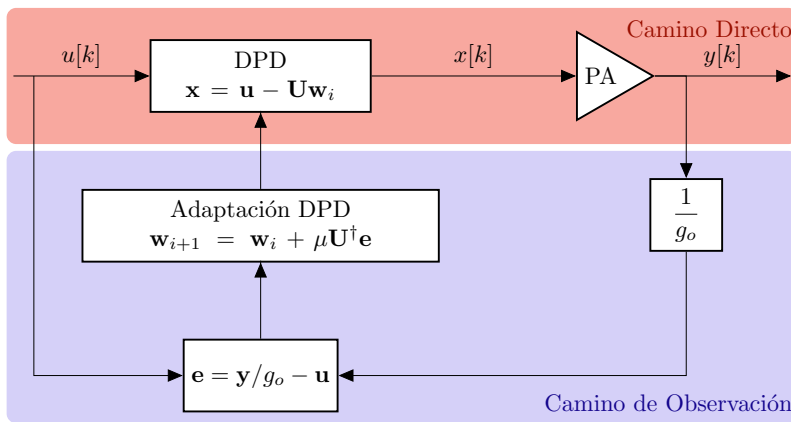


Figura 3.6 Arquitectura de aprendizaje directo.

Aun siendo válida e impecable teóricamente, ILA no es óptima en la práctica, hecho que motiva la aparición de la Arquitectura de Aprendizaje Directo (DLA) [42], cuyo esquemático se presenta en la Figura 3.6. Por tanto, ya no se trata de encontrar una relación en el camino de observación como se hacía en ILA, sino de obtenerla en el camino directo. En este caso, sí se emplea la misma matriz de regresores que en el caso de modelado, pero ahora no se desea encontrar un vector de coeficientes para estimar  $\mathbf{y}$ , sino que se pretende estimar directamente la señal de distorsión  $\mathbf{d} = \mathbf{U}\mathbf{w}$ , es decir, únicamente el efecto negativo que está teniendo el PA sobre la señal de entrada [32]. Esto se consigue iterativamente en la etapa de adaptación, donde  $\mu \in (0, 1]$  es el factor de aprendizaje. La interpretación del proceso iterativo es algo más compleja, ya que a la entrada del PA se va perfeccionando  $\mathbf{x}$  mediante la corrección que se realiza al substraer  $\mathbf{U}\mathbf{w}$  a señal original  $\mathbf{u}$  en cada paso. En cualquier caso, ahora se pretende encontrar  $\mathbf{w}$  t.q.  $|\mathbf{e}| = 0$  siendo este error la diferencia entre la salida normalizada por la ganancia objetivo y la señal original de entrada.

Prestando atención de nuevo al bloque de adaptación, se puede ver cómo realmente está compuesto por dos operaciones. La primera de ellas consiste en el cálculo de los incrementos del vector de pesos, esto es,  $\mathbf{d}\mathbf{w}_i = \mathbf{U}^T \mathbf{e}$ , para posteriormente actualizar  $\mathbf{w}_{i+1} = \mathbf{w}_i + \mu \mathbf{d}\mathbf{w}_i$ . Así se está llevando a cabo un descenso del gradiente, donde lo interesante se encuentra en el método de cálculo de los  $\mathbf{d}\mathbf{w}_i$ . La realización de la pseudoinversa de Moore-Penrose es justamente el resultado de la solución de LS. Así que, a continuación, se disocia esa etapa de cálculo para realizarla mediante ADMM, presentando el resultado esquemático en la Figura 3.7

Nótese que en el bloque de cálculo de  $\mathbf{d}\mathbf{w}_i$  se ha editado la notación, quedando ahora representada la solución de cada problema individual por la variable  $\mathbf{t}^{m,(k+1)}$ .

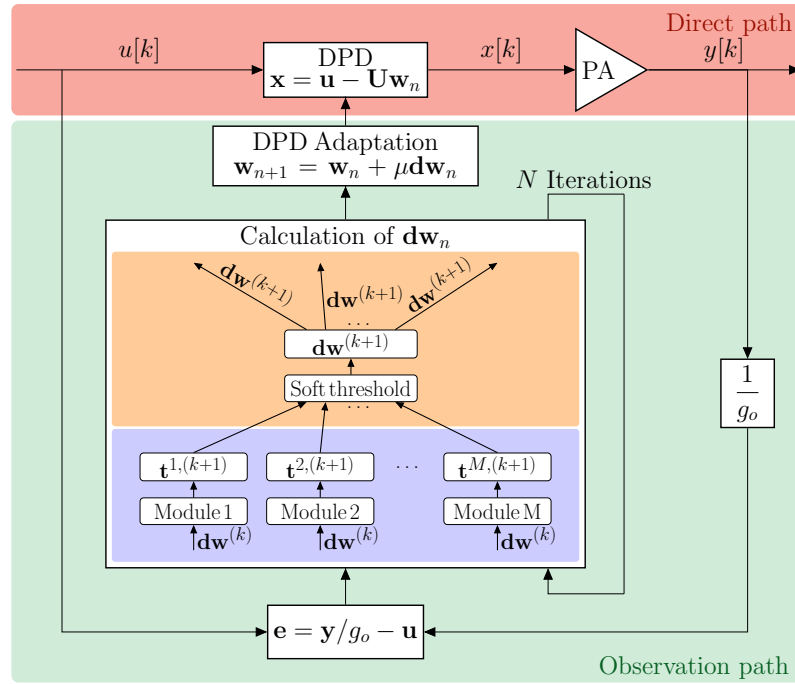


Figura 3.7 Implementación de ADMM para la arquitectura de aprendizaje directo [37].

### 3.7 Complejidad computacional

Llegados a este punto, es interesante evaluar teóricamente qué supone la división del problema en trozos más pequeños en cuanto a la computación de la solución se refiere. Para ello puede evaluarse el número de operaciones realizadas por un módulo computacional, esto es, el cálculo de los  $\mathbf{u}^{i,(k+1)}$ . Se recupera a continuación la operación:

$$\mathbf{u}^{i,(k+1)} = (\mathbf{X}_i^H \mathbf{X}_i + \rho \mathbf{I})^{-1} (\mathbf{X}_i^H \mathbf{y}_i + \rho (\mathbf{h}^{(k)} - \mathbf{s}^{i,(k)})), \quad (3.42)$$

que debe ser comparada con la solución de mínimos cuadrados:

$$\mathbf{h} = (\mathbf{X}^H \mathbf{X})^{-1} \mathbf{X}^H \mathbf{y}. \quad (3.43)$$

Puede observarse cómo realmente ambas ecuaciones tienen la misma estructura. El orden de las operaciones que se realiza es el siguiente, que se desarrolla para el caso de mínimos cuadrados:

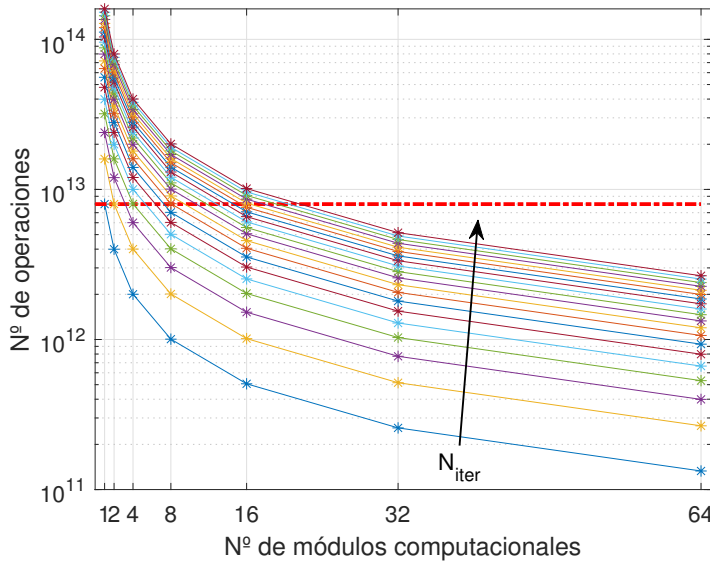
1. Cálculo de la matriz de autocorrelaciones muestrales ( $\mathbf{R}_\mathbf{X}$ ). Esto consiste en el producto  $\mathbf{X}^H \mathbf{X}$ , es decir, una matriz  $M \times N$  por una  $N \times M$ . En el producto de matrices se realizan  $M^2 N$  multiplicaciones y  $M^2(N-1)$  sumas, obteniendo un total de  $2M^2 N - M^2$  operaciones.
2. Cálculo de la matriz de precisión ( $\mathbf{P}_\mathbf{X} = \mathbf{R}_\mathbf{X}^{-1}$ ). Según [43], la inversa de una matriz cualquiera de  $M$  columnas siguiendo el algoritmo de Gauss requiere un total de  $M(M+1)/2$  divisiones,  $(2M^3 + 3M^2 - 5M)/6$  productos y  $(2M^3 + 3M^2 - 5M)/6$  restas. Se decide truncar aquellos términos de orden uno para simplificar, teniendo un total de  $2M^3/3 + 3M^2/2$  operaciones.
3. Cálculo de la matriz de autocorrelación cruzada ( $\mathbf{R}_{\mathbf{X}\mathbf{y}} = \mathbf{X}^H \mathbf{y}$ ). Se realizan  $MN$  multiplicaciones y  $M(N-1)$  sumas, alcanzando un total aproximado de  $2MN$  operaciones.
4. Producto  $\mathbf{P}_\mathbf{X} \mathbf{R}_{\mathbf{X}\mathbf{y}}$ . Producto de una matriz  $M \times M$  por un vector  $M \times 1$  que resulta en aproximadamente  $2M^2$  operaciones.

Habiendo simplificado los términos de orden uno, puede verse que se obtienen las mismas expresiones para el caso del cálculo mediante ADMM, donde el número de filas ya no es  $N$ , sino  $N/M_{\text{módulos}}$ . Así, la complejidad computacional de cada caso resulta:

$$\mathcal{O}_{LS} = 2M^2N + \frac{2}{3}M^3 + 2MN + \frac{5}{2}M^2 \quad (3.44)$$

$$\mathcal{O}_{ADMM} = N_{iter} \left( 2M^2 \frac{N}{M_{modulos}} + \frac{2}{3}M^3 + 2M \frac{N}{M_{modulos}} + \frac{5}{2}M^2 \right) \quad (3.45)$$

donde  $N_{iter}$  es el número de iteraciones que se ejecuta el método. Como era de esperar, el número de operaciones es inversamente proporcional al número de módulos computacionales ( $M_{modulos}$ ).



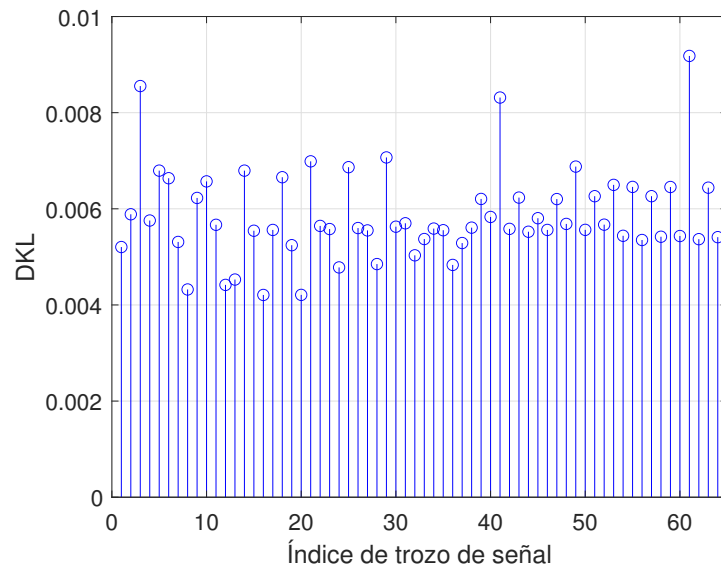
**Figura 3.8** Comparativa de complejidad computacional entre LS, representada en la línea roja discontinua y ADMM entre 1 y 20 iteraciones del método.

En la Figura 3.9 se muestra una comparativa del número de operaciones de ambos métodos, parametrizando en función del número de iteraciones que se realicen en ADMM.

### 3.8 Consenso de los distintos subproblemas

Otro aspecto interesante a tener en cuenta tiene que ver con el subproblema que observa cada módulo, como una parte aislada del total de muestras, obteniendo como es de esperar una solución distinta de cada módulo de cómputo. Este fenómeno ya se observaba en el ejemplo de la Figura 3.2. Cuanto menos se parezcan las soluciones entre los distintos subproblemas, peor calidad tendrá el consenso obligando al método a realizar más iteraciones, y obteniendo una peor calidad de la solución. También se debe tener en cuenta que no siempre el consenso es posible. Además, la capacidad de generalizar del modelo puede verse afectada si se emplean trozos de señal demasiado cortos, puesto que esa falta de muestras provocará que cada módulo tienda a tratar de aprenderse el trozo de señal que le corresponde. Es por esto que se considera que no es recomendable tomar trozos demasiado pequeños. Para abordar esta cuestión numéricamente, se plantea la divergencia de Kullback-Leibler para el caso en el que se distribuye la señal en 64 trozos, que supone trabajar con longitudes del 1.56% de la señal completa. En este caso, el modelo  $q$  será la señal completa, mientras que  $p_X^d$  será cada trozo. Para el caso de la señal con semilla aleatoria 1000 se obtuvo la siguiente distribución de divergencias.

Los valores obtenidos son lo bastante bajos y bastante uniformes como para esperar un parecido uniforme en las soluciones de cada módulo computacional.



**Figura 3.9** Divergencia de Kullback-Leibler de cada trozo de señal con la señal completa para la semilla aleatoria 1000, en un escenario de 64 trozos.

## 4 Diseño experimental y resultados

*El mundo exige resultados. No le cuentes a otros tus dolores del parto. Muéstrales al niño.*

INDIRA GANDHI

En los capítulos anteriores, se han comentado multitud de aspectos que se pretenden ahora poner a prueba para corroborar la idoneidad de los razonamientos planteados. Para ello, en este capítulo se recogen distintas experimentaciones que tratan de abordar estos puntos relativos al cálculo distribuido en problemas de modelado y linealización de amplificadores de potencia.

### 4.1 Banco de medidas

Todos los experimentos recogidos en este trabajo se realizaron mediante la plataforma RF WebLab de la Universidad de Chalmers [44]. El banco de medidas se compone de un transceptor vectorial de señal PXIe-5646R de National Instruments con un ancho de banda instantáneo de 200 MHz, con frecuencia central de 2,14 GHz. Esto permite trabajar con un ancho de banda de la señal modulada suficientemente elevado como para requerir el empleo de un modelo subyacente basado en el GVG. La etapa amplificadora está compuesta, en primer lugar, por un preamplificador lineal con ganancia de 40 dB, seguido del PA que se desea evaluar. El amplificador de potencia bajo prueba está basado en el transistor de GaN CGH40006P de Cree y está alimentado por un módulo PXI-4130 de National Instruments como fuente de alimentación. Tras la etapa amplificadora se coloca un atenuador de 30 dB para proteger al transceptor. Todo el sistema es accesible desde MATLAB®.

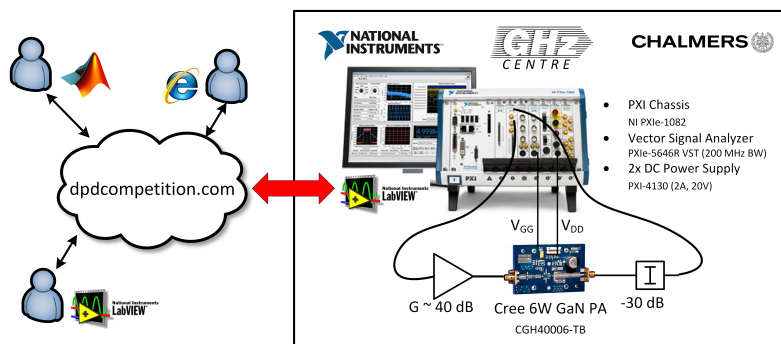


Figura 4.1 Setup de medidas del RF WebLab de la Universidad de Chalmers.

Como se ha comentado anteriormente, toda la experimentación se realiza con señales OFDM que siguen el formato de 5G-NR. Éstas presentan un ancho de banda de 50 MHz, con una separación entre portadoras de 30 kHz. Cada subportadora es modulada con símbolos 16-QAM, que son generados por una secuencia aleatoria de bits, controlada por una semilla de aleatorización. A las señales se les aplicó un factor de

sobremuestreo en torno a 3,3 para alcanzar la tasa de muestreo de 200 MHz empleada por el RF WebLab. Finalmente, se realizó un postprocesado para alinear la entrada con la salida. En el caso del modelado, se realizó un promediado adicional con 10 repeticiones para bajar el nivel de ruido. Esto no se planteó en DPD puesto que va en detrimento de una aplicación en tiempo real, aunque siempre es beneficioso para conseguir un mejor NMSE. El amplificador se situó en un punto de operación que devolviese una potencia promedio a la salida de unos 28,9 dBm en todas las pruebas de modelado. En las pruebas de linealización la potencia promedio fue de 29,5 dBm. Este cambio en la potencia promedio se debe a pequeños cambios de la Universidad de Chalmers que no se han podido controlar. En estas condiciones la compresión de ganancia del PA fue de aproximadamente 1,5 dB.

## 4.2 Figuras de mérito

A continuación se exponen las distintas figuras de mérito que se emplearán para juzgar la calidad de los resultados obtenidos. Estas figuras son ampliamente utilizadas en el campo de la predistorsión digital, debido a que dan una buena información sobre la distorsión de la señal. La principal figura de mérito que ayudará a extraer conclusiones en este trabajo es el NMSE (*Normalized Mean Square Error*). Se trata de una medida cuantitativa del parecido entre la señal modelada  $y_{mod}$ , es decir, la estimación, y la señal medida  $y_{med}$ .

$$\text{NMSE}[\text{dB}] = 10 \log_{10} \left\{ \frac{\sum_{n=1}^N |y_{mod}[n] - y_{med}[n]|^2}{\sum_{n=1}^N |y_{med}[n]|^2} \right\}. \quad (4.1)$$

De manera que, cuanto menor sea el NMSE más se parecerá la estimación a la medida real. Esta figura está asociada tanto al modelado como a la linealización, de modo que es un indicador de exactitud en ambos casos de uso. En el caso del DPD, el error se calcula entre la salida dividida por la ganancia objetivo y la entrada, colocando en el denominador la entrada, como se muestra a continuación

$$\text{NMSE}_{\text{DPD}}[\text{dB}] = 10 \log_{10} \left\{ \frac{\sum_{n=1}^N \left| \frac{y_{med}[n]}{g_0} - u[n] \right|^2}{\sum_{n=1}^N |u[n]|^2} \right\}. \quad (4.2)$$

De esta forma se mide cuánto se separa la salida real de la salida perfectamente lineal, que sería  $\mathbf{y} = g_0 \cdot \mathbf{u}$ .

Otra medida importante, más relacionada con la linealización, es el ACPR (*Adjacent Channel Power Ratio*). Esta cantidad da una idea la distorsión fuera de banda, que es muy perjudicial para el desarrollo de las comunicaciones en bandas adyacentes. Se define para el canal superior y el inferior como

$$\text{ACPR}_{\text{inf}}[\text{dBc}] = 10 \log_{10} \left( \frac{\int_{f_{inf}-B/2}^{f_{inf}+B/2} S_y(f) df}{\int_{f_c-B/2}^{f_c+B/2} S_y(f) df} \right) \quad (4.3)$$

$$\text{ACPR}_{\text{sup}}[\text{dBc}] = 10 \log_{10} \left( \frac{\int_{f_{sup}-B/2}^{f_{sup}+B/2} S_y(f) df}{\int_{f_c-B/2}^{f_c+B/2} S_y(f) df} \right),$$

donde  $f_c$  es la frecuencia central,  $f_{inf}$  la frecuencia central de la banda inferior y  $f_{sup}$  la frecuencia central de la banda superior.

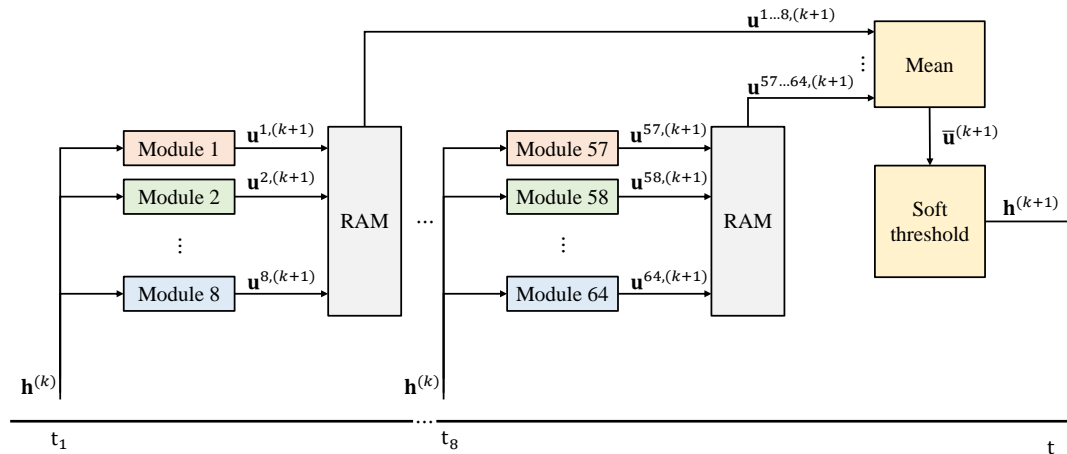
Finalmente, dado que se puede implementar un receptor digital porque se conoce el stream de bits que modula la señal, se puede calcular el EVM (*Error Vector Magnitude*). Se calcula como

$$\text{EVM}[\%] = \sqrt{\frac{\sum_{i=1}^N \|\vec{x}_i - \vec{x}_i'\|^2}{\sum_{i=1}^N \|\vec{x}_i'\|^2}} \times 100, \quad (4.4)$$

donde  $\vec{x}_i'$  es el símbolo demodulado recibido y  $\vec{x}_i$  el transmitido. Esto da una medida promedio de cuánto se desvían los puntos recibidos en la constelación del punto original transmitido. Ciertamente, existe un EVM por debajo del cual la probabilidad de error de símbolo resulta cero, ya que, se puede tener dispersión alrededor del punto original sin traspasar las fronteras de decisión. Esta medida da un valor de distorsión en banda, a diferencia del ACPR que se centra en caracterizar este efecto fuera de banda.

### 4.3 Implementación de ADMM

Se ha apuntado ya en las Secciones 3.5 y 3.6 la arquitectura en paralelo que posibilitan las ecuaciones de este método y cómo encajarlas en la resolución de los problemas de modelado y linealización. A continuación, se adaptan estas arquitecturas al equipo de computación empleado. El equipo dispone de una memoria RAM (*Random Access Memory*) de 128 GB, con un procesador Intel i9 con 24 núcleos de computación. El software de MATLAB®, en su *Parallel Computing Toolbox* limitó a 8 el número de trabajadores (*workers*) en paralelo. Atendiendo a lo comentado en la Sección 3.7, se ha decidido implementar un particionado de 64 trozos. Por tanto, el reparto de las operaciones entre los módulos computacionales es el que queda representado en la Figura 4.2.



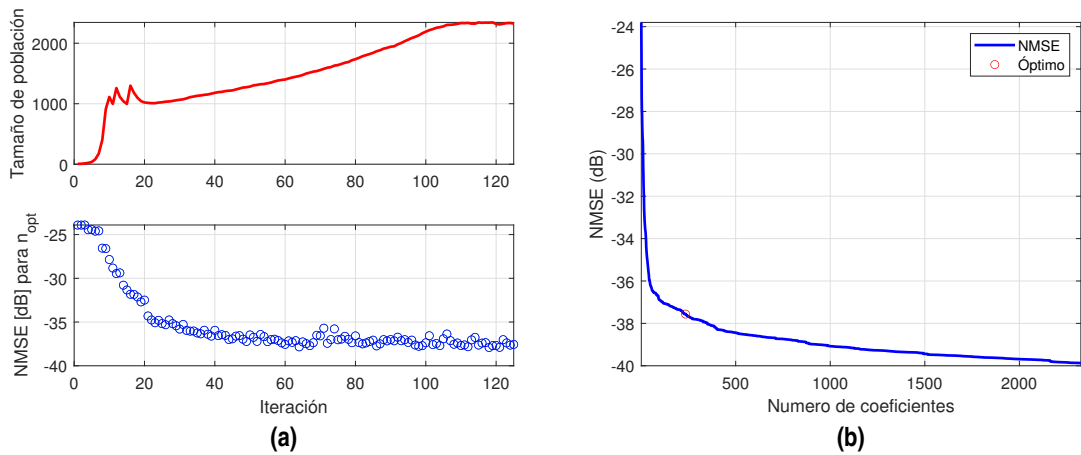
**Figura 4.2** Implementación llevada a cabo sobre el equipo. Las operaciones se realizan en 8 etapas en los tiempos  $t_1, \dots, t_8$ , guardando los resultados en memoria para finalmente calcular la media y aplicar el umbral suave.

Los valores de  $\lambda$  y  $\rho$  se obtuvieron buscando la minimización del NMSE, tomando los valores  $\lambda = 1,00 \cdot 10^{-7}$  y  $\rho = 1,42 \cdot 10^{-5}$ . Téngase en cuenta que el umbral queda definido por  $\kappa = \lambda / M\rho$ , obteniéndose en este caso  $\kappa = 1,1015 \cdot 10^{-4}$ . Esto quiere decir que tan solo quedan eliminados aquellos coeficientes que queden por debajo de este valor, y dado que es bastante bajo, se eliminan muy pocos regresores. Esto no quita que la regularización tenga efecto, puesto que en la resolución de la pseudoinversa, aparece el término  $\rho \mathbf{I}$ . Por último, se realizaron 20 iteraciones del método.

## 4.4 Modelado

El primer paso que da una idea de la idoneidad de la solución planteada es realizar una etapa de modelado, es decir, se plantea una estructura de regresores y se trata de calcular los coeficientes que mejor estimen las salidas  $y[n]$  del PA para las entradas  $x[n]$ . En esta etapa se quieren poner a prueba dos conceptos. El primero de ellos es la validez de ADMM como método de resolución, del cual se indicarán las especificaciones a continuación. El segundo concepto será la capacidad de generalizar del modelo basado en series de Volterra, que se ha visto teóricamente que es muy sensible a la PEP.

El modelo subyacente empleado es no estructurado y ha sido creado por GVG, ejecutando 125 generaciones. En la Figura 4.3 se muestra el avance de las iteraciones, donde se ve que el propio algoritmo incorpora la ejecución de DOMP en cada iteración. El generador de modelo devolvió una estructura de 2325 regresores, donde el set óptimo estaba compuesto por los 235 primeros regresores según el criterio BIC. Esta estructura se determinó con una señal a la que no se le hicieron promediados, y aunque es posible que la matriz de regresores cambie ligeramente, se espera que no influya demasiado en este caso, ya que se están seleccionando los más relevantes, que se espera estén poco correlados con el ruido. No obstante sería bueno realizar de nuevo la búsqueda. El modelo fue el expuesto en [11], donde se recoge una tabla con los primeros 50 regresores que ilustra la necesidad de emplear GVG en el escenario propuesto.



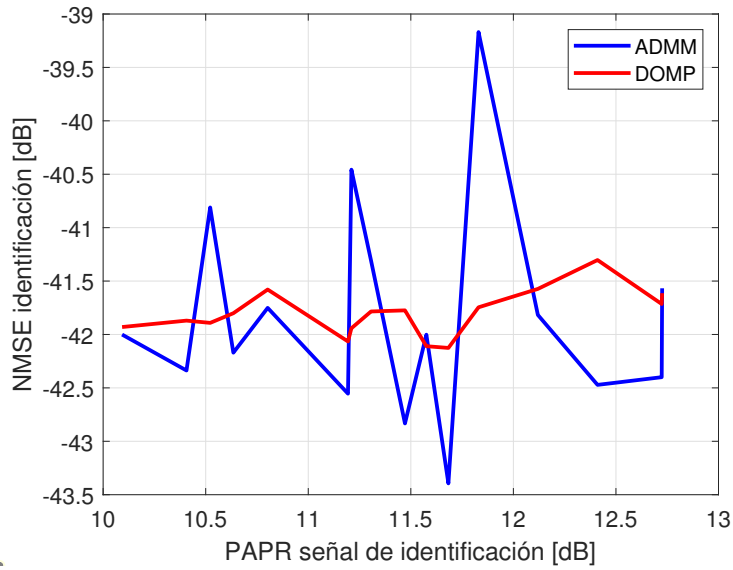
**Figura 4.3** Búsqueda del modelo mediante GVG (a) avance en las 125 iteraciones, (b) aporte al NMSE de cada regresor ordenado según DOMP, con el óptimo determinado por BIC ( $n_{opt}$ ).

Con esta configuración se plantea modelar 16 escenarios distintos, donde la característica diferenciadora es la PAPR. La comparativa se realiza entre el método propuesto y LS habiendo aplicado DOMP, a sabiendas de que DOMP proporciona una matriz mucho más adecuada en cuanto a número de regresores e información que aportan al modelo. Se incide en que, durante las pruebas, los resultados presentados con DOMP simbolizan la solución de mínimos cuadrados para la matriz acondicionada por DOMP.

Tras la ejecución de ADMM, la regularización incluida eliminó 4 regresores, trabajando con una matriz de 2321 columnas, frente a los 235 que mantiene DOMP. En la Figura 4.4 se muestran los distintos NMSE de identificación obtenidos para señales con PAPR variante entre 10.1 y 12.7 [dB], donde pueden discutirse dos aspectos. En primer lugar, el NMSE de identificación que proporciona ADMM se muestra más variante en función de la PAPR. Parece razonable que sea así, ya que la matriz con la que trabaja ADMM tiene más regresores y estos son más sensibles. Por otro lado, llama la atención la mejora de identificación en buena parte del barrido. Esto debe estar relacionado con el hecho de que, aunque la matriz de ADMM tenga regresores correlados, el número de condición de cada módulo es inferior al de la matriz completa, ya que por definición  $cond(\mathbf{R}) = \|\mathbf{R}^{-1}\| \|\mathbf{R}\|$  [10]. Por tanto, aunque los regresores no son óptimos, aportan información sin estropear demasiado la condición de la matriz a ser invertida, consiguiendo que esa información que aportan sea útil, al menos en modelado.

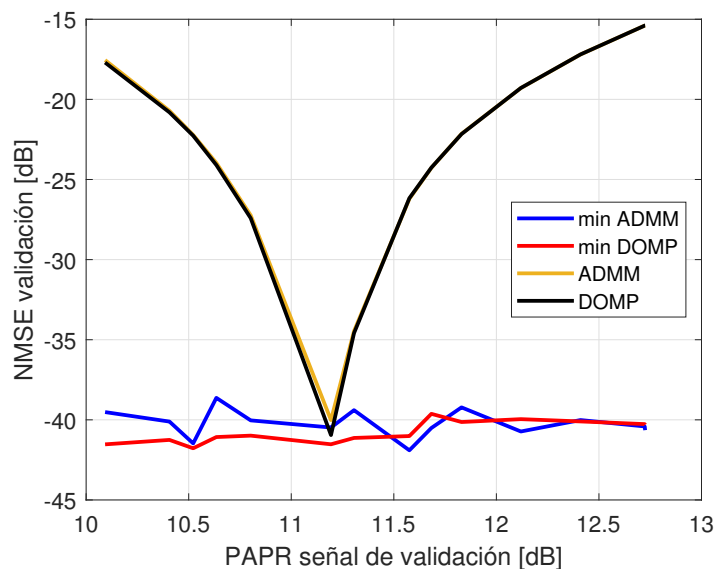
Una vez obtenidos los modelos, se plantea la validación. Para ello, se emplea otro set de 16 medidas, con un barrido en PAPR parecido al anterior y se pone a prueba cada modelo para cada nueva señal. En la Figura 4.7 se recoge el resultado obtenido con el modelo de la semilla 1000 frente a los mejores NMSE de





**Figura 4.4** NMSE de identificación de 16 modelos ordenados por PAPR, obtenidos para ADMM y para mínimos cuadrados tras la aplicación de DOMP.

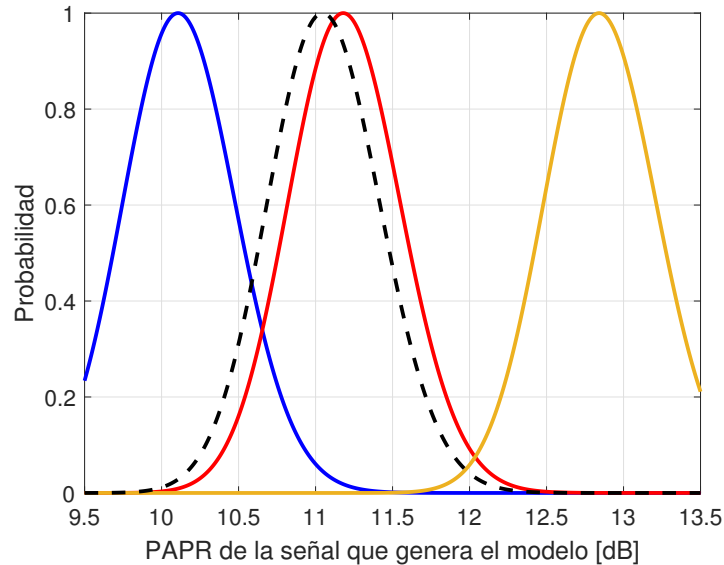
validación. Los mejores modelos para cada señal de validación eran los que provenían de un modelo con PAPR de señal de identificación más cercana. Además, la curva centrada en el modelo de semilla 1000 va en consonancia con las gráficas teóricas de probabilidad derivadas de la divergencia de Kullback-Leibler. Esto confirma que, al menos en modelado, el resultado obtenido en la regresión es muy sensible al cambio de la PEP, que se ha provocado por un barrido en PAPR ya que se ha mantenido fija la potencia promedio. Se ha observado el mismo efecto sobre el resto de modelos centrados en sus respectivas PAPR.



**Figura 4.5** NMSE de validación. Las curvas min ADMM y min DOMP representan los mejores resultados de NMSE, que se corresponden con modelos más cercanos a la PAPR que se valida. Las curvas ADMM y DOMP se corresponden con las validaciones obtenidas con el modelo obtenido a partir de la señal de semilla 1000.

#### 4.4.1 Ponderación de modelos

Se plantea una prueba adicional relativa al concepto de probabilidad que se ha extraído de la divergencia de Kullback-Leibler, y es la posibilidad de introducir una ponderación entre los modelos. Esto cobra un cierto sentido al observar lo sucedido en la Figura 4.7, donde la menor validación ha coincidido con la mayor probabilidad, que aunque no se ha computado para la variación de la PAPR en el modelo, se mostró teóricamente en la Sección 2.6.2 que esta característica de la señal marcará el parecido entre las distribuciones de probabilidad. Por tanto, podría entenderse que en la prueba anterior se ha realizado una elección de modelo en base a esta probabilidad. Ahora se propone una ponderación entre estos modelos, regida justamente por la probabilidad  $P(\bar{x}_q \sim p_X^d)$ .



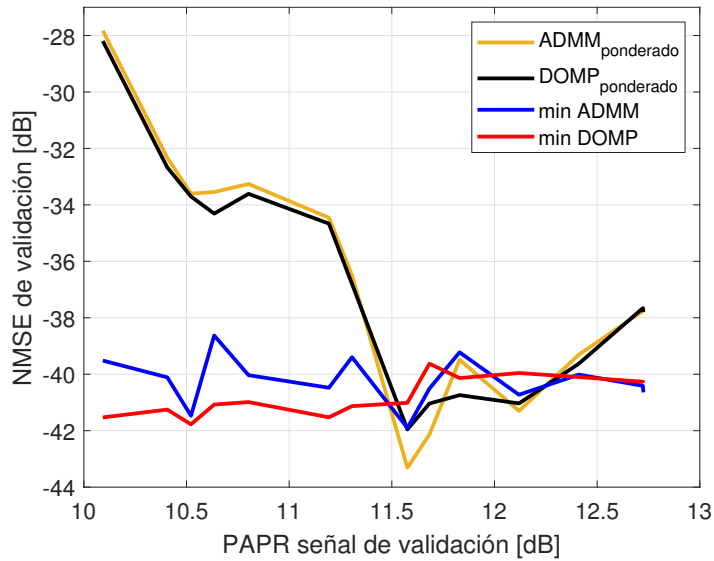
**Figura 4.6** Curvas de probabilidad computadas para tres de los modelos a color. La curva negra discontinua es la del modelo de Montecarlo.

Para ello, se propone partir de la curva de divergencia de Kullback-Leibler frente a PAPR que se expuso en la Sección 2.6.2, en la Figura 2.8. Puede observarse que tiene una cierta tendencia parabólica, de modo que para poder trabajar con una función a la que introducir nuevos valores de entrada se empleó la función *polyfit* de MATLAB®. Ajustando el polinomio a orden 4 se obtuvo un buen ajuste. Una vez se tuvo la expresión de la divergencia de Kullback-Leibler para el modelo de Montecarlo, se introduce una suposición por sencillez. Ésta consiste en que el comportamiento de la curva si se computa con un modelo de distinta PAPR ha de ser similar, de modo que con aplicar un desplazamiento debería ser suficiente. Así, se consigue centrar el mínimo de la divergencia en la PAPR del modelo que se evalúa. Siguiendo la ecuación (2.21) para computar la probabilidad se obtiene la Figura 4.6, con una curva de probabilidad para cada modelo. Nótese que se han escalado para tener probabilidades entre 0 y 1, y aprovechar así el máximo del rango.

Para generar el modelo conjunto, se introdujo la PAPR de la señal con la que se va a validar en las 16 funciones, obteniendo así 16 probabilidades. En la prueba anterior se eligió la máxima y se aplicó el modelo correspondiente. En este caso, se plantea una ponderación para cada modelo simplemente dividiendo entre la suma de todas las ponderaciones. De esta forma, se adquiere una media ponderada de los coeficientes  $\mathbf{h}$  de cada modelo, multiplicando la  $\mathbf{h}$  de cada modelo por su ponderación y sumándolas todas. Se obtienen así los resultados de NMSE de la Figura 4.7. A priori parece que la ponderación ha sido adecuada para algunos valores de PAPR, e incluso ha mejorado en algún caso.

## 4.5 Linealización

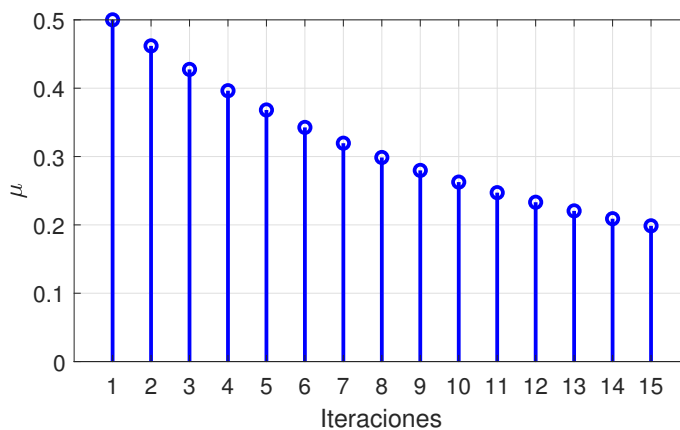
A continuación se ponen a prueba los conceptos en linealización, para ver hasta qué punto esta solución se adapta a la arquitectura y poder juzgar distintos aspectos sobre su rendimiento.



**Figura 4.7** NMSE de validación. Las curvas min ADMM y min DOMP representan los mejores resultados de NMSE, que se corresponden con modelos más cercanos a la PAPR que se valida. Las curvas  $ADMM_{ponderado}$  y  $DOMP_{ponderado}$  se corresponden con las validaciones obtenidas con el modelo obtenido a partir de la suma ponderada de los 16 modelos en base a la cercanía en PAPR.

#### 4.5.1 Implementación directa de ADMM en DLA

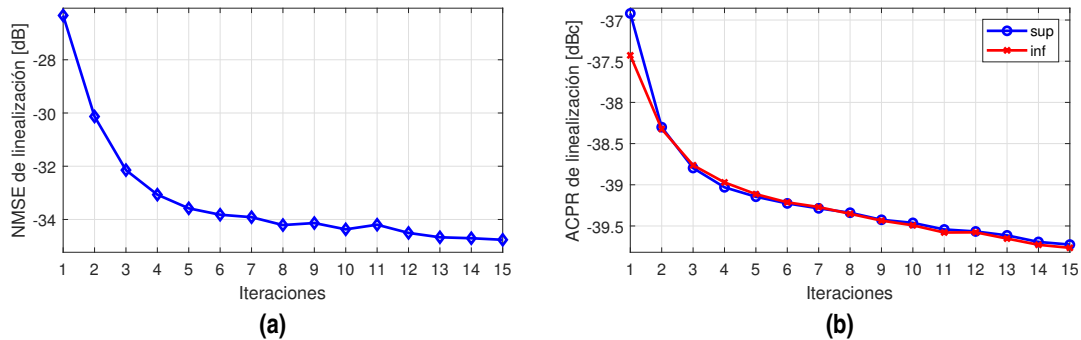
Siguiendo el procedimiento que se expuso en [37], sería esperable poder implementar la prueba de predistorsión directamente con los mismos  $\lambda$  y  $\rho$  que se obtuvieron en modelado. Lamentablemente se ha comprobado que los incrementos  $\mathbf{dw}$  son demasiado pequeños, quedando demasiados regresores por debajo del umbral mínimo. Además, la experiencia en modelado ha sugerido que el método necesita gran parte de los regresores para alcanzar una solución competitiva con DOMP. El cociente entre ambas variables controla la proporción de regresores que sobrevive al corte, pero no se pueden ajustar arbitrariamente libremente. Téngase en cuenta que la aparición de  $\rho$  en la operación de proyección durante el cálculo de  $\mathbf{u}^{i,(k+1)}$  ajusta la velocidad de convergencia del descenso del gradiente que lleva a cabo ADMM. Es por eso que se ha decidido ajustar  $\lambda$ , ya que solo aparece en la umbralización. Por tanto, se decidió hacer una primera reducción de  $\lambda$ , dándole el valor de  $\lambda = 10^{-10}$ . Se ejecutaron 15 iteraciones de entrenamiento con un factor de aprendizaje variante entre 0.5 y 0.2, con una caída exponencial, como se muestra en la Figura 4.8.



**Figura 4.8** Factor de aprendizaje empleado para la etapa de entrenamiento de DLA.

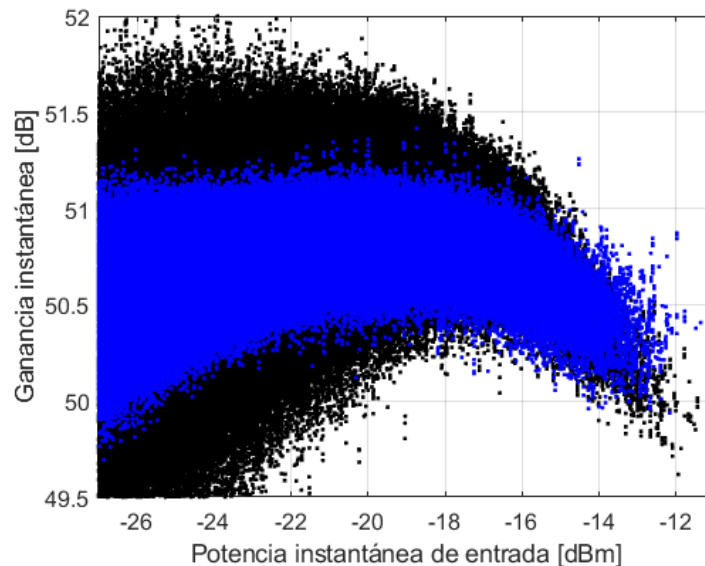
Se debe tener en cuenta que en cada iteración se ejecuta ADMM, de modo que el número de coeficientes supervivientes a la regularización de LASSO puede ir variando durante las iteraciones. En la última iteración,

quedó una matriz de 864 regresores, con la que se trabaja en la validación. En la Figura 4.9 (a) se muestra el avance del NMSE en las iteraciones del DLA, mientras que en Figura 4.9 (b), se observa el decrecimiento del ACPR.



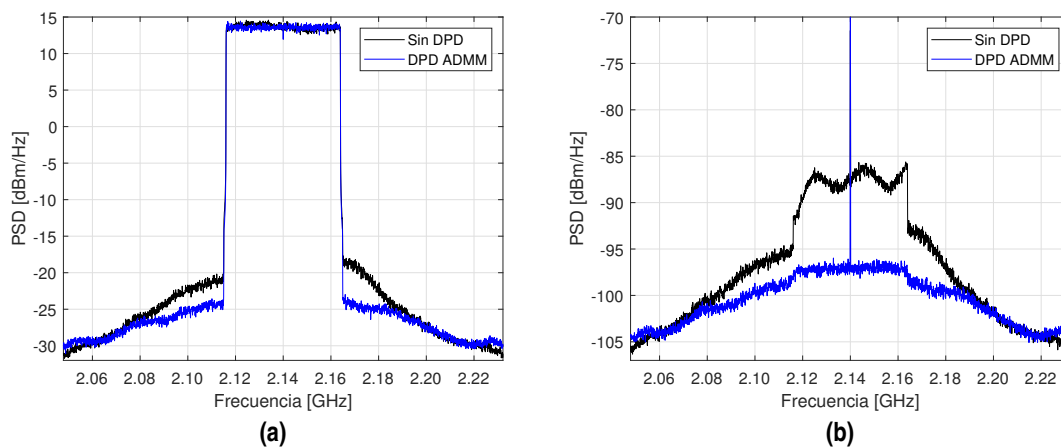
**Figura 4.9** Descenso del NMSE de linealización (a) y ACPR (b) durante la identificación en las 15 iteraciones que se llevaron a cabo de DLA.

Los resultados ciertamente quedan un poco lejos de lo deseable, dado que comparando con el reporte que se dió en [11] donde se trabaja con las mismas señales y se computan las dos técnicas, se consiguieron resultados 4 dB peores tanto de NMSE como de ACPR. Esto también se observa en la ganancia instantánea en la Figura 4.10 y el espectro en Figura 4.11 (a), donde en la primera se ve cómo, aunque se estabiliza ligeramente la ganancia no llega a tener comportamiento constante, mientras que en el espectro se ve cómo todavía se debería reducir más la emisión fuera de banda. También se muestra el espectro del error en Figura 4.11 (b), donde se entiende que aún se podría reducir este error.



**Figura 4.10** Ganancia instantánea frente a potencia instantánea de entrada. Aunque se ha corregido un poco el comportamiento aún queda compresión en la zona de alta potencia de entrada.

Estos resultados sugieren que, aunque se haya incluido la regularización, el problema numérico intrínseco es bastante dominante haciendo que no se llegue a una solución acorde con lo observado en modelado. Es muy probable que el motivo de este empeoramiento resida en la estructura de la propia arquitectura DLA, ya que se calculan incrementos de los pesos que se incorporan a la señal  $\mathbf{x}$  mediante restas. En cualquier caso, cabe destacar que la prueba es muy exigente, ya que la resolución mediante LS tras haber aplicado DOMP sobre la matriz de regresores parte de un punto numéricamente mucho más favorable al caso que se acaba de



**Figura 4.11** Comparativa antes y después de hacer DPD del espectro (a) y del espectro del error (b).

plantear, ya que el número de condición de la nueva matriz tras aplicar DOMP es mucho menor, como se indica en [23].

#### 4.5.2 Reajustes para alcanzar la linealización

En los resultados de la Sección 4.5.1 se ha visto que, pese a ser una buena herramienta de modelado, aplicar ADMM únicamente a la matriz de regresores completa no parece alcanzar la solución que se espera, y que seguir el procedimiento expuesto en [37] no parece dar resultado. Es por esto que se decide aplicar un par de cambios, incorporando una selección de coeficientes previa y recalculando los parámetros  $\lambda$  y  $\rho$  antes de ejecutar el método.

Por un lado, cabe destacar que, realmente no se requiere de una ejecución del algoritmo DOMP, puesto que el GVG ya devuelve los regresores ordenados en base a este criterio e incluso aporta el número óptimo de regresores en base a BIC. Por otro lado, este reajuste permite plantear una comparación justa entre la resolución por LS y ADMM, ya que ambos métodos tendrían exactamente el mismo punto de partida. Es decir, anteriormente se ha contrastado tanto el método de obtención de solución, ADMM contra LS, como la precisión tras la selección de coeficientes, esto es, regularización de LASSO contra DOMP. Como se comentó en la Sección 3.4 el podado de regresores realizado mediante DOMP es más inteligente que umbralizar respecto a la varianza del propio regresor. Por esto ahora, aunque se mantiene LASSO en ADMM, se facilita el punto de partida por la selección de coeficientes en ambos casos, dejando partir desde el mismo problema a las resoluciones mediante ADMM y LS. Por tanto, se va a partir de la matriz óptima de 235 regresores y se van a aplicar ambos métodos para poder comparar el rendimiento de ADMM con la solución de menor error.

Se realizaron 15 iteraciones de la arquitectura de aprendizaje directo con el factor de aprendizaje de la Figura 4.8. Además, como se ha apuntado al comienzo de esta sección, se busca el óptimo de  $\lambda$  y  $\rho$  previo al cálculo de  $\mathbf{dw}$ , haciendo uso de la función *fminsearch* de MATLAB<sup>®</sup>. En la Figura 4.12 se ve cómo han ido variando estos parámetros en cada iteración, hecho que confirma que era necesario recalcularlos. Resulta llamativo el comportamiento de los parámetros, ya que se alternan etapas donde se ubica un umbral muy alto, esto es  $\lambda \gg \rho$  donde se rechazan muchos regresores, con etapas con umbral bajo  $\lambda < \rho$  donde se emplea gran parte de los coeficientes. Esto revela que, de una iteración de DLA a la siguiente, la función a minimizar en la etapa de cálculo puede cambiar sustancialmente, alternándose etapas donde conviene eliminar regresores con etapas donde se usa la matriz completa.

A continuación, se expone la comparativa en rendimiento de ambos métodos. En primer lugar, puede verse en la Figura 4.13 el avance del NMSE de linealización y del ACPR en las sucesivas iteraciones de DLA. Es interesante matizar que en torno a la octava iteración ya se alcanza una buena solución. Además, ADMM queda unos 2 dB peor en prestaciones, dato que resulta bastante favorable, teniendo en cuenta que en la última iteración únicamente sobreviven 162 regresores por la aplicación de la regularización de LASSO, esto es, se trabaja con un 68.9% de los regresores óptimos.

Por otro lado, las gráficas de ganancia instantánea de la Figura 4.14, espectro de la Figura 4.15 (a) y espectro del error de la Figura 4.15 (b), confirman que con ambos métodos se alcanza la linealización.

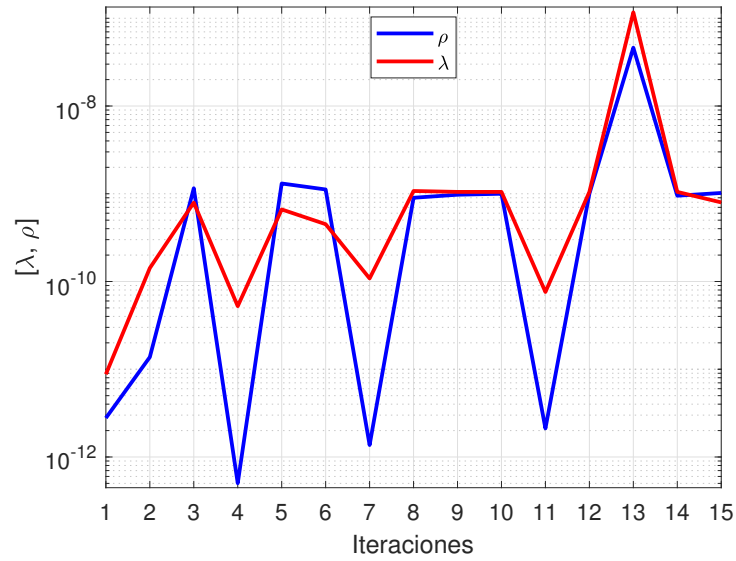


Figura 4.12 Variación de  $\lambda$  y  $\rho$  para cada iteración de DLA.

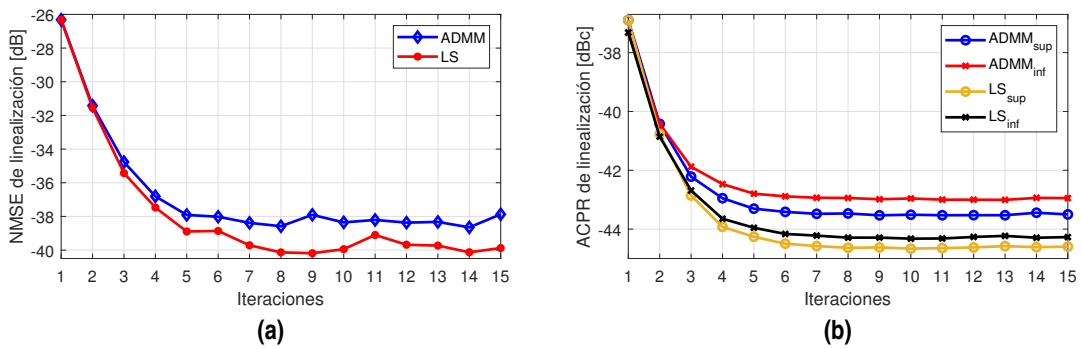


Figura 4.13 Comparativa del descenso del NMSE de linealización (a) y ACPR (b) para ambos métodos durante la identificación en las 15 iteraciones que se llevaron a cabo de DLA.

En este punto en el que se ve que ambos métodos son capaces de linealizar, se evalúa el concepto de generalización de la solución, planteado en la Sección 2.6 y evaluado para modelado en la Sección 4.4. Para comprobarlo, se realiza el mismo experimento, validando el sistema para el mismo set de 16 señales con barrido de PAPR entre 10.1 dB y 12.7 dB. Al hacer la sustracción de la señal de distorsión a la señal generada, generalmente se obtiene una PAPR algo superior, y dado que el RF WebLab tiene un sistema de protección para PAPR mayores de 14 dB se aplicó *clipping* solo si la señal resultante excedía este límite.

Quizás, los resultados mostrados en la Figura 4.16 parecen un poco contradictorios con lo obtenido al modelar, aunque ciertamente es positivo comprobar que con distintas señales de entrada se mantiene el rendimiento, dejando ver que el modelo de distorsión obtenido para la señal de semilla aleatoria 1000, es válido para el resto de valores de PAPR. Esto hace brillar una diferencia importante entre modelar y linealizar, que se debe tener en cuenta para entender ambos procesos. Cuando se modela, se pretende encontrar la relación entrada-salida del PA, lo que implica que se modela y como función de la combinación de las distintas componentes del espacio de regresores  $Col(\mathbf{X})$  ponderadas por el vector de coeficientes  $\mathbf{h}$ . Es natural que el modelo sea muy sensible a  $\mathbf{x}$  puesto que  $\mathbf{y} = f(\mathbf{x})$ . De manera que  $\mathbf{x}$  constituye tanto el espacio de representación de la señal como la propia entrada al sistema, que en la zona lineal es de la forma  $\mathbf{y} = G\mathbf{x}$ . Cuanto mayor es la excursión de  $\mathbf{x}$  por la zona de saturación del PA más lejos se estará de tener la relación  $\mathbf{y} = G\mathbf{x}$ . Por tanto, cuando se modela se trata de encontrar aquellas funciones base que mejor representen el comportamiento del PA. Esta estructura será la que se use en la etapa de DPD, pero el problema a resolver es distinto, ya que el objetivo ahora es estimar la señal de distorsión  $\mathbf{d}$  que hace que el comportamiento de la

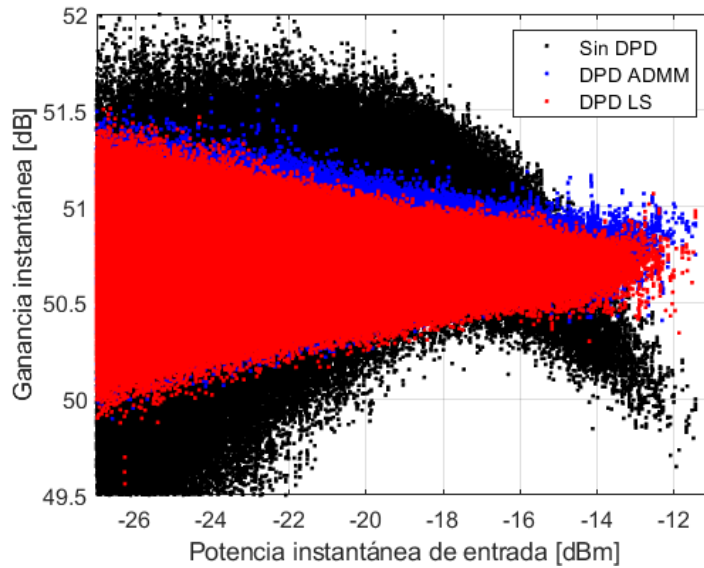


Figura 4.14 Ganancia instantánea frente a potencia instantánea de entrada para ambos métodos.

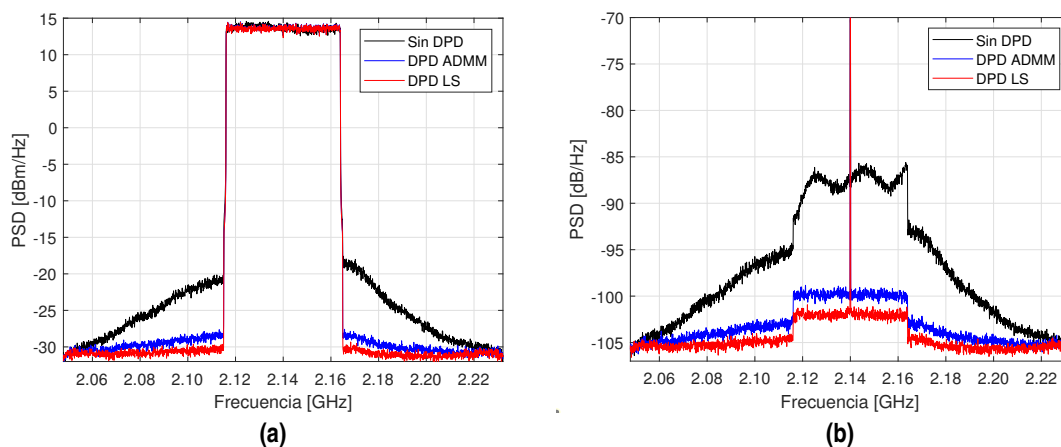
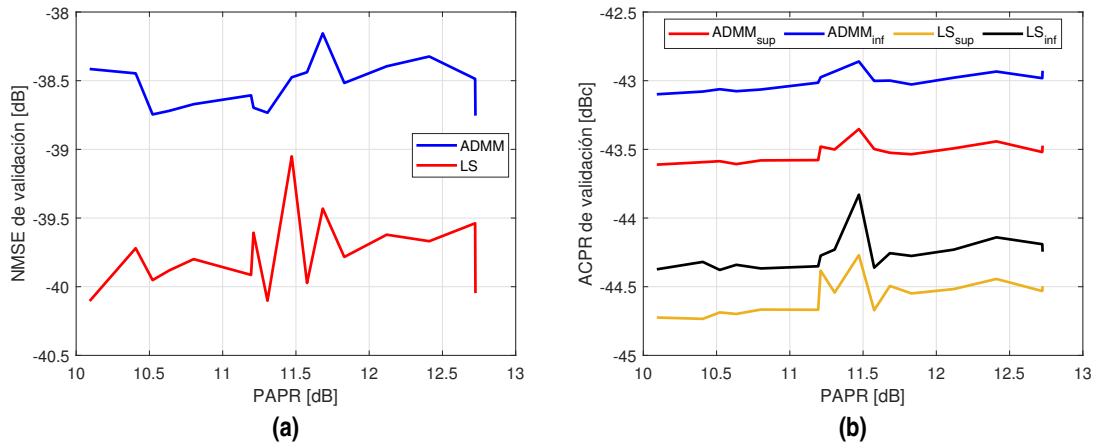


Figura 4.15 Comparativa antes y después de hacer DPD del espectro (a) y del espectro del error (b) para ambos métodos.

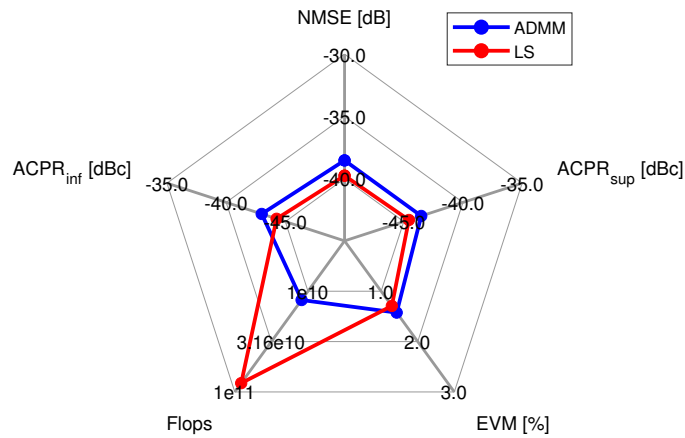
salida sea lineal. Es decir, se trata de encontrar la señal  $\mathbf{d}$  que hay que restar a la entrada  $\mathbf{u}$  para linealizar la salida  $\mathbf{y}$ . Aunque  $\mathbf{d}$  no es la distorsión del PA, sino una señal no lineal con efectos de memoria, estimarla adecuadamente dependerá, entre otras cosas, del punto de operación en el que se esté haciendo trabajar al amplificador. Esto es debido a que la distorsión en la salida  $\mathbf{y}$  se manifiesta en mayor o menor medida en función de la excursión de  $\mathbf{x}$  por la zona de saturación, observada en las Figuras 4.10 y 4.14. Por tanto, para que los pesos en  $\mathbf{d} = \mathbf{U}\mathbf{w}$  sean aplicables para predistorsionar otras señales OFDM de entrada, lo fundamental es que se hayan calculado en un caso en el que la excursión de  $\mathbf{x}$  por la zona de saturación sea suficiente como para generar distorsión a la salida. Además, aunque el PA es un sistema con memoria, se espera que no cambie demasiado de una señal a otra, ya que el comportamiento en el tiempo de una OFDM es estadísticamente similar para distintas realizaciones. Por tanto, se entiende que el motivo del comportamiento observado al validar en la Figura 4.16 (a) reside en que los pesos  $\mathbf{w}$  que generan la señal de distorsión  $\mathbf{d} = \mathbf{U}\mathbf{w}$ , que se aplica a la entrada del PA restándose a la señal original, son más extrapolables de una OFDM a otra, que la salida esperada cuando las señales de entrada tienen diferentes PAPR. Aunque es importante destacar que las prestaciones se mantienen más o menos estables para distintas PAPR, no es así para distintas potencias promedio  $P_{\text{avg}}$ . Aunque igualmente cambia la PEP, mover la  $P_{\text{avg}}$  genera un escenario con muchos más cambios que los introducidos al variar la PAPR. En cualquier caso y a la vista de los resultados, ambos



**Figura 4.16** NMSE de validación (a) y ACPR de validación (b) para ambos métodos de señales con distinta PAPR.

métodos consiguen linealizar el sistema para todo el rango de PAPR dentro del mismo punto de operación.

Por último, la Figura 4.17 muestra una comparativa de rendimiento medio alcanzado en validación, que se recoge numéricamente en la Tabla 4.1. Se destaca que el número de operaciones en el caso de ADMM se ha calculado para las 20 iteraciones, donde al igual que en el caso de modelado, se han implementado 64 módulos computacionales.



**Figura 4.17** Comparativa entre ADMM y LS en validación, incluyendo el número de operaciones necesarias para la identificación, que en el caso de ADMM se corresponden con 20 iteraciones del método.

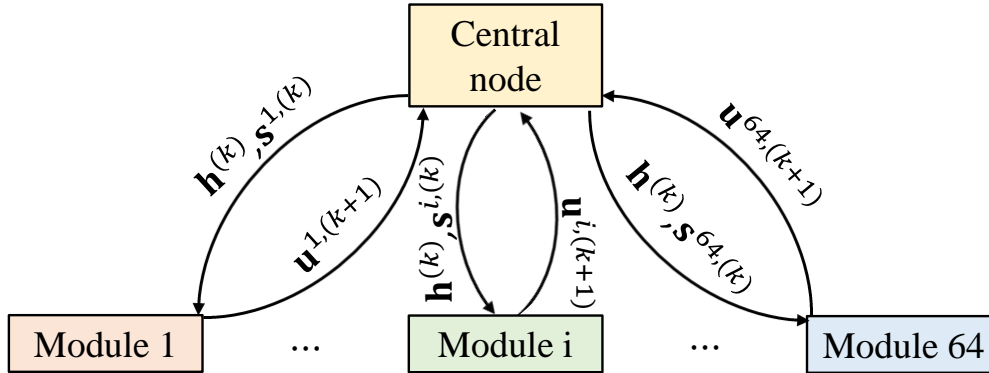
**Tabla 4.1** Resultados medios de linealización para una potencia promedio de salida de 29.5 dBm con señales 5G-NR de 50 MHz cuya PAPR varía entre 10.1 y 12.7 dB.

	NMSE	ACPR <sub>inf</sub>	ACPR <sub>sup</sub>	EVM	Nº operaciones
Unidades	[dB]	[dBc]	[dBc]	[%]	Flops
Sin DPD	-26.3	-37.3	-36.9	4.8	-
DPD DOMP+ADMM	-38.5	-43.0	-43.5	1.4	$1.2 \cdot 10^{10}$
DPD DOMP+LS	-39.7	-44.3	-44.6	1.2	$8.2 \cdot 10^{10}$

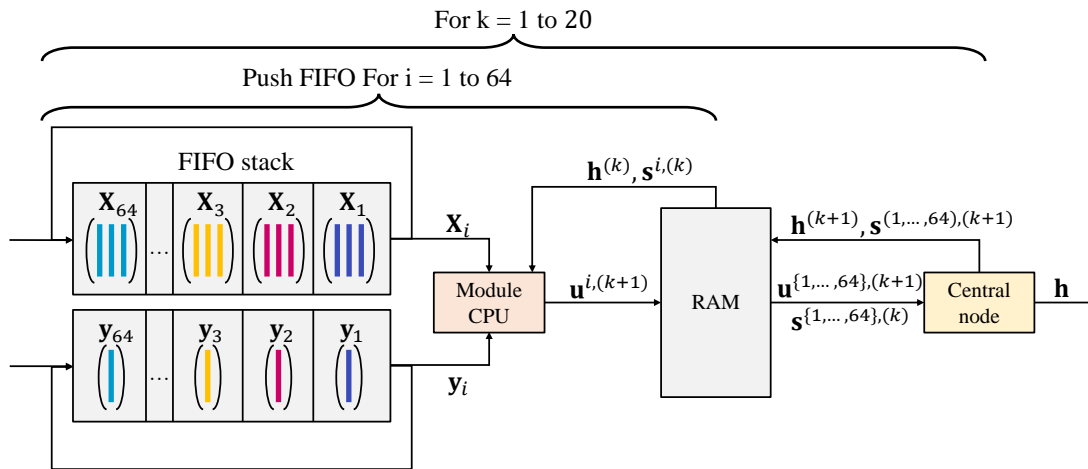


### 4.6 Ventajas de la reducción del coste computacional

Aunque la Tabla 4.1 recoge el número de operaciones al realizar 20 iteraciones de ADMM, es importante destacar que, en cada iteración, se requieren únicamente  $6.1 \cdot 10^8$  flops. Si se compara con las  $8.2 \cdot 10^{10}$  operaciones que realiza LS, se ve que ADMM requiere únicamente de un dispositivo que sea capaz de llevar a cabo el 0.75% de las operaciones que se llevan a cabo con la matriz completa. Esto, combinado con el hecho de que la matriz que maneja cada agente es 64 veces más pequeña para esta implementación, no solo permite una solución distribuida, sino que también habilita la resolución de DPD en dispositivos con menos prestaciones. Es decir, se podría alcanzar una solución cooperativa como la ejemplificada para el modelado en la Figura 4.18 o una secuencial como en la Figura 4.19.



**Figura 4.18** Implementación de ADMM en paralelo. El número de operaciones a realizar en un instante determinado por cada módulo es el 0.75% de las que implica hacer LS.



**Figura 4.19** Implementación de ADMM en paralelo. El número de operaciones a realizar en un instante determinado por el módulo de CPU es el 0.75% de las que implica hacer LS.

Ambas arquitecturas tienen sus ventajas y sus inconvenientes. Por ejemplo, la solución en paralelo es más rápida pero requiere de un mayor número de dispositivos, mientras que la secuencial es más lenta pero se puede realizar con un solo dispositivo. En ambos casos, la exactitud de las soluciones es la misma. La elección de la arquitectura dependerá de las exigencias de cada caso, pero deja de manifiesto que la habilitación de un enfoque distribuido aporta una flexibilidad en cuanto a la implementación que no son capaces de proporcionar soluciones como LS.



## 5 Conclusiones y líneas futuras

---

*El análisis lógico es la primera operación que debiera emprenderse al comprobar las hipótesis científicas, sean fácticas o no.*

MARIO BUNGE, 1959

En este trabajo se han ido presentando distintos aspectos relativos al modelado y la linealización de amplificadores de potencia. Con el fin de explorar nuevas soluciones que habiliten el procesado distribuido dentro de este campo, se ha pasado por los conceptos fundamentales del modelado, tratando de poner a prueba aspectos relativos a la capacidad de generalizar de los modelos, dejando así una intuición de lo que se espera obtener en la práctica. Además, se ha tratado de conducir la solución de mínimos cuadrados, que sigue un enfoque centralizado, hacia una solución implementable en una arquitectura distribuida. Cada paso hacia la solución final se ha dado en base a los razonamientos típicos dentro de este campo, adaptando la formulación típica de los problemas de optimización con restricciones al caso de estudio. Aunque la implementación teórica avanzaba la posibilidad de aplicar ADMM tanto en modelado como en linealización, los resultados experimentales han corroborado este hecho. Los resultados obtenidos en modelado revelan ciertos aspectos. El primero de ellos es que, pese a tratar con la matriz completa, la aplicación de ADMM se ha mostrado factible en este caso, obteniendo valores de NMSE en identificación y validación que compiten con una solución altamente competitiva y empleada en otros trabajos. Además, se ha corroborado la dependencia encontrada teóricamente entre la PAPR y la divergencia de Kullback-Leibler en el caso del modelado, donde las mejores prestaciones se han alcanzado cuando la señal de validación tenía una PAPR cercana a la de la señal que generó el modelo. No solo se ha probado la elección del mejor modelo en base a esta métrica, sino que también se ha propuesto una ponderación entre modelos, regida por la propia divergencia de Kullback-Leibler. Se ha comprobado que en algunos casos se han obtenido mejores prestaciones al ponderar los modelos, hecho que sugiere que sería adecuado afinar los pesos de cada modelo, quizás mediante el escalado de la probabilidad, siendo más exclusivo o no en función de la región de PAPR con la que se trabaje.

Aunque el modelado es interesante para conocer el comportamiento de los amplificadores de potencia, el resultado más relevante es la linealización. En la prueba de la Sección 4.5.1, se ha podido comprobar que en el caso planteado no es correcto suponer que los parámetros de ADMM encontrados durante el modelado sean válidos para linealización. Así, en la Sección 4.5.2, se ha comprobado que el método es válido para linealizar cuando se reajustan sus parámetros. Sería conveniente averiguar alguna relación determinista entre estos parámetros y las características de la señal de entrada, para ajustarlos sin necesidad de computar una minimización. También se observa que se alcanza la linealización en torno a la octava iteración. El NMSE de linealización de ADMM es aproximadamente 1.5 dB peor que el de LS, pero a cambio, ADMM sólo necesita un dispositivo que compute el 0.75% de las operaciones que se llevan a cabo en LS. De manera que, además de aportar flexibilidad en cuanto a la arquitectura de implementación, habilita la computación en el borde y la adaptabilidad a dispositivos más reducidos. Por tanto, el principal reto en un futuro cercano, consiste en tratar de ajustar esta solución a un hardware específico, e ir perfeccionando aspectos como los parámetros del algoritmo, para evitar realizar una optimización en cada iteración del bucle DLA.



# Apéndice A

## Código de MATLAB® empleado en el trabajo

---

La realización de los distintos experimentos presentados durante el trabajo conlleva bastante código. Aquí se recoge únicamente la implementación de ADMM, debido a su especial interés.

### Código A.1 Código de implementación de ADMM en paralelo.

```
function [h, nmse] = parADMMrwd_ponderado(X,y,rho,lambda,N_row_sections,pond)

[N,M] = size(X); %N: nº de muestras, M: nº de regresores
n_workers = 8;
N_row_sections = N_row_sections/n_workers; % nº de secciones por worker

%Matriz de indexación
for i = 1:N_row_sections
    Si(i,:) = 1+(i-1)*(N/n_workers)/N_row_sections:i*(N/n_workers)/
        N_row_sections;
end

Xc = Composite();
yc = Composite();
sc = Composite();
s = zeros(M,N_row_sections*n_workers);

%Reparto de las variables a los agentes
for i = 1:n_workers
    Xc{i} = X(1+(i-1)*N/n_workers:i*N/n_workers,:);
    yc{i} = y(1+(i-1)*N/n_workers:i*N/n_workers);
    sc{i} = s(:,1+(i-1)*N_row_sections:i*N_row_sections);
end

%Inicialización de las variables
h = zeros(M,1);
s_mean = zeros(M,1);

%Bucle del método
for k = 1:20
    %Cálculo de  $u^{i,(k+1)}$  por cada agente. Cada worker gestiona un grupo de
    agentes
    spmd
        for i = 1:N_row_sections
            uc(:,i) = inv(Xc(Si(i,:),:))*Xc(Si(i,:),:) + rho*eye(M)*...
                (Xc(Si(i,:),:))*yc(Si(i,:))+rho*(h - sc(:,i));
        end
    end
end
```

```

end

%Recopilación en el nodo central
u = [];
s = [];
for i = 1:n_workers
    eval("u = [u, uc{i}];");
    eval("s = [s, sc{i}];");
end

%Solución y Soft thresholding operation:
u_mean = squeeze(mean(u,2));
h = wthresh(u_mean + s_mean,'s',lambda/(N_row_sections*n_workers*rho));

%Actualización de la variable dual escalada:
s = s + u - repmat(h,1,N_row_sections*n_workers);

%Envío de  $u^{i,(k+1)}$  y  $s^{i,(k+1)}$  a los agentes
for i = 1:n_workers
    eval("sc{i} = s(:,1+(i-1)*N_row_sections:i*N_row_sections);");
    eval("uc{i} = u(:,1+(i-1)*N_row_sections:i*N_row_sections);");
end

%Media de s para la siguiente iteración
s_mean = squeeze(mean(s,2));
ymod = X*h;
nmse(k) = 20*log10(norm(ymod-y)/norm(y));
end

```

# Apéndice B

## Artículos de investigación

---

Este Trabajo Fin de Máster ha sido la culminación de dos años de trabajo, de la mano del grupo Grupo de Sistemas de Radiocomunicación del Departamento de Teoría de la Señal y Comunicaciones. Durante este tiempo, se han enviado y aceptado tres contribuciones a las ediciones XXXVII y XXXVIII del Simposio Nacional de la URSI, de las que he sido autor y coautor. Éstas se recogen en este apéndice para que sirvan de apoyo al lector.

# Modeling and Linearization of Wideband Power Amplifiers with Alternating Direction Method of Multipliers (ADMM)

Elías Marqués-Valderrama, María J. Madero-Ayora, Juan A. Becerra, Miguel Nogales  
 elimarval@alum.us.es, mjmadero@us.es, jabecerra@us.es, mignoggon@alum.us.es,

Dpto. de Teoría de la Señal y Comunicaciones. Universidad de Sevilla. Camino de los descubrimientos s/n. 41092. Sevilla

**Abstract**—This work provides a different solving method for modeling and linearization of power amplifiers (PA), applying the Alternating Direction Method of Multipliers (ADMM), and exploring its benefits. The ability of providing distributed computation is currently becoming a need, mainly on real-time scenarios where the resources are limited. Moreover, a large amount of regressors are typically needed to explain the behavior of the PA, presenting numerical issues at the resolution of the inverse when the regression is performed. The ADMM is also a compressing sensing (CS) technique, that ensures a solution since it involves a norm  $\ell_1$  regularization, translated into an underlying coefficient selection. Experimental results achieved with this approach are provided for a class AB power amplifier driven by 30 MHz 5G New Radio signals.

## I. INTRODUCTION

In the last years, the world of mobile communications has been suffering an important amount of changes. The demand in terms of binary rate and latency is increasing, and impulsed Fifth Generation (5G) to opt for orthogonal frequency division multiplexing (OFDM) waveforms with flexible numerology, which present a better spectrum efficiency. This kind of signals have the problem of exhibiting a high peak-to-average power ratio (PAPR), incurring in distortion issues after passing through a power amplifier (PA) when the operation point is set near to the saturation zone, where the device presents a better power efficiency. There are several techniques that deal with this problem, being digital predistortion (DPD) schemes, which allow real-time solutions, commonly used. Most of these approaches require an underlying model, where the Volterra-based models are widely extended on the literature due to their capability for representing physical behavior of PAs.

Furthermore, to describe this reality in mathematical formulation with the aim of understanding the behavior of the PA (modeling) as well as applying this comprehension of the device to avoid the distortion (DPD), it is necessary to solve equations that typically present numerical issues, causing instabilities in regression and yielding very sensitive solutions to variations in the input, or that are not even able to converge. This is one of the reasons behind possible modifications of the strategy in regression problems, going beyond least squares (LS) resolution, that enforces the minimal error estimator, to find more stable solutions. This problem has been studied in several works as [1] or [2], where the goal was to enhance the regression taking care of the conditioning of the algebraic operations. Following this line, [3] was presented, delving into the benefits of performing Ridge regularization instead

of LS. Further developing these approaches, this work tries to achieve a wider point of view about the meaning of imposing restrictions on these optimization problems to be able to design customized implementations in modeling and linearization. For this proposal, a different approach for solving the measurement equation will be presented in this work, taking advantage of the benefits derived from implementing a sparse solution based on Alternating Direction Method of Multipliers (ADMM).

This work is organized as follows. Section II contains the mathematical formulation, Section III extends the ADMM solution to modeling and DPD problems, Section IV shows experimental results and Section V wraps up the work with the conclusions.

## II. ADMM MODEL EQUATIONS FORMULATION

In this section, the modeling problem will be driven from the classical approach based on LS solution, to the distributed method allowed by ADMM technique, following the reasoning shown in [4]. Firstly, the measurement equation is

$$\mathbf{y} = \mathbf{X}\mathbf{h} + \mathbf{e}, \quad (1)$$

where  $\mathbf{y} = [y[q], y[q-1], \dots, y[q-m+1]]^T \in \mathbb{C}^{m \times 1}$  is the discrete set of output samples, represented by a linear weighted sum of the columns of the regressor matrix  $\mathbf{X} = [\phi_1, \phi_2, \dots, \phi_n] \in \mathbb{C}^{m \times n}$  governed by the Volterra model coefficients vector  $\mathbf{h} \in \mathbb{C}^{n \times 1}$  and quantifying the modeling errors in  $\mathbf{e} \in \mathbb{C}^{m \times 1}$ . LS method proposes to solve (1) by minimizing the residual sum of squares (RSS)

$$RSS = \|\mathbf{y} - \mathbf{X}\mathbf{h}\|_2^2 = \|\mathbf{y} - \hat{\mathbf{y}}\|_2^2, \quad (2)$$

where  $\hat{\mathbf{y}}$  represents the estimation of  $\mathbf{y}$ , yielding the solution

$$\hat{\mathbf{h}} = (\mathbf{X}^H \mathbf{X})^{-1} \mathbf{X}^H \mathbf{y} = \mathbf{X}^\dagger \mathbf{y}. \quad (3)$$

This equation can be understood as the projection of  $\mathbf{y}$  over the Volterra subspace represented by  $\mathbf{X}$  through the Moore-Penrose pseudo-inverse  $\mathbf{X}^\dagger$ .

Although the optimization of (2) reaches the minimal error solution, the ill-conditioning of the regressor matrix generally causes instabilities on the regression, yielding a model that is very sensitive to small variations on the input  $\mathbf{X}$ . Even sometimes, the problem is irresolvable due the numerical issues in  $(\mathbf{X}^H \mathbf{X})^{-1}$  operation. To avoid this behavior, a more flexible solution of the problem is sought, modifying (2),



leading on a multivariable Lagrangian optimization problem described as

$$\text{minimize } f_1(\mathbf{u}) + f_2(\mathbf{h}) \quad \text{s.t. } \mathbf{P}\mathbf{u} + \mathbf{Q}\mathbf{h} = \mathbf{c}, \quad (4)$$

where  $\mathbf{P} \in \mathbb{C}^{p \times n}$  and  $\mathbf{Q} \in \mathbb{C}^{p \times q}$  are known matrices, together with  $\mathbf{c} \in \mathbb{C}^{p \times p}$ , that fixes the constraint. Applying the multipliers method, the augmented Lagrangian form may be written as

$$L_p(\mathbf{u}, \mathbf{h}, \mathbf{d}) = f_1(\mathbf{u}) + f_2(\mathbf{h}) + \mathbf{d}^T (\mathbf{P}\mathbf{u} + \mathbf{Q}\mathbf{h} - \mathbf{c}) + \frac{\rho}{2} \|\mathbf{P}\mathbf{u} + \mathbf{Q}\mathbf{h} - \mathbf{c}\|_2^2 \quad (5)$$

being the Lagrangian multiplier represented in  $\mathbf{d}^T \in \mathbb{C}^{p \times p}$ , also called the dual variable, and  $\rho > 0$  is the augmented parameter. Through an algebraic transformation and defining  $\mathbf{s} = \mathbf{d}/\rho$  as the scaled dual variable, a general iterative ADMM algorithm can be described as:

$$\mathbf{u}^{(k+1)} := \underset{\mathbf{u}}{\text{argmin}} L_p(\mathbf{u}, \mathbf{h}^{(k)}, \mathbf{s}^{(k)}), \quad (6a)$$

$$\mathbf{h}^{(k+1)} := \underset{\mathbf{h}}{\text{argmin}} L_p(\mathbf{u}^{(k+1)}, \mathbf{h}, \mathbf{s}^{(k)}), \quad (6b)$$

$$\mathbf{s}^{(k+1)} := \mathbf{s}^{(k)} + (\mathbf{P}\mathbf{u}^{(k+1)} + \mathbf{Q}\mathbf{h}^{(k+1)} - \mathbf{c}). \quad (6c)$$

At this point, the meaning of ADMM acronym makes sense given that the alternation on the direction is visible in each iteration. It can be seen that, the first step (6a), minimizes the function in the direction of  $\mathbf{u}$  (fixing  $\mathbf{h}^{(k)}$  and  $\mathbf{s}^{(k)}$ ); and in the second step (6b), the operation is focused on the optimization in the  $\mathbf{h}$  direction (fixing the new vector  $\mathbf{u}^{(k+1)}$  calculated in (6a) and  $\mathbf{s}^{(k)}$ ). Finally, the Lagrange multipliers  $\mathbf{s}$  are updated in (6c).

There are a large amount of methods that can be shaped as (4). For example, the Ridge regularization—or norm  $\ell_2$  regularization—is a particularization where  $f_1 = \|\mathbf{y} - \mathbf{X}\mathbf{u}\|_2^2$  and  $f_2 = \lambda\|\mathbf{h}\|_2^2$  that does not need a Lagrangian augmentation, presenting a closed solution. Conversely, the Lasso regularization—or  $\ell_1$  regularization—, where  $f_2 = \lambda\|\mathbf{h}\|_1$ , does not present a closed form at the resolution of the optimization problem. This way, making the assumption that (1) allows a sparse solution, it seems reasonable to formulate the particular ADMM form for Lasso regularization as:

$$\begin{aligned} \text{minimize } & \frac{1}{2} \|\mathbf{y} - \mathbf{X}\mathbf{u}\|_2^2 + \lambda\|\mathbf{h}\|_1 \\ \text{s.t. } & \mathbf{u} - \mathbf{h} = \mathbf{0}, \end{aligned} \quad (7)$$

remaining a constraint  $\mathbf{P} = \mathbf{I}$ ,  $\mathbf{Q} = -\mathbf{I}$  and  $\mathbf{c} = \mathbf{0}$ , that forces  $\mathbf{h}$  to be equal to  $\mathbf{u}$ . Notice that in the Ridge case it is also  $\mathbf{u} = \mathbf{h}$ .

Once the problem has been approached to ADMM, some considerations should be regarded. In the first place, the implementation of Lasso regularization avoids several numerical issues in the calculation of the model. In the second place, the formulation of the problem such as (7) allows to solve it in a distributed fashion. Furthermore, to exploit the advantages of this point of view, a distributed resolution as a consensus-based ADMM carrying out a row wise division of the matrix  $\mathbf{X}$  is presented. For this proposal, it is convenient to rewrite (7) as

$$\begin{aligned} \text{minimize } & \frac{1}{2} \|\mathbf{y}_i - \mathbf{X}_i \mathbf{u}^i\|_2^2 + \lambda\|\mathbf{h}\|_1 \\ \text{s.t. } & \mathbf{u}^i - \mathbf{h} = \mathbf{0}, \quad \forall i \in \{1, \dots, M\}, \end{aligned} \quad (8)$$

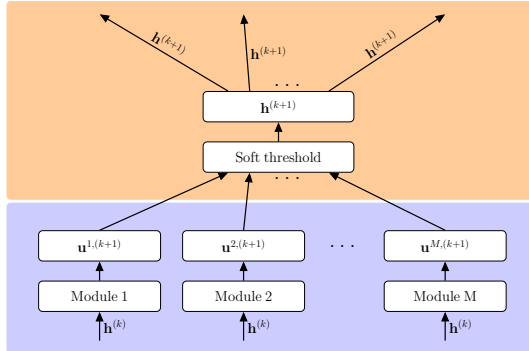


Fig. 1: Representation of one iteration of ADMM. Each "Module  $i$ " computes (10a) in parallel. Obtaining the dual augmented variable has been omitted in this representation.

where the subscript  $i$  represents the selection of the  $i$ -th segment of  $\mathbf{y}$  and  $\mathbf{X}$ , and the superscript  $i$  represents the  $i$ -th replica, enabling their independent resolution. Developing the augmented Lagrangian as

$$\begin{aligned} L_p(\mathbf{u}^1, \dots, \mathbf{u}^M, \mathbf{h}, \mathbf{s}^1, \dots, \mathbf{s}^M) = & \\ \frac{1}{2} \sum_{i=1}^M \|\mathbf{y}_i - \mathbf{X}_i \mathbf{u}^i\|_2^2 + \lambda\|\mathbf{h}\|_1 + & \\ \frac{\rho}{2} \sum_{i=1}^M \|\mathbf{u}^i - \mathbf{h} + \mathbf{s}^i\|_2^2 - \frac{\rho}{2} \sum_{i=1}^M \|\mathbf{s}^i\|_2^2, & \end{aligned} \quad (9)$$

the following iterative solution is reached

$$\mathbf{u}^{i,(k+1)} = (\mathbf{X}_i^H \mathbf{X}_i + \rho \mathbf{I}_n)^{-1} (\mathbf{X}_i^H \mathbf{y}_i + \rho (\mathbf{h}^{(k)} - \mathbf{s}^{i,(k)})), \quad (10a)$$

$$\mathbf{h}^{(k+1)} = \mathbf{S}_{\frac{\lambda}{M\rho}} \left( \bar{\mathbf{u}}^{(k+1)} + \bar{\mathbf{s}}^{(k)} \right), \quad (10b)$$

$$\mathbf{s}^{i,(k+1)} = \mathbf{s}^{i,(k)} + \mathbf{u}^{i,(k+1)} - \mathbf{h}^{(k+1)}, \quad (10c)$$

being  $\bar{\mathbf{u}}$  and  $\bar{\mathbf{s}}$  the mean of  $\mathbf{u}^i$  and  $\mathbf{s}^i$  respectively, and  $\mathbf{S}_\kappa(a)$  the smooth thresholding operator defined as

$$\mathbf{S}_\kappa(a) = \begin{cases} a - \kappa \text{sign}(a), & |a| > \kappa \\ 0 & |a| \leq \kappa. \end{cases} \quad (11)$$

In order to ease the comprehension of the algorithm, Fig. 1 gathers all operations in a more visual way.

### III. ADMM IN MODELING AND DPD

In the first instance, modeling appears to be immediate, i.e., simply applying (10a) - (10c) iteratively. Instead, to achieve a linearization solution it is necessary to chose an architecture and adapt it to the algorithm. This way, the direct learning architecture (DLA) [5] is selected, implementing the ADMM before the DPD adaptation stage, as shown in Fig. 2. Notice that the notation has been changed in the figure to highlight that now the error is the variable to be modeled, satisfying equation  $\mathbf{e}_i = \mathbf{U}_i \mathbf{t}^i$  on each module block, being now  $\mathbf{U}_i$  the input regressors matrix, and  $\mathbf{t}^i$  the coefficients vector replica. After  $N$  iterations, the resulting coefficients  $\mathbf{d}\mathbf{w}_n$  scaled by a learning factor  $\mu$ , update the weights vector  $\mathbf{w}_{n+1}$  that governs the underlying Volterra model in the DPD stage,

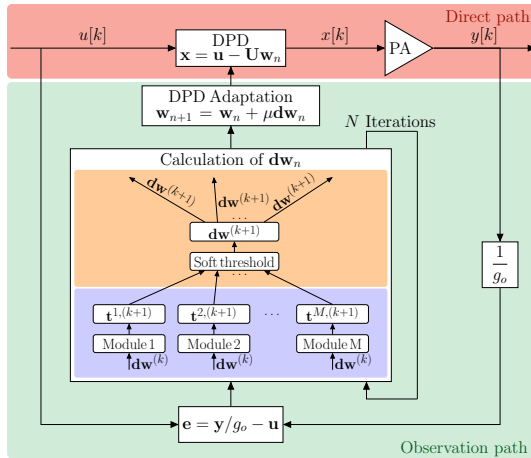


Fig. 2: Implementation of DLA using ADMM to calculate the model.

servicing now to convert the complex envelope at the input  $u[k]$  into the predistorted signal  $x[k]$ . The error vector is the difference between the normalized output  $y$  and its input  $u$ .

When modeling and linearizing using ADMM, a previous selection of  $\rho$  and  $\lambda$  is required. The larger  $\rho$  the faster should the method converge, and the relation between  $\lambda$  and  $M\rho$  will determine the amount of regressors that will be discarded. These two constants have been chosen by enforcing the minimization of the NMSE only once at the beginning.

#### IV. EXPERIMENTAL RESULTS

To achieve a better understanding about the benefits of ADMM, in this section its performance will be compared against the results obtained by the application of Doubly Orthogonal Matching Pursuit (DOMP) algorithm [6], which will be implemented by its reduced complexity version (RC-DOMP) [7].

##### A. Measurement test bench

The setup was composed by a Rohde & Schwarz SMU200A signal generator followed by two preamplifiers TVA-4W-422A+ of Minicircuits in cascade, measuring the output signal by a PXA-N9030A signal analyzer of Keysight Technologies, and using a power supply to drive the PA. The amplifier under test was a class AB PA based on the GaN transistor CGH400006P from Cree, with a working frequency of 3.6 GHz. The signal applied to the PA was an OFDM in accordance with the 5G-NR standard, having a 30 MHz bandwidth and a numerology of 1, i.e. a subcarrier spacing of 30 kHz, presenting a PAPR around 11 dB and generated with a sampling rate  $f_s = 92.16$  MHz. Then, the signal analyzer acquired samples of the complex envelope at the output of the PA. Moreover, the employed PC had 32 GB of RAM and an Intel i7 processor with 6 cores of CPU. For computing the ADMM, 16 computational modules were deployed, distributing the load over 4 cores of the CPU in parallel. Thus, ADMM was executed by the sequence of 4 turns with 4 modules acting in parallel.

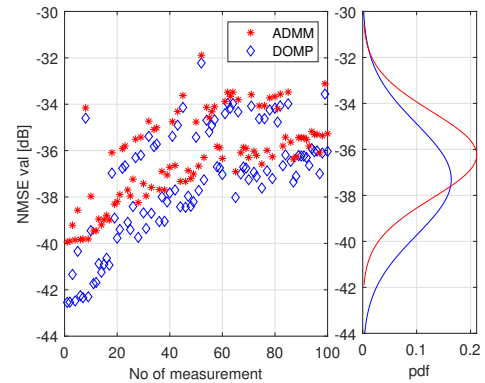


Fig. 3: Validation of the models with 100 different measurements. Probability density functions (pdf) have been estimated under the assumption that the empirical rule is applicable.

##### B. Modeling performance

For modeling this PA, a generalized memory polynomial (GMP) [8] model was chosen, building a subset of 121 regressors. Then, the model was estimated using a probe signal with random symbols generated with a fixed seed, and launching the computation of both methods, ADMM and DOMP. The values of  $\rho$  and  $\lambda$  used in ADMM were  $\rho = 4.6 \cdot 10^{-8}$  and  $\lambda = 2.8 \cdot 10^{-8}$ . The ADMM was programmed to run 20 iterations at each invocation of the method.

Once the coefficients were identified, the models were validated by computing the NMSE, using one hundred validation signals generated with different random seeds, so that, the only difference between identification and validation signals was the random seeds used to build the bit stream modulating the OFDM. In Fig. 3, it can be seen that the mean value of NMSE differs about 1 dB between both methods. An upward trend on NMSE can be appreciated as the iterations progress. A priori, this degradation is believed to be related to thermal effects, involving long term effects which could modify the PA distortion. Moreover, despite this behavior, it can be said that the solution reached by ADMM is slightly more stable by observing the NMSE probability distribution at the right panel of Fig. 3.

##### C. Predistortion evaluation

Then, maintaining the same structure of regressors, DPD was performed using DLA, choosing a learning factor  $\mu = 1/3$ . The DOMP was executed before evaluating it, selecting the most important regressors, and then, the calculation of  $d w$  was realized through the pseudo-inverse following (3). On the other hand, ADMM was implemented in DLA following the indications given in Section III.

The evolution of NMSE against the iterations of DLA is presented in Fig. 4 (a), where the iterative improvement in the coefficients of the underlying model can be observed. Furthermore, the evolution of the adjacent channel power ratio (ACPR) is shown in Fig. 4 (b). Taking into account the relationship between the NMSE decrease and the correction

TABLE I: Linearization performance in the different cases

Case	NMSE [dB]	ACPR <sub>L</sub> [dBc]	ACPR <sub>R</sub> [dBc]	EVM [%]	P <sub>out</sub> [dBm]
No DPD	-23.0	-31.9	-33.2	6.7	24.9
DLA-ADMM	-39.4	-45.0	-45.4	1.1	24.9
DLA-DOMP	-42.7	-48.2	-48.0	0.9	24.9

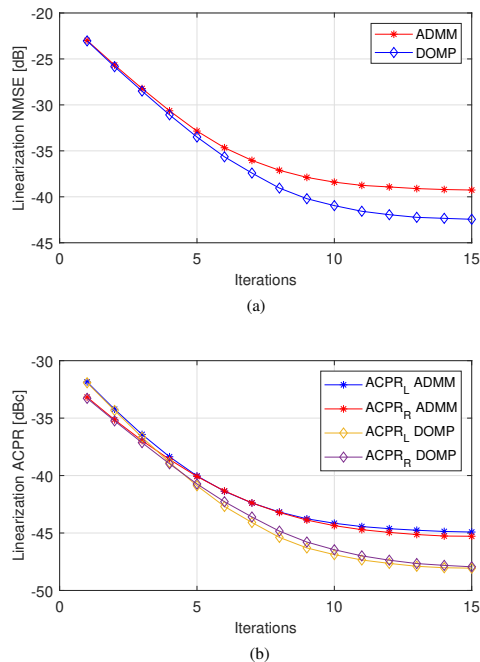


Fig. 4: Evolution of NMSE and ACPR during the DLA execution.

of the in-band distortion, and the ACPR reduction with the elimination of the out-band effects, it can be said that the two methods achieve a proper linearization. In addition, it can be considered that both techniques have reached convergence because the last slopes of the curves are very close to zero.

In particular, the effect on the ACPR can be evaluated by observing Fig. 5, where it can be seen how the spectral regrowth produced by the distortion has been reduced. Table I shows a validation result of both methods, that is, the result using a different input signal after calculation of the weights in a previous DLA performed with an identification signal. Information about the error vector magnitude (EVM) is also included in this table, which provides a notion about the scattering of the received symbols of the constellation. The lower the scatter, the less probability of bit error there will be, representing a measurement of the quality in terms of digital communications.

## V. CONCLUSIONS

The purpose of this paper was the presentation of a different technique to compute the regression involved on the

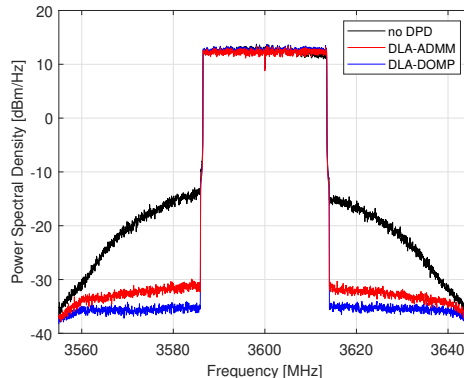


Fig. 5: Spectrum of the output signal with and without DPD.

resolution of both modeling and linearization problems with power amplifiers, exploring different benefits of it. After a theoretical development, it has been demonstrated that it is possible to apply ADMM in a distributed fashion to solve these problems and it has been exemplified using a class AB PA. The results show that ADMM achieves good performance both in modeling and in DPD, closely following the results achieved by a state-of-the-art method such as the DOMP.

## ACKNOWLEDGEMENTS

This work was supported by the Spanish National Board of Scientific and Technological Research (CICYT) under Grant TEC2017-82807-P, and by the European Regional Development Fund (ERDF) of the European Commission.

## REFERENCES

- [1] P. L. Gilibert, 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.
- [2] J. A. Becerra, M. J. Madero-Ayora, E. Marqués-Valderrama, M. Nogales, and C. Crespo-Cadenas, "Preconditioning the regression of power amplifier behavioral models and digital predistorters," in *2022 IEEE Topical Conference on RF/Microwave Power Amplifiers for Radio and Wireless Applications (PAWR)*, Jan. 2022, pp. 58–61.
- [3] M. Nogales, E. Marqués-Valderrama, J. A. Becerra, M. J. Madero-Ayora, and C. Crespo-Cadenas, "Arquitectura de aprendizaje directo en DPD de doble banda con regularización de Ridge," in *XXXVI Simposium Nacional de la Unión Científica Internacional de Radio (URSI 2021)*, Sep. 2021.
- [4] J. Heredia-Jueas, A. Molaei, L. Tirado, and J. A. Martínez-Lorenzo, "Consensus and sectioning-based ADMM with norm-1 regularization for imaging with a compressive reflector antenna," *IEEE Transactions on Computational Imaging*, vol. 7, pp. 1189–1204, Nov. 2021.
- [5] 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, Jul. 2004, pp. 2362–2366 Vol.4.
- [6] 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, Jun. 2018.
- [7] 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.
- [8] 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, 2004.

# Genetic Volterra Subspace Exploring of Wideband Power Amplifiers Models

Miguel Nogales, Juan A. Becerra, María José Madero-Ayora, and Elías Marqués-Valderrama  
mignoggon@alum.us.es, jabecerra@us.es, mjmadero@us.es, elimarval@alum.us.es

Dpto. de Teoría de la Señal y Comunicaciones. Universidad de Sevilla. Camino de los descubrimientos s/n. 41092. Sevilla

**Abstract**—In this work, the branch of machine learning known as genetic algorithms (GAs) is applied in the context of digital predistortion (DPD), specifically as a way to find the best subspace from the regressors generated by the Volterra series. Predistortion of power amplifiers (PAs) requires a model, usually Volterra-based, which is key in terms of performance. The classical approach of selecting a subset from the Volterra space is disputed, proposing a genetic-based Volterra-subspace generator (GVG) which finds the most suitable set of regressors from the whole Volterra basis. To contrast this proposal, the models generated with the GVG are benchmarked against state-of-the-art pruned structures in PA modeling and DPD with different configurations in terms of their hyperparameters. Measurement experiments demonstrate the generality of the proposed method and reveal the lack of precision behind the choice of conventional structures. The proposal, shaped as an open source software, is made available as a tool to generate sparse structures for PAs and DPDs, laying the foundation for a new perspective in DPD design.

## I. INTRODUCTION

The ever-increasing need of higher data rates and the rapid evolution of wireless communication schemes have positioned orthogonal frequency division multiplexing (OFDM) as the basis over which the fifth generation new radio (5G-NR) has been rolled out. Although OFDM signals are characterized by high efficiency in mobile scenarios, their time-domain signals exhibit a high peak-to-average power ratio (PAPR), which impacts the performance of the power amplifier (PA). PAs are efficient at high powers, where strong nonlinearities occur. The combination of high PAPR signals with the efficiency-versus-nonlinearity trade-off renders digital predistortion (DPD) as a unique scheme to diminish nonlinearity while still working at high power levels near saturation.

Within DPD, most of the flagship techniques are based on Volterra-based models. A priori pruned model structures, such as the memory polynomial (MP) [1] and the generalized memory polynomial (GMP) [2] were presented as subsets of the full Volterra space that offer fair performance in general situations. A posteriori selection techniques such as orthogonal matching pursuit (OMP) [3] or doubly orthogonal matching pursuit (DOMP) [4] are devoted to select the most relevant regressors from a set, while techniques like Ridge regularization deal with instability caused by ill conditioning. LASSO is an  $\ell_1$ -regularization example that performs coefficient selection.

This work explores a different approach powered by Genetic Algorithm (GA), and it is organized as follows. The background of GAs and DPD is provided in Section II. Section III covers the combination of both fields as a way to find the best subspace from the regressors generated in a

Volterra space. Next, Section IV reports the experimental design, and the results of the PA modeling and DPD experiments are exhibited. Finally, the conclusions wrap up the work in Section V, remarking the highlights and drawbacks of this approach.

## II. DIGITAL PREDISTORTION AND GENETIC ALGORITHMS: THEORETICAL BASIS

This section explains the basic concepts of DPD and GA and how they are interconnected in the areas that the proposal seeks to improve.

### A. Digital Predistortion and PA Modeling

In the context of DPD, the measurement equation relates the input and output of the PA and is defined as:

$$\mathbf{y} = \mathbf{X}\mathbf{h} + \mathbf{e}, \quad (1)$$

being  $\mathbf{y} = [y[k], y[k-1], \dots, y[k-m+1]]^T \in \mathbb{C}^{m \times 1}$  a stack of  $m$  PA output samples,  $k$  the discrete-time index, the model coefficients  $\mathbf{h} \in \mathbb{C}^{n \times 1}$ ,  $\mathbf{e}$  accounts for the model errors and the regressor matrix is denoted by  $\mathbf{X} \in \mathbb{C}^{m \times n}$ . This last term can be expressed as a row matrix of column elements, represented as

$$\mathbf{X} = [\xi_1 \quad \xi_2 \quad \dots \quad \xi_n] \in \mathbb{C}^{m \times n}, \quad (2)$$

where  $\xi_i$  are the Volterra regressors. The regressors act as axes over which the output signal is expressed by means of the least squares method. This regression is performed through the Moore-Penrose pseudoinverse  $\mathbf{X}^\dagger$  to estimate the coefficients  $\hat{\mathbf{h}}$ , yielding

$$\hat{\mathbf{h}} = \mathbf{X}^\dagger \mathbf{y} = (\mathbf{X}^H \mathbf{X})^{-1} \mathbf{X}^H \mathbf{y}, \quad (3)$$

where  $\mathbf{X}$  is required to be well conditioned.

The Volterra space can be represented by different models that account for certain variations in their regressors. First, the Volterra series, referenced here as full Volterra (FV) [5] model, defines the most general subspace for bandpass PAs, whose baseband expression is

$$y[k] = \sum_{p=0}^P \sum_{q_1=0}^{Q_1} \sum_{q_2=0}^{Q_2} \dots \sum_{q_{2p+1}=0}^{Q_{2p+1}} h_{2p+1}(q_1, q_2, \dots, q_{2p+1}) \cdot \prod_{r=1}^{p+1} x[k - q_r] \prod_{s=p+2}^{2p+1} x^*[k - q_s], \quad (4)$$

where  $P$  defines the model order and  $Q$  sets the maximum memory depth.

At the other extreme, in the case of PAs, one of the simplest Volterra-based models is the MP model. Another approach

is the GMP model, an evolved MP model which consists of the same structure adding components which are based on the lagging and leading envelope. Another model worth to mention is the Complex Volterra Series (CVS) [6], which yields a much denser space of regressors, so it can deal with errors due to both non linearity and I/Q impairments, but also its massive pool size makes it hard to manage.

Even though this approach will offer a richer model, its high computational complexity will not allow the extraction of the set of regressors in real-time executions such as the forward path of the predistorter.

### B. On Genetic Algorithms

The key aspects of this family of algorithms are the stages of each generation, identifying generations as iterations of the process. There will be some population of objects holding the data; this data will define how appropriate they are to solve a certain task, the problem proposed, and its aptitude to solve it will be quantified by a function, returning a value called fitness. This way, the stages that will occur in each iteration—or generation—are selection, crossover, and mutation.

Selection is performed by evaluating the quality of each member of the population, maintaining the members with the best fitness, and discarding the rest. That is the reason behind the *survival of the fittest* law, it is expected that the best performing data will not be discarded and even better individuals will be generated by means of the other two stages.

Next is crossover. Having performed the selection will lead to a better population, but smaller in size. This stage will deal with this, generating new members by mixing the data within the surviving ones. The crossover is an operation that can be implemented in many ways, but all of them center around having the parents' genetic data mixed (typically two parents) and then split (or not) into some children. It is expected that, by repeating this operation along the generations, the best genetics will arise and the worst ones will be discarded.

However, if only crossover was used, just the genetics present in the initial population could be the ones to be selected. This is the problem mutation tries to solve, it does so by creating a range of random modifications that when performed on the population will increase its diversity, being able to further explore the solution space.

### III. THE GENETIC-BASED VOLTERRA-SUBSPACE GENERATOR (GVG)

This section is devoted to match the GA basics with Volterra-based modeling. Each component of the GA framework is identified and translated to a modeling scheme to define the GVG. A high-level conceptual flow chart of the technique is presented in Fig. 1.

The regressor matrix takes the role of the population. This way, each regressor becomes an individual, or chromosome in the GA jargon, all these regressors being generated following a Volterra structure. Each regressor is unequivocally represented by a set of constituents  $x$ ,  $x^*$  and  $|x|$ , i.e., the signal, its conjugate, and its envelope, respectively. Each constituent is defined by its nonlinear order and memory depth.

One of the most important design choices is the fitness function. Since the DOMP technique returns a sorted list of regressors in decreasing relevance for the modeling, its

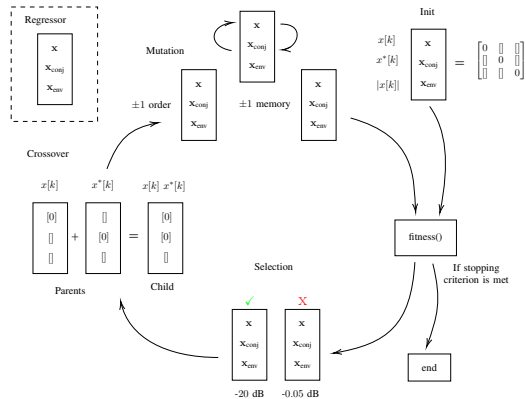


Fig. 1. Scheme of the genetic-based Volterra-subspace generator (GVG) stages applied on regressors, represented by their main constituents  $\mathbf{X}$ ,  $\mathbf{X}_{\text{conj}}$  and  $\mathbf{X}_{\text{env}}$  through the four phases scoring, selection, crossover and mutation.

output will be used to compute the regressor fitness. This way, the score of each regressor is defined as its improvement in modeling error, i.e., the decrease in normalized mean squared error (NMSE) after the inclusion of that regressor. The fitness function of the  $n$ -th individual follows

$$\text{fitness}_n = \text{NMSE}(\mathbf{X}_{1:n} \mathbf{h}, \mathbf{y}) - \text{NMSE}(\mathbf{X}_{1:n-1} \mathbf{h}, \mathbf{y}) \quad (5)$$

after the DOMP execution over the population at current generation.

The crossover between two regressors will produce only one child with all genetic data of their parents, whilst mutations are split into three types and consist of modifying the values of nonlinear order or memory depth in one constituent. These random modifications of the population ensure that new areas from within the Volterra space will be discovered and analyzed.

The code for the proposed algorithm is available at [7].

### IV. EXPERIMENTAL RESULTS

The features of the proposed GVG approach to explore the Volterra space are illustrated in this section, with experimental results presented for both modeling and linearization, comparing the above-mentioned models.

#### A. Modeling of a commercial power amplifier

This first part consists in the modeling of a commercial PA based on the CGH40006-TB testboard for the GaN transistor CGH40006P from Cree, operated at center frequency 2.14 GHz. The experimental setup employed in this case was the RF WebLab maintained by Chalmers University of Technology [8].

The probing signal was an OFDM signal with 20 MHz bandwidth, 30 kHz separation between subcarriers, and 16-QAM symbols over all the subcarriers, following the 5G-NR format. It presented a PAPR of 10.5 dB. A sampling frequency of 184.32 MSa/s was employed both for the generation of the probing signal and for its acquisition at the output of the PA under test. The average output power was 26.7 dBm with a gain compression of 1.8 dB.

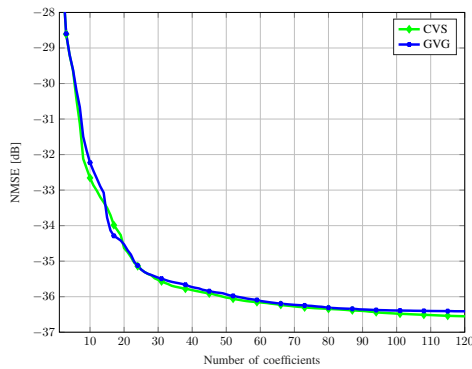


Fig. 2. Modeling NMSE versus the number of coefficients ordered by their relevance according to the DOMP algorithm for a complete CVS model structure with 6271 regressors and for the reduced set of regressors generated by the GVG approach.

Firstly, as a proof that the GVG approach can exhaustively explore the Volterra-based space, its modeling performance was compared to that obtained with the CVS model structure. Fig. 2 shows the modeling NMSE versus the number of coefficients ordered by their relevance according to the DOMP algorithm for the full stock of regressors of the CVS model structure and for the reduced set of regressors generated by the GVG approach. Since the CVS model is the most general Volterra-based subspace, only a set of 6271 regressors were considered for implementation reasons. For the sake of comparison, in this experiment, the GVG population was limited so that only the regressors composing the CVS subspace could appear. The similarity between the modeling performance in both curves of Fig. 2 demonstrates that the GVG approach is able to generate regressors with similar flexibility to the CVS model. Extending this feature to the exploration of a nonrestricted Volterra-based space, the GVG approach should be able to generate any valuable regressor in case no restrictions are imposed by memory resources or execution time.

Next, the modeling performance achieved by the GVG approach is compared to other classical model structures, which are: the MP model with 13-th order and memory depth 10; the GMP model with 13-th order for part  $a$  and 7-th order for parts  $b$  and  $c$ , and a maximum memory depth of 12; the FV model with 13-th order, memory depth 5 up to the 5-th order and memoryless for higher orders; the CVS model with 13-th order, memory depth 5 up to the 5-th order and memoryless for higher orders; and the model given by the GVG approach without any limitation in the type of regressors generated. The NMSE achieved by each model structure can be observed in Fig. 3 versus the number of components ordered by the DOMP algorithm. The GVG approach outperforms the other model structures due to the richer set of regressors it generates.

Finally, the normalized power spectral density (PSD) of the measured signal at the output of the commercial PA under test

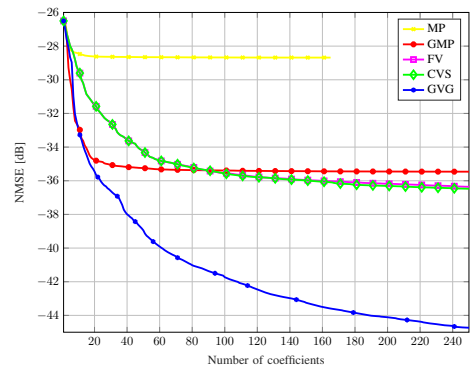


Fig. 3. Comparison of the modeling NMSE versus the number of coefficients ordered by their relevance according to the DOMP algorithm for the MP, GMP, FV and CVS models, and the model given by the GVG approach.

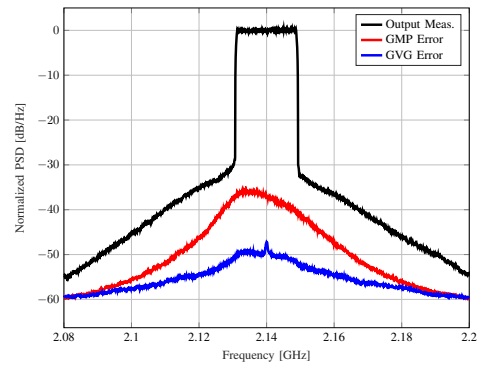


Fig. 4. Normalized power spectral density of the measured signal at the output of the commercial PA under test and normalized spectra of the error signals between the measured and the modeled outputs with the GMP model and with the set of regressors generated by the GVG approach.

is provided in Fig. 4, together with the normalized PSDs of the error signal between the measured and the modeled outputs with the GMP model and with the set of regressors generated by the GVG approach.

### B. Linearization of a commercial power amplifier

The capability of the GVG approach to determine the appropriate set of regressors for linearization is exemplified in this subsection for the same commercial PA indicated in Section IV-A. The linearization was achieved by means of a direct learning architecture (DLA) DPD [9], where a decreasing learning rate over the iterations was chosen for better convergence. DPD performance of the nonstructured model generated with the GVG approach was compared to that achieved by a reduced-order GMP model only containing the coefficients selected by the DOMP algorithm. Considering that the DOMP algorithm is integrated in the GVG approach,

TABLE I  
DPD PERFORMANCE AFTER 20 DLA ITERATIONS FOR THE COMMERCIAL PA UNDER TEST WITH AVERAGE OUTPUT POWER OF 26.7 DBM AND A 20-MHZ 5G-NR SIGNAL.

Case	NMSE [dB]	ACPR <sub>lower</sub> [dBc]	ACPR <sub>upper</sub> [dBc]	EVM [%]
No DPD	-26.6	-34.5	-36.0	4.0
GMP + DOMP	-34.8	-43.1	-45.1	1.7
GVG	-43.9	-51.1	-51.7	1.0

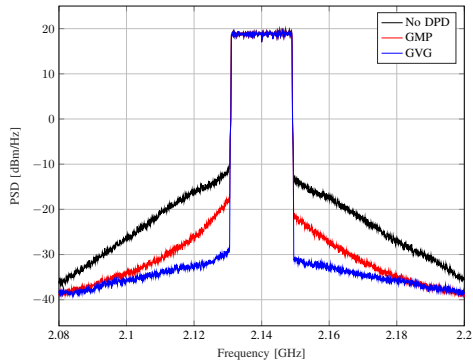


Fig. 5. Power spectral density at the output of the commercial PA under test without linearization and with DPD based on a reduced-order GMP model and a nonstructured model provided by the GVG approach.

its combination too with the GMP model allows a comparison in equivalent conditions. The configuration of the models was the same as in Section IV-A.

The PSDs of the linearized signals at the output of the commercial PA under test are presented in Fig. 5, together with the unlinearized output signal. It can be observed that the nonstructured model provided by the GVG approach outperforms the reduced-order GMP model. This is corroborated by the ACPR values provided in Table I, showing an improvement of 8.6 dB in the worst case with the reduced-order GMP model and of 15.7 dB in the worst case with the model provided by the GVG approach. It can be also observed that the linearization NMSE was improved in 8.2 dB with the reduced-order GMP and in 17.3 dB with the GVG approach, reaching in this second case a value of  $-43.9$  dB.

Another interesting result is the evolution of the linearization NMSE and ACPR provided by the two methods in comparison, as shown in Fig. 6.

## V. CONCLUSIONS

A new way of exploring and generating PA behavioral models by powering GAs with Volterra series is presented. The use of DOMP in the fitness calculation and selection processes ensures orthogonality, what provides the GVG a remarkable ability to find a sparse model.

Measurement experiments show how the GVG outperforms emblematic models after their pruning with the most recent coefficient selection techniques due to its capability of searching in the whole Volterra space without limitations imposed

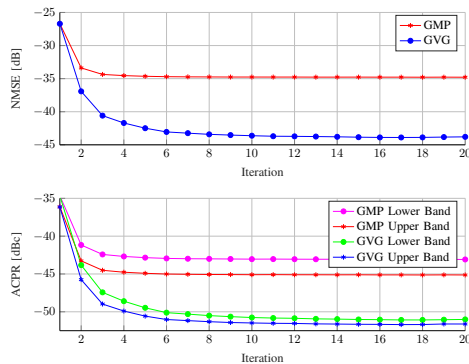


Fig. 6. Linearization NMSE and ACPR over the DLA DPD iterations for the commercial PA under test with a reduced-order GMP model and a nonstructured model provided by the GVG approach.

by a-priori pruned models. Since the technique relies on a high computational complexity to iterate the algorithm over its generations, it is not intended to be executed in real-time calculations but as a convenient tool to extract the model structure from a dataset throughout the DPD design process.

## ACKNOWLEDGMENTS

This project was supported by the Spanish National Board of Scientific and Technological Research (CICYT) under Grant TEC2017-82807-P, and by the European Regional Development Fund (ERDF) of the European Commission.

## REFERENCES

- [1] J. Kim and K. Konstantinou, "Digital predistortion of wideband signals based on power amplifier model with memory," *Electron. Lett.*, vol. 37, no. 23, p. 1417, Nov. 2001.
- [2] D. Morgan, Z. Ma, J. Kim, M. Zierdt, and J. Pastalan, "A generalized memory polynomial model for digital predistortion of RF power amplifiers," *IEEE Trans. Signal Process.*, vol. 54, no. 10, pp. 3852–3860, Oct. 2006.
- [3] 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 Trans. Microw. Theory Tech.*, vol. 63, no. 2, pp. 745–753, Feb. 2015.
- [4] 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 Microw. Wireless Compon. Lett.*, vol. 28, no. 8, pp. 726–728, Aug. 2018.
- [5] S. Benedetto, E. Biglieri, and R. Daffara, "Modeling and performance evaluation of nonlinear satellite links—a Volterra series approach," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-15, no. 4, pp. 494–507, Jul. 1979.
- [6] 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 Process.*, vol. 131, no. C, p. 245–248, Feb. 2017.
- [7] M. Nogales, J. A. Becerra, M. J. Madero-Ayora, and E. Marqués-Valderrama, "Genetic based Volterra space generator (GVG) sample code [source code]," Code Ocean, Feb. 2022, DOI: 10.24433/CO.328475.v1.
- [8] P. N. Landin, S. Gustafsson, C. Fager, and T. Eriksson, "Weblab: A web-based setup for PA digital predistortion and characterization [application notes]," *IEEE Microw. Mag.*, vol. 16, no. 1, pp. 138–140, Jan. 2015.
- [9] H. Paaso and A. Mammela, "Comparison of direct learning and indirect learning predistortion architectures," in *2008 IEEE International Symposium on Wireless Communication Systems*, Dec. 2008, pp. 309–313.

# Evaluación de regresores basados en modelos de Volterra para amplificadores de potencia con señales de banda ancha

Elías Marqués-Valderrama, María José Madero-Ayora, Juan A. Becerra, Miguel Nogales  
 emarques@us.es, mjmadero@us.es, jabecerra@us.es, mignoggon@alum.us.es

Dpto. de Teoría de la Señal y Comunicaciones. Universidad de Sevilla. Camino de los Descubrimientos, s/n. 41092 Sevilla.

**Resumen**—This work analyzes the accuracy achieved by two different Volterra-based modeling approaches when they are applied for the modeling and linearization of a power amplifier driven by a wideband signal. The conventional approach adopts a certain model structure, being the GMP model one of the most widely employed models in literature. An alternative approach is the genetic-based Volterra subspace generator (GVG), where the space containing the Volterra regressors is explored with a genetic algorithm, resulting in a set of regressors that compose a non-structured sparse model. Experimental results with a 50-MHz 5G-NR signal showed that the non-structured GVG approach outperforms a comprehensive set of GMP model settings with different nonlinear order and memory depth. The GVG approach was able to incorporate richer regressors that can be associated with a complex dynamic behavior of the power amplifier as the signal bandwidth increases.

## I. INTRODUCCIÓN

La apuesta radical en las últimas décadas hacia las comunicaciones móviles se ha visto reflejada en la irrupción de la quinta generación móvil (5G), cuyos principales objetivos consisten en mejorar prestaciones como el ancho de banda, el número de dispositivos conectados y la latencia. Dentro de la 5G *New Radio* (5G-NR), la multiplexación por división ortogonal de frecuencias (OFDM) constituye el pilar fundamental sobre el que se apoya el incremento de prestaciones, con aumentos considerables del ancho de banda de emisión.

Dentro de los sistemas de comunicaciones, el amplificador de potencia (PA) es un elemento fundamental para que se pueda llevar a cabo la comunicación, a la vez que acaba siendo el elemento que limita la eficiencia en potencia, debido a su comportamiento no lineal en zonas de alta eficiencia. Si bien la OFDM es la modulación por excelencia, presenta una relación potencia pico a potencia promedio (PAPR) demasiado elevada, de manera que cuando se hace trabajar al PA en el punto de operación más eficiente, esto es, cerca de la zona de saturación, proliferan los efectos no lineales, que degradan la señal y provocan el recrecimiento espectral que imposibilita las comunicaciones en las bandas de frecuencia adyacentes.

Este problema ha sido ampliamente estudiado a lo largo de la literatura, donde la linealización mediante predistorsión digital (DPD) se emplea en numerosos trabajos. En la mayoría de los casos se hace uso de las series de Volterra como modelo subyacente, ya que es capaz de modelar efectos no lineales y con memoria con funciones base de tipo polinómico. Dado que la serie de Volterra conforma un espacio de dimensiones inabarcables se suelen emplear subespacios más reducidos, donde típicamente prevalece el modelo *generalized memory*

*polynomial* (GMP) [1]. Normalmente, a mayor complejidad del conjunto dispositivo-señal mayor cantidad de regresores se necesita, degenerando en problemas numéricos, ocasionados por la correlación entre los regresores. Termina siendo muy difícil incorporar riqueza al modelo, ya que cada regresor aporta muy poca información, siendo más perjudicial el problema numérico ocasionado. Esto hace que la selección de coeficientes sea una tarea crítica en DPD, haciendo que destaquen técnicas como el *doubly orthogonal matching pursuit* (DOMP), que selecciona los regresores más relevantes en base a la correlación [2] y aplica el *Bayesian information criterion* (BIC) [3], para buscar el óptimo entre el número de regresores del modelo y la cantidad de información aportada, reduciendo así la matriz de regresores.

En los últimos años se ha extendido el uso de algoritmos genéticos (GA) para resolver distintos problemas relativos al DPD, como pueden ser la estimación de coeficientes o la optimización del podado de regresores, entre otros. En [4], se presentó el *genetic-based Volterra-subspace generator* (GVG), a fin de aprovechar la potencia de los GA para crear nuevos modelos no estructurados, aportando más riqueza con regresores no convencionales, y alcanzando mejores resultados que los modelos clásicos.

Ciertamente, el modelo GMP ofrece en general buenas prestaciones manteniendo una reducida complejidad computacional. En cambio, para anchos de banda elevados el modelo no consigue ajustar adecuadamente los nuevos comportamientos que se dan en la señal. Esto ha servido de motivación a algunos autores para la introducción de redes neuronales en la implementación de las técnicas de predistorsión [5], que si bien son capaces de linealizar el sistema, pierden capacidad de interpretación. En este artículo, se pretende comparar los regresores que componen el modelo no estructurado resultante del enfoque GVG con distintos modelos GMP, para así arrojar luz sobre la falta de rendimiento de los modelos GMP cuando se emplean anchos de banda más exigentes.

## II. MODELADO ESTRUCTURADO

La predistorsión digital no puede entenderse sin el modelado, siendo fundamental conocer bien las características del sistema antes de tratar llevar a cabo el DPD. Así pues, mediante la serie de Volterra, se explica la salida del amplificador como la suma ponderada de términos formados por productos de la envolvente compleja de la señal conjugada y retrasada. La expresión en banda base sin aplicar ninguna restricción se conoce como el modelo *Full Volterra* (FV), con la forma



$$y[n] = \sum_{p=0}^P \sum_{q_1=0}^{Q_1} \sum_{q_2=0}^{Q_2} \cdots \sum_{q_{2p+1}=0}^{Q_{2p+1}} h_{2p+1}(q_1, q_2, \dots, q_{2p+1}) \cdot \prod_{r=1}^{p+1} x[n - q_r] \prod_{s=p+2}^{2p+1} x^*[n - q_s], \quad (1)$$

donde el máximo orden no lineal contemplado es  $2P + 1$  y la profundidad de memoria para cada orden es  $Q_{2p+1}$ .

Para usar el modelo se emplea la ecuación de medida

$$\mathbf{y} = \mathbf{X}\mathbf{h} + \mathbf{e}, \quad (2)$$

donde las  $M$  muestras de la envolvente compleja de la salida  $\mathbf{y} = [y[n], y[n-1], \dots, y[n-M+1]]^T \in \mathbb{C}^{M \times 1}$  se pueden obtener como el producto de la matriz de regresores  $\mathbf{X} = [\phi_1, \phi_2, \dots, \phi_N] \in \mathbb{C}^{M \times N}$  por el vector de coeficientes  $\mathbf{h} \in \mathbb{C}^{N \times 1}$ , cuantificando el error de la estimación en  $\mathbf{e} \in \mathbb{C}^{M \times 1}$ . El modelo de Volterra queda pues embebido en la matriz  $\mathbf{X}$ , donde cada columna  $\phi_i$  representa uno de los  $N$  regresores, de manera que las columnas de  $\mathbf{X}$  contienen el espacio algebraico de representación del modelo.

Dado que existen infinidad de combinaciones de productos  $x[n - q_r] x^*[n - q_s]$ , tradicionalmente se aplican restricciones a la serie, a fin de crear subespacios de FV manejables. Sería el caso del modelo GMP [1], uno de los más empleados en la literatura, que viene dado por

$$y[n] = \sum_{k=0}^{K_a} \sum_{l=0}^{L_a} a_{kl} x[n-l] |x[n-l]|^k + \sum_{k=1}^{K_b} \sum_{l=0}^{L_b} \sum_{m=1}^{M_b} b_{klm} x[n-l] |x[n-l-m]|^k + \sum_{k=1}^{K_c} \sum_{l=0}^{L_c} \sum_{m=1}^{M_c} c_{klm} x[n-l] |x[n-l+m]|^k. \quad (3)$$

Téngase en cuenta que, con este enfoque, aquellos regresores que incorporen productos  $|x|^k$  con  $k$  impar quedarían fuera del espacio de FV, pero se permiten debido a que se ha probado que son beneficiosos para el ajuste del modelo.

Denominaremos modelado estructurado a esta forma de proceder. Principalmente, presenta la ventaja de que, siguiendo algún razonamiento *ad hoc* o basado en conocimiento del comportamiento físico del PA, se propone una estructura de modelo inicial que es una simplificación o podado de la estructura general dada en (1). Lamentablemente, estos modelos no siempre responden de forma adecuada en escenarios más complicados. El empleo del modelo GMP resulta inadecuado cuando, incluso ampliando mucho el espacio de regresores, no se consigue precisión en la estimación, como sucede al aumentar el ancho de banda de la señal. Esa circunstancia sugiere que quizás es conveniente buscar otro tipo de regresores y otras combinaciones que se encuentren fuera del subespacio generado por el GMP.

### III. MODELADO NO ESTRUCTURADO: GVG

En esta tesitura, coge fuerza la idea de emplear modelos no estructurados, apoyados en las capacidades de los algoritmos genéticos. Para ello, se plantea el uso del GVG, a fin de generar el subespacio de Volterra con aquellos regresores que

mejor se ajusten al comportamiento particular del PA. Esto da mayor libertad al modelo, pudiendo obtener combinaciones de regresores no incluidas en las estructuras GMP o FV.

Desde este punto de vista, cada regresor constituye un individual (o cromosoma) que se representa unívocamente como la combinación de los constituyentes  $x$ ,  $x^*$  y  $|x|$ , donde cada uno queda definido por su orden no lineal y su profundidad de memoria. La clave de este enfoque reside en que la obtención del nuevo individual se realiza aplicando DOMP. Así, sobrevivirán los nuevos descendientes que más reduzcan el error cuadrático medio normalizado (NMSE) y que presenten menor correlación con sus antecesores. De esta forma, se garantiza que se maximice la cantidad de información aportada por cada nuevo descendiente.

## IV. RESULTADOS EXPERIMENTALES

### A. Banco de medidas

La plataforma experimental empleada para la obtención de los resultados de este trabajo fue el RF WebLab mantenido por Universidad Tecnológica de Chalmers [6]. El banco de medidas incluye un transceptor vectorial de señal PXIe-5646R con ancho de banda instantáneo de 200 MHz y frecuencia central de 2,14 GHz, un módulo PXI-4130 como fuente de alimentación (ambos de National Instruments), un pre-amplificador lineal con una ganancia de 40 dB, el amplificador de potencia bajo prueba basado en el transistor de GaN CGH40006P de Cree y un atenuador de 30 dB. Se generó una señal OFDM de acuerdo con el formato de 5G-NR, con ancho de banda de 50 MHz, separación entre subportadoras de 30 kHz y símbolos 16-QAM sobre todas ellas. Dicha señal presentaba una PAPR de 11,2 dB. Se aplicó un factor de sobremuestreo en torno a 3,3 para adaptar la tasa de muestreo de la señal a la tasa de 200 MHz empleada en el RF WebLab. Las señales de entrada y salida se postprocesaron para alinearlas en el tiempo. En el punto de operación, el amplificador bajo prueba proporcionaba una potencia promedio a la salida de 29,5 dBm, con 1,5 dB de compresión de ganancia.

### B. Resultados en modelado

Para poder evaluar los distintos aspectos comentados, se propone una comparación exhaustiva, tratando de modelar el amplificador bajo prueba mediante los dos enfoques. En primer lugar, se propone una búsqueda intensiva de un modelo GMP, aumentando progresivamente el orden (valores de  $K_a$ ,  $K_b$  y  $K_c$  en (3)), la profundidad de memoria (valores de  $L_a$ ,  $L_b$  y  $L_c$  en (3)) y el desfase entre la entrada y su envolvente en los productos  $x[n-l] |x[n-l \pm m]|^k$  (valores de  $M_b$  y  $M_c$  en (3)). Así pues se han generado 231 modelos, donde el más extenso contiene 3603 regresores, alcanzando orden 21, profundidades de memoria de hasta 50 muestras con memoria decreciente en órdenes superiores y con desfases de hasta 10 *taps*. A cada modelo se le ha aplicado el algoritmo DOMP, ordenando así la matriz por regresores de mayor a menor importancia, y conservando únicamente el número de regresores óptimo indicado por el BIC.

En la Fig. 1 se presenta el NMSE de identificación alcanzado por cada modelo tras realizar la selección de coeficientes. Queda de manifiesto que, a partir de un orden y profundidad de memoria, no merece la pena seguir introduciendo regresores, puesto que no se mejora la precisión del modelo y

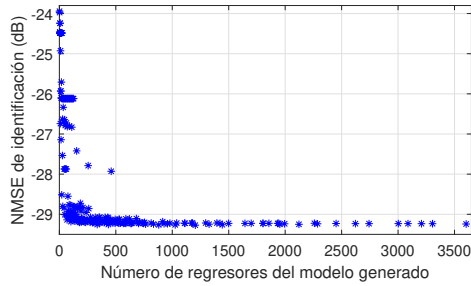


Fig. 1. NMSE de identificación frente a cantidad de regresores de cada modelo GMP. El NMSE fue calculado tras aplicar selección de coeficientes.

va en detrimento del coste computacional. El resultado no permite considerar que se ha encontrado un modelo aceptable, quedando al descubierto la necesidad de encontrar nuevos regresores que ayuden a mejorar estas prestaciones.

Por otro lado, se ha generado un modelo no estructurado a partir del enfoque GVG. Para ello, se han ejecutado 125 generaciones del algoritmo genético, alcanzando en la última una población de 2325 regresores. Igualmente, al llevar incorporada la selección de coeficientes mediante DOMP y BIC, el algoritmo devuelve un número óptimo de regresores. En la Fig. 2 se representa el avance del NMSE conforme se aportan los regresores ordenados según DOMP tanto para el modelo GMP más extenso como para el modelo encontrado por GVG. Se puede observar que la diferencia en la precisión de la estimación entre los dos modelos es considerable. El número óptimo de regresores según BIC para el enfoque GVG es de 235 con un porcentaje de regresores conservados del 10,1%, frente a los 79 del modelo GMP más extenso, que se traduce en una conservación del 2,2% de la matriz total. Con el número óptimo de regresores se ha obtenido un NMSE de identificación de  $-29,2$  dB y  $-29,0$  dB en validación para el modelo GMP, mientras que se alcanzan  $-37,6$  dB de NMSE de identificación y  $-37,1$  dB en validación para el caso de GVG. Estos números permiten confirmar que un escenario de estas características requiere de un modelo con regresores más ricos, que la estructura GMP no puede aportar.

Por último, en la Tabla I se recogen los 50 primeros regresores de cada modelo ordenados por el algoritmo DOMP, donde puede verse que la mayoría de los regresores seleccionados en el enfoque GVG son marcadamente diferentes de los presentes en la estructura del modelo GMP, por lo que aportan información relevante. El enfoque GVG genera regresores que contienen productos con más de dos retrasos diferentes para la señal y su envolvente. Además, algunos regresores muestran dependencia con el complejo conjugado de la señal  $x^*[n]$  o sus versiones retrasadas, lo que se relaciona con imperfecciones I/Q.

### C. Resultados en linealización

En este apartado se compararán las prestaciones alcanzadas en la linealización del amplificador bajo prueba con los enfoques de modelado estructurado, ejemplificado por el modelo GMP, y de modelado no estructurado empleando GVG. Para el punto de operación seleccionado, la distorsión no lineal del

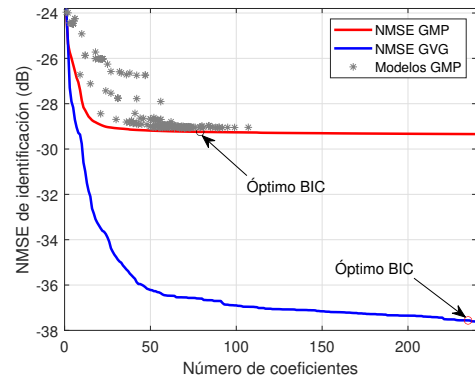


Fig. 2. Evolución del NMSE de identificación respecto a la cantidad de regresores incluidos en el modelo, ordenados en base a DOMP. La nube de puntos representa los NMSE de identificación alcanzados por el resto de modelos, cuando se aplica selección de coeficientes mediante DOMP.

amplificador viene dada en la Tabla II en términos del NMSE entre la señal de salida del PA y la señal de entrada escalada, la relación de potencia en el canal adyacente (ACPR) y la magnitud del vector error (EVM).

Para la comparación, se ha tomado el modelo GMP más extenso de los indicados en el apartado anterior, con un conjunto inicial de 3603 regresores, de los cuales solo 79 fueron seleccionados. Por otro lado, el conjunto no estructurado final de regresores que proporciona GVG está formado por 235. Basado en el conjunto de regresores proporcionados por cada enfoque, GMP+DOMP y GVG, se diseñó un DPD mediante aprendizaje directo (DLA) [7], para lo que se ejecutaron 15 iteraciones de entrenamiento con una tasa de aprendizaje de  $1/3$ . Después, los coeficientes del DPD se fijaron y se aplicó una señal de validación generada con una secuencia diferente de símbolos.

La densidad espectral de potencia de la señal de salida original junto con las señales linealizadas mediante los dos DPD se muestran en la Fig. 3, donde puede observarse una mayor reducción de las emisiones fuera de banda por parte del DPD basado en el enfoque no estructurado. Esto queda corroborado por los valores de ACPR proporcionados en la Tabla II, en la que se observa una mejora de 5,0 dB en el peor caso usando el modelo GMP+DOMP, mientras que la mejora es de 8,9 dB en el peor caso cuando se usa GVG. En cuanto a la distorsión en banda, cuyo indicador es el EVM, el DPD basado en GMP+DOMP alcanza un valor del 3,4%, mientras que el DPD basado en GVG reduce su valor hasta el 1,5%. Por último, la reducción del NMSE de linealización que produce el DPD basado en GMP+DOMP es de 5,6 dB, mientras que el DPD basado en GVG lo reduce en 14,6 dB.

### V. CONCLUSIONES

En este trabajo se ha analizado la precisión que consiguen dos enfoques diferentes basados en serie de Volterra para el modelado y linealización de un amplificador de potencia al que se le aplica una señal 5G-NR de 50 MHz. Por un lado, el enfoque convencional consiste en emplear modelado

TABLA I  
LISTA DE LOS 50 REGRESORES MÁS RELEVANTES DE CADA MODELO.

	GMP + DOMP	GVG
1	$x[n]$	$x[n]$
2	$x[n-5]$	$x[n-4]^2 x^*[n]$
3	$ x[n-5]   x[n-2] $	$x[n-5]$
4	$ x[n-2]   x[n-4] ^2$	$x[n-4] x[n-5] x^*[n]  x[n]   x[n-3] $
5	$ x[n-3]   x[n-2] ^2$	$x[n]  x[n] ^5$
6	$ x[n-4]   x[n+1] $	$x[n]^2 x^*[n-2]$
7	$ x[n-2]   x[n-3] ^2$	$ x[n-9] $
8	$ x[n-1]   x[n+1] ^5$	$x[n-6] x[n-8] x^*[n]$
9	$x[n]^*  x[n] $	$x^*[n]  x[n] ^5$
10	$x[n]^*  x[n] ^2$	$x[n]  x[n] $
11	$ x[n]   x[n-2] ^3$	$x[n]  x[n] ^2$
12	$ x[n-2]   x[n-1] $	$x[n-1]$
13	$ x[n-4] $	$x[n-6]$
14	$ x[n]   x[n-1] ^2$	$x[n]^3 x^*[n-4]^2$
15	$ x[n-3]   x[n-1] ^2$	$x[n-4] x[n-5] x^*[n]  x[n] $
16	$ x[n-1] ^*  x[n-1] $	$x[n-4] x[n-6] x^*[n]  x[n] ^3  x[n-3] $
17	$ x[n] ^*  x[n] ^5$	$x[n]^3 x^*[n-4]^2  x[n] $
18	$ x[n-3] $	$x[n-3] x[n-4] x^*[n]  x[n-4] $
19	$ x[n]   x[n-2] $	$x[n-6] x[n-9] x^*[n]  x[n]   x[n-7] $
20	$ x[n-5]   x[n+1] ^2$	$x[n-10]^2 x^*[n]$
21	$ x[n-6]   x[n] $	$ x[n-2] $
22	$ x[n-3]   x[n-4] $	$x[n-4] x[n-5] x^*[n]  x[n] ^2$
23	$ x[n]   x[n-3] $	$x[n]^3 x^*[n-4]^2  x[n-3] $
24	$ x[n-6]   x[n-2] $	$x[n-1] x[n-4] x^*[n-2]$
25	$ x[n-3]   x[n-1] $	$x[n-4] x[n-6] x^*[n]  x[n] $
26	$ x[n-9]   x[n-6] $	$x[n-4] x[n-5] x^*[n]  x[n-3] $
27	$ x[n-6]   x[n-3] ^5$	$x[n-4]^2 x^*[n]  x[n]   x[n-3] $
28	$ x[n-4]   x[n-3] ^4$	$x[n-3]^2 x^*[n]  x[n-4] ^2$
29	$ x[n]   x[n-5] ^4$	$ x[n-2]   x[n-9] $
30	$ x[n]   x[n-4] $	$x[n]^2 x^*[n-2]  x[n] $
31	$ x[n-5]   x[n-3] ^2$	$x[n]^3 x^*[n-4]^2  x[n] ^2$
32	$ x[n-5]   x[n+2] ^4$	$x[n-3]^2 x^*[n]$
33	$ x[n-1]   x[n+1] ^4$	$x[n-1]  x[n] ^3$
34	$ x[n-5]   x[n+1] $	$x[n-3] x[n-5] x^*[n] \cdot  x[n] ^3  x[n-3] ^2  x[n-5] $
35	$ x[n-6]   x[n-1] $	$x[n-10]$
36	$ x[n-6]   x[n-3] $	$x[n-2]$
37	$ x[n]   x[n-1] $	$x[n-3]$
38	$ x[n-4]   x[n-1] $	$x[n-3]^2 x^*[n]  x[n] $
39	$ x[n-4]   x[n+6] ^2$	$x[n-3] x[n-4] x^*[n]  x[n] ^2  x[n-4] $
40	$ x[n-13]   x[n-23] $	$x[n-5]  x[n-5] $
41	$ x[n-3]   x[n-7] ^3$	$x[n-3]^2 x^*[n]  x[n-3] $
42	$ x[n-3]   x[n-4] ^4$	$x[n-2] x[n-3] x^*[n]  x[n]   x[n-1] $
43	$ x[n-5]   x[n-4] $	$x[n-3] x[n-4] x^*[n]  x[n] ^3  x[n-4] ^2$
44	$ x[n]   x[n+2] ^2$	$x^*[n-1]  x[n] ^3  x[n-2] ^2$
45	$ x[n-6]   x[n-8] $	$x[n-4] x[n-5] x^*[n]$
46	$ x[n-5]   x[n-11] $	$x[n-10]^2 x^*[n]  x[n]   x[n-10] $
47	$ x[n-9]   x[n] $	$x[n-5] x[n-6] x^*[n] \cdot  x[n]   x[n-5]   x[n-6] $
48	$ x[n-6]   x[n+2] ^2$	$x[n]^3 x^*[n-3] x^*[n-4]  x[n] $
49	$ x[n-1]   x[n] $	$x[n-4]$
50	$ x[n-3]   x[n-2] $	$x[n-4] x[n-6] x^*[n]$

TABLA II  
RESULTADOS DE LINEALIZACIÓN PARA UNA POTENCIA PROMEDIO DE SALIDA DE 29,5 DBM CON UNA SEÑAL 5G-NR DE 50 MHZ.

Unidades	NMSE [dB]	ACPR <sub>inf</sub> [dBc]	ACPR <sub>sup</sub> [dBc]	EVM [%]
Sin DPD	-23,9	-33,7	-34,5	6,4
DPD GMP+DOMP	-29,5	-38,7	-39,8	3,4
DPD GVG	-38,5	-43,2	-43,4	1,5

estructurado, del que el GMP es una de las estructuras de modelo más extendidas en la literatura. Por otro lado, el enfoque GVG consiste en explorar el espacio de regresores de Volterra mediante un algoritmo genético, dando como resultado un modelo no estructurado de naturaleza dispersa.

Los resultados experimentales para la señal de banda ancha empleada indican que el enfoque no estructurado supera en modelado a un amplio número de configuraciones del modelo

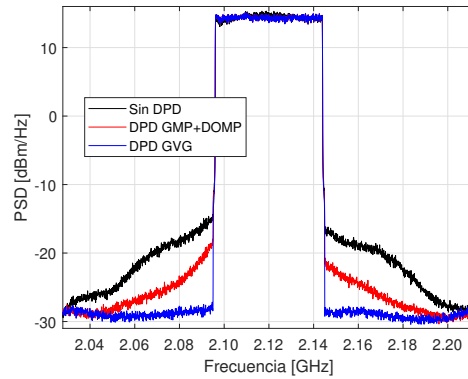


Fig. 3. Densidad espectral de potencia a la salida del amplificador sin linealización y con dos DPD, uno basado en el modelo estructurado GMP+DOMP y otro basado en el modelo no estructurado proporcionado por GVG.

GMP. En el caso del modelado estructurado que proporciona el modelo GMP, se ha comprobado que la incorporación de un mayor número de regresores al aumentar el orden o la profundidad de memoria no supone una mejora del ajuste ni de la capacidad de linealización.

La razón de las mejores prestaciones obtenidas con el enfoque GVG están relacionadas con su capacidad para encontrar el conjunto más adecuado de regresores sin que existan restricciones impuestas por una estructura predeterminada para el modelo, pudiendo de esa forma incorporar regresores con un formato más rico en términos de memoria que se asocian al comportamiento dinámico más complejo del amplificador a medida que aumenta el ancho de banda de la señal.

#### AGRADECIMIENTOS

Este trabajo es parte del proyecto de I+D+i PID2021-123090NB-I00, financiado por MCIN/AEI/10.13039/501100011033/ y "FEDER Una manera de hacer Europa".

#### REFERENCIAS

- [1] D. R. Morgan, Z. Ma, J. Kim, M. G. Zierdt, J. Pastalan, "A Generalized memory polynomial model for digital predistortion of RF power amplifiers," *IEEE Trans. Signal Process.*, vol. 54, no. 10, pp. 3852-3860, Oct. 2006.
- [2] J. A. Becerra, M. J. Madero-Ayora, J. Reina-Tosina, C. Crespo-Cadenas, "Sparse Identification of Volterra Models for Power Amplifiers Without Pseudoinverse Computation," *IEEE Trans. Microw. Theory Techn.*, vol. 68, no. 11, pp. 4570-4578, Nov. 2020.
- [3] J. A. Becerra, M. J. Madero-Ayora, R. G. Noguera, C. Crespo-Cadenas, "On the optimum number of coefficients of sparse digital predistorters: a Bayesian approach," *IEEE Microw. and Wireless Comp. Lett.* vol. 30, no. 12, pp. 1117-1120, Dec. 2020.
- [4] M. Nogales, J. A. Becerra, M. J. Madero Ayora, E. Marqués Valderrama, "Genetic Volterra subspace exploring of wideband power amplifiers models," *XXXVII Simposium Nacional de la Unión Científica Internacional de Radio*, 5-7 Sep. 2022, Málaga.
- [5] T. Kobal, A. Zhu, "Digital Predistortion of RF Power Amplifiers With Decomposed Vector Rotation-Based Recurrent Neural Networks," *IEEE Trans. Microw. Theory Techn.*, vol. 70, no. 11, pp. 4900-4909, Nov. 2022.
- [6] P. Landin, S. Gustafsson, C. Fager, T. Eriksson, "Weblab: A web-based setup for PA digital predistortion and characterization," *IEEE Microwave Mag.*, vol. 16, pp. 138-140, Feb. 2015.
- [7] D. Zhou, V. E. DeBrunner, "Novel adaptive nonlinear predistorters based on the direct learning algorithm," in *IEEE Trans. Signal Process.*, vol. 55, no. 1, pp. 120-133, Jan. 2007.



# Índice de Figuras

2.1	Ejemplo de característica AM/AM de un amplificador clase J	6
2.2	Saturación del NMSE de identificación tras realizar selección de coeficientes en un conjunto de 231 modelos GMP con diferentes parámetros, con un máximo de 3063 regresores [11]	6
2.3	Evolución del NMSE de identificación respecto a la cantidad de regresores incluidos en el modelo, ordenados en base a DOMP. La nube de puntos representa los NMSE de identificación alcanzados por distintos modelos GMP, cuando se aplica selección de coeficientes mediante DOMP [11]	7
2.4	Representación del PA. Las siglas SNLM simbolizan un sistema no lineal con memoria	8
2.5	Representación del PA según la ecuación (2.11)	8
2.6	La estimación de mínimo error $\hat{\mathbf{y}}_*$ se obtiene mediante una proyección ortogonal al espacio de representación	9
2.7	Función de distribución de probabilidad (f.d.p) del modelo de Montecarlo junto con las señales de máxima y mínima PAPR. La f.d.p del modelo de Montecarlo presenta valores en todo el dominio de $ \mathbf{x} $	12
2.8	Divergencia de Kullback-Leibler de la población de 4000 señales OFDM con el modelo de Montecarlo $D_{KL}(p_{X_i}^d    q)$	13
2.9	Probabilidad de que $q$ genere una secuencia $p_{X_i}^d$ de longitud $n = 737820$ , normalizada y escalada por $L/n$ con $L = 100$ . Este resultado se interpreta a partir de ahora como una probabilidad de extrapolación	13
3.1	Ejemplo gráfico de dualidad fuerte (izquierda) y dualidad débil (derecha)	16
3.2	Ejemplo de función de minimización RSS. En rojo se representa la curva obtenida con todos los puntos mientras que en negro se presentan las curvas correspondientes a dividir el problema en 10 partes con el mismo número de muestras	19
3.3	Esquemático de división sabia de filas, donde se ha de minimizar el conjunto de $\{f_1(\mathbf{u}_1), \dots, f_M(\mathbf{u}_M)\}$ funciones de minimización, según (3.22)	20
3.4	Topología de ADMM centralizado para la resolución del problema planteado [37]	22
3.5	Esquema conceptual de DPD	23
3.6	Arquitectura de aprendizaje directo	23
3.7	Implementación de ADMM para la arquitectura de aprendizaje directo [37]	24
3.8	Comparativa de complejidad computacional entre LS, representada en la línea roja discontinua y ADMM entre 1 y 20 iteraciones del método	25
3.9	Divergencia de Kullback-Leibler de cada trozo de señal con la señal completa para la semilla aleatoria 1000, en un escenario de 64 trozos	26
4.1	Setup de medidas del RF WebLab de la Universidad de Chalmers	27
4.2	Implementación llevada a cabo sobre el equipo. Las operaciones se realizan en 8 etapas en los tiempos $t_1, \dots, t_8$ , guardando los resultados en memoria para finalmente calcular la media y aplicar el umbral suave	29
4.3	Búsqueda del modelo mediante GVG (a) avance en las 125 iteraciones, (b) aporte al NMSE de cada regresor ordenado según DOMP, con el óptimo determinado por BIC ( $n_{opt}$ )	30
4.4	NMSE de identificación de 16 modelos ordenados por PAPR, obtenidos para ADMM y para mínimos cuadrados tras la aplicación de DOMP	31

4.5	NMSE de validación. Las curvas min ADMM y min DOMP representan los mejores resultados de NMSE, que se corresponden con modelos más cercanos a la PAPR que se valida. Las curvas ADMM y DOMP se corresponden con las validaciones obtenidas con el modelo obtenido a partir de la señal de semilla 1000	31
4.6	Curvas de probabilidad computadas para tres de los modelos a color. La curva negra discontinua es la del modelo de Montecarlo	32
4.7	NMSE de validación. Las curvas min ADMM y min DOMP representan los mejores resultados de NMSE, que se corresponden con modelos más cercanos a la PAPR que se valida. Las curvas $ADMM_{ponderado}$ y $DOMP_{ponderado}$ se corresponden con las validaciones obtenidas con el modelo obtenido a partir de la suma ponderada de los 16 modelos en base a la cercanía en PAPR	33
4.8	Factor de aprendizaje empleado para la etapa de entrenamiento de DLA	33
4.9	Descenso del NMSE de linealización (a) y ACPR (b) durante la identificación en las 15 iteraciones que se llevaron a cabo de DLA	34
4.10	Ganancia instantánea frente a potencia instantánea de entrada. Aunque se ha corregido un poco el comportamiento aún queda compresión en la zona de alta potencia de entrada	34
4.11	Comparativa antes y después de hacer DPD del espectro (a) y del espectro del error (b)	35
4.12	Variación de $\lambda$ y $\rho$ para cada iteración de DLA	36
4.13	Comparativa del descenso del NMSE de linealización (a) y ACPR (b) para ambos métodos durante la identificación en las 15 iteraciones que se llevaron a cabo de DLA	36
4.14	Ganancia instantánea frente a potencia instantánea de entrada para ambos métodos	37
4.15	Comparativa antes y después de hacer DPD del espectro (a) y del espectro del error (b) para ambos métodos	37
4.16	NMSE de validación (a) y ACPR de validación (b) para ambos métodos de señales con distinta PAPR	38
4.17	Comparativa entre ADMM y LS en validación, incluyendo el número de operaciones necesarias para la identificación, que en el caso de ADMM se corresponden con 20 iteraciones del método	38
4.18	Implementación de ADMM en paralelo. El número de operaciones a realizar en un instante determinado por cada módulo es el 0.75 % de las que implica hacer LS	39
4.19	Implementación de ADMM en paralelo. El número de operaciones a realizar en un instante determinado por el módulo de CPU es el 0.75 % de las que implica hacer LS	39

# Índice de Tablas

---

4.1	Resultados medios de linealización para una potencia promedio de salida de 29.5 dBm con señales 5G-NR de 50 MHz cuya PAPR varía entre 10.1 y 12.7 dB	38
-----	--	----





# Bibliografía

---

- [1] 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.
- [2] 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, Feb. 2017.
- [3] S. Benedetto, E. Biglieri, and R. Daffara, “Modeling and performance evaluation of nonlinear satellite links-A Volterra series approach,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-15, pp. 494–507, Jul. 1979.
- [4] C. Crespo-Cadenas, M. J. Madero-Ayora, and J. A. Becerra, “A bivariate Volterra series model for the design of power amplifier digital predistorters,” *Sensors*, vol. 21, no. 17, Sep. 2021.
- [5] ———, “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)*, Jan. 2021, pp. 4–7.
- [6] T. Adali, P. J. Schreier, and L. L. Scharf, “Complex-valued signal processing: The proper way to deal with impropriety,” *IEEE Transactions on Signal Processing*, vol. 59, no. 11, pp. 5101–5125, 2011.
- [7] B. Schubert, A. Gökceoglu, L. Anttila, and M. Valkama, “Augmented Volterra predistortion for the joint mitigation of power amplifier and I/Q modulator impairments in wideband flexible radio,” in *2013 IEEE Global Conference on Signal and Information Processing*, Dec. 2013, pp. 1162–1165.
- [8] 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.
- [9] 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.
- [10] E. Marqués-Valderrama, “Esquema de linealización con arquitectura de aprendizaje directo y regularización de Ridge. (Trabajo Fin de Grado Inédito),” Sevilla, 2021.
- [11] E. Marqués-Valderrama, M. J. Madero-Ayora, J. A. Becerra, and M. Nogales, “Evaluación de regresores basados en modelos de Volterra para amplificadores de potencia con señales de banda ancha,” in *XXXVIII Simposium Nacional de la Unión Científica Internacional de Radio (URSI 2023)*, 2023.
- [12] M. Nogales, J. A. Becerra, M. J. Madero-Ayora, and E. Marqués-Valderrama, “Genetic Volterra subspace exploring of wideband power amplifiers models,” in *XXXVII Simposium Nacional de la Unión Científica Internacional de Radio (URSI 2022)*, Sep. 2022.
- [13] ———, “Genetic based Volterra space Generator (GVG) sample code [source code],” in *XXXVII Simposium Nacional de la Unión Científica Internacional de Radio (URSI 2022)*. Code Ocean, Feb. 2022.

- [14] 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.
- [15] A. Abdelhafiz, A. Kwan, O. Hammi, and F. M. Ghannouchi, "Digital predistortion of LTE-A power amplifiers using compressed-sampling-based unstructured pruning of Volterra series," *IEEE Transactions on Microwave Theory and Techniques*, vol. 62, no. 11, pp. 2583–2593, Oct. 2014.
- [16] J. A. Becerra, D. Herrera, M. J. Madero-Ayora, and C. Crespo-Cadenas, "Sparse model selection of digital predistorters using subspace pursuit," in *2018 13th European Microwave Integrated Circuits Conference (EuMIC)*, Nov. 2018, pp. 190–193.
- [17] 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.
- [18] 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.
- [19] 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.
- [20] C. Crespo-Cadenas, M. J. Madero-Ayora, J. A. Becerra, and S. Cruces, "A fast sparse Bayesian pursuit approach for power amplifier linearization," in *2021 IEEE MTT-S International Wireless Symposium (IWS)*, Aug. 2021, pp. 1–3.
- [21] ———, "A sparse-Bayesian approach for the design of robust digital predistorters under power-varying operation," *IEEE Transactions on Microwave Theory and Techniques*, vol. 70, no. 9, pp. 4218–4230, Mar. 2022.
- [22] P. Schreier and L. Scharf, *Statistical Signal Processing of Complex-Valued Data: The Theory of Improper and Noncircular Signals*. Cambridge University Press, 2010.
- [23] J. A. Becerra, M. J. Madero-Ayora, E. Marqués-Valderrama, M. Nogales, and C. Crespo-Cadenas, "Preconditioning the regression of power amplifier behavioral models and digital predistorters," in *2022 IEEE Topical Conference on RF/Microwave Power Amplifiers for Radio and Wireless Applications (PAWR)*, Jan. 2022, pp. 58–61.
- [24] 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.
- [25] A. Molisch, *Wireless communications*. Chichester: John Wiley, Dec. 2010.
- [26] Z. Zhou, X. Chen, E. Li, L. Zeng, K. Luo, and J. Zhang, "Edge intelligence: Paving the last mile of artificial intelligence with edge computing," *Proceedings of the IEEE*, vol. 107, pp. 1738–1762, Jun. 2019.
- [27] Z. Lin, H. Li, and C. Fang, *Alternating Direction Method of Multipliers for Machine Learning*, Springer, Ed. Springer, Jun. 2022.
- [28] S. P. Boyd, *Convex optimization*. Cambridge: Cambridge University Press, 2004.
- [29] S. Cruces, *Apuntes de la asignatura Procesamiento Avanzado de Señal en Comunicaciones. Master Universitario en Ingeniería de Telecomunicación (ETSI). Departamento de Teoría de la Señal y Comunicaciones. Universidad de Sevilla*, Sep. 2021.
- [30] D. Gabay and B. Mercier, "A dual algorithm for the solution of nonlinear variational problems via finite element approximation," *Computers and Mathematics with Applications*, vol. 2, no. 1, pp. 17–40, 1976.

- [31] R. T. T. Hastie and R. Friedman, *The elements of statistical learning: data mining, inference, and prediction*, 2nd ed. Springer, 2017.
- [32] M. Nogales, E. Marqués-Valderrama, J. A. Becerra, M. J. Madero-Ayora, and C. Crespo-Cadenas, "Arquitectura de aprendizaje directo en DPD de doble banda con regularización de Ridge," in *XXXVI Simposium Nacional de la Unión Científica Internacional de Radio (URSI 2021)*, Sep. 2021.
- [33] 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.
- [34] J. Heredia-Juesas, L. Tirado, A. Molaei, and J. A. Martínez-Lorenzo, "ADMM based consensus and sectioning norm-1 regularized algorithm for imaging with a CRA," in *2019 IEEE International Symposium on Antennas and Propagation and USNC-URSI Radio Science Meeting*, Oct. 2019, pp. 549–550.
- [35] J. A. Becerra, M. J. Madero-Ayora, R. G. Noguera, 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.
- [36] J. Heredia-Juesas, A. Molaei, L. Tirado, and J. A. Martínez-Lorenzo, "Consensus and sectioning-based ADMM with norm-1 regularization for imaging with a compressive reflector antenna," *IEEE Transactions on Computational Imaging*, vol. 7, pp. 1189–1204, Nov. 2021.
- [37] E. Marqués-Valderrama, M. J. Madero-Ayora, J. A. Becerra, and M. Nogales, "Modeling and linearization of wideband power amplifiers with alternating direction method of multipliers (ADMM)," in *XXXVII Simposium Nacional de la Unión Científica Internacional de Radio (URSI 2022)*, Sep. 2022.
- [38] D. Bertsekas and J. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*. Athena Scientific, 2015.
- [39] K. Bredies and D. A. Lorenz, "Linear convergence of iterative soft-thresholding," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 813–837, Apr. 2008.
- [40] M. Schetzen, "Theory of pth-order inverses of nonlinear systems," *IEEE Transactions on Circuits and Systems*, vol. 23, no. 5, pp. 285–291, May 1976.
- [41] 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.
- [42] D. Zhou and V. DeBrunner, "A novel adaptive nonlinear predistorter based on the direct learning algorithm," in *2004 IEEE International Conference on Communications*, vol. 4, Jun. 2004, pp. 2362–2366.
- [43] R. Farebrother, *Linear Least Squares Computations*. Routledge, 1988.
- [44] P. Landin, S. Gustafsson, C. Fager, and T. Eriksson, "WebLab: A web-based setup for PA digital predistortion and characterization [application notes]," *IEEE Microwave Magazine*, vol. 16, no. 1, pp. 138–140, Feb. 2015.



# Glosario

---

**5G** Quinta generación móvil. 1, 11, 15

**5G-NR** *5G New Radio*. 1, 5, 7, 11, 27

**ACPR** Relación de potencia con el canal adyacente. 28, 29, 34, 35

**ADMM** Método de dirección alterna de multiplicadores. 1, 15, 17, 18, 22, 23, 24, 25, 30, 33, 35, 38, 39, 41

**BIC** *Bayesian Information Criterion*. 18, 30, 35

**CoSaMP** *Compressive sampling matching pursuit*. 8

**CVS** *Complex Volterra series*. 4, 7, 8, 12

**DKL** Divergencia de Kullback-Leibler. 10

**DLA** Arquitectura de Aprendizaje Directo. 23, 34, 35, 41

**DOMP** *Doubly Orthogonal Matching Pursuit*. 6, 7, 18, 30, 33, 34, 35, 59

**DPD** *Digital Predistortion*. 1, 3, 22, 23, 28, 36, 39

**eMBB** Banda ancha móvil mejorada. 1

**EVM** Magnitud del vector error. 29

**FFT** Transformada rápida de Fourier. 11

**FV** *Full Volterra*. 4, 5, 6, 7

**GMP** *Generalized Memory Polynomial*. 5, 7, 59

**GVG** *Genetic-based Volterra subspace generator*. 6, 7, 27, 30, 35

**IFFT** Transformada rápida de Fourier Inversa. 11

**ILA** Arquitectura de Aprendizaje Indirecto. 23

**LASSO** *Least absolute shrinkage and selection operator*. 18, 21

**LS** Mínimos cuadrados. 1, 9, 23, 30, 34, 35, 39, 41

**mMTC** Comunicaciones tipo máquina masivas. 1

**MP** *Memory Polynomial*. 5

**NMSE** Error cuadrático medio normalizado. 7, 28, 29, 30, 32, 34, 35, 41, 59

- OFDM** Multiplexación por división de frecuencias ortogonales. 1, 11, 12, 13, 27, 37, 59
- PA** Amplificador de potencia. 1, 3, 5, 6, 8, 10, 11, 12, 14, 18, 22, 23, 27, 28, 30, 36, 37, 59
- PAPR** Relación potencia pico a potencia promedio. 1, 10, 11, 13, 14, 30, 31, 32, 36, 37, 38, 41
- PEP** Potencia máxima de la envolvente. 11, 14, 30, 31, 37
- PRSS** Suma residual de cuadrados penalizada. 18
- RF** Radiofrecuencia. 4, 27
- RSS** Suma residual de cuadrados. 9, 15, 18, 19
- SP** *Subspace Pursuit*. 8
- URLLC** Comunicaciones ultra fiables de baja latencia. 1, 15