Sparse Identification of Volterra Models for Power Amplifiers Without Pseudoinverse Computation

Juan A. Becerra, *Senior Member, IEEE*, María José Madero Ayora, *Senior Member, IEEE*, Javier Reina-Tosina, *Senior Member, IEEE*, and Carlos Crespo-Cadenas, *Senior Member, IEEE*

***Abstract*— We present a new formulation of the doubly orthogonal matching pursuit (DOMP) algorithm for the sparse recovery of Volterra series models. The proposal works over the covariance matrices by taking advantage of the orthogonal properties of the solution at each iteration and avoids the calculation of the pseudoinverse matrix to obtain the model coefficients. A detailed formulation of the algorithm is provided along with a computational complexity assessment, showing a fixed complexity per iteration compared with its previous versions in which it depends on the iteration number. Moreover, we empirically demonstrate the reduction in computational complexity in terms of runtime and highlight the pruning capabilities through its application to the digital predistortion of a class J power amplifier operating under 5G-NR signals with the bandwidth of 20 and 30 MHz, concluding that this proposal significantly outperforms existing techniques in terms of computational complexity.**

***Index Terms*— Behavioral modeling, digital predistortion (DPD), doubly orthogonal matching pursuit (DOMP), greedy algorithm, model identification, power amplifier (PA), sparse regression, Volterra series.**

I. Introduction

THE CHOICE of the operation point of power amplifiers (PAs) poses a challenge in terms of the tradeoff between efficiency and linearity. On the one hand, PAs are known to show nonlinear behavior at high power levels, where efficiency is higher. This fact combined with the linearity requirements of new digital modulations, such as those included in the fifth generation (5G) of mobile communications, put the spotlight on signal processing techniques such as digital predistortion (DPD), which allows operating near saturation while mitigating nonlinearity [1].

DPD is based on behavioral models, on which there exists a vast literature about which is the best model structure. Volterra series are a widely adopted approach in this context since they allow to represent a nonlinear system with memory, while the coefficients of the model are still linear with respect to the system output. This mathematical signature leads to the classical theory of regression, in which the least-squares error minimization provides a general tool that enables the model coefficient retrieval.

Unfortunately, Volterra series suffer the curse of dimensionality since their number of coefficients largely grows with the memory depth and order, being this basis for the need of model selection and dimensionality reduction. In this context, drawing from the most-general full Volterra (FV) model [2], *ad hoc* behavioral models, such as the memory polynomial (MP) [3] or the generalized MP (GMP) [4] models, arise.

In contrast, *a posteriori* pruning techniques perform a selection of the most-relevant coefficients or model regressors by analyzing the signal properties. Within the family of greedy pursuits, which are characterized for selecting regressors by hard decision iteratively, the orthogonal matching pursuit (OMP) was applied to PA DPD in [5], followed by a reduced complexity version named by the authors as a simplified sparse parameter identification (SSPI) algorithm [6] that turned the calculation of the pseudoinverse matrix at each step into an iterative fashion. The modified Gram–Schmidt (MGS) technique [7], also known as orthogonal least squares (OLS) [8]–[10], and OMP were fused into the doubly OMP (DOMP) [11], which was shown to overcome the difficulties of choosing the correct regressors under a highly correlated basis such as the Volterra series. The main issue of the use of MGS against OMP is the significant increase in computational complexity [12], although the latter has shown to be outperformed by the DOMP in the pruning of both PA and DPD models [13].

In this framework, a special interest has risen in low-complexity DPD implementations [14]–[20] due to the desired use of limited resources and memory in field-programmable gate arrays (FPGAs) and the numerical stability benefits observed from the signal processing standpoint.

The DOMP algorithm in [11] presents a high computational complexity due to the pseudoinverse operation and the Kronecker product needed for the orthogonalization. In [21], an SSPI DOMP was contributed, which showed a lower computational complexity by transforming the pseudoinverse calculation in a recursive operation that depends on the result of the previous iteration. In this work, we provide a new formulation of the algorithm that takes advantage of the orthogonal properties to work in the covariance domain, achieving a reduced complexity (RC-DOMP) that is evidenced

Manuscript received May 9, 2020; revised July 15, 2020; accepted July 18, 2020. Date of publication August 27, 2020; date of current version November 4, 2020. This work was supported in part by the Spanish National Board of Scientific and Technological Research (CICYT) under Grant TEC2017-82807-P and in part by the European Regional Development Fund (ERDF) of the European Commission. *(Corresponding author: Juan A. Becerra.)*

The authors are with the Departamento de Teoría de la Señal y Comunicaciones, Escuela Técnica Superior de Ingeniería, Universidad de Sevilla, 41092 Seville, Spain (e-mail: jabecerra@us.es; mjmadero@us.es; jreina@us.es; ccrespo@us.es).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMTT.2020.3016967

by a fixed runtime per iteration. In addition, a transformation matrix from the Volterra domain to the equivalent orthogonal domain is defined as a result of the algorithm, enabling orthogonal and parallel signal processing while not losing the reference of the original Volterra regressors. Synergy of this approach exists with [22], where the covariance matrix was used to obtain a low computational complexity version of a direct-learning DPD. Note that the logic behind the algorithm versions remains the same; the evolution consists of different implementations of equivalent operations.

The remainder of this contribution is organized as follows. First, Section II introduces the notation and prerequisites of the work. Section III deals with the theoretical part of the RC-DOMP algorithm, followed by a computational complexity assessment in Section IV. Experimental design and results of the DPD of a class J PA working with 20- and 30-MHz 5G-NR signals are detailed in Section V. Finally, Section VI summarizes the main results and concludes this article.

II. Notation and Prerequisites

In this section, we review the fundamentals of the proposal and establish notational conventions that will be followed henceforth.

The baseband system measurement equation relates the complex envelope input of the signal $\mathbf{x}$ to its output $\mathbf{y}$ following the form:

$$\mathbf{y} = \mathbf{X}\mathbf{h}_{\mathbf{X}} + \mathbf{e} \tag{1}$$

where $\mathbf{y} = [y[q], y[q-1], \ldots, y[q-m+1]]^T$ is a buffered output set of m consecutive samples with discrete-time index q, $\mathbf{h}_{\mathbf{X}}$ is the vector holding the n coefficients of the model, $\mathbf{e}$ accounts for the measurement and modeling errors, and $\mathbf{X}$ is a centered design matrix defined as

$$\mathbf{X} = \begin{bmatrix} \boldsymbol{\xi}_1 & \boldsymbol{\xi}_2 & \cdots & \boldsymbol{\xi}_n \end{bmatrix} \in \mathbb{C}^{m \times n} \tag{2}$$

where $\boldsymbol{\xi}_i$ are the Volterra regressors, which are model-dependent, and in general will take the form of multiplications where the factors are delayed and/or conjugate versions of the input signal $\mathbf{x}$. The underlying Volterra model will shape the regressor structure, being amongst the most-general the FV, characterized by the input–output relationship

$$\boldsymbol{\xi}_i(\mathbf{x}) = \prod_{r_i=1}^{p_i+1} x[q-q_{r_i}] \prod_{r_i=p_i+2}^{2p_i+1} x^*[q-q_{r_i}] \tag{3}$$

where $2p_i + 1$ refers to the regressor order.

The classical theory of regression allows to perform an estimation of the Volterra kernel vector $\hat{\mathbf{h}}_{\mathbf{X}}$ through the well-known Moore–Penrose pseudoinverse matrix $\mathbf{X}^{\dagger}$ [23]

$$\hat{\mathbf{h}}_{\mathbf{X}} = \left(\mathbf{X}^H\mathbf{X}\right)^{-1}\mathbf{X}^H\mathbf{y} = \mathbf{X}^{\dagger}\mathbf{y} \tag{4}$$

that returns a complex value for each one of the components of $\hat{\mathbf{h}}_{\mathbf{X}}$.

The DOMP is an iterative technique that aims at choosing the regressor indicated by a local decision in each iteration k. The support set $S^{(k-1)}$ includes the selected coefficients up to the end of iteration $k-1$, and after a new regressor is chosen in the next iteration k, its index $s^{(k)}$ is included in $S^{(k-1)}$ to become $S^{(k)}$. Since one regressor is selected per iteration, the solution is also k-sparse, i.e., the estimation of the Volterra kernel vector has k nonzero entries at iteration k. A residual $\mathbf{r}^{(k)}$ is kept in the execution to account for the modeling error, following:

$$\mathbf{r}^{(k)} = \mathbf{y} - \hat{\mathbf{y}}^{(k)} \tag{5}$$

where $\hat{\mathbf{y}}^{(k)}$ is the output estimation provided by the Volterra kernel vector of k components. The DOMP chooses the regressor that exhibits a maximum correlation with the residual and performs the Gram–Schmidt orthogonalization of the selected vector with respect to the rest of the basis.

The reduction in computational complexity of the current proposal is achieved by working with the covariance matrices instead of the measurement matrices, operating over matrices of $n \times n$ dimension instead of $m \times n$, what eliminates the dependence of the regressor length.

III. Reduced-Complexity DOMP

The DOMP algorithm selects one Volterra component at each iteration k. This selection is performed indirectly in a linearly transformed subspace that in this work is denoted by $\mathbf{Z}^{(k)}$. The operation is conceptually simple; at the beginning of each iteration, the index $s^{(k)}$ that shows the highest correlation of the $s^{(k)}$th regressor with the residual is selected and consequently included in the support set $S^{(k-1)}$. After that, the algorithm orthogonalizes the unselected regressors to the selected component and normalizes the selected basis function so that the resulting basis is orthonormal. In this way, at each step k, the elements included in the support set are orthonormal, i.e., the rank of matrix $\mathbf{Z}^{(k)}$ is k and

$$\mathbf{Z}_{S^{(k)}}^{(k)^H}\mathbf{Z}_{S^{(k)}}^{(k)} = \mathbf{I} \tag{6}$$

where the subindex in $\mathbf{Z}_{S^{(k)}}^{(k)}$ represents the set of columns indicated by $S^{(k)}$. Also, the elements included in the support set and those that are not in it are mutually orthogonal. The transformed Volterra measurement matrix $\mathbf{Z}^{(k)}$ is composed of a set of transformed regressors $\boldsymbol{\phi}_i$

$$\mathbf{Z}^{(k)} = \begin{bmatrix} \boldsymbol{\phi}_1 & \boldsymbol{\phi}_2 & \cdots & \boldsymbol{\phi}_n \end{bmatrix} \in \mathbb{C}^{m \times n} \tag{7}$$

where the iteration-dependent superscript in $\boldsymbol{\phi}_i^{(k)}$ has been omitted for notation simplicity. The correlation matrix of the transformed measurement matrix is defined by

$$\mathbf{R}_{\mathbf{Z}^{(k)}} = \mathbf{E}[\mathbf{Z}^{(k)^H}\mathbf{Z}^{(k)}] \tag{8}$$

whose maximum likelihood estimator is its sample autocorrelation matrix

$$\mathbf{R}_{\mathbf{Z}^{(k)}} = \begin{bmatrix} \rho_{11} & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{21} & \rho_{22} & \cdots & \rho_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1} & \rho_{n2} & \cdots & \rho_{nn} \end{bmatrix} \in \mathbb{C}^{n \times n} \tag{9}$$

where the component cross correlation is defined as

$$\rho_{ij} = \boldsymbol{\phi}_i^H\boldsymbol{\phi}_j. \tag{10}$$

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

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

Fig. 1. DOMP—reduced complexity version pseudocode.

Following the same rationale, the cross correlation between the transformed measurement matrix and the output:

$$\mathbf{R}_{\mathbf{Z}^{(k)}\mathbf{y}} = \mathbf{E}[\mathbf{Z}^{(k)^H}\mathbf{y}] \tag{11}$$

is approximated by its estimator

$$\mathbf{R}_{\mathbf{Z}^{(k)}\mathbf{y}} = \begin{bmatrix}\gamma_1 & \gamma_2 & \cdots & \gamma_n\end{bmatrix}^T \in \mathbb{C}^{n\times 1} \tag{12}$$

where the sample cross correlation between the ith regressor and the output obeys

$$\gamma_i = \boldsymbol{\phi}_i^H\mathbf{y} \tag{13}$$

in which the iteration-dependent superscript is omitted again for simplicity.

Considering that the matrices defined earlier are sized by the number of model components, which is generally much lower than the number of samples in the regressors, the resulting algorithm shows a computational complexity that is fixed for every iteration compared with that of the previous versions of the algorithm [11], [13], [21] that required increasing computations with the number of selected coefficients. Sections III-A–III-F overview the steps of this contribution.

The technique is summarized in Fig. 1. Note that steps 10, 11, and 12, which correspond to obtaining the Volterra coefficients vector, the residual, and the modeled output, are not strictly necessary for the execution of the algorithm—these variables are not used in the algorithm loop—, so these may only be included when it is necessary to explicitly calculate the modeled signal at each iteration. The algorithm steps are detailed next.

Sample MATLAB code of the RC-DOMP algorithm in the modeling of the 20-MHz signal in Section V can be found in [24].

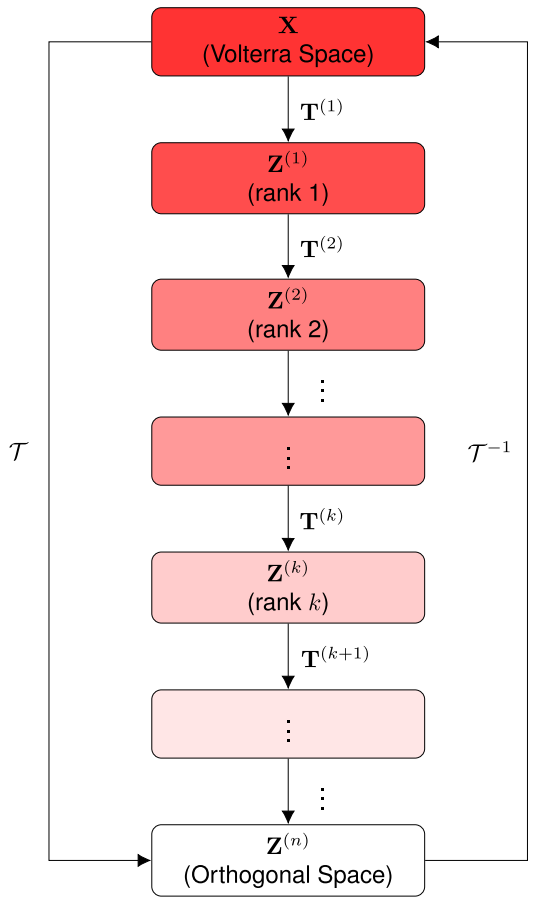


Fig. 2. Block diagram of the relation between the regressor and transformation matrices. The gradient shading in the blocks accounts for the orthogonalization process or whitening over the vector basis.

A. Initialization

Prior to the execution of the algorithm, the transformed measurement matrix is equal to the Volterra basis

$$\mathbf{Z}^{(0)} = \mathbf{X}. \tag{14}$$

The support set is empty since there are not yet components selected

$$S^{(0)} = \emptyset \tag{15}$$

and the residual $\mathbf{r}^{(k)}$, which keeps track of the modeling error in iteration k, is equal to the signal to model

$$\mathbf{r}^{(0)} = \mathbf{y}. \tag{16}$$

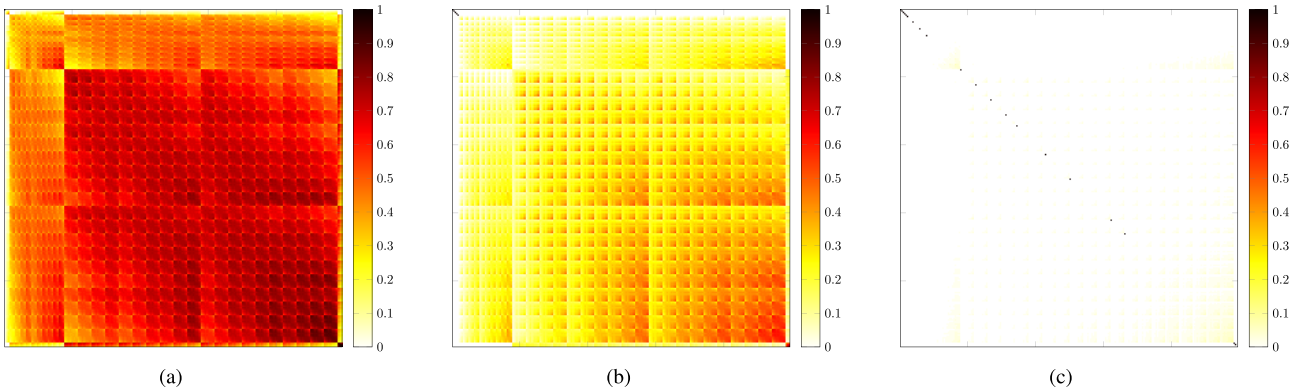
Consequently, since the model does not include any regressor, the estimation of the output prior to the beginning of the algorithm is

$$\hat{\mathbf{y}}^{(0)} = \mathbf{0}. \tag{17}$$

B. Regressor Selection

At the beginning of iteration k, the algorithm selects the component $s^{(k)}$ that maximizes the normalized cross correlation of the elements not included in the support set

$$s^{(k)} = \arg\max_{i \notin S^{(k)}} \frac{|\gamma_i|}{\sqrt{\rho_{ii}}} \tag{18}$$



(a) (b) (c)

Fig. 3. Autocorrelation matrix $\mathbf{R}_{\mathbf{Z}^{(k)}}$ displayed as an image for $n = 248$ in different iterations. It can be observed that the autocorrelation matrix tends to a diagonal, while the components are being orthogonalized. (a) $k = 1$. (b) $k = 5$. (c) $k = 20$.

and then, this index is added to the support set

$$S^{(k)} = S^{(k-1)} \cup s^{(k)}. \tag{19}$$

C. Orthogonalization

In the orthogonalization process, the not-yet-selected model basis functions are orthogonalized through Gram–Schmidt with respect to the selected regressor. For that, the transformation matrix $\mathbf{T}$ is defined

$$\mathbf{T}^{(k)} = \left(t_{ij}^{(k)}\right) \in \mathbb{C}^{n\times n} \tag{20}$$

whose elements t_{ij} for the ith row and the jth column follow the structure:

$$t_{ij}^{(k)} = \begin{cases} -\dfrac{\rho_{ij}}{\rho_{ii}}, & i = s^{(k)},\ j \notin S^{(k)} \quad (21) \\ \dfrac{1}{\sqrt{\rho_{ii}}}, & i = j = s^{(k)} \quad (22) \\ 1, & i = j \neq s^{(k)} \quad (23) \\ 0, & \text{elsewhere} \quad (24) \end{cases}$$

where (21) performs the orthogonalization of the unselected jth regressor and the regressor selected in the current iteration, (22) accounts for the normalization of the selected component, and (23) leaves the elements that belong to the support set unmodified. The transformation matrix allows to perform the orthogonalization and normalization processes in one matrix multiplication through

$$\mathbf{Z}^{(k)} = \mathbf{Z}^{(k-1)}\mathbf{T}^{(k)}. \tag{25}$$

Fig. 2 shows the relation between the transformation and correlation matrices. The accumulated transformation matrix $\mathcal{T}$ is defined in the Appendix. For the sake of illustration, the autocorrelation matrix $\mathbf{R}_{\mathbf{Z}^{(k)}}$ for a GMP model of 248 components is plotted as an image in Fig. 3 for the first, fifth, and twentieth iterations. The model structure can be organized as follows; there exist three blocks, being the first four regressors the first order, the second block of 40 regressors the third order, the next block of 200 coefficients the fifth order, and the last four coefficients correspond to orders seven to thirteen. In $k = 1$, the correlation between the regressors is clearly observed by the high values exhibited by the nondiagonal terms of the matrix. In the following iterations, the selected regressors are normalized and not correlated with the rest, showing a value of 1 in the diagonal and being 0 in the rest of its column and row. Finally, at iteration $k = 20$, the model structure is almost lost since it has been captured by the selected regressors, evidenced by the whitening of the matrix when only 20 regressors have been included in the model.

D. Output Correlation Update

The cross correlation between the nonselected regressors of matrix $\mathbf{Z}$ at iteration k and the output is updated by considering the orthogonality imposed at the previous step so that

$$\gamma_i^{(k)} = \begin{cases} \gamma_i^{(k-1)} - \left(\dfrac{\rho_{s^{(k)}i}}{\rho_{s^{(k)}s^{(k)}}}\right)^* \gamma_{s^{(k)}}^{(k-1)}, & i \notin S^{(k)} \quad (26) \\ \dfrac{\gamma_i^{(k-1)}}{\sqrt{\rho_{ii}}}, & i = s^{(k)} \quad (27) \\ \gamma_i^{(k-1)}, & \text{elsewhere} \quad (28) \end{cases}$$

is the corresponding transformation in the output correlation.

E. Autocorrelation Update

Since the basis is multiplied by the transformation matrix, the autocorrelation update can be obtained through the relation $(\mathbf{ZT})^H\mathbf{ZT} = \mathbf{T}^H\mathbf{Z}^H\mathbf{ZT}$. Hence

$$\mathbf{R}_{\mathbf{Z}^{(k)}} = \mathbf{T}^{(k)^H}\mathbf{R}_{\mathbf{Z}^{(k-1)^H}}\mathbf{T}^{(k)}. \tag{29}$$

F. Regression

All the regressors in the support set are orthogonal among themselves; therefore, the regression—that is, the estimation of the transformed Volterra kernel vector—consists of a simple selection of the projections of $\mathbf{y}$ on this space, which is already calculated

$$\hat{\mathbf{h}}^{(k)} = \gamma_{S^{(k)}}. \tag{30}$$

Note that this estimation corresponds to the basis $\mathbf{Z}^{(k)}$. Since the regression is performed on an orthonormal basis, there is

no need of using the pseudoinverse, hence reducing computational complexity. The Volterra kernel in the original Volterra subspace can be directly recovered through

$$\hat{\mathbf{h}}_{\mathbf{X}} = \mathbf{T}^{(1)}\mathbf{T}^{(2)} \cdots \mathbf{T}^{(k)}\hat{\mathbf{h}}^{(k)}. \tag{31}$$

The residual is updated by subtracting the contribution of the last selected component

$$\mathbf{r}^{(k)} = \mathbf{r}^{(k-1)} - \mathbf{Z}^{(k)}_{s^{(k)}}\gamma_{s^{(k)}} \tag{32}$$

and the estimation of the model output is attained by adding the same contribution

$$\hat{\mathbf{y}}^{(k)} = \hat{\mathbf{y}}^{(k-1)} + \mathbf{Z}^{(k)}_{s^{(k)}}\gamma_{s^{(k)}}. \tag{33}$$

Observe that the estimation is performed incrementally by adding a term to the value of the preceding iteration. This fact is key in the reduction of computational complexity against other algorithms that perform a full estimation per iteration.

IV. Complexity Assessment

In this section, we perform a computational complexity comparison of the three versions of the DOMP algorithm. The first advantage of the proposed method is observed in terms of memory requirements since moving the operations to the correlation domain removes the dependence on the time index and the vector sizes are much shorter. Compared with previous versions of the algorithm, the RC-DOMP works over the correlation matrix whose dimensions are $n \times n$ and the rest work with the measurement matrix characterized by dimensions of $m \times n$. Since the equation system that is being solved is overdetermined, that is, $m > n$, storage of the variables is highly improved.

Regarding running complexity, this assessment is performed by using the Bachmann–Landau notation in terms of the number of arithmetic multiplications, the latter being the one that has more impact on the overall complexity [25], required for each algorithm operation. The different versions of the DOMP algorithm are characterized by a computational complexity per iteration of

$$\mathcal{O}_{\text{iter,DOMP}} = k^3 + 4mk^2 - 4kmn + nm^2 + 2mn^2 \tag{34}$$

$$\mathcal{O}_{\text{iter,SSPI DOMP}} = 2mk^2 - 4kmn + nm^2 + 2mn^2 \tag{35}$$

$$\mathcal{O}_{\text{iter,RC-DOMP}} = 4n^3 \tag{36}$$

where $\mathcal{O}$ stands for the quantity of real-number multiplications at iteration k. It can be observed that the original DOMP [11] exhibits a complexity dominated by the cubic of the iteration number that results of the $m \times k$ pseudoinverse calculation at each iteration; in the SSPI DOMP [21], this relationship is reduced to quadratic dependence, and in the novel RC-DOMP, there is no dependence with the iteration number. Considering that m/n is the number of samples per regressor, it is straightforward to conclude that the RC-DOMP outperforms the rest of the algorithms under comparison. The relation between their total computational complexities follows:

$$\mathcal{O}_{\text{total,DOMP}} > \mathcal{O}_{\text{total,SSPI DOMP}} > \mathcal{O}_{\text{total,RC-DOMP}}. \tag{37}$$

In addition, The DOMP and SSPI DOMP methods can only be partially implemented in parallel since the inverse operation is difficult to implement in an FPGA or other hardware [22]. However, the proposed method can be implemented in parallel computing since the inverse operation is removed. Moreover, power consumption is reduced as a consequence of lowering computational complexity.

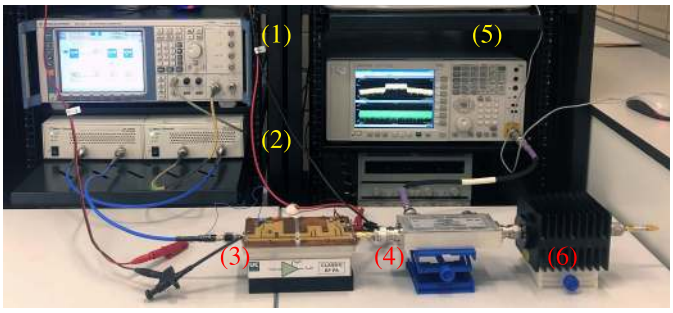


Fig. 4. Picture of the experimental setup composed of (1) VSG, (2) preamplifiers, (3) PA, (4) coupler, (5) VSA, and (6) power load.

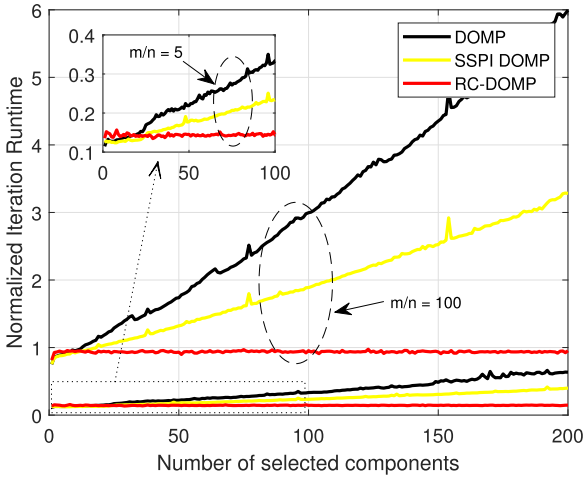


Fig. 5. Evolution of runtime per iteration with the number of selected components for a number of samples per component m/n of 5 and 100. Runtime is normalized with respect to the RC-DOMP execution of $m/n = 10$.

V. Experimental Results and Discussion

In this section, the proposed technique is benchmarked against its previous versions regarding their runtime, and the results are discussed considering the computational complexity assessment provided in Section IV. Since this contribution is intended to provide an equivalent formulation of the DOMP algorithm but with a lower computational complexity, the comparison in terms of pruning capabilities with other techniques would show the same results as those provided in [13]. In addition, for the sake of completeness of this work, the RC-DOMP approach is applied to the DPD of a class J PA working under two different bandwidths in order to highlight the details that differ in the predistorter pruning for both cases.

The experimental test bench, whose picture is shown in Fig. 4, was composed of an SMU200A vector signal generator (VSG) from Rohde & Schwarz, a PXA-N9030A vector signal analyzer (VSA) from Keysight Technologies

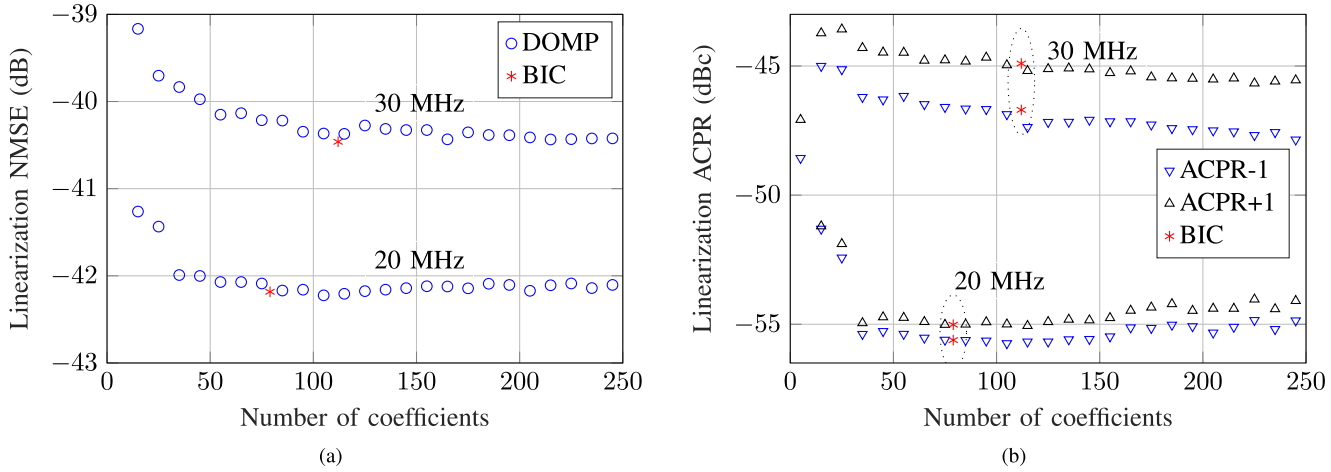


Fig. 6. Evolution of (a) linearization NMSE and (b) ACPR with the number of DPD coefficients sorted by their importance up to $n = 250$ components for signal bandwidths of 20 and 30 MHz.

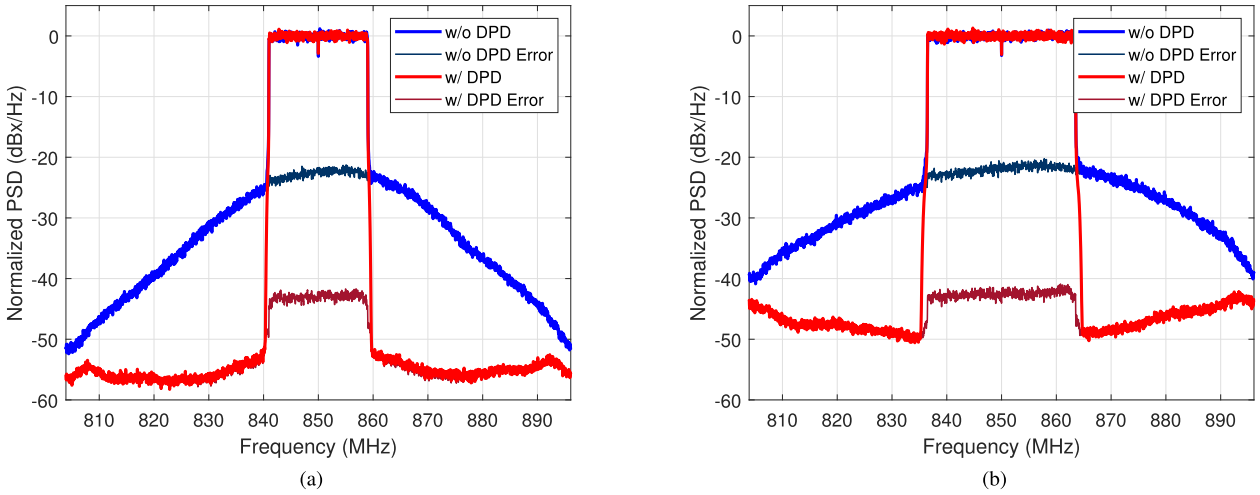


Fig. 7. Normalized PSDs of the output and error signals with and without DPD working with (a) 20- and (b) 30-MHz 5G-NR signals.

TABLE I
PARAMETER CONFIGURATION OF THE MODEL UNDER TEST

Parameter	Value
$\mathcal{K}_a$	$[0:16]$
$\mathcal{L}_a$	$[15\,15\,\cdots 15]$
$\mathcal{K}_b = \mathcal{K}_c$	$[1:16]$
$\mathcal{L}_b = \mathcal{M}_b = \mathcal{L}_c = \mathcal{M}_c$	$[5\,5\cdots 5]$

and a dc power supply. The instrumentation was controlled through LAN by using Standard Commands for Programmable Instruments (SCPI) from a PC with MATLAB. The device under test (DUT) was the cascade of two Mini-Circuits TVA-4W-422A+ preamplifiers and a continuous-mode class J PA [26] based on the Cree CGH35015F transistor working in the band of 800–900 MHz. The operating point was set to an average output power of +30.8 dBm (+41 dBm of peak output power), and it is characterized by a gain compression of 3.5 dB.

The probing signals were designed according to the 5G-NR standard with the bandwidths of 20 and 30 MHz. Different sets of signals were used for identification and pruning of the DPD model and performance validation. The signals exhibited a peak-to-average power ratio (PAPR) of 11 dB and contained over 360 000 samples corresponding to a sampling frequency of 92.16 MSample/s.

In the VSA, the output RF signal was downconverted to baseband by setting the appropriate range, span, and sampling rate in order to recover it. The measurement dynamic range was optimized by averaging 100 acquisitions of the output signal. Finally, the measured signal was time-aligned in order to synchronize the input and output data sets.

The model under test was a composite GMP of 17$^{\text{th}}$ order comprising its three parts, including both even and odd powers in the envelope. The parameter configuration is shown in Table I. On top of that, the conjugate of the linear regressors, i.e., regressors of the image signal, were added to the model in order to mitigate linear I/Q imbalance [27]–[29]. These regressors are based on widely linear [30] and widely nonlinear signal processing [31]. The regressor pool with this

TABLE II

MEASURED PERFORMANCE IN TERMS OF NMSE, ACPR, EVM, AND NUMBER OF COEFFICIENTS FOR THE PA WITH AND WITHOUT DPD

Experiment	NMSE (dB)	ACPR-1 (dBc)	ACPR+1 (dBc)	EVM (%)	Number of coef.	PAE (%)
20 MHz w/o DPD	-20.2	-30.1	-27.7	7.9	-	25
20 MHz w/ DPD	-42.2	-55.6	-55.0	1.0	79	25
30 MHz w/o DPD	-19.8	-29.9	-26.4	9.7	-	25
30 MHz w/DPD	-40.5	-47.6	-45.9	1.0	112	25

configuration reaches a total number of 1248 components. The reason why the model to prune has this intricate structure is justified by the PA behavior, which exhibits strong nonlinearity evidenced by gain expansion followed by compression and a mild rotation in the output constellation. The model equation follows:

$$\begin{aligned} y[q] = &\sum_{k=0}^{\mathcal{K}_a}\sum_{l=0}^{\mathcal{L}_a} a_{kl}x[q-l]|x[q-l]|^k \\ &+ \sum_{k=1}^{\mathcal{K}_b}\sum_{l=0}^{\mathcal{L}_b}\sum_{u=1}^{\mathcal{M}_b} b_{klu}x[q-l]|x[q-l-u]|^k \\ &+ \sum_{k=1}^{\mathcal{K}_c}\sum_{l=0}^{\mathcal{L}_c}\sum_{u=1}^{\mathcal{M}_c} c_{klu}x[q-l]|x[q-l+u]|^k \\ &+ \sum_{l=0}^{\mathcal{L}_a(1)} d_l x^*[q-l]. \end{aligned} \tag{38}$$

To emphasize the enhancement in computational complexity and validate Section IV, the iteration runtime in the first 200 iterations was measured in a desktop PC. The results, which are shown in Fig. 5, validate those previously reported in [21], where it was highlighted that the SSPI decreases the runtime of the OMP and DOMP techniques. Now, the RC-DOMP shows a constant computational complexity per iteration, which results in the lowest computational complexity of the benchmark. Results are shown for a number of samples per component of 5 and 100 that represent two opposite ends in choosing this ratio.

The pruning technique was executed over the DPD model in an indirect learning scheme to gather the 250 most relevant components of the model sorted by their importance. A DPD was calculated for each number of components, allowing to obtain the linearization normalized mean-squared error (NMSE), adjacent channel power ratio (ACPR), and error vector magnitude (EVM) as performance indicators. In addition, the Bayesian information criterion (BIC) was calculated over the model evolutions to indicate the optimum number of coefficients for each case [32]. Fig. 6 shows a decreasing error evolution with the number of components within the DPD as it is to be expected. Note that the required number of coefficients for achieving linearization is much higher than previous experiments in the literature (see [13]) due to the strong nonlinearities of the DUT. The difference in linearization results for the 20- and 30-MHz cases is based on having a greater effective oversampling for the 20-MHz signal and the PA frequency response, which is considerably more nonlinear in the 30-MHz bandwidth. The numerical value of the performance indicators without DPD and with DPD composed of the optimum number of coefficients is given in Table II. The linearity metrics without DPD are equivalent for both signal bandwidths, except for the EVM, where the bandwidth difference affects the in-band nonlinearity. The optimum number of components given by the BIC was 79 for the 20-MHz signal and 112 for the bandwidth of 30 MHz, which results in a reduction of 90% in a number of coefficients. The linearization with DPD exhibits a better behavior in the 20-MHz signal despite having an equivalent model than that with the 30-MHz signal. Fig. 7 shows the normalized power spectral densities (PSDs) of the output without DPD and with the BIC optimum DPD for both bandwidths. Finally, the amplitude-modulation-to-amplitude-modulation (AM/AM) and amplitude-modulation-to-phase-modulation (AM/PM) characteristics of the PA with and without DPD for the 30-MHz case are shown in Fig. 8. Note that the DUT gain corresponds to the cascade of two preamplifiers and a PA.

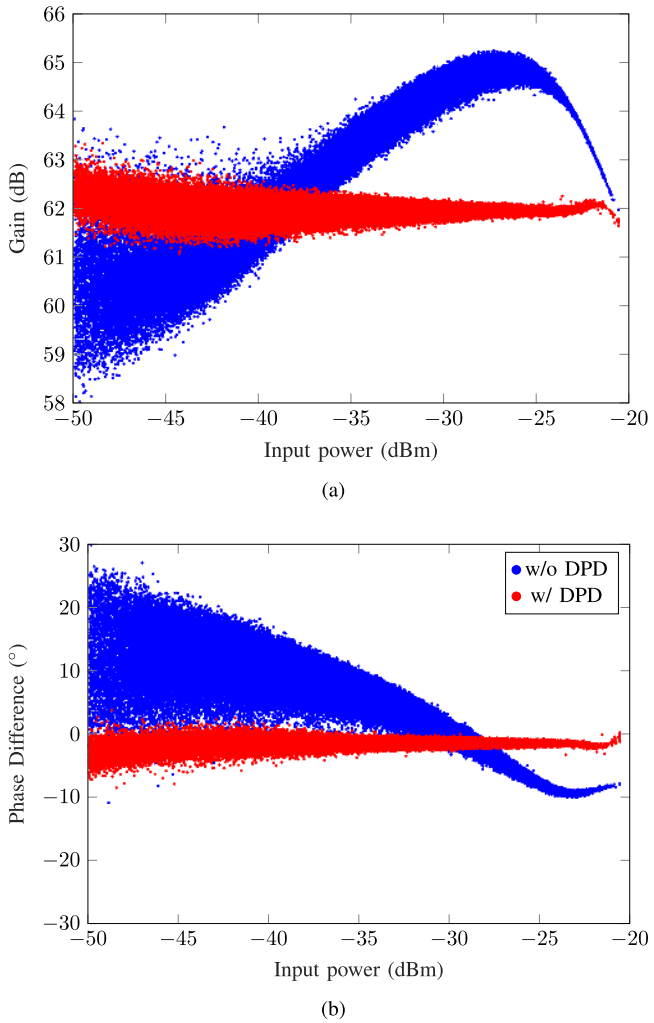


Fig. 8. Dynamic (a) gain and (b) AM/PM characteristics of the PA operating over a 30-MHz 5G-NR probe signal, with and without DPD. The number of coefficients of the DPD was given by the BIC rule, which obtained the optimal number of 112.

VI. Conclusion

In this article, an equivalent formulation of the DOMP algorithm has been proposed. It exploits the synergy of having the regressors in an orthogonal vector space to perform all the operations over the correlation matrices. Since the selected regressors are orthogonal, the calculation of the pseudoinverse matrix and its multiplication over the output to estimate the Volterra coefficients turns into a simple projection that transforms the regression into just a selection of coefficients in a vector. Also, a transformation matrix from the Volterra space to the equivalent orthogonal domain is defined as a result of the algorithm, enabling orthogonal signal processing while keeping the reference of the original Volterra regressors.

Experimental results showed the reduction in complexity while still having the same performance as the original algorithm. The whitening of the correlation matrix and sorting and classification of the selected regressors by the algorithms along with the modeling capabilities make DOMP a forward-looking proposal in the pruning of DPDs.

Appendix

Let $\mathcal{T}$ be the matrix that considers the transformations made at each step of the algorithm

$$\mathcal{T} = \mathbf{T}^{(1)}\mathbf{T}^{(2)}\cdots\mathbf{T}^{(n)}. \tag{39}$$

Following this definition, the Volterra matrix $\mathbf{X}$ and the linearly transformed orthonormal matrix $\mathbf{Z}$ are related by:

$$\mathbf{Z}^{(n)} = \mathbf{X}\mathcal{T} \tag{40}$$

which is invertible to

$$\mathbf{X} = \mathbf{Z}^{(n)}\mathcal{T}^{-1}. \tag{41}$$

This transformation allows to work simultaneously in the Volterra space and in an equivalent orthogonal space in which the elements in $S^{(n)}$ are sorted by their contribution to the model output.

Acknowledgment

The authors would like to thank the team of Prof. José Ángel García at the Universidad de Cantabria for providing the PA used in the experimental part of this article.

References

[1] J. Wood, *Behavioral Modeling and Linearization of RF Power Amplifiers*. Norwood, MA, USA: Artech House, 2014.

[2] V. Mathews and G. Sicuranza, *Polynomial Signal Processing*. Hoboken, NJ, USA: Wiley, 2000.

[3] 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, 2001.

[4] D. R. Morgan, Z. Ma, J. Kim, M. G. 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.

[5] J. Reina-Tosina, M. Allegue-Martinez, C. Crespo-Cadenas, C. Yu, and S. Cruces, "Behavioral modeling and predistortion of power amplifiers under sparsity hypothesis," *IEEE Trans. Microw. Theory Techn.*, vol. 63, no. 2, pp. 745–753, Feb. 2015.

[6] J. Peng, S. He, Z. Dai, and B. Wang, "A simplified sparse parameter identification algorithm suitable for power amplifier behavioral modeling," *IEEE Microw. Wireless Compon. Lett.*, vol. 27, no. 3, pp. 290–292, Mar. 2017.

[7] S. Chen, S. A. Billings, and W. Luo, "Orthogonal least squares methods and their application to non-linear system identification," *Int. J. Control*, vol. 50, no. 5, pp. 1873–1896, Nov. 1989.

[8] S. Maymon and Y. C. Eldar, "The Viterbi algorithm for subset selection," *IEEE Signal Process. Lett.*, vol. 22, no. 5, pp. 524–528, May 2015.

[9] A. Hashemi and H. Vikalo, "Sparse linear regression via generalized orthogonal least-squares," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Dec. 2016, pp. 1305–1309.

[10] A. Hashemi and H. Vikalo, "Accelerated orthogonal least-squares for large-scale sparse reconstruction," *Digit. Signal Process.*, vol. 82, pp. 91–105, Nov. 2018.

[11] J. A. Becerra, M. J. Madero-Ayora, J. Reina-Tosina, C. Crespo-Cadenas, J. Garcia-Frias, 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.

[12] T. Blumensath and M. E. Davies, "On the difference between orthogonal matching pursuit and orthogonal least squares," Univ. Edinburgh, Edinburgh, U.K., Tech. Rep., Mar. 2007.

[13] 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 Trans. Microw. Theory Techn.*, vol. 67, no. 9, pp. 3575–3585, Sep. 2019.

[14] D. Lopez-Bueno, Q. A. Pham, G. Montoro, and P. L. Gilabert, "Independent digital predistortion parameters estimation using adaptive principal component analysis," *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 12, pp. 5771–5779, Dec. 2018.

[15] S. Wang, M. Abi Hussein, O. Venard, and G. Baudoin, "Optimal sizing of two-stage cascaded sparse memory polynomial model for high power amplifiers linearization," *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 9, pp. 3958–3965, Sep. 2018.

[16] N. Guan, N. Wu, and H. Wang, "Model identification for digital predistortion of power amplifier with signed regressor algorithm," *IEEE Microw. Wireless Compon. Lett.*, vol. 28, no. 10, pp. 921–923, Oct. 2018.

[17] N. Kelly and A. Zhu, "Direct error-searching SPSA-based model extraction for digital predistortion of RF power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 3, pp. 1512–1523, Mar. 2018.

[18] Q. A. Pham, D. Lopez-Bueno, T. Wang, G. Montoro, and P. L. Gilabert, "Partial least squares identification of multi look-up table digital predistorters for concurrent dual-band envelope tracking power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 12, pp. 5143–5150, Dec. 2018.

[19] Q. A. Pham, G. Montoro, D. Lopez-Bueno, and P. L. Gilabert, "Dynamic selection and estimation of the digital predistorter parameters for power amplifier linearization," *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 10, pp. 3996–4004, Oct. 2019.

[20] C. Yu, K. Tang, and Y. Liu, "Adaptive basis direct learning method for predistortion of RF power amplifier," *IEEE Microw. Wireless Compon. Lett.*, vol. 30, no. 1, pp. 98–101, Jan. 2020.

[21] 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," in *Proc. IEEE Top. Conf. RF/Microw. Power Modeling Radio Wireless Appl. (PAWR)*, Jan. 2019, pp. 1–3.

[22] Z. Wang, W. Chen, G. Su, F. M. Ghannouchi, Z. Feng, and Y. Liu, "Low computational complexity digital predistortion based on direct learning with covariance matrix," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 11, pp. 4274–4284, Nov. 2017.

[23] A. Ben-Israel and T. N. Greville, *Generalized Inverses*. New York, NY, USA: Springer-Verlag, 2003.

[24] J. A. Becerra, M. J. Madero-Ayora, J. Reina-Tosina, and C. Crespo-Cadenas, "Reduced complexity doubly orthogonal matching pursuit (RC-DOMP) sample code," CodeOcean Capsule, Univ. Seville, Seville, Spain, Tech. Rep., Jul. 2020, doi: 10.24433/CO.1589600.v1.

[25] P. Bürgisser, M. Clausen, and A. Shokrollahi, *Algebraic Complexity Theory*. Berlin, Germany: Springer, Jan. 1997.

[26] M. J. Madero-Ayora, M. Allegue-Martinez, J. A. Garcia, J. Reina-Tosina, and C. Crespo-Cadenas, "Linearization and EVM enhancement of an efficient class j amplifier for 3G and 4G mobile communication signals," in *Proc. Workshop Integr. Nonlinear Microw. Millimetre-Wave Circuits*, Sep. 2012, pp. 1–3.

[27] B. Schubert, A. Gokceoglu, 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 *Proc. IEEE Global Conf. Signal Inf. Process.*, Dec. 2013, pp. 1162–1165.
[28] C. Crespo-Cadenas, M. J. Madero-Ayora, J. Reina-Tosina, and J. A. Becerra-Gonzalez, "A widely nonlinear approach to compensate impairments in I/Q modulators," in *Proc. Eur. Microw. Conf. (EuMC)*, Sep. 2015, pp. 506–509.
[29] M. Abdelaziz, L. Anttila, and M. Valkama, "Reduced-complexity digital predistortion for massive MIMO," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2017, pp. 6478–6482.
[30] B. Picinbono and P. Chevalier, "Widely linear estimation with complex data," *IEEE Trans. Signal Process.*, vol. 43, no. 8, pp. 2030–2033, Aug. 1995.
[31] C. Crespo-Cadenas, M. J. Madero-Ayora, J. Reina-Tosina, and J. A. Becerra-González, "Formal deduction of a volterra series model for complex-valued systems," *Signal Process.*, vol. 131, pp. 245–248, Feb. 2017.
[32] C. Crespo-Cadenas, M. J. Madero-Ayora, J. Reina-Tosina, and J. A. Becerra-Gonzalez, "Transmitter linearization adaptable to power-varying operation," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 10, pp. 3624–3632, Oct. 2017.

Juan A. Becerra (Senior Member, IEEE) received the B.S. and M.Sc. degrees in telecommunication engineering from the Universidad de Sevilla, Seville, Spain, in 2009 and 2012, respectively, the Ph.D. degree in electrical and computer engineering from the University of Delaware, Newark, DE, USA, in 2017, and the Ph.D. degree in telecommunication engineering from the Universidad de Sevilla in 2019.

Since 2017, he has been with the Department of Signal Theory and Communications, Universidad de Sevilla, where he is currently a Contract Professor. His main research areas include behavioral modeling and linearization of power amplifiers, and compressed-sensing signal processing.

María José Madero Ayora (Senior Member, IEEE) received the M.Sc. and Ph.D. degrees in telecommunication engineering from Universidad de Sevilla, Seville, Spain, in 2002 and 2008, respectively.

Since 2003, she has been with the Department of Signal Theory and Communications, Universidad de Sevilla, where she is currently an Associate Professor. Her current research interests include nonlinear analysis of active microwave devices, compensation of impairments in modulators and power amplifiers, and measurement techniques of nonlinear communication systems.

Javier Reina-Tosina (Senior Member, IEEE) was born in Seville, Spain. He received the Telecommunication Engineering and Ph.D. degrees from the Universidad de Sevilla, Seville, in 1996 and 2003, respectively.

Since 1997, he has been with the Department of Signal Theory and Communications, Universidad de Sevilla, where he is currently an Associate Professor. His research interests include the nonlinear analysis of active microwave devices, behavioral modeling of power amplifiers and linearization techniques, the integration of information technologies in biomedicine, intelligent devices for homecare, and bioelectromagnetics.

Carlos Crespo-Cadenas (Senior Member, IEEE) was born in Madrid, Spain. He received the degree in physics from the University of Havana, Havana, Cuba, in 1973, and the Ph.D. degree from the Polytechnique University of Madrid, Madrid, in 1995.

Since 1995, he has been with the Department of Signal Theory and Communications, Universidad de Seville, where he is currently a Professor. His research interests include the communication systems, nonlinear analysis of active microwave devices, power amplifier behavioral modeling, and linearization techniques.