Comparative Analysis of Greedy Pursuits for the Order Reduction of Wideband Digital Predistorters

Juan A. Becerra, *Member, IEEE*, María J. Madero-Ayora, *Member, IEEE*, and Carlos Crespo-Cadenas, *Senior Member, IEEE*

***Abstract*— This paper provides a review of greedy pursuits for optimizing Volterra-based behavioral models structure and estimating its parameters. An experimental comparison of the digital predistortion (DPD) linearization performance achieved by these approaches for model-order reduction, such as compressive sampling matching pursuit (CoSaMP), subspace pursuit (SP), orthogonal matching pursuit (OMP), and the novel doubly OMP (DOMP), is presented. A benchmark of the techniques in the DPD of a commercial class AB power amplifier (PA) and a class J PA operating over a 15-MHz Long-Term Evolution (LTE) signal is presented, giving a clear overview of their pruning characteristics in terms of linearization indicators and regressor selection capabilities. In addition, the benchmark is run in a cross-validation scheme by identifying the DPD with a 30-MHz 5G-new radio (NR) signal and validating with the same signal and a 20-MHz multicarrier wideband code division multiple access (WCDMA) signal. The DOMP is shown to be a promising technique since it achieves an enhanced model-order reduction for a similar linearization performance and precision.**

***Index Terms*— Adaptive algorithms, greedy algorithms, modeling, nonlinear systems, power amplifiers (PAs), predistortion, Volterra series.**

I. Introduction

EVERY generation of wireless communication systems brings new modulation schemes, higher bandwidths, and more challenging constrains for the transceivers. Many efforts have been devoted to the design of transmitters, in general, and to the power amplifier (PA) in particular, the latter being the most critical subsystem in terms of power efficiency and linearity.

Obtaining an accurate prediction of the nonlinear distortion produced by the PA with modern communication signals can be computationally too expensive if a transistor level description is adopted. Among different system level approaches that have been proposed to cope with these simulations, behavioral models based on Volterra series are quite popular [1], [2]. These Volterra-based behavioral models serve also as a basis for digital predistortion (DPD), an appealing linearization technique with the evolution of the processing capacity of electronic devices [3].

The Volterra series consists of a sum of multidimensional convolutions and can be seen as a generalization of a power series representation with memory, which is limited to a finite length for practical reasons. Since communication systems are bandpass, it is common to express the input–output relationship in terms of the signals complex envelopes. When the interest is placed on the fundamental frequency zone, as it is the case for the PA modeling, only odd-order terms are justified in the Volterra-based behavioral models. The baseband model derived for the fundamental frequency output is referred to as the full Volterra (FV) model [4] in order to distinguish it from other pruned Volterra representations with a simplified structure. Another general model is the complex-valued Volterra series (CVS) model presented in [5] as an extension of the Volterra series for the case of a nonlinear system with complex-valued signals.

However, the more general the considered behavioral model is, the larger the number of coefficients is included in its kernel structure. Pruning has been dealt with different approaches. In many cases, *ad hoc* pruning strategies are applied without knowledge of the internal structure of the PA, based on an *a priori* selection. This is the case of the widely adopted memory polynomial (MP) [6] and generalized MP (GMP) models [2]. Observe that the "even-order envelope power" regressors used in MP and GMP are not even-order terms. For example, $|x|^3x = (|x|^2x)|x|$ is not an homogeneous fourth-order term, because it is the product of the third-order term $|x|^2x$ and the nonlinear function $|x|$. These beneficial terms are also odd-order and can be explained from a Volterra series perspective. Near-diagonality pruning was first introduced in [7], producing a first concept of which components of the model are more relevant.

Recently, several *a posteriori* techniques for model order reduction based on signal processing have been proposed in the literature in order to tackle this issue, relying on the sparsity assumption for the kernels of the model. The least absolute shrinkage and selection operator (LASSO), which provides a

Manuscript received January 31, 2019; revised May 12, 2019; accepted June 13, 2019. Date of publication August 14, 2019; date of current version September 4, 2019. This research was funded by Ministerio de Economía, Industria y Competitividad (MINECO) of the Government of Spain, grant number TEC2017-82807-P and by the European Regional Development Fund (ERDF) of the European Commission. This article is an expanded version from the IEEE Latin America Microwave Conference 2018, Arequipa, Peru, December 12–14, 2018. *(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.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.2019.2928290

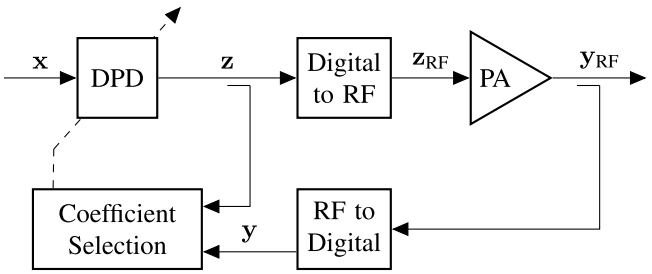


Fig. 1. Block diagram of an indirect learning DPD system.

sparse solution through the minimization of the ℓ_1-norm of the error, was first introduced in PA modeling in [8]. The principal component analysis (PCA), which enables the pruning in an orthogonal subspace, was first employed for DPD in [9] and recently given in an iterative fashion [10]. Pruning by the delay characteristics of the regressors is performed in [11], and a heuristic order-reduction method was introduced in [12].

Several algorithms within the family of greedy pursuits, which are a subset of techniques in the compressed-sensing field, have been also applied to DPD, as the compressive sampling matching pursuit (CoSaMP) [13] and a related technique, the subspace pursuit (SP) [14]. The basis pursuit was used for reducing the complexity of an MIMO DPD in [15]. The orthogonal matching pursuit (OMP) was first applied to PA DPD in [16], followed by a reduced-complexity version of the algorithm in [17]. To overcome the high correlation that appears in Volterra series, the doubly OMP (DOMP) was designed [18] as well as its low-complexity variant [19].

In [20], a review of widely adopted Volterra-based behavioral models and techniques for optimizing the model structure and estimating its parameters was performed. In this work, we focus on greedy algorithms applied to the order reduction of Volterra-based models, providing a standard formulation and a discussion of their similarities and differences. The experimental part covers the use of this set of techniques over an FV and GMP DPDs in three scenarios and evaluates their linearization performance.

The remainder of this contribution is organized as follows. First, Section II introduces the commons of Volterra models and regression. Section III deals with the theoretical part and discussion of greedy algorithms followed by a complexity assessment in Section IV. DPD experimental design and results are detailed in Section V. Finally, Section VI summarizes the main results and concludes this paper.

II. Framework

A general schematic of a DPD system with coefficient selection is shown in Fig. 1. Considering that the input and the output of the system are the vectors $\mathbf{x}$ and $\mathbf{y}$, constructed with the corresponding complex envelope samples, the objective of linearization is to compensate the resulting nonlinear distortion by processing $\mathbf{x}$ with a DPD before it is applied to the PA input. The success of the system linearization rests on a good model approach to the system, a search of the significant terms of the model with the consequent model-order reduction, and a precise procedure to estimate the model parameters.

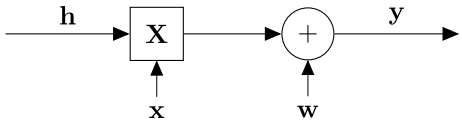


Fig. 2. Block diagram of Volterra series represented as a measurement process where $\mathbf{h}$ is the vector of Volterra coefficients to be estimated.

In the modeling of the nonlinear response of a PA, the aim is to find the mathematical relation $f(\cdot)$ between the input $x[q]$ and the output $y[q]$

$$y[q] = f(x[q]) + \epsilon[q] \tag{1}$$

where q is the discrete-time index and ϵ is the uncaptured error by the model. Although there exists an ample literature on how this behavior can be modeled [21]–[24]; in this work, we put the spotlight on Volterra series, in which the output is a linear combination of transformations $\phi_i(\cdot)$ of the input signal

$$y[q] = \sum_i h_i \phi_i(x[q]) + \epsilon \tag{2}$$

where h_i is the ith Volterra coefficient and $\phi_i(x[q])$ is its corresponding regressor or component. As an example of regressor structure, the most general is the FV that follows:

$$\phi_i(x[q]) = \prod_{r=1}^{p+1} x[q-q_r] \prod_{r=p+2}^{2p+1} x^*[q-q_r]. \tag{3}$$

The convenience of Volterra series resides in the fact that although the relation between the input and the output is inherently nonlinear, the Volterra regressors are linear to the output in the Volterra subspace. There exist other structures that exhibit these characteristics, such as the decomposed vector rotation (DVR) [25]. This linearity enables the model identification through all the classic signal processing, from the well-known least-squares (LS) technique to the most advanced sparse regression algorithms, allowing a diverse set of techniques that deal with how to extract the relevant coefficients. Representing (4) in its matrix form, we obtain the measurement equation

$$\mathbf{y} = \mathbf{X}\mathbf{h} + \epsilon \tag{4}$$

where $\mathbf{y} = [y[q], y[q-1], \ldots, y[q-m+1]]^T$ is a buffered output set of m consecutive samples, $\mathbf{X} \in \mathbb{C}^{m \times n}$ is the measurement matrix, which contains a regressor of size m per column, and $\mathbf{h} \in \mathbb{C}^n$ is the vector holding the n coefficients of the model. Following Fig. 2, matrix $\mathbf{X}$ can be interpreted as a vectorial basis (or more formally, since their columns are nonorthogonal, this is called a *frame* [22], [26]) in which $\mathbf{h}$ defines how much of each regressor is needed to represent the output $\mathbf{y}$.

The regression of the Volterra coefficients can be performed by minimizing the quadratic norm of the error. The estimation of the Volterra vector $\hat{\mathbf{h}}$ follows the normal equation:

$$\hat{\mathbf{h}} = (\mathbf{X}^H\mathbf{X})^{-1}\mathbf{X}^H\mathbf{y} = \mathbf{X}^\dagger\mathbf{y} \tag{5}$$

where H is the Hermitian transpose operator. The Moore–Penrose pseudoinverse $\mathbf{X}^{\dagger}$ computes the LS solution.

Regarding the structure of the measurement matrix $\mathbf{X}$, the baseband Volterra input–output relationship is

$$y[q] = \sum_{p=0}^{P} \sum_{\mathbf{q}_{2p+1}=\mathbf{0}}^{\mathbf{Q}_{2p+1}} h_{2p+1}[\mathbf{q}_{2p+1}] \times \prod_{r=1}^{p+1} x[q-q_r] \prod_{r=p+2}^{2p+1} x^*[q-q_r] \quad (6)$$

where $2P+1$ is the (odd) nonlinear order. The main disadvantage of the FV model is that its number of coefficients largely increase with the order and the memory depth—this fact is known as the *curse of dimensionality*—which affects negatively to the computational complexity of the final model and its properties through a regression. There exist several *ad hoc* models that are widely used. The most simple of them is the MP model [6], [27], which only retains the components of the diagonal, that is,

$$y[q] = \sum_{p=0}^{P} \sum_{l=0}^{L} a_{pl} x[q-l]|x[q-l]|^p \quad (7)$$

where $P+1$ is the nonlinear order and L is the memory depth. An extended version of the MP, which is very popular among researchers because of its efficiency in the modeling of memory effects, is the GMP, which adds nondiagonal terms up to a certain distance from the diagonal. The GMP follows the structure:

$$\begin{aligned} y[q] = & \sum_{p=0}^{P_a} \sum_{l=0}^{L_a} a_{pl} x[q-l]|x[q-l]|^p \\ & + \sum_{p=1}^{P_b} \sum_{l=0}^{L_b} \sum_{u=1}^{U_b} b_{plu} x[q-l]|x[q-l-u]|^p \\ & + \sum_{p=1}^{P_c} \sum_{l=0}^{L_c} \sum_{u=1}^{U_c} c_{plu} x[q-l]|x[q-l+u]|^p \end{aligned} \quad (8)$$

having a regressor form of $\phi(x) = x[q-l]|x[q-l \pm u]|^p$, where $P_i, i \in [a,b,c]$ control the nonlinear order, L_i, $i \in [a,b,c]$ fix the diagonal memory, and $U_i, i \in [a,b,c]$ set the maximum distance to the diagonal.

The main issue in the use of Volterra-based models is related to their high number of coefficients. Setting the order and memory parameters is a task that is commonly executed by trial and error, increasing them up until a desired level of performance is reached. Some indicators that relate the modeling performance and complexity, such as the complexity-aware-NMSE (CAN) metric [28] and the NMSE tolerance per coefficient μ [12], have been defined to aid this process. Also, since the models are usually defined in a series form depending on parameters that set order and memory depth, the selection of independent coefficients leads to losing their original mathematical structure.

Algorithm 1 General Pursuit Framework

Input: $\mathbf{X} \in \mathbb{C}^{m \times n}$, $\mathbf{y} \in \mathbb{C}^m$
Output: $S^{(\text{end})}$, $\hat{\mathbf{h}}^{(\text{end})}$
1: **for** $t = 1$ till stopping criterion is met **do**
2: Select columns of $\mathbf{X}$ (regressors of the model) based on some algorithm-dependent criteria and include them into $S^{(t)}$.
3: Update the estimate $\hat{\mathbf{y}}^{(t)}$ and $\mathbf{r}^{(t)}$.
4: **end for**

III. Greedy Pursuits

Greedy pursuits are a classic selection technique in the field of compressed-sensing signal processing [22]. They allow through an iterative fashion to select components of the model to be used by applying some local criteria at each iteration. Some of them also provide the sorting capability, returning an ordered list of the coefficients in which they are positioned by their relevance. In Algorithm 1, the general high-level framework of greedy pursuits [22] is shown. Greedy algorithms take as input the regressor matrix $\mathbf{X}$ and the output $\mathbf{y}$, returning the support set $S^{(\text{end})}$, which includes the indices of the selected regressors. The algorithm updates the residual

$$\mathbf{r}^{(t)} = \mathbf{y} - \hat{\mathbf{y}}^{(t)} \quad (9)$$

in each iteration t, which is defined as the difference between the output and its estimate $\hat{\mathbf{y}}^{(t)}$, and accounts for the amount of error to be captured by the not yet selected coefficients.

In this context, the next component will be included in the support set and depending on the particular greedy algorithm itself, the selection rule varies. Finally, once the stopping criterion is achieved, the algorithm halts the execution and returns the selection. The stopping criterion can be either a target objective, like a number of desired coefficients or an amount of maximum error, or may be calculated with order selection techniques.

In the mathematical formulation of the techniques, we rely on the nonlinear projection $H_k(\cdot)$ that sets all but the largest k elements of its arguments to zero and the supp$(\cdot)$ operation that returns the indices of the elements of the argument that are not equal to zero. For example, both functions concatenated as supp$(H_1(\cdot))$ return the index of the maximum value of the argument.

A. Orthogonal Matching Pursuit

The OMP [29] is a widely used greedy algorithm because of its simplicity and convergence properties. In its execution, it adds to the support set the regressor that has more resemblance with the residual, measured by their cross correlation. This way, the normalized correlation of regressor i (note that each column is divided by its module) is calculated

$$\mathbf{g}_{\{i\}}^{(t)} = \frac{\mathbf{X}_{\{i\}}^H}{\|\mathbf{X}_{\{i\}}\|_2} \mathbf{r}^{(t-1)} \quad (10)$$

where the subindex $\{i\}$ in the measurement matrix stands for its ith column. The index $s^{(t)}$ that has the maximum correlation

Algorithm 2 OMP

Input: $\mathbf{X}$, $\mathbf{y}$
Output: $S^{(\text{end})}$, $\hat{\mathbf{h}}^{(\text{end})}$
1: *Initialization* : $\mathbf{r}^{(0)} \leftarrow \mathbf{y}$, $S^{(0)} \leftarrow \varnothing$
2: **for** $t = 1$ till stopping criterion is met **do**
3: $\quad \mathbf{g}_{\{i\}}^{(t)} \overset{\forall i}{\leftarrow} \frac{\mathbf{X}_{\{i\}}^H}{\|\mathbf{X}_{\{i\}}\|_2}\mathbf{r}^{(t-1)}$
4: $\quad s^{(t)} \leftarrow \text{supp}\left(H_1\left|\mathbf{g}^{(t)}\right|\right)$
5: $\quad S^{(t)} \leftarrow S^{(t-1)} \cup s^{(t)}$
6: $\quad \hat{\mathbf{h}}_{S(t)} \leftarrow \mathbf{X}_{S(t)}^{\dagger}\mathbf{y}$
7: $\quad \hat{\mathbf{y}}^{(t)} \leftarrow \mathbf{X}_{S(t)}\hat{\mathbf{h}}$
8: $\quad \mathbf{r}^{(t)} \leftarrow \mathbf{y} - \hat{\mathbf{y}}^{(t)}$
9: **end for**

Algorithm 3 DOMP

Input: $\mathbf{X}$, $\mathbf{y}$
Output: $S^{(\text{end})}$, $\hat{\mathbf{h}}^{(\text{end})}$
1: *Initialization* : $\mathbf{r}^{(0)} \leftarrow \mathbf{y}$, $S^{(0)} \leftarrow \varnothing$, $\mathbf{Z}^{(0)} \leftarrow \mathbf{X}$
2: **for** $t = 1$ till stopping criterion is met **do**
3: $\quad \mathbf{g}_{\{i\}}^{(t)} \overset{\forall i}{\leftarrow} \frac{\mathbf{Z}_{\{i\}}^H}{\|\mathbf{Z}_{\{i\}}\|_2}\mathbf{r}^{(t-1)}$
4: $\quad s^{(t)} \leftarrow \text{supp}\left(H_1\left|\mathbf{g}^{(t)}\right|\right)$
5: $\quad S^{(t)} \leftarrow S^{(t-1)} \cup s^{(t)}$
6: $\quad \mathbf{p}^{(t)} \leftarrow \mathbf{Z}_{\{i^{(t)}\}}^{(t-1)H}\mathbf{Z}^{(t-1)}$
7: $\quad \mathbf{Z}^{(t)} \leftarrow \mathbf{Z}^{(t-1)} - \mathbf{p}^{(t)} \otimes \mathbf{Z}_{\{i^{(t)}\}}^{(t-1)}$
8: $\quad \hat{\mathbf{h}}_{S(t)} \leftarrow \mathbf{X}_{S(t)}^{\dagger}\mathbf{y}$
9: $\quad \hat{\mathbf{y}}^{(t)} \leftarrow \mathbf{X}_{S(t)}\hat{\mathbf{h}}$
10: $\quad \mathbf{r}^{(t)} \leftarrow \mathbf{y} - \hat{\mathbf{y}}^{(t)}$
11: **end for**

Algorithm 4 SP

Input: $\mathbf{X}$, $\mathbf{y}$, k
Output: $S^{(\text{end})}$, $\hat{\mathbf{h}}^{(\text{end})}$
1: *Initialization* : $\mathbf{r}^{(\mathbf{0})} \leftarrow \mathbf{y}$, $S^{(0)} \leftarrow \varnothing$
2: **for** $t = 1$ till stopping criterion is met **do**
3: $\quad \mathbf{g}_{\{i\}}^{(t)} \overset{\forall i}{\leftarrow} \frac{\mathbf{X}_{\{i\}}^H}{\|\mathbf{X}_{\{i\}}\|_2}\mathbf{r}^{(t-1)}$
4: $\quad s^{(t)} \leftarrow \text{supp}\left(H_k\left|\mathbf{g}^{(t)}\right|\right)$
5: $\quad S^{(t-0.5)} \leftarrow S^{(t-1)} \cup s^{(t)}$
6: $\quad \hat{\mathbf{h}}_{S^{(t-0.5)}} \leftarrow \mathbf{X}_{S^{(t-0.5)}}^{\dagger}\mathbf{y}$, $\hat{\mathbf{h}}_{\overline{S^{(t-0.5)}}} = 0$
7: $\quad S^{(t)} \leftarrow \text{supp}(\hat{\mathbf{h}}_{S^{(t-0.5)}})$
8: $\quad \hat{\mathbf{h}}_{S^{(t)}} \leftarrow \mathbf{X}_{S^{(t)}}^{\dagger}\mathbf{y}$
9: $\quad \hat{\mathbf{y}}^{(t)} \leftarrow \mathbf{X}_{S^{(t)}}\hat{\mathbf{h}}_{S^{(t)}}$
10: $\quad \mathbf{r}^{(t)} \leftarrow \mathbf{y} - \hat{\mathbf{y}}^{(t)}$
11: **end for**

Algorithm 5 CoSaMP

Input: $\mathbf{X}$, $\mathbf{y}$, k
Output: $S^{(\text{end})}$, $\hat{\mathbf{h}}^{(\text{end})}$
1: *Initialization* : $\mathbf{r}^{(\mathbf{0})} \leftarrow \mathbf{y}$, $S^{(0)} \leftarrow \varnothing$
2: **for** $t = 1$ till stopping criterion is met **do**
3: $\quad \mathbf{g}_{\{i\}}^{(t)} \overset{\forall i}{\leftarrow} \frac{\mathbf{X}_{\{i\}}^H}{\|\mathbf{X}_{\{i\}}\|_2}\mathbf{r}^{(t-1)}$
4: $\quad s^{(t)} \leftarrow \text{supp}\left(H_{2k}\left|\mathbf{g}^{(t)}\right|\right)$
5: $\quad S^{(t-0.5)} \leftarrow S^{(t-1)} \cup s^{(t)}$
6: $\quad \hat{\mathbf{h}}_{S^{(t-0.5)}} \leftarrow \mathbf{X}_{S^{(t-0.5)}}^{\dagger}\mathbf{y}$, $\hat{\mathbf{h}}_{\overline{S^{(t-0.5)}}} = 0$
7: $\quad S^{(t)} \leftarrow \text{supp}(\hat{\mathbf{h}}_{S^{(t-0.5)}})$
8: $\quad \hat{\mathbf{y}}^{(t)} \leftarrow \mathbf{X}_{S^{(t)}}\hat{\mathbf{h}}_{S^{(t-0.5)}}$
9: $\quad \mathbf{r}^{(t)} \leftarrow \mathbf{y} - \hat{\mathbf{y}}^{(t)}$
10: **end for**

is attained through

$$s^{(t)} = \text{supp}(H_1|\mathbf{g}^{(t)}|) \tag{11}$$

and then added to the support set in the next iteration

$$S^{(t)} = S^{(t-1)} \cup s^{(t)}. \tag{12}$$

Then, an estimation of the Volterra vector with just the columns that belong to the support set is performed with (5)

$$\hat{\mathbf{h}}_{S(t)} = \mathbf{X}_{S(t)}^{\dagger}\mathbf{y} \tag{13}$$

and the residual is updated with (9).

In the following iteration, the contribution of the last selected component will not be a part of the residual, and the next most similar regressor will be then selected. The OMP is summarized in Algorithm 2.

B. DOMP

The DOMP algorithm, whose summary is shown in Algorithm 3, was specifically designed to deal with highly correlated measurement matrices like Volterra series. It follows the same selection criterion that the OMP, but after the selection, it orthogonalizes the unselected regressors with respect to the chosen coefficient. That way it eliminates the effect of the correlation of the support set and its complement. In its operation, the algorithm keeps track of the matrix $\mathbf{Z}$, which is the orthogonal equivalent of $\mathbf{X}$. Therefore, the selection is performed in an orthonormal space that has a one-to-one relation to the Volterra space, allowing to indirectly select the coefficients in the measurement matrix. The orthogonalization is performed by using Gram–Schmidt, which is implemented in steps 6 and 7. In step 6, the projection of the selected regressor onto the rest of the basis is performed and then subtracted in step 7, where $\otimes$ denotes the Kronecker product.

C. Subspace Pursuit and CoSaMP

The SP and the CoSaMP are very similar conceptually. Algorithm 4 shows the SP pseudocode and Algorithm 5 summarizes the CoSaMP operation. Both perform a selection in two stages. First, they select the k and $2k$, respectively, coefficients with the highest correlation to the residual and add them to an intermediate support set $S^{(t-0.5)}$. The corresponding estimation $\hat{\mathbf{h}}_{S^{(t-0.5)}}$ is then performed and those coefficients belonging to the complement of the support set $\hat{\mathbf{h}}_{\overline{S^{(t-0.5)}}}$ are set to zero. Next, only the k coefficients of $\hat{\mathbf{h}}$ with the highest absolute value are retained into $S^{(t)}$. The SP performs a second pseudoinverse to update $\hat{\mathbf{h}}$ and the residual while the CoSaMP uses the estimation of the Volterra vector attained in the first selection.

TABLE I

COMPUTATIONAL COMPLEXITY ASSESSMENT MEASURED AS NUMBER OF MULTIPLICATION OPERATIONS WHERE t REFERS TO THE SIZE OF THE SUPPORT SET THAT IS EQUAL TO THE ITERATION INDEX t

Case	OMP	DOMP	SP	CoSaMP
Per iteration	$\mathcal{O}(mt^2)$	$\mathcal{O}(mt^2 + m^2n)$	$\mathcal{O}(2mk^2)$	$\mathcal{O}(4mk^2)$

Note that while both OMP and DOMP add one component to the support set per iteration, SP and CoSaMP run iteratively for a desired sparsity level, providing a k-sparse solution until reaching the stopping criterion.

IV. COMPLEXITY ASSESSMENT

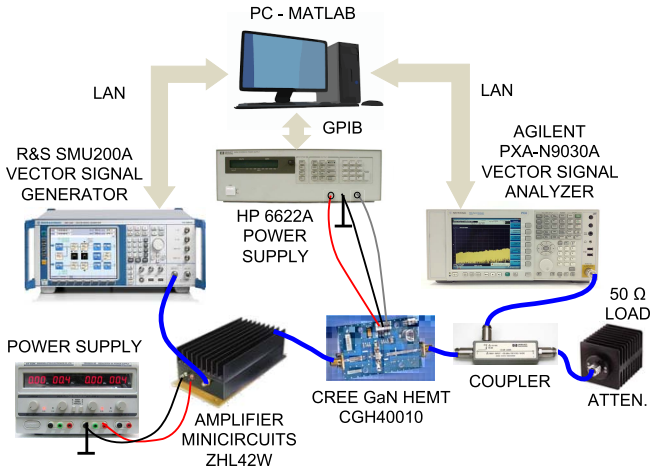
In this section, we provide a computational complexity comparison using the Bachmann–Landau notation in terms of the number of arithmetic multiplications, being this latter the one that has more impact on the overall complexity [30], required for each algorithm operation. The analysis is performed on the identification complexity of the techniques [31]. In Table I, the computational complexity per iteration for the algorithms in comparison is provided. The complexity is generally dominated by the pseudoinverse calculation involved in the regression. The OMP performs a $m \times t$ pseudoinverse per iteration, which results in a complexity of $\mathcal{O}(mt^2)$ considering that the pseudoinverse is calculated by using the QR algorithm. The DOMP shows an added complexity due to its Kronecker product, which is resembled by the term $\mathcal{O}(m^2n)$. The SP realizes two $m \times k$ pseudoinverses and the CoSaMP performs one $m \times 2k$ pseudoinverse, which follow $\mathcal{O}(2mk^2)$ and $\mathcal{O}(4mk^2)$, respectively.

Their fixed complexity per iteration, which depends on the desired sparsity, evidences the main advantage of SP and CoSaMP against the matching pursuits. Both OMP and DOMP have an increasing complexity with the number of selected components. On the other hand, in SP and CoSaMP, the number of iterations is not predefined, as it happens in the OMP and the DOMP that require k iterations to provide a k-sparse solution.

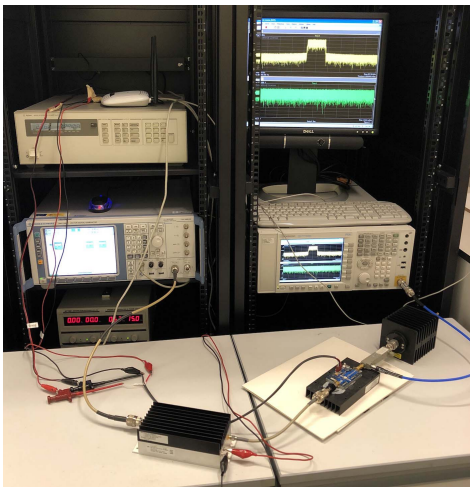
Finally, to wrap up the overview of greedy algorithms, a high-level qualitative comparative is provided in Table II, where the coefficient selection criterion used by each technique along with their advantages and drawbacks are summarized.

V. EXPERIMENTAL BENCHMARK

The selection techniques have been validated in a MATLAB-controlled hardware test bench shown in Fig. 3. Thus, the benchmark was carried out using a 15-MHz Long-Term Evolution (LTE) signal in different PAs, namely, a commercial class AB PA and a class J PA, which are presented in Sections V-A and V-B, respectively. Also, a signal cross validation was performed in Section V-C by identifying the model with a 30-MHz 5G-new radio (NR) signal and subsequently performing the validation operating over a 20-MHz multicarrier wideband code division multiple access (WCDMA) signal.



(a)



(b)

Fig. 3. Experimental benchmark. (a) Block-diagram. (b) Photograph of the actual setup.

The experimental setup is composed of a SMU200A vector signal generator (VSG) from Rohde & Schwarz, a PXA-N9030A vector signal analyzer (VSA) from Keysight Technologies, and two dc power supplies. The equipment, except one of the power supplies, is controlled through LAN and general-purpose instrumentation bus (GPIB) by using standard commands for programmable instruments (SCPI) from a PC with MATLAB.

A. DPD of a Commercial PA

A test signal was generated with an OFDM format and 15-MHz bandwidth, according to the LTE downlink standard, with a peak-to-average power ratio (PAPR) of 9.8 dB and more than 360 000 samples. A custom MATLAB script controlled the sending of the signal to the VSG and its settings to generate it with an oversampling factor of 6, i.e., a sampling frequency of 92.16 MSa/s.

A ZHL42W preamplifier from MiniCircuits was the first device of the measurement chain, followed by the evaluation board of a PA based on Cree's CGH40010 GaN HEMT. The ZHL42W drived the PA to a higher nonlinear operation

TABLE II
QUALITATIVE COMPARATIVE OF GREEDY ALGORITHMS

Technique	Coefficient selection criterion	Advantages	Disadvantages
OMP	Correlation with the residual.	• Returns a sorted list of regressors that can be pruned. • Fixed computational complexity for a desired sparsity.	• Does not take into account the correlation between regressors.
DOMP	Correlation with the residual after orthogonalization.	• Returns a sorted list of regressors that can be pruned. • Fixed computational complexity for a desired sparsity. • Enables the use of orthogonal signal processing.	• Performs an orthogonalization at each iteration.
PCA	Variance of the regressor.	• Returns a sorted list of regressors that can be pruned. • Fixed computational complexity for a desired sparsity.	• Needs a transformation into an equivalent space. • Does not take into account the PA output for selecting the coefficients.
SP	Correlation with the residual (k highest) and magnitude of the coefficient.	• Low computational complexity on updating the model structure.	• Returns a pruned solution without sorting the regressors.
CoSaMP	Correlation with the residual ($2k$ highest) and magnitude of the coefficient.	• Low computational complexity on updating the model structure.	• Returns a pruned solution without sorting the regressors.

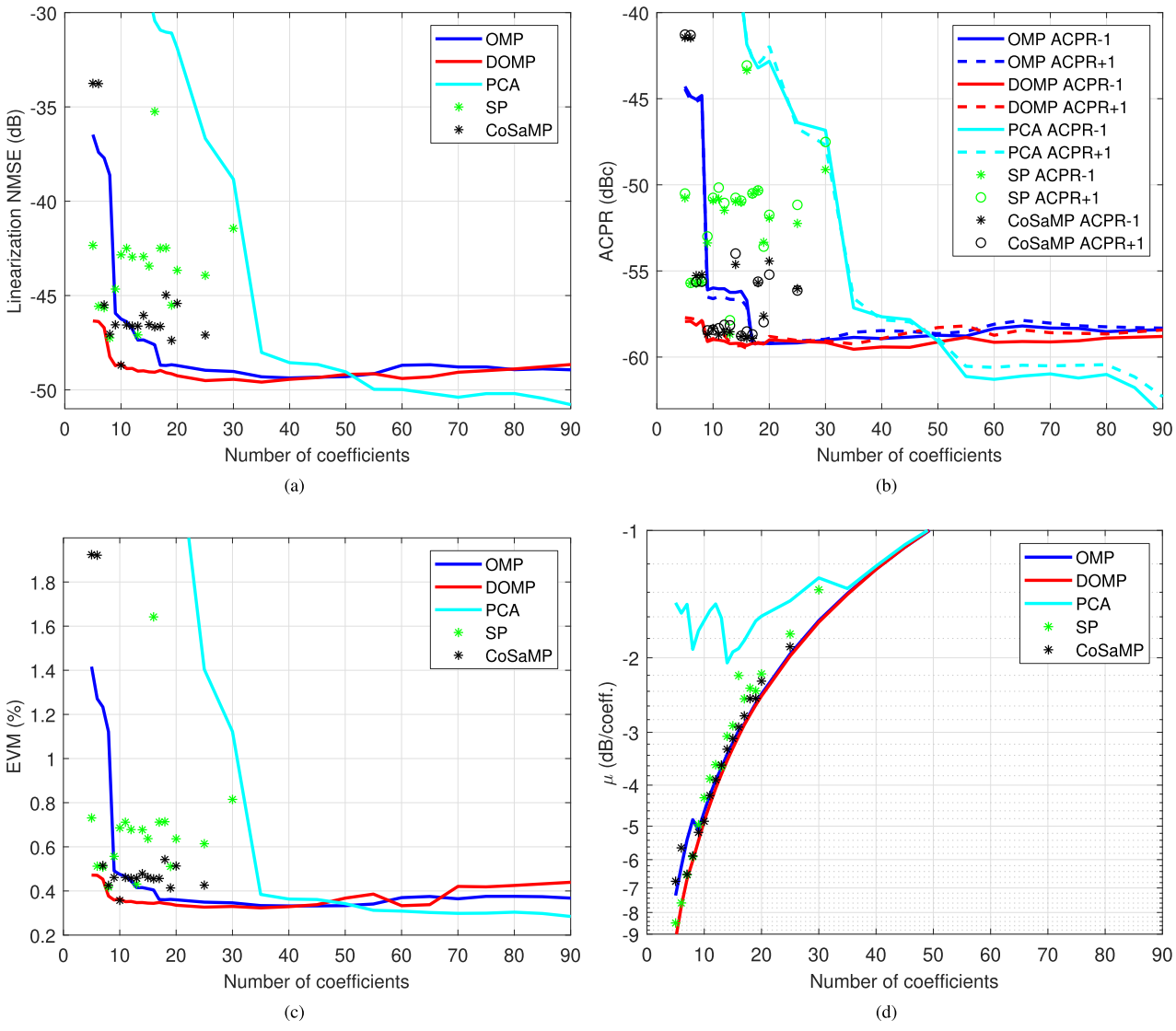


Fig. 4. Linearization performance for the models in comparison for a sweep of desired number of coefficients. (a) NMSE. (b) ACPR in the first lower and upper adjacent channels. (c) EVM. (d) NMSE tolerance per coefficient (μ).

point, delivering an average output power of +26.2 dBm (+36.0 dBm of peak output power) at 3.6 GHz. The average gain was 46.2 dB, which corresponds to a gain compression of 0.7 dB.

In the VSA, the output RF signal was downconverted to baseband and the appropriate range and sampling rate were set in order to recover it. The measurement dynamic range was optimized by averaging 300 acquisitions of the output signal.

TABLE III
BENCHMARK OF PRUNING TECHNIQUES FOR A TARGET NMSE OF −45 dB

Case	# coeff.	ACPR (dBc)				NMSE (dB)	EVM (%)	μ (dB/coeff.)
		-2	-1	+1	+2			
w/o DPD	-	−49.6	−38.6	−38.6	−49.3	−29.4	3.18	-
LS (FV model)	248	−67.3	−66.0	−66.0	−67.1	−54.4	0.19	−0.2
OMP	9	−59.2	−56.1	−56.5	−59.7	−45.9	0.49	−5.1
DOMP	5	−61.4	−57.9	−57.7	−61.3	−46.3	0.47	−9.3
PCA	35	−60.0	−57.1	−56.6	−59.9	−48.0	0.38	−1.4
SP	6	−59.3	−55.7	−55.7	−59.6	−45.6	0.51	−7.6
CoSaMP	7	−59.1	−55.3	−55.6	−59.4	−45.5	0.52	−6.5

TABLE IV
SELECTED REGRESSORS FOR EACH OF THE PRUNING TECHNIQUES FOR A MODEL OF 20 COEFFICIENTS

#. coeff.	Pruning Technique			
	OMP	DOMP	SP	CoSaMP
1	$x(q)$	$x(q)$	$x(q)$	$x(q)$
2	$x(q)\|x(q)\|^4$	$x(q)\|x(q)\|^2$	$x(q)\|x(q-1)\|^2$	$x(q)\|x(q-1)\|^2$
3	$x(q)\|x(q)\|^{12}$	$x(q)\|x(q)\|^4$	$x(q)\|x(q-2)\|^2$	$x(q)x(q-1)x^*(q-2)$
4	$x(q)\|x(q)\|^{10}$	$x(q)\|x(q)\|^6$	$x(q)x(q-3)x^*(q-2)$	$x(q)x(q-2)x^*(q-1)$
5	$x^2(q)x^*(q-2)$	$x(q)\|x(q)\|^8$	$x^2(q-3)x^*(q)\|x(q)\|^2$	$x(q-1)\|x(q-1)\|^2$
6	$x^3(q)x^{*2}(q-3)$	$x(q)\|x(q)\|^{10}$	$x^2(q-3)x^*(q-1)\|x(q)\|^2$	$x^2(q-1)x^*(q-2)$
7	$x(q-1)x(q-2)x^*(q)$	$x(q-3)$	$x^2(q-3)x^*(q-2)\|x(q)\|^2$	$x(q)\|x(q)\|^4$
8	$x(q-3)\|x(q-3)\|^2$	$x(q-1)$	$x(q-3)\|x(q-3)\|^2\|x(q)\|^2$	$x(q-1)\|x(q)\|^4$
9	$x(q)\|x(q)\|^2$	$x(q)\|x(q-1)\|^2$	$x(q-1)\|x(q-3)\|^4$	$x(q-2)\|x(q)\|^4$
10	$x(q-3)$	$x(q)\|x(q)\|^{12}$	$x(q-2)x^2(q-3)x^{*2}(q)$	$x^2(q-1)x^*(q)\|x(q)\|^2$
11	$x^2(q-3)x^*(q)$	$x(q-2)$	$x(q-2)x^2(q-3)x^*(q)x^*(q-1)$	$x(q-1)x(q-2)x^*(q)\|x(q)\|^2$
12	$x(q-3)\|x(q)\|^2$	$x(q-2)\|x(q)\|^2$	$x^2(q-3)x^*(q)\|x(q-2)\|^2$	$x(q-1)x(q-3)x^*(q)\|x(q)\|^2$
13	$x(q-1)$	$x(q)x(q-3)x^*(q-1)$	$x(q-2)x(q-3)x^*(q)\|x(q-3)\|^2$	$x^2(q-2)x^*(q)\|x(q)\|^2$
14	$x^2(q-3)x^*(q-1)\|x(q-3)\|^2$	$x^2(q-1)x^*(q)\|x(q-1)\|^2$	$x^3(q-3)x^*(q)x^*(q)$	$x^3(q-1)x^{*2}(q)$
15	$x^3(q-3)x^{*2}(q)$	$x^2(q-2)x^*(q-1)\|x(q)\|^2$	$x^3(q-3)x^*(q)x^*(q-1)$	$x^2(q-1)x(q-2)x^{*2}(q)$
16	$x(q)\|x(q)\|^6$	$x^2(q)x^*(q-3)$	$x^3(q-3)x^*(q)x^*(q-2)$	$x^2(q-1)x(q-3)x^{*2}(q)$
17	$x(q)\|x(q)\|^8$	$x(q)x(q-1)x^*(q-2)$	$x(q)\|x(q)\|^6$	$x(q-1)x^2(q-2)x^{*2}(q)$
18	$x^2(q-2)x^*(q)\|x(q-2)\|^2$	$x(q)x(q-1)x^*(q-3)$	$x(q)\|x(q)\|^8$	$x(q)\|x(q)\|^6$
19	$x(q)x(q-1)x^*(q-3)\|x(q-3)\|^2$	$x(q)x(q-2)x^*(q-3)$	$x(q)\|x(q)\|^{10}$	$x(q)\|x(q)\|^8$
20	$x^2(q)x^*(q-3)$	$x(q)x(q-3)x^*(q-2)$	$x(q)\|x(q)\|^{12}$	$x(q)\|x(q)\|^{10}$

Then, the signal was downloaded to the MATLAB workspace and it was postprocessed to adequate it to the algorithms. The postprocessing consisted on the following operations.

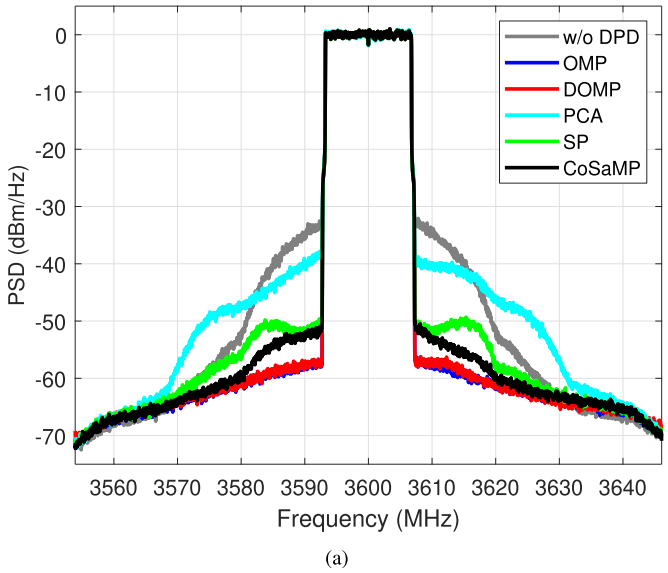
1) *Normalization:* Scaling of the signal taking into account the attenuations and gains of the measurement chain.
2) *Time Alignment:* Synchronization of the input and output signals in time.
3) *Partition Into Several Data Sets:* The DPDs were identified with a segment of consecutive samples with length equal to the 0.5% of the total signal length. This segment was chosen to include the sample with the highest absolute value at the output. The complete signal was applied to the DPD in order to get the predistorted input signal.

The model under test was a FV configured with a fifth order and memory depth of 3. The memoryless part up to 13th order was added to it in order to enhance its modeling capabilities, reaching a total number of coefficients of 248. The full basis model follow (6) with $P = 6$ and $Q_{2p+1} = 3$ for $p \in [0, 1, 2]$ and $Q_{2p+1} = 0$ for $p \in [3, 4, 5]$.

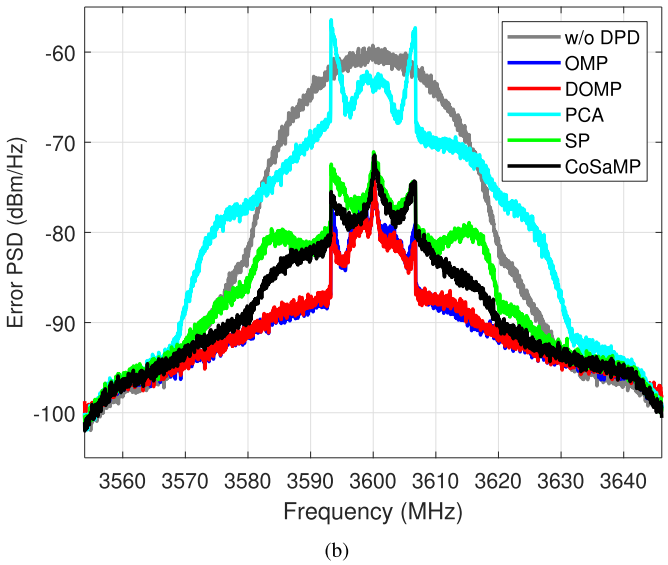
The OMP, DOMP, PCA, SP, and CoSaMP techniques were swept in number of desired coefficients to perform a benchmark. Fig. 4 shows the linearization performance of each of the algorithms. In this work, we selected the following order-reduction performance indicators.

1) Normalized mean square error (NMSE) between the input to the DPD and the scaled PA output.
2) Adjacent channel power ratio (ACPR) of the output signal.
3) Error vector magnitude (EVM) of the linearized signal.
4) NMSE tolerance per coefficient (μ) [12] defined as the achieved linearization NMSE divided by the number of coefficients of the DPD.

As it is reasonable to expect, NMSE, ACPR, and EVM follow the same pattern for all the techniques in comparison. Since OMP, DOMP, and PCA select the coefficients incrementally, the evolution of the error metrics is decreasing with the number of components. SP and CoSaMP are run independently for each number of coefficients; therefore, they do not guarantee models with increasing error mitigation capabilities, i.e., the

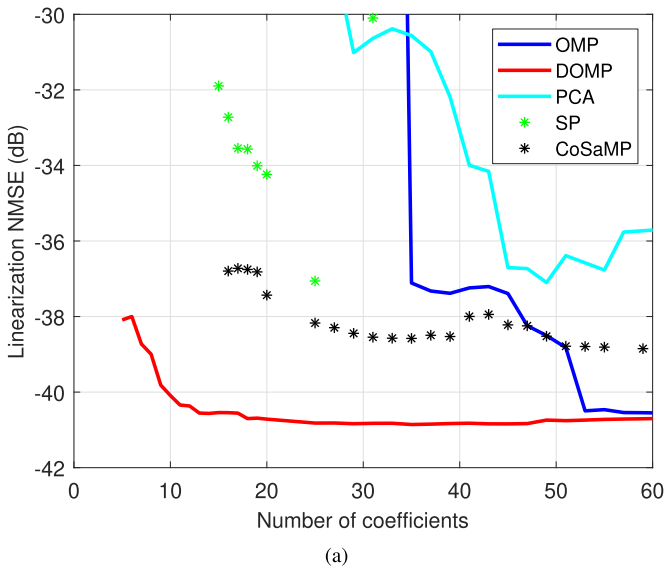


(a)

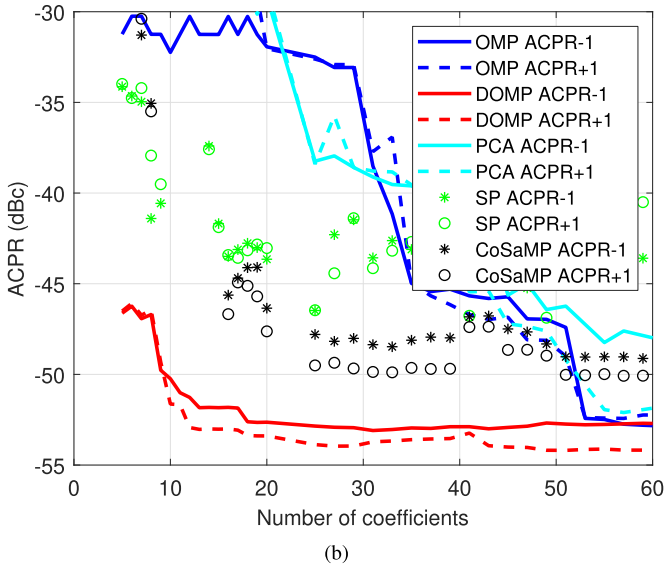


(b)

Fig. 5. Spectrum of a 15-MHz LTE signal for 20 coefficients. (a) PSD of the output. (b) PSD of the error.



(a)



(b)

Fig. 6. Linearization performance of a class J PA for the models in comparison for a sweep of desired number of coefficients. (a) NMSE. (b) ACPR.

DPDs for each number of coefficients are independent. PCA shows a slow decrease in the first iterations to achieve the best linearization at the end of the span in the number of coefficients. DOMP achieves the fastest pruning, reaching its best linearization performance in just the first iterations. Regarding the NMSE tolerance, the Pareto front—defined here as the values with lowest tolerance for each number of coefficients [32]—is conformed by the DOMP for $\mu < -1$ and by PCA for a number of coefficients greater than 50.

In order to highlight the pruning performance of the algorithms, a target linearization NMSE of -45 dB was fixed. The pruned model with the least number of coefficients that fulfills the target NMSE is given in Table III. All the algorithms provide similar linearization capabilities in terms of ACPR, while the reduction in the number of coefficients varies from a 96% in the DOMP to a 85% in the PCA. The maximum NMSE tolerance is produced by the DOMP, with over -9 dB per coefficient.

Table IV shows the selected regressors for a fixed number of 20 coefficients. Note that for SP and CoSaMP, these are provided without any particular order, while for OMP and DOMP, they are sorted as returned by the algorithms. Focusing on the memoryless part, all the algorithms select it, with the exception of the fifth-order term $x(q)|x(q)|^4$ in the SP case. The DOMP selects the whole set of the first-order regressors $x(q-1)$, $x(q-2)$, and $x(q-3)$ while the OMP misses the second and the SP and CoSaMP were not able to catch them. Also, regarding the sort order (than can only be compared between OMP and DOMP because these are the only ones that are incremental), DOMP chooses first the regressors that are reasonably expected to be important, such as the memoryless part, the linear terms and some other third

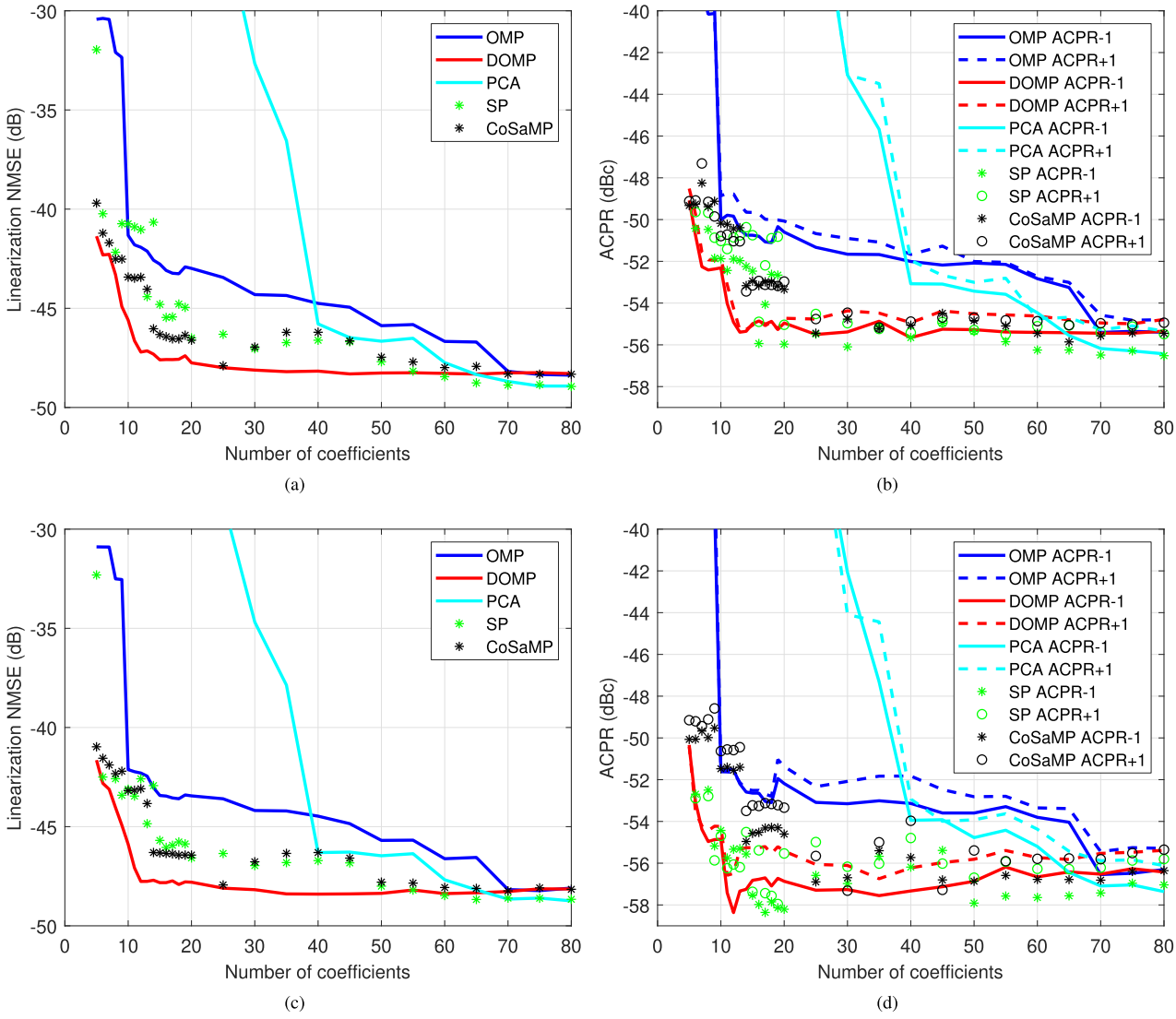


Fig. 7. Linearization performance for the models in comparison for a sweep of desired number of coefficients in a cross-validation scenario. The DPD was identified with a 30-MHz 5G-NR signal. (a) NMSE and (b) ACPR of a 30-MHz 5G-NR signal. (c) NMSE and (d) ACPR of a 20-MHz multicarrier WCDMA signal.

order terms, while OMP has dispersion in the criteria for the selection. Regarding the spectral regrowth reduction, the normalized power spectral density (PSD) of the linearized signal and of the error signal for the same number of coefficients are shown in Fig. 5. The out-of-band distortion is clearly reduced with respect to the case without DPD. The DOMP algorithm improves the performance of the other algorithms showing a linearization result similar to the OMP, followed by the rest of the techniques.

B. DPD of a Class J PA

The experimental benchmark was also run over a continuous mode class J PA [33] based on the CGH35015F device, which was designed for an operation frequency of 850 MHz. Respect to the previous experiment, the driver was replaced by two cascaded Mini-Circuits TVA-4W-422A+ preamplifiers.

The operation point was set to an output power of +32.8 dBm, in which the PA exhibits a gain expansion followed by a compression. This behavior produces a remarkable nonlinear distortion, illustrated by the fact that the operation point exhibits an ACPR of −28.7 and −27.8 dBc in the first adjacent channels, and an NMSE of −19.0 dB. In this experiment, the model under test was a GMP with 13th order in the diagonal and 7th order in the nondiagonal parts. The memory was set to decrease with the order, resulting in a set of 81 coefficients.

Fig. 6 demonstrates how the behavior of the benchmark holds after varying the PA and the model under test, in which the DOMP exhibits the fastest decay.

C. DPD of a Commercial PA and Signal Cross Validation

A cross-validation experiment was carried out in order to further probe the DPDs attained by the algorithms. In this experiment, the DUT was the evaluation board used in Section V-A. The driver was replaced by a Mini-Circuits TVA-4W-422A+, which brought the DUT to

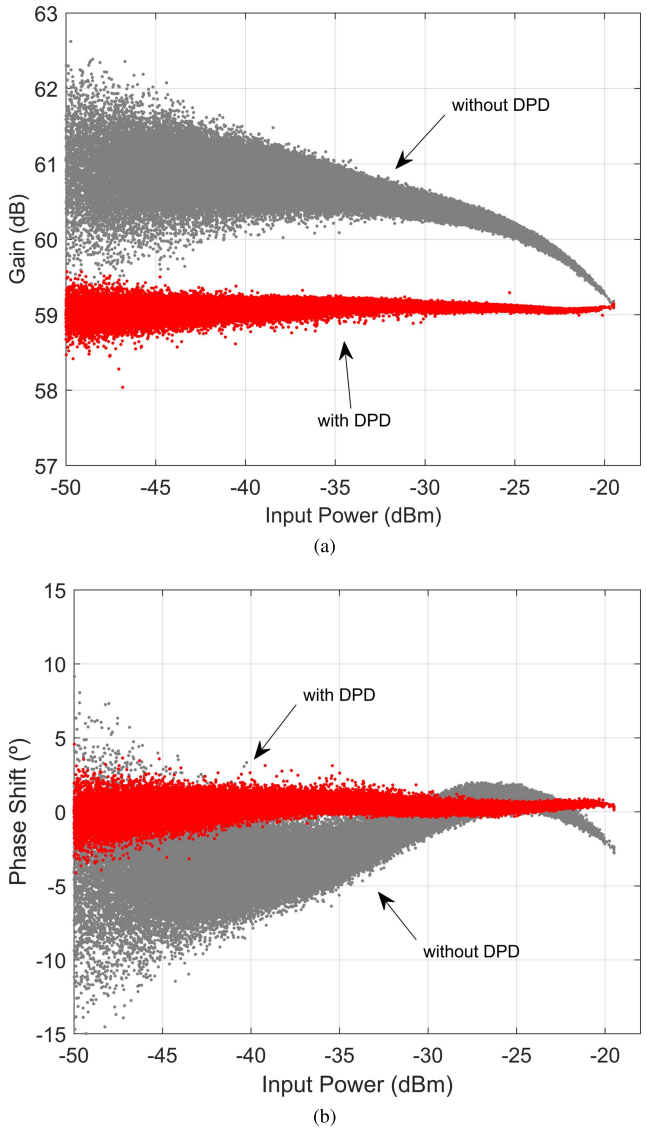


Fig. 8. AM/AM and AM/PM characteristics of the PA without DPD and the DOMP algorithm with 35 coefficients.

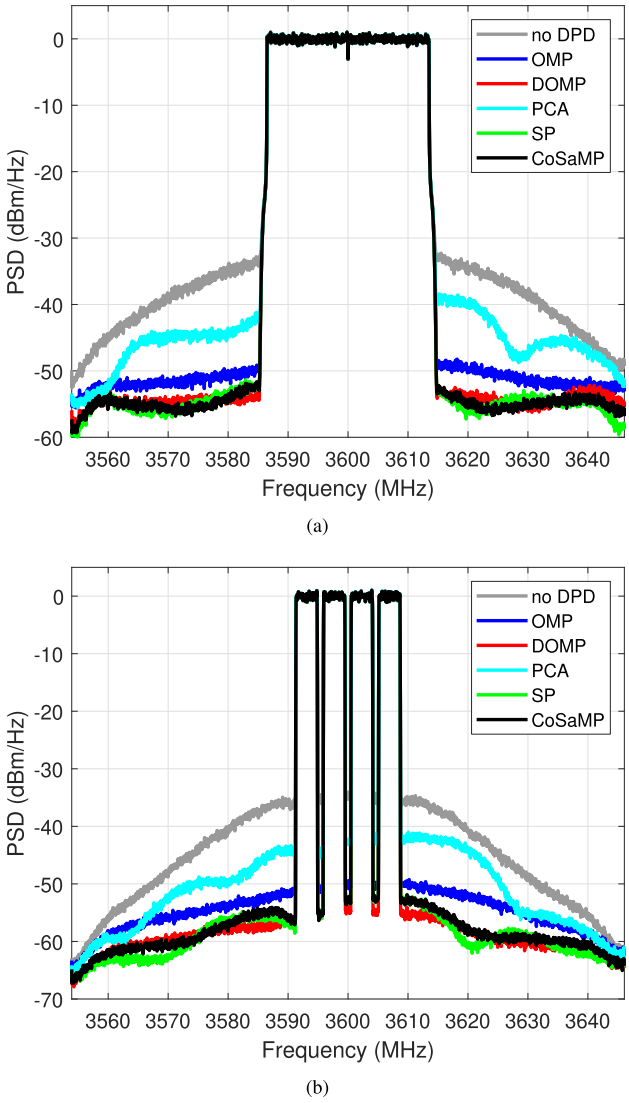


Fig. 9. PSD for 35 coefficients in a cross-validation scenario, having identified the DPD with a 5G-NR signal. (a) 30-MHz 5G-NR (b) 20-MHz multicarrier WCDMA.

a higher operation point and the sweep in desired sparsity was repeated. The output power was +29.2 dB, showing an average gain of 61.1 dB and a gain compression of over 2 dB.

The signal used for DPD identification was configured following the 5G-NR standard with a separation between subcarriers of 30 KHz and a canalization of 30 MHz. With the aim of highlighting the independence of the signal, the obtained DPDs were also validated with a 20-MHz multicarrier WCDMA. A crest factor reduction (CFR) technique was applied to both signals for exhibiting a PAPR of 10.5 dB. The model under test was the GMP used in Section V.B.

The linearization performance in NMSE and ACPR is shown in Fig. 7, where it can be observed that the resulting DPDs hold the performance after changing the signal type. The AM/AM and AM/phase modulation (PM) characteristics of the PA without DPD and with a 35-coefficient DPD attained by the DOMP are shown in Fig. 8. Finally, Fig. 9 shows the unlinearized and linearized spectra considering the resulting DPD for a target number of coefficients of 35.

VI. Conclusion

This paper has presented a review of greedy pursuits for the order reduction of behavioral models. A standard framework has been provided in which OMP, DOMP, SP, and CoSaMP algorithms have been formulated with a common notation. As a result, the slight differences in the algorithm equations have been highlighted. The computational complexity of the techniques has been assessed, and a high-level qualitative comparative has been provided.

Moreover, a benchmark of the techniques in the DPD of a commercial class AB PA and a class J PA have been presented, giving a clear overview of their pruning characteristics in terms of linearization indicators and regressor selection capabilities. The novel technique that takes into account the correlation of the Volterra basis, the DOMP algorithm, has shown an equivalent linearization performance with fewer coefficients than the rest of the algorithms in comparison.

REFERENCES

[1] A. Zhu, J. C. Pedro, and T. J. Brazil, "Dynamic deviation reduction-based volterra behavioral modeling of RF power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 54, no. 12, pp. 4323–4332, Dec. 2006.

[2] 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.

[3] L. Guan and A. Zhu, "Green communications: Digital predistortion for wideband RF power amplifiers," *IEEE Microw. Mag.*, vol. 15, no. 7, pp. 84–99, Nov. 2014.

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

[5] 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.

[6] J. Kim and K. Konstantinou, "Digital predistortion of wideband signals based on power amplifier model with memory," *Electron. Lett.*, vol. 37, no. 23, pp. 1417–1418, Nov. 2001.

[7] A. Zhu and T. J. Brazil, "Behavioral modeling of RF power amplifiers based on pruned volterra series," *IEEE Microw. Wireless Compon. Lett.*, vol. 14, no. 12, pp. 563–565, Dec. 2004.

[8] D. Wisell, J. Jalden, and P. Handel, "Behavioral power amplifier modeling using the LASSO," in *Proc. IEEE Instrum. Meas. Technol. Conf.*, May 2008, pp. 1864–1867.

[9] P. L. Gilabert *et al.*, "Order reduction of wideband digital predistorters using principal component analysis," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Jun. 2013, pp. 1–7.

[10] D. López-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.

[11] R. N. Braithwaite, "Digital predistortion of an RF power amplifier using a reduced volterra series model with a memory polynomial estimator," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 10, pp. 3613–3623, Oct. 2017.

[12] S. Wang, M. A. Hussein, O. Venard, and G. Baudoin, "A novel algorithm for determining the structure of digital predistortion models," *IEEE Trans. Veh. Technol.*, vol. 67, no. 8, pp. 7326–7340, Aug. 2018.

[13] 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 Trans. Microw. Theory Techn.*, vol. 62, no. 11, pp. 2583–2593, Nov. 2014.

[14] J. A. Becerra, D. Herrera, M. J. Madero-Ayora, and C. Crespo-Cadenas, "Sparse model selection of digital predistorters using subspace pursuit," in *Proc. 13th Eur. Microw. Int. Circuits Conf. (EuMIC)*, Sep. 2018, pp. 190–193.

[15] E. Zenteno, S. Amin, M. Isaksson, D. Rönnow, and P. Händel, "Combating the dimensionality of nonlinear MIMO amplifier predistortion by basis pursuit," in *Proc. 44th Eur. Microw. Conf.*, Oct. 2014, pp. 833–836.

[16] 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 Techn.*, vol. 63, no. 2, pp. 745–753, Feb. 2015.

[17] 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.

[18] 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.

[19] J. A. Becerra, M. J. Madero-Ayora, J. Reina-Tosina, C. Crespo-Cadenas, J. García-Frías, and G. Arce, "A reduced-complexity doubly orthogonal matching pursuit algorithm for power amplifier sparse behavioral modeling," in *Proc. IEEE Topical Conf. RF/Microw. Power Amplif. Radio Wireless Appl. (PAWR)*, Jan. 2019, pp. 1–3.

[20] C. Crespo-Cadenas, M. J. Madero-Ayora, and J. A. Becerra, "Volterra-based behavioral modeling, parameter estimation, and linearization," in *IEEE MTT-S Int. Microw. Symp. Dig.*, Dec. 2018, pp. 1–4.

[21] J. Wood, *Behavioral Modeling and Linearization of RF Power Amplifiers* (Artech House Microwave Library). Norwood, MA, USA: Artech House, 2014.

[22] Y. C. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications* (Compressed Sensing: Theory and Applications). Cambridge, U.K.: Cambridge Univ. Press, 2012.

[23] A. Zhu and T. J. Brazil, "An overview of Volterra series based behavioral modeling of RF/microwave power amplifiers," in *Proc. IEEE Annu. Wireless Microw. Technol. Conf.*, Dec. 2006, pp. 1–5.

[24] J. C. Pedro and S. A. Maas, "A comparative overview of microwave and wireless power-amplifier behavioral modeling approaches," *IEEE Trans. Microw. Theory Techn.*, vol. 53, no. 4, pp. 1150–1163, Apr. 2005.

[25] W. Cao and A. Zhu, "A modified decomposed vector rotation-based behavioral model with efficient hardware implementation for digital predistortion of RF power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 7, pp. 2443–2452, Jul. 2017.

[26] K. A. Okoudjou, *Finite Frame Theory: A Complete Introduction to Overcompleteness* (Proceedings of Symposia in Applied Mathematics). Providence, RI, USA: American Mathematical Society, 2016.

[27] Y.-J. Liu, W. Chen, J. Zhou, B.-H. Zhou, and F. M. Ghannouchi, "Digital predistortion for concurrent dual-band transmitters using 2-D modified memory polynomials," *IEEE Trans. Microw. Theory Techn.*, vol. 61, no. 1, pp. 281–290, Jan. 2013.

[28] O. Hammi and A. Miftah, "Complexity-aware-normalised mean squared error 'CAN' metric for dimension estimation of memory polynomial-based power amplifiers behavioural models," *IET Commun.*, vol. 9, no. 18, pp. 2227–2233, Dec. 2015.

[29] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. 27th Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, USA, vol. 1, Nov. 1993, pp. 40–44.

[30] P. Bürgisser, M. Clausen, and A. Shokrollahi, *Algebraic Complexity Theory*, vol. 315. Berlin, Germany: Springer-Verlag, 1997.

[31] A. S. Tehrani, H. Cao, T. Eriksson, M. Isaksson, and C. Fager, "A comparative analysis of the complexity/accuracy tradeoff in power amplifier behavioral models," *IEEE Trans. Microw. Theory Techn.*, vol. 58, no. 6, pp. 1510–1520, Jun. 2010.

[32] T. Keßler, F. Logist, and M. Mangold, "Use of predictor corrector methods for multi-objective optimization of dynamic systems," *Comput. Aided Chem. Eng.*, vol. 38, pp. 313–318, 2016.

[33] M. J. Madero-Ayora, M. Allegue-Martínez, J. A. García, 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.

Juan A. Becerra (S'12–M'18) 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 of 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 current research interests include behavioral modeling and linearization of power amplifiers and compressed-sensing signal processing.

María J. Madero-Ayora (S'06–M'09) received the M.Sc. and Ph.D. degrees in telecommunication engineering from the University of 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.

Carlos Crespo-Cadenas (M'93–SM'15) was born in Madrid, Spain. He received the Physics degree from the University of Havana, Havana, Cuba, in 1973, and the Doctor 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 Sevilla, where he is currently a Professor. His current research interests include the communication systems, nonlinear analysis of active microwave devices, power amplifier behavioral modeling, and linearization techniques.