Symmetrized Basis Function Approximation Network for Passive Intermodulation Cancellation

Changan Liu, Zhibin Yan

***Abstract*—The passive intermodulation (PIM) interference in simultaneous transmit-receive system is difficult to deal with because the process of generating PIM signal is both nonlinear and with memory. To cancel such PIM interference digitally, this paper proposes the symmetrized basis function approximation network. Firstly, we recognize and justify one methodology underlying the current cascaded model, by which the two physically coupled causes, nonlinearity and memory, are mathematically decoupled. After decoupling, the nonlinear block in the system model only needs to be a static nonlinear mapping without memory. This recognition to the role of nonlinear block, along with the concept of basis transformation in linear space, leads to the idea of basis function approximation network (BFAN). Secondly, we point out that the PIM generating process is an odd mapping. Building this prior information into system modelling, we propose further the symmetrized BFAN, which greatly improves cancellation performance and reduces computational complexity. Lastly, we believe that the number of the nonlinear functions used in the system model should correspond to the number of the PIM sources. Building this belief into system modeling, we let different nonlinear units share the same weights so that they represent the same nonlinear function corresponding to the same one PIM source. This turns out to be tremendously effective in PIM interference cancellation for the experiment data. Additionally, the critical point phenomenon is explained with experiment and data analysis by the virtue of our simple but rational way to deal with nonlinearity in PIM.**

***Index Terms*—passive intermodulation, nonlinearity, memory, system identification, digital predistortion**

I. Introduction

PASSIVE intermodulation (PIM) interference is an important technical challenge in the developing of modern wireless communication technology. The PIM phenomenon stems from the inherent nonlinear characteristic of passive components that withstand high-power transmitting signal. The non-linearity leads to the interaction between multiple frequencies of the transmitting signal, resulting in unexpected intermodulation products. These generated intermodulation signals can fall within the receiving frequency band of the simultaneous transmit-receive system, causing severe interference to received signals.

A. Prior Art and Related Research

Recent years witnessed the development of an advanced technology, the digital pre-distortion, in wireless communication [1]. In the scenario of PIM interference cancellation, the technology seeks to identify mathematically the nonlinear process of generating PIM signal occurring in receiving channel from transmitting signal, and then to cancel the PIM interference by algorithm [2].

Roughly speaking, the theoretical aspect of digital predistortion is in the scope of a research field—system identification [3]. The first thing is to construct a mathematical model to represent the real process of the signal transformation. Particularly in PIM interference cancellation, research from the viewpoint of system identification theory is growing. A segmented polynomial model is developed to calculate the powers of PIM in [4]. In [5], the authors propose a behavioral model, which takes into account the coexisting power amplifier nonlinearities. A composite exponential model is presented to characterize the weak nonlinearity of PIM in [6]. The work on [7] focuses on modelling the distributed nonlinearities of transmission lines.

To give the non-linearity an explicit mathematical expression with parameters to be identified features these models, among which the Volterra series has been widely adopted in literature [8], [9]. In spirit Volterra series is just Taylor series with infinite number of arguments. Since Taylor series is well known in approximating nonlinear function, this virtue makes the Volterra series the first choice at the beginning of digital pre-distortion technology. However, as communication system and signal processing technology develop rapidly, cascaded models, including Wiener-Hammerstein ones, are increasingly showing advantages in applications with real-time requirement for their structural simplicity and computational efficiency [10], [11].

Cascaded model consists of memory linear blocks and memoryless nonlinear blocks. In this modeling approach, the linear blocks capture the memory effects, while the nonlinear blocks address the nonlinearity without memory. The architecture simplifies nonlinear issues and reduces the computational burden during system identification by segregating the nonlinearity and memory property, both of which complicate system modeling and identification. In the aspect of modeling the memory-less non-linearity, spline-based methods have attracted much attention in recent years for their effectiveness in achieving lower computational complexity while maintaining model accuracy. Integrating all these considerations, cascaded model shows advantages compared with the early favored Volterra series models, which deal with nonlinearity and memory property together, not separately [11]–[14]. Still, it needs more justification to separate the PIM generating physical process into memory linear and memoryless nonlinear parts mathematically equivalently. Such reflection will inspire

This work was supported in part by HUAWEI under the project TC20210609011, "PIM digital modeling based on multi-dimensional nonlinear System", and by NSF of China under Grant 62273056.

The authors are with School of Sciences, Harbin Institute of Technology (Shenzhen), Shenzhen, Guangdong, China (e-mail: 22b358003@stu.hit.edu.cn; zbyan@hit.edu.cn)

further improvement.

The perspective of complex-valued signal processing results from the common application of I/Q modulation naturally. In [15], [16], the authors present a concept known as the injection-based nonlinear block, where the complex nonlinear gain of the system is determined by a deviation from unity gain. This method is effective in reducing the uncertainty of gain across cascaded blocks. It introduces square root calculations, thereby increasing local computational burden in the forward prediction of signal and the backward propagation of error. This matters much in the engineering realization of the technology.

Neural networks have become more and more commonly used for modeling the radio circuits and systems [17]. They provide an efficient alternative to traditional nonlinear blocks, but are often computationally expensive. In [18], the authors propose feedforward neural network with multiple hidden layers for digital predistortion of the power amplifier. A class of augmented complex-value functional link network adaptive algorithms are developed for complex-value signals in [19]. To add hidden layers is capable of approximating any nonlinear function in principle, but this increases computational burden, and lacks physical support when applied to PIM cancellation. There is a real technical demand to reduce the computational burden while maintaining accuracy. It is in hope to build more prior information into the architecture to simplify the network.

B. Novelty and Contributions

In this paper, we propose symmetrized basis function approximation network for PIM cancellation, which greatly reduces computational complexity and achieves competitive cancellation performance compared with the existing solutions in the state of the art. The novelty and contributions of this paper are summarized as follows:

- We use the myopic theorem [20], [21] to justify one specific cascaded model, whose nonlinear block, called basis function approximation network specially in this paper, is memoryless. The meaningful implication of the theorem is that the model only needs to contain one nonlinear layer, which is memoryless. Since the role of the nonlinear layer itself is to approximate a **static** nonlinear function, not a dynamical nonlinear operator, it can be taken as a basis function approximation network.
- We point out that the generating process of PIM interference has the inherent symmetry physically, and further symmetrize our basis function approximation network. Symmetrization is achieved through transforming the basis functions into odd-symmetric forms, thereby building the system's prior physical information into the network design. This innovative approach not only significantly reduces the computational complexity, but also enhances the cancellation performance.
- Using data gathered from the real simultaneous transmit-receive system, the identified nonlinear function in our model shows a critical point phenomenon: it begins with almost no interference in low power region and transitions to weak nonlinearity as the power transcends a critical point. This behavior aligns with the physically generating mechanism of real PIM, demonstrating the physical validity of the model.
- In the experiment, we let the different nonlinear units share the same weights with the understanding in mind that the nonlinear transformations distributed in different places in the process may be indeed produced by the same one PIM source. Intuitively, the needed number of different nonlinear functions in the model depends on the number of PIM sources in the real system. Building this prior information into the model greatly reduces the computational burden in the identification algorithm.

C. Notations used in this paper

In this article, matrices are represented by boldface capital letters. Ordinary transpose, Hermitian transpose, and complex conjugation are denoted by $(\cdot)^{\mathrm{T}}, (\cdot)^{\mathrm{H}}, (\cdot)^{*}$, respectively. All vectors are in column, denoted by boldface lowercase letters, like $\boldsymbol{w} \in \mathbb{C}^M = [\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \boldsymbol{w}_M]^{\mathrm{T}}$. In recursive algorithms, discrete time iteration index n is added. For example, we write $\boldsymbol{w}[n+1] = \boldsymbol{w}[n] + \Delta\boldsymbol{w}[n]$. Additionally, the convolution is represented as $*$.

D. Organization of the Paper

The rest of the paper is organised as follows. Section II presents the basis function approximation network, which justifies and illustrates the system modelling methodology of separating memory and non-linearity in PIM cancellation. Section III introduces the symmetrization method in the basis function approximation network. The method builds the prior odd-mapping information into the network structure. The real PIM cancellation results are provided and analyzed in Section IV. Finally, Section V concludes the main findings of this paper and provides remarks on future research direction.

II. Principle of Basis Function Approximation Network

In this section, we introduce the basis function approximation network to model the nonlinear behavior of the generating process of PIM interference signal.

A. Recognizing the Role of Nonlinear Block

Different from the normal nonlinear nodes in the general neural network, the role of the nonlinear block is appointed to approximate locally some static nonlinear function. This methodology is mainly based on the following "myopic theorem" [20].

Theorem II.1. Let the input signal $x(\cdot)$ and output signal $d(\cdot)$ take values in complex number field $\mathbb{C}$. Suppose that continuous nonlinear operator $\mathscr{A} : x(\cdot) \mapsto d(\cdot)$ is causal, time-invariant, and uniformly fading-memory. Then for every $\epsilon > 0$, there exist a sufficiently large integer M, a linear time-invariant causal operator $L : x(\cdot) \mapsto u(\cdot)$ where the intermediate signal $u(\cdot)$ is $\mathbb{C}^M$-valued or, in other words, has M components, and a nonlinear function $NL : \mathbb{C}^M \to \mathbb{C}$

which maps the n-dimensional complex vector $u(t)$ to the complex number $y(t)$, such that for every t,

$$\begin{aligned}&|(\mathscr{A}(x(\cdot)))(t) - NL((L(x(\cdot)))(t))|\\ =&|d(t) - NL(u(t))| = |d(t) - y(t)| < \epsilon.\end{aligned}$$

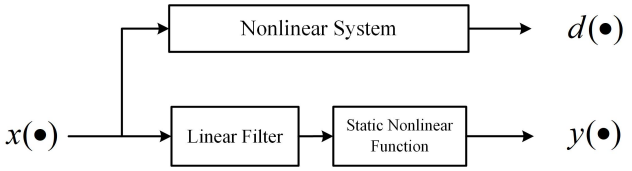

Fig. 1. Nonlinear system approximation.

According to Theorem II.1, we can represent logically equivalently the generating process of PIM interference signal by a cascaded model, which is decomposed into cascaded memory linear block and memoryless nonlinear block; see Fig. 1. One meaningful implication is that we can deal with the complicated problems caused by non-linearity and memory physically coupled through two mathematically decoupled model parts. The complexity of the memory problem is reduced because it is dealt with by linear blocks; the complexity of the nonlinear problem is reduced because it is dealt with by static nonlinear function without memory.

B. Linear Space of Piece-wise Linear Continuous Function

Now the design of the needed nonlinear block is to well approximate a static nonlinear function. In the scenario of PIM cancellation, no smoothness requirement is imposed on the approximation. In order to reduce computational complexity to the utmost extent (this matters much for engineering realization), we use piece-wise linear continuous function, the spline function of degree 1, to approximate the static nonlinear function in hope. This can be justified by the following theorem.

Theorem II.2. Let $f : [a, b] \to \mathbb{R}$ be a continuous function. For any given $\epsilon > 0$, there exists a $\delta > 0$ such that if the partition

$$a = x_0 < x_1 < \cdots < x_N = b \tag{1}$$

satisfies

$$\max_{i=1}^{N} |x_i - x_{i-1}| < \delta,$$

then

$$\max_{a \le x \le b} |f_N(x) - f(x)| < \epsilon,$$

where

$$f_N(x) = f(x_{i-1}) + \frac{f(x_i) - f(x_{i-1})}{x_i - x_{i-1}}(x - x_{i-1})$$

for

$$x_{i-1} \le x \le x_i, i = 1, 2, ..., N.$$

The proof is straightforward from the uniform continuity of continuous function on bounded closed interval.

To learn a spline function of degree 1 from data, it is a convention to represent the function as the linear combination of the spline basis functions

$$\phi_i(x) = \begin{cases} \frac{x - x_{i-1}}{x_i - x_{i-1}}, & x_{i-1} \le x \le x_i, \\ \frac{x - x_{i+1}}{x_i - x_{i+1}}, & x_i \le x \le x_{i+1}, \\ 0, & \text{otherwise}, \end{cases} \tag{2}$$

for $i = 1, 2, ..., N - 1$,

$$\phi_0(x) = \begin{cases} \frac{x - x_1}{x_0 - x_1}, & x_0 \le x \le x_1, \\ 0, & \text{otherwise}, \end{cases}$$

and

$$\phi_N(x) = \begin{cases} \frac{x - x_{N-1}}{x_N - x_{N-1}}, & x_{N-1} \le x \le x_N, \\ 0, & \text{otherwise}. \end{cases}$$

Using these spline basis functions to construct nonlinear block

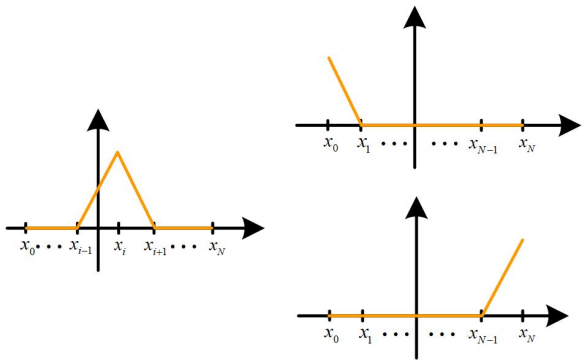

Fig. 2. Graph of basis function $\phi_i(x)$.

leads to the well known Look-Up Table (LUT) method in digital predistortion research field in wireless communication [15].

The set of all piece-wise linear continuous functions with the fixed partition (1) forms an $(N + 1)$-dimensional linear space. Therefore we realize that what type of nonlinear block to construct depends upon what basis to select in the linear space. One significant point of our basis function approximation network is that we do not use the conventional spline basis functions. Instead, we use others.

Denote the Rectified Linear Unit (ReLU) function

$$\text{ReLU}(x) = \begin{cases} x, & x \ge 0, \\ 0, & \text{otherwise}. \end{cases}$$

Let

$$h_i(x) = \text{ReLU}(x - x_i) \tag{3}$$

for $i = 0, 1, ..., N - 1$, and

$$h_N(x) = 1.$$

Theorem II.3. Let the partition (1) of the interval $[a, b]$ be fixed.

1) The set of all piece-wise linear continuous functions on $[a, b]$ forms an $(N + 1)$-dimensional linear space over the real number field $\mathbb{R}$.

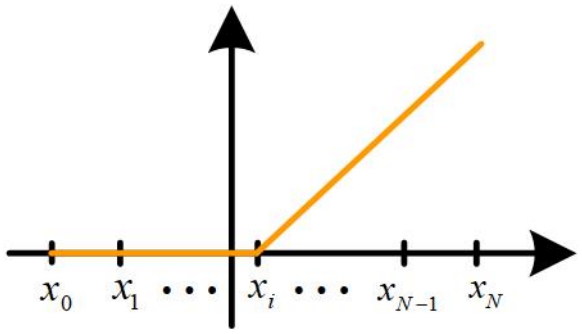

Fig. 3. Graph of basis function h_i.

2) The group $\{\phi_0(x), \phi_1(x), ..., \phi_N(x)\}$ is a basis of the linear space.
3) The group $\{h_0(x), h_1(x), ..., h_N(x)\}$ is another basis of the linear space.

We use the basis $\{h_0(x), h_1(x), ..., h_N(x)\}$. Consequently, the approximating function takes the form

$$v(x) = \sum_{i=0}^{N} w_i h_i(x),$$

where w_i is the corresponding weight parameter, as depicted in the following Fig. 4. This is our basis function approximation network (BFAN), one unit to be used for representing a static nonlinear function.

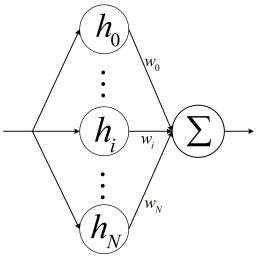

Fig. 4. Basis function approximation network (BFAN).

It is worthy to point out that, in Fig. 4, the same signal, no longer with different weights, enters all the $N+1$ basis functions $h_i(x), i = 0, 1, ..., N$. These basis functions, although each of them is exactly one type of "activation functions"–ReLU–in the usual neural network, are not used to imitate the behavior of neurons individually, but are used together to approximate locally a static nonlinear function hoped for in Theorem II.1. This is a main start point of our methodology.

C. Applying BFAN to Complex Valued Signal

Up to now, BFAN is defined to process real valued signal. In the context of communication system, complex I/Q signals are utilized. One complex valued signal can be simply seen as two real valued signals, the real and imaginary components. However, in our PIM cancellation practice, we use the following nonlinear structure Fig.5 to process complex valued signal, sufficiently considering the two components being coupled.

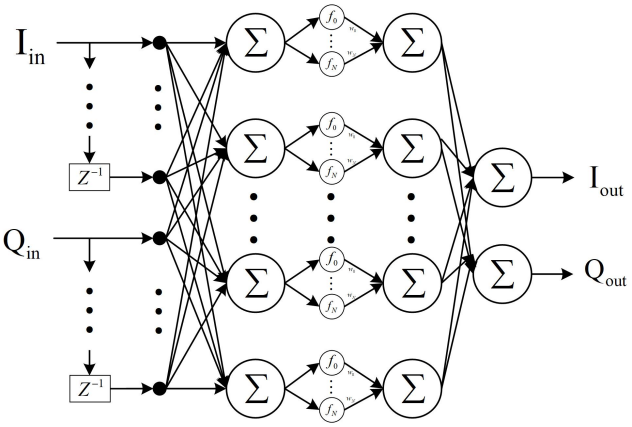

Fig. 5. BFAN for complex valued signal.

III. Building Symmetry Information into Network Structure

In this section, we propose two symmetrized networks based on the physical properties of PIM interference. We present the detailed mathematical expressions and identification algorithms for the cascaded models that utilize the symmetrized basis function approximation network to construct the nonlinear block.

A. Symmetrization Principle

A mapping from signal $x(\cdot)$ to signal $y(\cdot)$, denoted

$$y(\cdot) = \mathscr{A}[x(\cdot)],$$

is called to be odd, if

$$\mathscr{A}[-x(\cdot)] = -\mathscr{A}[x(\cdot)],$$

and even, if

$$\mathscr{A}[-x(\cdot)] = \mathscr{A}[x(\cdot)].$$

For clarity, we write the following simple facts as a lemma.

Lemma 1.

1) Linear mapping is odd.
2) If two mappings are odd, then the cascaded (composite) mapping of them is odd.

Firstly, we point out that, the system model in now PIM cancellation practice represents an odd mapping from input signals to output signals. In current state-of-the-art of PIM cancellation, LUT serves as the nonlinear block [10], whose input-output mapping relation between the complex valued signals $u(n)$ and $v(n)$ can be mathematically described by

$$v(n) = u(n) \sum_{i=0}^{N} \theta_i \phi_i(|u(n)|)$$

with the complex coefficients θ_i to be identified, where ϕ_i is defined in Eq. (2) and $|u(n)|$ means complex absolute value.

Although the mapping $u(\cdot) \mapsto v(\cdot)$ is nonlinear, it is an odd mapping. This can be seen from the simple calculation

$$\begin{aligned}&(-u(n))\sum_{i=0}^{N}\theta_i\phi_i(|-u(n)|)\\ =&-\left[u(n)\sum_{i=0}^{N}\theta_i\phi_i(|u(n)|)\right].\end{aligned}$$

Other linear blocks are automatically odd. Therefore, the total system model represents an odd mapping.

Secondly, from physical mechanism, it is reasonable to use an odd mapping to model the process from the transmitting signal to arouse PIM interference signal in receiving channel in wireless communication: if the input signal $x(\cdot)$ produces output signal $y(\cdot)$, then $-x(t)$ produces $-y(\cdot)$. Indeed, the odd-mapping phenomenon for PIM generating considered in this paper is observed by specifically designed experiment in laboratory and data analysis.

It is far from being trivial to incorporate the almost trivial fact of odd mapping into the architecture of system model in PIM cancellation technology. To do this we introduce the symmetrization skills.

Theorem III.1. Let $h(x)$ be an arbitrary function.

1) $h(x) - h(-x)$ is an odd function.
2) $h(x) + h(-x)$ is an even function, and the product $(h(x) + h(-x)) \cdot x$ is an odd function.

The proof is straightforward. Theorem III.1 1) and 2) lead to the odd and even symmetrization skills described in the two later subsections respectively.

Now we need to render odd our nonlinear blocks, which are not the LUT in now cancellation practice. To make our BFAN represent an odd mapping, the key is to symmetrize its nonlinear basis functions. We explain the principle in following.

Suppose that we want to learn a piece-wise linear continuous function defined on the interval $[-a, a]$ with partition

$$-a = x_{-N} < \cdots < x_{-1} < x_0 < x_1 < \cdots < x_N = a.$$

If we do not use the fact that the function to be identified is indeed an odd function, then from Theorem II.3 the dimension of linear space of all piece-wise linear continuous functions is $2N + 1$, and we have $2N + 1$ parameters $\theta_i, i = -N, -N + 1, ..., N$, to be learned. Here the parameters appear in the expression

$$\sum_{i=-N}^{N-1}\theta_i \cdot \mathrm{ReLU}(x - x_i) + \theta_N \cdot 1.$$

However, if we use the prior information that the function to be identified is indeed an odd function, then we can take the partition with $x_{-i} = -x_i$ and there are only N parameters $\lambda_i, i = 1, ..., N$, to be learned. Here the parameters appear in the expression

$$\sum_{i=0}^{N-1}\lambda_i \cdot [\mathrm{ReLU}(x - x_i) - \mathrm{ReLU}(-x - x_i)].$$

By Theorem III.1 1), $[\mathrm{ReLU}(x - x_i) - \mathrm{ReLU}(-x - x_i)]$ is an odd symmetrization of $\mathrm{ReLU}(x - x_i)$. In neural network language, we can also say that the network nodes $\mathrm{ReLU}(x - x_i)$ and $-\mathrm{ReLU}(-x - x_i)$ share the same weight λ_i.

We have explained how to successfully incorporate the prior odd-mapping information into the network design. Weight-sharing is one golden rule for neural network architecture design [22]. It absorbs the prior physical information on the real system; it reduces the number of parameters to be learned and then enhances the model's approximation accuracy and training efficiency.

B. Odd-Symmetrized Basis Function Approximation Network

Fig. 6. Symmetrized BFAN as nonlinear block.

1) Method Description: Denote by $x[n]$ and $y[n]$ the input and output signals of the cascaded model, by $a[n]$ the first intermediate signal after the first linear block, by $c[n]$ the second intermediate signal after the nonlinear block. The first linear block is a linear filter

$$\begin{aligned}u[n] &= (K * x)[n]\\ p[n] &= \begin{bmatrix} H(Z)I_u \\ H(Z)Q_u \end{bmatrix}\\ a[n] &= \Phi p[n],\end{aligned} \tag{4}$$

where $K \in \mathbb{C}^{M_K}$, M_K denotes the number of taps of the first linear filter, $H(Z) = [Z^{M_N-1}, \cdots, 1, Z^{-1}, \cdots, Z^{-M_N+1}]$, $\Phi \in \mathbb{C}^{M_a \times M_P}$, and $I_u = \frac{u+u^*}{2}$ and $Q_u = \frac{u-u^*}{2}$ represent the in-phase and quadrature components of $u[n]$.

The second nonlinear block maps the value $a[n]$ of the first intermediate signal $a(\cdot)$ to the value $c[n]$ of the second intermediate signal $c(\cdot)$ according to

$$c[n] = [f(a_1[n]), \cdots, f(a_{M_a}[n])]^{\mathrm{T}}$$

where

$$\begin{aligned}f(a_i[n]) &= \sum_{k=0}^{N} w_k f_k(a_i[n])\\ &= \sum_{k=0}^{N} w_k (h_k(a_i[n]) - h_k(-a_i[n])).\end{aligned} \tag{5}$$

Note that h_k is the basis function defined in Eq. (3), and f_k **is the odd symmetrization of** h_k.

The last block is a linear filter

$$\begin{aligned}v[n] &= \Psi c[n]\\ y[n] &= (G * v)[n],\end{aligned} \tag{6}$$

where $G \in \mathbb{C}^{M_G}$, M_G denotes the number of taps of the second linear filter, and $\Psi \in \mathbb{C}^{M_a \times 1}$.

In summary, there are five sets of parameters to be identified. They are Φ, W, and Ψ for the nonlinear block, and K and G for the two linear blocks. The detailed model is depiced in Fig. 6.

2) Learning rules and complexity analysis: Our goal is to use the known data of reference input $x[\cdot]$ and reference output $d[\cdot]$ to estimate the five sets of parameters. Let

$$e[n] = d[n] - y[n],$$

the error of the model output $y[\cdot]$ relative to the reference output data $d[\cdot]$, and take the cost index

$$J(n) = \frac{1}{2} e[n] e^*[n].$$

Then we use the least mean square (LMS) algorithm to adaptively estimate the parameters [23]. The update for each corresponding parameters from step n to step $n+1$ is defined by

$$\begin{aligned}
K_i(n+1) &= K_i(n) - \mu_{K_i} \nabla_{K_i} J(n) \\
\Phi_{i,j}(n+1) &= \Phi_{i,j}(n) - \mu_{\Phi_{i,j}} \nabla_{\Phi_{i,j}} J(n) \\
W_i(n+1) &= W_i(n) - \mu_{W_i} \nabla_W J(n) \\
\Psi_i(n+1) &= \Psi_i(n) - \mu_{\Psi_i} \nabla_{\Psi_i} J(n) \\
G_i(n+1) &= G_i(n) - \mu_{G_i} \nabla_{G_i} J(n).
\end{aligned}$$

These μ's represent the learning rates for each parameter.

Note that the gradient is in the sense of Wirtinger complex derivative [24]. We write them according to the order of backward propagation as follows.

$$\begin{aligned}
\nabla_{G_i} J(n) &= \frac{\partial e[n] e^*[n]}{\partial G_i^*} = -e[n] \frac{\partial y^*[n]}{\partial G_i^*} - 0 \cdot e^*[n] \\
&= -e[n] v^*[n+1-i]. \qquad (7)
\end{aligned}$$

$$\begin{aligned}
\nabla_{\Psi_i} J(n) &= \frac{\partial e[n] e^*[n]}{\partial \Psi_i^*} = -e[n] \left(G * \frac{\partial v^*}{\partial \Psi_i^*} \right)[n] \\
&= -e[n] \left(G^* * c_i^* \right)[n]. \qquad (8)
\end{aligned}$$

$$\begin{aligned}
\nabla_{W_i} J(n) &= \frac{\partial e[n] e^*[n]}{\partial W_i^*} = -e[n] \left(G^* * \left(\Psi^* \frac{\partial c^*}{\partial W_i^*} \right) \right) \\
&= -e[n] \left(G^* * \left(\Psi^* \begin{bmatrix} f_k^*(a_1) \\ \vdots \\ f_k^*(a_{M_a}) \end{bmatrix} \right) \right)[n]. \qquad (9)
\end{aligned}$$

$$\begin{aligned}
\nabla_{\Phi_{i,j}} J(n) &= \frac{\partial e[n] e^*[n]}{\partial \Phi_{i,j}^*} \\
&= -e[n] \left(G^* * \left(\Psi^* \frac{\partial c^*}{\partial \Phi_{i,j}^*} \right) \right)[n] \\
&= -e[n] \left(G^* * \left(\Psi^* \begin{bmatrix} 0 \\ f^{*'}(a_j) p_i^* \\ 0 \end{bmatrix} \right) \right)[n].
\end{aligned}$$

$$\begin{aligned}
\nabla_{K_i} J[n] &= \frac{\partial e[n] e^*[n]}{\partial K_i^*} \\
&= -e[n] \frac{\partial y^*[n]}{\partial K_i^*} - \frac{\partial y[n]}{\partial K_i^*} e^*[n],
\end{aligned}$$

where $\frac{\partial y^*[n]}{\partial K_i^*}$ and $\frac{\partial y^{\lfloor}[n]}{\partial K_i^*}$ are computed as in (10).

C. Even-Symmetrized Basis Function Approximation Network

1) Method Description: While the symmetrization method of subsection III-B is based on Theorem III.1 1), the one here on Theorem III.1 2). So except the following two formulae

$$c[n] = \left[f\left(a_1[n]\right) a_1[n], \cdots, f\left(a_{M_a}[n]\right) a_{M_a}[n] \right]^{\mathrm{T}}$$

and

$$\begin{aligned}
f\left(a_i[n]\right) &= \sum_{k=1}^{N} w_k f_k(a_i[n]) \\
&= \sum_{k=1}^{N} w_k \left(h_k(a_i[n]) + h_k(-a_i[n]) \right),
\end{aligned}$$

all other ones are the same as in subsection III-B. Note that h_k is the basis function defined in Eq. (3), and f_k **is the even symmetrization of** h_k.

2) Learning rules and complexity analysis: The gradient computations for parameter G and Ψ are identical to (7) and (8). The other three are computed as follows.

$$\begin{aligned}
\nabla_{W_i} J(n) &= \frac{\partial e[n] e^*[n]}{\partial W_i^*} = -e[n] \left(G^* * \left(\Psi^* \frac{\partial c^*}{\partial W_i^*} \right) \right) \\
&= -e[n] \left(G^* * \left(\Psi^* \begin{bmatrix} f_k^*(a_1) a_1^* \\ \vdots \\ f_k^*(a_{M_a}) a_{M_a}^* \end{bmatrix} \right) \right)[n].
\end{aligned}$$

$$\begin{aligned}
\nabla_{\Phi_{i,j}} J(n) &= \frac{\partial e[n] e^*[n]}{\partial \Phi_{i,j}^*} \\
&= -e[n] \left(G^* * \left(\Psi^* \frac{\partial c^*}{\partial \Phi_{i,j}^*} \right) \right)[n] \\
&= -e[n] \left(G^* * \left(\Psi^* \begin{bmatrix} 0 \\ f^{*'}(a_j) a_j^* p_i^* \\ 0 \end{bmatrix} \right) \right)[n].
\end{aligned}$$

$$\begin{aligned}
\nabla_{K_i} J(n) &= \frac{\partial (e[n] e^*[n])}{\partial K_i^*} \\
&= -e[n] \frac{\partial y^*[n]}{\partial K_i^*} - \frac{\partial y[n]}{\partial K_i^*} e^*[n],
\end{aligned}$$

where $\frac{\partial y^*[n]}{\partial K_i^*}$ and $\frac{\partial y[n]}{\partial K_i^*}$ are computed according to (11).

IV. Test Experiments

In this section we evaluate our symmetrized BFAN through comparing it with other various models in terms of PIM cancellation performance. The data are collected in a controlled laboratory environment using actual available equipment. These data undergo a series of preprocessing steps, including filtering, frequency shifting, and normalization, before being used in experimental tests for evaluation. The reference PIM signals are measured in an environment isolated from external signals.

Our symmetrized BFAN helps to observe two interesting phenomena on the specific nonlinearity related to PIM in wireless communication. One is that the PIM nonlinear interference

TABLE I
COMPUTATIONAL COMPLEXITY OF ODD SYMMETRIZED BFAN IN CANCELLATION AND PARAMETER UPDATE STAGE

Operation		Additions/sample	Multiplications/sample
Cancellation	$u(n)$	$2M_K$	$2M_K-2$
	$v(n)$	$(M_p+N+2)M_a$	$(M_p+N+1)M_a-2$
	$y(n)$	$2M_G$	$2M_G-2$
Parameter updates	$\nabla_G J(n)$	$2M_G$	0
	$\nabla_\Psi J(n)$	$(2M_G+1)M_a$	$(2M_G-1)M_a$
	$\nabla_W J(n)$	$(2M_G+2M_a+1)N$	$(2M_G+2M_a-3)N$
	$\nabla_\Phi J(n)$	$(2M_G+4)M_pM_a$	$(2M_G-1)M_pM_a$
	$\nabla_K J(n)$	$2M_K(2M_G+M_aM_p+3M_a+1)$	$2M_K(2M_G+M_aM_p+M_a-1)$

occurs only when the power of the transmitting signal transcends a threshold value; the other is that the needed number of the nonlinear units in the network may physically corresponds to the number of the PIM sources. These phenomena can clearly emerge to us thanks to our symmetrized BFAN being able to isolate the simple memoryless nonlinearity.

A. SISO Case

The transmitting signals are in the frequency bands of 70-75 MHz and 117.5-122.5 MHz. The PIM interference, primarily composed of third-order and fifth-order intermodulation distortion, are found within the receiving frequency band, which ranges from -40-40 MHz. The level of the PIM interference signal reads 19.13 dB.

We compare our proposed symmetrized BFAN model with two other ones, the memory polynomial (MP) and Wiener-Hammerstein LUT-based (WHLUT) models. These two are high-performance models in PIM interference cancellation; see [10].

Fig. 7 shows the signals spectra for the three canceller models, each with the same number of memory taps so that they have the same memory depth. It can be seen that the cancellation achieved by the symmetrized BFAN reaches 17.33 dB. Meanwhile, the cancellation of the WHLUT model is around 15.11 dB, and the MP model is 13.80 dB. Table III presents the cancellation results with three models, clearly indicating that our proposed model surpasses both the MP model and WHLUT model in performance. Additionally, the number of parameters for each model is detailed in Table III, demonstrating that our proposed model achieves the best performance with the fewest parameters to be estimated.

The computational complexities of the odd-symmetrized BFAN model are presented in Table I in terms of the number of addition and multiplication operations per processed sample. Note that the numbers M_K, M_p, M_a, N and M_G are defined

$$\frac{\partial y^*[n]}{\partial K_i^*} = \left(G^* * \left(\Psi^* \begin{bmatrix} f^{'}(a_1) & & \\ & \ddots & \\ & & f^{'}(a_{M_a}) \end{bmatrix}^* \Phi^* \begin{bmatrix} H(Z)\frac{x^*[n+1-i]}{2} \\ H(Z)\frac{x^*[n+1-i]}{2} \end{bmatrix} \right) \right)[n]$$

$$\frac{\partial y[n]}{\partial K_i^*} = \left(G * \left(\Psi \begin{bmatrix} f^{'}(a_1) & & \\ & \ddots & \\ & & f^{'}(a_{M_a}) \end{bmatrix} \Phi \begin{bmatrix} H(Z)\frac{x^*[n+1-i]}{2} \\ H(Z)\frac{-x^*[n+1-i]}{2} \end{bmatrix} \right) \right)[n] \tag{10}$$

$$\frac{\partial y^*[n]}{\partial K_i^*} = \left(G^* * \left(\Psi^* \begin{bmatrix} f^{'}(a_1)a_1+f(a_1) & & \\ & \ddots & \\ & & f^{'}(a_{M_a})a_{M_a}+f(a_{M_a}) \end{bmatrix}^* \Phi^* \begin{bmatrix} H(Z)\frac{x^*[n+1-i]}{2} \\ H(Z)\frac{x^*[n+1-i]}{2} \end{bmatrix} \right) \right)[n]$$

$$\frac{\partial y[n]}{\partial K_i^*} = \left(G * \left(\Psi \begin{bmatrix} f^{'}(a_1)a_1+f(a_1) & & \\ & \ddots & \\ & & f^{'}(a_{M_a})a_{M_a}+f(a_{M_a}) \end{bmatrix} \Phi \begin{bmatrix} H(Z)\frac{x^*[n+1-i]}{2} \\ H(Z)\frac{-x^*[n+1-i]}{2} \end{bmatrix} \right) \right)[n] \tag{11}$$

TABLE II
COMPUTATIONAL COMPLEXITY OF EVEN-SYMMETRIZED BFAN IN CANCELLATION AND PARAMETER UPDATE STAGE

Operation		Additions/sample	Multiplications/sample
Cancellation	$u(n)$	$2M_K$	$2M_K - 2$
	$v(n)$	$(M_p + N + 3)M_a$	$(M_p + N + 1)M_a - 2$
	$y(n)$	$2M_G$	$2M_G - 2$
Parameter updates	$\nabla_G J(n)$	$2M_G$	0
	$\nabla_\Psi J(n)$	$(2M_G + 1)M_a$	$(2M_G - 1)M_a$
	$\nabla_W J(n)$	$(2M_G + 3M_a + 1)N$	$(2M_G + 2M_a - 3)N$
	$\nabla_\Phi J(n)$	$(2M_G + 5)M_p M_a$	$(2M_G - 1)M_p M_a$
	$\nabla_K J(n)$	$2M_K(2M_G + M_a M_p + 4M_a + 1)$	$2M_K(2M_G + M_a M_p + 2M_a - 1)$

TABLE III
PERFORMANCE AND NUMBER OF PARAMETERS IN CASE 1

Model	Cancellation	Parameters
BFAN	17.33 dB	116
WHLUT	15.11 dB	154
MP	13.80 dB	131

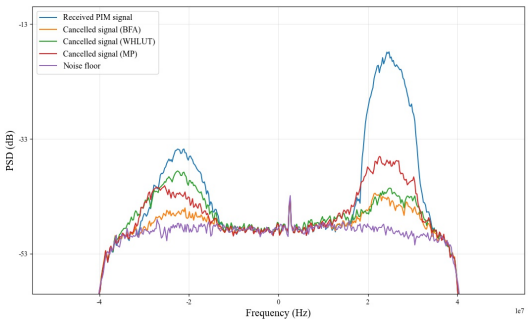

Fig. 7. PIM cancellation performance comparison in a SISO system.

in (4), (5), and (6) respectively.

B. MIMO Case

Fig. 8. Symmetrized BFAN in MIMO system.

Fig. 9 shows the power spectral densities of the received PIM signal and the signals after cancellation with the models in an eight-input-eight-output system. In all channels, both the odd-symmetrized BFAN and the even-symmetrized BFAN models achieve better cancellation compared to the BFAN model without symmetrization, maintaining the same level of computational complexity. Especially, the odd-symmetrized BFAN model shows markedly improved cancellation performance over the BFAN model without symmetrization in Channel 4, illustrating the considerable advantages of building the symmetry information into network structure. Additionally,

TABLE IV
CANCELLATION PERFORMANCE IN CASE 2

Model	Even BFAN	Multiple basis	Odd BFAN	BFAN
Channel 0	18.61 dB	18.55 dB	18.55 dB	18.39 dB
Channel 1	12.79 dB	12.78 dB	12.78 dB	12.36 dB
Channel 2	20.02 dB	19.93 dB	19.98 dB	19.84 dB
Channel 3	14.31 dB	14.29 dB	14.30 dB	14.04 dB
Channel 4	19.14 dB	19.01 dB	19.08 dB	17.11 dB
Channel 5	17.20 dB	17.12 dB	17.14 dB	16.86 dB
Channel 6	19.01 dB	18.94 dB	18.88 dB	18.49 dB
Channel 7	16.84 dB	16.79 dB	16.84 dB	16.24 dB
Average	17.24 dB	17.18 dB	17.19 dB	16.67 dB

the detailed cancellation results of all models in MIMO system are presented in Table IV.

In general, the nonlinear blocks located in different places in the BFAN represent different nonlinear functions, necessitating different parameters for each nonlinear basis. This means that, as shown in Fig.8, the weights assigned to each basis in different blocks should ideally vary. However, in the scenario of PIM cancellation, the nonlinear characteristics may be generated from a single or a few PIM sources. If so, to let some different blocks share the same nonlinear parameters can effectively reduce computational complexity. As per (9), calculating each gradient necessitates $(2M_G + 2M_a + 1)N$ multiplications and $(2M_G + 2M_a - 3)N$ additions. Then to let all nonlinear blocks share the same weights can decrease multiplications and additions required in the identification algorithm by $(M_a - 1)(2M_G + 2M_a + 1)N$ and $(M_a - 1)(2M_G + 2M_a - 3)N$ times respectively. For the specific PIM data in our experiments, it turns out that multiple different nonlinear functions do not necessarily lead to performance improvement relative to single one shared by all nonlinear blocks.

Since we use only one shared static nonlinear function in the network, this provides us a valuable opportunity to observe the nonlinear characteristic of PIM generating process. Fig. 10 shows the identified nonlinear function in MIMO model. It validates with experiment and data analysis one physical property of PIM generating in folklore: PIM is one kind of weak nonlinear phenomenon; it is aroused only when the power of the transmitting signal transcends a critical point. The curve exhibits a transition from almost zero in low power region to nonlinearity in high power region, demonstrating

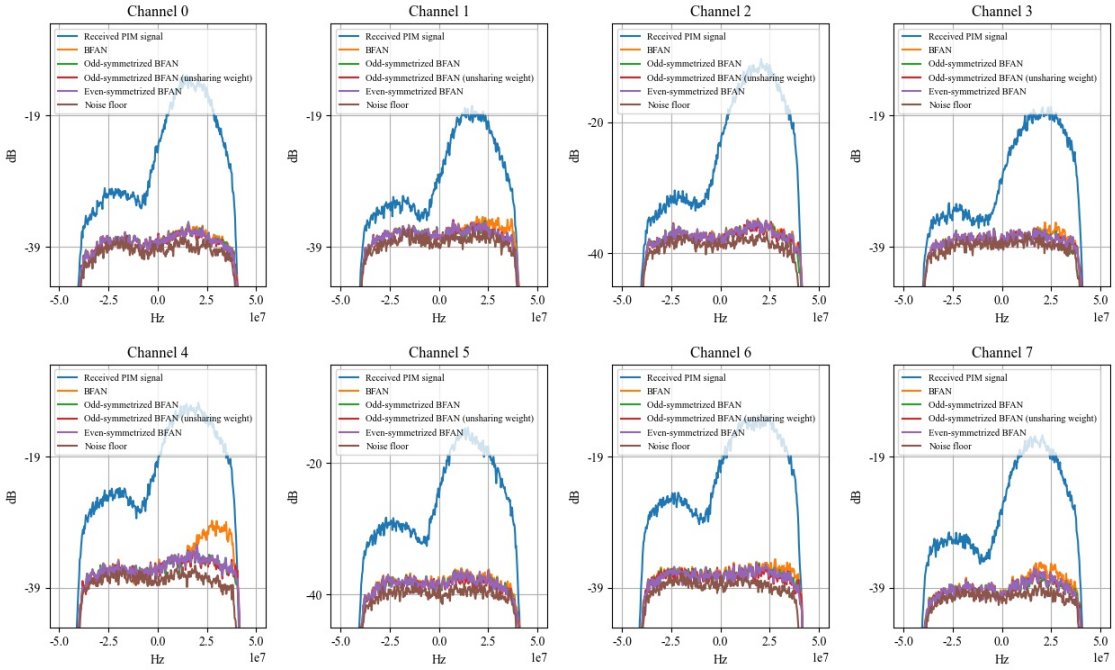

Fig. 9. Comparison of PIM cancellation performance: odd-symmetrized, even-symmetrized, and non-symmetrized BFANs in a MIMO system.

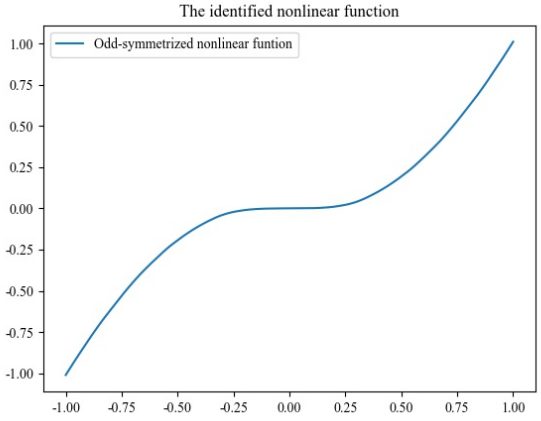

Fig. 10. Nonlinear function identified from actual data.

a critical point inflection and weak nonlinear characteristics. This trend aligns with the physical mechanisms of PIM in actual communication systems, where transmitting signal with low power hardly induces PIM, and the power of PIM signal is observed clearly as transmitting power increases significantly.

V. Conclusion

We propose the symmetrized basis function approximation network to cancel digitally the PIM interference in simultaneous transmit-receive system in wireless communication. The work contains four contributions as follows.

- Upon rigorous mathematical analysis, we propose the basis function approximation network.
- We build symmetry into basis function approximation network using symmetrization of function, a simple mathematical skill.
- The critical point phenomenon is explained with experiment and data analysis.
- A modelling rule is suggested that the number of the nonlinear functions used in the system model should correspond to the number of the PIM sources.

The symmetrized basis function approximation network provides a worthwhile solution for digital cancellation of PIM interference in wireless communication. Its design principle is advanced in that it uses rigorous mathematical notions, such as nonlinear operator representation, basis transformation of linear space, symmetrization of mapping, and it builds important prior physical information, such as the symmetry of PIM generating process, the number of PIM sources, into system modelling.

The paper focuses on the nonlinearity. Currently, the memory issue is dealt with by finite memory model. It is an interesting future research direction along this work to adopt infinite memory model, such as infinite impulse response linear filter, in the linear block, while the nonlinear block uses the symmetrized basis function approximation network.

Acknowledgement

The authors would like to thank Lei Wang with HUAWEI for his help in carrying out the practical signal measurements. He also provided various valuable comments and suggestions to the early version of the paper. We learn much from him about the technical status on PIM cancellation practice in wireless communication.

REFERENCES

[1] H. E. Hamoud, T. Reveyrand, S. Mons, and E. Ngoya, "A comparative overview of digital predistortion behavioral modeling for multi-standards applications," in *2018 International Workshop on Integrated Nonlinear Microwave and Millimetre-wave Circuits (INMMIC)*, 2018, pp. 1–3.

[2] M. Z. Waheed, D. Korpi, L. Anttila, A. Kiayani, M. Kosunen, K. Stadius, P. P. Campo, M. Turunen, M. Allén, J. Ryynänen, and M. Valkama, "Passive intermodulation in simultaneous transmit–receive systems: Modeling and digital cancellation methods," *IEEE Transactions on Microwave Theory and Techniques*, vol. 68, no. 9, pp. 3633–3652, 2020.

[3] L. Ljung, *System Identification: Theory for the User*, 2nd ed. Upper Saddle River, NJ, USA: Prentice Hall, 1999.

[4] L. Zhang, H. Wang, S. He, H. Wei, Y. Li, and C. Liu, "A segmented polynomial model to evaluate passive intermodulation products from low-order pim measurements," *IEEE Microwave and Wireless Components Letters*, vol. 29, no. 1, pp. 14–16, 2019.

[5] M. Z. Waheed, D. Korpi, L. Anttila, A. Kiayani, and M. Valkama, "Passive intermodulation in simultaneous transmit-receive systems: Modeling and digital cancellation methods," *IEEE Transactions on Microwave Theory and Techniques*, vol. PP, no. 99, pp. 1–1, 2020.

[6] L. Zhang, H. Wang, J. Shen, H. Wei, X. Wang, Y. Li, and C. Liu, "A composite exponential model to characterize nonlinearity causing passive intermodulation interference," *IEEE Transactions on Electromagnetic Compatibility*, vol. 61, no. 2, pp. 590–594, 2019.

[7] D. S. Kozlov, A. P. Shitvov, A. G. Schuchinsky, and M. B. Steer, "Passive intermodulation of analog and digital signals on transmission lines with distributed nonlinearities: Modelling and characterization," *IEEE Transactions on Microwave Theory and Techniques*, vol. 64, no. 5, pp. 1383–1395, 2016.

[8] B. Fehri and S. Boumaiza, "Baseband equivalent volterra series for behavioral modeling and digital predistortion of power amplifiers driven with wideband carrier aggregated signals," *IEEE Transactions on Microwave Theory and Techniques*, vol. 62, no. 11, pp. 2594–2603, 2014.

[9] A. Varano, D. Peumans, and Y. Rolain, "Rova: A practical, parsimonious volterra model for modulated microwave systems," *IEEE Transactions on Microwave Theory and Techniques*, vol. 72, no. 6, pp. 3701–3710, 2024.

[10] P. P. Campo, L. Anttila, D. Korpi, and M. Valkama, "Cascaded spline-based models for complex nonlinear systems: Methods and applications," *IEEE Transactions on Signal Processing*, vol. 69, pp. 370–384, 2021.

[11] A. S. Tehrani, H. Cao, S. Afsardoost, T. Eriksson, M. Isaksson, and C. Fager, "A comparative analysis of the complexity/accuracy tradeoff in power amplifier behavioral models," *IEEE Transactions on Microwave Theory and Techniques*, vol. 58, no. 6, pp. 1510–1520, 2010.

[12] M. Scarpiniti, D. Comminiello, R. Parisi, and A. Uncini, "Novel cascade spline architectures for the identification of nonlinear systems," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 62, no. 7, pp. 1825–1835, 2015.

[13] S. Michele, C. Danilo, P. Raffaele, and U. Aurelio, "Hammerstein uniform cubic spline adaptive filters: Learning and convergence properties," *Signal Processing*, vol. 100, pp. 112–123, 2014.

[14] F. M. Ghannouchi and O. Hammi, "Behavioral modeling and predistortion," *IEEE Microwave Magazine*, vol. 10, no. 7, pp. 52–64, 2009.

[15] P. P. Campo, A. Brihuega, L. Anttila, M. Turunen, D. Korpi, M. Allén, and M. Valkama, "Gradient-adaptive spline-interpolated lut methods for low-complexity digital predistortion," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 68, no. 1, pp. 336–349, 2021.

[16] S. Chen, X. Hong, J. Gao, and C. J. Harris, "Complex-valued b-spline neural networks for modeling and inverting hammerstein systems," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 9, pp. 1673–1685, 2014.

[17] Q.-J. Zhang, K. Gupta, and V. Devabhaktuni, "Artificial neural networks for rf and microwave design - from theory to practice," *IEEE Transactions on Microwave Theory and Techniques*, vol. 51, no. 4, pp. 1339–1350, 2003.

[18] H. Wu, W. Chen, X. Liu, Z. Feng, and F. M. Ghannouchi, "A uniform neural network digital predistortion model of rf power amplifiers for scalable applications," *IEEE Transactions on Microwave Theory and Techniques*, vol. 70, no. 11, pp. 4885–4899, 2022.

[19] Z.-Y. Luo, J.-L. Zhou, Y.-F. Pu, and L. Li, "A class of augmented complex-value flann adaptive algorithms for nonlinear systems," *Neurocomputing*, vol. 520, pp. 331–341, 2023.

[20] I. W. Sandberg and L. Xu, "Uniform approximation of multidimensional myopic maps," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 44, no. 6, pp. 477–500, 1997.

[21] I. Sandberg, "Uniform approximation and the circle criterion," *IEEE Transactions on Automatic Control*, vol. 38, no. 10, pp. 1450–1458, 1993.

[22] Y. L. Cun, B. Boser, J. S. Denker, R. E. Howard, W. Habbard, L. D. Jackel, and D. Henderson, *Handwritten digit recognition with a back-propagation network*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1990, p. 396–404.

[23] S. Haykin, *Adaptive Filter Theory*. Upper Saddle River, NJ: Prentice Hall, 1996.

[24] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*. Englewood Cliffs, NJ: Prentice Hall, 2000.