Model Identification for Digital Predistortion of Power Amplifier With Signed Regressor Algorithm

Ning Guan\textsuperscript{\textcircled{C}}, Nan Wu\textsuperscript{\textcircled{D}}, Member, IEEE, and Hua Wang\textsuperscript{\textcircled{D}}, Member, IEEE

Abstract—In this letter, a low-complexity approach is proposed to estimate model parameters of digital predistortion (DPD). It takes the advantage of low implementation cost of the signed regressor algorithm and eliminates most multiplications required. A Walsh–Hadamard transform is employed to avoid the correlation between signed DPD basis functions. Performance of the proposed approach is evaluated via experiments, with a two-carrier long term evolution-advanced signal on a 3.5-GHz power amplifier. It is shown that compared with the conventional least-square-based method the proposed approach can achieve similar linearization and convergence performance with much lower computational complexity.

Index Terms—Digital predistortion (DPD), model identification, signed regressor algorithm (SRA).

I. INTRODUCTION

RAPIDLY increasing user data rate is pushing the modern communication system towards wider bandwidth and higher spectral efficiency. Accordingly, the transmission signal becomes more sensitive to a nonlinear distortion of the RF power amplifier (PA). To enhance the linearity of PA while maintaining a high power efficiency, digital predistortion (DPD) technique [1] has been widely studied and employed. The basic principle of DPD is to introduce proper distortion on the transmitted signal, so that the whole transmitter path (DPD + PA) becomes linear. As it is well known, accurately finding the parameters of DPD model is crucial to guarantee a good linearization performance. Conventionally, indirect learning [2] and direct learning [3] approaches are utilized and both of them are mainly based on the least-square (LS) algorithm. However, the LS algorithm incorporates computational intensive operations, such as matrix multiplication and inversion, leading to large chip area and/or high power consumption. Future challenging scenario, such as small cell base station and mobile transmitter, requires low power consumption of DPD itself, and the resource demanding LS-based estimator becomes less affordable. Reducing the complexity of parameter estimation without sacrificing linearization performance is becoming an important issue in DPD.

Several recently published papers have been focusing on this problem [4]–[6]. In [4], a decorrelation-based approach is proposed to estimate DPD parameters in a concurrent dual-band scenario. This approach is, in fact, a least-mean-square (LMS)-based algorithm and its convergence speed is affected by mutual correlation between DPD basis functions. Accordingly, an additional orthogonalization procedure is required before learning each iteration to guarantee fast convergence. Wang et al. [5], propose a modified version of the conventional direct learning DPD, in which the time-varying input signal-filled matrix is replaced by a constant autocovariance matrix, which reduces the number of multiplications. While this approach significantly simplifies the algorithm, the ability of DPD to track signal characteristic variation is lost as well. In [6], a simultaneous perturbation stochastic approximation algorithm is introduced to avoid gradient calculation. However, the convergence speed of this approach is slow.

In this letter, based on our previous work on finding DPD parameters with undersampled and real-valued feedback signal [7], we propose a signed regressor algorithm (SRA)-based approach to reduce the computational overhead of DPD model identification. Experimental results show that compared with the conventional LS-based algorithm, the proposed approach can achieve similar linearization performance and convergence speed, with much lower computational complexity.

II. SYSTEM MODEL

The simplified block diagram of the DPD system is illustrated in Fig. 1. We consider a DPD model that has $K$-parameters and its output is linear with respect to its parameters. Denoting the input and output of DPD, and the full-rate output of PA as $x(n) = x_r(n) + jx_i(n)$, $x_{pd}(n)$, and $y(n) = y_r(n) + jy_i(n)$, respectively, the input–output relationship of PA can be described as

\[
y(n) = g(x_{pd}(n)) = g(\bar{x}(n), w)
\]

where $g(\cdot)$ is the unknown PA characteristic, $\bar{x}(n) = [x(n), \ldots, x(n-M)]^T$, $M$ is the memory depth of DPD, and $w = [w_0^T, \ldots, w_K^T]^T$ is the DPD model parameters vector.
To estimate $\mathbf{w}$, the LS-based algorithm using undersampled and real-valued feedback signal $y_r(Dn)$ can be adopted [7]

$$\hat{\mathbf{w}}_{p+1} = \hat{\mathbf{w}}_p + \mu [\mathbf{I}_K, j\mathbf{I}_K] [\mathbf{X}^T \mathbf{X}]^{-1} \mathbf{X}^T e$$  \hspace{1cm} (2)

where $\hat{\mathbf{w}}_p$ is the estimated parameters in the $p$th iteration, $\mu$ is the step size, $\mathbf{I}_K$ is the $K$-order identity matrix. Regression matrix $\mathbf{X} \in \mathbb{R}^{N \times 2K}$, and error vector $e \in \mathbb{R}^{N \times 1}$ are constructed as

$$\mathbf{X} = [\text{Re}\{\Phi\}, -\text{Im}\{\Phi\}]$$
$$\Phi = [\phi_1(\mathbf{x}(Dn)), \ldots, \phi_K(\mathbf{x}(Dn))], \quad n = 0, \ldots, N - 1$$
$$e = \left[ x_r(0) - \frac{y_r(0)}{G}, \ldots, x_r[D(N - 1)] - \frac{y_r[D(N - 1)]}{G} \right]^T$$  \hspace{1cm} (3)

where $\text{Re}\{\cdot\}$ and $\text{Im}\{\cdot\}$ denote the real and imaginary parts, respectively. $D$ is the undersampling factor, $\phi_k(\cdot)$ are the basis functions of DPD model, and $G$ is the expected gain of linearized PA.

The problem with the LS-based algorithm (2) is it's high computational complexity. To ensure the accuracy of estimation and avoid overfitting, $N \gg 2K$ should be provided for $\mathbf{X}$ and consequently a large number of multiplications are required to calculate the matrix product. As an alternative, LMS-based implementation can be used to avoid the calculation of inverse autocovariance matrix $(\mathbf{X}^T \mathbf{X})^{-1}$ in (2)

$$\hat{\mathbf{w}}_{p+1} = \hat{\mathbf{w}}_p + \mu [\mathbf{I}_K, j\mathbf{I}_K] \mathbf{X}^H e.$$  \hspace{1cm} (4)

However, as only the first-order approximation is utilized, and regression matrix $\mathbf{X}$ is usually ill-conditioned, (4) experiences a much slower convergence speed, which is not applicable in tracking the fast variation of PA behavior.

III. Signed Regressor Algorithm-Based DPD

A. Proposed Approach

To reduce the complexity of the LS algorithm while maintaining an excellent linearization and convergence performance, we propose to modify (2) using SRA [8]. Originally proposed for linear LMS adaptive filtering, the main idea of SRA is to replace $\mathbf{X}^T$ by its clipped version, that is

$$\hat{\mathbf{w}}_{p+1} = \hat{\mathbf{w}}_p + \mu [\mathbf{I}_K, j\mathbf{I}_K] \text{sgn}(\mathbf{X}) \text{sgn}(\mathbf{X})^T e$$  \hspace{1cm} (5)

where $\text{sgn}(\cdot)$ is the signum function, and $\text{sgn}(\mathbf{X})$ denotes applying a signum function on each element of $\mathbf{X}$. By clipping the regression matrix $\mathbf{X}$ to binary value ±1, multiplications among basis functions and error vector are eliminated, which considerably reduce the implementation cost.

Unfortunately, directly applying SRA in DPD leads to iteration divergence because the function $\text{sgn}(\cdot)$ creates duplicated columns in a signed regression matrix $\text{sgn}(\mathbf{X})$. To understand this issue, consider the linear term $x(n)$ and third-order memoryless nonlinear term $x(n)x(n)^2$, which are commonly included in the DPD model. Since $|x(n)|^2 \geq 0$, we always have

$$\text{sgn}(\text{Re}(x(Dn)|x(Dn)|^2)) = \text{sgn}(\text{Re}(x(Dn)))$$  \hspace{1cm} (6)

which renders two or more columns in $\text{sgn}(\mathbf{X})$ to become perfectly the same. Under such circumstance, the estimated parameters $\hat{\mathbf{w}}_p$ are unreliable.

To avoid this problem, we propose to apply SRA in a Walsh–Hadamard transformation (WHT) domain, rather than in time domain. More specifically, both $\mathbf{X}$ and $e$ are first processed by a unitary WHT matrix $\mathbf{T} \in \mathbb{R}^{N \times N}$, whose $[n,m]$th element $T_{nm}$ is either +1 or −1 and is only determined by $N$, which is the row size of $\mathbf{X}$. Then, the transformed regression matrix $\mathbf{Z} = \mathbf{TX}$ is clipped by the $\text{sgn}(\cdot)$, resulting in

$$\hat{\mathbf{w}}_{p+1} = \hat{\mathbf{w}}_p + \mu [\mathbf{I}_K, j\mathbf{I}_K] [\text{sgn}(\mathbf{Z}) \mathbf{Z}]^{-1} \text{sgn}(\mathbf{Z}) \mathbf{T} e.$$  \hspace{1cm} (7)

The reason for introducing WHT is that it gaussianizes the distribution of basis functions. The $(n,k)$th element of $\mathbf{Z}$ is

$$Z_{n,k} = \sum_{m=0}^{N-1} T_{nm} X_{m,k}$$  \hspace{1cm} (8)

where

$$X_{m,k} = \begin{bmatrix} \text{Re}[\phi_k(\mathbf{x}(Dm))] \\ \text{Im}[\phi_k(\mathbf{x}(Dm))] \end{bmatrix}, \quad k = 1, \ldots, K$$  \hspace{1cm} (9)

For $X_{m,k}$ with the same column index $k$, these elements are generated by the same basis function; therefore, all of them are identically distributed. Moreover, when the undersampling factor $D \gg 1$, $X_{m,k}$ with a different row index $m$ is widely spaced in the time axis, the elements from adjacent rows $X_{m,k}$ and $X_{m+1,k}$ become independent. Hence, as $T_{nm} = \pm 1$ does not affect the distribution of $X_{m,k}$, $Z_{n,k}$ can be interpreted as sum of $N$ i.i.d random variables. According to the central limit theorem, with sufficiently large $N$, all $Z_{n,k}$ become Gaussian variables, regardless of the original distribution of $X_{m,k}$, that is

$$Z_{n,k} \sim N(0, \sigma_n^2), \quad n = 0, \ldots, N - 1$$  \hspace{1cm} (10)

where $\sigma_n$ is the standard deviation of column $k$. Similar conclusion can also be applied for transformed error vector $\mathbf{T} e$. Since all elements of $\mathbf{Z}$ and $\mathbf{T} e$ become Gaussian variables, we have [9]

$$\text{sgn}(\mathbf{Z})^T \mathbf{Z} \approx \sqrt{2/\pi} \Sigma \Sigma^T = \sqrt{2/\pi} N \Sigma \Sigma^T \mathbf{X}$$  \hspace{1cm} (11)

$$\text{sgn}(\mathbf{Z})^T \mathbf{T} e \approx \sqrt{2/\pi} \Sigma \Sigma^T \mathbf{T} e = \sqrt{2/\pi} N \Sigma \Sigma^T \mathbf{X}$$  \hspace{1cm} (12)

where $\Sigma = \text{diag}(\sigma_1^{-1}, \ldots, \sigma_K^{-1})$. Substituting (11) and (12) into (7), we have

$$\hat{\mathbf{w}}_{p+1} = \hat{\mathbf{w}}_p + \mu [\mathbf{I}_K, j\mathbf{I}_K] [\text{sgn}(\mathbf{Z})^T \mathbf{Z}]^{-1} \text{sgn}(\mathbf{Z}) \mathbf{T} e \approx \hat{\mathbf{w}}_p + \mu [\mathbf{I}_K, j\mathbf{I}_K] [\sqrt{2/\pi} N \Sigma \Sigma^T \mathbf{X}]^{-1} \sqrt{2/\pi} N \Sigma \Sigma^T \mathbf{X} \mathbf{T} e \approx \hat{\mathbf{w}}_p + \mu [\mathbf{I}_K, j\mathbf{I}_K] (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{T} e$$  \hspace{1cm} (13)

which means that the iteration procedure of the proposed approach (7) is equivalent to (2), thus its linearization and convergence performance actually remains the same with the LS-based algorithm.

B. Complexity Comparison

The computational complexities of LS-DPD, LMS-DPD, and the proposed SRA-DPD, in terms of real multiplications and additions in each iteration, are listed in Table I. Benefited from the nature of a signum function, most multiplications in the proposed approach are eliminated compared with LS-DPD, although additional $K + 1$ WHTs are introduced. Fortunately, since WHT requires only additions/subtractions and can be efficiently implemented using a fast algorithm (FWHT), these WHTs lead to a slight increase in times of additions. Consequently, the proposed SRA-DPD requires far lesser number of multiplications than LS-DPD, and the total operations required in each iteration are also reduced. For instance, consider a DPD system with $K = 76$ and $N = 4096$, the required real multiplications and total operations per
The proposed SRA-DPD can effectively linearize the PA and have fast convergence capability that achieve the optimal performance until the end of test. Meanwhile, both LS-DPD and the proposed SRA-DPD can effectively linearize the PA and have fast convergence capability that achieve the optimal performance within six iterations. In summary, the proposed SRA-DPD has nearly the same linearization and convergence performance with LS-DPD, which supports the analysis in Section III-A.

### V. CONCLUSION

In this letter, a new DPD model identification approach based on SRA is presented. Taking advantage of the signed regression matrix, the proposed approach requires much less multiplications per iteration compared to the conventional LS-based algorithm. Walsh–Hadamard transform is employed to avoid duplicated DPD basis functions. Experimental results show that compared to LS-based algorithm, the proposed approach can significantly reduce the computational complexity of DPD model identification, without sacrificing the linearization performance and the fast convergence capability.

### REFERENCES


