

# Weighted Least Squares with Orthonormal Polynomials and Numerical Integration for Estimation of Memoryless Nonlinearity

Kazuki Komatsu, Yuichi Miyaji, and Hideyuki Uehara,

**Abstract**—The nonlinearity of amplifiers is one of the major impairments in wireless communications. In this letter, we propose a novel estimation method for the memoryless nonlinearity of amplifiers using weighted least squares and provide its theoretical error analysis on complex Gaussian signals. In the proposed method, the input signal and weight value are obtained via numerical integration formulas. Simulation results show that the proposed method can achieve a sufficiently low reconstruction error with 10 measurement samples on the estimation of the 13th-order nonlinearity. In addition, the simulation and theoretical results are consistent with each other.

**Index Terms**—Communication system nonlinearities, complex Gaussian process, least squares, nonlinear distortion, numerical integration

## I. INTRODUCTION

WIRELESS communication systems suffer from the nonlinearities of amplifiers or other radio-frequency (RF) circuits. Accordingly, the estimation and compensation of these nonlinearities are important research objectives. The memoryless nonlinearity of an amplifier needs to be estimated accurately to achieve better pre-distortion [1] or better self-interference cancellation [2]. In the simplest model of the nonlinearity estimation problem, the relation between the input signal  $x_n$  and the output signal  $y_n$  can be written as

$$y_n = f(x_n) + z_n, \quad (1)$$

where  $f(x)$  is the nonlinear transfer function of the target amplifier and  $z_n$  is additive white Gaussian noise that is independent of  $x_n$  and distributed on  $\mathcal{CN}(0, \sigma_z^2)$ . In the model of (1), we focus on the accurate estimation of the transfer function of the amplifier with a small number of observation samples under the assumption that the output of the amplifier can be observed directly. The simplest solution of this problem is achieved using polynomial approximation and least squares estimation. Accordingly, the transfer function of the amplifier  $f(x)$  is approximated to the following  $P$ -th order memoryless polynomial:

$$f(x) \approx a_1 x + a_3 x|x|^2 + \cdots + a_P x|x|^{P-1}, \quad (2)$$

This work was supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI Grant Numbers JP18K04138, JP19K14979, and JP19J12727.

In this work, we used the cluster computer system of the Information Media Center (IMC), Toyohashi University of Technology.

The authors are with the Department of Electrical and Electronic Information Engineering, Toyohashi University of Technology, Japan (e-mail: komatsu.kazuki.op@tut.jp; miyaji@ee.tut.ac.jp; uehara@tut.jp).

and the coefficients  $a_1, a_3, \dots, a_P$  are estimated using the following least squares method with  $N$  measurement samples:

$$\hat{\mathbf{a}} = [\hat{a}_1 \quad \hat{a}_3 \quad \cdots \quad \hat{a}_P]^T = (\mathbf{X}^H \mathbf{X})^{-1} \mathbf{X}^H \mathbf{y}, \quad (3)$$

where  $(\cdot)^T$  and  $(\cdot)^H$  denote the transpose and Hermitian transpose of a matrix, respectively,  $\hat{a}_p$  represents the estimated coefficients, and

$$\mathbf{X} = \begin{bmatrix} x_1 & x_1|x_1|^2 & \cdots & x_1|x_1|^{P-1} \\ x_2 & x_2|x_2|^2 & \cdots & x_2|x_2|^{P-1} \\ \vdots & \vdots & \ddots & \vdots \\ x_N & x_N|x_N|^2 & \cdots & x_N|x_N|^{P-1} \end{bmatrix}, \quad (4)$$

$$\mathbf{y} = [y_1 \quad y_2 \quad \cdots \quad y_N]^T. \quad (5)$$

However, this solution has severe numerical instability due to the large condition number of the Gram matrix  $\mathbf{X}^H \mathbf{X}$  on high-order nonlinearity estimation [1], [3], [4]. Existing literature [1], [3], [4] provides an improved method for mitigating the instability, using orthonormal polynomials instead of  $x|x|^{p-1}$ . Because of their advantages of orthonormality and orthogonality, orthogonal polynomials have been applied to not only estimation problems but also the latest studies on nonlinearities in a wide range of wireless communications such as the analysis of massive multiple-input multiple-output (MIMO) systems [5] and nonlinear equalizers [6].

In the improved version of the least squares estimation, the transfer function  $f(x)$  is approximated to the following  $P$ -th order expansion:

$$f(x) \approx b_1 \psi_1(x) + b_3 \psi_3(x) + \cdots + b_P \psi_P(x), \quad (6)$$

where  $\psi_p(x)$  is a  $p$ -th order orthonormal polynomial. For the expansion (6), among the various types of orthogonal polynomials, a polynomial that satisfies the following orthonormality is used, to achieve better stability:

$$\mathbb{E} [\psi_p(x) \psi_q^*(x)] = \int_{\mathcal{C}} \psi_p(x) \psi_q^*(x) p_x(x) dx = \delta_{pq}, \quad (7)$$

where  $\delta_{pq}$  is the Kronecker delta,  $p_x(x)$  is the probability density function of  $x$ , and  $\int_{\mathcal{C}} dx$  indicates integration on the complex plane. The orthonormal polynomial  $\psi_p(x)$  depends on the distribution of the communication signal because the expectation of (7) depends on it. Most current communication systems use orthogonal frequency-division multiplexing (OFDM) as the modulation scheme. The complex amplitude of the OFDM signal is distributed on the complex Gaussian distribution due to a high number of subcarriers and the central

limit theorem [3], [7]. When the complex Gaussian signal has a unit variance, i.e., unit power, the orthonormal polynomial  $\psi_p(x)$ , which satisfies the orthonormality of (7), can be written as

$$\psi_{2m+1}(x) = \frac{(-1)^m}{\sqrt{m+1}} L_m^1(|x|^2)x, \quad (8)$$

where  $L_m^1(z)$  is the following generalized Laguerre polynomial:

$$L_m^\alpha(z) = \sum_{n=0}^m \frac{(-1)^n}{n!} \binom{m+\alpha}{m-n} z^n. \quad (9)$$

Therefore, the estimated coefficient vector of the orthonormal expansion (6) obtained using the improved least squares can be expressed as

$$\hat{\mathbf{b}} = [\hat{b}_1 \quad \hat{b}_3 \quad \cdots \quad \hat{b}_P]^T = (\mathbf{\Psi}^H \mathbf{\Psi})^{-1} \mathbf{\Psi}^H \mathbf{y}, \quad (10)$$

where

$$\mathbf{\Psi} = \begin{bmatrix} \psi_1(x_1) & \psi_3(x_1) & \cdots & \psi_P(x_1) \\ \psi_1(x_2) & \psi_3(x_2) & \cdots & \psi_P(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_1(x_N) & \psi_3(x_N) & \cdots & \psi_P(x_N) \end{bmatrix}. \quad (11)$$

In (10), the  $(i, j)$  element of the Gram matrix  $\mathbf{\Psi}^H \mathbf{\Psi}$  can be written as

$$(\mathbf{\Psi}^H \mathbf{\Psi})_{i,j} = \sum_{n=1}^N \psi_{2i-1}^*(x_n) \psi_{2j-1}(x_n). \quad (12)$$

When the number of measurements  $N$  is sufficiently large, the equation

$$\lim_{N \rightarrow \infty} \frac{1}{N} (\mathbf{\Psi}^H \mathbf{\Psi})_{i,j} = \mathbb{E} [\psi_{2i-1}^*(x) \psi_{2j-1}(x)] = \delta_{ij} \quad (13)$$

holds due to the orthonormality of  $\psi_p(x)$  because  $\frac{1}{N} (\mathbf{\Psi}^H \mathbf{\Psi})_{i,j}$  is a sample average, and it converges to the expected value when  $N \rightarrow \infty$ . Thus, (13) indicates that, if a sufficiently large number of measurement samples is available, the condition number of the Gram matrix converges to 1.

However, the convergence speed of (13) is very low when the measurement signal  $x_n$  is randomly generated from a complex Gaussian distribution. The intuitive reason is that (12) is the Monte Carlo integration. It is known that the error of the Monte Carlo integration decreases as  $1/\sqrt{N}$ , and it is much slower than other numerical integration schemes. The same issue arises on  $\mathbf{\Psi}^H \mathbf{y}$  of (10). The  $i$ -th element of  $\mathbf{\Psi}^H \mathbf{y}$  is written as

$$(\mathbf{\Psi}^H \mathbf{y})_i = \sum_{n=1}^N \psi_{2i-1}^*(x_n) y_n. \quad (14)$$

When the number of measurements  $N$  is sufficiently large, the equation

$$\begin{aligned} \lim_{N \rightarrow \infty} \frac{1}{N} (\mathbf{\Psi}^H \mathbf{y})_i &= \mathbb{E} [\psi_{2i-1}^*(x) f(x)] \\ &= \frac{1}{\pi} \int_{\mathbb{C}} f(x) \psi_{2i-1}^*(x) e^{-|x|^2} dx = b_{2i-1} \end{aligned} \quad (15)$$

holds due to the orthonormality of  $\psi_p(x)$ . However, the convergence speed of (15) is very low due to the Monte Carlo integration.

To summarize, the conventional least squares method has the following problems:

- Large condition number: When the number of measurement signals is not sufficient, the condition number of the Gram matrix becomes a large value.
- Low convergence speed: When the number of measurement signals is not sufficient, the estimated value does not converge to a true value.

These problems are related to random sampling observation, and the authors of the paper [8] proposed a sample selection method based on the genetic algorithm to solve these problems on digital pre-distortion with non-orthogonal polynomials. In contrast, in this letter, we propose a weighted least squares method with orthonormal polynomials and numerical integration. In the proposed method, the measurement input signal  $x_n$  and the weights of the least squares are easily obtained via numerical integration formulas.

The details of the proposed method are described in Section II. In Section III, the proposed scheme and the conventional least squares method are compared via numerical simulations. Section IV presents the conclusion of the letter.

## II. PROPOSED METHOD

The proposed method uses the weighted least squares method. The measurement samples and weights are obtained using a numerical integration formula. In the proposed method, we approximate the transfer function to the orthonormal polynomial expansion of (6). Then, the vector of the estimated coefficients is expressed as

$$\hat{\mathbf{b}} = (\mathbf{\Psi}^H \mathbf{W} \mathbf{\Psi})^{-1} \mathbf{\Psi}^H \mathbf{W} \mathbf{y}, \quad (16)$$

where  $\mathbf{\Psi}$  and  $\mathbf{y}$  are the same as (11) and (5), respectively, and  $\mathbf{W} = \text{diag}\{w_1, w_2, \dots, w_N\}$  is a diagonal weight matrix. The main difference between the conventional and proposed methods is that the measurement samples  $x_n$  and the weight  $w_n$  are obtained using a numerical integration formula that can calculate the following two integrals:

$$\mathbb{E} [\psi_{2i-1}^*(x) \psi_{2j-1}(x)] = \frac{1}{\pi} \int_{\mathbb{C}} \psi_{2i-1}^*(x) \psi_{2j-1}(x) e^{-|x|^2} dx, \quad (17)$$

$$\mathbb{E} [f(x) \psi_{2i-1}^*(x)] = \frac{1}{\pi} \int_{\mathbb{C}} f(x) \psi_{2i-1}^*(x) e^{-|x|^2} dx, \quad (18)$$

with a high accuracy even if the number of measurements  $N$  is very small. Generally, a numerical integration formula with  $N$  samples can be written as

$$\mathbb{E} [g(x)] \approx \sum_{n=1}^N g(x_n) w_n, \quad (19)$$

where  $g(x)$  is an arbitrary function, and  $x_n$  and  $w_n$  are the computing points and weights of the numerical integration, respectively. In the proposed method, we use the computing points as measurement samples, and the  $n$ -th element of the

diagonal weight matrix  $\mathbf{W}$  is  $w_n$ . Then, we can expect that the following two equations:

$$(\Psi^H \mathbf{W} \Psi)_{i,j} = \sum_{n=1}^N \psi_{2i-1}^*(x_n) \psi_{2j-1}(x_n) w_n, \quad (20)$$

$$(\Psi^H \mathbf{W} \mathbf{y})_i = \sum_{n=1}^N y_n \psi_{2i-1}^*(x_n) w_n, \quad (21)$$

rapidly converge to the expected values of (17) and (18), respectively, if the noise is ignored. Therefore, the Gram matrix  $\Psi^H \mathbf{W} \Psi$  becomes the identity matrix, and the estimated vector  $\hat{\mathbf{b}}$  stably converges to the true coefficient vector  $\mathbf{b}$ , even if the number of measurement samples is very small. In addition, if the Gram matrix is approximately the identity matrix, the estimate can be given as  $\hat{\mathbf{b}} \approx \Psi^H \mathbf{W} \mathbf{y}$ .

#### A. Example 1: Gauss–Laguerre quadrature

The integrations of (17) and (18) can be rewritten as

$$\begin{aligned} \mathbb{E}[g(x)] &= \frac{1}{\pi} \int_0^{2\pi} \int_0^\infty g(re^{i\theta}) e^{-r^2} r dr d\theta \\ &= \int_0^\infty g(r) \cdot 2re^{-r^2} dr = \int_0^\infty g(\sqrt{t}) e^{-t} dt, \end{aligned} \quad (22)$$

where  $g(x) = \psi_{2i-1}(x) \psi_{2j-1}^*(x)$  for (17), and  $g(x) = f(x) \psi_{2i-1}^*(x)$  for (18). The reason for the above transformation is that  $g(x) = g(|x|)$  holds in both cases. The last term of (22) is a semi-infinite integral with an exponentially decaying weight function, and the Gauss–Laguerre quadrature is a good choice for integrating it with high accuracy. In the Gauss–Laguerre quadrature, the computing point  $t_n$  is the  $n$ -th root of the Laguerre polynomial  $L_N(x) = L_N^0(x)$ , and the weights are given by [9, Eq. 25.4.45]

$$w'_n = \frac{x_n}{(N+1)^2 [L_{N+1}(x_n)]^2}. \quad (23)$$

Then, the measurement samples and weights of the proposed method are  $x_n = \sqrt{t_n}$  and  $w_n = w'_n$ , respectively. Moreover, the error of the Gauss–Laguerre quadrature is given by [9, Eq. 25.4.45]

$$R_N[g] = \frac{(N!)^2}{(2N)!} \left. \frac{d^{2N}}{dt^{2N}} g(\sqrt{t}) \right|_{t=\xi}. \quad (0 < \xi < \infty) \quad (24)$$

Thus, the convergence speed of the Gauss–Laguerre quadrature is much higher than that of the Monte Carlo integration because  $\frac{(N!)^2}{(2N)!} \ll 2^{-N}$ .

In Section III, when the number of measurements is larger than 100, the 100 measurement samples and weights are repeated  $N/100$  times to generate  $N$  measurement samples and weights because the Gauss–Laguerre quadrature has very high accuracy, even if  $N = 20$ . Thus, we use  $x_n = \sqrt{t_{(n\%100)}}$  and  $w_n = \frac{100}{N} w'_{(n\%100)}$  when  $N > 100$ , where the binary operator  $\%$  indicates the remainder after division, and  $t_n$  and  $w_n$  are obtained from 100-points Gauss–Laguerre quadrature.

#### B. Example 2: Rectangular rule

Generally, the rectangular rule is not a highly accurate integration method, but it is practical for the proposed method. The middle term of (22) is an integration with a rapidly decreasing weight  $e^{-r^2}$ , and we can obtain sufficient accuracy even if the integration interval is only  $[0, 5]$  instead of  $[0, \infty)$ . Then, the measurement samples and weights of the proposed method with the rectangular rule can be written as

$$x_n = \frac{5}{N} n, \quad w_n = \frac{10}{N} x_n e^{-x_n^2}. \quad (25)$$

In (22), the integrand can be approximated to zero at both ends of the integration domain, i.e.,  $g(r) \cdot 2re^{-r^2} \approx 0$  at  $r = 0$  and  $r = 5$ . Then, the rectangular rule for (22) is almost equal to the trapezoidal rule, and the error of the integration can be expressed as [9, Eq. 25.4.2]

$$R_N[g] \approx \frac{125}{12N^2} \left. \frac{d^2}{dr^2} [g(r) \cdot 2re^{-r^2}] \right|_{r=\xi}. \quad (0 < \xi < 5) \quad (26)$$

Thus, the convergence speed of the rectangular rule is higher than that of the Monte Carlo integration because  $R_N[g] \sim \mathcal{O}(N^{-2})$ .

It can be observed from (25) that the measurement samples can be viewed as a ramp signal. This is an interesting aspect of the rectangular rule for the proposed method.

#### C. Theoretical error analysis

In this subsection, we analyze the following total reconstruction error:

$$E_{\text{tot}}^2 = \mathbb{E} \left[ \left| f(x) - \hat{f}(x) \right|^2 \right], \quad (27)$$

where  $\hat{f}(x)$  is the reconstructed nonlinearity defined as

$$\hat{f}(x) = \hat{b}_1 \psi_1(x) + \hat{b}_3 \psi_3(x) + \cdots + \hat{b}_P \psi_P(x). \quad (28)$$

Furthermore, the nonlinear function  $f(x)$  can be expanded to an infinite series as

$$f(x) = b_1 \psi_1(x) + b_3 \psi_3(x) + \cdots = \sum_{p=1,3,\dots}^{\infty} b_p \psi_p(x). \quad (29)$$

Thus, the total error can be rewritten as

$$\begin{aligned} E_{\text{tot}}^2 &= \mathbb{E} \left[ \left| \sum_{p=1,3,\dots}^P (b_p - \hat{b}_p) \psi_p(x) + \sum_{p=P+2,P+4,\dots}^{\infty} b_p \psi_p(x) \right|^2 \right] \\ &\stackrel{(a)}{=} \underbrace{\sum_{p=1,3,\dots}^P \mathbb{E} \left[ |b_p - \hat{b}_p|^2 \right]}_{\text{Estimation error: } E_{\text{est}}^2} + \underbrace{\sum_{p=P+2,P+4,\dots}^{\infty} |b_p|^2}_{\text{Approximation error: } E_{\text{app}}^2}. \end{aligned} \quad (30)$$

The transform of  $\stackrel{(a)}{=}$  is due to the orthonormality of (7). The proposed method has two errors: the estimation error  $E_{\text{est}}^2$  and the approximation error  $E_{\text{app}}^2$ . The approximation error is the error caused by approximating the series expansion of  $f(x)$  in finite dimensions. The conventional method also has this error

due to the approximation of (6). The approximation error can be written as [10, eq. (3.2.7), section 3.2, p. 217]

$$E_{\text{app}}^2 = \mathbb{E} [|f(x)|^2] - \sum_{p=1,3,\dots}^P |\mathbb{E} [f(x)\psi_p^*(x)]|^2. \quad (31)$$

In (31), the approximation error monotonically decreases as the order  $P$  increases. This error has the same value in both the proposed and conventional methods if the order  $P$  is the same.

In contrast, the estimation error  $E_{\text{est}}^2$  leads to a performance variation between the proposed and conventional methods. When the Gram matrix converges to the identity matrix sufficiently, the estimated coefficient  $\hat{b}_p$  can be written as

$$\begin{aligned} \hat{b}_{2i-1} &= (\Psi^H \mathbf{W} \mathbf{y})_i \\ &= \sum_{n=1}^N f(x_n) \psi_{2i-1}^*(x_n) w_n + \sum_{n=1}^N z_n \psi_{2i-1}^*(x_n) w_n. \end{aligned} \quad (32)$$

Thus, the estimation error  $E_{\text{est}}^2$  can be rewritten as

$$\begin{aligned} E_{\text{est}}^2 &= \sum_{p=1,3,\dots}^P \left| \sum_{n=1}^N f(x_n) \psi_p^*(x_n) w_n - b_p \right|^2 \\ &+ \sum_{p=1,3,\dots}^P \mathbb{E} \left[ \left| \sum_{n=1}^N z_n \psi_p^*(x_n) w_n \right|^2 \right]. \end{aligned} \quad (33)$$

In the right-hand side of (33), the first term indicates the square of the quadrature error, and the second term indicates the error caused by the noise. The square of the quadrature error is  $(R_N[f(r)\psi_p^*(r)])^2$ , and the error rapidly decays at a rate of  $\mathcal{O}(N^{-4})$ , even for the rectangular rule. In addition, the noise error can be rewritten as

$$\sum_{p=1,3,\dots}^P \mathbb{E} \left[ \left| \sum_{n=1}^N z_n \psi_{2i-1}^*(x_n) w_n \right|^2 \right] = \sigma_z^2 \sum_{p=1,3,\dots}^P \sum_{n=1}^N |\psi_p(x_n) w_n|^2. \quad (34)$$

Equation (34) shows that the effect of noise depends on the values of the measurement samples and weights determined by the numerical integration method employed, and the sum of their squares is an indicator of the influence of noise. When the rectangular rule is used for the proposed method, the summation of the right-hand side of (34) with a large  $N$  can be asymptotically expressed as

$$\sum_{n=1}^N |\psi_p(x_n) w_n|^2 \approx \frac{5}{N} \int_0^\infty \left\{ \psi_p(x_n) \cdot 2re^{-r^2} \right\}^2 dr \sim \mathcal{O}(N^{-1}). \quad (35)$$

The noise error decays at a rate of  $\mathcal{O}(N^{-1})$ , and the quadrature error is negligibly small compared with the noise error.

To summarize this section, we can estimate the error of the proposed method as

$$E_{\text{tot}}^2 \approx E_{\text{app}}^2 + \sigma_z^2 \sum_{p=1,3,\dots}^P \sum_{n=1}^N |\psi_p(x_n) w_n|^2. \quad (36)$$

The first term is a constant for the number of measurements  $N$ , and the second term decays at a rate of  $\mathcal{O}(N^{-1})$ . Thus, the

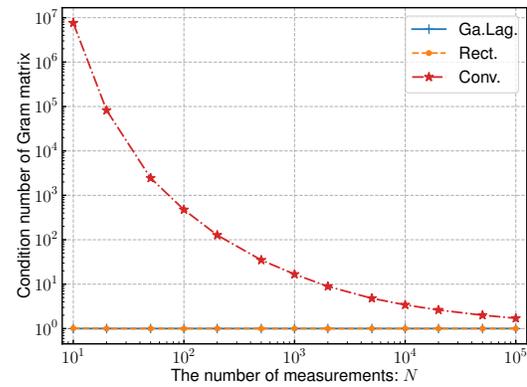


Fig. 1. Condition number of the Gram matrix of each method when  $P = 7$ . The value is averaged over  $10^4$  times independent Monte Carlo simulation.

convergence rate of the proposed method is the same as that of the conventional Monte-Carlo-based least squares method whose rate of square error is  $\mathcal{O}(N^{-1})$ . This is because the conventional method is exactly same as the proposed method with randomly generated samples  $x_n$  and weights  $w_n = 1/N$ . However, the rate is a characteristic of  $N \rightarrow \infty$ , and we compare the characteristics of each method from a small to a large number  $N$  in numerical experiments in the following section.

### III. RESULTS OF NUMERICAL EXPERIMENTS

In this section, we evaluate and compare the condition number and total reconstruction error, which is defined as (27), for the proposed and conventional methods using  $10^4$  times Monte Carlo simulation. In the simulation, we use the Rapp model [11] as an amplifier, and its transfer function can be written as

$$f(x) = \frac{x}{\left(1 + (|x|/B)^{2s}\right)^{\frac{1}{2s}}}, \quad (37)$$

where  $B$  indicates the input back-off (IBO), and  $s$  is the smoothness factor. Furthermore, we use  $B = \sqrt{10}$  (i.e., 10 dB IBO) and  $s = 3$ .

Figure 1 shows the condition number of the Gram matrix of each method. As mentioned in the Introduction, the condition number of the conventional method is much larger than that of the proposed method because it is based on the Monte Carlo integration. The proposed method successfully reduces the condition number because of its high accuracy of numerical integration. This indicates that the proposed method can achieve better stability than the conventional least squares method.

Figures 2, 3, 4, and 5 show the simulation results and theoretical results of the total reconstruction error for each method, with  $P = 7$  and  $P = 13$ , from a very noisy case to an almost noise-free case for a wide range of applications such as pre-distortion, post-distortion, and self-interference cancellers. The error of the proposed method is much smaller than that of the conventional method, as the proposed method has good stability and better accuracy of integration. Surprisingly, even when only 10 measurements are used, the proposed method maintained the reconstruction error below the noise.

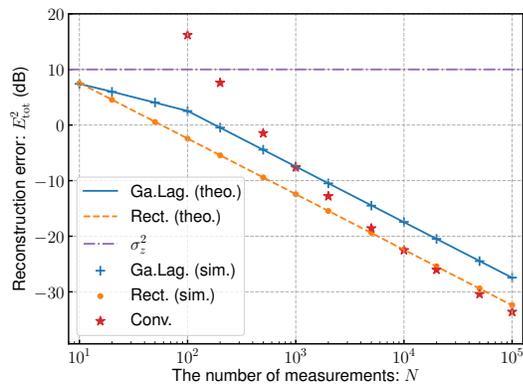


Fig. 2. Total reconstruction error  $E_{\text{tot}}^2$  of each method with  $P = 7$  and  $\sigma_z^2 = 10^{-1}$ . The lines indicate theoretical results, and the markers indicate simulation results.

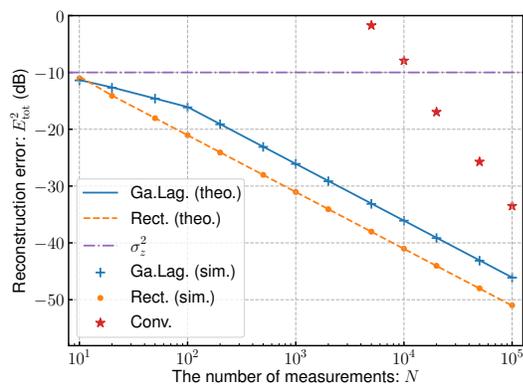


Fig. 3. Total reconstruction error  $E_{\text{tot}}^2$  of each method with  $P = 13$  and  $\sigma_z^2 = 10^{-1}$ . The lines indicate theoretical results, and the markers indicate simulation results.

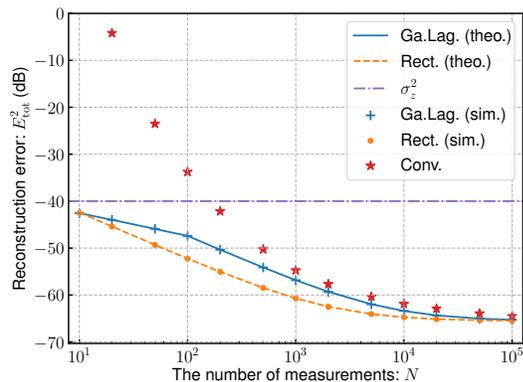


Fig. 4. Total reconstruction error  $E_{\text{tot}}^2$  of each method with  $P = 7$  and  $\sigma_z^2 = 10^{-4}$ . The lines indicate theoretical results, and the markers indicate simulation results.

In contrast, in Fig. 2, the error of the conventional method is smaller than that of the proposed method when  $N > 10^4$ . Therefore, the conventional method is more effective than the proposed method when a sufficient number of samples are used under low signal-to-noise ratio (SNR). In addition, the theoretical and simulation results are consistent with each other in these figures. Thus, the analysis in this paper is useful for the error estimation of the proposed method. Moreover, if the error is to be reduced further, a numerical integration method that reduces the value of (34) needs to be used.

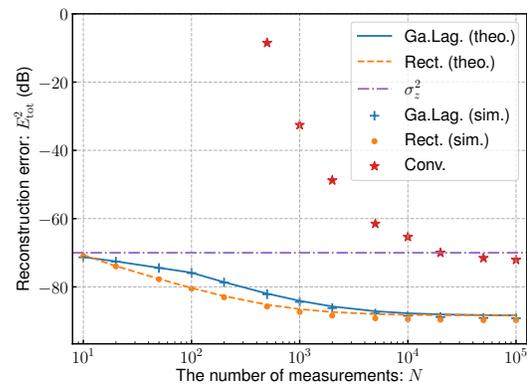


Fig. 5. Total reconstruction error  $E_{\text{tot}}^2$  of each method with  $P = 13$  and  $\sigma_z^2 = 10^{-7}$ . The lines indicate theoretical results, and the markers indicate simulation results.

#### IV. CONCLUSION

In this letter, we proposed a novel estimation method for the memoryless nonlinearity of an amplifier. The method uses weighted least squares with orthonormal polynomials and numerical integration. The measurement signal and weights of the proposed method were designed based on the numerical integration method to converge the Gram matrix to the unit matrix with high accuracy, even with a small number of observations. Moreover, we derived the theoretical error of the proposed method. The simulation results showed that the proposed method dramatically improved the accuracy of the conventional least squares method and achieved sufficient accuracy with 10 measurement samples. The theoretical results and simulation results were consistent with each other.

#### REFERENCES

- [1] H. Qian, S. Yao, H. Huang, and W. Feng, "A low-complexity digital predistortion algorithm for power amplifier linearization," *IEEE Trans. Broadcast.*, vol. 60, no. 4, pp. 670–678, Dec. 2014.
- [2] K. Komatsu, Y. Miyaji, and H. Uehara, "Basis function selection of frequency-domain Hammerstein self-interference canceller for in-band full-duplex wireless communications," *IEEE Trans. Wireless Commun.*, vol. 17, no. 6, pp. 3768–3780, Jun. 2018.
- [3] R. Raich and G. Zhou, "Orthogonal polynomials for complex Gaussian processes," *IEEE Trans. Signal Process.*, vol. 52, no. 10, pp. 2788–2797, Oct. 2004.
- [4] R. Dallinger, H. Ruotsalainen, R. Wichman, and M. Rupp, "Adaptive pre-distortion techniques based on orthogonal polynomials," in *Proc. 44th Asilomar Conf. Signals, Syst., Comput.*, Nov. 2010, pp. 1945–1950.
- [5] S. Teodoro, A. Silva, R. Dinis, F. M. Barradas, P. M. Cabral, and A. Gameiro, "Theoretical analysis of nonlinear amplification effects in massive MIMO systems," *IEEE Access*, vol. 7, pp. 172 277–172 289, 2019.
- [6] L. Shen, B. Henson, Y. Zakharov, and P. D. Mitchell, "Adaptive nonlinear equalizer for full-duplex underwater acoustic systems," *IEEE Access*, 2020, (Early Access, DOI:10.1109/ACCESS.2020.3000590).
- [7] P. Banelli and S. Cacopardi, "Theoretical analysis and performance of OFDM signals in nonlinear AWGN channels," *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 430–441, Mar. 2000.
- [8] J. Kral, T. Gotthans, R. Marsalek, M. Harvanek, and M. Rupp, "On feedback sample selection methods allowing lightweight digital predistorter adaptation," *IEEE Trans. Circuits Syst. I*, vol. 67, no. 6, pp. 1976–1988, 2020.
- [9] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. New York City: Dover, 1964.
- [10] W. Gautschi, *Orthogonal Polynomials: Computation and Approximation*, ser. Numerical mathematics and scientific computation. Oxford University Press, 2004.
- [11] C. Rapp, "Effects of HPA-nonlinearity on a 4-DPSK/OFDM-signal for a digital sound broadcasting system," in *Proc. the Second European Conf. on Satellite Commun.*, Oct. 1991, pp. 179–184.