Step-Size Control for Width Adaptation in Radial Basis Function Networks for Nonlinear Channel Equalization

Namyong Kim

Abstract: **A method of width adaptation in the radial basis function network (RBFN) using stochastic gradient (SG) algorithm is introduced. Using Taylor's expansion of error signal and differentiating the error with respect to the step-size, the optimal time-varying step-size of the width in RBFN is derived. The proposed approach to adjusting widths in RBFN achieves superior learning speed and the steady-state mean square error (MSE) performance in nonlinear channel environment. The proposed method has shown enhanced steady-state MSE performance by more than 3 dB in both nonlinear channel environments. The results confirm that controlling over step-size of the width in RBFN by the proposed algorithm can be an effective approach to enhancement of convergence speed and the steady-state value of MSE.**

Index Terms: **Equalization, nonlinear channel, radial basis function network (RBFN), step-size, stochastic gradient (SG), width.**

I. INTRODUCTION

The performance of radial basis function network (RBFN) is highly dependent on the choice of parameters of the RBFN. In [1], a simple learning algorithm that simultaneously adapts all the network parameters- centers, widths, and weights was proposed. The algorithm applies stochastic-gradient (SG) method to the RBF parameter adaptation. This RBFN-stochastic gradient (SG) algorithm has proved superior to many of the existing algorithms, with less computational requirements in nonlinear channel equalization applications [2]. Recently, the RBFN-SG has also been applied to odor identification problems for future multimedia systems equipped with odor sensors [3]. Another hybrid training algorithm [4] has been introduced that adapts the parameters using gradient descent procedure while employing singular value decomposition to compute the optimum linear weights at each iteration. The authors in the paper pointed that due to high sensitivity of the widths, the widths appear a major source of ill-conditioning in RBF networks. The width of a node controls the shape of the basis function or the response of the associated node of the RBFN. In learning process, the width measures the extent to which neighbor centers near the node concerned participate in the learning process [5]. Large or small values make the node response too flat or too peaked, respectively, and, therefore, both of these two extreme conditions should be avoided. In differential equation applications, N. May-Duy and T. Tran-Cong [6] adjusted the width as the only adjustable parameter according to the rule that a larger width is assigned where the centers are widely separated from each other and a smaller width where the centers are closer. In contrast to the approach taken by other authors as reviewed above, in the present method the width is updated by gradient descent method employing time-varying step-size at each iteration. The learning speed of the algorithms using stochastic-gradient descent method is dependent on the step-size. Our basic idea is that if we force the slope of the squared error with respect to the step-size to get close to zero, the optimum step-size for the widths of the RBFN-SG algorithm can be obtained and we can acquire more accurate width values and faster convergence speed. In this paper, we introduce a method of width adaptation in the RBFN-SG algorithm by applying the Taylor's expansion approach to error signal and differentiating it with respect to the step-size in order to obtain optimal time-varying step-size for the widths. Two common nonlinear channel models have been simulated to show that the proposed method performs better than the RBFN-SG algorithm.

Manuscript received March 12, 2009; approved for publication by Hyuncheol Park, Division I Editor, June 11, 2010.

N. Kim is with the Department of Information & Communication Engineering, Kangwon National University, email: namyong@kangwon.ac.kr.

This paper is organized as follows. In Section II, we briefly describe RBFN-SG algorithm. The optimal time-varying step-size of the width in RBFN-SG method is proposed in Section III. Section IV reports simulation results and discussions. Finally, concluding remarks are presented in Section V.

II. RBFN-SG ALGORITHM FOR NETWORK PARAMETER ADAPTATION

The RBFN-SG algorithm adapts all the free parameters of the network using gradient descent of the instantaneous output error power. Input vector having L elements is defined as

$$x^{(n)} = [x(n)\, x(n-1) \cdots x(n-L+1)]^T. \tag{1}$$

Let the error be denoted by $e^{(n)} = d^{(n)} - y^{(n)}$, where $d^{(n)}$ is desired output and $y^{(n)}$ is the RBFN output, all at the training time n. For a network parameter ϕ, the RBFN-SG algorithm adapts its value $\phi^{(n)}$ at time n according to

$$\phi^{(n+1)} = \phi^{(n)} - \mu_\theta \frac{\partial {e^{(n)}}^2}{\partial \phi^{(n)}} \tag{2}$$

where μ_θ is the convergence coefficient or step-size. Among localized basis functions, the Gaussian is the most popular choice for RBFN-SG. The output of RBFN-SG with M Gaussian basis functions is

$$y^{(n)} = \sum_{j=1}^{j=M} w_j^{(n)} \exp\left(-\frac{\|x^{(n)} - c_j^{(n)}\|^2}{\sigma_j^{(n)^2}}\right). \tag{3}$$

In the RBFN-SG algorithm, the center, width, and weight of hidden unit j at time n, are adapted according to the following equations [2]

$$c_j^{(n+1)} = c_j^{(n)} + \mu_c e^{(n)} w_j^{(n)} \exp\left(\frac{-\|x^{(n)} - c_j^{(n)}\|^2}{\sigma_j^{(n)2}}\right) \frac{(x^{(n)} - c_j^{(n)})}{\sigma_j^{(n)2}} \tag{4}$$

$$\sigma_j^{(n+1)} = \sigma_j^{(n)} + \mu_s e^{(n)} w_j^{(n)} \exp\left(\frac{-\left\|x^{(n)} - c_j^{(n)}\right\|^2}{\sigma_j^{(n)2}}\right) \frac{\left\|x^{(n)} - c_j^{(n)}\right\|^2}{\sigma_j^{(n)3}} \tag{5}$$

$$w_j^{(n+1)} = w_j^{(n)} + \mu_w e^{(n)} \exp\left(\frac{-\left\|x^{(n)} - c_j^{(n)}\right\|^2}{\sigma_j^{(n)2}}\right) \tag{6}$$

where μ_s, μ_c, and μ_w are step-sizes that control the speed of adaptation. Using firstly Taylor's expansion of error signal and then differentiating it with respect to the widths, μ_s, a method of obtaining optimum step-size value for the widths can be derived as described in the following section.

III. STEP-SIZE CONTROL FOR WIDTH ADAPTATION

Defining the activation level of hidden unit j at time n as

$$O_j^{(n)} = \exp\left(\frac{-\left\|x^{(n)} - c_j^{(n)}\right\|^2}{\sigma_j^{(n)2}}\right), \tag{7}$$

the width adaptation (5) becomes

$$\sigma_j^{(n+1)} = \sigma_j^{(n)} + \mu_s e^{(n)} w_j^{(n)} O_j^{(n)} \frac{\left\|x^{(n)} - c_j^{(n)}\right\|^2}{\sigma_j^{(n)3}}. \tag{8}$$

The difference between $\sigma_j^{(n+1)}$ and $\sigma_j^{(n)}$, $\Delta\sigma_j^{(n)}$ is

$$\Delta\sigma_j^{(n)} = \mu_s e^{(n)} w_j^{(n)} O_j^{(n)} \frac{\left\|x^{(n)} - c_j^{(n)}\right\|^2}{\sigma_j^{(n)3}}. \tag{9}$$

Using the difference (9) between the two consecutive time indexes, we can write $e^{(n+1)}$ as a Taylor's expansion of $e^{(n)}$,

$$e^{(n+1)} = e^{(n)} + \sum_{j=1}^{M} \frac{\partial e^{(n)}}{\partial \sigma_j^{(n)}} \Delta\sigma_j^{(n)} + \sum_{i=1}^{M}\sum_{j=1}^{M} \frac{\partial^2 e^{(n)}}{\partial \sigma_i^{(n)} \partial \sigma_j^{(n)}} \Delta\sigma_i^{(n)} \Delta\sigma_j^{(n)} + \cdots \tag{10}$$

where

$$\frac{\partial e^{(n)}}{\partial \sigma_j^{(n)}} = -\frac{\partial y^{(n)}}{\partial \sigma_j^{(n)}} = -\frac{\partial \sum_{j=1}^{j=M} w_j^{(n)} O_j^{(n)}}{\partial \sigma_j^{(n)}} = -2 w_j^{(n)} O_j^{(n)} \frac{\left\|x^{(n)} - c_j^{(n)}\right\|^2}{\sigma_j^{(n)3}} \tag{11}$$

and

$$\frac{\partial^2 e^{(n)}}{\partial \sigma_i^{(n)} \partial \sigma_j^{(n)}} = 0, \quad \text{for } i \neq j. \tag{12}$$

The second term in (10) can be rewritten as $\sum_{j=1}^{M} \partial^2 e^{(n)} / \partial \sigma_j^{(n)2} \cdot [\Delta\sigma_j^{(n)}]^2$ and it can be negligible in the steady state.

Substituting (9), (11), and (12) into (10) yields

$$e^{(n+1)} = e^{(n)} - \sum_{j=1}^{M} 2\mu_s e^{(n)} w_j^{(n)2} O_j^{(n)2} \frac{\left\|x^{(n)} - c_j^{(n)}\right\|^4}{\sigma_j^{(n)6}}. \tag{13}$$

Also,

$$e^{(n+1)^2} = e^{(n)^2}\left[1 - \sum_{j=1}^{M} 2\mu_s w_j^{(n)2} O_j^{(n)2} \frac{\left\|x^{(n)} - c_j^{(n)}\right\|^4}{\sigma_j^{(n)6}}\right]^2. \tag{14}$$

By differentiating (14) with respect to μ_s, we can obtain

$$\frac{\partial e^{(n+1)^2}}{\partial \mu_s} = 2e^{(n)^2}\left[1 - \sum_{j=1}^{M} 2\mu_s w_j^{(n)2} O_j^{(n)2} \frac{\left\|x^{(n)} - c_j^{(n)}\right\|^4}{\sigma_j^{(n)6}}\right] \cdot \left[-\sum_{j=1}^{M} 2 w_j^{(n)2} O_j^{(n)2} \frac{\left\|x^{(n)} - c_j^{(n)}\right\|^4}{\sigma_j^{(n)6}}\right]. \tag{15}$$

Letting (15) be equal to zero yields the optimum time-varying convergence coefficient, ${}^*\mu_s^{(n)}$.

$${}^*\mu_s^{(n)} = \frac{1}{\sum_{j=1}^{M} 2 w_j^{(n)2} O_j^{(n)2} \frac{\left\|x^{(n)} - c_j^{(n)}\right\|^4}{\sigma_j^{(n)6}}}. \tag{16}$$

When the denominator of (16) is too small, numerical difficulties may arise because then the step-size for the widths becomes big enough to diverge. To overcome this kind of problem, (16) is slightly modified as follows

$${}^*\mu_s^{(n)} = \frac{1}{a + \sum_{j=1}^{M} 2 w_j^{(n)2} O_j^{(n)2} \frac{\left\|x^{(n)} - c_j^{(n)}\right\|^4}{\sigma_j^{(n)6}}} \tag{17}$$

where constant $a > 0$. This modification using a small positive constant to avoid numerical problems for small node signals can also be referred in [7].

Consequently, the width which is the major parameter that affects the performance, is updated according to the following rule for the hidden unit j at time n.

$$\sigma_j^{(n+1)} = \sigma_j^{(n)} + \frac{e^{(n)} w_j^{(n)}}{a + \sum_{j=1}^{M} 2 w_j^{(n)2} O_j^{(n)2} \frac{\left\|x^{(n)} - c_j^{(n)}\right\|^4}{\sigma_j^{(n)6}}} \cdot \exp\left[\frac{-\left\|x^{(n)} - c_j^{(n)}\right\|^2}{\sigma_j^{(n)2}}\right] \frac{\left\|x^{(n)} - c_j^{(n)}\right\|^2}{\sigma_j^{(n)3}}. \tag{18}$$

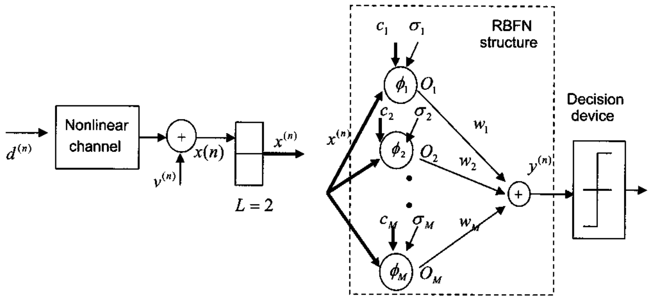

Fig. 1. Schematic of RBFN equalizer for $L = 2$.

The time-varying convergence coefficient $^*\mu_s^{(n)}$ governs stability and rate of convergence. It is necessary to evaluate whether (17) is in the range of stability by extracting the stability condition of (13).

From (13),

$$e^{(n+1)}=e^{(n)}\left[1-2\mu_s\sum_{j=1}^{M}w_j^{(n)^2}O_j^{(n)^2}\frac{\left\|x^{(n)}-c_j^{(n)}\right\|^4}{\sigma_j^{(n)6}}\right]. \quad (19)$$

Starting with the initial guess $e^{(0)}$, we obtain the nth error $e^{(n)}$

$$e^{(n)}=e^{(0)}\left[1-2\mu_s\sum_{j=1}^{M}w_j^{(n)^2}O_j^{(n)^2}\frac{\left\|x^{(n)}-c_j^{(n)}\right\|^4}{\sigma_j^{(n)6}}\right]^n. \quad (20)$$

The result in (20) shows that the proposed algorithm is stable and convergent when

$$\lim_{n\to\infty}\left[1-2\mu_s\sum_{j=1}^{M}w_j^{(n)^2}O_j^{(n)^2}\frac{\left\|x^{(n)}-c_j^{(n)}\right\|^4}{\sigma_j^{(n)6}}\right]^n=0. \quad (21)$$

This indicates

$$\left|1-2\mu_s\sum_{j=1}^{M}w_j^{(n)^2}O_j^{(n)^2}\frac{\left\|x^{(n)}-c_j^{(n)}\right\|^4}{\sigma_j^{(n)6}}\right|<1. \quad (22)$$

This condition can be also expressed as

$$0<\mu_s<\frac{1}{\sum_{j=1}^{M}w_j^{(n)^2}O_j^{(n)^2}\frac{\left\|x^{(n)}-c_j^{(n)}\right\|^4}{\sigma_j^{(n)6}}}. \quad (23)$$

By comparing the proposed convergence coefficient, $^*\mu_s^{(n)}$ in (17) with (23), it is apparent that $^*\mu_s^{(n)}$ satisfies the stability condition.

IV. RESULTS AND DISCUSSION

In this section, we present simulation results and discussion for nonlinear channel equalization using the linear tapped delay line (TDL) equalizer structure and nonlinear RBFN for performance comparison. The RBFN equalizer we considered is depicted in Fig. 1. The performance is measured by the mean-squared-error (MSE) between the equalizer output and the correct symbols. The transmitted training symbol is a random sequence of bipolar signals $(+1, -1)$. The initial centers are formed from the first few successive channel output samples of the training set. The RBFN weights are initialized to zeros. The additive white Gaussian noise $v^{(n)}$ has zero mean, variance 0.001. The equalizer input dimension is set to $L = 2$ and the initial common values of the spread parameter, width, were $\sigma_j^{(0)} = 2$ for all j. We set this initial width value based on a consideration of the distance between the two transmitted symbol values. As in [2], we also have found that the performance is rather robust to variation over a significant range of values for the initial width. The RBFN equalizer has 30 hidden nodes ($M = 30$). The values of step-size for RBFN-SG are all 0.05 with no momentum. Constant a for the proposed method is set to 0.7. The TDL equalizer structure with the least mean square (LMS) algorithm, referred to as TDL-LMS in this paper, has 30 taps and its step-size is also the same 0.05. The TDL structure can be replaced in stead of the RBFN structure in Fig. 1 for linear equalizer performance evaluation. For learning performance comparison, nonlinear channel environment described in [8] is used. The nonlinear channel 1 for the first simulation is given by

$$x(n)=h^{(n)}-0.9h^{(n)^3}+v^{(n)},\qquad h^{(n)}=d^{(n)}+0.5d^{(n-1)}. \quad (24)$$

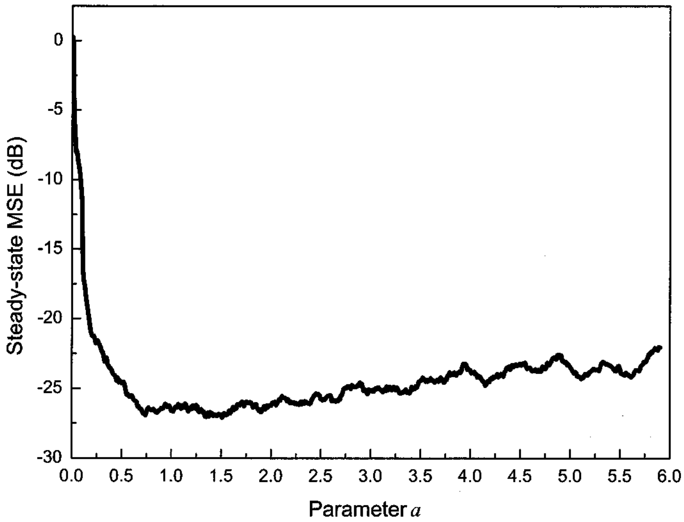

Fig. 2. Minimum MSE performance with respect to the constant a.

To investigate the effect of the constant a on the equalizer performance, minimum MSE is depicted for different values ranging from 0 to 6 under channel 1 environment in Fig. 2. The curve of minimum MSE with respect to the constant a is concave and the optimum value a can be obtained at the saddle point that makes the algorithm converge and produces the lowest steady-state MSE. The optimum point in Fig. 2 justifies our choice of 0.7 for the constant a. As the constant a increases, we see the proposed algorithm becomes the conventional RBFN-SG algorithm which uses a constant step-size.

The convergence of the various equalizers for nonlinear channel 1 is shown in Fig. 3. It can be seen that the linear equal-

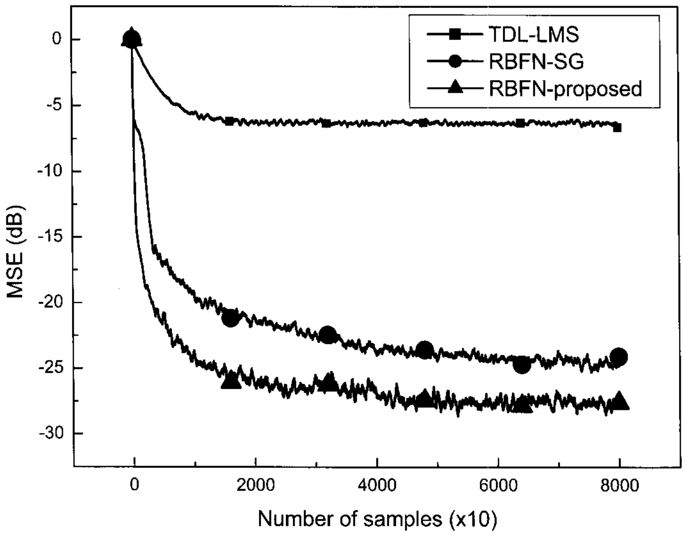

Fig. 3. MSE learning performance for nonlinear channel 1.

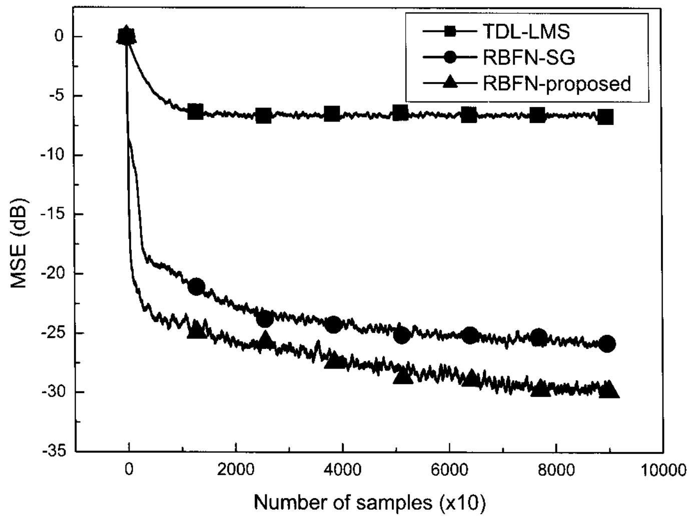

Fig. 4. MSE learning performance for nonlinear channel 2.

izer, TDL-LMS performs very poorly due to the nonlinearity of the channel. All the RBFN equalizers have given better performance than the linear equalizer. Most impressive is the superiority of the proposed RBFN equalizer to RBFN-SG. It is particularly noteworthy that more accurate width values can improve the RBFN performance significantly.

To verify the results of enhanced performance, we performed the simulation for another nonlinear channel model. The second simulation is with the nonlinear channel 2 [9] as follows

$$x(n) = h^{(n)} + 0.1h^{(n)^2} + 0.05h^{(n)^3} + v^{(n)},$$
$$h^{(n)} = 0.5d^{(n)} + d^{(n-1)}. \qquad (25)$$

Fig. 4 shows that the proposed RBFN equalizer has faster learning performance than the linear TDL-LMS and RBFN-SG equalizer. From the results depicted in Fig. 3 and 4, it can also be observed that the difference between the steady-state MSE values of the RBFN-SG and that of the proposed method exceeds 3 dB in both cases. The results confirm that controlling over step-size of the width in RBFN has significant effects on convergence speed as well as the steady-state value of MSE.

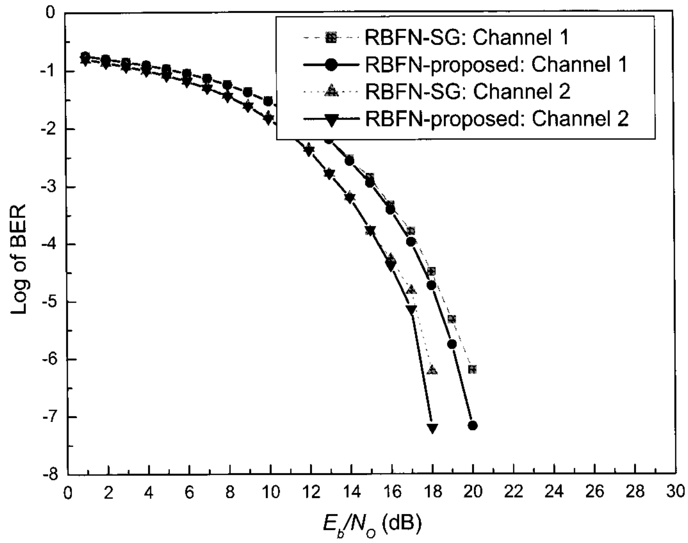

Fig. 5. BER performance with respect to E_b/N_o.

To give simulation results for more severe noise case, we have given the bit error ratio (BER) performance comparisons versus signal-to-noise ratio by increasing the variance of the noise. In general, nonlinear equalizers with lower MSE do not always reflect the lower BER. In the simulation results shown in Fig. 5, we have a similar phenomenon. The proposed algorithm gives only small BER performance enhancement in both channel environments. For error rate 10^{-6}, about 1 dB gain can be obtained but in the region of bigger noise variances, below 15 dB of E_b/N_o, two algorithms have almost the same BER performance.

The optimum step-size approach to the width parameter of RBFN can be considered to be extended to center and weight. From our results of applying this technique to all the RBFN parameters, we could not get considerably better performance than only the width. This implies that the width appears a major parameter that controls the convergence performance in RBF networks as described in Section I.

V. CONCLUSION

In this paper, we introduce a new method of width adaptation for the RBFN-SG algorithm by applying the Taylor's expansion approach to error signal and differentiating it with respect to the step-size in order to obtain optimal time-varying step-size for the widths in RBFN. The proposed approach to adjusting widths in RBFN can achieve superior learning performance in nonlinear channel environments. In the aspect of the steady-state MSE performance, the proposed method has shown increased performance by more than 3 dB, but in the aspect of BER performance, the proposed algorithm gives only small BER performance enhancement. The results confirm that controlling over step-size of the width in RBFN by the proposed algorithm gives small BER enhancement but has significant effects on convergence speed. This superior performance indicates that the proposed method for width adaptation can be a promising candidate for RBFN nonlinear channel equalization applications that require high speed convergence performance.

REFERENCES

[1] I. Cha and S. A. Kassam, "Time-series prediction by adaptive radial basis function networks," in *Proc. IEEE Twenty-Sixth Ann. Conf. Inf. Sci., Syst.*, 1993, pp. 818–823.

[2] I. Cha and S. A. Kassam, "Interference cancellation using radial basis function networks," in *Proc. IEEE Signal processing*, vol. 47, 1995, pp. 247–268.

[3] K. H. Kwon, N. Kim, and H. G. Byun, "On training neural network algorithms for odor identification for future multimedia communication system," in *Proc. IEEE ICME2*, July 2006, pp. 1309–1312.

[4] S. McLoone and G. Irwin, "Nonlinear optimization of RBF networks", *Int. J. Syst. Sci.*, vol. 29, pp. 179–189, 1998.

[5] S. Haykin, *Neural Networks: A Comprehensive Foundation*, Prentice-Hall, 1999.

[6] N. Mai-Duy and T. Tran-Cong, "Numerical solution of differential equations using multiquadratic radial basis function networks," *Neural Netw.*, vol. 14, pp. 185–199, 2001.

[7] D. Havelock, S. Kuwano, and M. Vorlander, *Handbook of Signal Processing in Acoustics*, Springer, 2008.

[8] S. Chen, G. Gibson, C. Cowan, and P. Grant, "Adaptive equalization of finite nonlinear channels using multilayer perceptrons," in *Proc. IEEE Signal Process.*, vol. 20, 1990, pp. 107–119.

[9] D. Sebald and Bucklew, "Support vector machine techniques for nonlinear equalization," *IEEE Trans. Signal Process.*, vol. 48, pp. 3217–3226, 2000.

Namyong Kim received the B.S., M.S., and Ph.D. degree from Yonsei University, all in Electronic Engineering in 1986, 1988, and 1991, respectively. From 1992 to 1997, he was with Kwandong University, Korea. In 1998–2002, he was an Associate Professor in the department of Information & Communications Engineering at Kangwon National University and currently he serves as a Professor. His current research interests are in adaptive signal processing in mobile communications and odor sensing-identification systems.