Identification of Nonlinear Systems in Presence of Outliers Using Robust Norm and Differential Evolution

Babita Majhi, H. Pal Thethi, and Jeetamitra Satpathy

***Abstract*—Conventional error based cost function provides unsatisfactory weight update of an adaptive system when outliers are present in the training signal. To alleviate this problem in this paper a hybrid approach using differential evolution (DE) and Wilcoxon norm is proposed to provide robust training in identification of complex nonlinear systems. Exhaustive simulation study shows superior performance of the new method compared to the conventional square error based minimization method.**

***Index Terms*—Differential evolution, robust system identification, wilcoxon norm.**

I. Introduction

The identification refers to determination of a suitable mathematical model of a given system, from which the behavior of the system under different operating conditions can be predicted. The identification also helps to develop a controller which enables the system to perform in a desired manner. The practical plants and systems are mostly nonlinear and dynamic in nature and hence their identification is complex and challenging task. Many research efforts still ongoing to develop accurate and appropriate model of complex plants when outliers are present in the training signal. Under such practical situations the training of model using conventional learning algorithms such as the least mean square (LMS) or recursive least square (RLS) [1] becomes ineffective and inaccurate. But all these derivative based algorithms involve squared of the error term as the fitness function. Further many evolutionary computing techniques such as genetic algorithm (GA) [2], particle swarm optimization (PSO) [3], bacterial foraging optimization (BFO) [4], ant colony optimization (ACO) [5] and differential evolution (DE) [6] have been employed as optimization tools in identification tasks. This new approach employs the mean square error (MSE) as the fitness or cost function. When outliers are present in the training signal, the convergence performance of MSE based learning algorithm is poor. Therefore there is a need to suggest improved method of nonlinear system identification under such adverse situation of outliers in the training signal. Conventionally an adaptive filter minimizes the Euclidean norm of the error, while an robust estimator minimizes a robust cost function (RCF) called Wilcoxon norm [7]. Accordingly robust training of artificial neural network (ANN) have recently been proposed for approximation of nonlinear functions [8]. In this paper a new method of robust identification of nonlinear dynamic systems or plants is proposed by progressively minimizing the RCF [7], of errors using a derivative free DE technique. The identification potentiality of the new method is assessed through simulation study and is compared with the results obtained from corresponding Euclidean norm based DE technique. The main contribution of the paper is the formulation of robust identification task as an optimization problem. The second contribution is the effective minimization of a robust norm by employing a population based derivative free DE technique which essentially adjusts the weights of the model.

The organization of the paper is outlined in six sections. Section II introduces the identification problem. An overview of DE is dealt in Section III. The new learning algorithm required for robust identification of the models using DE is developed and presented in Section IV. To validate the performance of the new model simulation study of different nonlinear systems is carried out in Section V. The conclusion of the proposed investigation is drawn in Section VI.

II. Robust Identification of Nonlinear Systems

A scheme of identification of a dynamic nonlinear system is shown in Fig.1 in which $x(n), \hat{y}(n), y(n)$ and $e(n)$ denote the input, output of the model, the output of the system and the error between the two at n th time instant respectively. The input $x(n)$ is an uniformly generated white signal. About 10% to 50% of the samples of the system are contaminated with strong outliers of magnitude as high as ± 5 to ± 20. The error term is thus given by

$$e(n) = y(n) - \hat{y}(n) \qquad (1)$$

From the error samples the Wilcoxon norm is computed which is then minimized by changing the weights of the model by using DE algorithm. The Wilcoxon norm of the errors is a robust norm against outliers and is explained here.

Robust norm of error terms (Wilcoxon Norm) [7],[8]

A score function, $\phi(u)$ of error terms has two characteristics

$$\int_0^1 \phi(u)du = 0 \text{ and } \int_0^1 \phi^2(u)du = 1 \qquad (2)$$

Manuscript received May 10, 2012; revised June 25, 2012.

B. Majhi is with the Dept. of CS & IT, G. G. Vishwavidyalaya, Central University, Bilaspur, India(e-mail: babita.majhi@gmail.com).

H. P. Thethi is with School of Electronics, KIITS University, Bhubaneswar, India (e-mail: harpalnitr@gmail.com).

J. Satpathy is with Dept. of Electronics Engg., at Govt. Polytechnic, Bhubaneswar , India (e-mail: jeetams@sify.com).

Let an error vector be defined as

$$E = [e(1)\ e(2) ....e(K)]^T \tag{3}$$

Rearrange the error vector in an increasing order and let each element v_i of the ordered vector is ranked serially as $1 \le i \le K$. In other words let $R(v_i) = i$ denotes the rank or order value of v_i such that $v_1 < v_2 < v_3 .... < v_K$.

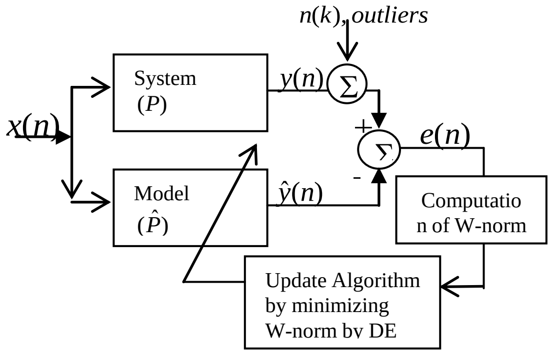


Fig. 1. Robust identification scheme of a dynamic system using DE

The score associated with the score function is $a(i)$ and is given by

$$a(i) = \phi(u) = \phi\left(\frac{i}{K+1}\right) = \sqrt{12}(u - 0.5) \\ = \sqrt{12}\left(\frac{i}{K+1} - 0.5\right) \tag{4}$$

where $u = \dfrac{i}{K+1}$. Then the Wilcoxon norm, C_i is defined as

$$C_i = \sum_{i=1}^{K} a(i) v_i \tag{5}$$

III. Differential Evolution

In this section the basic principle of DE is outlined. The initial target population is denoted by *NP*. Each individual of the population has *D* parameter values which are initially selected randomly between predefined search ranges $x_{ij}^{\min}$ and $x_{ij}^{\max}$. Thus

$$x_{ij}^0 = x_{ij}^{\min} + (x_{ij}^{\max} - x_{ij}^{\min}) \times r \tag{6}$$

where x_{ij}^t represent the i^{th} individual for j^{th} dimension at t^{th} generation and r is a uniform random number between 0 and 1. Mutant individuals are generated from the target population by adding the weighted difference between two randomly selected target population members to a third member in the target population. It is given by

$$v_{ij}^t = x_p^{t-I} + F \times (x_q^{t-I} - x_r^{t-I}) \tag{7}$$

where p, r and r are three random chosen individuals from the target population such that $(p \ne q \ne r \ne i \in (1,....., NP))$ and $(j = 1,2,......, D)$. F is a mutation scale factor which affects the differential variation between two individuals and its value lies between 0 to 2. Mutation phase is followed by the recombination of mutant individual with its corresponding target individual. To achieve this, a crossover operation is performed to obtain the trial individual as

$$u_{ij}^t = \begin{cases} v_{ij}^t & if\ r_{ij}^t \le CR\ or\ j = D_j \\ x_{ij}^t & otherwise \end{cases} \tag{8}$$

where D_j denotes to a randomly chosen dimension $(j = 1,2,......, D)$, which ensures that at least one parameter of each trial individual u_i^t differs from its counterpart in the previous generation u_i^{t-1}. CR is crossover constant in the range [0,1] and r_{ij}^t is a uniformly distributed random number between 0 and 1.

During the reproduction of the offspring, it is possible that the trial individual may exceed the search space. For this reason, parameter values violating the search range are restricted according to (6)

To include a trial individual u_i^t as a member of the target population for the next generation, it is compared to its counterpart x_i^{t-I} at the previous generation. The selection is based on the survival of the fittest among the trial and target individuals. Hence

$$x_i^t = \begin{cases} u_i^t & if\ \ f(u_i^t) \le f(x_i^{t-I}) \\ x_i^{t-I} & otherwise \end{cases} \tag{9}$$

where $f(.)$ denotes the fitness value of an individual.

IV. DE Based Robust System Identification : The Algorithm

Referring to Fig.1,

$$y(k) = P[x(k)] + n(k) \tag{10}$$

where $n(k)$ represents the outliers with amplitude of ± 5 or ± 20 present with 10% to 50% samples of the plant outputs. The DE based model consists of an equal order FIR system with unknown coefficients. In the present adverse situation the learning algorithm to be developed should be immune to the presence of outliers in generating output response of the model which is in close agreement with the system response $y(k)$. The DE based robust parameter updating rule is given in following steps :

1) The coefficients of the model are initially chosen from a population of *M* target vectors. Each target vector constitutes *P* number of parameters and each parameter represents one coefficient of the adaptive filer.
2) *K* number of input samples uniformly distributed

between -0.5 to $+0.5$ is used as input to the model of Fig. 1. In case of DE, K= 100.

3) Each of the input samples is passed through the plant $P(z)$ and at 10-50% of randomly chosen samples of the output are added with outliers. In this way K training samples are produced.
4) Each input sample is also passed through the model using each target vector as model parameters. Thus M sets of K estimated outputs are obtained.
5) Each of the desired output is compared with corresponding estimated output and K errors are produced for each target vector.
6) Let the error vector of P^{th} target vector at n th generation due to application of K input samples to the model be represented as $[e_{1,p}(n), e_{2,p}(n), \ldots\ldots\ldots, e_{K,p}(n)]^T$. The error values are then arranged in an increasing manner from which the rank $R\{e_{n,p}(n)\}$ of each n th error term is obtained. The score associated with the rank of the each error term is evaluated as

$$a(i) = \sqrt{12}\left(\frac{i}{K+1} - 0.5\right) \quad (11)$$

where $i,\ (1 \le i \le K)$ denotes the rank of an error. At n th generation of each p th target vector the Wilcoxon norm is then calculated as

$$C_p(n) = \sum_{i=1}^{K} a(i)\, e_{i,p}(n) \quad (12)$$

7) Since the objective is to minimize Wilcoxon norm the DE based optimization strategy is used.
8) The mutation and crossover operations are carried out sequentially as per DE rule given in Section III. Selection operator is finally used to select the best M target vectors
9) The learning process is stopped when Wilcoxon norm (WN) at an iteration reaches the minimum level.
10) Then the elements of the corresponding best target vector represent the desired robust estimation of parameters.

V. SIMULATION STUDY

In this section, simulation study carried out to assess the identification performance of the proposed algorithm is outlined and the results of nonlinear identification of dynamic plants in presence of 10% to 50% of outliers in the desired signal are presented. The outliers are uniformly distributed random values within the range from -5 to 5 up to -20 to +20. These are added to 10% to 50% randomly selected training samples. The following nonlinear channel models are used in the simulation study :

Example-1

(a) Parameter of the linear system of the plant [0.2600, 0.9300, 0.2600]

Example-2

(a)Parameter of the linear system of the plant [0.3040, 0.9030, 0.3040]

(b) Nonlinearity of the plants

$$NL1: y_n(k) = \tanh\{y(k)\}$$

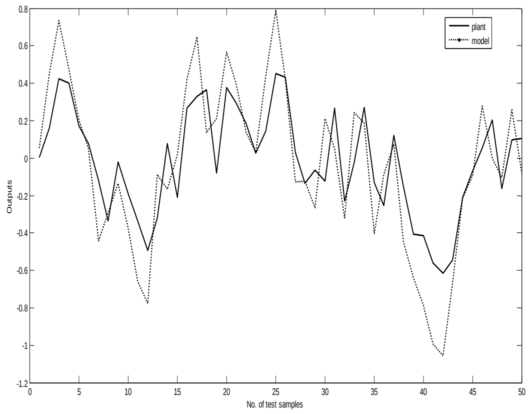
$$NL2: y_n(k) = y(k) + 0.2y^2(k) - 0.1y^3(k)$$

where $y(k)$ is the output of the linear part of the plant and $y_n(k)$ is the output of the overall system.

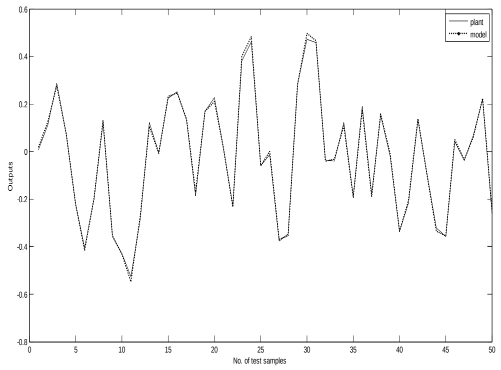
In scheme-1, the mean square error (MSE) is used as the cost function where as in schemes-2 it is Wilcoxon norm defined in Section-III. The performance of the proposed scheme are obtained from simulation studies and compared with that obtained by scheme-1. Sum of squared error (SSE) is used as quantitative measure for performance evaluation. The parameters used in the study are no. of particles=30, no. of input samples=100, population size = 30, no. of generations = 100, no. of ensample average = 10, and $F = 0.9, CR = 0.9$. These typical values are selected as it offers best possible simulation performance. Simulation is carried out at different conditions of outliers but the results presented in Figs. 2(a) – 2(d) are comparison of responses of the plant and the model for 50% outlier only. It is evident from these plots that robust model provides accurate response matching in presence of outlier whereas the conventional scheme-1 based model exhibits poor identification performance. In both examples, the SSE obtained from scheme-2 is listed in Table 1 to Table 4 is much lower than that obtained from scheme-1 model.

TABLE I: COMPARISON OF SUM OF SQUARED ERROR FOR EXAMPLE-1 WITH NL1

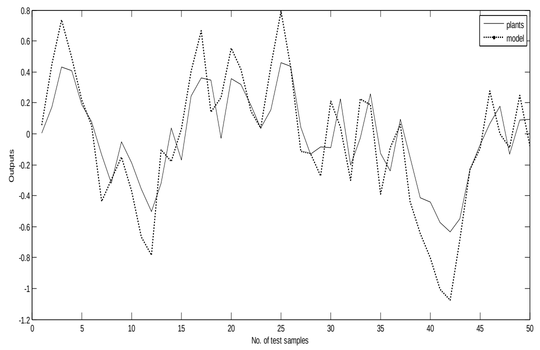
Percentage of outliers	Sum of squared error	
	Scheme-1	Scheme-2
Outlier range within -5 to 5		
10%	0.0832	0.0098
20%	0.6055	0.0095
30%	0.1280	0.0064
40%	1.4764	0.0063
50%	2.5348	0.0055
Outlier range within -10 to 10		
10%	0.3077	0.0098
20%	0.6055	0.0095
30%	0.2986	0.0064
40%	2.5314	0.0063
50%	2.5085	0.0055
Outlier range within -20 to 20		
10%	0.6130	0.0098
20%	0.6055	0.0095
30%	0.9826	0.0064
40%	2.6469	0.0063
50%	2.4586	0.0055



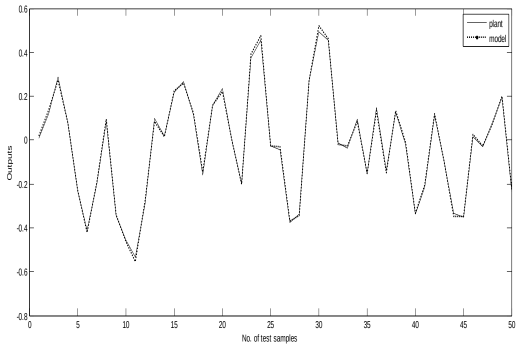
2(a) Comparison of responses of example-1 with NL1 in presence of 50% outliers within the range -20 to 20 using scheme-1



2(b) Comparison of responses of example-1 with NL1 in presence of 50% outlier within the range -20 to 20 using scheme-2



2(c) Comparison of responses of example-2 with NL2 in presence of 50% outliers within the range -20 to 20 using scheme-1



2(d) Comparison of responses of example-2 with NL2 in presence of 50% outliers within the range -20 to 20 using scheme-2

TABLE II: COMPARISON OF SUM OF SQUARED ERROR FOR EXAMPLE-1 WITH NL2

Percentage of outliers	Sum of squared error	
	Scheme-1	Scheme-2
Outlier range within -5 to 5		
10%	0.1243	0.0270
20%	0.6472	0.0306
30%	0.1179	0.0259
40%	1.4832	0.0242
50%	2.5942	0.0280
Outlier range within -10 to 10		
10%	0.2903	0.0270
20%	0.6472	0.0306
30%	0.2821	0.0259
40%	2.4343	0.0242
50%	2.5620	0.0280
Outlier range within -20 to 20		
10%	0.5884	0.0270
20%	0.6472	0.0306
30%	0.9534	0.0259
40%	2.5551	0.0242
50%	2.5066	0.0280

TABLE III: COMPARISON OF SUM OF SQUARED ERROR FOR EXAMPLE-2 WITH NL1

Percentage of outliers	Sum of squared error	
	Scheme-1	Scheme-2
Outlier range within -5 to 5		
10%	0.0842	0.0109
20%	0. 8370	0.0105
30%	0.1632	0.0066
40%	1.4760	0.0066
50%	2.2872	0.0059
Outlier range within -10 to 10		
10%		0.0109
20%		0.0105
30%		0.0066
40%		0.0066
50%		0.0059
Outlier range within -20 to 20		
10%		0.0109
20%		0.0105
30%		0.0066
40%		0.0066
50%		0.0059

TABLE IV: COMPARISON OF SUM OF SQUARED ERROR FOR EXAMPLE-2 WITH NL2

Percentage of outliers	Sum of squared error	
	Scheme-1	Scheme-2
Outlier range within -5 to 5		
10%	0.0882	0.0272
20%	0.8956	0.0309
30%	0.1437	0.0272
40%	1.4826	0.0256
50%	2.3769	0.0292
Outlier range within -10 to 10		
10%	0.2892	0.0272
20%	0.8956	0.0309
30%	0.3092	0.0272
40%	2.1657	0.0256
50%	2.3461	0.0292
Outlier range within -20 to 20		
10%	0.7399	0.0272
20%	0.8956	0.0309
30%	0.9831	0.0272
40%	2.2879	0.0256
50%	2.2935	0.0292

VI. Conclusion

The paper has introduced the DE based training of the parameters of nonlinear identification models using robust norm. Extensive simulation studies demonstrate that more accurate and robust models can be generated using the DE and Wilcoxon norm compared to those obtained by conventional squared error based norm. Thus when outliers are present in the training samples the Wilcoxon norm based training outperforms the conventional squared error norm based training algorithms

References

[1] B. Widrow and S. D. Sterns, *Adaptive Signal Processing*, Pearson Education Englewood Cliffs: Prentice-Hall, 1985.

[2] T. Kumon, M. Iwasaki, T. Suzuki, T. Hashiyama, N. Matsui, and S. Okuma, "Nonlinear system identification using Genetic algorithm," in Proc. of *IEEE 26th Annual conference on Industrial Electronics Society (IECON 2000)*, vol. 4, 22-28 Oct. 2000, pp. 2485 - 2491.

[3] J. Kennedy , R. C. Eberhart, and Y. Shi, *Swarm Intelligence*, San Francisco: Morgan Kaufmann Publishers, 2001.

[4] K. M. Passino, "Biomimicry of Bacterial Foraging for distributed optimization and control," *IEEE Control system magazine*, vol 22, issue 3, pp. 52-67, June 2002.

[5] X. Qiang, L. Jihai, and Y. Jia, "Identification and control based on neural networks and ant colony optimization algorithm," *IEEE Int. conf. on Logistics systems and intelligent management*, vol. 2, 2010, pp. 1255-1258.

[6] M. M. Olama, S. M. Djouadi, and C. D. Charalambous, "Stochastic differential equations for modeling, estimation and identification of mobile to mobile communication channels," *IEEE Trans. on Wireless Communications*, vol. 8, issue 4, pp. 1754-1763, 2009.

[7] J. W. McKean, "Robust analysis of Linear models," *Statistical Science*, vol. 19, no. 4, pp. 562-570, 2004.

[8] J. G. Hsieh, Y. L. Lin, and J. H. Jeng, "Preliminary study on Wilcoxon learning machines," *IEEE Trans. on neural networks*, vol. 19, no. 2, pp. 201-211, Feb. 2008

B. Majhi is working as Asst. Professor in the Dept. of CS & IT, G. G. Vishwavidyalaya, Central University, Bilaspur, India. She completed her Ph.D.(CS&E) from National Institute of Technology Rourkela and Master (CA) from Sambalpur University in 2005 and 2009 respectively. She is continuing her Postdoctoral work at University of Sheffield, UK under the Boyscast Fellowship of DST, Govt. of India during 2011-2012. To her credit she has published 77 research papers in various referred journals and international conferences. She has guided one Ph. D. and 06 M. Tech. theses. Her research interests are digital signal processing, soft and evolutionary computing, computational finance and data mining.

H. P. Thethi is working as Asst. Professor in the School of Electronics, KIIT University, Bhubaneswar, India. He did his Ph. D. from KIIT University in 2011 and M. Tech. from National Institute of Technology Rourkela. Presently he is also the Innovation Chair Professor, for Mission1x Concept Center from Wipro Technologies at KIIT University. He has reviewed two books on Signal Processing from TMH and published several papers in various national and international conferences and journals. His research interests are Signal Processing, Adaptive Signal Processing and VLSI Design.

J. Satpathy is working as a Lecturer in the Dept. of Electronics Engg., at Govt. Polytechnic, Bhubaneswar since 1996. She completed her B. Tech. (Electronics and Telecommunication) from Orissa Engg. College affiliated to Utkal University and M. Tech. (Electronics and Communication Engg.) from C. V. Raman College of Engg. affiliated to BPUT, Orissa in 2010. Her research interests are digital signal processing and Soft computing

.