Annals of DAAAM for 2010 & Proceedings of the 21st International DAAAM Symposium, Volume 21, No. 1, ISSN 1726-9679
ISBN 978-3-901509-73-5, Editor B. Katalinic, Published by DAAAM International, Vienna, Austria, EU, 2010
Make Harmony Between Technology and Nature, and Your Mind will Fly Free as a Bird
Annals of DAAAM International

A LEVENBERG-MARQUARDT ALGORITHM WITH INSTRUMENTAL VARIABLES AND ITS APPLICATION ON THE IDENTIFICATION OF DYNAMIC SYSTEM

DOKOUPIL, J[akub] & PIVONKA, P[etr]

Abstract: *This article deals with process identification, using nonlinear ARX model via feed-forward multilayer neural network. Estimation of network parameters is achieved using the Levenberg-Marquardt (LM) method in iterative batch mode adaptation. In order to obtain consistent estimate, original implementations of LM algorithm − which include instrumental variables (IV) technique − are suggested. Basic and extended IV methods are presented as some of the IV methods. Advantages of the proposed approach are illustrated in the example simulations on the real process, using B&R PLC.*
Key words: *Neural networks, Levenberg-Marquardt, NARX, Instrumental variables, System identification*

1. INTRODUCTION

Building a representative mathematical model of the process plays an important role in the controller synthesis. Input-output identification in the field of linear least squares methods is, as far as consistent estimate is concerned, sufficiently developed. These methods, however, do not completely have to approximate the dynamics of generally nonlinear system. Motivation of this work is therefore to use nonlinear feed-forward neural network, which is able to obtain consistent estimate of the process. Consistency of the estimate is achieved using instrumental variables into algorithms in order to update the network parameters. Neural network training is based on a LM search direction, which is often in many aspects the best gradient method for the training. The next area of research will introduce instruments which allow for direct identification in closed-loop with controller.

2. IV METHODS IN NEURAL NETWORKS

A predicted output of the dynamic feed-forward neural network is generally determined by a smooth function $g(\cdot,\cdot)$ which maps the input values to the output of the network (Šíma & Neruda, 1998)

$$\hat{y}(k,\theta) = \hat{g}(\varphi(k),\theta) \tag{1}$$

Where $\hat{y}(k,\theta)$ is the predicted output value, $\varphi(k) \in \mathbb{R}^{n_\varphi}$ is the regression vector and $\theta \in \mathbb{R}^n$ ($n \geq n_\varphi$) is the parameter vector. Approximation of nonlinear SISO process using the NARX model via neural network will be dealt with in the article. Network input vector is therefore filled in conformity with the regressor vector for the optimal ARX predictor

$$\varphi(k) = [u(k-1) \ \cdots \ u(k-n_b) \ -y(k-1) \ \cdots \\ -y(k-n_a)]^T \tag{2}$$

where $u(k)$ and $y(k)$ are sequences of measured values of inputs and outputs of the process with orders n_b, n_a.

The concept of batch adaptation requires minimizing the value of the cost function $V_k(\theta)$ for the input-output data batch $\{u(i) \ \ y(i)\}_{i=1}^N$ in form

$$V_N(\theta) = \frac{1}{N}\sum_{i=1}^N \varepsilon^2(i,\theta) \tag{3}$$

where $\varepsilon(k,\theta) = y(k) - \hat{y}(k,\theta)$.

If we proceed from the initial estimate $\hat{\theta}_N^{(0)}$, general off-line gradient algorithm minimizing the cost function $V_k(\theta)$ is given by the iterative calculation (Nelles, 2001)

$$\hat{\theta}_N^{(k+1)} = \hat{\theta}_N^{(k)} - \eta^{(k)} \left[R_N^{(k)}\right]^{-1} V_N'\left(\hat{\theta}_N^{(k)}\right) \tag{4}$$

Where $\hat{\theta}_N^{(i)}$ is the parameter estimation in ith iteration of the calculation, $R_N^{(i)} \in \mathbb{R}^{n\times n}$ is the directional matrix which modifies the local search direction proportional to some learning rate $\eta^{(i)}$ into the gradient direction $V_N'\left(\hat{\theta}_N^{(i)}\right) \in \mathbb{R}^n$. An individual gradient algorithms just differ from each other in realization of the directional matrix $R_N^{(i)}$. The calculations of the $V_N'\left(\hat{\theta}_N^{(k)}\right)$ and of the $R_N^{(i)}$ for the LM are given by

$$V_N'\left(\hat{\theta}_N^{(k)}\right) = -\sum_{i=1}^N \psi\left(i,\hat{\theta}^{(k)}\right)\varepsilon\left(i,\hat{\theta}^{(k)}\right) \tag{5}$$

$$R_N^{(k)} = \sum_{i=1}^N \psi\left(i,\hat{\theta}^{(k)}\right)\psi\left(i,\hat{\theta}^{(k)}\right)^T + \delta^{(k)} I \tag{6}$$

Where we introduced $\delta^{(k)} \geq 0$ to ensure regularity of the matrix $R_N^{(k)}$ and the vector $\psi(k,\theta) \in \mathbb{R}^n$

$$\psi(i,\theta) = \frac{\partial \hat{g}(\varphi(i),\theta)}{\partial \theta} \tag{7}$$

In the analysis which follows it is assumed that the true system is given by

$$y(k) = g(\varphi(k),\theta_0) + v(k) \tag{8}$$

where θ_0 is the vector of true parametres and $v(k)$ represents zero-mean stochastic disturbance term. We demand equality of the true and the estimated system parameters, thus

$$\begin{aligned}\hat{\theta}_N^{(k+1)} - \theta_0 &= \hat{\theta}_N^{(k)} - \theta_0 + R^{-1}\Big[\sum_{i=1}^N \psi\left(i,\hat{\theta}^{(k)}\right)\varepsilon\left(i,\hat{\theta}^{(k)}\right) \\ &\quad -\sum_{i=1}^N \psi(i,\theta_0)\left(g(\varphi(k),\theta_0) - g(\varphi(k),\theta_0)\right)\Big] \\ &= R^{-1}\left[\sum_{i=1}^N \psi\left(i,\hat{\theta}^{(k)}\right) v(i)\right]\end{aligned} \tag{9}$$

The obtained estimate will be consistent ($\lim_{N\to\infty} \hat{\theta}$=$\theta_0$) unless

$$E\left(\psi\left(i,\hat{\theta}^{(k)}\right)\psi\left(i,\hat{\theta}^{(k)}\right)^T + \delta^{(k)} I\right) \text{ is nonsingular} \tag{10}$$

$$E\left(\psi\left(i,\hat{\theta}^{(k)}\right)v(i)\right) = 0 \tag{11}$$

To keep the validity of (10) even in numerically ill-conditioned cases is guaranteed by a suitable choice of $\delta^{(k)}$ (Hristev, 1998). Validity of the relation (11) is conditioned by the properties of the quantity $v(i)$. Ideally, properties of white noise are assumed, i.e. independence of $v(k)$ on their previous values. In case of colored noise, it is possible to use two basic approaches in order to fulfill the validity (11). These are whitening of the prediction error or just introduction of instrumental variables.

2.1 Basic IV method

The idea of the method is to introduce the vector of instruments $\zeta(k) \in \mathbb{R}^n$ which does not correlate with the disturbance $v(k)$ (Ljung, 1999). Consistent estimate $\hat{\theta}$ is obtained as a solution to the overdetermined set of linear equations representing the gradient of the cost function

$$\frac{1}{N}\sum_{i=1}^{N} \zeta(i)\, \varepsilon\big(i, \hat{\theta}^{(k)}\big) = 0 \qquad (12)$$

With regard to (12), we assume appropriate matrices for the LM method

$$V_N'\left(\hat{\theta}_N^{(k)}\right) = -\sum_{i=1}^{N} \zeta(i)\, \varepsilon\big(i, \hat{\theta}^{(k)}\big) \qquad (13)$$

$$R_N^{(k)} = \sum_{i=1}^{N} \zeta\big(i, \hat{\theta}^{(k)}\big)\, \psi\big(i, \hat{\theta}^{(k)}\big)^T + \delta^{(k)} I \qquad (14)$$

2.2 Extended IV method

Extended IV method is a generalization of the basic IV method as it uses augmented vector $\zeta(k) \in \mathbb{R}^{n_\zeta}$ $(n_\zeta \geq n)$ and its prefiltering by the stable filter $F(q)$ (Söderström & Stoica, 1989). By developing a weighted cost function into the Taylor series and its minimization,

$$V_N(\theta) = \left\| \frac{1}{N}\left[\sum_{i=1}^{N} \zeta(i) F(q) y(i)\right] - \frac{1}{N}\left[\sum_{i=1}^{N} \zeta(i) F(q) \varphi(i)^T\right]\theta \right\|_Q^2 \qquad (15)$$

where $\|x\|_Q^2 = x^T Q x$ with positive definite weighting matrix Q, we obtain the relations for the LM method

$$V_N'\left(\hat{\theta}_N^{(k)}\right) = -\sum_{i=1}^{N} \psi\big(i, \hat{\theta}^{(k)}\big)\, W(i) \varepsilon\big(i, \hat{\theta}^{(k)}\big) \qquad (16)$$

$$R_N^{(k)} = \sum_{i=1}^{N} \psi\big(i, \hat{\theta}^{(k)}\big) W(i)\, \psi\big(i, \hat{\theta}^{(k)}\big)^T + \delta^{(k)} I \qquad (17)$$

$$W(i) = F(q) \zeta^T(i) Q(i) \zeta(i) F(q) \qquad (18)$$

Some possibilities of how to choose a vector of instruments are described in the relations (19), (20) and (21)

$$\zeta(k) = \frac{\partial \hat{g}(\varphi(k), \theta)}{\partial \theta} \qquad (19)$$

According to the choice of vector $\varphi(k)$ we distinguish IV method with delayed observations (IVd) with $n_c \in \mathbb{Z} \cup \{0\}$ or IV method with the additional model (IVm)

$$\varphi(k) = [u(k-1) \;\; \cdots \;\; u(k-n_b) \;\; -y(k-n_c-1) \;\; \cdots \;\; -y(k-n_c-n_a)]^T \text{ (IVd)} \qquad (20)$$

$$\varphi(k) = [u(k-1) \;\; \cdots \;\; u(k-n_b) \;\; -\hat{y}(k-1, \theta) \;\; \cdots \;\; -\hat{y}(k-n_a, \theta)]^T \text{ (IVm)} \qquad (21)$$

3. REAL DATA EXPERIMENTS

Verification of individual modifications of the LM algorithm was performed in real process in the form of air tunnel with nonlinear static characteristic shown in Fig. 1. Sampling period of the process was chosen as $T_s = 0.2$s. Fig. 2 shows responses of the process as well as of the mathematical model created using the LM-IVd method. Fig. 3 then shows detailed progress on chosen data horizon, pointing to the significant influence of the disturbing stochastic part of process. Summary of results of individual LM method modifications, as far as the value of cost function $V_N(\theta)$ according to (3) is concerned, is given in Tab. 1.

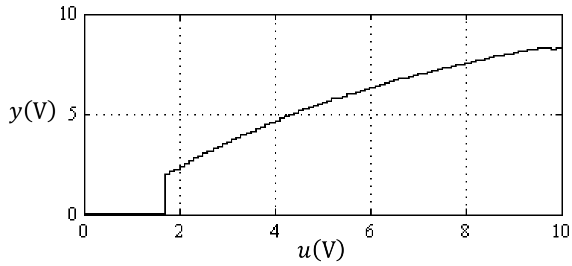

Fig. 1. Static characteristics of the real process

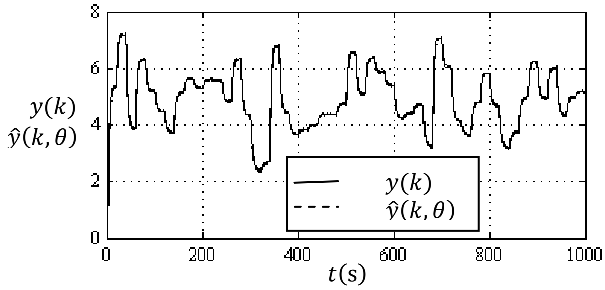

Fig. 2. The process output $y(k)$ and its corresponding multi-step model prediction $\hat{y}(k,\theta)$

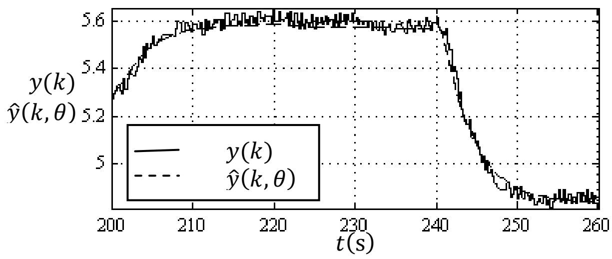

Fig. 3. Detail of the process and its model response

	LM-IVd	LM-IVm	LM
$V_N(\theta)$	0.0076	0.0065	0.0179

Tab. 1. Obtained cost function values

4. CONCLUSION

This paper presents and develops the LM algorithm with instrumental variables for the use of dynamic modeling. The simulation results for parameter estimation of the NARX model achieved in the real process demonstrate significant increase in identification quality in cases when IV was introduced, as opposed to the classical LM method used.

5. ACKNOWLEDGEMENTS

This work was supported as a part of the solution of Czech Science Foundation GAČR project No. 102/09/1680 Control Algorithm Design by Means of Evolutionary Approach and by the Czech Ministry of Education in the frame of MSM 0021630529 Intelligent Systems in Automation.

6. REFERENCES

Söderström, T. & Stoica, P. (1989). *System Identification,* Prentice-Hall Int., ISBN 0-13-881236-5, Cambridge

Nelles, O. (2001). *Nonlinear System Identification,* Springer, ISBN 3-540-67369-5, Berlin

Šíma, J.; Neruda, R. (1998). *Theoretical Issue of Neural Networks.* MATFYZPRESS, ISBN 80-85863-18-9, Prague

Ljung, L. (1999). *System Identification – Theory for the User,* Prentice-Hall, ISBN 0-13-656695-2, London

Hristev, R. M. (1998). *The ANN BOOK.* GNU Public License