ANN-Based Parameter Estimation with Minimum Variance

Andrzej Materka, Michal Strzelecki
Technical University of Lodz, Institute of Electronics
Stefanowskiego 18, 90-924 Lodz, Poland
[materka,mstrzel]@ck-sg.p.lodz.pl

Abstract: The concept of using ANN-like approximators for estimation of dynamic system parameters is considered. It is shown that the modular, classifier-approximator architecture offers new possibilities of real-time observation point selection for optimal ANN performance - in terms of minimum parameter variance.

INTRODUCTION

Parametric modeling is a widespread methodology used to quantitatively describe properties and state of physical systems. It is especially useful for diagnostic purposes, e.g. in medicine, electronic circuit testing for prototyping and manufacture, control engineering, scientific research, and others [1]. The basic assumptions to this methodology are that there exists an adequate mathematical model for the physical dynamic system under consideration, given in terms of mathematical equations with constant coefficients (parameters). The model structure is known but actual values of model parameters are unknown. To find parameter values, the system is excited by a stimulus signal (either at input or in the form of nonzero initial conditions) and the system response is measured (observed). A set of system response samples forms the observation vector (Fig. 1).

It is assumed throughout this paper, without any loss of generality, that the sampled time-domain response of a system under test forms the observation vector. The stimulus and observations have to be defined such that there exists a unique mapping \( \theta = g(y) \) between the observation vector \( y = (y_1, y_2, \ldots, y_n) \)' and the parameter vector \( \theta = (\theta_1, \theta_2, \ldots, \theta_p)' \), \( p \leq n \), for any \( \theta \in \Theta \) and \( y \in \Omega \), where \( \Theta \) and \( \Omega \) are, respectively, the parameter and the observation spaces. This inverse mapping is usually unknown. It is traditionally determined pointwise - by iterative numerical calculation [2], performed each time a new observation vector is acquired [3], [4]. The iterative, time-consuming approach is not suitable for real-time, fast parameter identification.

A new technique has recently been developed for the parameter estimation task [5]. In this approach, an artificial neural network (ANN) is trained on examples to learn the mapping of interest. Once trained, the ANN recalls the parameter values in a short time, with no need for any iterative calculations (Fig. 2). Thus high speed is the advantage of the technique in a number of applications [6, 7]. The method applies to any linear or nonlinear quasi-stationary system, provided the mapping of interest is unique which can be ensured by known means [8]. The present research effort is focused on optimizing the ANN architecture for the task. A modular ANN was proposed in [9] that is easy to train in a highly repeatable manner for a wide range of unknown parameters - in the
The problem of system parameter estimation using noise-corrupted observations is considered in this paper.

\[ \hat{\theta} = \theta + \delta \theta \]  

The presence of the noise-induced parameter error \( \delta \theta \) is an undesired effect. Due to this effect, the parameter estimate \( \hat{\theta} \) is a random variable. The lower the variance of this random variable, the better the estimator under consideration. In this paper, we consider the issue of selection of the system response sampling moments such that the resulting observation vectors provide small variance of the parameter estimates. To illustrate this problem let us consider a one-dimensional system first, characterized by a single scalar parameter \( \theta_a \leq \theta \leq \theta_b \), where \( \theta_a \) and \( \theta_b \) are real numbers. The system response is given by

\[ y(t) = f(t, \theta) + v(t) \]  

for \( t \geq 0 \). In the noiseless case \( [v(t) = 0] \), the scalar observation \( y(t_o) = f(t_o, \theta) \) taken at any point \( t_o \geq 0 \), can be used to estimate the unknown parameter \( \theta \), provided that the inverse function

\[ \theta = g[f(t_o, \theta)] \]  

exists. In the one-dimensional case, the existence of the inverse function is guaranteed by a nonzero magnitude of the derivative \( \left| \frac{\partial^2 \theta}{\partial \theta} \right| > 0 \), \( \theta_a \leq \theta \leq \theta_b \). Therefore, the larger the magnitude of the derivative (5), the lower the variance of the estimated parameter. Thus the goal of the experiment design is to select such an observation point that maximizes the observation sensitivity to the parameter.

\[ x : t_o = 0.2, \sigma_v = 0.7, * : t_o = 1.2 \]

**Example 1** Consider a first-order system whose time-domain response to a test signal is

\[ y(t) = \exp(-t/\theta) \]  

Assume the actual value of parameter \( \theta \) is unknown; however, it is known that \( \theta \) is a real number from the interval \( [0.2, 1.2] \). The derivative of (6) with respect to \( \theta \) is given by

\[ \frac{\partial^2 y}{\partial \theta^2} = \frac{-t}{\theta^2} \exp(-t/\theta) \]  

Its magnitude attains a maximum at \( t = \theta \). Thus, there is no single observation moment \( t \) able to minimize the parameter variance for all \( \theta \epsilon [0.2, 1.2] \). Plots of the derivative (7) calculated for 3 different values of the observation moment are shown in Fig. 3. It is
seen in this example that for \( \theta<0.4 \), a good observation moment could be \( t_o=0.2 \). It produces a 6-times increase of the derivative at \( \theta=0.2 \), compared to the observation taken at \( t_o=0.7 \), which in turn gives the biggest observation sensitivity to \( \theta \) for moderate values of \( \theta \), \( 0.4<\theta<0.9 \). For large values of \( \theta \), \( \theta>0.9 \), the observation measured at \( t_o=1.2 \) results in the smallest parameter variance.

**OBSERVATION SUBSET SELECTION USING A MODULAR ANN**

The above analysis and Example 1 demonstrate that there is no single sampling moment that would be optimal for the whole parameter domain. In general, every value of the unknown system parameter may need its own observation times which correspond to a minimum parameter variance. It is postulated here that a modular classifier-approximator ANN architecture can be used as a means of real-time observation point selection to reduce the variance of parameter estimates.

![Fig. 4](image-url)  
**Fig. 4** First stage of information processing in the proposed modular ANN architecture

According to the approach, the system response is uniformly sampled in time to form an \( n \)-element observation vector, \( n \geq p \). A subset of this vector, comprising observations taken at appropriately selected \( p \) time instants forms a ‘nominal’ observation vector \( N \). This vector is used to represent (approximately) the whole parameter range and is applied to the input of a classifier network (Fig. 4). The classifier decides to which preselected parameter subrange \( S_j \), \( j=1,2,...,m \), the measured system response corresponds. An appropriate \( p \)-element subset \( O_j \) of the observation vector is then invoked, which is (in the assumed sense) optimal for the subrange selected. This vector is applied to the input of a neural-network-like approximator which produces the parameter estimate at its output (Fig. 5).

![Fig. 5](image-url)  
**Fig. 5** Second stage of information processing in the proposed modular ANN architecture

The variance of the parameter is on the average lower that it would be in the case of an approximator driven directly by the vector \( N \), because the vector \( N \) is not optimal for all the regions \( S_j \).

**Example 2** Suppose the nominal observation time for the system (6) is \( t_{nom}=0.7 \). The observation \( y(t_{nom}) \) changes from 0.00674 at \( \theta=0.2 \) through 0.08208 at \( \theta=0.4 \) and 0.32919 at \( \theta=0.9 \), to 0.4346 at \( \theta=1.2 \). Define three nonoverlapping subranges of \( \theta \): \( S_1=[0.2,0.4) \), \( S_2=[0.4,0.9) \) and \( S_3=[0.9,1.2) \), such that their union covers the whole parameter range \([0.2,1.2)\). The optimal sampling times for the subranges are selected to be, respectively, \( t_1=0.2 \), \( t_2=0.7 \), \( t_3=1.2 \). When substituted to (6), they define the observations \( y(t_i) \).

A network comprising simple threshold-nonlinearity classifiers was used to select the proper observation for the subrange identified. The Matlab-computed [11] standard deviation of the parameter \( \theta \), as obtained using the proposed technique is illustrated in Fig. 6 (lower curves). The upper curves in Fig. 6 show the noise performance of a standard neural-network estimator of Fig. 2 driven by the observation sampled at the nominal time.
Up to 10dB SNR improvement can be noticed at \( \theta = 0.2 \) in this example.

In the general \( p \)-dimensional case, the noise-induced parameter variance can be minimized by proper observation points selection, \( t_i, \ i = 1, 2, \ldots, n \), such that they maximize the determinant \( D \) of the so called information matrix that is defined based on observation sensitivity to parameters [8]. For most systems of practical interest, there is no single choice of observations that would be optimal for the whole parameter space, especially for a wide parameter range. A two-stage ANN-based parameter estimation procedure proposed in this paper can be used to deal with this problem. In this approach, the whole parameter space is divided into \( m > 1 \) lower-volume regions. Each region is allocated a subset of all available observation points, such that it maximizes \( D \) - locally within the region. In the recall phase, an arbiter ANN classifier first decides which of the predefined observation vectors should be given the highest confidence, and then a preoptimised approximator module is used, driven by the selected vector, to give the lowest-variance parameters.

To illustrate the method suitability to multiparameter systems, consider an exponential model

\[
y(t) = \theta_3 \left[ \exp\left(\frac{-t}{5\theta_1}\right) - \exp\left(\frac{-t}{\theta_2}\right) \right] + \nu(t) \tag{8}
\]

where \( \theta_1, \theta_2 \) and \( \theta_3 \) are 3 unknown parameters. The response (8) is quite ubiquitous, eg. it may represent biological multicompartamental systems [12] or frequency-selective electronic circuits [7]. To facilitate graphical presentation of the results, but without any loss of generality, the value of 1.0 was assumed for \( \theta_3 \). The other parameters take their values from a rectangle as follows

\[
\{ \Theta: 0.5 \leq \theta_1 \leq 1.5, 0.2 \leq \theta_2 \leq 1.2, \ \theta_3 = 1 \} \tag{9}
\]

For any \( \theta \in \Theta \), Equation (8) describes the system under test response to a test stimulus.

In the discussed case, there are \( p=2 \) unknown parameters, so 2 samples of the signal \( y(t) \) are necessary to determine the parameter values.

![Fig. 7 The reduction (in dB) of the variance of \( \theta_2 \) obtained with this technique compared to the estimation using single 'nominal' observation vector (Results of Monte-Carlo experiment using Matlab®)](image)
Figures 4 and 5 explain the role of the classifier and the approximator in the proposed ANN architecture. The system under test is excited by a stimulus $x(t)$. The system response is sampled to form the observation set $\{y_i, i=1,2,\ldots,n\}$ whose ‘nominal’ subset $N$ is applied to the input of the classifier. The classifier makes the decision about the number $j$, $j=1,2,\ldots,m$, of the parameter-space subregion $S_j$, to which a given observation region corresponds. Based on this decision, an optimal observation subset $O_j$ is selected and applied to the input of the approximator. The weights $w_j$ of the approximator are also adjusted to the subregion identified. The approximator produces the estimate $\hat{\theta}$ of the unknown parameters $\theta$.

**DISCUSSION AND CONCLUSION**

A modular ANN architecture has been proposed for fast and robust parameter estimation with a minimum variance. In the proposed approach, the observation space is split into $m$ non-overlapping subregions. An optimal observation vector is defined for each subregion. In the recall phase, one of these vectors is selected by the classifier and is applied to the input of the approximator. The architecture differs from the one presented in [9] in that there is the same input applied to all the ANN approximators in [9], whereas their inputs are optimally selected in this approach. Results of computer simulation are presented for a 3-parameter multi-exponential system. They demonstrate that by splitting the observation domain into 4 regions, and optimising the inputs to the approximator, the additive-noise-induced rms error can be reduced by as much as 5dB in this example. Further modifications of the proposed technique are under investigation, aimed at incorporating fuzzy-logic concepts for intelligent postprocessing of the outputs obtained from parallel-working approximators, each driven by respective preselected observation vector.

**Acknowledgment:** Support from Polish Scientific Research Committee Grant 8T11F 01010 is much appreciated.

**REFERENCES**