Adaptive neural networks for robust estimation of signal parameters

Tadeusz Łobos, Paweł Kostyła, Zbigniew Wacławek
Wroclaw University of Technology, Wroclaw, Poland, and
Andrzej Cichocki
FRP Riken ABS Laboratory, Institute of Physical and Chemical Research, Japan

Received June 1999
Revised January 2000
Accepted January 2000

Keywords *Neural networks, Estimation, Signal processing*

Abstract *In many applications, very fast methods are required for estimating of parameters of harmonic signals distorted by noise. Most of the known digital algorithms are not fully parallel, so that the speed of processing is quite limited. In this paper new parallel algorithms are proposed, which can be implemented by analogue adaptive circuits employing some neural networks principles. Algorithms based on the least-squares (LS) and the total least-squares (TLS) criteria are developed and compared. The problems are formulated as optimization problems and solved by using the steepest descent continuous-time optimization algorithm. The corresponding architectures of analogue neuron-like adaptive processors are also shown. The developed networks are more robust against noise in the measured signal than other known neural network algorithms. The network based on the TLS criterion optimizes the estimation under the assumption that the signal model can also be perturbated (frequency or sampling interval fluctuation and so forth). The TLS estimates are better and more reliable than the corresponding LS estimates, when applying a higher sampling frequency and a wider sampling window. The TLS algorithm is a generalization of the well known LMS rule and could be in some applications superior to the family of LMS algorithms. Extensive computer simulations confirm the validity and performance of the proposed algorithms.*

Introduction

Estimation of parameters (amplitudes) of harmonic signals is important in electric power systems and power electronics due to increasing use of nonlinear dynamic loads. The harmonics produced have usually varying amplitudes caused by the dynamic nature of nonlinear loads. Fast estimation of the parameters is essential for the control and protection of electric power systems. It is also useful in modelling, measurements and compensation of higher harmonics. Various digital and analogue (neural networks) algorithms have been proposed for the estimation of parameters of harmonic signals. They range from simple least-squares (LS) methods (Osowski, 1992; Cichocki and Łobos, 1994), least absolute value (LAV) technique (Van Den Bos, 1988; Cichocki and Łobos, 1994), minimax (Chebyshev norm) technique (Van Den Bos, 1988; Cichocki and Łobos, 1994; Cichocki *et al.*, 1994), methods based on the singular value decomposition

COMPEL : The International Journal for Computation and Mathematics in Electrical and Electronic Engineering, Vol. 19 No. 3, 2000, pp. 903-912.

(SVD) (Osowski, 1994), DFT (Łobos, 1989), to the Kalman filtering approach (Łobos, 1989).

The purpose of this paper is to present novel on-line techniques for estimation of parameters of harmonic signals based on the least-squares (LS) and the total least-squares (TLS) criteria (Golub and Van Loan, 1980). In the literature, Cichocki and Unbehauen (1993, 1994) presented methods for solving systems of linear equations are modified and adapted for signal processing problems. The problems are formulated as optimization problems and solved by using the steepest descent continuous-time optimization algorithm (Amari, 1990; Cichocki and Unbehauen, 1992, 1993). The solution of the optimization problem is based on some principles of neural network techniques given by Tank and Hopfield (1986). The developed networks contain elements which show a certain similarity to the adaptive threshold elements of the perception presented by Widrow and Lehr (1990). The corresponding architectures of analogue neuron-like adaptive processors are also shown. The developed methods are more robust against random noise in comparison with other known algorithms.

Mathematical formulation of the problem

The following standard problem has been analysed and solved.

Let $y(t)$ denote a measured noisy signal

$$y(t) = \sum_{i=1}^{n} a_i \ \sin(i\omega t) + \sum_{i=1}^{n} b_i \ \cos(i\omega t) + r(t) \tag{1}$$

where

$\omega = 2\pi f$ is a known (or rather approximately estimated) angular frequency,

$r(t)$ is unknown noise or error (residual),

a_i, b_i are unknown amplitudes of harmonic signals.

On the basis of values $y(t)$ it is necessary to find or estimate in real time the amplitudes a_i, b_i. Assuming that a continuous-time signal $y(t)$ is sampled and held with a sampling interval T, the problem can be mathematically reformulated as solving a large overdetermined system of linear equations

$$\mathbf{Dx} = \mathbf{y} \tag{2}$$

where

$$\mathbf{x} = [a_1, b_1, a_2, b_2, \ldots, a_n, b_n]^{\mathrm{T}} \in \mathbf{R}^{2n}$$
$$\mathbf{y} = [y(T), y(2T), \ldots, y(mT)] \in \mathbf{R}^{m}$$

and

$$\mathbf{D} = \begin{bmatrix} sin(\omega T) & cos(\omega T) & \cdots & sin(n\omega T) & cos(n\omega T) \\ sin(2\omega T) & cos(2\omega T) & \cdots & sin(2n\omega T) & cos(2n\omega T) \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ sin(m\omega T) & cos(m\omega T) & \cdots & sin(mn\omega T) & cos(mn\omega T) \end{bmatrix}$$

$y(T), y(2T), \ldots, y(mT)$ – sampled values of noisy signal $y(t)$.

Least squares approach

The above formulated problem can be solved using the standard LS approach. According to this approach the energy (objective) function, defined as

$$E_2(\mathbf{x}) = \frac{1}{2}\|\mathbf{e}\|_2^2 = \frac{1}{2}\mathbf{e}^{\mathrm{T}}\mathbf{e}, \quad \text{where } \mathbf{e} = \mathbf{D}\mathbf{x} - \mathbf{y} \tag{3}$$

should be minimized.

In this case the estimated vector **x** will be given by

$$\mathbf{x} = [\mathbf{D}^{\mathrm{T}}\mathbf{D}]^{-1}\mathbf{D}^{\mathrm{T}}\mathbf{y} \tag{4}$$

Formula (4) requires computation of an inverse matrix, which is rather time consuming. Another possibility is to use an on-line identification LS algorithm (Soderstrom and Stoica, 1989). To find the LS estimates in real time, Hopfield-type neural networks (Amari, 1990; Cichocki and Unbehauen, 1992, 1993) can be employed. According to this approach, the system of differential equations

$$\frac{d\mathbf{x}}{dt} = -\mu\mathbf{D}^{\mathrm{T}}(\mathbf{D}\mathbf{x} - \mathbf{y}) = -\mu\mathbf{D}^{\mathrm{T}}\mathbf{e} \tag{5}$$

where $\mu > 0$ is the appropriate learning rate, should be solved. To ensure a fast convergence of the algorithm the learning rate μ was setting in the range $10^3 \div 10^5$.

The matrix differential equation (5) can be written in a scalar form as

$$\frac{da_i}{dt} = -\mu \sum_{k=1}^{m} e_k \sin(ik\omega T) \tag{6}$$

$$\frac{db_i}{dt} = -\mu \sum_{k=1}^{m} e_k \cos(ik\omega T) \tag{7}$$

where

$$e_k = \sum_{i=1}^{n} [a_i \ \sin(ik\omega t) + b_i \ \cos(ik\omega t] - y(kT)$$

$i = 1, 2, \ldots, n$ – number of harmonic components;

$k = 1, 2, ..., m$ – number of samples.

For estimating the amplitudes of the basic components, the above system of differential equations can be implemented by an adaptive analogue network, as shown in Figure 1. The network can be considered as a single neuron with synopses a_1, b_1, which are learned (adjusted, tuned) according to equations (6) and (7). This kind of analogue networks is called in the literature adaptive or optimisation neural networks (Cichocki and Unbehauen, 1992, 1993, 1994; Fa Long and Unbehauen, 1997; Osowski, 1992; Tank and Hopfield, 1986). The network consists of basic computing units: integrators, summers and multipliers. The network shown estimates only the amplitudes a_1 and b_1.

The LS technique is relatively simple. However, the approach is optimal only if matrix **D** is exactly known and the vector **y** is perturbated by a Gaussian noise.

Total least squares approach

The standard LS method assumes that the matrix **D** is exactly determined and only the vector **y** is contaminated by noise. In practice, the matrix **D** is also perturbated by error. In fact, the frequency ω is not exactly known. Moreover, it can slightly fluctuate during the measurement, and these fluctuations are unknown. Furthermore, the sampling period is sometimes not fixed but also fluctuates (i.e. the sampling of the signal is not ideally regular). For these reasons, to a obtain more reliable and robust solution, the total least squares (TLS) approach was applied. The approach is known in the statistics literature as orthogonal regression or errors-in-variable regression (Cichocki and Unbehauen, 1993). The TLS criterion assumes errors both in the matrix **D** and

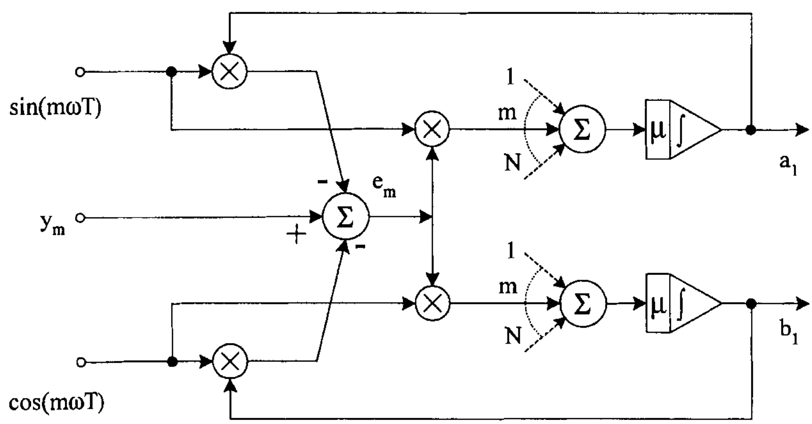

Figure 1. Adaptive neural network for estimating the amplitudes of the basic component according to the least-squares criterion

in the vector $\mathbf{y}$. Therefore, whereas LS minimizes the prediction error, TLS minimizes the error normal to the graph of the linear predictor. Applying the TLS criterion, the following linear matrix equation is obtained (Cichocki and Unbehauen, 1994)

$$(\hat{\mathbf{D}} + \mathbf{R})\mathbf{x}_{TLS} = \hat{\mathbf{y}} + \mathbf{r} \tag{8}$$

where

$$\mathbf{D} = \hat{\mathbf{D}} + \mathbf{R} \text{ and } \mathbf{y} = \hat{\mathbf{y}} + \mathbf{r}$$

$\hat{\mathbf{D}} \in R^{m\times n}, \hat{\mathbf{y}} \in R^m$ are exact but unknown matrices,

$\mathbf{R} \in R^{m\times n}, \mathbf{r} \in R^m$ are corresponding errors.

In other words, the TLS problem can be formulated as the optimization problem: to find the vector $\mathbf{x}*_{TLS}$ that minimizes

$$\|\mathbf{R}\|_F^2 + \|\mathbf{r}\|_F^2 \tag{9}$$

subject to the equality constraints (8), where $\|\mathbf{R}\|_F$ denotes the Frobenius norm of $\mathbf{R}$. The main numerical tool for solving the TLS problem is the singular value decomposition (SVD) of the extended matrix

$$\tilde{\mathbf{D}} = [\mathbf{D}, \mathbf{y}] = \mathbf{U}\Sigma\mathbf{V}^{\mathrm{T}} \tag{10}$$

The TLS solution is computed as (Golub, 1980)

$$\mathbf{x}^*_{TLS} = -\frac{1}{v_{2n+1,2n+1}} \cdot \left[v_{1,2n+1}, v_{2,2n+1}, v_{2n,2n+1}\right]^{\mathrm{T}} \tag{11}$$

where $v_{2n+1} = \left[v_{1,2n+1}, v_{2,2n+1}, \ldots, v_{2n,2n+1}\right]^{\mathrm{T}}$ is the right singular vector associated to the smallest singular value σ_{n+1} of the extended matrix $[\mathbf{D}, \mathbf{y}]$.

As the singular value σ_{n+1} goes to zero, the LS and TLS approach each other. It is important to point out that the LS solution is based on the minimization of the sum of the squared errors (3), while the TLS solution is based on the minimization of the sum of weighted squared errors (Golub and Van Loan, 1980).

$$e(\mathbf{x}) = \frac{\mathbf{D}\mathbf{x} - \mathbf{y}}{[1 + \mathbf{x}^{\mathrm{T}}\mathbf{x}]^{\frac{1}{2}}} \tag{12}$$

In other words, the TLS problem can be formulated as the minimization of the energy function

$$E_{TLS}(\mathbf{x}) = \frac{\|\mathbf{D}\mathbf{x} - \mathbf{y}\|_2^2}{1 + \mathbf{x}^{\mathrm{T}}\mathbf{x}} = \sum_{k=1}^{m} \frac{\left(\sum_{j=1}^{2n} d_{kj}x_j - y_k\right)^2}{1 + \mathbf{x}^{\mathrm{T}}\mathbf{x}} \tag{13}$$

In comparison with the standard LS technique, obtaining the solution of the TLS problem is generally quite burdensome and very time consuming. This is probably because the TLS approach has not been as widely used as the usual LS approach, although the TLS approach was investigated in robust statistics long ago.

In order to simplify the algorithm, an instantaneous error was introduced, defined as

$$e(t) = \mathbf{S}^{\mathrm{T}}(t)(\mathbf{Dx} - \mathbf{y}) = \mathbf{S}^{\mathrm{T}}(t)\mathbf{e} = \sum_{k=1}^{m}\left(\sum_{j=1}^{2n} d_{kj}x_j - y_k\right)S_k(t)$$
$$= \sum_{j=1}^{2n} \tilde{d}_j x_j(t) = \tilde{y}(t) \tag{14}$$

where

$\mathbf{S}(t) = [S_1(t), S_2(t), \ldots, S_m(t)]^{\mathrm{T}}$ is the vector of zero-mean independent identically distributed (i.i.d.) externally excitation signals (e.g. zero-mean noise sources),

$$\tilde{d}_j(t) \stackrel{df}{=} \sum_{k=1}^{m} d_{kj}S_k(t) \tag{15}$$

$$\tilde{y}(t) \stackrel{df}{=} \sum_{k=1}^{m} y_k S_k(t) \tag{16}$$

For more explanation see the Appendix in Cichocki and Unbehauen (1994). The TLS problem can be now reformulated as minimization of the following instantaneous energy function

$$E_{TLS}[\mathbf{x}(t)] = \frac{1}{2}\frac{e^2(t)}{\mathbf{x}^{\mathrm{T}}\mathbf{x} + 1} \tag{17}$$

Applying the gradient descent approach, the system of differential equations is obtained

$$\frac{dx_j(t)}{dt} = -\mu(t)\frac{\partial E_{TLS}[\mathbf{x}(t)]}{\partial x_j} = -\mu(t)e(t)\frac{\tilde{d}_j(t)\left(\mathbf{x}^{\mathrm{T}}\mathbf{x}+1\right) - e(t)x_j(t)}{\left(\mathbf{x}^{\mathrm{T}}\mathbf{x}+1\right)^2} \tag{18}$$

where $\mu(t) > 0$.

The above set of differential equations can be further simplified, after linearization, as

$$\frac{dx_j}{dt} = -\mu(t)e(t)\left[\tilde{d}_j(t) + \tilde{y}(t)x_j(t)\right] \tag{19}$$

A functional block diagram illustrating implementation of the algorithm (19) is shown in Figure 2. The block diagram can be considered as a single neuron with synapses x_j, learned (adjusted) according to equation (19).

Robust tls algorithm (RTLS)

The TLS algorithm is rather sensitive to noise and kind of distributed errors, especially in the presence of outliers.

This sensitivity implies that we need to modify or generalize the TLS algorithm to eliminate, as far as possible, outlying points or large spiky noise. This fact was the main motivation for development and investigation of a new generalized algorithm called Robust Total Least Squares (RTLS) algorithm.

The learning algorithm (19) can be extended as follows:

$$\frac{dx_j(t)}{dt} = -\mu(t)\Psi[e(t)]\left[\tilde{d}_j(t) + \alpha\tilde{y}(t)x_j(t)\right] \tag{20}$$

where $\alpha \geq 0$ is non-negative coefficient, $\Psi(e)$ is non-linear activation function enabling suppression or neglect of large error, e.g.

$$\Psi(e) = tanh(\gamma e) \text{ or } \Psi(e) = \begin{cases} e & \text{for} |e| \leq \beta \\ 0 & \text{otherwise} \end{cases},$$

β is so called cut-off parameter, depending on the problem to be solved.

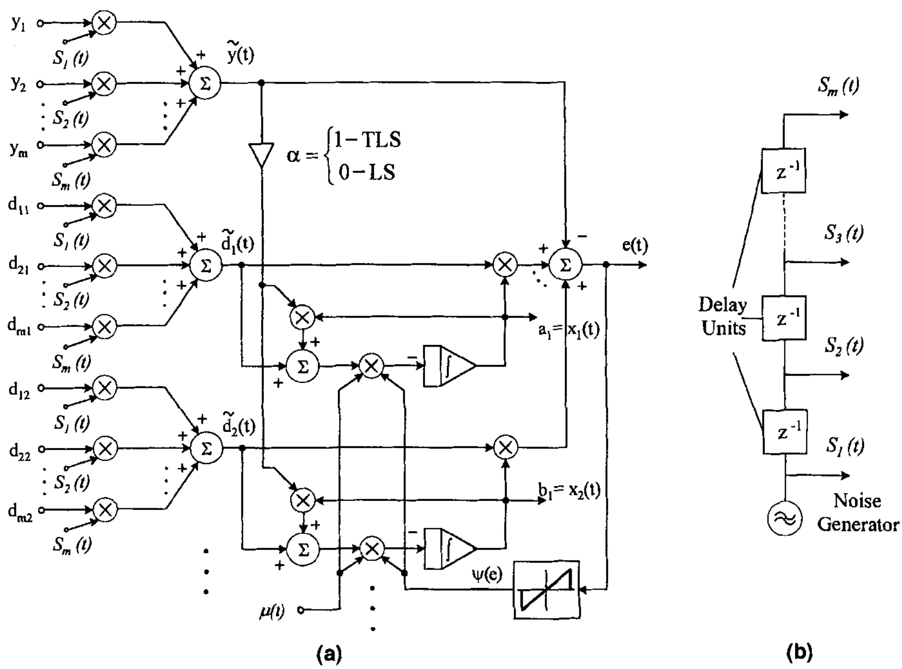

Figure 2.
(a) Adaptive neural network for solving the estimation problem (see equation (20)); (b) exemplary method for generation signals S

In the special case $\Psi(e) \equiv e$ and $\alpha = 0$ we obtain the standard LS algorithm and for $\Psi(e) \equiv e$ and $\alpha = 1$ the standard TLS algorithm. It should be noted, that by changing the value of the parameter α more or less emphasis can be given to errors of the matrix $\mathbf{D}$ with respect to errors of the vector $\mathbf{y}$. If $\alpha = 0$, we assume that error is only in the vector $\mathbf{y}$. On the other hand, for large α (say 100) it can be assumed that the vector $\mathbf{y}$ is almost free of error and the all error lies in the data matrix $\mathbf{D}$. Such an extreme case is referred to as the so-called DLS (data least squares) problem (since error occurs only in $\mathbf{D}$ but not in $\mathbf{y}$).

The RTLS algorithm (20) can be transformed to time-discrete form as

$$x_j(k+1) = x_j(k) - \eta(k)\Psi[e(k)]\left[\tilde{d}_j(k) + \alpha\tilde{y}(k)x_j(k)\right] \tag{21}$$

Simulation experiments

Extensive computer simulation experiments have confirmed the validity and performance of the proposed algorithms. The associated networks were simulated on computer. Owing to limited place, we shall present only some illustrative results.

First, a signal

$$y(t) = \sin(\omega t) + r(t) \tag{22}$$

was simulated. The algorithms were investigated for the frequency of the basic component equal to 50Hz. The investigations were carried out when the frequency of the simulated signal changed. The sampling window was $NT = 0.02$; 0.03; 0.04 and 0.06s. The number of samples $N = 20 \div 100$. When the frequency of the simulated signal was 49Hz and 51Hz the estimation errors were less than 1 per cent. Slightly better results were obtained using the TLS method (Figure 3).

The influence of the random (white) noise in the measured (simulated) signal on the estimation error (Figure 3) was also investigated.

The new LS and especially TLS methods show a great immunity against the noise. It can be stated that the accuracy of the TLS method depends strongly on the sampling frequency and sampling window. It is better for higher sampling

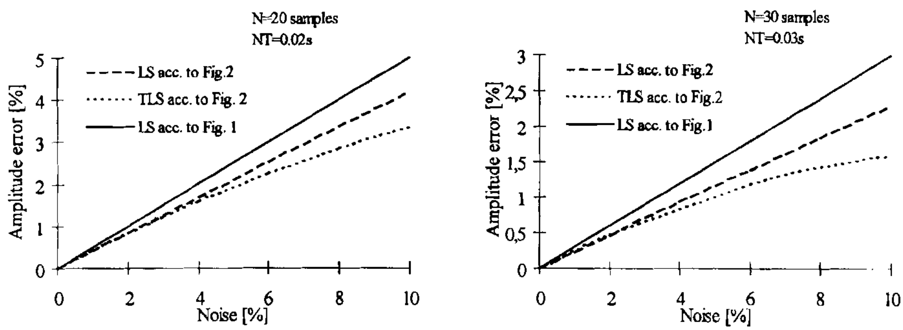

Figure 3.
Errors of the amplitude estimation when using the networks acc. to Figure 1 and acc. to Figure 2

frequency and wider sampling window. In our investigations the most accurate results were obtained for the sampling frequency f_S = 2,500Hz and the sampling window NT = 0.04s (Figure 4).

The investigations were made for the learning rate $\mu(t)$ = 500 and for $\mu(t)$ = 1,500 exp(– 50t). In the second case the estimation errors were smaller. The non-linear activation function $\Psi(e)$ enables suppression of large impulsive errors. When comparing the two networks applying the LS criterion (Figure 3), more accurate results were obtained using the new network (Figure 2). The level of the auxiliary noise $S_n(t)$ was about 1 per cent of the amplitudes of the simulated signal. When using the new LS algorithm the trajectories of the estimated parameter converge in less than 5ms, and for the TLS algorithm in less than 2ms.

Conclusions

Adaptive analogue neural networks represent a very promising approach for high-speed estimation of parameters of signals. In this paper, new algorithms and architectures of neuron-like adaptive circuits were developed, according to the LS and TLS optimization criteria, applying the gradient descent approach. They are more robust against noise in the measured signal than other known LS neural network algorithms. The network based on the TLS criterion optimizes the estimation under the assumption that the signal model can be also perturbated (frequency or sampling interval fluctuation and so forth). The

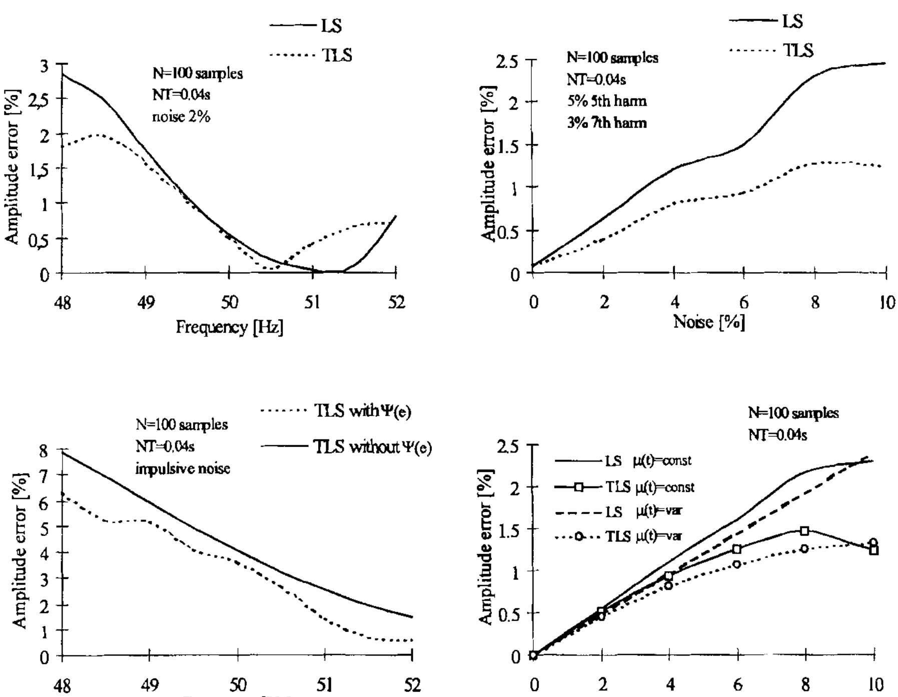

Figure 4.
Errors of the amplitude estimation of the new method (Figure 2)

TLS estimates are better and more reliable than the corresponding LS estimates when applying a higher sampling frequency and a wider sampling window.

Extensive computer simulation experiments confirmed the validity and performance of the proposed algorithms.

References

Amari, S.-I. (1990), "Mathematical foundations of neurocomputing", *Proc. IEEE*, Vol. 78, pp. 1443-63.

Cichocki, A. and Łobos, T. (1994), "Artificial neural networks for real-time estimation of basic waveforms of voltages and currents", *IEEE Trans. on Power Systems*, Vol. 9, pp. 612-18.

Cichocki, A. and Unbehauen, R. (1992), "Neural networks for solving systems of linear equations and related problems", *IEEE Trans. Circ. Syst.*, Vol. 39, pp. 910-23.

Cichocki, A. and Unbehauen, R. (1993), *Neural Networks for Optimization and Signal Processing*, Teubner-Wiley, Stuttgart.

Cichocki, A. and Unbehauen, R. (1994), "Simplified neural networks for solving linear least squares and total least squares problems in real time", *IEEE Trans. on Neural Networks*, Vol. 5 No. 6, pp. 910-23.

Cichocki, A., Kostyla, P., Łobos, T. and Wacławek, Z. (1994), "Neural networks for real-time estimation of signals encountered in power systems", *IMACS Int. Symp. on Signal Processing, Robotics and Neural Networks*, Lille, France, pp. 317-20.

Fa-Long, L. and Unbehauen, R. (1997), *Applied Neural Networks for Signal Processing*, Cambridge University Press, Cambridge.

Golub, G.H. and Van Loan, C.F. (1980), "An analysis of the total least squares problem", *SIAM J. Numer. Anal.*, Vol. 17, pp. 883-93.

Łobos, T. (1989), "Nonrecursive methods for real-time determination of basic waveforms of voltages and currents", *IEE Proc.- C*, Vol. 136, pp. 347-51.

Osowski, S. (1992), "Neural networks for estimation of harmonic components in a power system", *IEE Proc.- C*, Vol. 139, pp. 129-35.

Osowski, S. (1994), "SVD technique for estimation of harmonic components in a power system: a statistical approach", *IEE Proc.- C*, Vol. 141, pp. 473-9.

Soderstrom, T. and Stoica, P. (1989), *System Identification*, Prentice-Hall, London.

Tank, D.W. and Hopfield, J. (1986), "Simple neural optimization networks: an A/D converter, signal decision circuit and a linear programming circuit", *IEEE Transactions on Circuits and Systems*, Vol. 33, pp. 533-41.

Van Den Bos, A. (1988), "Nonlinear least-absolute values and minimax model fitting", *Automatica*, Vol. 24, pp. 803-8.

Widrow, B. and Lehr, M. (1990), "30 years of adaptive neural networks: perceptron, madaline and back propagation", *Proc. IEEE*, Vol. 78, pp. 1415-42.