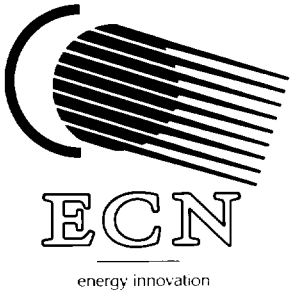
MARCH 1996

ECN-RX--96-011



NEURAL NETWORK TRAINING BY KALMAN FILTERING IN PROCESS SYSTEM MONITORING

Ö. CIFTCIOGLU
E. TÜRKCAN

The Netherlands Energy Research Foundation ECN is the leading institute in the Netherlands for energy research. ECN carries out basic and applied research in the fields of nuclear energy, fossil fuels, renewable energy sources, policy studies, environmental aspects of energy supply and the development and application of new materials.

ECN employs more than 800 staff. Contracts are obtained from the government and from national and foreign organizations and industries.

ECN's research results are published in a number of report series, each series serving a different public, from contractors to the international scientific world.

This RX-series is used for publishing pre-prints or reprints of articles that will be or have been published in a journal, or in conference or symposium proceedings.

Please do not refer to this report but use the reference provided on the title page: 'Submitted for publication to ...' or 'Published in ...'.

Het Energieonderzoek Centrum Nederland (ECN) is het centrale instituut voor onderzoek op energiegebied in Nederland. ECN verricht fundamenteel en toegepast onderzoek op het gebied van kernenergie, fossiele-energiedragers, duurzame energie, beleidsstudies, milieuaspecten van de energievoorziening en de ontwikkeling en toepassing van nieuwe materialen.

Bij ECN zijn ruim 800 medewerkers werkzaam. De opdrachten worden verkregen van de overheid en van organisaties en industrieën uit binnen- en buitenland.

De resultaten van het ECN-onderzoek worden neergelegd in diverse rapportenseries, bestemd voor verschillende doelgroepen, van opdrachtgevers tot de internationale wetenschappelijke wereld.

Deze RX-serie wordt gebruikt voor het uitbrengen van pre-prints of reprints van artikelen die in een tijdschrift of in proceedings van conferenties of symposia (in definitieve vorm) zullen verschijnen of al zijn verschenen.

Gelieve niet te refereren aan het rapportnummer, maar de verwijzing te gebruiken die hiernaast op de titelpagina figureert: 'Voor publikatie aangeboden aan ...' of 'Verschenen in ...'.

Netherlands Energy Research Foundation ECN
P.O. Box 1
NL-1755 ZG Petten
the Netherlands
Telephone : +31 2246 49 49
Fax : +31 2246 44 80

This report is available on remittance of Dfl. 35 to:
ECN, Facility Services,
Petten, the Netherlands
Postbank account No. 3977703.
Please quote the report number.

Energieonderzoek Centrum Nederland
Postbus 1
1755 ZG Petten
Telefoon : (02246) 49 49
Fax : (02246) 44 80

Dit rapport is te verkrijgen door het overmaken van f 35,-- op girorekening 3977703 ten name van:
ECN, Faciliteiten
te Petten
onder vermelding van het rapportnummer.

W-KS/K: KS00219232X

M-TIO(L): **EN** •Neural network training by Kalman filtering in process system monitoring•
M-LA: •EN•
M-FI/K(K): **ED** (FI00101932X) •Netherlands Energy Research Foundation (ECN), Petten (NL)• 1
PH •Netherlands Energy Res. Found., Petten (Netherlands)•
(IN) 4593600
(PH) NENEPE0000
(SI) 006757V
M-FI/K(K): **AF** (FI001342496) •Istanbul Technical Univ. (TR). Faculty of Electrical and Electronic Engineering• 2
PH •Technical Univ. Istanbul (Turkey). Fac. of Electr. & Electron. Eng.•
(PH) TUTE00EL00
M-FI/K(K): **AF** (FI00101932X) •Netherlands Energy Research Foundation (ECN), Petten (NL)• 3
PH •Netherlands Energy Res. Found., Petten (Netherlands)•
(IN) 4593600
(PH) NENEPE0000
(SI) 006757V
M-AU(A): **AU** •Ciftcioglu, Ö.• 2
M-AU(A): **AU** •Türkcan, E.• 3
M-NR(X): **NR** ECN-RX–96-011
M-PDS: 19960300
M-COLL: •13 p.•
M-AV(D): **ST** •Also available from the authors at the Netherlands Energy Research Foundation (ECN), Petten (NL)•
E-ORDST-D/ks: **R** 19970512
E-ORDST-D/ks: **X** 19970513
X-EXSERV/ks: •INIS/NL•
M-FK1/K: FK001290658
M-MTIOFK1(L): **EN** •5. Turkish symposium on artificial intelligence and neural networks (TIANN-5)•
M-MLFK1(O): **TR** Istanbul
M-MDSFK1: 19960627 M-MDEFK1: 19960628

HSN: DE01008014X

(180/191)

H-OC: KA H-TDB: KA EN
H-STC(S): **N** 19970512 H-STDE(V): **V** 19970512
H-REG(R): **CY** NL H-LI: K
H-TYP: R H-LEVELS: M
H-STPR: I
I-CC(F): **IN** E3200
I-CT: 1 E •REACTOR MONITORING SYSTEMS• -; E
I-CT: 2 E •ON-LINE CONTROL SYSTEMS• -; E
I-CT: 3 E •NEURAL NETWORKS• -; E
I-CT: 4 PE •EXPERT SYSTEMS• -; PE
I-CT: 5 PE •TRAINING• -; PE
I-CT: 6 PE •REAL TIME SYSTEMS• -; PE
I-CT: 7 EZ •ALGORITHMS• -; EZ
I-CT: 8 PEZ •STOCHASTIC PROCESSES• -; PEZ
I-CT: 9 E •BORSSELE REACTOR• -; E
I-CTMQ(D): **EN** 1:9
I-CTMQ(D): **EN** 9:2
I-CTMQ(D): **EN** 6:3

I-ABEN(D): **ST**

•Kalman filtering approach for neural network training is described. Its extended form is used as an adaptive filter in a nonlinear environment of the form a feedforward neural network. Kalman filtering approach generally provides fast training as well as avoiding excessive learning which results in enhanced generalization capability. The network is used in a process monitoring application where the inputs are measurement signals. Since the measurement errors are also modelled in Kalman filter the approach yields accurate training with the implication of accurate neural network model representing the input and output relationships in the application. As the process of concern is a dynamic system, the input source of information to neural network is time dependent so that the training algorithm presents an adaptive form for real-time operation for the monitoring task. (orig.)•

W-SELECT(D): **EN** EN/INIS/WS
V-IP: •BKTEM97071• •BKTEM97071•
V-IPD: •19970603• •19970603•
V-TECH: •F4BKO• •19970512•
V-TECHOK: •F4RKD/S• •19970516•
V-INDEX: •F4DGX• •19970711•
V-ABWRITE: •TEM97071• •19970603•
V-ABENOK: •F4RLX/S• •19970623•
V-SAFORMD: •00000000•
V-IOK(D): **ST** QPROD/S 19970711
V-TECHKORR: •F4EN0• •19970711•

Full Information KD-CONF FK001290658

$DG0 0001

H-SN : FK001290658
H-TDB : MN H-OC: KA
H-STDE(V) : V 19970512
H-STC(S) : N 19970512
D-MTIO(L) : The 5. Turkish symposium on artificial intelligence and neural networks (TAINN '96)
D-MDS : 19960627 D-MDE: 19960628 D-ML(O): Istanbul D-MDDISPL:
V-IP : KOPP/KA
V-IPN : KA001850299
V-IPD : 19970512
V-TECH : KOPP/KA
V-TECHD : 19970512

*** E N D O F D O C U M E N T ***

___ <-- H, COL, SI, F, O, P, MN, MDEL, S, L, E, B, Q, U, MU

Anzahl der gefundenen Informationen: 1 1997-05-13 10:54:55

Dt. Abstract Seite: ____ Bearbeiter: DG TIB: ____
Engl. Abstract Seite: 1 Online erledigt: ____ Fernleihe: ____
Short Communic.: ____ Uebersetzer: ____ Out of scope: ____

H-SN: DE01008014X H-LEVELS:M

M-TIO(L): EN Neural network training by Kalman filtering in process system monitoring

H-TDB: KA
H-TDB: EN
H-LI: K____ H-TC(F): ____

I-TIAUG(L): __ ____

I-CC(F): PB ____ I-CCSI(F): PB ____
I-AO: PB ____
E 32.02
~~St A0 P 102~~
NT 970 774

I-MAI(F): __ ____

I-CT:
1 EN reactor monitoring systems
2 " on-line control systems
3 " neural networks
4 " expert systems
5 " training
6 " real time systems
7 " algorithms
8 " stochastic processes
9 " borssele reactor
10 __ ____
11 __ ____
12 __ ____

I-CTMQ(D): EN 1:9 EN 9:2 EN 6:3 __ ____ __ ____
__ ____ __ ____ __ ____ __ ____ __ ____

I-CTLT(D): __ ____ __ ____ __ ____ __ ____ __ ____

MARCH 1996

KS00219232X
R: FI
DE01008014X

ECN-RX--96-011

DE01008014X

NEURAL NETWORK TRAINING BY KALMAN FILTERING IN PROCESS SYSTEM MONITORING

Ö. CIFTCIOGLU
E. TÜRKCAN

To be presented at the TAINN'96, The Fifth Turkish Symposium on Artificial Intelligence and Neural Networks, 27–28 June 1996, Istanbul, Turkey

Prepared by	: E. Türkcan	Initial:
Position	: Author	
Date	:	
Reviewed by (on behalf of co-readers)	: V.A. Wichers	Initial:
Position	: 1st co-reader	
Date	:	
Approved by	: H. Gruppelaar	Initial:
Position	: Manager Nuclear Sciences	
Date	: 16-4-'96	

Neural Network Training by Kalman Filtering in Process System Monitoring

Ö. Ciftcioglu
Istanbul Technical University, Electrical-Electronics Engineering Faculty,
80626 Maslak-Istanbul, Turkey.

and

E. Türkcan
Netherlands Energy Research Foundation ECN,
P.O. Box 1, NL-1755 ZG Petten, The Netherlands.

Abstract

Kalman filtering approach for neural network training is described. Its extended form is used as an adaptive filter in a nonlinear environment of the form a feedforward neural network. Kalman filtering approach generally provides fast training as well as avoiding excessive learning which results in enhanced generalization capability. The network is used in a process monitoring application where the inputs are measurement signals. Since the measurement errors are also modelled in Kalman filter the approach yields accurate training with the implication of accurate neural network model representing the input and output relationships in the application. As the process of concern is a dynamic system, the input source of information to neural network is time dependent so that the training algorithm presents an adaptive form for real-time operation for the monitoring task.

Introduction

Conventionally, standard backpropagation algorithm for feedforward neural network training is rather popular due to its simplicity. For small nets or simple problems it may be justified. However the convergence becomes poor if the problem is relatively complicated requiring increased dimensionality in the network structure. Additionally, in the case information provided to the network is measurement data, the measurements errors add to the gradient noise so that the existing poor convergence properties near a minimum may be aggravated. Therefore an improved training algorithm is of interest for effective, that is, approach to minimum well defined together with avoiding local minima and efficient, that is, fast convergence training.

Training feedforward neural networks can be viewed as an identification problem for a nonlinear distributed dynamic system. In such a system each perceptron is subject to modelling by adaptive Kalman filtering while the nonlinearity introduced by the activation function (usually a sigmoid) is additionally considered. For linear dynamic systems with white process noise and measurement noise, the Kalman [1] algorithm is optimum. For a nonlinear system, the extended version of the Kalman algorithm can be applied by linearizing the system around the current estimate of the parameters. As the measurement noise is already in the model, the algorithm becomes also appealing especially for the case where training information source is measurement data from process sensors. For real-time neural network estimations for the purpose of monitoring for instance, the algorithm is further of interest using the algorithm in the form of recursive estimator [2].

Kalman filtering

By means of Kalman filtering optimal state estimation for a linear dynamic system is performed. Extended versions of the Kalman algorithm can be applied to non-linear dynamic systems by linearizing the system around the current estimate of the

This work is partly funded by the EU for the CEC-Accident Management Support project.

parameters. Although computationally complex, Kalman algorithm update parameters recursively that is consistent with all previously introduced data and usually converges in a few iterations. Derivation of Kalman equations are widely available in the literature [3,4,5]. A brief description of Kalman equations by means of which neural network parameters are computed adaptively, are presented below.

Consider a linear discrete system and assume that modelling technique has produced an adequate description in the form of a linear stochastic system to describe the propagation in time of a state vector $\underline{X}(k)$

$$\underline{X}(k) = \underline{\underline{\phi}}(k, k-1)\underline{X}(k-1) + \underline{\underline{B}}(k)\underline{U}(k) + \underline{\underline{G}}(k)\underline{W}(k) \tag{1}$$

$$\underline{Y}(k) = \underline{\underline{C}}(k)\underline{X}(k) + \underline{V}(k) \tag{2}$$

Here, $\underline{X}(k)$ is an n-vector state process; $\underline{\underline{\phi}}(k, k-1)$ the non-singular $n \times n$ system dynamics matrix; $\underline{\underline{B}}(k)$ the $n \times r$ input matrix; $\underline{U}(k)$ the r-vector deterministic input; $\underline{\underline{G}}(k)$ the $n \times p$ noise input matrix; $\underline{W}(k)$ the p-vector white Gaussian noise process; $\underline{Y}(k)$ the m-vector measurement process; and $\underline{\underline{C}}(k)$ the $n \times m$ measurement matrix; $\underline{V}(k)$ the m-vector white Gaussian measurement noise process. The statistics of the noise process $\underline{W}(k)$ and $\underline{V}(k)$ are assumed to be

$$E[\underline{W}(k)] = 0 \tag{3}$$

$$E[\underline{W}(k_1)\underline{W}(k_2)^T] = \underline{\underline{Q}}\delta_{k_1 k_2} \tag{4}$$

$$E[\underline{V}(k)] = 0 \tag{5}$$

$$E[\underline{V}(k_1)\underline{V}(k_2)^T] = \underline{\underline{R}}\delta_{k_1 k_2} \tag{6}$$

where δ is the Kronecker delta, $\underline{\underline{Q}}(k)$ and $\underline{\underline{R}}(k)$ are $p \times p$ symmetric positive-semidefinite and positive-definite matrices, respectively. The system noise $\underline{W}(k)$ includes the effect of variability in the natural system as well as model structure errors. The measurement noise $\underline{V}(k)$ represents the uncertainty associated with the measurement process. The initial condition $\underline{X}(0)$ is assumed to be Gaussian with statistics:

$$E[\underline{X}(0)] = \hat{\underline{X}}(0) \tag{7}$$

$$E\{[\underline{X}(0) - \hat{\underline{X}}(0)]\, E[\underline{X}(0) - \hat{\underline{X}}(0)]^T\} = \underline{\underline{P}}(0) \tag{8}$$

where $\underline{\underline{P}}(0)$ is the $n \times n$ symmetric positive-definite matrix.

The estimate of the system state $\underline{X}(k)$ can be obtained by the help of the information provided by the system model and the measurements $\underline{Y}(k)$ obtained from the actual system. For the solution of this filtering problem the Bayesian approach is used and the conditional probability density of the state $\underline{X}(k)$, conditioned on the entire history of the measurements, is identified. Once this density is explicitly described, an optimal estimate of the state $\underline{X}(k)$ can be defined. Under the assumption of the model given above, the conditional density is Gaussian and it is completely characterized by its mean and covariance matrix. Hence, the estimate of $\underline{X}(k)$ based on the conditional density will results in the same estimated $\hat{\underline{X}}(k)$ and the same covariance matrix of the estimation error $\underline{\underline{P}}(k)$. The optimal state estimate is propagated from measurement time $k-1$ to measurement time k by the equations

$$\hat{\underline{X}}(k|k-1) = \underline{\underline{\phi}}(k, k-1)\underline{X}(k-1|k-1) + \underline{\underline{B}}(k)\underline{U}(k) \tag{9}$$

$$\begin{aligned} \underline{\underline{P}}(k|k-1) \;=\; & \underline{\underline{\phi}}(k, k-1)\underline{\underline{P}}(k-1|k-1)\underline{\underline{\phi}}(k, k-1)^T + \\ & \underline{\underline{G}}(k)\underline{\underline{Q}}(k)\underline{\underline{G}}(k)^T \end{aligned} \tag{10}$$

At measurement time k, the measurement $\underline{Y}(k)$ becomes available. The estimate is updated by the equations:

$$\hat{\underline{X}}(k|k-1) = \hat{\underline{X}}(k|k-1) + \underline{\underline{K}}(k) \cdot [\underline{Y}(k) - \underline{\underline{C}}(k)\hat{\underline{X}}(k|k-1)] \tag{11}$$

$$\underline{\underline{P}}(k|k) = \underline{\underline{P}}(k|k-1) - \underline{\underline{K}}(k)\underline{\underline{C}}(k)\underline{\underline{P}}(k|k-1) \tag{12}$$

where

$$\underline{\underline{K}}(k) = \underline{\underline{P}}(k|k-1)\underline{\underline{C}}(k)^T \cdot [\underline{\underline{C}}(k)\underline{\underline{P}}(k|k-1)\underline{\underline{C}}(k)^T + \underline{\underline{R}}(k)]^{-1} \tag{13}$$

is the filter gain.

Extended Kalman Filtering

If the linear model is not adequate to represent the process, the system state can be represented by the non-linear stochastic system

$$\underline{X}(k) = \underline{\Phi}[\underline{X}(k-1), k-1, k)] + \underline{\underline{B}}(k)\underline{U}(k) + \underline{\underline{G}}(k)\underline{W}(k) \tag{14}$$

$$\underline{X}(0) = \underline{X}_0 \tag{15}$$

where $\underline{\Phi}[\underline{X}(k-1), k-1, k]$ is the state vector representing the system dynamics. Let the measurements be modelled by the non-linear equation

$$\underline{Y}(k) = \underline{C}[\underline{X}(k), k] + \underline{V}(k) \tag{16}$$

where $\underline{C}[\underline{X}(k), k]$ is a vector describing the relation between the state and the measurements. Now, let us denote a discrete reference state for $\underline{\bar{x}}(k)$. The state equation may be written as:

$$\begin{aligned} \underline{X}(k) \quad = \quad & \underline{\Phi}[\underline{X}(k-1), k-1, k] - \\ & \underline{\Phi}[\underline{\bar{x}}(k-1), k-1, k] + \\ & \underline{\Phi}[\underline{\bar{x}}(k-1), k-1, k] + \\ & \underline{\underline{B}}(k)\underline{U}(k) + \underline{\underline{G}}(k)\underline{W}(k) \end{aligned} \tag{17}$$

and the observation equation as:

$$\underline{Y}(k) = \underline{C}[\underline{X}(k), k] - \underline{C}[\underline{\bar{x}}(k), k] + \underline{C}[\underline{\bar{x}}(k), k] + \underline{V}(k) \tag{18}$$

Considering the small deviations of $\underline{X}(k) - \underline{\bar{x}}(k)$, we can expand $\underline{\Phi}[\underline{\bar{x}}(k-1), k-1, k]$ in the neighbourhood of $\underline{\bar{x}}(k)$ and keep only the linear terms, so that

$$\begin{aligned} & \underline{\Phi}[\underline{X}(k-1), k-1, k] - \underline{\Phi}[\underline{\bar{x}}(k-1), k-1, k] \approx \\ & \quad \underline{\underline{\Phi}}'[\underline{\bar{x}}(k-1), k-1, k][\underline{X}(k-1) - \underline{\bar{x}}(k-1)] \end{aligned} \tag{19}$$

$$\underline{C}[\underline{X}(k), k] - \underline{C}[\underline{\bar{x}}(k), k] \approx \underline{\underline{C}}'[\underline{\bar{x}}(k), k][\underline{X}(k) - \underline{\bar{x}}(k)] \tag{20}$$

where

$$(\underline{\underline{\Phi}}'[\underline{\bar{x}}(k-1),k-1,k])_{ij} = \frac{\partial(\underline{\Phi}[\underline{\bar{x}}(k-1),k-1,k])_i}{\partial(\underline{\bar{x}}(k-1))_j} \tag{21}$$

$$(\underline{\underline{C}}'[\underline{\bar{x}}(k),k])_{ij} = \frac{\partial(\underline{C}[\underline{\bar{x}}(k),k])_i}{\partial(\underline{\bar{x}}(k))_j} \tag{22}$$

are the matrices of partial derivatives. Substitution of the approximate Taylor series expansions in the non-linear vector state equation yields

$$\begin{aligned} \underline{X}(k) \;=\; & \underline{\underline{\Phi}}'[\underline{\bar{x}}(k-1),k-1,k]\underline{X}(k-1) - \\ & \underline{\underline{\Phi}}'[\underline{\bar{x}}(k-1),k-1,k]\underline{\bar{x}}(k-1) + \\ & \underline{\Phi}[\underline{\bar{x}}(k-1),k-1,k] + \\ & \underline{\underline{B}}(k)\underline{U}(k) + \underline{\underline{G}}(k)\underline{W}(k) \end{aligned} \tag{23}$$

and the approximate linear observation equation

$$\begin{aligned} \underline{Y}(k) \;=\; & \underline{\underline{C}}'[\underline{\bar{x}}(k),k]\underline{X}(k) - \underline{\underline{C}}'[\underline{\bar{x}}(k),k]\underline{\bar{x}}(k) + \\ & \underline{C}[\underline{\bar{x}}(k),k] + \underline{V}(k) \end{aligned} \tag{24}$$

Given the linearized model above, the standard Kalman filter can be employed to obtain the estimate of the state $\underline{X}(k)$ and its covariance matrix. In this model, the reference state vector is taken to be

$$\underline{\bar{x}}(k) = \underline{\Phi}[\underline{\bar{x}}(k-1),k-1,k] + \underline{\underline{B}}(k)\underline{U}(k) \tag{25}$$

$$\underline{\bar{x}}(0) = \underline{\hat{X}}_0 \tag{26}$$

so that the reference trajectory is completely determined by the prior estimate of the state. This estimator is known as the linearized Kalman filter.

The basic idea of the extended Kalman filter is to relinearize about each estimate $\underline{\hat{X}}(k|k)$ of the form

$$\begin{aligned} \underline{\bar{x}}(p) = \underline{\Phi}[\underline{\bar{x}}(p-1),p-1,p)] + \underline{\underline{B}}(p)\underline{U}(p) \quad ; \\ p = k+1, k+2 \end{aligned} \tag{27}$$

$$\underline{\bar{x}}(k) = \underline{\hat{X}}(k|k) \tag{28}$$

As soon as a new measurement is available and a new state estimate has been obtained, a new and better reference trajectory is incorporated into the estimation process. With this choice of reference trajectory large initial estimation errors are not allowed to propagate through time and therefore the linearity assumption is less likely to be violated. It is important to note that the extended filter gain, unlike the linearized Kalman filter gain, depends on the measurements and therefore cannot be precomputed.

Neural Network Training by Kalman Filtering

Kalman filtering can be used for adaptive parameter estimation as well as optimal state estimation for a linear dynamic system. In the application of Kalman filtering to neural network training the parameters are the network weights which are considered

as states and the estimation of the states is performed using the information applied at the network 's input. However this information cannot be used directly as the size of the covariance matrix is equal to the square of the number of weights involved. To circumvent the problem we partition the global problem into a number of subproblems each of which is at the level of a single neuron. A generic model for a single neuron is shown in Fig. 1. With the exclusion of the sigmoidal non-linearity part, the model is linear with respect to synaptic weights. In the Kalman filtering model the measurement information is not explicitly known. However, the residual noise is given by

$$\underline{e} = \underline{Y}(k) - \underline{\underline{C}}(k) \cdot \underline{\hat{X}}(k|k-1) \tag{29}$$

and the influence of the single linear subsystem on the global network error can be estimated as

$$g = -\frac{\partial E}{\partial y} \tag{30}$$

where the energy (or cost) function E defined by

$$E = \frac{1}{2} \sum_{n=1}^{N} \sum_{k=1}^{M} [d_n(k) - Y_n(k)]^2 \tag{31}$$

where M is the number of pair of input/output patterns; n the dimension of the networks output; $d(k)$ and $\underline{Y}(k)$ the network's desired and actual outputs, respectively. The representation of a single perceptron is given in Fig. 1 where $f(x)$ is the sigmoidal nonlinearity function called logistic function.

In a neural network with Kalman filtering each perceptron output is estimated by the Kalman estimator. Because of the nonlinearity introduced by the logistic function $f(x)$, extended Kalman filtering algorithm is required. However, since the analytic form of the nonlinearity is known, in place of the linearized filter algorithm, one can linearize the perceptron model so that a considerable saving in computation is gained. The two approaches are nearly equivalent. However, by means of the latter approach the stability characteristics of the estimator is enhanced which yields faster convergence in training. The linearized perceptron model for Kalman estimator is described below. Note that in this approach, each weight connected to the neuron is considered to be a state and estimation turns out to be an adaptive learning.

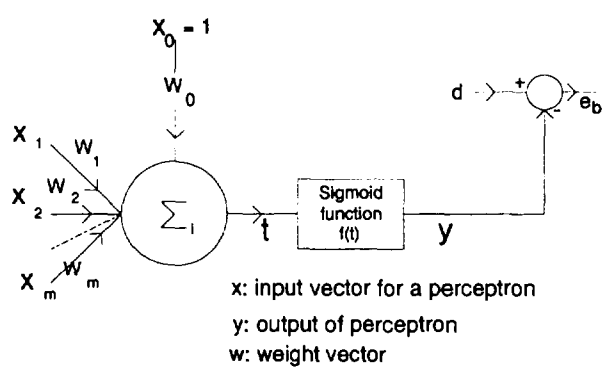


Fig. 1.: *Representation of a perceptron*

Considering the sigmoidal nonlinearity of the form

$$y = f(t) = \frac{1}{1 + e^{-t}} \tag{32}$$

for an arbitrary perceptron, the influence of this single linear subsystem on the global network error takes the form

$$g = \frac{\partial E}{\partial f(t)} \cdot \frac{\partial f(t)}{\partial t} = e_b \, y(1 - y) \tag{33}$$

Above

$$e_b = \frac{\partial E}{\partial y} \tag{34}$$

and g can be let play the role of measurement error, the case being similar to the conventional approach in neural network training. However the measurement error can be estimated differently as follows. We write the measurement equation as

$$d = f(\underline{x}) + e_b \tag{35}$$

Expanding the sigmoidal function in Taylor series and linearizing it, we write

$$f(\underline{x}) = f(\hat{\underline{x}}) + f'(\hat{\underline{x}})(\underline{x} - \hat{\underline{x}}) \tag{36}$$

where

$$f'(\hat{\underline{x}}) = \frac{\partial y}{\partial t} \cdot \frac{\partial t}{\partial \underline{x}} = y(1-y)\underline{w} = \underline{F} \tag{37}$$

so that the measurement equation becomes

$$d = \underline{F}^T \underline{x} + r \tag{38}$$

Above the cap sign indicates the estimate; r is the modified measurement error which is the sum of both modelling and the backpropagated errors:

$$r = f(\hat{\underline{x}}) - \underline{F}^T \hat{\underline{x}} + e_b \tag{39}$$

In the case where we do not allow for any modelling error in the measurement equation, we take

$$f(\hat{\underline{x}}) - \underline{F}^T \hat{\underline{x}} = 0 \tag{40}$$

for each estimate of the state vector $\underline{x}$, this results in

$$r = e_b \tag{41}$$

during the recursive estimation process where the variance estimate of r is equal to the measurement error variance to avoid the excessive learning. Although this approach is not strictly the Extended Kalman Filtering (EKF) algorithm described, it is still the extension of the Kalman algorithm, the extension (linearization) and the algorithm being virtually the same.

From the above equations i.e. Eqs. 37 and 38, it is interesting to note that, at each intermediary, known as hidden layer, each neuron receives its own effective input. In other words, it will have its own copy of covariance matrix $\underline{\underline{P}}$ even the input from the preceding layer is shared with other neurons present at the hidden layer. The price paid for this enhancement over the linear Kalman filtering is the higher computational and storage cost as compared to the former case.

The Kalman filtering approach for neural network training can be viewed as follows. For linear dynamic systems with Gaussian white process and observation noises, the Kalman algorithm is optimum. As opposed to gradient techniques, the Kalman algorithm computes the optimum value of the system parameters as each new data point is seen. Extended versions of Kalman algorithm is applied to nonlinear dynamic system by linearizing the system around the current estimate of the parameters. By doing so the parameters are updated consistently with all previously seen data and usually converges in a few iterations. This reduces the risk of trappings at the local minima as well as excessive learning which are not controlled in gradient-based adaptive approaches. With respect to the latter, it is interesting to note that, in the Kalman filter model the error e_b includes also the measurement noise if the data introduced to the network's input is obtained from a noisy measurement or it is in the form of sensory information coming from a process such as an operating nuclear reactor. Hence, this noise and the modelling errors are accounted in the residual estimation. Therefore, as the algorithm converges, the energy (or cost)

function approaches in the least-mean-square sense to a minimum variance which is approximately equal to the measurement error variance. Due to this, network cannot learn more than a certain limit so that overfitting, that is, learning beyond the measurement errors in the network training is avoided. This is particularly important as the overfitting causes decrease in the generalization capability of the network. Additionally the fast convergence is the most desirable feature in neural network training for stochastic signals where the statistical parameters are estimated both the inputs and the training algorithm being stochastic.

Application

The neural network approach for process monitoring was considered for the Borssele nuclear power plant in the Netherlands as a process system and the neural network subjected to Kalman filtering training was operated using the actual measurement data from the plant. Two data sets are used in the present investigation. Both of them are obtained during normal operation in different times. The first set is used for training the feedforward neural net with one hidden layer. The number of input signals are 12, the number of hidden-layer nodes are 8 and the output nodes are 12, where the network is used in the autoassociative mode. The network topology and the associated signals used are given in Fig. 2. The network having been trained, the performance of the network is tested with the other operational data in which two simulated sensor failures are deliberately introduced into one of the sensory signals among the twelve inputs. One failure is the gradual decrease of sensor response and other one is abrupt disconnection for a short a while. The variation of the modified signal and its estimation by the network is shown in Fig. 3a. The estimation of some other selected signals are shown in Fig. 3b-f. In Fig. 3b-f, the effects of the sensor failure of one sensor on the other ones are presented. From the monitoring view point, these effects are minute and the results indicate the satisfactory performance of the network, concerning the right estimation of the signal in spite of failed relevant sensor.

The Kalman algorithm is designed in an adaptive structure adaptive estimation of process signals where the algorithm accepts the real-time sensory information as inputs to the neural network and real-time recursive estimation is performed. In the case the difference between the measured and the estimated value is within the prescribed limits the real-time adaptive process remains in progress. Otherwise error message is issued and the adaptive process is stopped by equating Kalman gain zero while the real-time process continues without interruption.

Conclusions

In this study two data sets are intentionally selected from an operating nuclear power reactor as a process system to exercise information processing by neural network using measurement data. For measurement data Kalman filtering is especially suitable for recursive estimation, the estimation being optimal. In this respect, the performance of Kalman filtering-based neural network estimation is expected to be superior to the conventional simple approaches as this was demonstrated with the satisfactory results presented in this work. Additionally, as the convergence by Kalman filtering is very fast compared to other existing training approaches, it is also very suitable for real-time neural network applications using stochastic process signals as well as gross process signals themselves as the latter was the case in this study.

References

[1] Kalman, R.E. *A new approach to linear filtering and prediction problems*, J.Basic Eng. Trans ASME, D., Vol. 82. No 1, 1060, pp. 35–45, 1960.

[2] Ciftcioglu Ö. and Türkcan E., *Adaptive Training of Feedforward Neural Networks By Kalman Filtering*, ECN report ECN-R--95-001, The Netherlands, 1995.

[3] Anderson, B.D.O., **Optimal Filtering**, Prentice-Hall, 1979.

[4] Chui, C.K. and Chen G., **Kalman Filtering**, Springer Verlag, 1987.

[5] Gelp A., Ed., **Applied Optimal Estimation**, MIT Press, 1974.

	Sensor	Identification
1.	Steam Generator 1, Feedwater Pressure	SG1FWP (kg/cm^2)
2.	Steam Generator 2, Feedwater Pressure	SG2FWP (kg/cm^2)
3.	Steam Generator 1, Water level	SG1WL (cm)
4.	Steam Generator 2, Water Level	SG2WL (cm)
5.	Temperature difference: Core Exit-Core Inlet	L1,(YQ32-T01)-(YA01-T090) (^oC)
6.	Steam Generator 1, Feedwater Flow	SG1FWF (Ton/h)
7.	Steam Generator 2, Feedwater Flow	SG2FWF (T/h)
8.	Steam Generator 1, Steam Pressure	SG1SP (kg/cm^2)
9.	Steam Generator 2, Steam Pressure	SG2SP (kg/cm^2)
10.	Steam Generator 1, Steam Flow	SG1SF (T/h)
11.	Steam Generator 2, Steam Flow	SG2SF (T/h)
12.	Generated Electric Power	GEP (MWe)

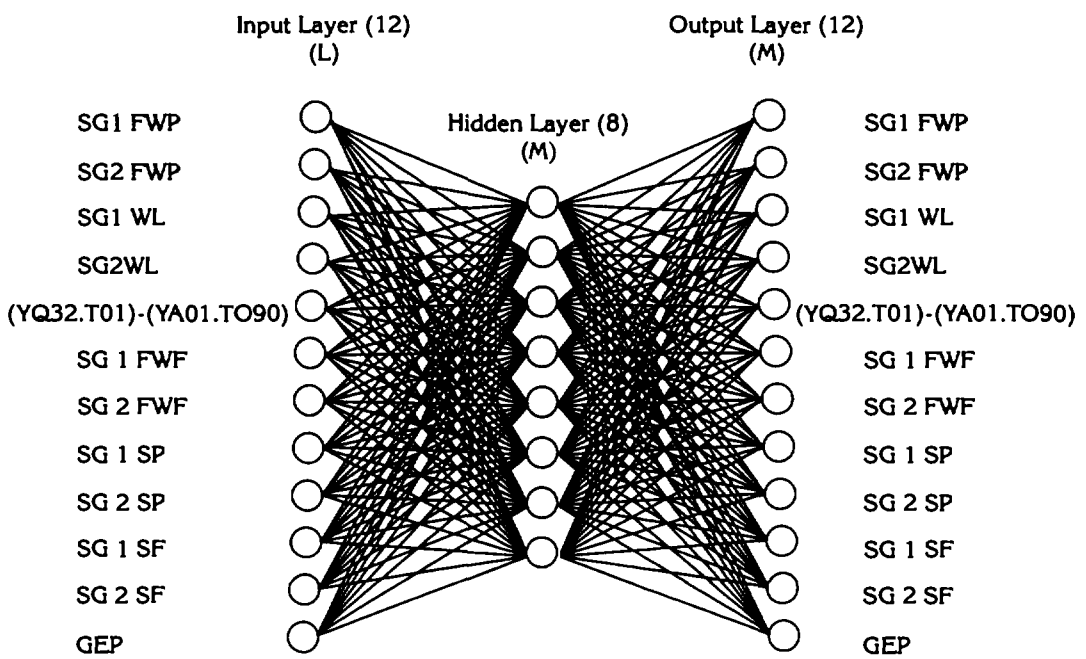


Fig. 2.: *Sensory signals/Identification and Feedforward neural network topology used in this study.*

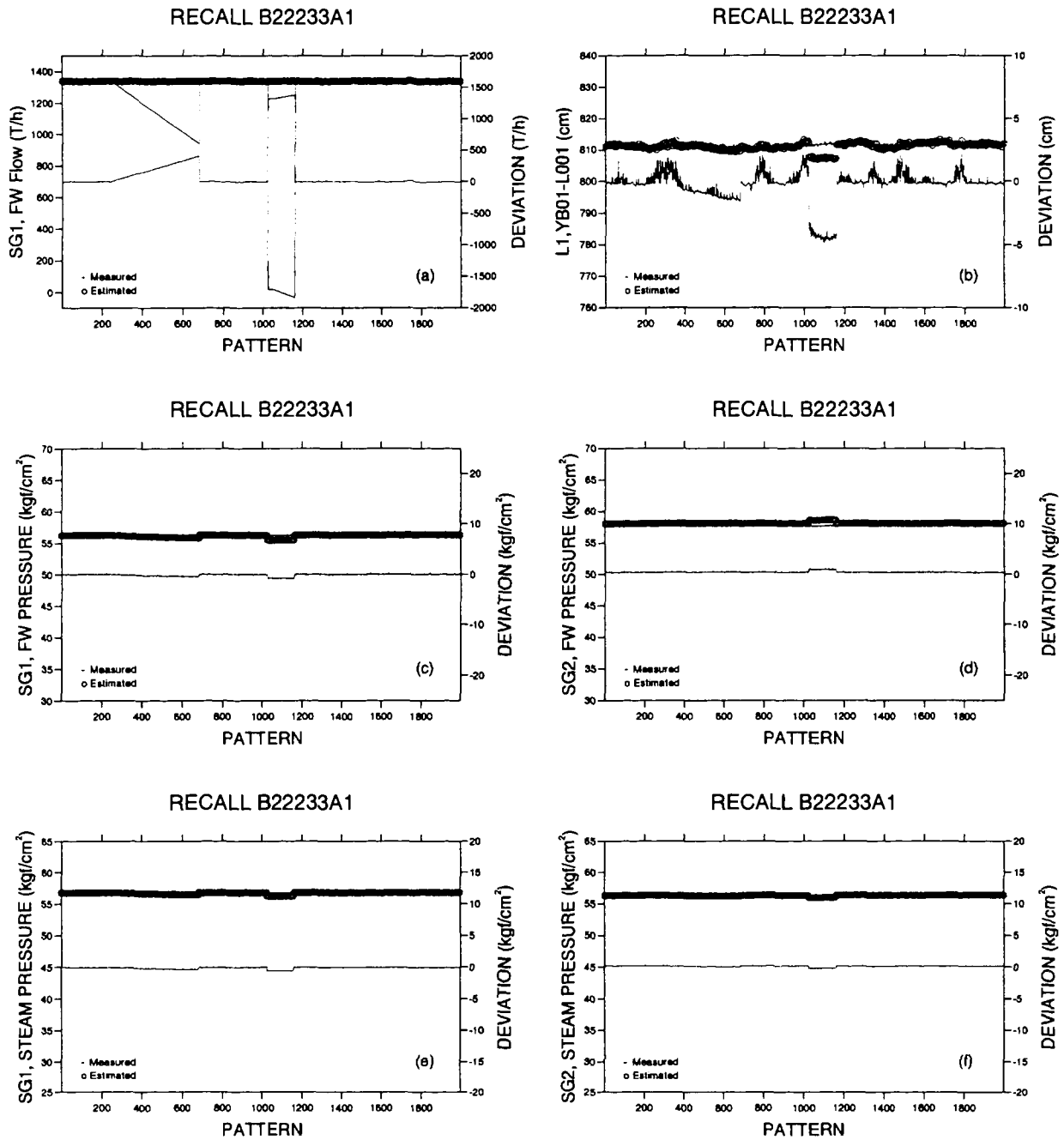


Fig. 3.:

Estimation of process signals by means of trained feedforward neural network. Fig. 3a is the signal where two sensor failures are presented as clearly seen from the magnitude of the errors (right scale). Fig. 3b–f indicate the corresponding estimates and the relevant estimation errors. (Note the correct estimate Fig. 3a of the signal in spite of failed sensor and some minor reflections of this failure on the other signal estimates Fig. 3b–f).

PRELIMINARY CALL FOR PAPERS

TAINN '96

The Fifth Turkish Symposium on Artificial Intelligence and Neural Networks

27-28 June, 1996
İstanbul Technical University, Maçka Campus
İstanbul, Turkey

Organized by

İstanbul Technical University
and
Boğaziçi University

Scope

Theory: Search, Knowledge Representation, Computational Learning Theory, Complexity Theory, Dynamical Systems, Combinatorial Optimization, Function Approximation, Estimation.
Algorithms and Architectures: Learning algorithms, Multilayer Perceptrons, Recurrent Networks, Decision Trees, Evolutionary Algorithms, Fuzzy Logic, Heuristic Search Methods, Symbolic Reasoning.
Applications: Expert Systems, Natural Language Processing, Computer Vision Image Processing, Speech Recognition, Coding and Synthesis, Handwriting Recognition, Time-Series Prediction, Medical Processing, Financial Analysis, Music Processing, Control, Navigation, Path Planning.
Cognitive and Neuro Sciences: Human Learning, Memory, and Language, Perception, Psychophysics, Computational Models.
Implementation: Simulation Tools, Parallel Processing, Analog and Digital VLSI, Neurocomputing Systems.

Paper Submission

Submit three hard-copies of **full papers** (in English or Turkish) limited to 10 pages (in 12 pt size) or **poster papers** limited to 4 pages along with 5 keywords by March31, 1996 to

YZYSA'96/TAINN'96
Department of Computer Engineering
Boğaziçi University
Bebek TR-80815 Istanbul

The symposium will also host *special sessions* on certain subtopics. Proposals by qualified individuals interested in chairing one of these is solicited. The goal is to provide a forum for researchers to better focus on a certain topic and discuss important issues. Individuals proposing have the following responsibilities: (1) Arranging presentations by experts of the topic, (2) Moderating or leading the session, and (3) Writing an overview of the topic and the session for the proceedings. Mail proposals by January 1, 1995 to

YZYSA'96/TAINN'96
Faculty of Electrical and Electronics Engineering
Istanbul Technical University
Maslak TR-80626 Istanbul