Buletinul Ştiinţific al Universităţii "Politehnica" din Timişoara

Seria ELECTRONICĂ şi TELECOMUNICAŢII
TRANSACTIONS on ELECTRONICS and COMMUNICATIONS

Tom 49(63), Fascicola 2, 2004

Gradient Algorithms with Improved Convergence

Cezar Partheniu[1]

Abstract-A generalized normalized gradient descend (GNGD) algorithm for linear finite-impulse response is presented and analized. The GNGD is an extension of the normalized least mean square (NLMS) algorithm by means of an additional gradient adaptive term in the denominator of the learning rate of NLMS. GNGD has better convergence in linear prediction configuration than other algorithms, good performances in system identification configuration in some conditions, worse response in interferences cancelling configuration and similar results with NLMS in reverse modelling configuration.

Keywords: Gradient adaptive learning rate, adaptive filter configuration, generalized normalized gradient descent

I. INTRODUCTION

The generalized normalized gradient descent (GNGD) algorithm is an extension of the normalized least mean square (NLMS) algorithm by means of an additional gradient adaptive term in the denominator of the learning rate of NLMS. GNGD adapts its learning rate according to the dynamics of the input signal with the additional adaptive term compensating for the simplifications in the derivation of NLMS. GNGD is robust to the initialisation of its parameters.

The NLMS is described by the following equations:

$$y[n] = x^H[n]w[n] \tag{1}$$

$$e[n] = d^*[n] - y[n] \tag{2}$$

$$\mu[n] = \frac{\overline{\mu}}{\|x[n]\|^2 + e} \tag{3}$$

$$w[n+1] = w[n] + \mu[n]x[n]e[n] \tag{4}$$

where e[n] is the error of the output signal, d[n] is the desired signal, $x[n]=[x[n-1],\ldots,x[n-n]]^H$ is the input signal vector, N is the filter length, H is the vector transpose operator, w[n] is the filter coefficients vector and μ is the learning rate which defines de convergence speed of the algorithm on the error surface defined with cost function

$$\mathrm{E[n]} = \frac{e^2[n]}{2}, \tag{5}$$

a very important parameter for the LMS algorithm.

The usual independence assumptions leed to a unitar μ for the fastest convergence. Practically the NLMS rate is smaller.

The input signals with unknown and variate dynamics, the "ill-conditioned" self-correlation matrix and the correlation between signals may determinate the divergence or low performances for the NLMS algorithm. As a solution, new algorithms have recently been developed [1], [2], [3]. The Mathews' and Benveniste algorithms are presented in the Appendix. These algorithms are based on the δE[n]/δμ estimators.

A major disadvantage of these algorithms is their sensitivity to the time correlation between the input signal samples and to the value of the additional adaptive rate. To this cause, a generalized normalized gradient descent (GNGD) algorithm has been developed. The stability and the improved convergence are introduced by the gradient adaptive compensation term e from the denominator of the learning rate of NLMS.

Due to noise, "ill-conditioned" correlation matrix, close-to-zero value of the input vector or a large learning rate, the NLMS algorithm (6) is not optimal for many practical settings.

$$w[n+1] = w[n] + \frac{\mu}{\|x[n]\|^2 + \varepsilon} e[n]x[n] = \\ = w[n] + \eta[n]e[n]x[n] \tag{6}$$

[1] Facultatea de Electronica si Telecomunicatii, Catedra de Comunicatii, Bucuresti, Bd.Iuliu Maniu 1-3, phone: 021.331.18.17, e-mail: cezarpart@yahoo.com

To that cause, the ε parameter from (6) is made gradient adaptive as

$$e[n+1] = e[n] - \rho \nabla_{\varepsilon[n-1]} e[n] \qquad (7)$$

Using the chain rule, the gradient $\nabla_{\varepsilon[n-1]} e[n]$ can be evaluated as

$$\frac{\partial E[n]}{\partial \varepsilon[n-1]} = \frac{\partial E[n]}{\partial \varepsilon[n]} \frac{\partial \varepsilon[n]}{\partial y[n]} \frac{\partial y[n]}{\partial w[n]} *$$

$$* \frac{\partial w[n]}{\partial \mu[n-1]} \frac{\partial \mu[n-1]}{\partial \varepsilon[n-1]} = \qquad (8)$$

$$= \frac{e[n]e[n-1]x^H[n]x[n-1]}{(\|x[n-1]\|^2 + \varepsilon[n-1])^2}$$

The GNGD algorithm is therefore described by

$$y[n] = x^H[n]w[n] \qquad (9)$$

$$e[n] = d[n] - y[n] \qquad (10)$$

$$w[n+1] = w[n] + \mu[n]e[n]x[n] \qquad (11)$$

$$w[n] = \frac{\mu}{\|x[n]\|^2 + \varepsilon[n]} \qquad (12)$$

$$e[n] = e[n-1] - \rho \frac{e[n]e[n-1]x^H[n]x[n-1]}{(\|x[n-1]\|^2 + \varepsilon[n-1])^2} \qquad (13)$$

The adaptive rate of GNGD is essentialy bounded by the stability limits of the NLMS algorithm. The compensation term ε is lower bounded according to (14)[4] for μ=1.

$$\varepsilon[n] > -\frac{\|x[n]\|^2}{2} \qquad (14)$$

The complexity of GNGD lies in between the complexity of Mathews' and Benveniste's algorithms and is roughly twice that of NLMS. To reduce the complexity and prevent disturbance in the steady state, it is possible to impose bounds on ε[n] or to stop its adaptation aftrer convergence.

4.Adaptive system configurations

There are four adaptive system configurations defined by the function realized.

System identification (Fig. 1). We want to create a model for an unknown system. This system and the adaptive filter have the same test signal *x*. The output signal of the unknown system is the "desired signal" for the adaptive filter. When *y* and *d* are close, the transfer function of the unknown system is approached with the transfer function of the adaptive filter. The dynamics of the system determine a time variability for the model.

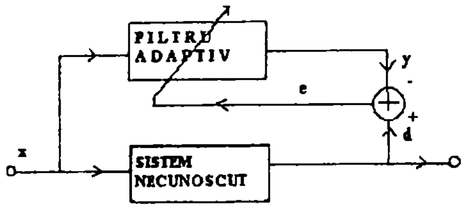

Fig. 1 System identification

Reverse modelling (fig. 2). The model also identifies an unknown system. When the error is zero,the global transfer function of both unknown system and adaptive filter is reduced to a delay. The transfer function of the adaptive filter is the reverse transfer function of the unknown system with a small difference caused by the unavoidable noise. The model can also eliminate the result of an unknown function (eg. Automate equalisation of communications channels).

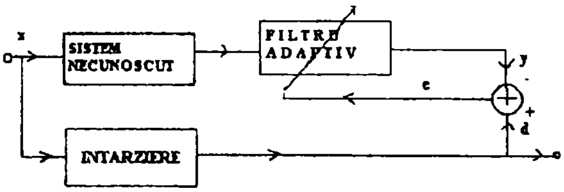

Fig 2 Reverse modelling

Linear prediction (fig. 3). The response of the filter for a delayed input sequence is compared with the actual sample. The error minimisation realise an optimal prediction of the input signal. The 1 output realise the "prediction error filter" and the 2 output realise the prediction.

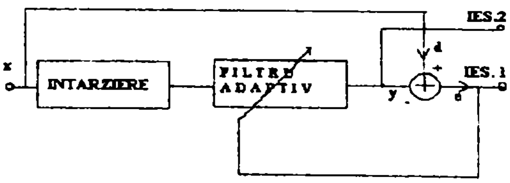

Fig. 3 Linear prediction

Interferences cancelling (fig. 4). The primary signal is the useful signal. It has an unuseful perturbing signal overlapped. There must be created a similar signal which will be substracted from the primary signal using a reference. This signal results from the adaptive filter.

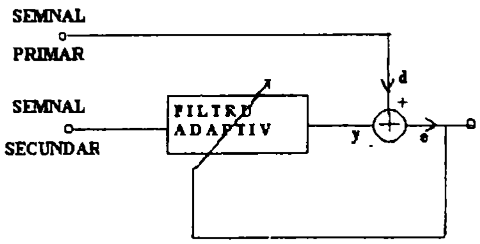

Fig.4 Interference cancelling

II. EXPERIMENTS

The analysis of the adaptive algorithms is made using all the four adaptive systems configurations: linear prediction, system identification, interferences cancelling and reverse modelling.

Linear prediction

The first comparison is made with GNGD and NLMS algorithms. The order of the filter is ord=[illegible] and the input sequence has N=3500 samples. The other parameters of the algorithms have usual values which made possible the comparison.

The experiment is made using a liniar stationary filtered signal given by

$$y[n]=1.79y[n-1]-1.85y[n-2]+ \\ 1.27y[n-3]-0.41y[n-4]+x[n]. \quad (15)$$

where x[n], a white noise with a zero average and unitary variance, is passed through a AR filter.

We observe in Fig.5 that GNGD converges faster than NLMS with 500 iterations. This improved convergence results from the gradient adaptive ε in the denominator of the learning rate of the algorithm.

In [4] we find a comparison between GNGD and Mathews' and Benveniste's algorithms. Using usual values of the parameters, it is evaluated that GNGD has faster convergence than Mathews' and Benveniste's algorithms. This result is shown in Fig.6.

System identification

The parameters of LMS, NLMS and GNGD used in system identification are the number of iterations N, the order of the filter *ord* and the specific parameters δ, λ for RLS, μ and ρ.

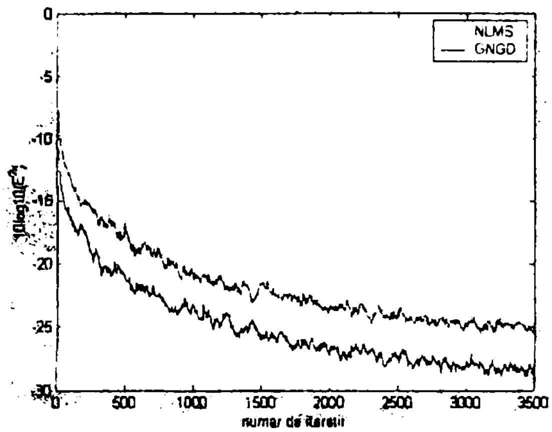

Fig 5 Convergence of GNGD and NLMS algorithms

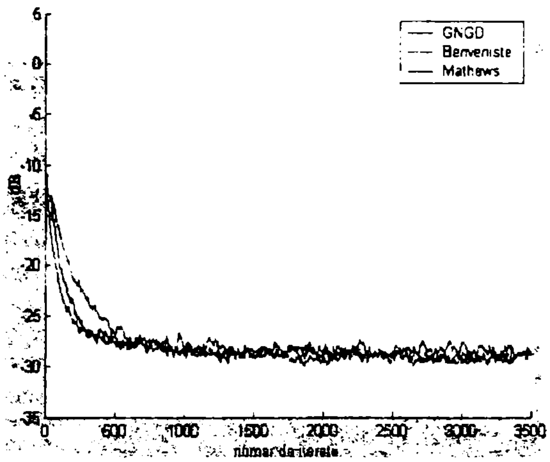

Fig 6. Convergence of GNGD, Mathews' and Benveniste's algorithms

We set N=1000, ord=7, δ=0.001 and λ=0.9. The mean square error for the algorithms is shown in Fig. 7.

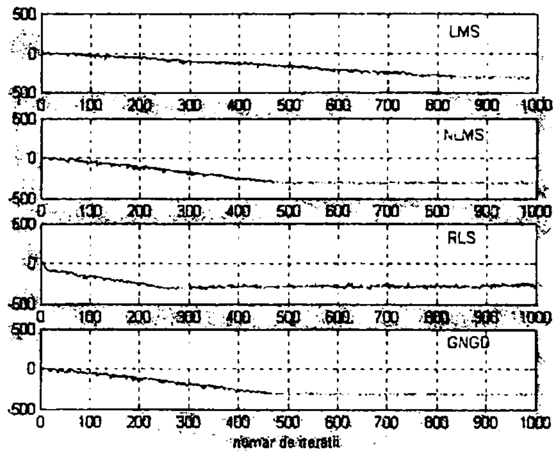

Fig 7 Mean square error in system identification(dB)

The optimization of the parameters results in a faster convergence for GNGD comparing to NLMS.

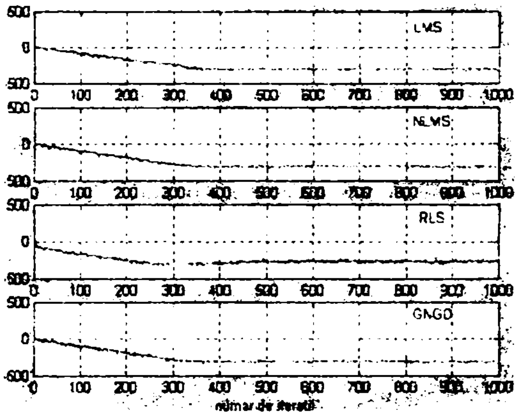

Fig. 8 Mean square error in system identification for optimized parameters (dB)

A value of μ close to 0.1 leads to the same performances for all the algorithms studied (Fig. 9).

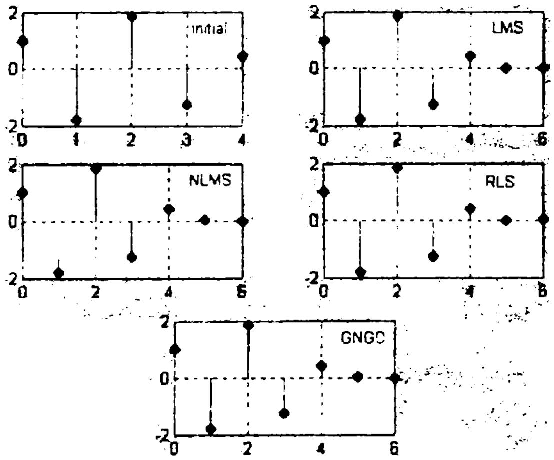

Fig. 9 System identification for μ=0.1

A value of μ less than 0.1 determines a worse response of the adaptive filter for GNGD and NLMS algorithms, the output signal being affected (Fig. 10). GNGD prooves to be sensitive to the values of its parameters.The algorithm has a good output signal for μ close to 0.1 and ρ less than 5.5.

For an optimal set of parameters values GNGD has a response better than NLMS.

Interferences cancelling

The algorithms studied here may be used in interferences cancelling configuration.

We consider a primary signal given by

$$d[n] = \sin(n\omega_0), \omega_0 = 0.05 * \pi \qquad (16)$$

and a perturbation with the following recursive relation

$$v_1[n] = 0.8v_1[n-1] + g[n] \qquad (17)$$

where $g[n]$ is a white noise with zero average and unitary variance. The following signal results

$$x[n] = d[n] + v_1[n] \qquad (18)$$

We also consider a second signal $v_2[n]$ defined as

$$v_2[n] = -0.6v_2[n-1] + g[n] \qquad (19)$$

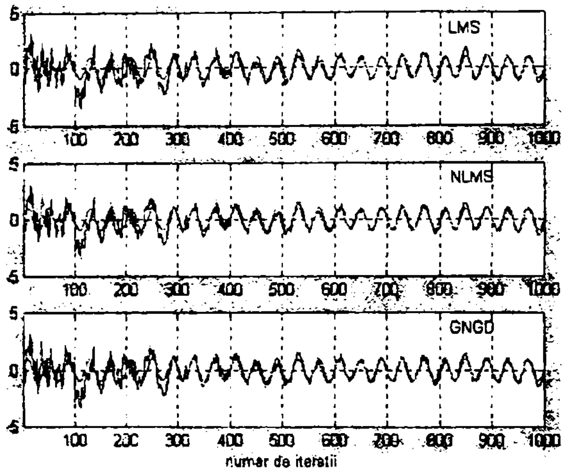

Fig. 10 System identification for μ<0.1

The error for N=1000 is made with 100 runs of independent trials performed and averaged. If we have no noise, we obtain the graphics in Fig. 11.

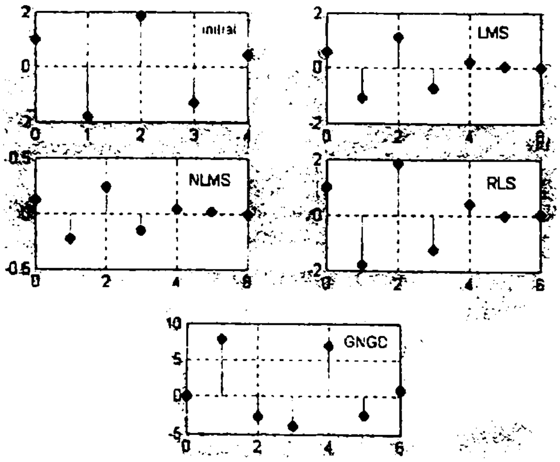

Fig 11. Interferences cancelling (n=0)

A noise factor n=0.5 leeds to an inpossible interferences cancelling for all the algorithms (Fig. 12).

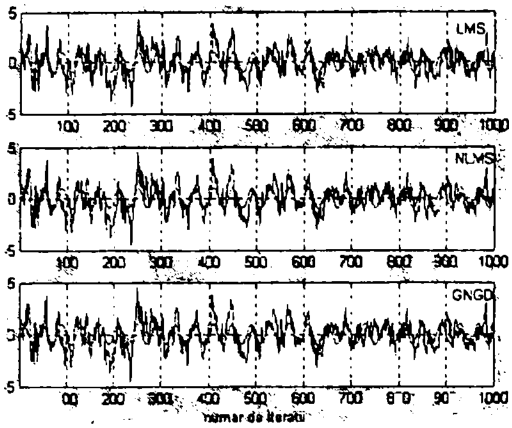

Fig. 12 Interferences cancelling (n=0.5)

A value of μ=0.1 prooves that GNGD and NLMS have a similar performance better than LMS.

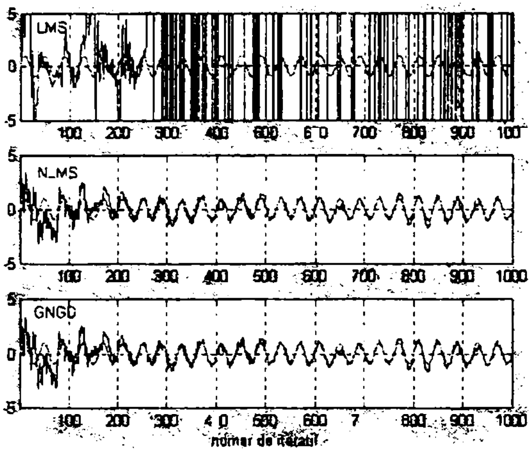

Fig.13 Interfence cancelling

All the algorithms have good results for filter length less than 25.

Reverse modelling

The reverse modelling configuration cancels the results of an unknown transfer function (eg. Automate equalization of communication channels). For this configuration we use an adaptive channel equalizor with a general design as Fig.14 describes.

The input signal has the ±1 values randomly distributed. The channel transfer function $H_c[z]$ is given by (15). The output signal has a white noise overlapped and the adaptive filter realizes the equalization. The switch is used in position "1" with training sequence.

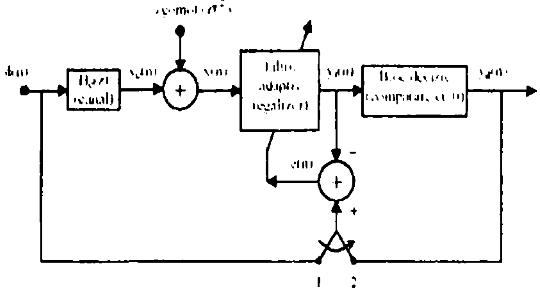

Fig. 14 Channel adaptive equalizor

The length of the input randomly sequence is N=1000 [illegible] filt[illegible] 5. Th[illegible] represent the input data sequence $d[n]$, the channel output sequence $x[n]$ and $y_f[n]$ şi $y_d[n]$ sequences obtained after equalization and decision in histograms for LMS, NLMS and GNGD algorithms. A noise makes the e ualization im ossible.

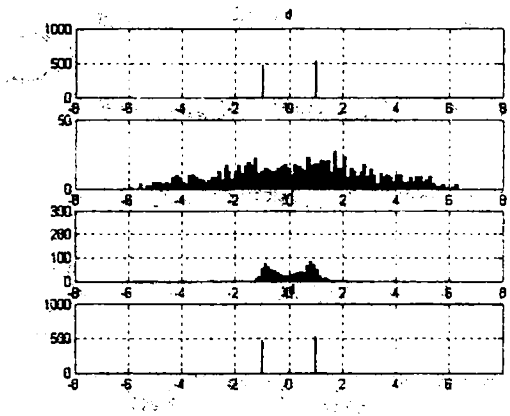

Fig. 15 LMS channel equalization

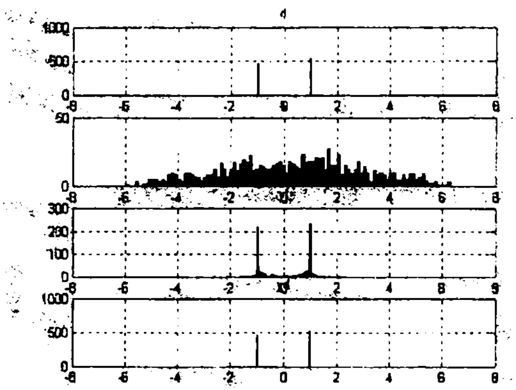

Fig.16 NLMS channel equalization

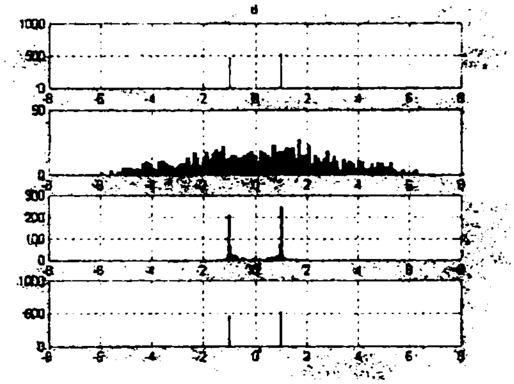

Fig. 17 GNGD channel equalization

As Fig.15-17 shows, NLMS and GNGD have better performances than LMS in reverse modelling.

III. CONCLUSION

The GNGD algorithm has better convergence than the other LMS algorithms studied. GNGD has better performances than NLMS in linear prediction configuration and similar in interferences cancelling design. In system identification and reverse modelling GNGD is sensitive to its parameters, conclusion which makes GNGD unuseful in these situations.

APPENDIX

Mathews' algorithm:

$$y[n] = x^H[n]w[n] \quad (20)$$

$$e[n] = d^*[n] - y[n] \quad (21)$$

$$\mu[n] = \mu[n-1] + \rho e[n]e[n-1] * x^H[n]x[n-1] \quad (22)$$

$$w[n+1] = w[n] + \mu[n]x[n]e[n] \quad (23)$$

Benveniste's algorithm:

$$y[n] = x^H[n]w[n] \quad (24)$$

$$e[n] = d^*[n] - y[n] \quad (25)$$

$$\mu[n] = \mu[n-1] + \rho \mathrm{Re}\{e[n]x^H[n]\Psi[n]\} \quad (26)$$

$$\Psi[n] = [I - \mu[n-1]x[n-1] x^H[n-1]]\Psi[n-1] + e^*[n-1]x[n-1] \quad (27)$$

$$w[n+1] = w[n] + \mu[n]x[n]e[n] \quad (28)$$

REFERENCES

[1] A.M. Kuzminskiy, "Self-adjustement of an adaptive coefficient of a noise compensator in a nonstationary process", *Iszvestiya VUZ Radioelektron.*, vol 29, no.3,pp.103-105,1986

[2] V.J.Mathews and Z.Xie, "A stochastic gradient adaptive filter with gradient adaptive step size", *IEEE Trans.Signal Processing*, vol.41, pp.2075-2087, June 1993

[3] A.Benveniste, M.Metivier andP.Priouret," *Adaptive algorithms and stochastic approximation*", New York: Springer-Verlag, 1990

[4] D.P.Mandic, "A generalized normalizede gradient descent algorithm", *IEEE Signal Processing Letters*, vol.11, no.2, February 2004

[5] C.Partheniu, "*Gradient adaptive algorithms with improved convergence*", diploma project, June 2004