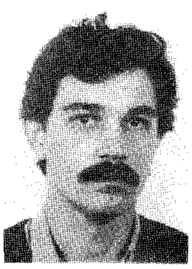
L_k norm adaptive transversal filters

Filtres transversaux adaptatifs avec norme L_k

Aníbal R. FIGUEIRAS-VIDAL

ETSI TELECOMUNICACION-UPM, Ciudad Universitaria, 28040 MADRID, Spain.

Aníbal R. Figueiras-Vidal obtained his Telecommunication Engineer title from ETSIT-UPM in 1973, and his Doctor degree from ETSIT-UPB in 1976. He has been an Assistant and an Associate Professor at ETSIT-UPB from 1973 to 1978, and a Professor at ETSIT-UPM from 1978. His present positions are Head of the Signal Theory and Processing Group and responsible for the creation of ETSIT-USC. He dedicates his research efforts to Digital Signal Processing theory and applications to Radar, Sonar, Image, and Communications. Prof. Figueiras-Vidal has published around 70 international papers on these subjects, and directed more than 20 research programs under industry contract.



José M. PAEZ-BORRALLO

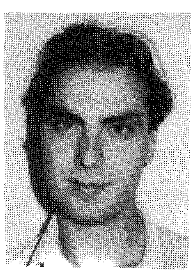
ETSI Telecomunicacion-UPM, Ciudad Universitaria, 28040 MADRID, Spain.

He received the Telecommunication Engineer degree from the ETSI Telecomunicación-UPM, Madrid, Spain in 1982. From 1982 to 1985 he was employed with OTEMA SA (spanish company of AEG-Telefunken) where he was involved with the development of microcontrolled systems for Electronic Warfare. In 1986 he joined Electrónica Ensa SA where he currently works in the development of an automatic classifier for RF-modulated signals. Furthermore, since 1984 he is pursuing the Doctor degree in Telecommunication Engineering at the Department of Signal Theory and Processing of the ETSI Telecomunicación-UPM, Madrid. His research activities include the areas of spectral estimation and adaptive filtering, mainly using optimization criteria alternative to the classical quadratic one.

Ramón GARCIA GOMEZ

ETSI Telecomunicación, Universidad Politecnica de Madrid, Ciudad Universitaria, 28040 MADRID, Spain.

Dr. Ing. Telecommunication. My professional interest is around Digital Signal Processing for Speech Processing, Data Communication, Image Compression and Spectral Analysis. Since 1983 I'm working at the ETSIT, Politechnic University of Madrid, where I teach Linear Systems Theory and Digital Signal Processing.



Luis A. HERNANDEZ-GOMEZ

ETSI Telecomunicación-UPM, Ciudad Universitaria, 28040 MADRID, Spain.

He received the Telecommunication Engineer degree from the ETSI Telecomunicación, Universidad Politécnica de Madrid, Spain in 1982. In 1984 he joined the Department of Signal Theory and Processing, from the same University, where he is now Associate Professor. Since 1985 he is pursuing the Doctor degree in Telecommunication Engineering. His current research interest include speech coding, signal processing, and the development of software for signal processing.

SUMMARY

A whole family of L_k norm adaptive transversal filters is introduced and analyzed, in the context of plant identification, under hypotheses that are validated by simulation results. The analysis allows to establish general convergence conditions and to compare the performance of the elements of the family from the point of view of their speed of convergence-degree of convergence (final residual error variance) compromise; the results of these comparisons depend on the plant noise distribution characteristics. The deterministic optimization of the adaption step is also formulated and evaluated by means of simulation. Finally, open research lines in this area are indicated.

KEY WORDS

L_k-norm, adaptive, identification, plant noise, convergence, optimization.

RÉSUMÉ

On introduit et analyse une famille complète de filtres transversaux adaptatifs avec norme L_k *dans le contexte de l'identification de systèmes (plant identification), sous des hypothèses qui sont validées par les résultats des simulations. L'analyse permet d'établir des conditions générales de convergence et comparer la prestation (performance) des éléments de la famille du point de vue du compromis entre la vitesse de convergence et le degré de convergence (variance de l'erreur résiduale finale); les résultats de cette comparaison dépendent des caractéristiques de la distribution du bruit du système. L'optimization déterministique de l'incrément d'adaptation est aussi formulée et évaluée au moyen des simulations. Finalement, on présente des lignes ouvertes de recherche.*

MOTS CLÉS

Norme L_k*, adaptatif, identification, bruit de système, convergence, optimisation.*

1. Introduction

The three most usual adaptive techniques employed in control and signal processing are recursive least squares, Kalman, and the family of stochastic gradient algorithms: many texts on modern control and [2] illustrate this fact.

Among these possibilities, stochastic gradient algorithms offer simplicity, robustness, and good performances (whenever the input signal does not possess a large eigenvalue spread). The LMS algorithms is particularly interesting [8, 9, 3], since it emphasizes these advantages. Then, any study about similar procedures will be useful in both control and signal processing.

In fact, the LMS algorithm can be considered as an element of a family which uses the instantaneous estimate of the gradient of a general error cost function to change the coefficients of a transversal filter to decrease the error; here, we will analyze those corresponding to the k-th power of the absolute error, i. e., an L_k norm family [7, 4, 5] because they are simple and they serve to build a first approach to other cost functions.

To keep our approach as general as possible, we will refer to a widely applicable scheme for control and signal processing applications: the plant identification depicted in Figure 1. We assume the plant and the adaptive filters being equal length (N) FIRs, and we use $x(n)$, $u(n)$, $\varepsilon(n)$, to indicate the input, plant noise, and identification error sequences, respectively.

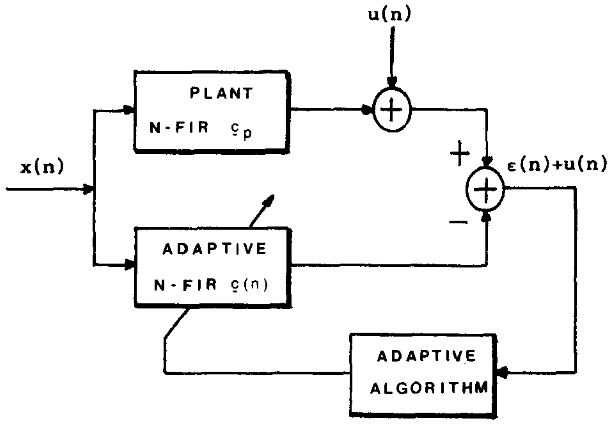


Fig. 1. — Plant identification. $x(n)$**: input process;** $u(n)$**: plant noise;** $\varepsilon(n)$**: residual error.**

Section 2 will be dedicated to establish the adaption equations of these L_k norm adaptive algorithms and to analyze the evolution of the identification error variance under some usual hypotheses and considering noise-free and representative plant noise cases. The conditions to guarantee convergence are bounds for the adaptation constant that we discuss in Section 3. A comparative analysis of the speed of convergence is presented in Section 4. Section 5 proposes and evaluates a deterministic optimization of the adaption constant, along with a first discussion of its stochastic optimization. Finally, the main conclusions of this study are summarized.

2. The algorithms and their error variance analysis

As the LMS does, the general algorithms we are discussing try to minimize $\mathrm{E}[|\varepsilon(n)+u(n)|^k]$ by using the local estimate

$$\hat{\nabla}\mathrm{E}\{[|\varepsilon(n)+u(n)|^k]\}=\nabla[|\varepsilon(n)+u(n)|^k];$$

then, since

$$(1)\quad \begin{aligned}\nabla[|\varepsilon(n)+u(n)|^k] &= \nabla\{|[\mathbf{c}_P-\mathbf{c}(n)]^T\mathbf{x}(n)+u(n)|^k\}\\ &= -k[\varepsilon(n)+u(n)]\,|\varepsilon(n)+u(n)|^{k-2}\,\mathbf{x}(n)\end{aligned}$$

the L_k algorithm is

$$(2)\quad \begin{aligned}\mathbf{c}(n+1)=\mathbf{c}(n)+\mu k[\varepsilon(n)+u(n)]\\ \times|\varepsilon(n)+u(n)|^{k-2}\,\mathbf{x}(n)\end{aligned}$$

To analyze the behaviour of (2) without excess of mathematical difficulties we will assume the following usual hypotheses:

(a) $x(n)$ and $u(n)$ are white, zero mean, mutually independent, wide sense stationary random processes, having variances σ_x^2 and σ_u^2, respectively;

(b) $\mathbf{c}(n)$ and $\mathbf{x}(n)$ are (approximately) independent random vectors;

(c) the convergence is enough to consider that $\mathbf{x}(n)$ and $\varepsilon(n)$ are approximately uncorrelated: we accept this approximate uncorrelation as independence (to calculate means of products as products of means).

(*b*) and (*c*) are, in fact, false hypotheses; but (*b*) is nearly true if μ is small enough, and (*c*), although theoretically weak when the convergence error is high, behaves well enough even in these cases (*see* simulations).

Now, we will obtain a general recursive formula for the mean square identification error $E[\varepsilon^2(n)]$, that equals $\sigma_\varepsilon^2(n)$, the variance of $\varepsilon(n)$, since $x(n)$ [and thus $\varepsilon(n)$] has zero mean.

Substracting both terms of (2) from $\mathbf{c}_P$, and multiplying term by term the transpose of the result by this result, we have, using $\mathbf{v}(n)=\mathbf{c}_p-\mathbf{c}(n)$

$$(3)\quad \mathbf{v}^T(n+1)\mathbf{v}(n+1) = \mathbf{v}^T(n)\mathbf{v}(n)-2\mu k[\varepsilon(n)+u(n)]|\varepsilon(n)+u(n)|^{k-2} \times \mathbf{v}^T(n)\mathbf{x}(n)+\mu^2 k^2|\varepsilon(n)+u(n)|^{2k-2}\mathbf{x}^T(n)\mathbf{x}(n)$$

taking expectations and considering that, using hypothesis (*a*) and (*b*)

$$(4a)\quad \sigma_\varepsilon^2(n)=E\{[\mathbf{v}^T(n)\mathbf{x}(n)]^2\} = E\left\{\left[\sum_{i=0}^{N-1} v_i(n)x(n-i)\right]^2\right\} = \sigma_x^2\sum_{i=0}^{N-1}E[v_i^2(n)]=\sigma_x^2 E[\mathbf{v}^T(n)\mathbf{v}(n)]$$

and also that, under hypothesis (*a*),

$$(4b)\quad E[\mathbf{x}^T(n)\mathbf{x}(n)]=E\left[\sum_{i=0}^{N-1}x^2(n-i)\right]=N\sigma_x^2$$

and, finally, regarding hypothesis (*b*) and (*c*) and multiplying both terms by σ_x^2, we arrive to

$$(5)\quad \sigma_\varepsilon^2(n+1)=\sigma_\varepsilon^2(n)-2\mu k\sigma_x^2 \times E\{\varepsilon(n)[\varepsilon(n)+u(n)]|\varepsilon(n)+u(n)|^{k-2}\} +\mu^2k^2N\sigma_x^4E[|\varepsilon(n)+u(n)|^{2k-2}]$$

that can we written in the form

$$(6)\quad \sigma_\varepsilon^2(n+1)=\sigma_\varepsilon^2(n)[1-\mu S_1(n)+\mu^2 S_2(n)]$$

where

$$(7a)\quad S_1(n)=2k\sigma_x^2E\{\varepsilon(n)[\varepsilon(n)+u(n)] \times|\varepsilon(n)+u(n)|^{k-2}\}/\sigma_\varepsilon^2(n)$$

$$(7b)\quad S_2(n)=k^2N\sigma_x^4E[|\varepsilon(n)+u(n)|^{2k-2}]/\sigma_\varepsilon^2(n)$$

Now, for simplicity, we develop these formulas under two different assumptions.

2.1. Noise-free plant case

We accept a Gaussian $x(n)$, and, then, a Gaussian $\varepsilon(n)$. Thus, since [6]

$$(8a)\quad E[\varepsilon^{2m}(n)]=(2m-1)!!\,\sigma_\varepsilon^{2m}(n)$$

$$(8b)\quad E[|\varepsilon(n)|^{2m-1}]=\sqrt{2/\pi}\,2^{m-1}(m-1)!\,\sigma_\varepsilon^{2m-1}(n)$$

where $(2m-1)!!$ equals $1.3.5...(2m-1)$. Making $u(n)=0$ in (7*a*), (7*b*) and using (8*a*), (8*b*), we arrive to

$$\begin{array}{ll}(9a) & \\ (9b) & \end{array} S_1(n)=\begin{cases}2k(k-1)!!\,\sigma_x^2\sigma_\varepsilon^{k-2}(n), & \text{even } k\\ 2\sqrt{2/\pi}\,k(k-1)!!\,\sigma_x^2\sigma_\varepsilon^{k-2}(n), & \text{odd } k\end{cases}$$

$$\begin{array}{ll}(10a) & \\ (10b) & \end{array} S_2(n)=\begin{cases}N\sigma_x^4/\sigma_\varepsilon^2(n), & k=1\\ Nk^2(2k-3)!!\,\sigma_x^4\sigma_\varepsilon^{2(k-2)}(n), & k>1\end{cases}$$

These formulas are also good approximations for large N.

2.2. Noisy plant cases

Now, we will assume again Gaussianity for $x(n)$ ($\varepsilon(n)$), and we will accept independence between $\varepsilon(n)$ and $u(n)$ [the same comments that those corresponding to general hypothesis (*c*) can be applied here].

To allow easier calculations, it is convenient to write

$$\begin{array}{ll}(11a) & \\ (11b) & \end{array} S_1(n)=\begin{cases}[2k\sigma_x^2/\sigma_\varepsilon^2(n)]\displaystyle\sum_{j=0}^{k-1}\binom{k-1}{j} \times E[\varepsilon^{j+1}(n)]E[u^{k-1-j}(n)], & \text{even } k\\ [2k\sigma_x^2/\sigma_\varepsilon^2(n)]\displaystyle\sum_{j=0}^{k-1}\binom{k-1}{j}\times\varphi_j(n), & \text{odd } k\end{cases}$$

where

$$(12)\quad \varphi_j(n)=E\{\varepsilon^{j+1}(n)u^{k-1-j}(n)\operatorname{sgn}[\varepsilon(n)+u(n)]\}$$

and

$$(13)\quad S_2(n)=[k^2N\sigma_x^4/\sigma_\varepsilon^2(n)] \times\sum_{j=0}^{2k-2}\binom{2k-2}{j}E[\varepsilon^j(n)]E[u^{2k-2-j}(n)]$$

To compute the above formulas, it is necessary to assume statistical distributions for u, independently of n, since we have established a stationary plant noise [$\varepsilon(n)$ has been assumed Gaussian, and it has a variance function of n, $\sigma_\varepsilon^2(n)$]. It is classical in these cases to explore distributions having long, intermediate, and short tails, since this characteristic is the essential to get representative results. Here we will use exponential, Gaussian, and uniform distributions for u; remarking that the qualitative results can be extended in the form the above reasoning indicates. The respective probability density functions are

$$(14a)\quad f_1(u)=(1/\sqrt{2}\,\sigma_u)\exp(-\sqrt{2}|u|/\sigma_u)$$

$$(14b)\quad f_2(u)=(1/\sqrt{2\pi}\sigma_u)\exp(-u^2/2\sigma_u^2)$$

$$(14c)\quad f_3=\begin{cases}1/2\sqrt{3}\,\sigma_u, & |u|\leqq\sqrt{3}\,\sigma_u\\ 0, & |u|>\sqrt{3}\,\sigma_u\end{cases}$$

To calculate (11*a*) and (13) under (14*a*), (14*b*), (14*c*) is easy. To solve (12), and consequently (11*b*), we write it in the form

$$(15a)\quad \varphi_j(n)=\int_{-\infty}^{\infty}\varepsilon^{j+1}(n)g_j[\varepsilon(n)]f[\varepsilon(n)]\,d\varepsilon(n)$$

where

$$(15b)\quad g_j[\varepsilon(n)]=\int_{-\infty}^{\infty} u^{k-1-j}\,\mathrm{sgn}[\varepsilon(n)+u]\,f(u)\,du = \int_{-\infty}^{-\varepsilon(n)} u^{k-1-j} f(u)\,du+\int_{-\varepsilon(n)}^{\infty} u^{k-1-j} f(u)\,du$$

Since we are assuming even densities for u, we have finally

$$(15'b),\ (15''b)\quad g_j[\varepsilon(n)]=\begin{cases} 2\,\mathrm{sgn}[\varepsilon(n)] \times \int_0^{|\varepsilon(n)|} u^{k-1-j} f(u)\,du, & \text{even } k-1-j \\ 2\int_{|\varepsilon(n)|}^{\infty} u^{k-1-j} f(u)\,du, & \text{odd } k-1-j \end{cases}$$

$(15'b)$, $(15''b)$ have been solved for $k=1, 3$, and the plant noise distributions $(14a)$, $(14b)$, $(14c)$. Introducing also the results of $(11a)$ for $k=2$, 4, and (13) for $k=1, 2, 3, 4$, under the same distributions, we have the following results for the four lower order algorithms

$\Big[$we use $s^2=\sigma_\varepsilon^2(n)/\sigma_u^2$, and

$$\mathrm{erf}(.)=(2/\pi)\int_0^{\cdot}\exp(-t^2)\,dt\Big]:$$

$k=1$

$$(16a),(16b),(16c)\quad S_1(n)=\begin{cases} (2\sqrt{2}\,\sigma_x^2/\sigma_u)\,[1-\mathrm{erf}(s)]\,e^{s^2}, & \text{exponential } u \\ 2\sqrt{2/\pi}\,\sigma_x^2/\sigma_u\sqrt{1+s^2}, & \text{Gaussian } u \\ 2\sigma_x^2\,\mathrm{erf}(\sqrt{3/2s^2})/\sqrt{3}\,\sigma_u, & \text{uniform } u \end{cases}$$

$$(17)\quad S_2(n)=N\,\sigma_x^4/\sigma_u^2 s^2$$

$k=2$

$$(18)\quad S_1(n)=4\,\sigma_x^2$$

$$(19)\quad S_2(n)=4N\,\sigma_x^4(1+s^2)/s^2$$

$k=3$

$$(20a),(20b),(20c)\quad S_1(n)=\begin{cases} 6\sqrt{2}\,\sigma_x^2\sigma_u\,\{2s/\sqrt{\pi}+[1-\mathrm{erf}(s)]\,e^{s^2}\}, & \text{exponential } u \\ 6\sqrt{2/\pi}\,\sigma_x^2\sigma_u \times[1+2s^2+s^2/(1+s^2)]/\sqrt{1+s^2}, & \text{Gaussian } u \\ 6\sigma_x^2\sigma_u[(\sqrt{3}+s^2/\sqrt{3}) \times\mathrm{erf}(\sqrt{3/2s^2})+\sqrt{2/\pi}\,e^{-3s^2/2}], & \text{uniform } u \end{cases}$$

$$(21a),(21b),(21c)\quad S_2(n)=\begin{cases} 27N\,\sigma_x^4\sigma_u^2(2+2s^2+s^4)/s^2, & \text{exponential } u \\ 27N\,\sigma_x^4\sigma_u^2(1+s^2)^2/s^2, & \text{Gaussian } u \\ 27N\,\sigma_x^4\sigma_u^2(3/5+2s^2+s^4)/s^2, & \text{uniform } u \end{cases}$$

$k=4$

$$(22)\quad S_1(n)=24\,\sigma_x^2\sigma_u^2(1+s^2)$$

$$(23a),(23b),(23c)\quad S_2(n)=\begin{cases} 240N\,\sigma_x^4\sigma_u^4(6+6s^2+3s^4+s^6)/s^2, & \text{exponential } u \\ 240N\,\sigma_x^4\sigma_u^4(1+s^2)^3/s^2, & \text{Gaussian } u \\ 48N\,\sigma_x^4\sigma_u^4 \times(9/5+15s^2+15s^4+5s^6)/s^2, & \text{uniform } u \end{cases}$$

Using $(16a)$ to $(23c)$ [along with $(9a)$, $(9b)$, and $(10a)$, $(10b)$] in (6) serves to evaluate the dynamic behaviour of the identification error variance $\sigma_\varepsilon^2(n)$ as a function of the adaption parameter μ, the plant noise, σ_x^2 and σ_u^2 (we will check the adequation of the theoretical expressions to simulated cases in a further Section). Thus, these results can be employed as aids for particular case design and evaluation. But we are interested in a general comparison; then, we will start by discussing the convergence conditions.

3. Convergence conditions

Introducing the quadratic polynomial in μ

$$(24)\quad P(\mu; N, \sigma_x^2, \sigma_u^2, \sigma_\varepsilon^2(n))=1-\mu S_1(n)+\mu^2 S_2(n)$$

(6) becomes

$$(25)\quad \sigma_\varepsilon^2(n+1)=\sigma_\varepsilon^2(n)\,P(\mu; N, \sigma_x^2, \sigma_u^2, \sigma_\varepsilon^2(n))$$

and it is clear that the convergence conditions depend on the characteristics of P in the range of $\sigma_\varepsilon^2(n)$ included in $[\sigma_\varepsilon^2(0), \sigma_\varepsilon^2(\infty)]$.

Before discussing these conditions, we must remark that P is a positive polynomial. Considering

$$(26)\quad \beta=\frac{S_1^2(n)}{4S_2(n)} = \frac{E^2\{\varepsilon(n)\,\mathrm{sgn}[\varepsilon(n)+u]\,|\varepsilon(n)+u|^{k-1}\}}{NE[\varepsilon^2(n)]\,E[|\varepsilon(n)+u|^{2k-2}]}$$

where we have used $(7a)$, $(7b)$ and $E[\varepsilon^2(n)]$ for $\sigma_\varepsilon^2(n)$, and since $|\varepsilon(n)|\geqq\varepsilon(n)\,sgn[\varepsilon(n)+u)]$ in any case, and putting $\varepsilon^2(m)=|\varepsilon(m)|^2$

$$(27)\quad \beta\leqq\frac{E^2[|\varepsilon(n)|\;|\varepsilon(n)+u|^{k-1}]}{NE[|\varepsilon(n)|^2]\,E[|\varepsilon(n)+u|^{2k-2}]}$$

according to the Cauchy-Schwartz inequality $(E^2[XY]\leqq E[X^2]\,E[Y^2])$, we obtain $\beta\leqq 1/N$; if $N\geqq 2$,

$\beta<1$; thus, the discriminant of P is negative; therefore P is a positive polynomial, since $S_2(n)$ is always positive, *see* (7 *b*).

Since $P>0$, $(0<)\,P<1$ ($\forall n$) is the convergence condition for the sequence of error variances. This condition is equivalent to

$$(28)\qquad (0<)\,\mu<\min_n\,[S_1(n)/S_2(n)]$$

To make easier the discussion of (28), we will divide again our study into noise-free and noisy plant cases.

3.1. NOISE-FREE PLANT CASE

We will use here $C_k(n)=S_1(n)/S_2(n)$ for the k-th order algorithm. Starting from (9 *a*), (9 *b*), (10 *a*), (10 *b*), it is trivial to obtain

$$C_k=\begin{cases} 2\sqrt{2/\pi}\,\sigma_\varepsilon(n)/N\,\sigma_x^2, & k=1 \quad (29\,a)\\ 2(k-1)!!/N\,k\,(2k-3)!!\,\sigma_x^2\,\sigma_\varepsilon^{k-2}(n), & \text{even } k \quad (29\,b)\\ 2\sqrt{2/\pi}\,(k-1)!!/N\,k\,(2k-3)!!\,\sigma_x^2\,\sigma_\varepsilon^{k-2}(n), & \text{odd } k>1 \quad (29\,c)\end{cases}$$

and, from these expressions, convergence condition (28) in the segment $[0,\infty)$ becomes

$$(30\,a)\qquad \mu_1<C_1(\infty)\qquad (k=1)$$

$$(30\,b)\qquad \mu_2<C_2=1/N\,\sigma_x^2\qquad (k=2)$$

$$(30\,c)\qquad \mu_k<C_k(0)\qquad (k>2)$$

Note the apperance of the well-know $1/N\,\sigma_x^2$ constant bound when $k=2$. For $k>2$, the convergence depends on the initial conditions. When $k=1$, we will see that the variance always converges to a value determined by the step we use.

If we decide to implement a (deterministically) variable step size (to speed up the convergence), we should not surpass locally $C_k(n)$.

Assuming the convergence bounds are verified, we have a variable $P_k(n)$ *if* $k\neq2$, and a constant $P_2<1$ if $k=2$; then, the mean square error convergence rate is constant if $k=2$ and variable if $k\neq2$. Since $\lim_{n\to\infty} P_k(n)\to1$ as we will see at the end of this Section, the asymptotic convergence rate of the non-quadratic algorithms is zero: i. e., there is a "saturation" effect for these algorithms. This does not exclude the possibility of local convergences faster then the $k=2$ one (in fact, they occur frequently in the first iterations), but the saturation effect is a drawback, in particular when the starting point is good enough.

We will discuss now the asymptotic variance value under convergence conditions.

3.1.1. *Asymptotic variance in* $k>2$ *cases*

If we accept $\sigma_\varepsilon^2(\infty)>0$, from (6)

$$(31)\qquad \sigma_\varepsilon^2(\infty)=\sigma_\varepsilon^2(\infty)\,[1-\mu\,S_1(\infty)+\mu^2\,S_2(\infty)]$$

therefore, using (29 *b*), (29 *c*)

$$(32)\qquad \mu=S_1(\infty)/S_2(\infty)=g(k)/\sigma_\varepsilon^{k-2}(\infty)$$

since $\sigma_\varepsilon^{k-2}(\infty)<\sigma_\varepsilon^{k-2}(0)$, (32) is not compatible with (30 *c*); hence, $\sigma_\varepsilon^2(\infty)=0$.

3.1.2. *Asymptotic variance in* $k=2$ *case*

Since $P_2(<1)$ is a constant, $\sigma_\varepsilon^2(n+1)=P_2\,\sigma_\varepsilon^2(n)$; then

$$(33)\qquad \sigma_\varepsilon^2(n+1)=P_2^n\,\sigma_\varepsilon^2(0)$$

from which

$$(34)\qquad \lim_{n\to\infty}\sigma_\varepsilon^2(n+1)=\sigma_\varepsilon^2(0)\lim_{n\to\infty}P_2=0$$

3.1.3. *Asymptotic variance in* $k=1$ *case*

Let μ be

$$(35)\qquad \mu=(2\sqrt{2/\pi}/N\,\sigma_x^2)\,T=[S_1(n)/S_2(n)]\,T/\sigma_\varepsilon(n)$$

with $0<T^2<\sigma_\varepsilon^2(0)$. This is equivalent to a bound

$$(36)\qquad \mu<(2\sqrt{2/\pi}/N\,\sigma_x^2)\,\sigma_\varepsilon(0)$$

Introducing (35) into (6), we have the first equation

$$(37)\qquad \begin{aligned}\sigma_\varepsilon^2(1)&=\sigma_\varepsilon^2(0)\,[1-TS_1^2(0)/\sigma_\varepsilon(0)\,S_2(0)\\&\quad+T^2\,S_1^2(0)/\sigma_\varepsilon^2(0)\,S_2(0)]\\&=\sigma_\varepsilon^2(0)-[\sigma_\varepsilon(0)-T]\,TS_1^2(0)/S_2(0)<\sigma_\varepsilon^2(0)\end{aligned}$$

The last inequality being due to the bound of T. Substracting T^2 from both sides

$$(38)\qquad \begin{aligned}&\sigma_\varepsilon^2(1)-T^2\\&=\sigma_\varepsilon^2(0)-T^2-[\sigma_\varepsilon(0)-T]\,TS_1^2(0)/S_2(0)\\&=[\sigma_\varepsilon(0)-T]\,[\sigma_\varepsilon(0)+T-TS_1^2(0)/S_2(0)]\end{aligned}$$

but

$$(39)\qquad S_1^2(0)/S_2(0)=(8\,\sigma_x^4/\pi)/N\,\sigma_x^4=8/N\,\pi$$

therefore, for $N\geqq2$, $S_1^2(0)/S_2(0)<2$; and according to it, we can obtain from (38)

$$(40)\qquad \sigma_\varepsilon^2(1)-T^2>[\sigma_\varepsilon(0)-T]^2>0$$

i. e., $T^2<\sigma_\varepsilon^2(1)$. Repeating this argument,

$$(41)\qquad T^2<\sigma_\varepsilon^2(n+1)<\sigma_\varepsilon^2(n)$$

and it becomes clear that $\sigma_\varepsilon^2(n)$ is a monotonic decreasing sequence limited by T^2 as long as we choose μ as given by (35).

After this, we can verify that $\lim_{n\to\infty} P_k(n)\to1$, as previously stated.

If $k>2$, since $\sigma_\varepsilon^2(\infty)=0$, and $S_1(n)$ is given by (9 *a*), (9 *b*), $S_2(n)$ corresponds to (10 *b*), it is clear that $\lim_{n\to\infty} S_1(n)\to0$, $\lim_{n\to\infty} S_2(n)\to0$, and, then according to (6), the above limit for $P_k(n)$ holds.

When $k=1$, according to (41), $\sigma_\varepsilon^2(\infty)=T^2$, and $\mu=[S_1(\infty)/S_2(\infty)]\,T/\sigma_\varepsilon^2(\infty)$; substituting into (6) with $n \to \infty$, we obtain

$$(43) \quad 1 - S_1^2(\infty)\,T/S_2(\infty)\,\sigma_\varepsilon(\infty) + S_1^2(\infty)\,T^2/S_2(\infty)\,\sigma_\varepsilon^2(\infty) = 1$$

considering the equality between $\sigma_\varepsilon^2(\infty)$ and T^2.

3.2. Noisy plant cases

First at all, we will remark that, when the plant noise is not zero, the identification error variance will not tend to zero if a finite μ is used. Considering (7 *a*), (7 *b*), it is clear that, when $\sigma_\varepsilon^2(n) \to 0$, i. e., $\varepsilon(n) \to 0$

$$(44\,a) \quad \lim_{\sigma_\varepsilon^2(n) \to 0} \sigma_\varepsilon^2(n)\,S_1(n) \to 0$$

$$(44\,b) \quad \lim_{\sigma_\varepsilon^2(n) \to 0} \sigma_\varepsilon^2(n)\,S_2(n) \to k^2 N \sigma_x^4 E[u^{2k-2}(n)]$$

then, (6) is not compatible under $\sigma_\varepsilon^2(\infty) \to 0$, unless $\mu \to 0$.

Now, we will consider only the $k \mp 1, 2, 3$, and 4 cases assuming exponential, Gaussian, and uniform distributions for u. Using the previously calculated values for $S_1(n)$ and $S_2(n)$ we can compute $C_k[s(n)] = S_1(n)/S_2(n)$, where $s^2(n) = \sigma_\varepsilon^2(n)/\sigma_u^2$, and discuss condition (28).

3.2.1. $k=1$

It is easy to find

$$\begin{aligned} &(45\,a) \\ &(45\,b) \quad C_1[s(n)] = \\ &(45\,b) \end{aligned} \begin{cases} K_1 s^2 [1 - \operatorname{erf}(s)]\, e^{s^2}, & \text{exponential } u \\ K_2 s^2/\sqrt{1+s^2}, & \text{Gaussian } u \\ K_3 s^2 \operatorname{erf}(\sqrt{3/2 s^2}), & \text{uniform } u \end{cases}$$

These three functions grow monotically with s.

If we choose $\mu = C_1(s_0)$, we can write (6) in the form

$$(46) \quad \sigma_\varepsilon^2(n+1) = \sigma_\varepsilon^2(n)[1 - C_1(s_0)\,S_2(n) \times \{C_1[s(n)] - C_1(s_0)\}]$$

and, assuming convergence, this is reached when $C_1[s(\infty)] = C_1[s_0]$; therefore, the steady-state corresponds to

$$(47) \quad \sigma_\varepsilon^2(\infty) = s_0^2\,\sigma_u^2 = \grave{e}C_1^{-1}(\mu)]^2\,\sigma_u^2$$

The convergence is guaranteed because $C_1[s(n)] > C_1(s_0] = \mu$ for any finite n.

3.2.2. $k=2$

Here, we have

$$(48) \quad C_2[s(n)] = K'\,s^2/(1+s^2)$$

independently of the distribution of u. This function also increases monotonically with s; hence; we can repeat the discussion of the previous $k=1$ case. Note that, in both $k=1$ and $k=2$ cases, the selection of the adaption constant fixes the final convergence.

3.2.3. $k=3$ *and* $k=4$

In these cases, $C_3[s(n)]$ and $C_4[s(n)]$ have not monotonic character; they offer a maximum in:

$k=3$, exponential u: $s \simeq 4$ dB, Gaussian u: $s \simeq 4$ dB, uniform u: $s \simeq 2$ dB,
$k=4$, exponential u: $s \simeq 4$ dB, Gaussian u: $s \simeq 0$ dB, uniform u: $s \simeq -2$ dB

and they decrease from this maximum to zero when $s \to 0$ or ∞. Their forms are represented in Figure 2 ($k=3$) and 3 ($k=4$) between $s = 20$ dB and $s = -30$ dB for Gaussian u (the other cases show equivalent forms).

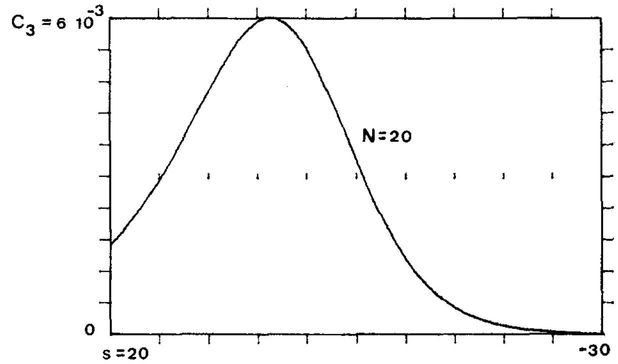


Fig. 2. — $C_3(s)$. **Gaussian noise** ($\sigma_x^2 = \sigma_u^2 = 1$).

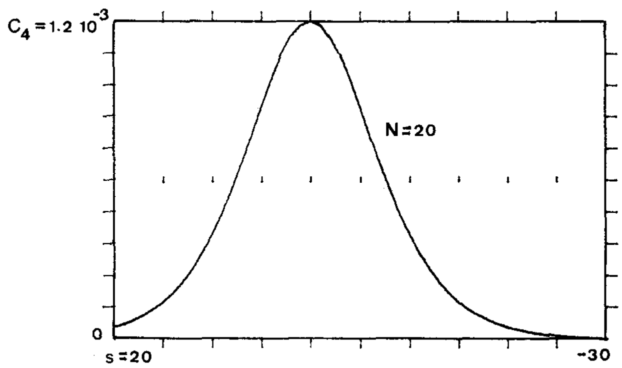


Fig. 3. — $C_4(s)$. **Gaussian noise** ($\sigma_x^2 = \sigma_u^2 = 1$).

These forms force a local examination of each particular initial and final conditions, since the $C_k[s(n)]$ can reach their smallest value at one of these points.

If we select $\mu = C_k[s_0]$ and $C_k[s_0] < C[s(0)]$, we can offer the same discussion as previously; when the initial condition is such that $C_k[s(0)] < C_k[s_0]$, thesystem will diverge from its initial condition. If we select $\mu = C_k[s(0)]$ to avoid it, we will have the equation (6) taking the form

$$(49) \quad \sigma_\varepsilon^2(n+1) = \sigma_\varepsilon^2(n)\{1 - C_k[s(0)]\,S_2(n) \times \{C_k[s(n)] - C_k[s(0)]\}\}$$

that forces the steady state $s(\infty) = s(0)$ and

$$(50) \quad \sigma_\varepsilon^2(\infty) = s^2(0)\,\sigma_u^2 = \grave{e}C_k^{-1}(\mu)]^2\,\sigma_u^2 < s_0^2\,\sigma_u$$

the process being not finished at s_0, to stop here, it is necessary to change μ to $\mu' = C_k[s_0]$ after arriving to $n = n_0$ such that $C_k[s(n_0)] > C_k[s_0]$.

In any case, note that situations having $C_k[s(0)] < C_k[s_0]$ are not usual in the practice.

Note that, on the other hand, we have the possibility of reducing the final identification error variance as much as desired by reducing the adaption constant.

4. Discussing the relative performances

To compare the speed of convergence versus the degree of convergence of the different algorithms, we introduce the parameter [1]

$$(51)\quad \lambda[\sigma_\varepsilon(n)] = -10\log[\sigma_\varepsilon^2(n+1)/\sigma_\varepsilon^2(n)]$$
$$= -10\log \mathrm{P}(\mu;\ \mathrm{N},\ \sigma_x^2,\ \sigma_u^2,\ \sigma_\varepsilon^2(n))$$

that measures the reduction of the identification error variance at each iteration.

As (51) shows, λ depends on several parameters. Since $\mu = \mathrm{S}_1(\infty)/\mathrm{S}_2(\infty)$ is proportional to σ_x^2, we will avoid the dependence on σ_x^2 by normalising it to 1; then, everything depends only on μ, s_0, N, the ratio $\sigma_\varepsilon^2(n)/\sigma_u^2$, and the plant noise distribution.
We will select a representative case, in which the final convergence is forced to $s_0 = -30\,\mathrm{dB}$, starting from $s(0) = 20\,\mathrm{dB}$, and we will discuss the speeds of convergence by looking the $\lambda[\sigma_\varepsilon(n)]$ vs $s(n) = \sigma_\varepsilon(n)/\sigma_u$ curves for $k = 1$, 2, 3, and 4, under exponential, Gaussian, and uniform plant noises, and using N as a parameter: N = 10, 20 and 40 (to check the tendency with the plant length).

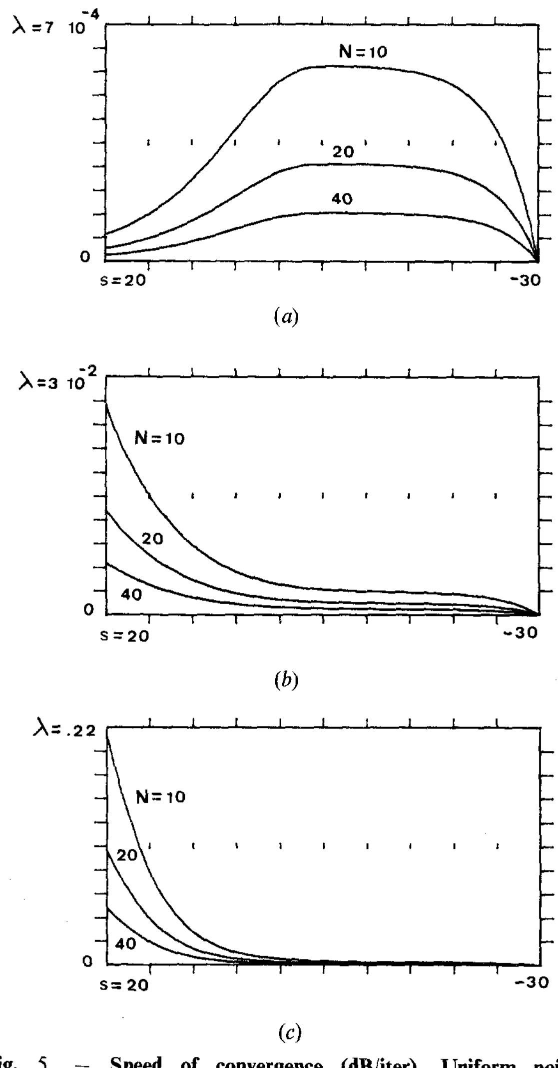


Fig. 5. — Speed of convergence (dB/iter). Uniform noise ($\sigma_x^2 = \sigma_u^2 = 1$). *(a)* L_1 algorithm. *(b)* L_3 algorithm. *(c)* L_4 algorithm.

Figure 4 shows the L_2 case, that is independent of the plant noise distribution. Note that the speed of convergence remains constant until the plant noise becomes important [$s(n) \simeq -15\,\mathrm{dB}$ in our case], decreasing after this.

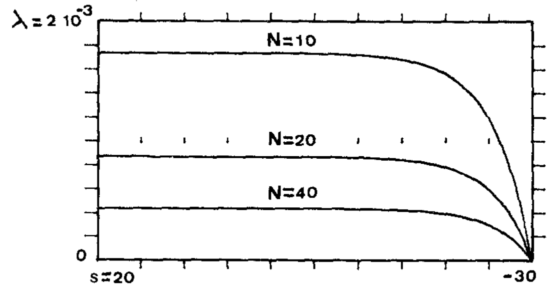


Fig. 4. — Speed of convergence for L_2^* (dB/iter., $\sigma_x^2 = \sigma_u^2 = 1$).

Figures 5 *a*, *b*, *c*, show the behaviours of L_1, L_3 and L_4 schemes, respectively, under uniform plant noise; Figures 6 *a*, *b*, *c*, are equivalent assuming Gaussianity, and Figures 7 *a* *b*, *c*, correspond to exponential plant

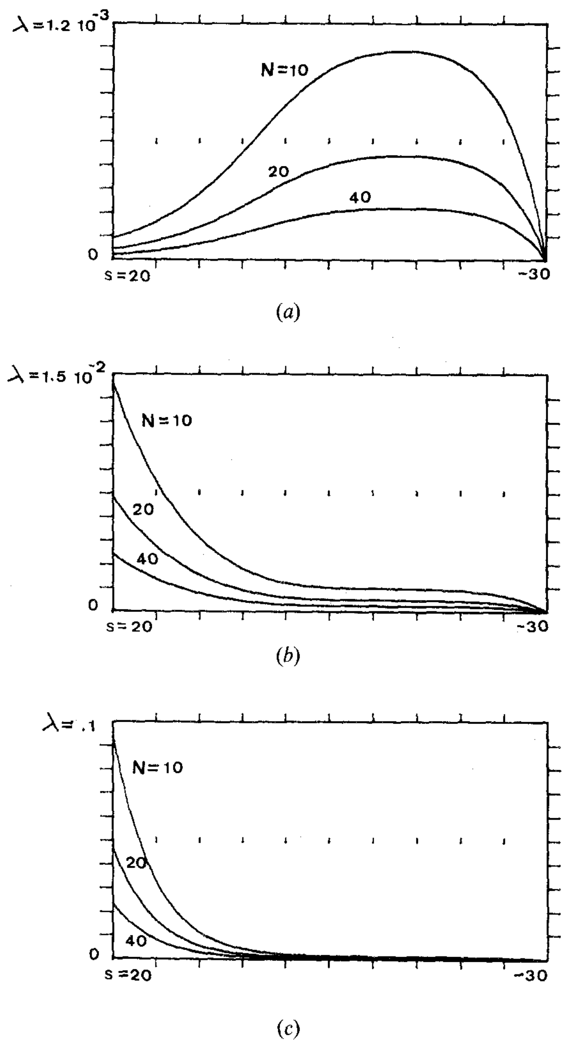


Fig. 6. — Speed of convergence (dB/iter). Gaussian noise ($\sigma_x^2 = \sigma_u^2 = 1$). *(a)* L_1 algorithm. *(b)* L_3 algorithm, *(c)* L_4 algorithm.

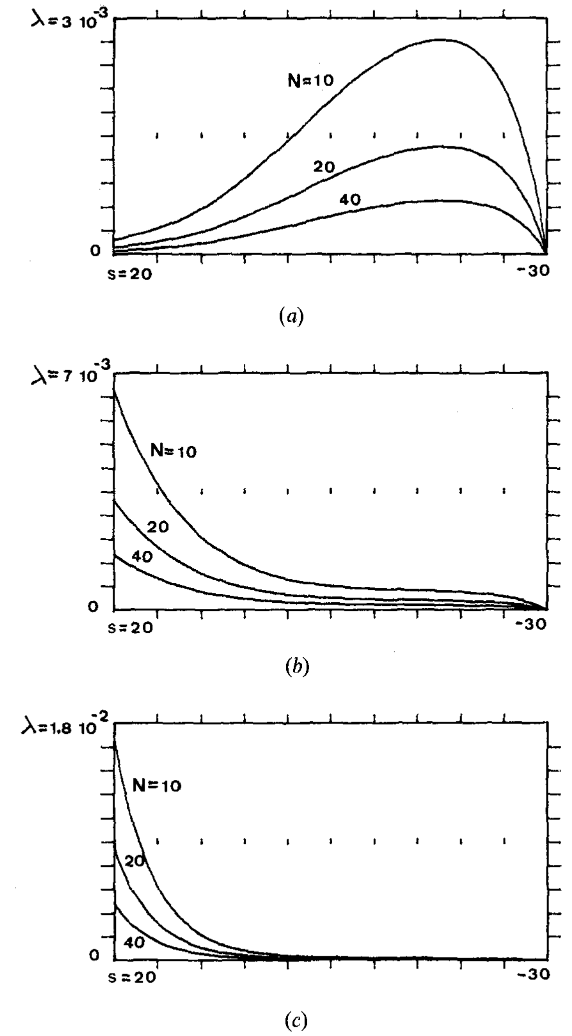


Fig. 7. — Speed of convergence (dB/iter). Exponential noise ($\sigma_x^2=\sigma_u^2=1$). (*a*) L_1 algorithm. (*b*) L_3 algorithm. (*c*) L_4 algorithm.

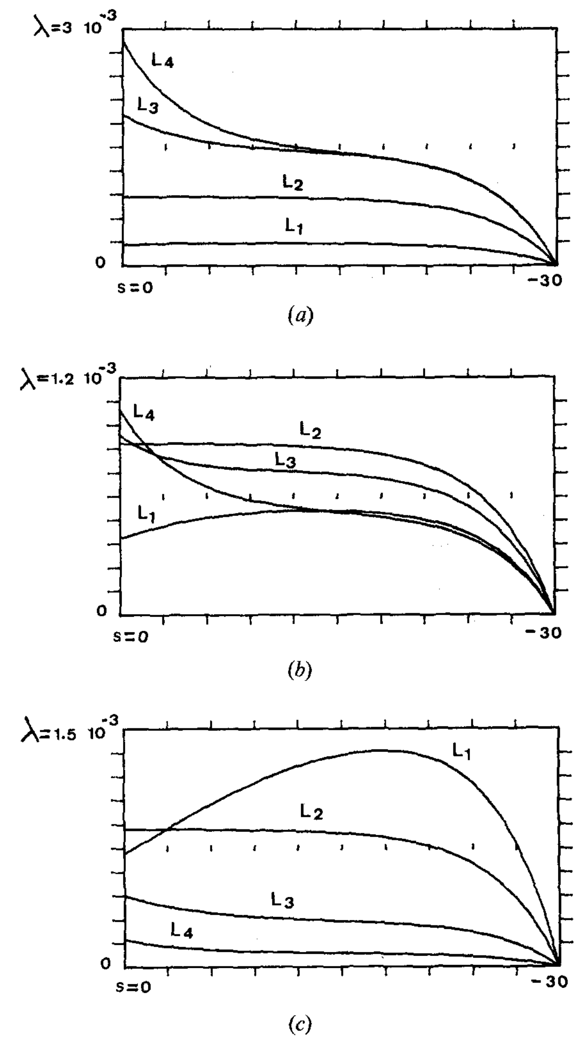


Fig. 8. Speed comparison (dB/iter., $\sigma_x^2=\sigma_u^2=1$, N=20). (*a*) Uniform noise. (*b*) Gaussian noise. (*c*) Exponential noise.

noise distribution. Note the following general characteristics:

– the L_1 algorithm shows a nearly flat local speed for identification errors of similar or slightly lower level that the plant noise, and lower in other cases;

– the L_3 and L_4 algorithms have very high speed when the identification error is much greater than the plant noise, but their speed decreases very quickly when this situation disappears.

All the schemes offer saturation effects.

On the other side, as expected, the speed of convergence decreases when the plant length increases.

To show clearly the relative advantages, Figures 8 *a*, *b*, *c* indicate comparatively the performances of the four algorithms for N=20, the other parameters being as above, for uniform, Gaussian, and exponential plant noises, respectively. The main conclusions are:

– for uniform (short tail) plant noise, higher the order is faster the convergence results;

– for Gaussian (intermediate tail) plant noise, L_2 is the best algorithm, except in (initial) high identification error situations;

– for exponential (long tail) plant noise, the lower order algorithms offer faster convergence rates, with the same exception indicated in the previous case.

5. Optimizing the adaption constant

5.1. DETERMINISTICALLY VARIABLE ADAPTION STEP

The previous discussion reveals that it is possible to increase the speed of convergence by using at each step an appropriate adaption constant; i. e., selecting an optimal $\mu_0(n)$ which minimizes $P(\mu;\, N,\, \sigma_x^2,\, \sigma_u^2)$, considering (24),

$$(52)\quad \partial P(\mu;\, N,\, \sigma_x^2,\, \sigma_u^2,\, \sigma_\varepsilon^2(n))/\delta\mu\,|_{\mu=\mu_0(n)} = -S_1(n)+2\,\mu_0(n)\,S_2(n)=0$$

and, then,

$$(53)\qquad \mu_0(n)=S_1(n)/2\,S_2(n)=C_k(n)/2$$

(according to this, convergence is guaranteed). This value needs to be computed step by step, and requires, obviously, to apply formulas (16 *a*) to (23 *c*), according to the case; note that $s=\sigma_\varepsilon(n)/\sigma_u$ can be computed if we know $\sigma_\varepsilon^2(0)$ by applying the basic formula (6); to know $\sigma_\varepsilon^2(0)$ (or a good estimate of it) is not very usual.

The values or forms of $\mu_0(n)$ correspond, according to (53), to (half) those discussed in Section 3 for $C_k(n)$.

The optimal reduction of $\sigma_\varepsilon^2(n)$ is, using $\mu_0(n)$,

$$(54)\quad P_0(N, \sigma_x^2, \sigma_u^2, \sigma_\varepsilon^2(n)) = P(\mu; N, \sigma_x^2, \sigma_u^2, \sigma_\varepsilon^2(n))\big|_{\mu=\mu_0(n)} = 1 - S_1^2(n)/4\,S_2(n)$$

and, according to the first discussion of Section 3, $P_0(.)<1$ if $N \geqq 2$; i. e., convergence is guaranteed, as previously said.

We will present our comparative discussion considering the usual cases.

5.1.1. *Noise-free plant case*

It is easy to obtain from (9 *a*), 9 *b*), 10 *a*), (10 *b*), (10 *c*) and (54)

$$P_0(.)=\begin{cases} 1-[(k-1)_*]^2/N(2k-3)_*, & \text{even } k \quad (55a)\\ 1-2/\pi N, & k=1 \quad (55b)\\ 1-2^k\{[(k-1)/2]!\}^2/\pi N(2k-3)_{*}, & \text{odd } k, k\neq 1 \quad (55c)\end{cases}$$

Note the expressions do not depend on σ_x^2. Table show values of (55 *a*), (55 *b*), (55 *c*) for different algorithm orders and plant lengths.

TABLE

Values of $P_0(.)$

N \ k	1	2	3	4	5	6
10.	.936	.900	.915	.940	.961	.976
20.	.968	.950	.958	.970	.981	.988
40.	.984	.975	.979	.985	.990	.994

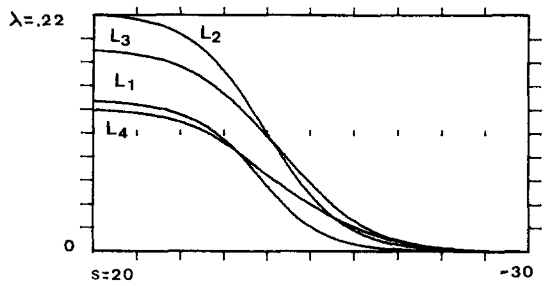
Note that the optimized L_2 algorithm presents the best behaviour (the convergence rate is $-10 \log P_0$), and that to increase k ($k>2$) and N decreases the speed of convergence. For large N, all the algorithms tend to be equivalent.

We must remark that, according to (53) and (9 *a*), (9 *b*) and (10 *a*), (10 *b*) the formal expression of the optimum adaption constant is

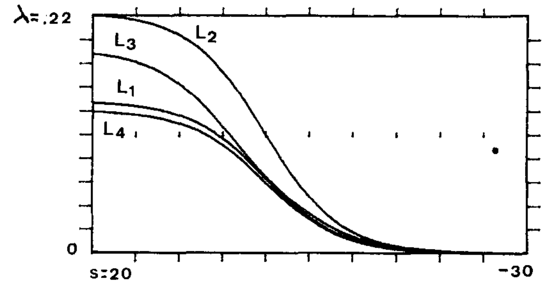
$$(56)\qquad \mu_0(n)=F(k)/N\,\sigma_x^2\,\sigma_\varepsilon^{k-2}(n)$$

being $F(k)$ a function of the order. This optimum step size depends on n except when $k=2$, in which case it becomes a constant, $\mu_0=1/2N\sigma_x^2$: this is an advantage of the L_2 algorithm, that can be optimized without additional computations.

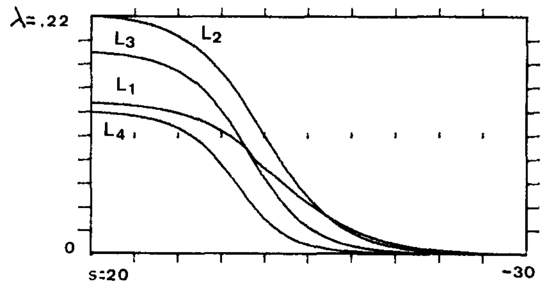
When $k=1$, μ_0 decreases with n [it varies with $\sigma_\varepsilon(n)$], forcing a zero final identification error variance.



(*a*)



(*b*)



(*c*)

Fig. 9. — Optimum speed comparison (dB/iter., $\sigma_x^2=\sigma_u^2=1$, N=20). (*a*) Univrom noise, (*b*) Gaussian noise, (*c*) Exponential noise.

When $k>2$, μ_0 is proportional to a value increasing with n, $1/\sigma_\varepsilon^{k-2}(n)$: this forces an unbounded sequence of step sizes, sonce $\sigma_\varepsilon^2(n)\to 0$.

According to the previous fact, we can say the optimized L_2 algorithm is the best choice assuming a noise-free plant.

Note this discussion cannot be applied when the plant noise is not zero.

5.1.2. *Noisy plant cases*

We will use again representative plots, since the analytical expressions do not allow an easy discussion.

Figures 9 *a*, *b*, *c*, show $\lambda_0(n)=-10\log P_0(.)$ for uniform, Gaussian, and exponential plant noises, respectively, using N=20 and comparing, thus, the speed of the L_1 to L_4 algorithms. We explore the band from 20 dB to −30 dB for $s=\sigma_\varepsilon(n)/\sigma_u$.

For high values of s (near the noiseless case), it can be seen that the performances follow the conclusions of the noise-free case: the optimized L_2 is the best algorithm, and the optimized L_4 worse. However,

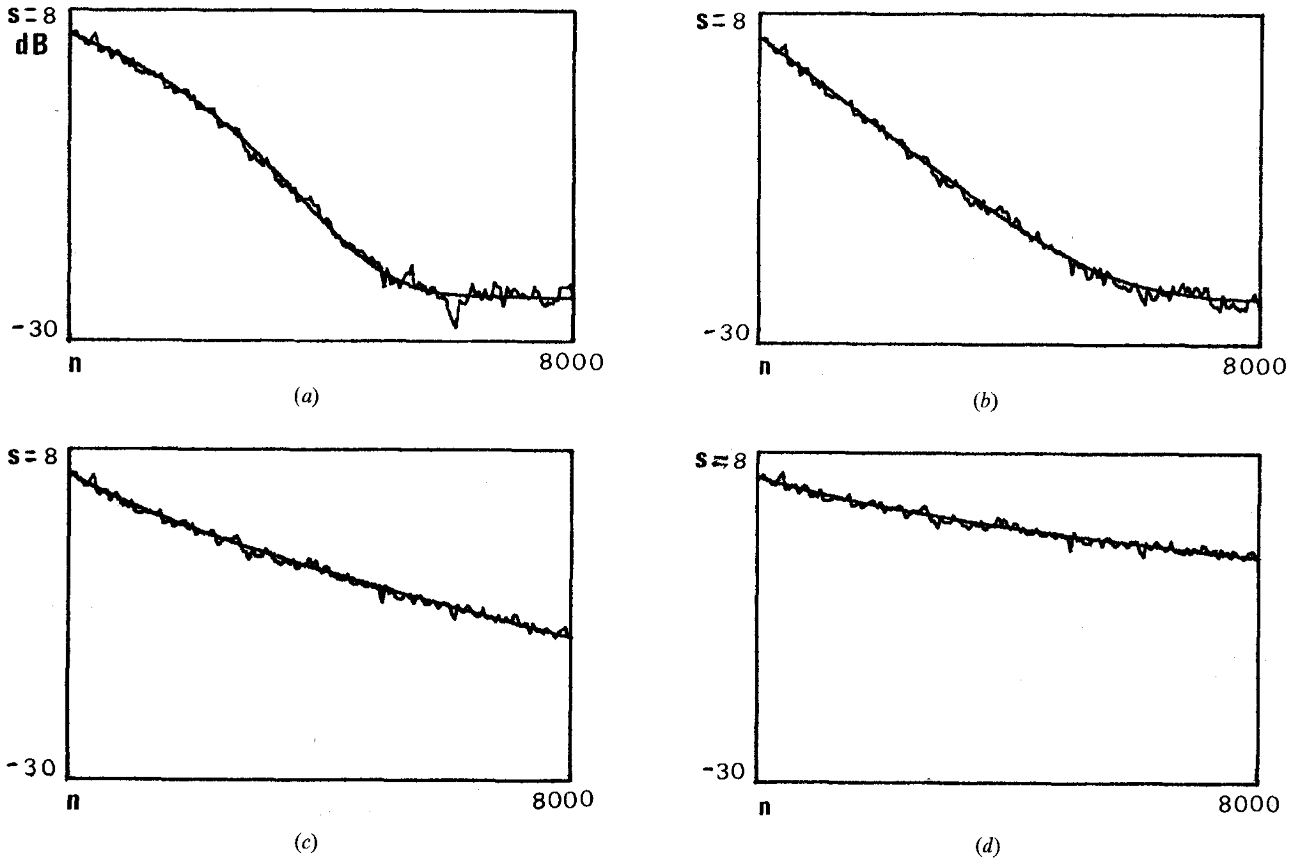


Fig. 10. – **Theoretical vs. simulated results. Exponential noise.** (*a*) L_1 **algorithm** ($\mu=8.40\times10^{-5}$). (*b*) L_2 **algorithm** ($\mu=2.15\times10^{-5}$). (*c*) L_3 **algorithm** ($\mu=4.97\times10^{-6}$). (*d*) L_4 **algorithm** ($\mu=5.27\times10^{-7}$).

when s decreases, there are different behaviours according to the assumed plant noise distribution:

– for uniform (short tail) noise, higher order algorithms tend to offer convergence rates better than those of L_1 and L_2 algorithms (when s is low enough (under $\simeq -8$ dB for $k=4$ and $\simeq -1$ dB for $k=3$ in the case of the Figure 9 *a*);

– for Gaussian (intermediate tail) noise, L_2 performs better in any circumstance;

– for exponential (long tail) noise, L_1 tends to become the best when s decreases (under $\simeq -7$ dB in the case of the Figure 9 *c*).

These results are coherent with those of the previous Section, and serve to the designer to select the appropriate algorithm.

5.2. First comments on an (stochastically) estimated step size

We must start by saying that it possible to think in modifying the value of μ according to the estimated (instantaneous) value of $\sigma_\varepsilon^2(n)$, trying to select the best adaption constant at each time; this is a very natural way of thinking, since it is clear from the previous discussion that there is an optimum value of μ depending on $\sigma_\varepsilon^2(n)$ (in different forms).

This approach will avoid the need of knowing $\sigma_\varepsilon^2(0)$ and using the theoretical formula (6) to compute μ_0 at each step. But note that all the previous analysis cannot be applied, since we have assumed a deterministic μ in our statistical analysis. Then, an additional formulation is needed; we will dedicate our further research effort to this possibility, since we have verified by simulations its potential advantages.

Hovever, a series of simple approaches can be explored without much reelaboration of our previous analysis; for instance, if we use instantaneous estimates for μ, and we accept a low-order approach for its variation, we will have

$$\mu_0' = \sum_{k=1}^{K} \mu_k |\varepsilon|^k \tag{57}$$

and the study of this possibility is equivalent to a combination of the schemes we are considering.

6. Simulation results

To verify the main aspects of our analysis and discussions, we present here some simulations results.

As a first example, consider the plant

$$c_p(i) = (-1)^i [1\text{-}0.05\, i], \qquad 0 \leqq i \leqq 9 \tag{58}$$

excited by a white, Gaussian sequence having $\sigma_x^2=1$. The initial setting for the adaptive filter is $\mathbf{c}(0)=\mathbf{0}$.

The first group of simulations correspond to the use of constant values of μ. These values are selected to

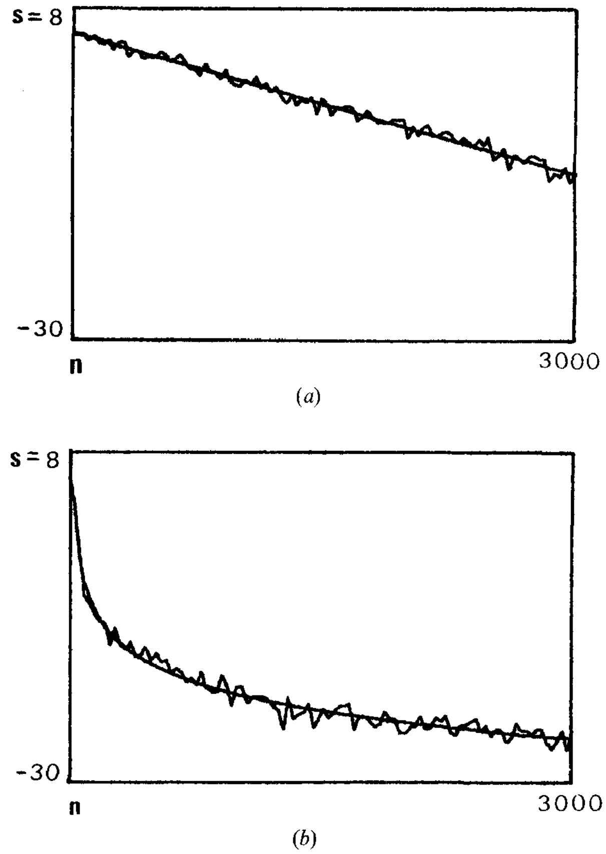


Fig. 11. — **Constant vs. optimum μ convergence comparison. Gaussian noise.** *(a)* **L_2 algorithm ($\mu = 3.15 \times 10^{-5}$).** *(b)* **$L_2$ algorithm [optimum $\mu_0(n)$].**

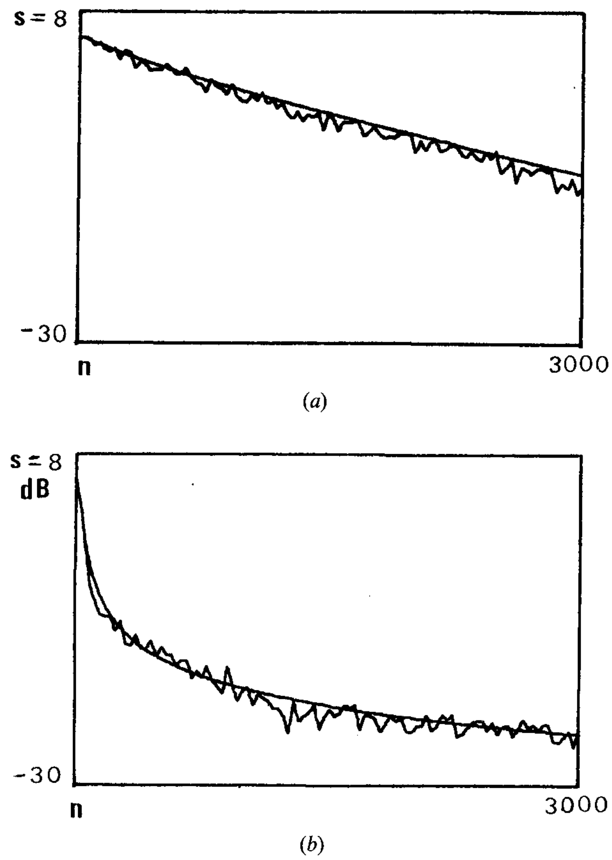


Fig. 12. — **Constant vs. optimum μ convergence comparison. Gaussian noise.** *(a)* **L_3** *algorithm* **($\mu = 1.11 \times 10^{-5}$).** *(b)* **$L_3$ algorithm optimum $\mu_0(n)$.**

obtain $\sigma_\varepsilon^2(\infty)/\sigma_u^2 = -25$ dB. All the figures show 4 averaged results for 8,000 iterations; we have also smoothed the curves with a constant moving-average window of length 50. The simulated results are presented jointly with the corresponding theoretical curves.

Figures 10 *a*, *b*, *c*, *d*, present results for exponential plant noise cases. The higher order algorithms converge slower than the L_1 when the presence of the plant noise is appreciable compared with the identification error. For brevity, cases with Gaussian and uniform plant noises will not be presented here, nevertheless their simulated results fit, in a similar way of Figure 10, the theoretical ones.

Thus, the observed characteristics agree, then, with the theoretical analysis. We must remark that the rule of having better performance of high/low order algorithms when the plant noise has short/long tails is just the predicted by classical works when applying the L_k norm to the objective $\varepsilon(n) + u$.

The second group of simulations is carried out using optimum values of μ, computed from $\mu(n) = S_1[s(n)]/2 S_2[s(n)]$ starting from a first estimate of $\sigma_\varepsilon^2(0)$ given by the average of four values of $\varepsilon^2(0)$ obtained with $\mathbf{c}(0) = \mathbf{0}$. We show only 3,000 iterations of 4 averaged realizations, smoothing with a window of length 30. Again for brevity, we present only results corresponding to Gaussian plant noise and L_2, L_3 algorithms, showing the agrement with the theoretical curves and the advantage respect to the constant algorithms [μ to obtain $\sigma_\varepsilon^2(\infty)/\sigma_u^2 = -25$ dB]: this agreement and advantage are similar for the other plant noises.

Figures 11 *a* , 12 *a* correspond to deterministically optimozed μ, and Figures 11 *b*, 12 *b* to constant μ. It can be observed that the optimized convergences are very similar, and that they are much faster than the constant μ convergences, mainly in the first steps (nearly noiseless condition). In these figures, the convergence to zero of the optimized algorithms cannot be observed.

As a general result, we can conclude that the closeness of theoretical and simulated results is so high that we can validate the hypotheses and analysis used.

7. Conclusion

We have studied a family of L_k-norm adaptive transversal filters which use the gradient of the k-th absolute moment of the overall error to update their coefficients in a plant identification application.

The assumed hypotheses and the analysis of the evolution of the identification error variance are validated by simulation results.

The comparative conclusions among the elements of this family are:

(1) Assuming a constant adaption step and zero plant noise, all the algorithms converge to zero in the mean

square sense except the L_1, which converges to an identification error variance depending on the adaption step used. The L_2 algorithm converges for step sizes under $1/N\,\sigma_x^2$, and the $k>2$ schemes for bounds related with the initial error variance $\sigma_\varepsilon^2(0)$. The convergence rate of the L_2 algorithm is constant, and the $k>2$ cases show a saturation effect.

(2) Considering a constant adaption step but non-zero plant noise, all the algorithms converge to a non-zero identification error variance which depends on the adaption step. When $k>2$, the initial state has to be considered to avoid divergence. The convergence speed is related to the plant noise distribution: when it has short/long tails, higher/lower order schemes perform better, respectively. For a Gaussian distribution, the L_2 algorithm is the best in a wide range of the identification error variance.

(3) Using a deterministic rule to optimize the adaption step, and assuming zero plant noise, the optimized L_2 algorithm offers the best performance, and also a constant value for the optimum setp size.

(4) Deterministically optimizing the step size and considering non-zero plant noise cases, all the algorithms clearly increase their convergence rates with respect to their constant step versions. The differences in performance according to the plant noise distribution are reduced a lot, all the algorithms offering very similar results: this observation and the easier computation of the adaption step for the L_2 algorithm, that does not depend on the plant noise, make this scheme preferable.

We must remark that, assuming a Gaussian input or a long plant filter, all these conclusions do not depend on the norm we use to measure the performance, since there is a direct relation between a L_k norm and the L_2 norm for a Gaussian error: that used in our analysis to give the k absolute moment as a function of the variance.

The line of introducing a stochastic optimization of the adaption step, as well as the possibility of using a general cost function to be selected according to optimization or robustness requirements, remain open. To extend this generalization to other adaptive schemes is also an interesting research field. Finally, the possibility of applying these L_k schemes to practical applications, such as equalizing and echo concelling in communications, needs to be considered.

Manuscrit reçu le 7 octobre 1986.

REFERENCES

[1] Claasen, TACM, Mecklenbraüker, WFG: Comparison of Two Algorithms for Adaptive FIR Digital Filters, *IEEE Trans. Acoustics, Speech and Signal Proc.*, ASSP-29, No. 3, Apr. 1981, pp. 670-678.

[2] Claasen, TACM, Mecklenbraüker, WFG: Overview of Adaptive Techniques in Signal Processing, *Proc. EUPSICO*-83, Erlangen, West Germany, Sept. 1983, pp. 747-754.

[3] W. A. Gardner, Learning Characteristics of Stochastic-Gradient-Descent Algorithms: A General Study, Analysis and Critique, *Sognal Processing*, 6, No., 2, Apr. 1984, pp. 113-133.

[4] J. M. Páez-Borrallo, A. R. Figueiras-Vidal and L. A. Hernández-Gómez, L-odd versus L-even Adaptive Algorithms: An Introduction, *Proc. IASTED Symp. on Applied Digital Filtering*, Paris, June 1985, pp. 32-35.

[5] J. M. Páez-Borrallo, A. R. Figueras-Vidal, Docampo-Amoedo, and J. R. Casar-Corredera, Convergence of Minimum L_k-norm Gradient-Type Adaptive Algorithms in Noise-Free Plant Identification, to be presented at *EUPSICO*-86, The Hague, The Netherlands, Sept. 1986.

[6] A. Papoulis, *Probability Random Variables and Stochastic Processes*, 2nd ed., New York, McGraw Hill, 1984.

[7] E. Walach and B. Widrow, The Least Mean Fourth (LMF) Adaptive Algorithm and Its Family: A General Discussion, *IEEE, Trans. Inf. Theory*, IT-30, No. 2, pt. 1, Mar. 1984, pp. 275-283.

[8] B. Widrow and M. E. Hoff, Adaptive Switching Circuits, *WESCON Conv. Rec.*, pt. 4, 1960, pp. 96-140.

[9] B. Widrow, J. M. Mcool, M. G. Larimore, and C. R. Johnson Jr., Stationary and Nonstationary Learning Characteristics of the LMS Adaptive Filter, *Proc. IEEE*, 64, No. 8, Aug. 1976, pp. 1151-1162.