Signal Processing 31 (1993) 1-15
Elsevier

Soft-constrained LMS algorithms for decoder stability in backward adaptive predictive systems*

J.C. Pesquet, O. Macchi and G. Tziritas
Laboratoire des Signaux et Systèmes, CNRS/UPS, Ecole Supérieure d'Électricité, Plateau de Moulon, 91192 Gif-sur-Yvette Cédex, France

Received 22 August 1991
Revised 26 June 1992

Abstract. This paper deals with the robustness of ADPCM systems versus transmission errors. To secure the stability of the decoder, it is necessary to modify the form of the LMS algorithm used to adapt the predictor. Solutions introducing soft constraints are investigated. While the leakage algorithm is proved to be not fully satisfactory, a new stabilizing algorithm is presented which allows to achieve good performances. Compared to existing methods, the main advantage of this algorithm is its low computational complexity. From a theoretical point of view, the effect of transmission errors is described by a set of nonlinear recurrent equations and an analysis is carried out in the deterministic second-order case. In more general cases, simulation results are given. Especially in image coding, it is shown that the method is efficient to fight against relatively high error rates.

Zusammenfassung. Diese Arbeit behandelt die Robustheit von ADPCM Systemen genenüber Übertragungsfehlern. Um die Stabilität des Dekodierers sicher zu stellen, muß die Form des LMS-Algorithmus, der für die Adaption des Prädiktors benutzt wird, verändert werden. Es werden Lösungen untersucht, wobei schwache Randbedingungen eingeführt werden. Während bewiesen wird, daß der 'Leakage'-Algorithmus nicht voll befriedigend arbeitet, wird ein neuer stabilisierender Algorithmus vorgestellt, der es erlaubt, gute Leistungen zu erzielen. Verglichen mit existierenden Methoden ist der Hauptvorteil dieses Algorithmus seine geringe rechnerische Komplexität. Der Effekt von Übertragungsfehlern wird vom theoretischen Standpunkt aus durch ein System nichtlinearer rekursiver Gleichungen beschrieben und es wird eine Analyse im deterministischen Fall zweiter Ordnung durchgeführt. In allgemeineren Fällen werden Simulationsergebnisse angegeben. Besonders bei der Bildkodierung wird gezeigt daß die Methode relativ hohe Fehlerraten wirkungsvoll bekämpft.

Résumé. Cet article traite de la robustesse des systèmes ADPCM vis-à-vis des erreurs de transmission. Pour assurer la stabilité du décodeur, il est nécessaire de modifier la forme de l'algorithme LMS qui est utilisé pour adapter le prédicteur. Des solutions introduisant des contraintes douces sont étudiées. Alors que l'algorithme avec facteur de fuite ne donne pas complète satisfaction, un nouvel algorithme stabilisant est présenté qui permet d'obtenir de bonnes performances. Comparativement au méthodes existantes, le principal avantage de cet algorithme est sa faible complexité en charge de calculs. D'un point de vue théorique, l'effet des erreurs de transmission est décrit par un système d'équations récurrentes non-linéaires et une analyse est menée dans le cas déterministe d'ordre 2. Dans des situations plus générales, des résultats de simulations sont fournis. En particulier en codage d'images, il apparaît que la méthode est efficace pour lutter contre des taux d'erreurs relativement élevés.

Keywords. Predictive coding; adaptive filtering; DPCM structure; transmission errors; constrained algorithms.

Correspondence to: Dr Jean-Christophe Pesquet, Laboratoire des Signaux et Systèmes, Ecole Supérieure d'Electricité, Plateau de Moulon, 91192 Gif-sur-Yvette Cédex, France.

* This paper has won the EURASIP/ELSEVIER Best Student Paper Award 1992.

1. Introduction

ADPCM (Adaptive Differential Pulse Code Modulation) is a well-known technique to achieve compression of correlated signals [10]. Indeed, it is a satisfactory solution to code both speech and images with low complexity. Yet, in the most widespread structure using a backward adapted predictor with the LMS algorithm, it appears that in the presence of transmission errors the decoder suffers from misalignment with respect to the encoder, which induces a coding noise. The decoder may even become unstable. To solve this problem, standard algorithms combine several means and in particular a modification of the adaptation and stability checks [5]. This last requirement makes the method substantially more complex when the order of the predictor is high and/or when two-dimensional fields are processed.

A simpler alternative solution is proposed here. It is based on a regularization of the prediction criterion leading to a class of soft-constrained LMS algorithms. According to the form of the regularization, either the usual LMS with a Leakage Factor (LF) [2, 6] or a new algorithm called the LMS with a Stabilizing Factor (SF) [11] is obtained. In this paper, the capabilities of these two algorithms to improve the robustness versus transmission errors are compared. The SF algorithm is proved to give the best performances.

In Section 2, the principles of adaptive predictive coding are recalled. Section 3 shows that the effect of transmission errors may be described by a set of nonlinear stochastic recurrent equations. In Section 4, the failure of the LMS algorithm is analysed and Section 5 explains the interest of including soft constraints in this algorithm. Then, in Section 6, a comparative study of the LF and SF algorithms is made when the recursion introduced in Section 3 is deterministic and of order two. To do so, some tools of nonlinear dynamical systems are used. Finally, in Section 7, computer simulations illustrate the advantages of the SF algorithm in more general cases and, in particular, for image coding.

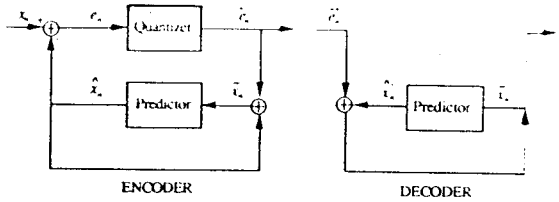

Fig. 1. Structure of the ADPCM system.

2. Adaptive predictive coding

The structure of the ADPCM system is presented in Fig. 1. The original signal is denoted by x_n, $\tilde{x}_n$ and $\hat{x}_n$ are respectively the reconstructed and the predicted samples, e_n is the prediction error and $\bar{e}_n$ is its quantized value.

The predictor is a FIR filter driven by $\tilde{x}_n$, which is adapted in order to deal with non-stationarities of x_n. Moreover, a backward adaptation is used to have no information to transmit other than a digital code corresponding to $\bar{e}_n$. Finally, the equations at the encoder are

$$\hat{x}_n = \boldsymbol{H}_{n-1}^{\mathrm{T}} \tilde{\boldsymbol{X}}_n, \tag{2.1a}$$

$$e_n = x_n - \hat{x}_n, \tag{2.1b}$$

$$\tilde{x}_n = \hat{x}_n + \bar{e}_n, \tag{2.1c}$$

$$\boldsymbol{H}_n = \mathscr{A}(\tilde{\boldsymbol{X}}_n, \boldsymbol{H}_{n-1}, \bar{e}_n), \tag{2.1d}$$

where $\boldsymbol{H}_n \in \mathbb{R}^N$ is the vector of predictor weights, $\tilde{\boldsymbol{X}}_n \in \mathbb{R}^N$ (respectively $\boldsymbol{X}_n \in \mathbb{R}^N$) is the corresponding vector of past values of $\tilde{x}_n$ (respectively x_n)[1] and $\mathscr{A}$ denotes the adaptation algorithm. Similarly, the equations at the decoder are

$$\hat{x}'_n = \boldsymbol{H}'^{\mathrm{T}}_{n-1} \tilde{\boldsymbol{X}}'_n, \tag{2.2a}$$

$$\tilde{x}'_n = \hat{x}'_n + \bar{e}'_n, \tag{2.2c}$$

$$\boldsymbol{H}'_n = \mathscr{A}(\tilde{\boldsymbol{X}}'_n, \boldsymbol{H}'_{n-1}, \bar{e}'_n), \tag{2.2d}$$

where the quantities used are distinguished by a prime from the equivalent ones at the encoder, because of possible transmission errors.

[1] For one-dimensional applications, we have $\tilde{\boldsymbol{X}}_n = (\tilde{x}_{n-1}, \ldots, \tilde{x}_{n-N})^{\mathrm{T}}$.

The decoder is said to be misaligned when $(\tilde{x}_n, \boldsymbol{H}_n) \neq (\tilde{x}'_n, \boldsymbol{H}'_n)$. The worst case is when recursion (2.2) becomes unstable in such a way that the decoded signal $\tilde{x}'_n$ bursts. The objective of the paper is to present and analyse a satisfactory adaptation rule $\mathscr{A}$, such that the decoder not only remains stable, but its misalignment vanishes with time. This is what we mean by 'robustness versus transmission errors'.

To simplify the analysis, the quantizer is assumed to be fixed. In fact, subject to some mild conditions [5, 7], the adaptation of this device is not a limiting factor for the robustness.

Moreover, to evaluate the quality of coding, we will use the prediction gain G [10] which is defined as

$$G = P_x / P_e, \tag{2.3}$$

where P_x is the power of x_n and P_e is the power of e_n.

3. Result of transmission errors

When there is no transmission error, identical initial values at the encoder and at the decoder ensure that $(\tilde{x}_n, \boldsymbol{H}_n) = (\tilde{x}'_n, \boldsymbol{H}'_n)$ for every $n \geq 0$. Conversely, transmission errors lead to the following misalignments:

$$\Delta\tilde{x}_n = \tilde{x}'_n - \tilde{x}_n, \tag{3.1a}$$

$$\Delta\boldsymbol{H}_n = \boldsymbol{H}'_n - \boldsymbol{H}_n. \tag{3.1b}$$

Subsequently, the dynamical behaviour of these misalignments is studied. For this purpose, it is assumed that if $n > n_0$, $\Delta\bar{e}_n = \bar{e}'_n - \bar{e}_n = 0$, which means that the errors have occurred before the iteration $n_0 + 1$. Then, when $n > n_0$, the following equations can be obtained from (2.1) and (2.2):

$$\Delta\tilde{x}_n = \boldsymbol{H}_{n-1}^{\mathrm{T}}\Delta\tilde{\boldsymbol{X}}_n + \tilde{\boldsymbol{X}}_n^{\mathrm{T}}\Delta\boldsymbol{H}_{n-1} + \Delta\tilde{\boldsymbol{X}}_n^{\mathrm{T}}\Delta\boldsymbol{H}_{n-1}, \tag{3.2a}$$

$$\Delta\boldsymbol{H}_n = \mathscr{A}_n\Delta\tilde{\boldsymbol{X}}_n + \mathscr{B}_n\Delta\boldsymbol{H}_{n-1} + \boldsymbol{C}_n, \tag{3.2b}$$

where

$$\mathscr{A}_n = (\partial\mathscr{A}/\partial\tilde{\boldsymbol{X}}_n)_{\tilde{\boldsymbol{X}}_n, \boldsymbol{H}_{n-1}, \bar{e}_n},$$

$$\mathscr{B}_n = (\partial\mathscr{A}/\partial\boldsymbol{H}_{n-1})_{\tilde{\boldsymbol{X}}_n, \boldsymbol{H}_{n-1}, \bar{e}_n},$$

and $\boldsymbol{C}_n$ is a vector depending on second- or higher-order terms in $\Delta\tilde{\boldsymbol{X}}_n$ and $\Delta\boldsymbol{H}_{n-1}$. Hence, the previous system is equivalent to a recurrence giving $(\Delta\tilde{\boldsymbol{X}}_{n+1}, \Delta\boldsymbol{H}_n)$. The problem is to get

$$(\Delta\tilde{\boldsymbol{X}}_{n+1}, \Delta\boldsymbol{H}_n) \xrightarrow[n \to +\infty]{} (\boldsymbol{0}, \boldsymbol{0}).$$

When this property is reached, the misalignments existing at iteration n_0 vanish and robustness versus transmission errors is achieved.

In general, the analysis is difficult for two reasons. The first one is the nonlinearity of system (3.2). The second one is the stochasticity of these equations since they are dependent on $\tilde{\boldsymbol{X}}_n$, $\boldsymbol{H}_{n-1}$ and $\bar{e}_n$, which themselves are functions of the random signal x_n. It must be emphasized that the difficulty stems from the use of an adaptive algorithm. Indeed, for the DPCM system where there is no adaptation of the predictor, $\mathscr{A}(\tilde{\boldsymbol{X}}_n, \boldsymbol{H}_{n-1}, \bar{e}_n)$ is a constant $\boldsymbol{H}$ and (3.2) reduces to

$$\Delta\tilde{x}_n = \boldsymbol{H}^{\mathrm{T}}\Delta\tilde{\boldsymbol{X}}_n, \tag{3.3a}$$

$$\Delta\boldsymbol{H}_n = \boldsymbol{0}. \tag{3.3b}$$

So, $\Delta\tilde{x}_n$ is the output of an IIR filter with a zero input. The inverse of this filter is usually called an innovator. Then, the alignment is reached if and only if the inverse of the innovator is a strictly stable linear filter. It is well-known that this condition is satisfied when $\boldsymbol{H}$ belongs to a stability domain of $\mathbb{R}^N$. For one-dimensional (respectively first quadrant quarter-plane two-dimensional) predictors, this domain is such that the zeros of the z-transfer function of the corresponding innovator are inside the unit circle [3] (respectively bi-circle [12]). For one-dimensional applications where $\tilde{\boldsymbol{X}}_n = (\tilde{x}_{n-1}, \tilde{x}_{n-2})^{\mathrm{T}}$, the domain is a triangle [3].

4. Failure of the LMS algorithm

A possible solution to adapt the predictor weights is the LMS algorithm [17]. In coding

applications, it has the advantage of providing low complexity and good tracking properties. It corresponds to the minimization of the mean square error (MSE):

$$J(\boldsymbol{H}) = \mathrm{E}\{(x_n - \boldsymbol{H}^{\mathrm{T}}\boldsymbol{X}_n)^2\}. \tag{4.1}$$

The form of the adaptation algorithm is then

$$\mathscr{A}(\tilde{\boldsymbol{X}}_n, \boldsymbol{H}_{n-1}, \bar{e}_n) = \boldsymbol{H}_{n-1} + \mu \bar{e}_n \tilde{\boldsymbol{X}}_n, \tag{4.2}$$

where $\mu > 0$ is the adaptation step-size. Thus (3.2b) becomes

$$\Delta \boldsymbol{H}_n = \mu \bar{e}_n \Delta \tilde{\boldsymbol{X}}_n + \Delta \boldsymbol{H}_{n-1}. \tag{4.3}$$

Let us now consider a signal x_n such that

$$x_n = \boldsymbol{H}_{\mathrm{opt}}^{\mathrm{T}} \boldsymbol{X}_n, \tag{4.4}$$

where $\boldsymbol{H}_{\mathrm{opt}} \in \mathbb{R}^N$ belongs to the border of the stability domain of IIR linear filters of order N (when x_n is non-vanishing). Such a signal is a predictable process which might be a tone in telephony or a uniform area in an image. Furthermore, let us assume that the algorithm has converged at the encoder when $n > n_0$ (by choosing n_0 large enough), i.e., $\boldsymbol{H}_n = \boldsymbol{H}_{\mathrm{opt}}$ and that the quantizer allows us to get $\bar{e}_n = e_n = 0$. Then, $\tilde{x}_n = x_n$ and (3.2a) and (4.3) lead to

$$\Delta \tilde{x}_n = (\boldsymbol{H}_{\mathrm{opt}} + \Delta \boldsymbol{H}_{n_0})^{\mathrm{T}} \Delta \tilde{\boldsymbol{X}}_n + \Delta \boldsymbol{H}_{n_0}^{\mathrm{T}} \tilde{\boldsymbol{X}}_n, \tag{4.5a}$$

$$\Delta \boldsymbol{H}_n = \Delta \boldsymbol{H}_{n_0}. \tag{4.5b}$$

Therefore, there exists a constant misalignment on $\boldsymbol{H}_n'$. Moreover, $\Delta \tilde{x}_n$ is the output of an IIR linear filter whose recursive part is $\boldsymbol{H}_{\mathrm{opt}} + \Delta \boldsymbol{H}_{n_0}$. Because of the form of $\boldsymbol{H}_{\mathrm{opt}}$, two situations may occur: this filter can either be stable or unstable according to the direction of $\Delta \boldsymbol{H}_{n_0}$. This means that $\Delta \tilde{x}_n$ is not surely bounded. Hence, the LMS algorithm fails to secure robustness. It could be noted that, in the derivation of (4.3) and (4.5a) from (3.2) and (4.2), no higher-order terms in $\Delta \tilde{\boldsymbol{X}}_n$ and $\Delta \boldsymbol{H}_{n-1}$ have been neglected, i.e., the resulting linear equations are exact in this case.

5. Use of soft constrained LMS algorithms

To solve the problem mentioned in the previous section, an approach of regularization can be followed. The criterion of minimization generalizes the classical MSE as follows:

$$J'(\boldsymbol{H}) = J(\boldsymbol{H}) + \alpha(\boldsymbol{H} - \boldsymbol{H}^f)^{\mathrm{T}}(\boldsymbol{H} - \boldsymbol{H}^f) + \beta(\boldsymbol{H} - \boldsymbol{H}^f)^{\mathrm{T}} \mathrm{E}\{\boldsymbol{X}_n \boldsymbol{X}_n^{\mathrm{T}}\}(\boldsymbol{H} - \boldsymbol{H}^f), \tag{5.}$$

where $\boldsymbol{H}^f$ is a fixed vector of $\mathbb{R}^N$, $\alpha \geqslant 0$ is the leakage factor and $\beta \geqslant 0$ is called the stabilizing factor. The minimum of $J'(\boldsymbol{H})$ can be reached by the stochastic gradient technique. It is easily found that the update equation is

$$\boldsymbol{H}_n = (1 - \mu\alpha)\boldsymbol{H}_{n-1} + \mu[x_n - (1+\beta)\boldsymbol{H}_{n-1}^{\mathrm{T}} \boldsymbol{X}_n + \beta \boldsymbol{H}^{f\mathrm{T}} \boldsymbol{X}_n] \boldsymbol{X}_n + \mu\alpha \boldsymbol{H}^f. \tag{5.2}$$

Of course, the LMS algorithm appears as a special case of (5.2), when $\alpha = \beta = 0$. The terms in α and β introduce implicit or soft contraints since they become of primary importance in the criterion when $\|\boldsymbol{H} - \boldsymbol{H}^f\|$ or $(\boldsymbol{H} - \boldsymbol{H}^f)^{\mathrm{T}} \boldsymbol{X}_n$ take large values. It is expected that this property will help preventing a divergence of $\boldsymbol{H}_n$ or $\hat{x}_n$. At the same time, in a stationary context, these constraints lead to a bias in the estimation of the vector of weights. Its asymptotic mean value can be specified, on the condition that $\boldsymbol{H}_n$ is convergent (see Appendix A),

$$\mathrm{E}\{\boldsymbol{H}_n\} - \boldsymbol{H}_{\mathrm{opt}} \xrightarrow[n \to +\infty]{} \boldsymbol{H}_\infty - \boldsymbol{H}_{\mathrm{opt}} = -[\alpha \boldsymbol{I} + (1+\beta)\mathscr{R}]^{-1} \times (\alpha \boldsymbol{I} + \beta \mathscr{R})(\boldsymbol{H}_{\mathrm{opt}} - \boldsymbol{H}^f), \tag{5.3}$$

where $\mathscr{R} = \mathrm{E}\{\boldsymbol{X}_n \boldsymbol{X}_n^{\mathrm{T}}\}$ and $\boldsymbol{H}_{\mathrm{opt}}$ is the best mean square estimate of the vector of weights minimizing $J(\boldsymbol{H})$. The above relation shows that the bias may be reduced owing to $\boldsymbol{H}^f$ which appears as an a priori estimate of $\boldsymbol{H}_{\mathrm{opt}}$. At the same time, the second and third terms of the criterion (5.1) tend to keep $\boldsymbol{H}_n$ close to $\boldsymbol{H}^f$. So, $\boldsymbol{H}^f$ must be well inside the stability domain to favour a stable algorithm. In practice, the fixed predictor corresponding to $\boldsymbol{H}^f$ must therefore be chosen 'sufficiently' stable and

must also be fitted to the long term statistics of the input x_n.

When a backward adaptation is used for the ADPCM system, x_n must be replaced by $\tilde{x}_n$ in (5.2) so that the decoder is able to duplicate the encoder updating. This yields

$$\mathcal{A}(\tilde{\boldsymbol{X}}_n, \boldsymbol{H}_{n-1}, \bar{e}_n) = (1-\mu\alpha)\boldsymbol{H}_{n-1} + \mu[\bar{e}_n - \beta(\boldsymbol{H}_{n-1} - H^f)^{\mathrm{T}}\tilde{\boldsymbol{X}}_n]\tilde{\boldsymbol{X}}_n + \mu\alpha H^f, \tag{5.4}$$

and the quantities appearing in (3.2b) are

$$\mathcal{A}_n = \mu\{[\bar{e}_n - \beta(\boldsymbol{H}_{n-1} - H^f)^{\mathrm{T}}\tilde{\boldsymbol{X}}_n]\boldsymbol{I} - \beta\tilde{\boldsymbol{X}}_n(\boldsymbol{H}_{n-1} - H^f)^{\mathrm{T}}\}, \tag{5.5b}$$

$$\mathcal{B}_n = (1-\mu\alpha)\boldsymbol{I} - \mu\beta\tilde{\boldsymbol{X}}_n\tilde{\boldsymbol{X}}_n^{\mathrm{T}}, \tag{5.5b}$$

$$\boldsymbol{C}_n = -\mu\beta\{\tilde{\boldsymbol{X}}_n\Delta\boldsymbol{H}_{n-1}^{\mathrm{T}}\Delta\tilde{\boldsymbol{X}}_n + [\tilde{\boldsymbol{X}}_n^{\mathrm{T}}\Delta\boldsymbol{H}_{n-1} + (\boldsymbol{H}_{n-1} - H^f)^{\mathrm{T}}\Delta\tilde{\boldsymbol{X}}_n + \Delta\boldsymbol{H}_{n-1}^{\mathrm{T}}\Delta\tilde{\boldsymbol{X}}_n]\Delta\tilde{\boldsymbol{X}}_n\}. \tag{5.5c}$$

Two cases will be of particular interest. The first one is the LMS with a Leakage Factor (LF) which is obtained when $\alpha \neq 0$ and $\beta = 0$. This algorithm is often the solution to misbehaviours that may occur with the LMS algorithm [14]. The second case corresponds to $\alpha = 0$ and $\beta \neq 0$ and is the new algorithm which was first proposed (but not analysed) in [11]: the LMS with a Stabilizing Factor (SF). Subsequently, this solution will be proved better than the leakage for the misalignment in our stability problem.

6. Analysis of a second-order recurrence

6.1. Background

For the reasons given in Section 3, the general form of the system defined by (3.2) and (5.5) seems intractable. Nevertheless, some analytical conclusions can be drawn for this nonlinear problem when it reduces to the following special case. On the one hand, it is considered that the order of the predictor is $N=1$, which leads to a second-order recurrence. Note that, in the analysis of nonlinear systems, it is a common practice to study in detail the low-dimension problems whereas the high-dimension ones are untractable. On the other hand, the input x_n is chosen predictable in such a way that the recurrence is deterministic. This choice is also justified by the analysis of Section 4 which has highlighted that this kind of signal pushes the standard LMS algorithm into instability. For a predictor of order 1, the predictable processes are of the form

$$x_n = (H_{\mathrm{opt}})^n X, \tag{6.1.1}$$

with $H_{\mathrm{opt}} \in \{-1, 1\}$.[2] The corresponding vector of past samples reduces to $\boldsymbol{X}_n = H_{\mathrm{opt}}x_n$. Additionally, it is assumed that the quantization error is negligible so that $\bar{e}_n = e_n$. Ideas about the hard nonlinearity introduced by the quantizer can be found in [16]. It is also assumed that the fixed predictor is stable, namely

$$-1 < H^f < 1. \tag{6.1.2}$$

6.2. Characteristics at the encoder

Under the previous assumptions, (5.2) yields

$$H_n = [1 - \nu(1+\sigma)]H_{n-1} + \nu(H_{\mathrm{opt}} + \sigma H^f), \tag{6.2.1a}$$

with

$$\nu = \mu X^2, \tag{6.2.1b}$$

$$\sigma = \frac{\alpha}{X^2} + \beta. \tag{6.2.1c}$$

The parameter ν is the normalized step size of the adaptive algorithm and σ is a factor giving a global evaluation of the regularization. If the following condition of convergence is satisfied:

$$0 < \nu < \nu_0 = \frac{2}{1+\sigma}, \tag{6.2.2}$$

[2] Non-bold variables are used to specify that scalar values are now considered.

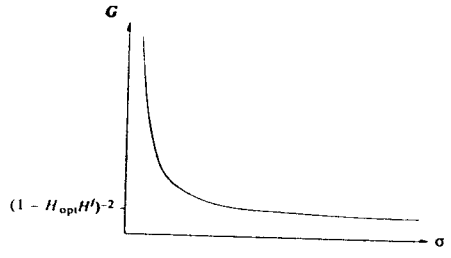

Fig. 2. Prediction gain as a function of σ.

the limiting value of H_n is

$$H_\infty = \frac{1}{1+\sigma}(H_{opt} + \sigma H^f). \tag{6.2.3}$$

Moreover, the prediction error becomes asymptotically

$$e_n^\infty = (H_{opt})^n X \frac{\sigma}{1+\sigma}(1 - H_{opt}H^f), \tag{6.2.4}$$

and the prediction gain which was defined in (2.3) is

$$G = \frac{(1+\sigma)^2}{(1-H_{opt}H^f)^2\sigma^2}. \tag{6.2.5}$$

As illustrated in Fig. 2, G gets poorer when σ increases (through an increase of α or β). Note also that the expression of G shows an improvement of the quality of the prediction as H^f gets closer to H_{opt}, which is the optimum value of H_n. It confirms the interest of the a priori knowledge.

6.3. Form of the recurrence for misalignments

To simplify the notations, let us define

$$\chi_n = \frac{\Delta \tilde{x}_n}{x_n}, \qquad \theta_n = \frac{\Delta H_n}{H_{opt}}, \tag{6.3.1a}$$

$$\delta = \frac{\alpha - \beta X^2}{\alpha + \beta X^2}. \tag{6.3.1b}$$

The variables χ_n and θ_n correspond to relative misalignments and δ allows us to compare the relative influences of the stabilizing and leakage factors. This parameter is a real number ranging from -1 to 1, the extreme values corresponding respectively to the SF and LF algorithms. Then, after convergence at the encoder, the misalignment equations (3.2) and (5.5) become

$$\chi_n = H_{opt}H_\infty\chi_{n-1} + \theta_{n-1} + \chi_{n-1}\theta_{n-1}, \tag{6.3}$$

$$\theta_n = H_{opt}(H_\infty - H^f)\delta\nu\sigma\chi_{n-1} + (1-\nu\sigma)\theta_n - \frac{1-\delta}{2}\nu\sigma[2\theta_{n-1} + H_{opt}(H_\infty - H^f)\chi_n + \theta_{n-1}\chi_{n-1}]\chi_{n-1}. \tag{6.3.}$$

This system is a nonlinear second-order recurrence for which, obviously, $(0, 0)$ is a fixed point. This point may be unique or not, according to the values of the parameters. Indeed, let us denote (χ, θ) another fixed point of the recurrence. It is straightforward to show that if $\sigma \neq 0$, (6.3.2) gives

$$H_{opt}(H_{opt} - H_\infty)\chi = (1+\chi)\theta, \tag{6.3.3}$$

$$(1-\delta)(H_{opt} - H^f)\chi^2 + [2H_{opt} - H_\infty - H^f - \delta(2H_{opt} + H_\infty - 3H^f)]\chi + 2[H_{opt} - H_\infty - \delta(H_\infty - H^f)] = 0. \tag{6.3.3}$$

In particular, it appears that the LF algorithm ($\delta = 1$) possesses a second fixed point

$$(\chi, \theta) = \left(\frac{H_{opt} - 2H_\infty + H^f}{H_\infty - H^f}, 1 - 2H_{opt}H_\infty + H_{opt}H^f\right). \tag{6.3.4}$$

This means that there is a risk for the algorithm to converge at the decoder towards a state for which the misalignment is nonzero. On the contrary, (6.3.3) shows that $(0, 0)$ is the only fixed point for the SF algorithm ($\delta = -1$). Hence the algorithm cannot lead to a finite misaligned decoder state. This is a first indication that the SF algorithm is superior to the LF one.

We will now consider the decoder stability when the misalignment variable (χ_n, θ_n) is close to the aligned state $(0, 0)$.

6.4. Local stability

6.4.1. Conditions of local stability

The recurrence (6.3.2) may be linearized in the neighbourhood of (0, 0) in the following way:

$$\begin{bmatrix} \chi_n \\ \theta_n \end{bmatrix} = \mathcal{M} \begin{bmatrix} \chi_{n-1} \\ \theta_{n-1} \end{bmatrix}, \tag{6.4.1.1a}$$

$$\mathcal{M} = \begin{bmatrix} H_{opt} H_\infty & 1 \\ H_{opt}(H_\infty - H^f)\delta\nu\sigma & 1-\nu\sigma \end{bmatrix}. \tag{6.4.1.1b}$$

The recurrence (6.3.2) is locally stable iff the linear recurrence (6.4.1.1) is stable, i.e., the matrix $\mathcal{M}$ has eigenvalues with moduli less than 1. It is well-known that this is equivalent to the inequalities

$$D < 1, \tag{6.4.1.2a}$$

$$|T| < 1 + D, \tag{6.4.1.2b}$$

for the determinant D and the trace T of $\mathcal{M}$. Then, the necessary and sufficient conditions for local stability are

$$H_{opt}[H_\infty(1-\nu\sigma) - (H_\infty - H^f)\delta\nu\sigma] < 1, \tag{6.4.1.3a}$$

$$|H_{opt} H_\infty + 1 - \nu\sigma| < 1 + H_{opt}[H_\infty(1-\nu\sigma) - (H_\infty - H^f)\delta\nu\sigma]. \tag{6.4.1.3b}$$

Moreover, the algorithm must be convergent, according to (6.2.2). If this condition is satisfied, it is shown in Appendix B that the set of inequalities (6.4.1.3) is equivalent to

$$\sigma > \delta, \tag{6.4.1.4a}$$

$$-[2\sigma - (\sigma - \delta)\nu\sigma]H_{opt} H^f < 2(2+\sigma) - (2+\sigma+\delta)\nu\sigma. \tag{6.4.1.4b}$$

It is useful to secure local stability for both kinds of predictable signals of order 1. Thus, introducing the two values $H_{opt} = 1$ and $H_{opt} = -1$ into (6.4.1.4b) yields

$$|2\sigma - (\sigma - \delta)\nu\sigma|\,|H^f| < 2(2+\sigma) - (2+\sigma+\delta)\nu\sigma. \tag{6.4.1.5}$$

Besides, if (6.4.1.4a) is satisfied, (6.2.2) yields

$$2\sigma - (\sigma - \delta)\nu\sigma > 0. \tag{6.4.1.6}$$

So, inequality (6.4.1.5) may be rearranged as

$$\nu < \nu_1 = \frac{2[2 + (1 - |H^f|)\sigma]}{[2 + (1 + |H^f|)\delta + (1 - |H^f|)\sigma]\sigma}. \tag{6.4.1.7}$$

Finally, subject to (6.2.2), the conditions of local stability are (6.4.1.4a) and (6.4.1.7). To proceed further, we will discuss these inequalities according to the values of the parameters.

6.4.2. Case $\delta \leqslant 0$

When δ is negative ($\alpha/X^2 \leqslant \beta$). (6.4.1.4a) holds naturally. Furthermore, it is easy to show that $\nu_1\sigma$ is greater than 2, so it is greater than $\nu_0\sigma$. Therefore, inequality (6.4.1.7) is satisfied. Hence, there is no constraint over σ. So, there is no limitation in the prediction performances.

6.4.3. Case $\delta > 0$

Let us now study the case where δ is positive ($\alpha/X^2 > \beta$).

(i) First, let us emphasize the importance of condition (6.4.1.4a).

Because of this inequality, there is a lower bound over σ and (6.2.5) shows that an upper bound for the prediction gain results:

$$G < G_M = \frac{(1+\delta)^2}{(1 - H_{opt} H^f)^2 \delta^2}. \tag{6.4.2.1}$$

This means that the local stability is obtained at the price of a poorer prediction. An equivalent way to visualize the performance degradation is the lower bound on the bias of the weights which is obtained as a result of (6.2.3),

$$|H_\infty - H_{opt}| > \frac{\delta}{1+\delta}|1 - H_{opt} H^f|. \tag{6.4.2.2}$$

According to (6.4.2.1) and (6.4.2.2), the prediction degradation is even stronger as δ is large and $H_{opt}H^f$ is small. For instance, for the LF algorithm ($\delta = 1$) and in the absence of an a priori knowledge about $H_{opt}(H^f = 0)$, we get $G_M = 4$ and $|H_f - H_{opt}| > 0.5$. The prediction error is only 6 dB below the input signal.

(ii) Secondly, let us show that condition (6.4.1.7) is not restrictive in practice.

This inequality is relevant only when ν_1 is lower than the convergence bound ν_0. According to the expressions of these quantities in (6.2.2) and (6.4.1.7), this is equivalent to

$$[-1 + |H^f| + (1 + |H^f|)\delta]\sigma > 2. \qquad (6.4.2.3)$$

Then, δ must be greater than $(1 - |H^f|)/(1 + |H^f|)$. This situation can only happen if $H^f \neq 0$, i.e. in the presence of some a priori knowledge about H_{opt}. Moreover, it is not difficult to check that the lower bound thus found for σ is then greater than δ and therefore leads to a prediction gain lower than G_M. In other words, to achieve the best possible prediction, it is interesting to choose σ close to δ in such a way that (6.4.2.3) is generally satisfied. So, from a practical point of view, if the convergence of the algorithm is taken for granted, the local stability of the decoder is ensured iff σ is greater than δ.

6.4.4. Conclusions

The main results of this section are summarized in Table 1, which demonstrates the critical role played by the sign of δ. It is clear that the case $\delta \leqslant 0$ is more interesting because the convergence condition is sufficient to ensure the local stability and there is no need to limit the prediction gain. In particular, it is shown that, for a predictable input of order 1, the SF algorithm is superior t the LF algorithm to reach the local stability of th decoder. Note also that, since better properties ar obtained when the factor β is greater than the leak age effect αX^2, the name of stabilizing factor i justified for β.

Table 1
Characteristics of locally stable constrained algorithms

	$\delta \leqslant 0$	$\delta > 0$
Convergence condition	$\nu < \nu_0$	$\nu < \nu_0$
Locally stability condition	none	$\sigma > \delta$, $\nu < \nu_1$
Maximum prediction gain	$+\infty$	G_M

6.5. Domains of attraction

When the conditions of local stability are ful filled, we can assert that there exists a domain o attaction $\mathcal{D}$ belonging to the phase plane such tha if $(\chi_{n_0}, \theta_{n_0}) \in \mathcal{D}$, $(\chi_n, \theta_n) \xrightarrow[n \to +\infty]{} (0, 0)$. Accordin to (6.2.3) and (6.3.2), this set is parametrized by δ, ν, σ and $H_{opt}H^f$. Numerical methods may b used to find the domain $\mathcal{D}$. For instance, a com puter allows us to determine the domain $\mathcal{D}_P$ defined as

$$\mathcal{D}_{P,\varepsilon} = \{(\chi_{n_0}, \theta_{n_0}) \in \mathbb{R}^2 | \exists p \in \{0, \ldots, P\} \quad |\chi_{n_0+p}| + |\theta_{n_0+p}| < \varepsilon\}. \qquad (6.5.1$$

An initial condition $(\chi_{n_0}, \theta_{n_0})$ in $\mathcal{D}_{P,\varepsilon}$ leads, afte maximum P iterations, to a state which is at distance of less than ε to the fixed point. By choos ing the integer number P large enough and the rea number ε small enough, a good approximation o $\mathcal{D}$ is obtained. Other methods to get the attractio domain are presented in [8].

In Figs. 3 and 4, the border of $\mathcal{D}_{P,\varepsilon}$ is drawn respectively for the LF and SF algorithms, when $P = 5000$ and $\varepsilon = 0.025$. It must be noted that there is no significant change in the form of this domain for 'reasonable' choices of the parameters ν, σ and $H_{opt}H^f$. Two comments may also be given on these

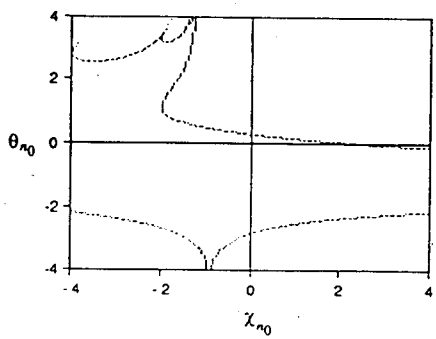

Fig. 3. Domain of attraction $\mathcal{D}_{5000,0.025}$ for the LF algorithm when $\nu = 0.1$, $\sigma = 1.5$, $H_{opt}H^f = 0$, $G = 4.4$ dB.

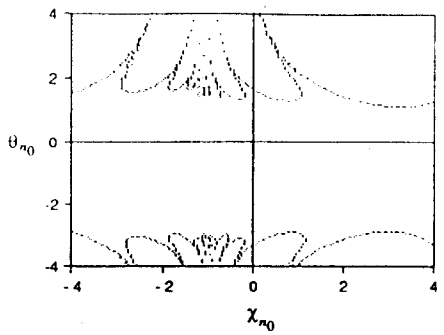

Fig. 4. Domain of attraction $\mathscr{D}_{5000,0.025}$ for the SF algorithm when $\nu=0.1$, $\sigma=0.3$, $H_{opt}H^f=0$, $G=12.7$ dB.

plots. At first, it may be observed that the largest part of the domains belongs to the lower half-plane. Indeed, according to (6.3.1a), a positive value of θ_n corresponds to a misalignment on the weights ΔH_n of the same sign as H_{opt}. So, the instability is more easily reached in this case. The other point to be emphasized is that the domain of attraction is much larger in the SF case than in the LF case. This fact is obviously in favour of the SF algorithm. This is all the more noteworthy as, according to relation (6.2.5), the SF algorithm illustrated in Fig. 4, corresponds to a higher value of the prediction gain ($G\cong 13$ dB) than the LF algorithm, illustrated in Fig. 3 ($G\cong 4$ dB). It is indeed not possible to reach a satisfactory prediction gain with the latter algorithm as a result of the basic constraint $\sigma>\delta=1$.

Based on the previous plots of the attraction domain, it is possible to predict whether the constrained algorithms will be robust to transmission errors or whether they will fail to realign the decoder. This is because the parameters χ_{n_0} and θ_{n_0} are related to the transmission errors arising before n_0.

Let us first consider the case of a single transmission error at n_0. Then, we clearly have

$$\chi_{n_0}=\Delta\bar{c}_{n_0}/X_{n_0}, \tag{6.5.2a}$$

$$\theta_{n_0}=\nu\chi_{n_0}. \tag{6.5.2b}$$

Therefore, $\chi_{n_0}\theta_{n_0}$ is positive and the points $(\chi_{n_0}, \theta_{n_0})$ are in the first and the third quadrant of the phase plane. At a first glance, it appears in Fig. 3 that the LF algorithms only have a small fraction of $\mathscr{D}$ in the first quadrant. However, this does not mean that it will fail to reach the alignment. Indeed, we usually have $|\Delta c_{n_0}|/X\leqslant 1$ and $\nu\mu X^2\ll 1$ and, thus, according to (6.5.2), $|\theta_{n_0}|\ll 1$. Now, domain $\mathscr{D}$ allows values around 0.1 for θ_{n_0}. Therefore, we can conclude that the algorithm will generally realign after a single error. The same conclusion evidently holds for the SF algorithm, which is much less critical.

When several consecutive transmission errors occur, the expressions of χ_{n_0} and θ_{n_0} are not readily accessible. So, our study does not allow us to give quantitative results when there is an accumulation of errors rather than the propagation of a single error. In a qualitative way, it is understood that $|\theta_{n_0}|$ will reach greater values. According to the plots of Fig. 4, such values are likely to be included in the stability domain of the SF algorithm, if there are not too many errors. Thus, we can conclude that this algorithm generally recovers after multiple errors. In view of Fig. 3, a similar behaviour cannot be expected for the LF algorithm. These conclusions will be supported by some simulations in the next section.

As was already noted, the previous plots do not allow us to clearly show the influences of the parameters ν, σ and H^f. To proceed further, we will plot $\mathscr{D}_{P,\varepsilon}$ for a relatively small value of P. The size of the corresponding domain gives an indication of the average speed of alignment. Indeed, the points belonging to $\mathscr{D}$ which are not included in $\mathscr{D}_{P,\varepsilon}$ correspond to initial conditions requiring more than P iterations to reach the alignment. In this way, the transient behaviour of the alignment phenomenon is evidenced. Figures 5 and 6 show the form of $\mathscr{D}_{P,\varepsilon}$ when $P=50$ and $\varepsilon=0.02$. Although these domains look quite different in the LF and SF cases, the influences of ν, σ and H^f seem to be qualitatively similar. In particular, it appears that an increase of ν or σ tends to enlarge $\mathscr{D}_{P,\varepsilon}$, which means that the alignment speed is globally improved. The introduction of H^f has an opposite effect. This result is in agreement with the intuition that H^f must not be chosen too close to

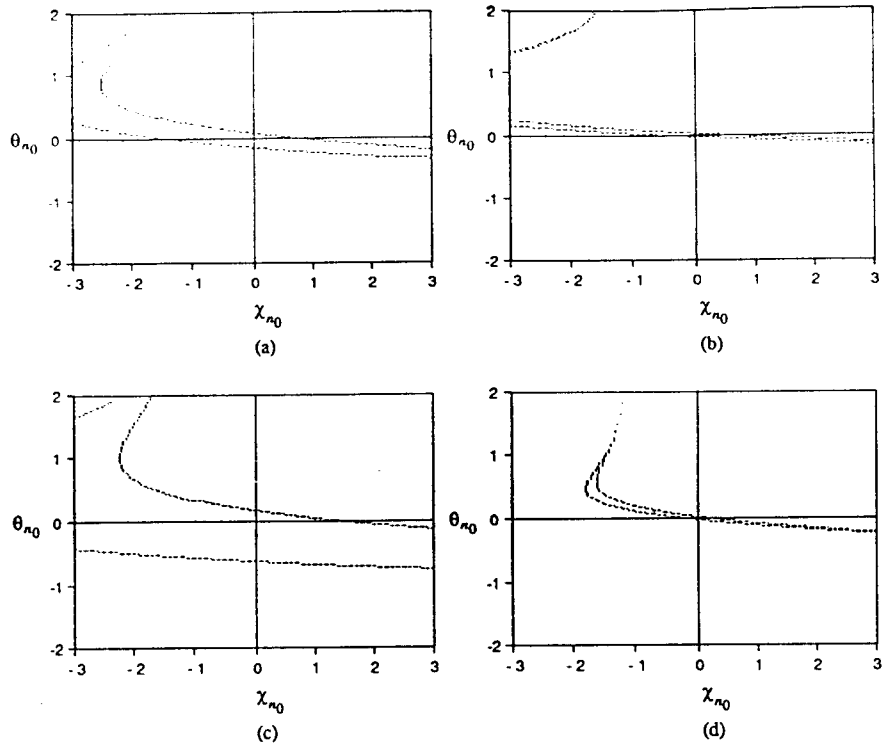

Fig. 5. Domain of attraction $\mathcal{D}_{50,0.02}$ for the LF algorithm when (a) $\nu=0.1$, $\sigma=1.5$, $H_{opt}H^f=0$, $G=4.4$ dB; (b) $\nu=0.05$, $\sigma=1.5$, $H_{opt}H^f=0$, $G=4.4$ dB; (c) $\nu=0.1$, $\sigma=1.7$, $H_{opt}H^f=0$, $G=4.0$ dB; (d) $\nu=0.1$, $\sigma=1.5$, $H_{opt}H^f=0.6$, $G=12.4$ dB.

instability. Indeed, as noted in Section 5, the presence of the a priori knowledge in criterion (5.1) tends to bring the estimates of the weights close to H^f. Hence, the choice of this parameter must be made cautiously. In practice, there will be a trade-off between the ability to compensate the bias of the stabilizing algorithms and an increase of the sensitivity to transmission errors.

7. Simulation results

Further results show that the conclusions of the previous section about the decoder alignment are valid in a more general framework.

7.1. Extension to a higher order

In the case of a predictor of order two and a sinusoidal input $x_n=X\sin(n\varphi)$, the recurrence given by (3.2) and (5.5) is of order four. For example, some simulations are provided when $\varphi=\pi/4$ and a perturbation $\Delta\tilde{x}_{n_0}=0.5X$, $\Delta H_{n_0}=[0.05, -0.05]^T$ arises at $n_0=50$. The normalized adaptation step-size $\nu=\mu X^2$ is set to 0.1 and there is no a priori knowledge ($H^f=\mathbf{0}$). The parameter σ is still defined as in (6.2.1c). Furthermore, the prediction gain has been evaluated empirically but it is easy to check its values theoretically.

In Figs. 7 and 8, the variations of $\Delta\tilde{x}_n/X$ are plotted for the LF and SF algorithms, respectively. For the LF algorithm, it appears that the decoder becomes unstable if $\sigma\leqslant 0.1$. The alignment is reached for $\sigma=0.2$ but the prediction gain is 5.2 dB, which is a poor value. For the SF algorithm, much higher values of G may be obtained, but if σ is chosen too small, the alignment speed is slow. By taking $\sigma=0.3$, the convergence is almost comparable to the LF algorithm with $\sigma=0.2$ and the prediction gain is improved at the more satisfactory value of 12.7 dB.

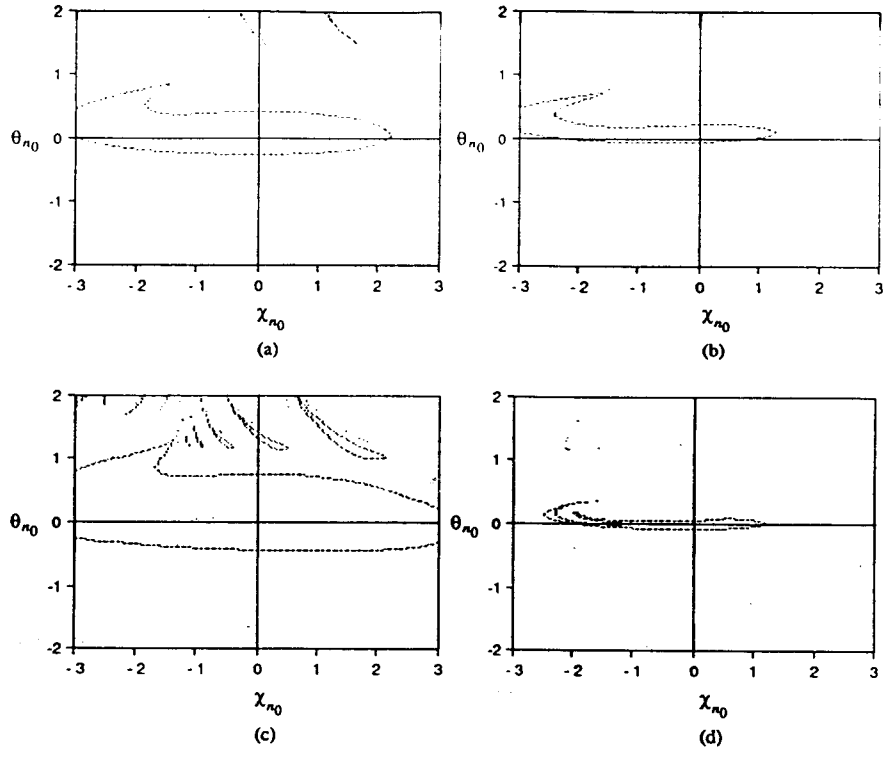

Fig. 6. Domain of attraction $\mathscr{D}_{50,0.02}$ for the SF algorithm when (a) $\nu=0.1$, $\sigma=0.3$, $H_{opt}H^f=0$, $G=12.7$ dB; (b) $\nu=0.05$, $\sigma=0.3$, $H_{opt}H^f=0$, $G=12.7$ dB; (c) $\nu=0.1$, $\sigma=0.5$, $H_{opt}H^f=0$, $G=9.5$ dB; (d) $\nu=0.1$, $\sigma=0.3$, $H_{opt}H^f=0.6$, $G=20.7$ dB.

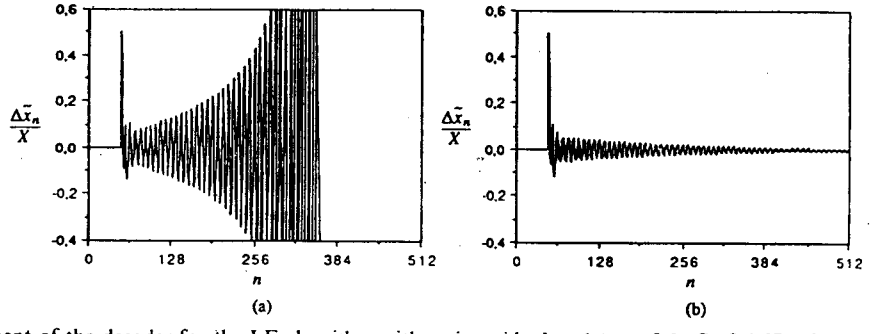

Fig. 7. Alignment of the decoder for the LF algorithm with a sinusoid when (a) $\sigma=0.1$, $G=8.4$ dB; (b) $\sigma=0.2$, $G=5.2$ dB.

7.2. *Application to image coding*

We consider images ordered by a raster scanning. Therefore, if the two-dimensional field is $x_{p,q}$ where p and q are respectively the vertical and horizontal coordinates, it may also be denoted x_n, where n is the scanning order. The two-dimensional nature of the signal can be taken into account by a two-dimensional prediction. Generally, the predictor is a FIR filter with three coefficients [9] such that

$$\boldsymbol{X}_n=[x_{p,q-1} \quad x_{p-1,q} \quad x_{p-1,q-1}]^{\mathrm{T}}. \tag{7.2.1}$$

Furthermore, by adapting the weights with an LMS algorithm, the performances of DPCM are

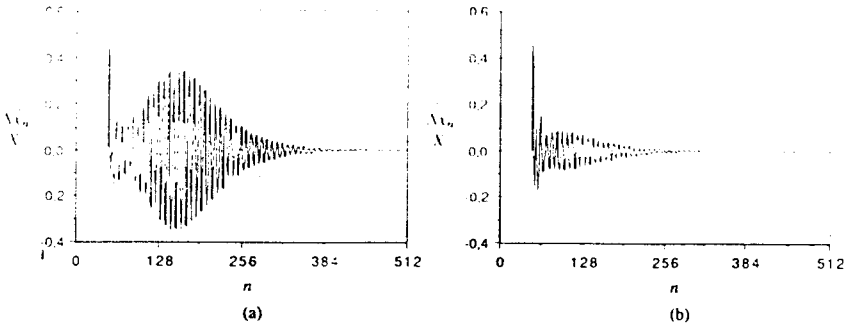

Fig. 8. Alignment of the decoder for the SF algorithm with a sinusoid when (a) $\sigma = 0.1$, $G = 20.8$ dB; (b) $\sigma = 0.3$, $G = 12.7$ dB.

improved, compared to the performance of a fixed predictor [1, 15].

Simulations are given for the first image of the test sequence Car (COST 211bis) which is typical of digital TV applications (see Fig. 9). This image is characterized by 528 × 674 8-bit pels. To decrease the transmission rate, the ADPCM system is used with a five-level fixed quantizer. The quality of the decoded images is evaluated in the presence of transmission errors both visually and by the signal-to-distortion ratio defined by

$$\mathrm{SNR} = \frac{255^2}{\mathrm{E}\{(x_n - \tilde{x}'_n)^2\}}. \qquad (7.2.2)$$

Fig. 9. First image of the Car sequence.

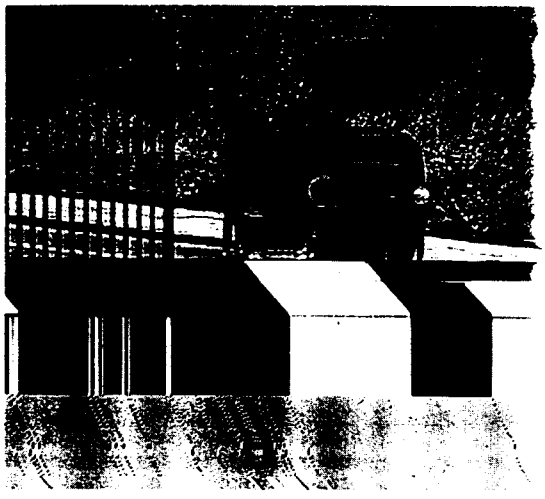

Fig. 10. Reconstructed image for the LMS algorithm with o transmission error at (264, 337).

Figure 10 illustrates the disastrous effect on t decoding process of a single transmission err arising at the centre of the image, when the LM algorithm is used to adapt the predictor. By usi the LF and SF algorithms the behaviour is mu improved. Figures 11 and 12 show the differen between the reconstructed image and the origin when the error rate is 10^{-5} (four errors). The val of the difference is translated by 127, so that the l grey level corresponds to zero. On these images, t superiority of the SF algorithms clearly appea This method also gives satisfactory performanc when the error rate reaches 10^{-3} (400 errors), evidenced by Figs. 13 and 14. In the first figure, t

Fig. 11. Reconstruction error for the LF algorithm with four uniformly distributed transmission errors ($\alpha = 5000$, SNR = 815, $H = 0$).

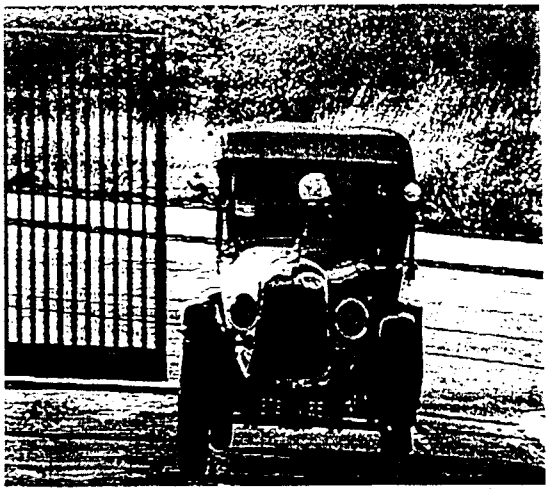

Fig. 13. Reconstructed image for the SF algorithm with 400 uniformly distributed transmission errors ($\beta = 0.8$, SNR = 1690, $H' = [0.6, 0.6, -0.36]^t$).

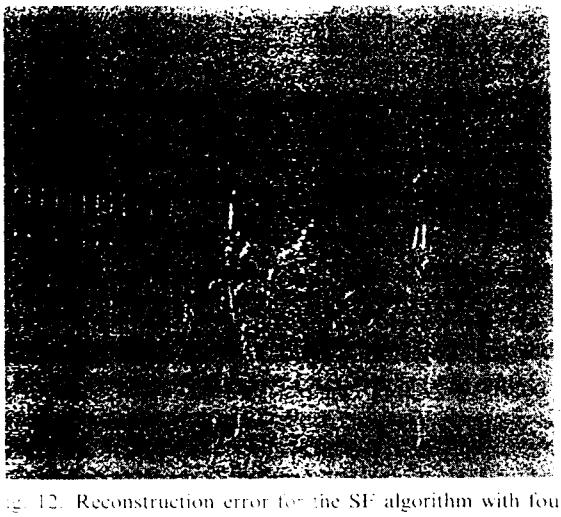

Fig. 12. Reconstruction error for the SF algorithm with four uniformly distributed transmission errors ($\beta = 0.3$, SNR = 1600, $H' = 0$).

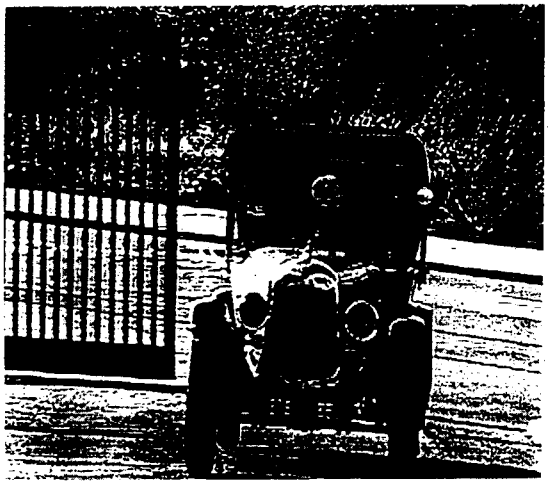

Fig. 14. Reconstructed image for the SF algorithm with a block of 400 transmission errors at the beginning of line 264 ($\beta = 0.8$, SNR = 1570, $H' = [0.6, 0.6, -0.36]^t$).

errors are uniformly distributed all over the image whereas, in the second figure, a block of errors is considered. In both cases, the value taken for H' is $[0.6, 0.6, -0.36]^t$ [3] which allows us to improve the prediction slightly. Note that, in all simulations, the smallest values of α and β allowing us to reach the alignment were chosen, in order to minimize the distortion. Furthermore, the errors were introduced by imposing an equiprobable transition of the coded prediction error towards one of the four other possible values.

[3] This two-dimensional filter corresponds to a separable symmetric innovator whose transfer function is $(1 - 0.6z_1^{-1})(1 - 0.6z_2^{-1})$.

It must be emphasized that similar behaviours may be observed in speech coding [4].

8. Conclusions

In an ADPCM system, transmission errors lead to misalignments between the variables at the encoder and at the decoder, which have been modelled by nonlinear recurrent equations. As the LMS algorithm fails to ensure the alignment of predictable processes, two constrained algorithms of low complexity were introduced. There is then a bias on the estimation of weights but it can be decreased if an a priori estimate of the weights is available.

In the case of a prediction of order one and a predictable input, the dynamical behaviour of misalignments has been studied. The analysis of local stability and the evaluation of domains of attraction evidence that the robustness is improved by the constrained algorithms. Compared to the LF algorithm, less restrictive stability conditions are obtained for the new SF algorithm which allows a better prediction.

It could be interesting to extend these theoretical results as computer simulations have shown that similar conclusions can also be drawn for sinusoids, speech and image signals.

Appendix A

Taking the expectation of each side of (5.2) yields

$$\mathrm{E}\{\boldsymbol{H}_n\} = [(1-\mu\alpha)\boldsymbol{I} - \mu(1+\beta)\mathscr{R}]\mathrm{E}\{\boldsymbol{H}_{n-1}\} + \mu[\mathscr{R}\boldsymbol{H}_{\mathrm{opt}} + (\alpha\boldsymbol{I} + \beta\mathscr{R})\boldsymbol{H}^f], \quad \text{(A.1)}$$

by making the usual assumption that $\boldsymbol{H}_{n-1}$ and $\boldsymbol{X}_n$ can be considered as independent and by substituting $\mathscr{R}\boldsymbol{H}_{\mathrm{opt}}$ for $\mathrm{E}\{x_n\boldsymbol{X}_n\}$. If the following conditions are satisfied:

$$\mu < \frac{2}{\alpha + (1+\beta)\lambda_{\max}}, \quad \text{(A.2a)}$$

$$\alpha + (1+\beta)\lambda_{\min} > 0, \quad \text{(A.2b)}$$

where $\lambda_{\min}$ and $\lambda_{\max}$ denote the minimum and maximum eigenvalues of $\mathscr{R}$, respectively, the sequence $\boldsymbol{H}_n$ admits an asymptotic mean value according to

$$\mathrm{E}\{\boldsymbol{H}_n\} \xrightarrow[n\to+\infty]{} \boldsymbol{H}_\infty = [\alpha\boldsymbol{I} + (1+\beta)\mathscr{R}]^{-1} \times [\mathscr{R}\boldsymbol{H}_{\mathrm{opt}} + (\alpha\boldsymbol{I} + \beta\mathscr{R})\boldsymbol{H}^f]. \quad \text{(A.3)}$$

Note that, when $\mathscr{R}$ is a singular matrix, (A.2b) implies that a unique asymptotic mean value is obtained only if $\alpha \neq 0$.

Appendix B

In this appendix, it is proved that inequalities (6.4.1.3) reduce to inequalities (6.4.1.4) under condition (6.2.2).

Let us recall that the definitions of δ given in (6.3.1b) imply that

$$-1 \leqslant \delta \leqslant 1. \quad \text{(B.1)}$$

Moreover, by using expression (6.2.3), conditions (6.4.1.3) may be rewritten as follows:

$$\sigma(1 - H_{\mathrm{opt}}H^f) + \nu\sigma[1 + \sigma H_{\mathrm{opt}}H^f + \delta(1 - H_{\mathrm{opt}}H^f)] > 0, \quad \text{(B.2a)}$$

$$(1 - H_{\mathrm{opt}}H^f)(\sigma - \delta) > 0, \quad \text{(B.2b)}$$

$$[2\sigma - \nu\sigma(\sigma - \delta)](1 + H_{\mathrm{opt}}H^f) + 2[2 - (1+\delta)\nu\sigma] > 0. \quad \text{(B.2c)}$$

Let us first consider conditions (B.2b) and (B.2c). It is straightforward from (6.1.2) that (B.2b) is equivalent to (6.4.1.4a). Moreover, (6.4.1.4b) is obviously equivalent to (B.2c).

Let us now prove that condition (B.2a) is always satisfied. To this purpose, two cases may be examined.

(i) If $1+\sigma H_{\text{opt}}H^f+\delta(1-H_{\text{opt}}H^f)<0$, (6.2.2) leads to

$$\begin{aligned}&\sigma(1-H_{\text{opt}}H^f)\\&\quad+\nu\sigma[1+\sigma H_{\text{opt}}H^f+\delta(1-H_{\text{opt}}H^f)]\\&\quad>\frac{\sigma}{1+\sigma}[3-H_{\text{opt}}H^f+2\delta(1-H_{\text{opt}}H^f)\\&\qquad+\sigma(1+H_{\text{opt}}H^f)].\end{aligned}\tag{B.3}$$

According to (6.1.2) and (B.1),

$$\delta(1-H_{\text{opt}}H^f)\geq-1+H_{\text{opt}}H^f,\tag{B.4}$$

and we then have

$$\begin{aligned}&\sigma(1-H_{\text{opt}}H^f)\\&\quad+\nu\sigma[1+\sigma H_{\text{opt}}H^f+\delta(1-H_{\text{opt}}H^f)]\\&\quad>\sigma(1+H_{\text{opt}}H^f).\end{aligned}\tag{B.5}$$

The above inequality associated to (6.1.2) shows that (B.2a) is true.

(ii) If $1+\sigma H_{\text{opt}}H^f+\delta(1-H_{\text{opt}}H^f)\geq0$, it is obvious that the same conclusion holds.

Acknowledgments

The authors would like to thank anonymous reviewers for their helpful comments.

References

[1] S.T. Alexander and S.A. Rajala, "Image compression results using the LMS adaptive algorithm", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. ASSP-33, No. 3, June 1985, pp. 712–714.

[2] M. Bellanger, *Adaptive Digital Filters and Signal Analysis*, Dekker, New York, 1987, pp. 112–114.

[3] M. Bellanger, *Traitement Numérique du Signal*, Masson, Paris, 1987, Chapter VI, pp. 195–224.

[4] M. Bonnet, *Codage Numérique des Signaux par Quantification et Prédiction Adaptatives Couplées*, Thèse d'Etat, Université de Paris-Sud, 1988.

[5] M. Bonnet, O. Macchi and M. Jaïdane-Saïdane, "Theoretical analysis of the ADPCM CCITT algorithm", *IEEE Trans. Comm.*, Vol. COM-38, No. 6, June 1990, pp. 847–858.

[6] R.D. Gitlin, H.C. Meadors, Jr. and S.B. Weinstein, "The tap-leakage algorithm: An algorithm for the stable operation of a digitally implemented fractionally spaced adaptive equalizer", *Bell System Tech. J.*, Vol. 61, October 1982, pp. 1817–1839.

[7] D.J. Goodman and R.M. Wilkinson, "A robust adaptive quantizer", *IEEE Trans. Comm.*, Vol. COM-22, August 1974, pp. 1037–1045.

[8] I. Gumowski and C. Mira, *Recurrences and Discrete Dynamic Systems*, Springer, Berlin, 1980.

[9] A.K. Jain, "Image data compression: A review", *Proc. IEEE*, Vol. 69, No. 3, March 1981, pp. 349–389.

[10] N.S. Jayant and P. Noll, *Digital Coding of Waveforms*, Prentice Hall, Englewood Cliffs, NJ, 1984, Section 6, pp. 252–350.

[11] O. Macchi and M. Jaïdane-Saïdane, "Predictor stability", European Patent 8710018, 1987.

[12] B.T. O'Connor and T.S. Huang, "Stability of general two-dimensional recursive filters", in: *Two-dimensional Digital Signal Processing I, Linear Filters*, Springer, Berlin, 1981, pp. 85–154.

[13] J.C. Pesquet, O. Macchi and G. Tziritas, "Second-order statistical analysis of two constrained LMS algorithms", *Proc. EUSIPCO*, September 1990, pp. 193–196.

[14] W.A. Sethares, D.A. Lawrence, C.R. Johnson, Jr. and R.R. Bitmead, "Parameter drift in LMS adaptive filters", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. ASSP-34, No. 4, August 1986, pp. 868–879.

[15] G. Tziritas and J.C. Pesquet, "A hybrid image coder: Adaptive intra-interframe prediction using motion compensation", *Proc. Internat. Conf. Acoust. Speech Signal Process.*, 1989, pp. 1878–1881.

[16] C. Uhl and O. Macchi, "When is DPCM a stable system?", *Proc. Internat. Conf. Acoust. Speech Signal Process.*, 1990, pp. 1747–1750.

[17] B. Widrow and S.D. Stearns, *Adaptive Signal Processing*, Prentice Hall, Englewood Cliffs, NJ, 1985, pp. 99–116.