CONVERGENCE PROPERTIES OF SOME ADAPTIVE FILTERING ALGORITHMS

C.C. EVCİ

TRT - 5, Avenue Réaumur 92350 Le Plessis Robinson
FRANCE

RESUME

Le propriétés de convergence d'un algorithme du gradient souple (VLMS) pour un critère de la valeur absolue de l'erreursont analyseés. VLMS vise à minimiser une fonction de l'erreur. Il calcule aussi un estimé du vrai gradient et, par conséquent, est plus performant que l'algorithme du gradient usuel qui minimise la valeur absolue de l'erreur (CLMS). La supériorité de l'algorithme du gradient souple (VLMS) sur l'algorithme usuel (CLMS) pour la prédiction est illustreé par les mesures des gains de prédiction obtenus pour trois séquences de parole différentes.

SUMMARY

The convergence properties of a versatile LMS algorithm (VLMS) for an absolute error criterion are analysed. VLMS attempts to minimize an error function and estimates the true gradient. Therefore, outperforms the usual gradient algorithm that minimizes absolute error (CLMS). The superiority of VLMS over CLMS in a prediction mode is illustrated by prediction gain performance measures for speech sentences. Finally, it is shown that for each coefficient 80% of the time, VLMS equations take the degression from CLMS algorithm.

I. INTRODUCTION

Adaptive filters are exploited in numerous applications, e.g. noise cancellation, linear prediction and echo cancellation in communication systems. The most familiar type for the adaptation of the coefficients is the least-mean-square (LMS) algorithm [1,2] that often applied to nonstationary signals. The filter (FIR type) output for a given input sample x_i in a prediction mode is given by,

$$y_i = \sum_{k=1}^{N} a_k . x_{i-k} . \qquad (1)$$

The number N of coefficients a_k may be selected according to the value of the mean-square prediction error, i.e. $<(x_i - y_i)^2>$, where $<(.)>$ means time average of (.). LMS algorithm when minimizing the mean-square prediction error adjusts the value of the kth coefficient a_k at the (i+1)th instant according to

$$a_{i+1,k} = a_{i,k} + h.e_i.x_{i-k} \qquad (2)$$

where h controls the predictor's rate of convergence and $e_i = x_i - y_i$ is the prediction error at the ith instant. When the absolute error, instead of mean-square error, is minimized, Equation 2 becomes,

$$a_{i+1,k} = a_{i,j} + h.\mathrm{sgn}(e_i).x_{i-k} \qquad (3)$$

where sgn(.) denotes the polarity of (.).

The parameter h is often replaced by $P_i(x)$ where $P_i(x)$ is a function of the power in the speech signal computed over a duration of approximately a pitch period. $P_i(x)$ is expressed by,

$$P_i(x) = \frac{\delta}{\frac{1}{M} \sum_{j=i-1-M}^{i-1} x_j^2 + B} \qquad (4)$$

The denominator of $P_i(x)$ behaves as an automatic gain control which tends to equalize the adaptation rate of prediction algorithm to the variation in the mean square value σ_x^2 of the speech sequence $\{x_j\}$ computed over the immediate past M samples. Thus when σ_x^2 increases the second term in Equation 2 is reduced and over-corrections of the $a_{i+1,k}$ coefficient avoided, preventing the occurrence of a large prediction error. The constant B in Equation 4 maintains a finite value of $P_i(x)$ during silence intervals. It should be mentioned that when Equation 4 is associated with Equation 3, x_j^2 in $P_i(x)$ is usually replaced by $|x_j|$.

The second algorithm (CLMS) defined by Equation 3 is simple for hardware implementation and studied by Cummiskey [3]. Also, in many applications such as ADPCM codec for speech signals, in the presence of transmission errors, CLMS is less affected than the LMS algorithm. This is because the former one uses the polarity of e_i rather than the actual amplitude of e_i as in the LMS algorithm. However, both algorithms defined by Equations 2,3 have their own disadvantages. For example, the first one has a slow convergence rate and deteriorates rapidly its performance for a small change in an adaptation constant h (or δ in Equation 4) The second one has also a slow convergence rate, but in the prediction context, has less multiplications and more robust to the variations in δ despite the small degradation in the peak prediction gain. In order to improve the prediction gain for wider range of δ , we propose versatile-LMS (VLMS) algorithm that is truely instantaneous [4]. The advantages of VLMS are two-fold. First, it attempts to minimize an error function. This function can be dependent on a variety of criteria such as the modulus of the prediction error, the square of this error, the differential of the error. Secondly, as the algorithm at a given instant computes several errors with respect to pre-selected error function and then picks and chooses the minimum error, it measures the true gradient and therefore outperforms the former ones.

In this article, we analyse the convergence properties of VLMS for an absolute error criterion and further, we show that for each coefficient 80% of the time VLMS equations take the degression from the CLMS algorithm, i.e., see Equation 3. In order to support our mathematical analysis computer simulations are carried out for various speech sentences with the predictor order of 12.

II. CONVERGENCE OF THE CLMS ALGORITHM

In this section the sequentially updated CLMS algorithm is compared with a modified autocorrelation algorithm in a prediction mode (MAP). MAP is defined by considering a sliding-block autocorrelation predictor (SBAP) whose coefficients are re-calculated from a block of W_B samples every sampling instant using the autocorrelation method. Although SBAP algorithm can not be used in a practical system because of the amount of side information that must be transmitted, it does achieve more accurate predictions than those of LMS or CLMS, and here it is used as a bench-mark.

Figure 1 depicts the adaptation logic of CLMS algorithm.

At the ith sampling instant, SBAP has a set of coefficients $\hat{A}_s$, and a prediction error, $e_{i,s}$, while CLMS in vector form, represented by

$$\hat{A}_{i+1} = \hat{A}_i + P.\hat{X}_i \,\mathrm{sgn}(e_i) \tag{5}$$

$P_i(x)$ is replaced by P, i.e., will be assumed constant over a small number of sampling intervals. The hat (^) above the symbol means the symbol is a vector, viz.

$$\hat{A}_k^T = [a_1, a_2, a_3, \ldots, a_N]$$

$$\hat{X}_i^T = [x_{i-1}, x_{i-2}, \ldots, x_{i-N}]$$

where the raised T implies transpose of the vector. The predicted output from CLMS in vector form is,

$$y_i = \hat{A}_i^T \hat{X}_i \tag{6}$$

For SBAP algorithm

$$e_{i,s} = x_i - \hat{A}_s^T \hat{X}_i \; . \tag{7}$$

Hence

$$e_{i,s} - e_i = -(\hat{A}_s - \hat{A}_i)^T \hat{X}_i \tag{8}$$

Now, let $\hat{\gamma}_i$ be as a difference vector, $(\hat{A}_i - \hat{A}_s)$ and write Equation 5 as

$$\hat{A}_{i+1} - \hat{A}_s = \hat{A}_i - \hat{A}_s + P.\hat{X}_i \,\mathrm{sgn}(e_i) \tag{9}$$

or

$$\hat{\gamma}_{i+1} = \hat{\gamma}_i + P.\hat{X}_i .\mathrm{sgn}(e_i) \tag{10}$$

then

$$e_i = e_{i,s} - \hat{\gamma}_i^T .\hat{X}_i \tag{11}$$

and substitution of Equation 11 in Equation 10 yields

$$\hat{\gamma}_{i+1} = \hat{\gamma}_i + P.\hat{X}_i \,\mathrm{sgn}(e_{i,s} - \hat{\gamma}_i^T .\hat{X}_i) \tag{12}$$

The norm of Equation 12 is

$$||\hat{\gamma}_{i+1}||^2 = ||\hat{\gamma}_i||^2 + 2P < \hat{\gamma}_i^T \hat{X}_i .\mathrm{sgn}(e_{i,s} - \hat{\gamma}_i^T . \hat{X}_i) > + P^2 < \hat{X}_i^2 \{\mathrm{sgn}(e_i)\}^2 > . \tag{13}$$

For $e_{i,s} < \hat{\gamma}_i^T \hat{X}_i$, the coefficients of CLMS are not near optimum, with P<<1, Equation 13 becomes

$$||\hat{\gamma}_{i+1}||^2 \cong ||\hat{\gamma}_i||^2 + 2P\hat{\gamma}_i^T \hat{X}_i \,\mathrm{sgn}(-\hat{\gamma}_i^T \hat{X}_i) \tag{14}$$

since $z.\mathrm{sgn}(z) = |z|$, Equation 14 is rewritten as

$$||\hat{\gamma}_{i+1}||^2 = ||\hat{\gamma}_i||^2 - 2P|\hat{\gamma}_i^T \hat{X}_i| \tag{15}$$

Finally, it is obvious that $2P|\hat{\gamma}_i^T \hat{X}_i|$ is always positive, hence

$$||\hat{\gamma}_{i+1}||^2 < ||\hat{\gamma}_i||^2 \; . \tag{16}$$

Consequently, CLMS algorithm converges. Also the convergence is slowed to stop when $|\hat{\gamma}_i^T .\hat{X}_i| \to |e_{i,s}|$, i.e. Equation 13-14 becomes

$$||\hat{\gamma}_{i+1}||^2 = ||\hat{\gamma}_i||^2 \tag{17}$$

Figure 2 shows the variation of prediction (dB) as a function of adaptation parameter δ (taken as a power of 2) for three different speech sentences sampled at 8kHz, bandlimited to 3.4kHz. The value of B in Equation 4 is 0.5 and M=100, N=12.

III. VLMS ALGORITHM AND ITS PERFORMANCE

In this section, we describe the versatile LMS algorithm which attempts to minimize an error function G. This function can be made dependent on a variety of criteria such as modulus of the prediction error, the square of this error etc. Thus G is a function of the prediction error, e_i. In VLMS, two values of G are formed every sampling instant in the updating of each predictor coefficients. For an N-order predictor, VLMS calculates 2N values of G every sampling instant. The values of G used in updating each coefficient at the ith instant are based on two estimates of the prediction error. For each coefficient the difference between the appropriate G values is formed as:-

$$\Delta_{i-1,k} = G_{i-1,2k-1} - G_{i-1,2k}, \quad k=1,2,3,\ldots,N. \tag{18}$$

Then coefficient adaptation equation at the ith instant is given by

$$a_{i,k} = a_{i-1,k} - P_{i-1}(x).\Delta_{i-1,k}.k^{-\alpha} \tag{19}$$

The term $k^{-\alpha}$ results in a small modification to the higher order coefficients then to the lower coefficients. Experimental results show that by setting α to a value just less than unity good prediction is obtained. The operation of VLMS algorithm at ith instant is illustrated in Figure 3.

The operation of the predictor employing VLMS algorithm can be explained with the aid of 2nd order predictor (N=2).

The second-order VLMS predictor at ith sampling instant forms four predictions $y_{i-1,1}$, $y_{i-1,2}$, $y_{i-1,3}$ and $y_{i-1,4}$ which are generated from x_{i-2} and x_{i-3}. These intermediary predictions are

$$y_{i-1,1} = (a_{i-1,1}+s_1).x_{i-2} + a_{i-1,2}x_{i-3} \tag{20}$$

$$y_{i-1,2} = (a_{i-1,1}-s_1).x_{i-2} + a_{i-1,2}.x_{i-3} \tag{21}$$

$$y_{i-1,3} = a_{i-1,1}.x_{i-2} + (a_{i-1,2}+s_2).x_{i-3} \tag{22}$$

$$y_{i-1,4} = a_{i-1,1}.x_{i-2} + (a_{i-1,2}-s_2).x_{i-3} \tag{23}$$

where s_1 and s_2 are system parameters $(s_2 < s_1)$ and defined by

$$s_k = [D.k^{\beta}]^{-1} \; , \tag{24}$$

where D>1 and β<1. The values of D and β will be subsequently quoted. Notice that s_1 has been added and subtracted from $a_{i-1,1}$ to yield $y_{i-1,1}$ and $y_{i-1,2}$ while a smaller change of $\pm s_2$ has been made to $a_{i-1,2}$ to give $y_{i-1,3}$ and $y_{i-1,4}$. We select the absolute error criterion in order to compare the results with previously described CLMS algorithm. Therefore, we form the moduli of these prediction errors, viz:

$$G_{i-1,j} = |x_{i-1} - y_{i-1,j}|, \quad j=1,2,3,4 \tag{25}$$

and then compute

$$\Delta_{i-1,1} = G_{i-1,1} - G_{i-1,2} \tag{26}$$

$$\Delta_{i-1,2} = G_{i-1,3} - G_{i-1,4} \tag{27}$$

As $G_{i-1,1}$ and $G_{i-1,2}$ are the moduli of the prediction error when $a_{i-1,1}$ is increased and decreased by s_1 respectively, it follows that if $\Delta_{i-1,1}>0$ then $a_{i-1,1}$ should be decreased and vice versa. Similar remarks apply for $\Delta_{i-1,2}$ and $a_{i-1,2}$. Consequently, the two coefficients in this example are updated according to Equation 19 and predicted sample is found from Equation 1.

The reason for naming VLMS is now apparent. The coefficients are sequentially updated every sampling instant with respect to pre-selected error function and the predictor estimates the gradient of the fastest descent of the error function with time.

Figure 4 shows the computer simulation results for the same speech sentences used in Section II. The system parameters α,β,B,D are 0.5, 0.5, 0.1 and 10.0 respectively. As can be seen in Figure 4, the variation of prediction gain as a function of δ shows more robust behaviour when compared to Figure 2.

The VLMS algorithm defined by Equation 19 converges as well [4]. As before, we assume $P_{i-1}(x)$ is constant over a small number of sampling intervals and ignore the optimizing term $k^{-\alpha}$. Hence

$$\hat{A}_i = \hat{A}_{i-1} - P.\hat{\Delta}_i$$

where $\hat{\Delta}_i$ is the vector representation of Equation 18. However, due to space limitations in this paper, we can not show all the necessary steps, interested readers may refer to reference [4].

In the next section, we are to demonstrate the differences between CLMS and VLMS that minimizes the absolute error, $G=<|e_i|>$. Also this analysis gives insights for the convergence behaviour of VLMS algorithm.

IV. A COMPARATIVE STUDY OF CLMS AND VLMS ALGORITHMS

In many ways, VLMS algorithm behaves like CLMS algorithm. However, in this section we will show that about 80% of the time, the prediction coefficients resulted from VLMS algorithm are different from those calculated from CLMS algorithm. This procedure can be described as follows:
Using Equation 26 together with Equations 20-23 and 25, we get (N=2)

$$\Delta_{i-1,1} = |x_{i-1}-a_{i-1,1}.x_{i-2}-a_{i-1,2}.x_{i-3}-s_1.x_{i-2}|-|x_{i-1}-a_{i-1}.x_{i-2}-a_{i-1,2}.x_{i-3}+s_1x_{i-2}| \qquad (28)$$

and since

$$e_{i-1} = x_{i-1}-a_{i-1,1}.x_{i-2}-a_{i-1,2}.x_{i-3} \qquad (29)$$

Equation 28 is rewritten as

$$\Delta_{i-1,1} = |e_{i-1}-s_1.x_{i-2}|-|e_{i-1}+s_1.x_{i-3}| \qquad (30)$$

Similarly,

$$\Delta_{i-1,2} = |e_{i-1}-s_2.x_{i-3}|-|e_{i-1}+s_2.x_{i-3}| \qquad (31)$$

Letting "w" be e_{i-1} and "v" be $s_1.x_{i-2}$ or $s_2.x_{i-3}$, Equation 30 is rewritten as

$$\Delta_{i-1,1} = |w-v|-|w+v| \qquad (32)$$

Equation 32 can be analysed with the aid of Figure 5. That is,

In REGION I: $|w-v|-|w+v| = -2v$ (33)

In REGION II: $|w-v|-|w+v| = -2w$ (34)

In REGION III: $|w-v|-|w+v| = 2v$ (35)

In REGION IV: $|w-v|-|w+v| = 2w$ (36)

Hence, in REGION I and III, $|w|>|v|$ and

$$|w-v|-|w+v| = -2v\ \mathrm{sgn}(w) \qquad (37)$$

Similarly, in REGIONS II and IV, $|v|>|w|$ and

$$|w-v|-|w+v| = -2w\ \mathrm{sgn}(v). \qquad (38)$$

Thus Equation 30 yields,

$$\Delta_{i-1,1} = \begin{cases} -2s_1x_{i-2}\mathrm{sgn}(e_{i-1}), & \text{if } |e_{i-1}|>|s_1x_{i-2}| \\ -2e_{i-1}\mathrm{sgn}(s_1x_{i-2}), & \text{if } |s_1x_{i-2}|>|e_{i-1}| \end{cases} \qquad (39)$$

while Equation 31 gives,

$$\Delta_{i-1,2} = \begin{cases} -2s_2x_{i-3}\mathrm{sgn}(e_{i-1}), & \text{if } |e_{i-1}|>|s_2x_{i-3}| \\ -2e_{i-1}\mathrm{sgn}(s_2x_{i-3}), & \text{if } |s_2x_{i-3}|>|e_{i-1}|. \end{cases} \qquad (40)$$

Further, substitution of Equations 39 and 40 in Equation 19, provides the adaptation equations of VLMS. When $|e_{i-1}|>|s_1x_{i-2}|$ and $|e_{i-1}|>|s_2x_{i-3}|$, VLMS is defined by (N=2)

$$a_{i,1} = a_{i-1,1}+2P_{i-1}(x).s_1.x_{i-2}.\mathrm{sgn}(e_{i-1}) \qquad (41)$$

$$a_{i,2} = a_{i-1,2}+2P_{i-1}(x).s_2.2^{-\alpha}.x_{i-3}\mathrm{sgn}(e_{i-1}). \qquad (42)$$

Finally, we know that Equation 3 for N=2 corresponds CLMS algorithm and two coefficients at ith instant are given by

$$a_{i,1} = a_{i-1,1}+h.x_{i-2}.\mathrm{sgn}(e_{i-1}) \qquad (43)$$

$$a_{i,2} = a_{i-1,2}+h.x_{i-3}.\mathrm{sgn}(e_{i-1}) \qquad (44)$$

Thus the basic design of the adaptation process of VLMS with absolute error has no significant departure from CLMS algorithm, if $|e_{i-1}|>|s_1.x_{i-2}|$ and $|e_{i-1}|>|s_2.x_{i-3}|$.

However, when the $|s_1.x_{i-2}|>|e_{i-1}|$ or $|s_2.x_{i-3}|>|e_{i-1}|$ inequalities are satisfied, VLMS assumes a very different form. That is

$$a_{i,1} = a_{i-1,1}+2P_{i-1}(x).e_{i-1}\mathrm{sgn}(s_1x_{i-2}) \qquad (45)$$

$$a_{i,2} = a_{i-1,2}+2P_{i-1}(x).2^{-\alpha}.e_{i-1}.\mathrm{sgn}(s_2x_{i-3}). \qquad (46)$$

Equations 45-46 indicate the degression from the CLMS algorithm for k=1,2.

Although, the above analysis is done for the second-order predictor, it is also valid for Nth order predictor. Figure 6 depicts the simulation results for the percentage of diversions of VLMS algorithm from the CLMS algorithm for the same three speech sentences (a,b,c) with N=12.

As is seen in Figure 6, during the adaptation of $a_{i,1}$ for the three sentences, 90% to 80% of the time, diversions from CLMS algorithm are obtained. Obviously, as we approach lower order prediction coefficients, the percentage is decreased (especially for sentence "c"). This is in agreement with experimental observations which support the importance, in the performance of most algorithms, of the first few prediction coefficients.

V. CONCLUSIONS

Our study of VLMS-versatile LMS that minimizes an error function has shown that VLMS for the same error criterion offers better prediction gain (dB) than Cummiskey's LMS (CLMS) algorithm. This is due to the prediction coefficients converging faster to their optimum value in VLMS. Then, the percentage of the departure of VLMS coefficients from the CLMS algorithm has been studied. In further investigations we hope to exploit the versatility of the VLMS to operate with different error functions and compare the performance to recently proposed dual-sign algorithm [5].

REFERENCES

[1] WIDROW, B. and STEARNS, S.D., "Adaptive Signal Processing", Englewood Cliffs: Prentice-Hall,1985.

[2] BELLANGER, M.G., "Digital Processing of Signals-Theory and Practice", John Wiley, London, 1984.

[3] CUMMISKEY, P., "Adaptive DPCM for Speech Processing", Ph.D. Dissertation, Newark College of Engineering, Newark, N.J., 1973.

[4] EVCI, C.C., "Prediction Techniques applied to DPCM for Encoding Speech Signals", Ph.D. Thesis, Loughborough Uni., England, 1982.

[5] KWONG, C.P., "Dual-Sign Algorithm for Adaptive Filtering", IEEE Trans. Vol., COM-34, No.12, Dec.86, pp.1272-1275.

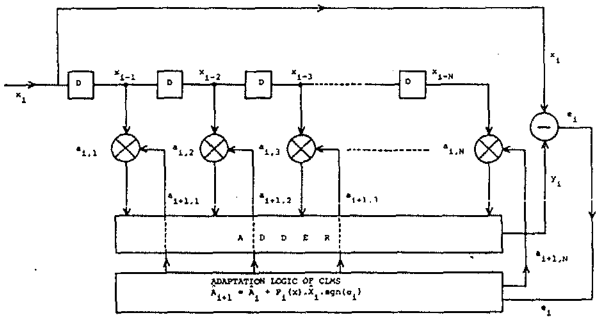

FIGURE 1

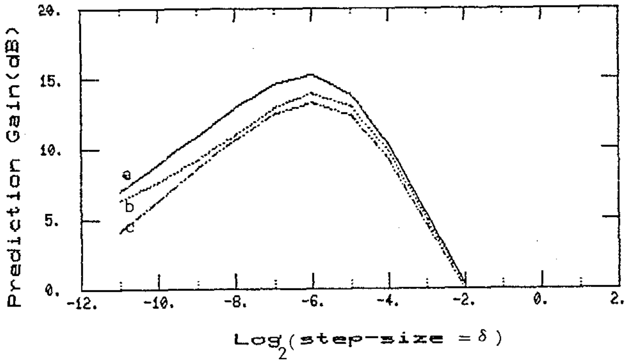

FIGURE 2

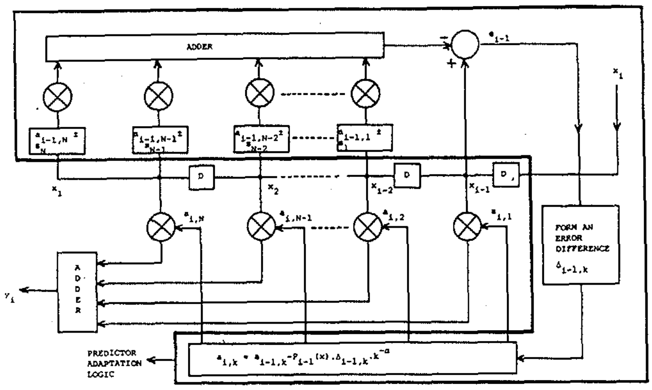

FIGURE 3

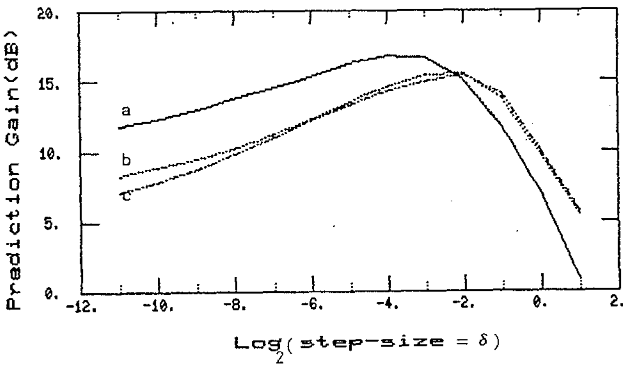

FIGURE 4

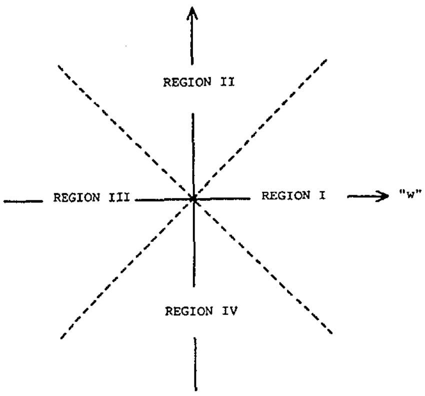

FIGURE 5

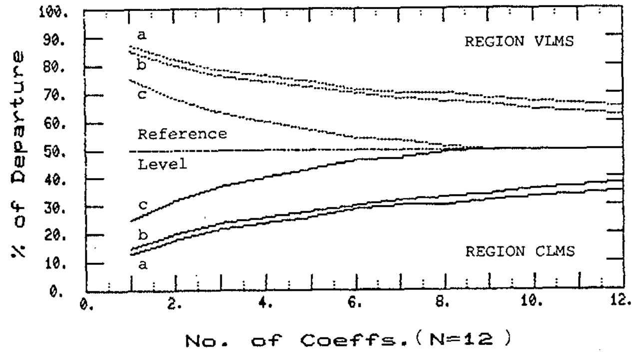

FIGURE 6