

STRUCTURE DE FILTRES A ARITHMÉTIQUE DISTRIBUÉE POUR UNE
 OPTIMISATION EN TERME DE BRUIT NUMÉRIQUE.

V. DEVLAMINCK*, F. WAUQUIER*, M. BENJELLOUN**, J.P. DUBUS*

*Laboratoire de Mesures Automatiques-Université de Lille-Flandres-Artois-VILLENEUVE D'ASCQ

**Institut National des Télécommunications - EVRY

RESUME

Le travail présenté par les auteurs consiste en l'étude d'une nouvelle structure de réalisation de filtres numériques à l'aide de la technique de l'arithmétique distribuée. En s'inspirant d'un article de F.J. TAYLOR (IEEE Transactions on Acoustics, Speech and Signal processing, vol. ASSP-34, Octobre 1986) qui proposait la réalisation, à l'aide de l'arithmétique distribuée, de filtres ayant une représentation dans l'espace d'état sous forme de matrices Compagnon. Les auteurs, proposent une structure permettant la réalisation de filtres à partir de n'importe quel type de représentations dans l'espace d'état. Une analyse du bruit d'arrondi généré dans la structure introduite conduit à l'obtention d'une valeur théorique de la puissance de ce bruit. Ces résultats théoriques sont vérifiés au moyen de simulations pour différents types de filtres du deuxième et troisième ordre. Les résultats montrent que cette structure n'est pas plus bruyante que celles réalisées à l'aide de l'arithmétique classique. L'avantage principal de cette nouvelle structure réside dans la possibilité d'arriver à une optimisation en terme de bruit sans complication de la structure initiale et ceci au moyen de techniques analogues à celles utilisées pour l'optimisation des filtres numériques à arithmétique classique.

SUMMARY

A new digital filter structure is proposed using distributed arithmetic. From a publication of F.J. TAYLOR (IEEE Transactions on Acoustics, Speech and Signal processing, vol. ASSP-34, October 1986) who suggests a digital filter realization with distributed arithmetic for a particular state space representation, the authors give a solution for a new structure usable with any state space representation. An error analysis for this structure is reported and verified by simulations which demonstrates the superiority of the new structure over traditional lumped parameter filters. But the major advantage of this realization is the possibility to minimize roundoff error effects without structure complications.

Introduction

L'approche de la réalisation des filtres numériques par leur représentation dans l'espace d'état présente de nombreux avantages. Parmi ceux-ci, on peut citer notamment les possibilités d'optimisation des filtres en terme de bruit. Cette optimisation se fait par la recherche d'une transformation T non singulière qui à un ensemble de matrices d'état (A, B, C, D) fait correspondre les matrices (A', B', C', D') avec

les relations : $A' = T^{-1} A T$, $B' = T^{-1} B$, $C' = C T$ et $D' = D$ de manière à laisser invariante la fonction de transfert du filtre. Les nouvelles matrices d'état devant correspondre à une structure de filtre à bruit minimal. Le problème principal vient alors du fait que la structure optimale en terme de bruit est en général beaucoup plus compliquée que la structure de départ, ceci du moins lorsque l'on réalise le filtre à l'aide d'une arithmétique classique utilisant des opérateurs



arithmétiques "localisés". L'utilisation de l'arithmétique distribuée va permettre de trouver une solution à ce dilemme.

I - Rappel sur l'arithmétique distribuée

Soit donc une fonction de filtrage numérique définie par la relation suivante :

$$y(n) = \sum_{i=0}^L a_i x(n-i) + \sum_{j=1}^L b_j y(n-j) \quad (1)$$

x et y étant les séquences respectives d'entrée et de sortie du filtre. En supposant que tous les nombres sont pris dans leur représentation en complément à 2 sur un format de B bits, on adaptera la notation suivante : $V(i,n) \in \{0,1\}$ ième bit de la variable V à l'instant nT. On définit alors une fonction ϕ par :

$$\phi_n(i) = \sum_{j=1}^L a_j x(i,n-j) + \sum_{j=1}^L b_j y(i,n-j) \quad (2)$$

Il devient alors possible de ré-écrire la relation (1) sous la forme :

$$y(n) = -\phi_n(0) + \sum_{i=1}^{B-1} \phi_n(i) 2^{-i} \quad (3)$$

(La distinction en $i = 0$ venant du fait qu'il s'agit du bit de signe en complément à 2). Il est donc possible de pré-calculer toutes les valeurs de la fonction ϕ et de les stocker dans une table mémoire. On obtient alors une réalisation de la fonction de filtrage sous forme d'un filtre à arithmétique distribuée (FAD) dont le schéma est donné en figure 1.

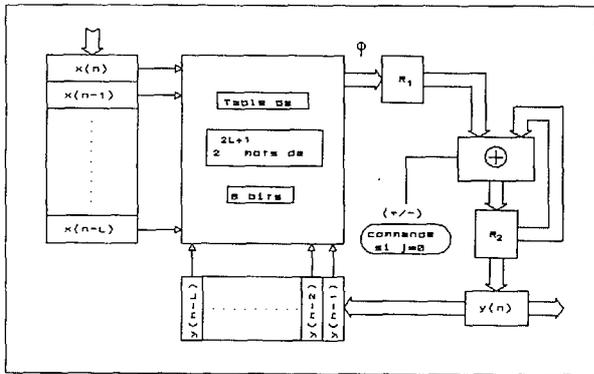


Figure 1 : Structure de filtre à arithmétique distribuée.

II - Structure FAD "direct II"

En s'inspirant des travaux de [1] et en utilisant le fait que la représentation dans l'espace d'état des structures de filtres dites directes d'ordre L conduit à des matrices d'état de type Compagnon, F.J. TAYLOR [2] a proposé un autre type de structures de FAD. Pour cela, il réalise à l'aide de 2 tables-mémoires d'une part l'équation d'état du système et d'autre part l'équation de sortie en remarquant que dans ce type de représentation, les variables d'état vérifient : $x_i(n+1) = x_{i+1}(n)$. Ceci conduit alors au schéma de filtre donné en figure 2.

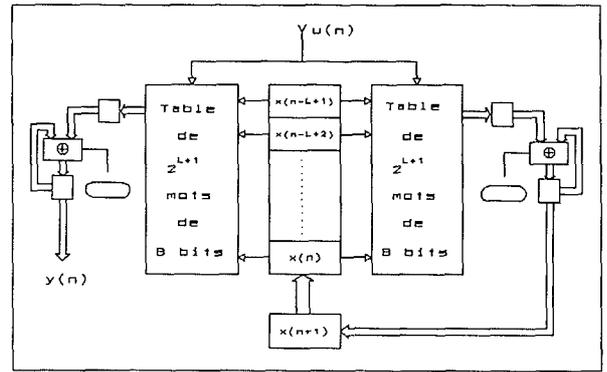


Figure 2 : Structure directe à arithmétique distribuée

III - Structure FAD pour une représentation quelconque dans l'espace d'état

La structure précédente n'est utilisable que pour un type particulier de matrice d'état. Elle n'est donc pas optimisable en terme de bruit. Pour cela, il faut une structure permettant la réalisation d'un filtre ayant une représentation quelconque dans l'espace d'état. Une telle structure est représentée en figure 3. Elle est obtenue en réalisant d'une part les équations d'état (E1 à EL) et d'autre part l'équation de sortie (ES) telles qu'elles sont données ci-après et ceci au moyen de fonctions ϕ définies de façon identique à ce qui a été vu précédemment et auxquelles sont associées les tables mémoires correspondantes.

$$\begin{aligned} x_1(n+1) &= \sum_{i=1}^L a_{1i} \cdot x_i(n) + b_1 \cdot u(n) & (E1) \\ &\vdots & \vdots \\ x_k(n+1) &= \sum_{i=1}^L a_{ki} \cdot x_i(n) + b_k \cdot u(n) & (Ek) \\ &\vdots & \vdots \\ x_L(n+1) &= \sum_{i=1}^L a_{Li} \cdot x_i(n) + b_L \cdot u(n) & (EL) \\ y(n) &= \sum_{i=1}^L c_i \cdot x_i(n) + d \cdot u(n) & (ES) \end{aligned}$$

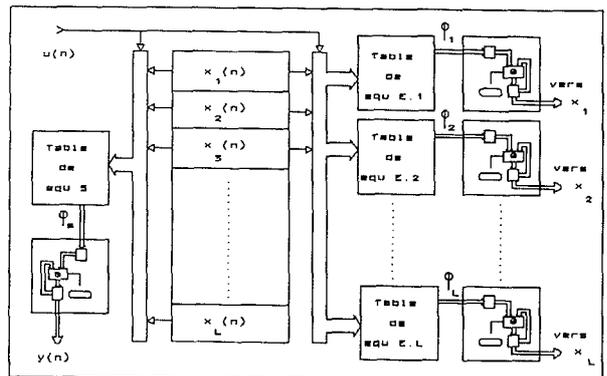


Figure 3 : Structure de FAD proposée



Au total, la complexité de la structure ne dépend que d'un seul paramètre, l'ordre du filtre. Un changement des matrices d'état du système se traduit uniquement par un changement du contenu des tables mémoires et ceci sans aucune modification de la structure.

IV - Analyse du bruit dans la structure proposée

Considérons d'abord les L équations d'état du système correspondant à la partie récursive de la structure. Il est possible d'identifier 3 sources de bruits (une variable surlignée signifie une valeur réellement calculée par rapport à la valeur théorique) :

$e_k^i = \bar{x}_k - x_k$ erreur sur les variables d'état

$e_k^u = \bar{\phi}(\bar{i}) - \phi(i)$ erreur sur le résultat de ϕ

$e_k^{u''}$ erreur en sortie des registres à décalage, vous avons pour la k^{ème} variable d'état en théorie

$$x_k(n+1) = \sum_{i=1}^{B-1} \phi_{nk}(i) \cdot 2^{-i} - \phi_{nk}(0) \quad (4)$$

Mais en réalité on calcule :

$$x_k(n+1) = \sum_{i=1}^{B-1} \phi_{nk}(i) \cdot 2^{-i} - \phi_{nk}(0) \quad (5)$$

On montre alors [3] que l'erreur correspondante peut s'exprimer sous la forme :

$$e_k^i(n+1) = \sum_{i=1}^L a_{ki} e_i^i(n) + \delta_k(n) \quad (6)$$

où les a_{ij} sont les termes de la matrice d'état A et

$$\delta_k(n) = e_k^{u''}(n) + \sum_{i=1}^{B-1} e_i^u(n) \cdot 2^{-i} + e_0^u(n) \quad (7)$$

Après transformation en z, il est possible de ré-écrire la relation (6) sous la forme : (8)

$$e^i(z) = \sum_{i \neq k} \frac{a_{ki} z^{-1}}{1 - a_{kk} z^{-1}} e^i(z) + \frac{z^{-1}}{1 - a_{kk} z^{-1}} \delta_k(z)$$

On obtient alors un système de L équations à L inconnues que l'on peut résoudre par rapport aux δ_k . On arrive donc à des équations du type :

$$e_k^i(z) = \sum_{i=1}^L H_{ki}(z) \cdot \delta_i(z) \quad (9)$$

Il reste maintenant à traiter l'équation de sortie de manière à obtenir l'expression du bruit $e(n)$ en sortie du filtre. De façon analogue à ce qui précède, on montre que $e(n)$ donne après transformation en z :

$$e(z) = \sum_{i=1}^L c_i e_i^i(z) + \delta(z) \quad (10)$$

où les c_i sont les termes de la matrice d'état C.

On peut alors remplacer les $e^i(z)$ par leur expression (9) ce qui conduit à :

$$e(z) = \sum_{i=1}^L T_i(z) \cdot \delta_i(z) + \delta(z) \quad (11)$$

soit encore sous forme matricielle :

$$e(z) = C^T(zI - A)^{-1} \cdot R + d(z) \quad (12)$$

avec $R = [\delta_1(z) \delta_2(z) \delta_3(z) \dots \delta_L(z)]^T$

Comme il est possible par ailleurs, de montrer que sous réserve que B ne soit pas trop petit ($B \gg 8$), la variance de tous les $\delta_i(z)$ est donnée par la relation (indépendante de i) :

$$\Delta^2 = \frac{7}{3} \cdot \frac{q^2}{12} \quad (q \text{ quantum de base}) \quad (13)$$

on arrive alors à une expression de la puissance du bruit en sortie du filtre qui est donnée par la relation suivante (on admet de plus que toutes les variables δ_i sont indépendantes) :

$$\sigma^2 = \sum_{i=1}^L \|T_i(z)\|^2 \cdot \Delta^2 + \Delta^2 \quad (14)$$

V - Résultats

Les résultats théoriques ont été vérifiés par simulation pour différents types de filtres du deuxième et troisième ordre. A titre d'exemple, nous donnons figure 4 le résultat de ces comparaisons pour un filtre du deuxième ordre réalisé au moyen de la structure proposée (la réalisation correspondante au moyen d'opérateurs localisés est également mentionnée dans la figure).

Les résultats montrent une bonne concordance entre la pratique et la théorie. On peut montrer également que la structure proposée n'est pas plus bruyante que la structure classique correspondante. Mais l'avantage principal de cette structure tient du fait qu'une optimisation en terme de bruit s'obtient sans aucune complication de la structure ce qui ne serait pas le cas pour une structure classique. Cette optimisation pouvant être obtenue grâce à des techniques analogues à celles énoncées précédemment à savoir par la recherche d'une transformation T non singulière qui laisse invariante la fonction de transfert du filtre et minimise l'expression (14).

Conclusion

L'étude présentée ici montre l'intérêt que peut avoir la structure proposée dans le cadre de l'optimisation des filtres en terme de bruit. Seul le cas du filtrage mono-dimensionnel a été abordé. Une étude est en cours dans le cadre du filtrage multidimensionnel en vue d'une application au traitement des images. En effet, dans ce domaine, la complexité arithmétique est particulièrement cruciale compte tenu des cadences d'échantillonnage et l'optimisation des structures prend alors tout son sens.

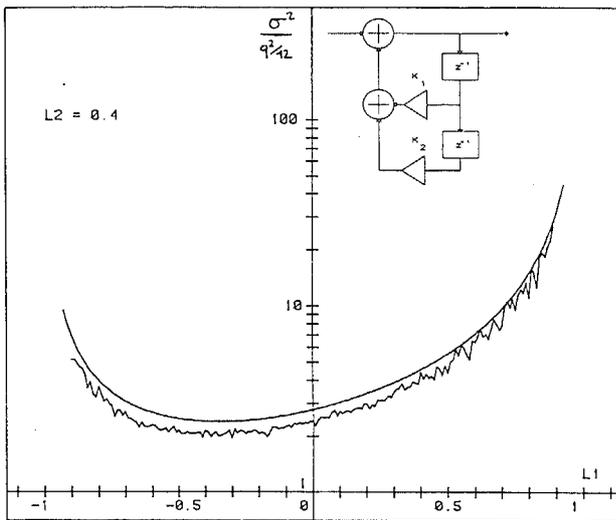


Figure 4 : Puissance réduite de bruit en fonction du pôle L_1 pour $B = 16$ et $L_2 = 0,4$. Valeurs théorique et simulées pour un filtre qui aurait une réalisation classique du type représenté.

Bibliographie

- [1] - A. PELED et B. LIU - "A new hardware realization of digital filters". IEEE Transac. on acoustics Speech and Signal Processing, vol. ASSP-22, Décembre 1974.
- [2] - F.J. TAYLOR - "An analysis of distributed arithmetic digital filter". IEEE Transac. on acoustics Speech and Signal Processing, vol. ASSP-34 octobre 1986.
- [3] - V. DEVLAMINCK - "Etude des bruits numériques dans les structures de filtres à coefficients fixes. Applications à l'arithmétique classique et à l'arithmétique distribuée". Thèse de doctorat en électronique. Université des Sciences et Techniques de Lille-Flandres-Artois. Décembre 1988.