IIR Structure and Its Application

Er. Rajeev Ranjan

(Assistant Professor, Department of Electrical Engineering, Womens Institute of Technology,
L. N. Mithila University, Darbhanga, India)
(A part of MSc.(Engg.) Research under guidance of Prof. U. C. Verma,
Retd. H.O.D. Deptt. of Electrical Engg. MIT, Muzaffarpur)

ABSTRACT : *The proposed research work on the topic Infinite Impulse Response "IIR Structure and Its Application" is basically state space representation for the digital filter of different form obtained from analog filter and transformation to digital one. In the normal course of analysis the digital filter transfer function are realized in different forms using time delay elements and multipliers.*

This realization can be eased if digital filter are represented in state space techniques opening a new field for computation analysis. This aspect has been incorporated in the proposed research work things outlining the introduction of IIR filter and different structure including ladder and wave structure realization, state space description of IIR filter is presented in detail including the normal form which has complex structure but simple form of state equation.

IIR filter design is based on state space technique using Lattice structure. The generalized technique of state space equation realized from general form of digital filter transfer function has been developed then the technique is demonstrated using differential form of various filter like LPF, HPF and BPF for the required order of analog filters. Statistical analysis of digital filter like round off error and dynamic scaling has been incorporated, where the two direct computes have been lumped together in the common state space representation. Autocorrelation and uncorrelation between noise and input samples gives round off noise and dynamic scaling give expression for the factor forms, round off error and dynamic range. This aspect on statistical analysis of digital filter has opened a new field for research.

Keywords *–correlation, filter realization, Infinite Impulse Response, ladder, lattice, noise, wave.*

1 INTRODUCTION

Digital filter got its significance when it was discovered as less complex filter circuit with great flexibility compared to that of analog filter. Digital filter has less complex structure. Also simple alternation of its coefficients can change cut off frequencies of a designed filter as per requirement. This flexibility does not hold for analog filter.

Digital filter has two basic structure namely

1. FIR and
2. IIR

Among these IIR requires stores of large data due to long stream of input required. However this can be realized with linear phase character. IIR filter has recursive in nature which is capable of generating outputs for inputs recursively. These have comparatively less complicated structure.

IIR filter can be converted to algebraic form in discrete time domain. This has opened in new chapter in digital filter designed as state variable approach. This aspect of filter design forms the basis for proposed thesis. This approach is used for various structures of digital filter like direct structure, cascade and parallel structure etc.

Round off noise and dynamic range are two important aspects in filter design. These can also be incorporated in state variable approach.

The proposed thesis begins with brief study of IIR digital filter structures like direct realization, cascade realization, ladder realization and some other structures as well.

State variable output in digital filter design has been described for all the structure discussed together with state variable approach for IIR filter which are complex in nature. This approach has been developed for all forms of filter discussed earlier. Lattice structure is also discussed here. This structure gives very complex filter. A band pass digital filter has been designed as an illustration of state space techniques at the end.

Digital filter designed remains incomplete without analysis of round of noise and dynamic range of this filter coefficient. These problems are discussed as noise by various authors [4-6]. This aspect of digital filter design has been incorporated in state variable approach. In normal method of filter design these two aspects of noise are analyzed separately but in state variable approach both are lumped in the state equation for filter.

2 INFINITE IMPULSE RESPONSE (IIR) DIGITAL FILTER

A digital filter (DFT) selects the required frequency components in a given discrete Signal and rejects the unwanted frequency. Its output is also discrete signal with selected frequency components. Thus a digital filter has input and output signals both discrete in nature (fig. 2.1)

Fig. 2.1

A time invariant digital filter has internal parameters not varying with time. For initial relaxed such system filter will have input and output relationship as

$\Re$ x(nT – kT) = y(nT–kT) …..(2.1)

For all possible excitation where $\Re$ denotes filter operation and for initially relaxed system

x(nT) = y(nT) = 0 for all n < 0

A digital filter characterized as

y(nT) = $\Re$ [x(nT)] = 2nT x(nT)

is not time invariant as

$\Re$ [x(nT–kT)] = 2nT [x(nT–kT)] ≠ y(nT–kT)

A digital filter on the other hand characterized as

y(nT)= $\Re$ [x(nT)=12x(nT–T)+11x(nT–2T)………..(2.2)

represents time invariant digital filter.

On the similar steps properties of causality and linearity can be defined for digital filter.

From equation (2.1) and (2.2) It can be seen that digital filter can be realized with delay blocks. Unit limit time delay refers to one sampling time delay or some time called one clock delay.

2.1 Characterization of digital filter

Analog filters are characterized in terms of differential equation. Digital filters on the other hand are characterized by difference equation. Two types of digital filter can be identified as 1) Recursive and 2) Non recursive

2.1.1 Recursive digital filter

In this digital filter output at any instant of time depends on present and past inputs as well as on past outputs. General from of equation for such filter can be expressed as

$$y(nT) = \sum_{i=0}^{N} a_i x(nT - iT) - \sum_{i=1}^{N} b_i y(nT - iT)$$
…..(2.3)

Such equation can be used to generate outputs from present input and past records of input/output. Evidently such systems have to be system with memory. Simulation of such filter will have delay blocks in the feedback path as well. Simulation of equation 2.3 can be expressed in block diagram as in Fig 2.2

2.1.2 Non Recursive filter

Such filter response does not require records of past output. Evidently its output will be linear combination of present and past inputs only as

$$\left[y(nT) = \sum_{i=0}^{N} a_i x(nT - iT) \right]$$
…….(2.4)

This will denote Nth order digital filter It will have only feed forward path. This can be seen from simulation of equation 2.4 in block diagram in fig. 2.3

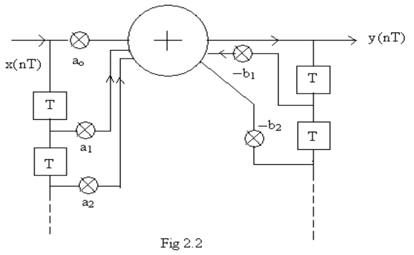


Fig 2.2

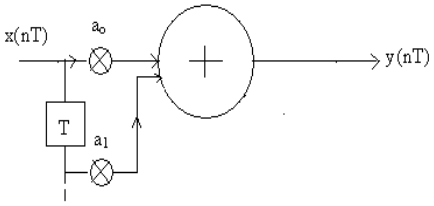


Fig 2.3

2.2 Units impulse response of a digital filter

Digital filter response to unit input can be obtained from Z-transform techniques since filter transfer function is in z domain as

$$H(z) \triangleq \frac{\mathcal{Z}[y(nT)]}{\mathcal{Z}[x(nT)]} = \frac{Y(z)}{X(z)}$$
………..(2.5)

Ƶ Where denotes z transformation

For unit impulse input

δ(nT) = 1 for n = 0
= 0 elsewhere ………..(2.6)

It’s z-transform is unity.i.e. X(z) = 1

Response of filter then becomes

Y(z) = H(z) …………(2.7)

Thus unit impulse response will be simply Z^{-1} of transfer function of H(z) as h(nT)

As an illustration

$$H(z) = \frac{z^2 - z + 1}{(z - p_1)(z - p_2)} \quad ……..(2.8)$$

Will give impulse response after partial fraction for $\frac{H(z)}{z}$ as

$$h(n) = \frac{1}{p_1 p_2}\delta(n) + \frac{(p_1{}^2 - p_1 + 1)}{p_1(p_1 - p_2)}(p_1)^n + \frac{(p_2{}^2 - p_2 + 1)}{p_2(p_2 - p_1)}(p_2)^n$$
……(2.9)

2.3 Digital filter realization

For this realization unit time’s delay in time domain is represented as Z^{-1} in Z domain. This concept is used in digital filter realization. Digital filter realization techniques can be classified according to its structure as

1. Direct 2. Direct canonic 3. Cascade

4. Prallel 5. Ladder 6. Wave structure.

2.3.1 Direct realization

This realization directly implement the filter transfer function H(z) in z-domain. This can be seen from fig 2.2 and fig. 2.3

As simple demonstration example

$$H(z) = \frac{a_o + a_1 z^{-1}}{1 + b_1 z^{-1} + b_2 z^{-2}} \quad ………..(2-10)$$

Can be considered whose direct realization is obtained from

$$H(z) = \frac{Y(z)}{X(z)} = \frac{a_o + a_1 z^{-1}}{1 + b_1 z^{-1} + b_2 z^{-2}}$$

Which gives

$$Y(z) = a_o X(z) + a_1 z^{-1} X(z) - b_1 z^{-1} X(z) - b_2 z^{-2} Y(z)$$
…(2.11)

This gives structure for direct realization as in Fig. 2.4

2.3.2 Direct Canonic form

Direct realization is said to be canonic if number of delay elements is equal to the order of transfer function. In fig. 2.4 number of delay used for numerator is one which for denominator is two confirming the order of numerator and denominator of polynomial of order 2.

2.3.3 Cascade structure

This structure is called simplest structure for high order transfer function of digital filter. In this realization high order digital filter transfer function is factorized into first and second order section. Then each such section is realized separately for their cascade connection. Transfer function H(z) of high order is expressed into M cascade section as

$$H(z) = \prod_{i=1}^{M} H_i(z)$$
……..(2.11)

where

$$H_i(z) = \frac{a_{oi} + a_{1i} z^{-1} + a_{2i} z^{-2}}{1 + b_{1i} z^{-1} + b_{2i} z^{-2}}$$
……..(2.12)

This gives i'th section as

$$Y(z) = a_{oi} X(z) + a_{1i} z^{-1} X(z) + a_{2i} z^{-2} X(z) - b_{1i} z^{-1} Y(z) - b_{2i} z^{-2} Y(z)$$
…..(2.13)

This gives realization structure as in fig. 2.5

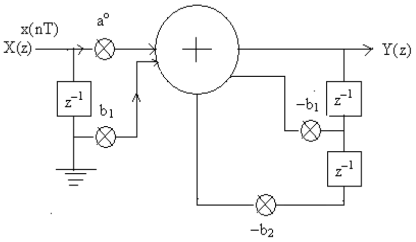


Fig 2.4 Direct realization

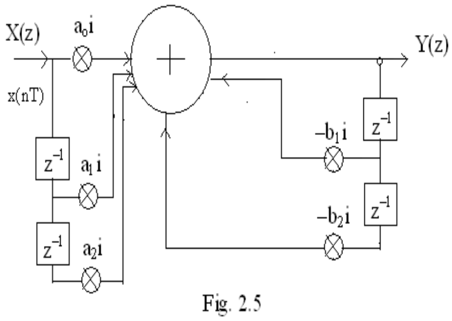


Fig. 2.5

Filter section in Cascaded structure.

2.3.4 Parallel realization

In this realization transfer function H(z) in expressed as summation of several (M sections) H(z) as

$$H(z) = \sum_{i=1}^{M} H_i(z)$$

………(2-14)

It's structure is shown in fig. 2.6

2.3.5 Ladder realization

In these techniques continued fraction can be used to realize this structure. Transfer function H(z) can be written as

$$\frac{Y(z)}{X(z)} = H(z) = \frac{H_2(z)}{1+H_1(z)}$$

Then Y(z) + Y(z) H$_1$(z) = X(z) H$_2$(z)

Y(z) = –Y(z) H$_1$(z) + X(z) H$_2$(z)

= Y^1(z) H$_1$(z) + X(z) H$_2$(z) …….. (2.15)

WhereY^1(z) = –Y(z)

Transfer function H$_1$(z) will be of the form,

$$H_1(z) = \frac{N_1(z)}{D_1(z)} = \frac{m_4 m_3 m_2 z^3 + (m_4 + m_2)z}{m_4 m_3 m_2 m_1 z^4 + (m_4 m_3 + m_4 m_1 + m_2 m_1)z^2 + 1}$$

In Continued fraction form

$$H_1(z) = \cfrac{1}{m_1 z + \cfrac{1}{m_2 z + \cfrac{1}{m_3 z + \cfrac{1}{m_4 z}}}}$$

…… (2.16)

$$H_2 z = \frac{(-1)^K}{D_1(z)}$$

…..(2.17)

Where K is integer having highest value $\frac{N}{2}$ $or < \frac{N}{2}$, thus

$$(-1)^K = \begin{cases} +1 \; for \; K = 1,4,5,8,9,.......... \\ -1 \; for \; K = 2,3,6,7,............ \end{cases}$$

Consider for example a transfer function

$$H(z) = \frac{10^{-2}\left(-3.517 + 0.665z + 0.665z^2 - 3.517z^3\right)}{1 - 3.266z + 3.739z^2 - 1.53z^3}$$

for ladder structure realization, in this H(z) both numerator and denominator have same highest order of z. A constant is therefore taken out to reduce highest part of z in numerator that is,

$$H(z) = 0.2299 + \frac{\left(-0.0582 + 0.0817z - 0.0793z^2\right)}{\left(1 - 3.266z + 3.739z^2 - 1.53z^3\right)}$$

Expressing numerator N(z) = $\sum_i a_i z^i$

gives a$_o$ = –0.0582, a$_1$ = 0.0817, a$_2$ = –0.0723

Again expressing N(z) = $\sum_i d_i \; n_i(z)$

Gives

n$_1$(z) = –1

n$_2$(z) = –(c$_1$z+1)

n$_3$(z) = c$_1$c$_2$z^2+c$_2$z+1

Now $H_N(z) = \frac{-0.0582 + 0.0817z - 0.0793z^2}{1 - 3.266z + 3.739z^2 - 1.53z^3} = \frac{H_2(z)}{1+H_1(z)}$

Denominator

$1–3.266z +3.739z^2 –1.53z$

$$=(1.53z^3+3.266z)\left[1-\frac{1+3.739z^2}{(1.53z^3+3.266z)}\right]$$

$$H_1(z)=-\frac{(3.739z^2+1)}{(1.53z^3+3.266z)}$$

$$=\cfrac{1}{-0.4092z+\cfrac{1}{-1.309z+\cfrac{1}{-2.856z}}}$$

Thus $m_1=c_1=–0.4092,\ m_2=c_2=-1.309$

and $m_3=c_3=–2.856$

Here N = 3 then

$N_1(z) = –1$

$n_2(z) = (c_1z+1)$

$n_2(z) = c_1c_2z^2 + c_2z+1$

Coefficients d_1, d_2 and d_3 are given by

$$\begin{bmatrix}(c_1c_2) & 0 & 0\\ c_2 & -c_1 & 0\\ 1 & -1 & 1\end{bmatrix}\begin{bmatrix}d_3\\ d_2\\ d_1\end{bmatrix}=\begin{bmatrix}a_2\\ a_1\\ a_o\end{bmatrix}$$

$$d_3=\frac{a_2}{c_1c_2}=\frac{-0.0793}{0.4092\times1.309}=–0.148$$

$$d_2=\frac{d_3c_3-d_1}{c_1}=–0.274,\ d_1=d_3-d_2-a_o=0.184$$

With evaluation of these coefficient the ladder structure can be drawn as shown in fig 2.7

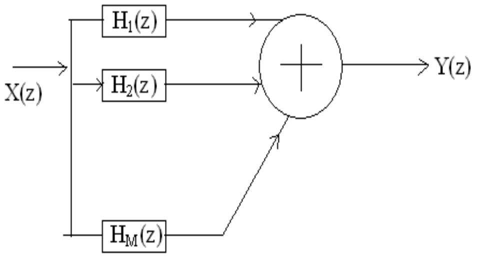


Fig. 2-6 Parllel realization

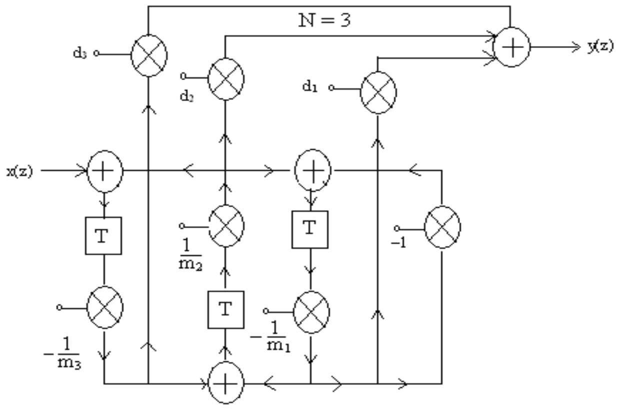


Fig 2.7 Ladder Realization (N =3)

2.3.6 Wave realization

Wave filter in case of analog signal use resonant cavity represented by equivalent LC tuned circuit. If these elements could be simulated through digital circuits then the along wave filter becomes digital wave filter for this purpose bilinear transformation.

$$S=\left(\frac{2}{T}\right)\left(\frac{z-1}{z+1}\right) \quad\text{.........(2.18)}$$

will be very useful.

Detailed analysis of this structure will be taken in next chapter. Where conversion table will be established as in table 2.1 these digital filter structures are derived from state space analysis of digital filter described in detail in next unit.

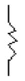
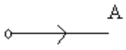
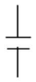
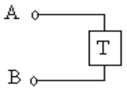
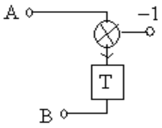
Element	R	Restructure
R 	R	 A
C 	$\frac{T}{2C}$	 A T B
L	$\frac{2L}{T}$	 A −1 T B

Table 2.1 Conversion Table for wave filter

3 STATE DESCRIPTION OF I I R FILTER

IIR filter called recursive filter can be represented by recursive difference equations. These equation together form state equation in matrix form. This form of IIR filter is derived from generalized form of filter transfer function as

$$H(z) = \frac{Y(z)}{X(z)} = \frac{\sum_{i=0}^{N} a_i z^{-i}}{1+\sum_{i=1}^{N} b_i z^{-i}}$$

……(3.1)

3.1 State equation derivation

IIR filter transfer function in (3.1) can be expressed as linear difference equation of high order (Nth order) as

$$y(n) = \sum_{i=0}^{N} a_i\ x(n-i) - \sum_{i=1}^{N} b_i y(n-i)$$

…….(3.2)

Where delay i denotes i'th sample period delay (= iT)

Equation (3.1) can be expressed in cascaded form as

$$Y(z) = X(z)\left[\sum_{i=0}^{N} a_i z^{-i}\right] \times \frac{1}{\left[1+\sum_{i=1}^{N} b_i z^{-i}\right]}$$

If first block has the response v(n) as shown in fig 3.1

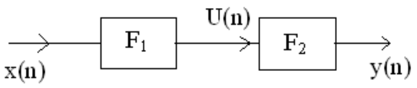


Fig. 3.1 Cascaded form

Then, $v(n) = \sum_{i=0}^{N} a_i x(n-i)$

…..(3.3)

and $y(n) = v(n) - \sum_{i=0}^{N} b_i\ y(n-i)$

Now assume that x(n) is applied directly to F_2 then response $y^1(n)$ is given by

$$y'(n) = x(n) - \sum_{i=1}^{N} b_i\ y'(n-i)$$

…….(3.4)

With variable $x_1(n)$, $x_2(n)$, ……………$x_N(n)$ defined as

$x_1(n) = y'(n–N)$ ………..(3.5)

$x_2(n) = y'(n–N+1)$ ………..(3.6)

………………

$x_N(n) = y'(n–1)$ …………(3.7)

Then state equations can be expressed as

$x_1\ (n+1) = y'\ (n–N+1) = x_2(n)$ ………..(3.8)

$x_2\ (n+1) = y' = (n–N+2) = x_3(n)$ ………..(3.9)

………………………………………………………
……

$x_{N–1}(n+1) = y'\ (n-1) = x_N(n)$ …………(3.10)

and finally

$x_N(n+1) = y'(n) = x(n) – b_1\ y'(n-1) –b_2\ y'(n–2)$

……………$b_N\ y^1(n–N)$

$= x(n) – b_1\ x_{N(n)} –b_2\ x_{N–1}(n)………b_N x_1(n)$

They can be expressed in matrix form as……….(3.11)

$$\begin{bmatrix} x_1\ (n+1) \\ x_2\ (n+1) \\ x_2\ (n+1) \\ x_N\ (n+1) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & & …….0 \\ 0 & 0 & 1 & 0 & ……..0 \\ 0 & 0 & 1 & 0 & ……..0 \\ -b_N & -b_{N-1} & & & ……-b_1 \end{bmatrix} \begin{bmatrix} x_1(n) \\ x_2(n) \\ x_2(n) \\ x_N(n) \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ x(n) \end{bmatrix}$$

Now response y(nT) from F_2 block due to input V(n) is given by,

$$y(n) = \Re_2 V(n) = \Re_2 \sum_{i=0}^{N} a_i x(n-i)$$

$$= \sum_{i=0}^{N} a_i \Re_2 x(n-i)$$

………(3.12)

$$= \sum_{i=0}^{N} a_i y'(n-i)$$

Where $\Re_2$ is filter operation from block F_2. Using (3.5) to (3.7) equation (3.12) becomes

$y(n) = a_o\ x(n) + c_1x_1(n) + ………+ c_N\ x_N\ (n)$ ……….(3.13)

Where, $c_1 = a_N – a_ob_N$

$c_2 = a_{N–1} – a_ob_{N–1}$

………………………

$c_N = a_1 – a_o – a_ob_1$ ……………..(3.14)

Then complete response is given by

$y(n) = [X]\ [c]$ ………..(3.15)

Thus a digital filter in general can be characterized by state and input equations as

$\underline{X}(n+1) = A\ \underline{X}\ (n) + B\ y(n)$ ……….(3.16)

$Y(n) = C\ \underline{X}\ (n) + D\ x(n)$ ……….(3.17)

Where $\underline{X}\ (n) = [x_{1(n)}……….x_N(n)]^T$

A is state matrix

B is input vector for single input x(n)

C is output matrix in terms of constant C_1, C_2 ………..

D is input vector for output equation.

As an illustration a digital filter characterized by

$y(n) = 1.5\ x(n) + 2x(n) + 0.5\ x\ (n-2) –0.5y(n–1) + 0.25\ y(n-2)$

Gives state space representation with

$$A = \begin{bmatrix} 0 & 1 \\ 0.25 & -0.5 \end{bmatrix},\ B = \begin{bmatrix} 0 \\ 1 \end{bmatrix},\ C = \begin{bmatrix} \frac{7}{8} & \frac{5}{4} \end{bmatrix},$$
$$D = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$

3.2 Direct form realization from state space representation

Output y(n) can be expressed in terms of state variables as

$y(n) = a_N\ x_1(n) + a_{N–1}\ x_2(n) + ………+a_1x_N(n) + a_ox_N(n+1)$ …..(3.18)

$x_n\ (n+1) = –b_N\ x_1(n) –b_{N–1}\ x_2(n)\ …….–b_1\ x_N(n) + x(n)$…(3.19)

These two equations give direct form of filter realization as shown in Fig 3.2 This is in canonic form due to number of delays equal to order of filter transfer function.

3.3 Cascaded form

In this form each block is second order or first order filter with state equation of the form

$x_1(n+1) = x_2(n)$

$x_2(n+1) = –b_2\ x_1(n) – b_1x_2(n) + x(n)$

output equation is of the form

$y(n) = a_2x_1(n) + a_1x_2(n) +a_ox_2(n+1)$

These equations give filter block diagram for each section as shown in fig. 3.3

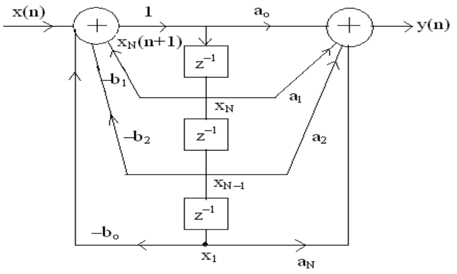


Fig. 3.2 Direct form of realization

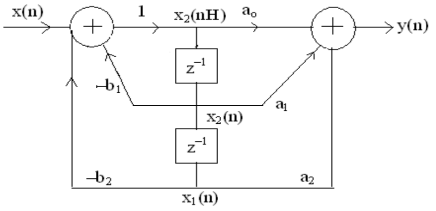


Fig 3.3 Filter block diagram for each casdade section

3.4 Parallel form

In this form the denominator is factorized into product of quadratic and first order expression of z for the purpose of partial fraction and each section is realized. The sum of these section outputs give the final output (fig 3.4) Each factored block will have transfer function of the form

$$\frac{Y_i(z)}{X(z)} = \frac{a_o + a_1 z^{-1}}{1 + b_1 z^{-1} + b_2 z^{-2}} \qquad \text{........(3.20)}$$

This can be realized in terms of state variable $x_1(n)$ and $x_2(n)$ as

$$\begin{bmatrix} x_1(n+1) \\ x_2(n+1) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -b_2 & -b_1 \end{bmatrix} \begin{bmatrix} x_1(n) \\ x_2(n) \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} x(n)$$

........(3.21)

Output equation then becomes

$y(n) = a_1x_2(n) + a_ox_2(n+1)$ ....(3.22)

Substituting for $x_2(n+1)$ from (3.21) gives

$y(n) = a_1x_2(n) + a_o[-b_2x_1(n)-b_1x_2(n)+x(n)]$

$= (-a_ob_2)\, x_1(n)+(a_1-a_ob_1)x_2(n)+a_ox(n)$ ....(3.23)

The circuit realization for the section is shown in fig. 3.5

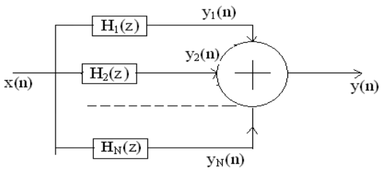


Fig 3.4 Parallel Realization

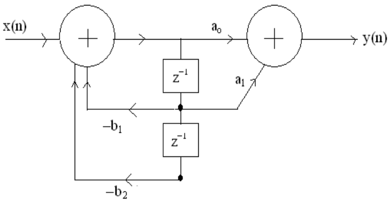


Fig 3.5 one section of Parallel realization

3.5 Normal form

In partial fraction form transfer function H(z) can be expressed as

$$\text{H(z)} \ d_o + \sum_{i=1}^{N} \frac{\alpha_i}{(z - \lambda i)}$$

........(3.24)

Where do, α_i are constant while λ_i may be complex as

$\lambda_i = \sigma + j\omega$ or in conjugate form

Consider two conjugate terms in summation with complex conjugate poles as

$$\left(\frac{\alpha_1}{z - \lambda_1} + \frac{\alpha_2}{z - \lambda_1^*} \right)$$

Suppose response to transfer function $\frac{\alpha_1}{z - \lambda_1}$ is $x_1 + jx_2$

while for $\frac{\alpha_2}{z - \lambda_1^*}$ is $x_1 - jx_2$ then for input 'x' the two terms together gives

$$\frac{\alpha_1 z^{-1}}{1 - \lambda_1 z^{-1}} = \frac{x_1 + jx_2}{x} \; and \; \frac{\alpha_2 z^{-1}}{1 - \lambda_1^* z^{-1}} = \frac{x_1 - jx_2}{x}$$

Substituting for $\lambda_1 = \sigma - j\omega$ above responses are expressed as

$$\alpha_1 z^{-1} x = \left[1 - z^{-1}(\sigma - j\omega)\right](x_1 + jx_2)$$

or

$$x_1 + jx_2 = \alpha_1 z^{-1} x + z^{-1}(\sigma - j\omega)\,(x_1 + jx_2)$$

after simplification it becomes

$H_1\, z^{-1}x = x_1 + jx_2 - z^{-1}\,(x_1\sigma + \omega x_2) - \quad z^{-1}j(x_2\sigma - \omega x_1)$ ....(3.24)

Similarly for conjugate response

$H_2\, z^{-1}x = x_1 - jx_2 - z^{-1}\,(x_1\sigma + \omega x_2) - z^{-1}j(x_2\sigma - \omega x_1)$ ....(3.25)

Adding them and on simplification

$x_1 = (H_1+H_2)z^{-1}x + z^{-1}\, x_1\sigma - z^{-1}\, \omega x_2$ ....(3.26)

In time domain it becomes

$x_1(n+1) = (H_1+H_2)\, x(n) + \sigma x_1(n) + \omega x_2(n)$ ....(3.27)

Similarly subtracting (3.26) from (3.25) gives

$x_2\,(n+1) = (H_1-H_2)\, x(n) + \sigma x_2(n) - \omega x_1(n)$ ....(3.28)

Suppose response to two sections is Y(z) then

$$\frac{Y(z)}{X(z)} = \frac{\alpha_1}{z - \lambda_1} + \frac{\alpha_2}{z - \lambda_1^*}$$

Or $$Y(z) = \frac{\alpha_1 X(z)}{z - \lambda_1} + \frac{\alpha_2 X(z)}{z - \lambda_1^*}$$

Then $y(n) = H_1\, [x_1(n) + jx_2(n)] + H_2\, [x_1(n)-jx_2(n)]$

$= (H_1+H_2)\, x_1(n) + j(H_1-H_2)\, x_2(n)$

Or, $y(n) = c_1x_1(n) + c_2x_2(n)$ …………(3.29)

Equation (3.27) to (3.29) gives the normal form of digital filter realization.

State and output equation in matrix form becomes.

$$\begin{bmatrix} x_1(n+1) \\ x_2(n+1) \end{bmatrix} = \begin{bmatrix} \sigma & \omega \\ -\omega & \sigma \end{bmatrix} \begin{bmatrix} x_1(n) \\ x_2(n) \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} x(n)$$

.(3.30)

and

$$y(n) = \begin{bmatrix} c_1 & c_2 \end{bmatrix} \begin{bmatrix} x_1(n) \\ x_2(n) \end{bmatrix}$$

………….(3.31)

Normal form realization gives structure of the form shown in fig. 3.6

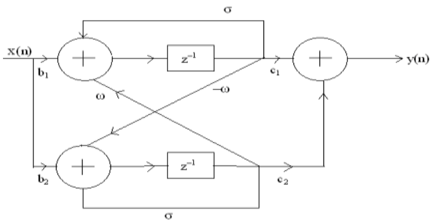


Fig. 3.6 Normal form

With this concept of state space techniques for IIR filter structure realization next chapter is dedicated to IIR filter design.

4 IIR FILTER DESIGN STATE SPACE TECHNIQUES

State space techniques for IIR filter realization was discussed in previous chapter. The same approach can be extended to filter design. Standard method of IIR filter design is lattice filter which has many advantages.

Digital lattice filter plays an important role in finite word length problems. Some of the major obstacles in realizing a digital lattice filter are the efficient use of hardware and an efficient method for directly transforming a Direct II structure into corresponding lattice or ladder structure.

4.1 General solution to problem of realizing lattice structure

Gray and Marked [8] have considered a more general solution to problem of realizing lattice structure. It is canonic both in multiplies and delays. Lattice structure for all pole system Transfer function

$$H(z) = \frac{1}{1+\sum_{K=1}^{N}\alpha_N(k)z^{-k}}$$ ……..

(4.1)

can be realized in the following steps.

Difference equation for (4.1) becomes

$$y(n) = -\sum_{k=1}^{N}\alpha_N(k)\, y(n-k) + x(n)$$

………(4.2)

This gives

$$x(n) = y(n) + \sum_{k=1}^{N}\alpha_N(k)\, y(n-k)$$

……….(4.3)

For N = 1(single pole)

$x(n) = y(n) + a_1(1)\, y(n-1)$ ………(4.4)

This can be realized in lattice structure with

$x(n) = f_1(n)$ ……….(4.5)

$y(n) = f_o(n)= f_1(n) - k_1\, g_o(n-1)$

$= x(n) - k_1\, y(n-1)$ …………(4.6)

$g_1(n)= k_1f_o(n)+ g_o(n-1)$

$= k_1\, y(n) + y(n-1)$ ………..(4.7)

(4.5) to (4.7) give single stage all pole lattice filters as shown in fig. 4.1

For N = 2

$x(n) = f_2(n)$

$y(n) = x(n) - a_2(1)\, y(n-1) - a_2(2)\, y(n-2)$

so

$f_2(n) = x(n)$ ………..(4.8)

$f_1(n) = f_2(n) - k_2g_1(n-1)$…..(4.9)

$g_2(n) = k_2f_1(n) + g_1(n-1)$ …..(4.10)

$f_o(n) = f_1(n) - k_1g_o(n-1)$ ….(4.11)

$g_1(n) = k_1f_o(n)+g_o(n-1)$ ….(4.12)

$y_o(n)= f_o(n) = g_o(n)$

$= f_1(n) - k_1g_o(n-1)$ ...(4.13)

(4.8) to (4.13) gives two stage lattice structure shown in fig. 4.2

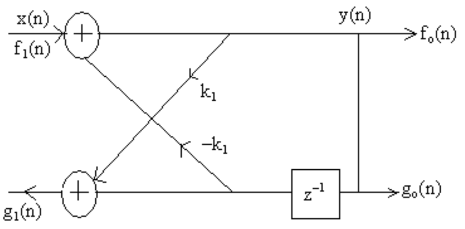


Fig 4.1 Single stage all pole lattice filter

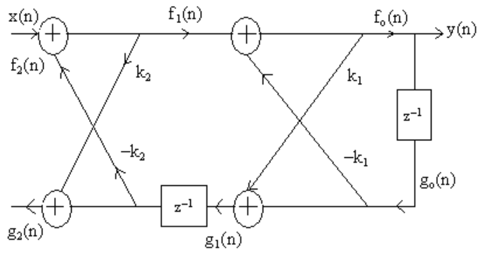


Fig 4.2 Two stage lattice all pole filter

For N – stage IIR lattice structure

$f_N(n) = x(n)$ ....(4.14)

$f_{m-1}(n) = f_m(n) - k_m\ g_{m-1}(n-1)$....(4.15)

$g_m\ (n) = k_m\ f_{m-1}(n) + g_{m-1}(n-1)$...(4.16)

For m = N, N-1, ...,

(4.14 to (4.16) for various m gives state equations as

$x(n) = f_N(n)$ ....(4.17)

$f_m(n) = f_{m-1}(n) + k_m\ g_{m-1}(n\text{-}1)$...(4.18)

$g_m(n) = g_{m-1}(n) + k_m\ f_{m-1}(n)$ ....(4.19)

for m = N, N-1, ..., 1

(4.18) and (4.19) gives state equation in matrix form as

$$\begin{bmatrix} f_1(n) \\ f_2(n) \\ . \\ . \\ . \\ f_{N(n)} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \text{-------} & 0 \\ 0 & 1 & 0\text{-----} & 0 \\ 0 & 0 & 1\text{----} & 0 \\ . & & & \\ . & & & \\ 0 & 0 & 0\text{----} & 1 \end{bmatrix} \begin{bmatrix} f_o(n) \\ f_1(n) \\ \\ . \\ . \\ f_{N-1}(n) \end{bmatrix} + \begin{bmatrix} k_1 & 0 & 0 & \text{----}0 \\ 0 & k_2 & 0 & \text{-----}0 \\ 0 & 0 & k_3 & \text{----}0 \\ . & & & \\ . & & & \\ 0 & 0 & 0 & \text{---}k_N \end{bmatrix} \begin{bmatrix} g_o(n\text{-}1) \\ g_1(n-1) \\ . \\ . \\ . \\ g_{N-1}(n-1) \end{bmatrix}$$

..(4.20)

$$\begin{bmatrix} g_1(n) \\ g_2(n) \\ . \\ . \\ . \\ g_{N(n)} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \text{-------} & 0 \\ 0 & 1 & 0\text{-----} & 0 \\ 0 & 0 & 1\text{----} & 0 \\ . & & & \\ . & & & \\ 0 & 0 & 0\text{----} & 1 \end{bmatrix} \begin{bmatrix} g_o(n) \\ g_1(n) \\ \\ . \\ . \\ g_{N-1}(n) \end{bmatrix} + \begin{bmatrix} k_1 & 0 & 0 & \text{----}0 \\ 0 & k_2 & 0 & \text{-----}0 \\ 0 & 0 & k_3 & \text{----}0 \\ . & & & \\ . & & & \\ 0 & 0 & 0 & \text{---}k_N \end{bmatrix} \begin{bmatrix} f_o(n\text{-}1) \\ f_1(n-1) \\ . \\ . \\ . \\ f_{N-1}(n-1) \end{bmatrix} \quad ..(4.21)$$

output equation is given by

$y(n) = f_o(n) = g_o(n)$ ....(4.22)

On expansion of recursive equation and comparing with the all pass expressions

For N = 2 which gives

$y(n)= z(n) - a_2(1)\ y(n-1) - a_2(2)\ y(n\text{-}2)$ ....(4.23)

And from recursive equations (4.9) to (4.12)

$y(n) = x(n) - k_1\ (1+k_2)\ y(n-1)\ -k_2\ y(n\text{-}2)$.........(4.24)

Comparison of (4.23) and (4.24) gives

$a_2(2) = k_2$, $a_2(1) = k_1\ (1+k_2)$,

$a_2(o) = 0$

4.2 General form of state space matrix equation for IIR filter

IIR filter in discrete form domain can be expressed in general form as

$$y(n) = \sum_{i=0}^{N} a_i x(n-i) - \sum_{i=1}^{N} b_i y(n-i)$$

............(4.25)

When sampling time T is incomplete for each sample instant i.

This filter f_o can be decomposed into pair of cascade filter (fig 4.2) as F_1 & F_2

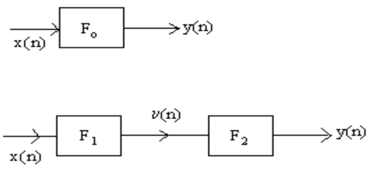


Fig. 4.3

v(n) is response of F_1 filter as operator $\Re_1$ describes. as v(n) = $\Re_1[x(n)]$

$$= \sum_{i=0}^{N} a_i x(n-i) \qquad \text{……..(4.26)}$$

And output y(n) of F_2 is characterized with operator $\Re_2$ as

$$y(n) \quad = \quad \Re_2[\upsilon(n)] = \upsilon(n) - \sum_{i=1}^{N} b_i \; y(n-i)$$

Let us assume x(n) to be applied directly as input to F_2 then it's response $y'(n)$ will be defined as

$$y'(n) = \Re_2[x(n)] = x(n) - \sum_{i=1}^{N} b_i \; y'(n-i)$$

Let new variables *$q_1(n)$, $q_2(n)$……. ,$q_N(n)$ be* defined as

$q_1(n)$ = $y'(n-N)$

= $y'(n-N+1)$

= $y'(n-N)$

$y'(n-N+1)$ = $q_2(n)$

This will give

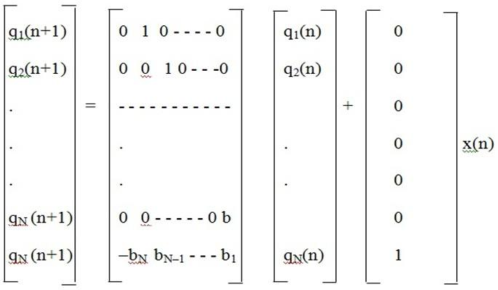
$q_2(n+1)=q_3(n)$

$q_{N-1}(n+1) = q_N(n)$

and

$q_N(n+1)$= $y'(n)$ = x(n)–
$b_1 \, y'(n-1) - b_2 y'(n-2) ……… b_N y'(n-N)$

These equation give state space representation of general IIR filter as …(4.27)



$$\begin{bmatrix} q_1(n+1) \\ q_2(n+1) \\ . \\ . \\ . \\ q_N(n+1) \\ q_N(n+1) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \text{- - - - -} 0 \\ 0 & 0 & 1 \; 0 \text{- - -} 0 \\ \text{- - - - - - - - - - - -} \\ . \\ . \\ 0 & 0 \text{- - - - - -} 0 & b \\ -b_N & b_{N-1} \text{- - -} & b_1 \end{bmatrix} \begin{bmatrix} q_1(n) \\ q_2(n) \\ \\ . \\ . \\ \\ q_N(n) \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} x(n)$$

Output equation can be expressed as

$$y(n) = \Re_2[v(n)] = \Re_2\left[\sum_{i=0}^{N} a_i x(n-i)\right]$$

$$= \sum_{i=0}^{N} a_i \Re_2[x(n-i)] = \sum_{i=0}^{N} a_i y'(n-i)$$

or $\quad y(n) = \sum_{i=0}^{N} a_i y'(n-i) \qquad$ …..(4.28)

(3.29) to (3.31) output equation (4.28) can be re-written as

$Y(n) = a_o x(n) + c_1 \, q_1(n)$ + ……… + $c_{N.} \, q_N(n)$

Where, $\quad c_1 = a_N - a_o b_N$

$c_2 = a_{N-1} - a_o \, b_{N-1}$

$c_N = a_1 - a_o b_1$

Output equation in matrix form can be expressed, as

$$y(n) = [C_1 \; C_2 …….C_N] \begin{bmatrix} q_1(n) \\ q_2(n) \\ . \\ . \\ . \\ q_N(n) \end{bmatrix} + a_o \, x(n)$$

……(4.29)

Thus for the Nth order IIR filter State space representation can be written as

q (n–1) = A q_o(n) + B x(n)

y(n) = C q(n) + D x(n)
……….(4.30)

Where A, B, C and D are matrices.

And the N auxiliary variables q_1(n), q_2(n), ……. Q_N(n) are called state variables. This state space concept will be described for Low pass and High pass IIR filter in next section

4.3 Low Pass Filter (LPF)

Consider second order Butterworth LPF whose transfer function is given by

$$H(s) = \frac{1}{s^2 + \sqrt{2}s + 1} \qquad \text{……..(4.31)}$$

Bilinear transformation

$$S \rightarrow \frac{2}{T}\frac{(1-z^{-1})}{(1+z^{-1})}$$

It gives,

$$H(z) = \frac{1}{\frac{4}{T^2}\left(\frac{1-z^{-1}}{1+z^{-1}}\right)^2 + \sqrt{2}\left(\frac{1-z^{-1}}{1+z^{-1}}\right) + 1}$$

$$= \frac{(1+z^{-1})^2}{\frac{4}{T^2}(1-z^{-1})^2 + \sqrt{2}(1-z^{-2}) + (1+z^{-1})}$$

$$= \frac{1+2z^{-1}+z^{-2}}{\left(\frac{4}{T^2}+\sqrt{2}+1\right) + \left(\frac{8}{T^2}+1\right)z^{-1} + \left(\frac{4}{T^2}-\sqrt{2}\right)z^{-2}}$$

For simplicity let sampling time be normalized to unity then discrete transfer function becomes

$$H(z) = \frac{1+2z^{-1}+z^{-2}}{6.414-7z^{-1}+2.586\,z^{-2}}$$

Or

$$\frac{Y(z)}{X(z)} = H(z) = \frac{0.156+0.31z^{-1}+0.156z^{-2}}{1-1.09z^{-1}+0.4z^{-2}} \quad \text{…….(4.32)}$$

This gives difference equation of output as

$$y(n) = 0.156\,x(n)+0.31\,x(n-1)+0.156\,x(n-2)$$

$$+1.09\quad y(n-1)-0.4z^{-2} \quad \text{…….(4.33)}$$

Comparing it with (4.25) coefficient become as

$a_o = 0.156$, $a_1 = 0.31$, $a_2 = 0.156$

$b_1 = 1.09$, $b_2 = -0.4$

then state equation become

$$\begin{bmatrix} q_1(n+1) \\ q_2(n+1) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ b_2 & b_1 \end{bmatrix} \begin{bmatrix} q_1(n) \\ q_2(n) \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} x(n) \quad \text{…….(4.34)}$$

Output equation become

$$y(n) = a_o x(n)+c_1 q_1(n)+c_2 q_2(n) \quad \text{….(4.35)}$$

where

$c_1 = a_2 - a_o b_2 = 0.156 - 0.156\,(-0.4) = 0.218$

$c_2 = a_1 - a_o b_1 = 0.31 - 0.156 \times 1.09 = 0.14$

Solution of (4.34) and (4.35) together gives solution in time domain for state variable $q_1(n)$ and $q_2(n)$ and output $y(n)$ for input $x(n)$

4.4 High Pass Filter (HPF)

In s – domain HPF is derived from LPF transfer function by replacing s by $\frac{1}{s}$. Then second order Butterworth HPF transfer function is obtained from (4.3) as

$$H(s) = \frac{1}{\left(\frac{1}{s}\right)^2 + \frac{\sqrt{2}}{s} + 1} = \frac{s^2}{s^2+\sqrt{2}s+1}$$

Using bilinear transformation

$$s \rightarrow \frac{2(1-z^{-1})}{(1+z^{-1})}$$

Assuming T to be normalized to unity we get discrete transfer function as

$$H(z) = \frac{4\left(\frac{1-z^{-1}}{1+z^{-1}}\right)^2}{4\left(\frac{1-z^{-1}}{1+z^{-1}}\right)^2 + \sqrt{2}\times 2\left(\frac{1-z^{-1}}{1+z^{-1}}\right) + 1}$$

$$= \frac{4-8z^{-1}+4z^{-2}}{(5+2\sqrt{2})-6z^{-1}+(5-2\sqrt{2})z^{-2}}$$

$$= \frac{4-8z^{-1}+4z^{-2}}{7.828-6z^{-1}+2.2z^{-2}}$$

$$\frac{Y(z)}{X(z)} = H(z) = \frac{0.51-1.02z^{-1}+0.51z^{-2}}{1-0.766z^{-1}+0.28z^{-2}} \quad \text{….(4.36)}$$

This gives coefficient a_o, a_1 a_2 etc as

$a_o = 0.51$, $a_1 = -1.02$, $a_2 = 0.15$

$b_1 = -0.766$, $b_2 = 0.28$

This gives state equation as

$$\begin{bmatrix} q_1(n+1) \\ q_2(n+1) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ b_2 & b_1 \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} x(n) \quad \text{……(4.37)}$$

and output equation as

$$y(n)= a_o x(n)+c_1 q_1(n)+c_2 q_2(n) \quad \text{…(4.38)}$$

where,

$c_1 = a_2 - a_o b_2 = 0.15 - 0.51 \times 0.28$

$= 0.007$

$c_2 = a_1 - a_o b_1 = -1.02 - 0.51 \times (-0.766) = 0.63$

4.5 Band Pass filter (BPF)

Second order Butterworth LPF in s–domain gives equivalent BPF by Quadratic transformation

$$s \to \frac{s^2 + \omega_l \omega_h}{s(\omega_h - \omega_l)} \qquad \text{........(4.39)}$$

Whose ω_n and ω_l are higher and lower cut off frequencies; this gives transfer function for BPF is s–domain as

$$H(s) = \frac{1}{s^2 + \sqrt{2}s + 1}$$

$$s \to \frac{s^2 + \omega_l \omega_h}{s(\omega_h - \omega_l)} \qquad \text{.......(4.40)}$$

$$= \frac{s^2(\omega_h - \omega_l)^2}{s^4 + \sqrt{2}(\omega_h + \omega_l)s^3 + (\omega_l^2 + \omega_h^2)s^2 + \sqrt{2}\omega_l\omega_h(\omega_h - \omega_l)s + \omega_l^2\omega_h^2}$$

For normalized centre frequency $\omega_o = 1$ linear and high frequency is assumed as $\omega_l = 0.4$ and $\omega_h = 1.1$ then BPF in (4.40) becomes

$$H(s) = \frac{0.04\, s^2}{s^4 + \sqrt{2} \times 2s^3 + (0.81 + 1.21)s^2 + \sqrt{2} \times 0.9 \times 1.1 \times 0.2s + 0.81 \times 1.21}$$

$$= \frac{0.04\, s^2}{s^4 + 2.83s^3 + 2.02s^2 + 0.28s + 0.98} \qquad \text{....(4.41)}$$

when converted to discrete with bilinear transformation

$$s \to \frac{2(1 - z^{-1})}{(1 + z^{-1})} \qquad \text{....(4.42)}$$

becomes

$$H(z) = \frac{0.4\left(\frac{1 - z^{-1}}{1 + z^{-1}}\right)^2}{16\left(\frac{1 - z^{-1}}{1 + z^{-1}}\right)^4 + 2.83 \times 8\left(\frac{1 - z^{-1}}{1 + z^{-1}}\right)^3 + 2.02 \times 4\left(\frac{1 - z^{-1}}{1 + z^{-1}}\right)^2 + 0.28 \times 2\left(\frac{1 - z^{-1}}{1 + z^{-1}}\right) + 0.98}$$

on simplification it becomes

$$\frac{Y(z)}{X(z)} = H(z) = \frac{0.033 - 0.0033z^{-2} + 0.0033z^{-4}}{1 - 2.16z^{-1} + 1.776z^{-2} - 0.33z^{-3} + 0.0385z^{-4}}$$

...(4.43)

This gives coefficient as

$a_o = 0.0033$, $a_1 = 0$, $a_2 = -0.0033$, $a_3 = 0$, $a_4 = 0.0033$

$b_1 = -2.16$, $b_2 = 1.776$, $b_3 = -0.33$, $b_4 = 0.0385$

This gives state equation as

$$\begin{bmatrix} q_1(n+1) \\ q_2(n+1) \\ q_3(n+1) \\ q_4(n+1) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -b_4 & -b_3 & -b_2 & -b_1 \end{bmatrix} \begin{bmatrix} q_1(n) \\ q_2(n) \\ q_3(n) \\ q_4(n) \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \text{x(n)}$$

......(4.44)

and output equation becomes

$y(n) = a_o x(n) + c_1 q_1(n) + c_2 q_2(n) + c_3 q_3(n) + c_4 q_4(n)$......(4.45)

Where *$c_1 = a_4 - a_o b_4 = 0.00317$*

$c_2 = a_3 - a_o b_3 = 0.001$

$c_3 = a_2 - a_o b_2 = 0.009$

$c_4 = a_1 - a_o b_1 = 0.007$

Then (4.45) becomes

$y(n) = 0.0033\, x(n) + 0.00317\, q_1(n) + 0.0011\, q_2(n) + 0.009\, q_3(n) + 0.007 q_4(n)$

$= 10^{-3}\,[3.33\, \text{x(n)} + 3.17 q_1(n) + 1.1 q_2(n) + 9 \times q_3(n) + 7 q_4(n)]$ ......(4.46)

On the same procedure bond stop filter can be designed in terms of state equations.

State equation approach of filter design gives real time solution for the filter. This method requires solution of state matrix equation whose order increases with the order of filter.

In matrix equation solution coefficient may have wide spread values then quantization is required beyond a finite decimal places. It's may affect filter response to some extent.

5 STATE SPACE METHOD FOR ANALYSIS OF ROUND OFF ERROR & SCALING

Digital filters designed in the frequency domain from well established techniques for analog filter. This designed filter when implemented on a general purpose computer, the word length is generally fixed. One is interested in estimation of accuracy of filter operation with this word length. This estimation of optimum word length helps in hardware implementation.

Specific sources of quantization error in the implementation and operation of digital filter are

a. **The filter coefficient:** These real number must be
quantized to some finite number of binary bits.

b. **The input samples:** these real or complex numbers must be quantized to a finite number of binary digits before being introduced into filter.

c. **The results of the multiplication of input data by coefficient within the filter:** these must be truncated or rounded off to a specific number of bits.

Input quantization error is integrant in any system where A/D conversion takes place. The error due to quantization of the filter coefficient is deterministic and can be analyzed on the basis set up by designer. The third (3) source of error is called round off noise and complex in nature.

5.1 Round off noise in multiplication of input data with coefficient

This error is dependent on filter architecture: This affects word length requirements for the coefficient quantization and operating characteristic such as the round off noise. This requires selection of optimum filter architecture for achieving minimum word length requirement. This should posses low coefficient sensitivity and a high immunity to round-off error at the output.

If signal through the filter architecture is much larger than Q, the quantization step size (reference level for quantization of sampled signal). Few assumptions can be made such as

1. $E[e_i(n)\, e_j(n+k)] = \dfrac{Q^2}{12},\ for\ i = j$

This shown auto correlation to be constant and gives noise characteristic as,

2. $E[e_i(n)\, e_j(n+k)] = 0\ \ for\ i \neq j$

This gives zero cross correlation or uncorrelated noise source.

3. $E[e_i(n)\ x(n+k)] = 0$

This gives uncorrelation between noise and input samples.

Digital filter model with noise is shown in Fig (5.1)

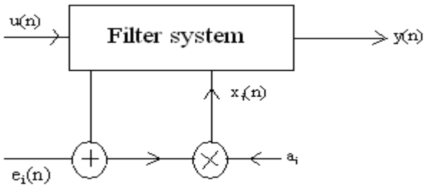


Fig. 5.1 Round-off Error source

Round off noise at filter output due to noise input e$_i$(n) is given by convolution

$$e_x(n) = \sum_{k=0}^{n} h_i(k) e_i(n-k) \qquad ……..(5.2)$$

Where h$_i$(k) is the impulse response of single input/single output time shift invariant filter.

Round off error variance due to e$_i$(n) is given by

$$\sigma_x^2(n) = E[e_{x_i}(n-m)\ e_{x_i}(n-k)]$$

$$= \sum_{m=0}^{n}\sum_{k=0}^{n} h_i(m)\ h_i(k) E\ [e_i(n-m)\ e_i(n-k)]$$

$$= \sum_{m=0}^{n}\sum_{k=0}^{n} h_i(m)\ h_i(k)\ \sigma_{e_i}^2\ \delta_k(k-m)$$

This follows from property (3) of uncorrelation property of $\delta_k(k-m)$ gives

$$\sigma_x^2(n) = \sigma_e^2 \sum_{m=0}^{n} h_i(m) \qquad …….(5.3)$$

When σ_e^2 is the error variance of round-off noise from error source, Round off noise for various noise inputs can be calculated using state variable techniques.

5.2 Scaling

If amplitude of internal signal in a fixed-point digital filter excels dynamic range overflow occurs. On the other hand if amplitude is kept at much lower level it will give poor S/N ratio. Then for optimum filter operations scaling of amplitudes are required. Suppose input signal is bounded by M in l_2 space then

$$||2(n)|| \leq \text{M} \qquad …..(5.4)$$

For filter transfer function F(z) roundness of output v(z) to M in time domain can be obtained by a scaling constant λ obtained from filter relationship.

$v(z) = \lambda F(z)\, X(z)$ …….(5.5)

Where v(z) is output of system with transfer function F(z) and input X(z)

Using schwartz inequality

$|v(n)| \leq \|X\|_2\, \|\lambda F\|_2$ …….(5.6)

Where lower suffix 2 denotes l_2 space, and

$|x(n)| \leq \|X\|_2 \leq M$

Then $|v(n) \leq M\, \|\lambda F\|_2$.(5.7)

For $|v(n)| \leq M$, condition gives limitation on λ as

$\|\lambda F\|_2 \leq 1$

or $\lambda \leq \dfrac{1}{\| F \|_2}$ …….(5.8)

5.3 State space method for round off error

In round off error and scaling analysis of filter two separate analysis are required to be done : Analysis of round off error requires computation of a series of impulse response while the determination of dynamic range. Constraints are done in l_2 space.

State variable technique determines both in algebraic form. Consider a linear time shift invariant digital filter characterized in algebraic form as

$\underline{X}(n+1) = A\underline{X}(n) + bu(n)$ …….(5.9)

And output equation as

$y(n) = c\underline{X}(n) + d\,u(n)$ …….(5.10)

Whose $\underline{X}(n)$ is n-dimensional vector, A is state matrix, u(n) is scalar input, b, c and d are real constant metrics. For n-dimensional state equation impulse response is also n-dimensional as

$h(k) = \{d(0)\ for\ k = 0; CA^k b\ for\ k \geq 1\}$ ……..(5.11)

and state variable $\underline{X}(n)$ is given by

$$\underline{X}(n) = \sum_{k=0}^{n} A^{k-1} b\, u(n-k) \text{ ……..(5.12)}$$

Where i'th row of $(A^{k-1}\, b)$ will be denoted as $f_i(n)$. For unit impulse input state variable is bounded in l_2 space as

$\|X_i\|_2 = \|F_i\|_2$ ……..(5.13)

A scale factor $\delta \geq 1$ determines probability of overflows for state variable as

$$\| X_i \|_2^2 \leq \delta^2 \| f_i \|_2^2 = \delta^2 \left(\sum_{k=1}^{\infty} f_i^2(k) \right) = [Q(2^{n_i} - 1)]^2,$$

for all i ……..(5.14)

Where Q is the quantization reference size and n_i is the word length of i'th state variable.

5.4 Analysis of round off error

Assume errors are uncorrelated, uniformly distributed (while noise) random process out [–Q/2 , Q/2] such that each error source contributes an error variance of $\dfrac{Q^2}{12}$ for each multiplication.

Suppose there are m_i non-integer multiplication in computation of $x_i(n)$ then m_i round off error sources corrupt $x_i(n)$. With proper scaling the sources error saves are not allowed to overflow.

Impulse response in (5.11), can then be analyzed in terms of a parameter g_i, the i'th component of row vector CA^{k-1}. The noise appearing at the output due to $x_i(n)$ is given by

$$\sigma_i^2 = \frac{m_i Q^2 \| g_i \|^2}{12} \text{ ……(5.15)}$$

Total output round off noise is then simply the sum of all the individual error statistic r

$$\sigma_x^2 = \sum_{i=1}^{n} m_i \frac{\| g_i \|^2\, Q^2}{12} \text{ ……(5.16)}$$

5.5 Mullis and Roberts [9] method of computing σ_x^2

Two new matrices K and W are defined as

$$K = AKA^T + bb^T = \sum_{k=0}^{\infty} (A^k b)(A^k b)^T$$

…..(5.17)

$$W = \overset{T}{A}WA + \overset{T}{C}C = \sum_{k=0}^{\infty}(CA^{K})^{T}(CA^{K}) \quad \text{.....(5.18)}$$

These matrices are positive definite for stable filters with no pole-zero cancellation K_{ij} is the inner product of f_i and f_j while W_{ij} is the inner product of g_i and g_j then scaling constraint in (5.14) can be interpreted in terms of diagonal elements of k as

$$\delta^2 k_{ii} = \left[Q(2^{ni} - 1)\right]^2; \quad i = 1, 2, \text{........} n \quad \text{........(5.19)}$$

While for Output errors variance in (5.16) becomes ($\delta = 1$)

$$\sigma_x^2 = \sum \frac{\sum_{i=1}^{n} m_i Q^2}{12} W_{ii} \quad \text{......(5.20)}$$

In general a non-singular matrix T can be obtained to transform A into diagonal matrix by replacing state variable X by new state variable X defined as

$$\underline{X'} = T^{-1}\,\underline{X} \quad \text{......(5.21)}$$

Then metrics (A, b,c) are transformed to (A′, b′, c′) and element.

$$X_1' = \frac{x_1}{T_{11}} \text{ and } (k', w') = (\overset{-1}{T}\, k\, \overset{T_r}{T}, \overset{-1}{T}\, WT) \quad \text{.......(5.22)}$$

Where T_r denotes transpose & T_{11} is 1st diagonal element of matrix T.so that,

$$k'_{11} = \frac{K_{11}}{T_{11}^2} \text{ and } W'_{11} = \frac{W_{11}}{T_{11}^2} \quad \text{.......(5.23)}$$

If word length is assumed uniform as

$N_i = m$ for all i

Then output error variance becomes

$$\sigma_x^2 = \frac{(n+1)n}{3}\left(\frac{\delta}{2^m}\right)^2 \frac{1}{n}\left[\sum_{i=1}^{n} k_{ii} w_{ii}\right]$$

k′ and w′ are explicit functions of T. It is then evident that scaling and round off errors are architecture dependent.

Thus state variable approach combines problem of round off error and dynamic range into single set of state variable. They are computed separately from state equation solution.

6 COMMENTS & CONCLUSION

Various structure of digital filter like direct structure were presented, in this work, cascade structure was most simple form for hardware realization point of view.

In article 3 state space representation of cascade structure in simple matrix form was observed to be most suitable for computational analysis. Each cascade section will have second order state matrix. Normal form of realization discussed here had very complex structure although they had simple form of matrix representation.

Simplicity of state space representation by cascade structure was demonstrated in article 4 where general form of higher order digital filter was represented in state space work various simple state space elements which all the state having “0” or “1” element except the last one with denominator coefficient b_1, b_2, ...b_N of general transfer function, biqued section of each LPF, HPX etc, are related in the thesis by the method of bilinear transformation arranging them be fined that state variable form requires large calculation but the final form is very simple.

In round off and scaling problem discussed in article 5 it was observed that both can be simultaneous analyzed in state space representation, closed form analysis best and some assumption like white noise character and uncorrelated noise simplify the problem.

FUTURE SCOPE OF RESEARCH WORK

State space analysis of round off and dynamic scaling lumped together opens a new field of state space statistical analysis for digital filter. This field of research can provide further scope of research work.

ACKNOWLEDGEMENT

The research paper is a part of MSc.(Engg.) research work submitted for degree at B R A Bihar University, and Its application has been done under the guidance of Dr. U.C. Verma of M.I.T., Muzaffarpur.

Dr. B. M. Mishra, Director Women’s Institute of Technology, Darbhanga and Dr. S.K.Jha, OSD, Darbhanga College of Engineering, Darbhanga guided me and made simple the lab-work in the premises so they are thankful for the work.

The research paper for degree M.Sc.(Electrical Engg.),is dedicated to my mother Smt. Manju Srivastava for continuous moral support.

REFERENCES

[1] Antonion, A, Digital filter analysis and Design, TMH Edition

[2] Babu, Ramesh P, Digital signal Processing, Sci-Tech Publication, Chennai.

[3] Oppenheim and Schafer, Digital signal Processing, PHI, Delhi.

[4] Jackson, L-B, “On the Introduction of Round off Noise and Dynamic range in Digital Filter” , Bell Syst. Tech. J, Vol 49, pp 159-184, February 1970

[5] Avenhauss, A, “ on the design of Digital filter with coefficients of lumped word Length” , IEEE Trans. , Audio Electro, Vol AU-20, pp 206-212, August 1972.

[6] Taylor, F.J & Marshall, J.W., “Computer Aided Design and analysis of standard IIR Structures” , Part 1 , IEEE, Circuit & System Magazine, Vol. 3, No.4, pp 2-6, July 1981.

[7] Taylor, F.J & Marsall, J.W., “Computer Aided Design and analysis of Standard IIR Structures”, Part 2, IEEE, Circuit & System Magazine, Vol. 4, No. 1, pp 5-10, March 1982.

[8] A.H.Gray and J.D.Markd, “ Digital Lattice and Ladder Filter Synthesis” IEEE Trans, Audio Electroacoust, AU-21, pp 491-500, December 1973.

[9] Mullis C.T. and Roberts R.A., “Synthesis of Minimum Round-off Noise Fixed point Digital filter”, IEEE Trans., Circuit system, Vol. (AS-23), pp 551-562,September 1976.