Sep 26, 2006

Scaling in Fixed Point Implementations of IIR System

* Scale input so that no overflows in intermediate node.

* $W_k(n)$ = value of the kth node variable

* $h_k(n)$ = impulse response from input $x(n)$ to the node variable $W_k(n)$

$$|w_k(n)| = \left| \sum_{m=-\infty}^{+\infty} x(n-m) h_k(m) \right|$$

$$\leq X_{max} \sum_{m=-\infty}^{+\infty} |h_k(m)|$$

To ensure $|W_k(n)| < 1$, sufficient condition is

$$X_{max} \sum_{m=-\infty}^{+\infty} |h_k(m)| < 1 \implies$$

$$X_{max} < \frac{1}{\sum_{m=-\infty}^{+\infty} |h_k(m)|} \qquad \forall k$$

If scale input by s,

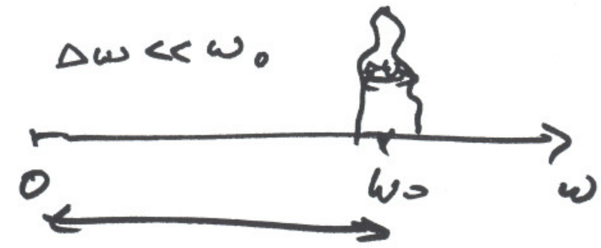
$$\boxed{s X_{max} < \frac{1}{\sum_{m=-\infty}^{+\infty} |h_k(m)|}}$$

Conservative
pessimistic
loose upper bound.

<u>Approach</u>: Assume ~~is~~ signal is narrowband.

$x(n) = X_{max} \cos \omega_0 n$

$\Delta\omega \ll \omega_0$



$$w_k(n) = |H_k(\omega_0)| X_{max} \cos\left[\omega_0 n + \measuredangle H_k(\omega_0)\right]$$

avoid overflow for <u>all</u> sinusoidal signals.

$$\max_{k,\,|\omega|<\pi} X_{max} |H_k(\omega)| < 1$$

scaling factor s:

$$\boxed{s\, X_{max} < \frac{1}{\max\limits_{k,\,|\omega|<\pi} |H_k(\omega)|}}$$

<u>Approach 3</u> : scale input so that total energy @ each node is smaller than total energy of the input signal.

$$\sum_{n=-\infty}^{+\infty} |W_k(n)|^2 < \sum_{n=-\infty}^{+\infty} |x(n)|^2$$

$$\sum_{n=-\infty}^{+\infty} |W_k(n)|^2 = \frac{1}{2\pi} \int_{-\pi}^{+\pi} |H_k(\omega)\, X(\omega)|^2 \, d\omega$$

$$\leq \frac{1}{(2\pi)^2} \int_{-\pi}^{+\pi} |X(\omega)|^2 d\omega \int_{-\pi}^{+\pi} |H_k(\omega)|^2 d\omega$$

$$\leq \sum_{n=-\infty}^{+\infty} |x(n)|^2 \; \frac{1}{2\pi} \int_{-\pi}^{+\pi} |H_k(\omega)|^2 d\omega$$

schwartz

parseval

To ensure $\sum |w_k(n)|^2 \leq \sum |x(n)|^2$

Scale $x(n)$ by scale factor s s.t.

$$s^2 \leq \frac{1}{\frac{1}{2\pi}\int_{-\pi}^{+\pi} |H_k(\omega)|^2 d\omega} = \frac{1}{\sum_{n=-\infty}^{+\infty} |h_k(n)|^2}$$

Can be shown that

$$\left\{ \sum_{n=-\infty}^{+\infty} |h_k(n)|^2 \right\}^{1/2} \leq \max_{k,\omega} |H_k(\omega)| \leq \sum_n |h_k(n)|$$

least conservative

most conservative i.e. most pessimistic

$$x(n) \rightarrow \boxed{\text{flows}} \rightarrow y(n)$$

$$x(n) \rightarrow \boxed{s} \rightarrow \boxed{\text{flow}} \rightarrow \quad s\, y(n)$$

Signal to noise ratio at the output =

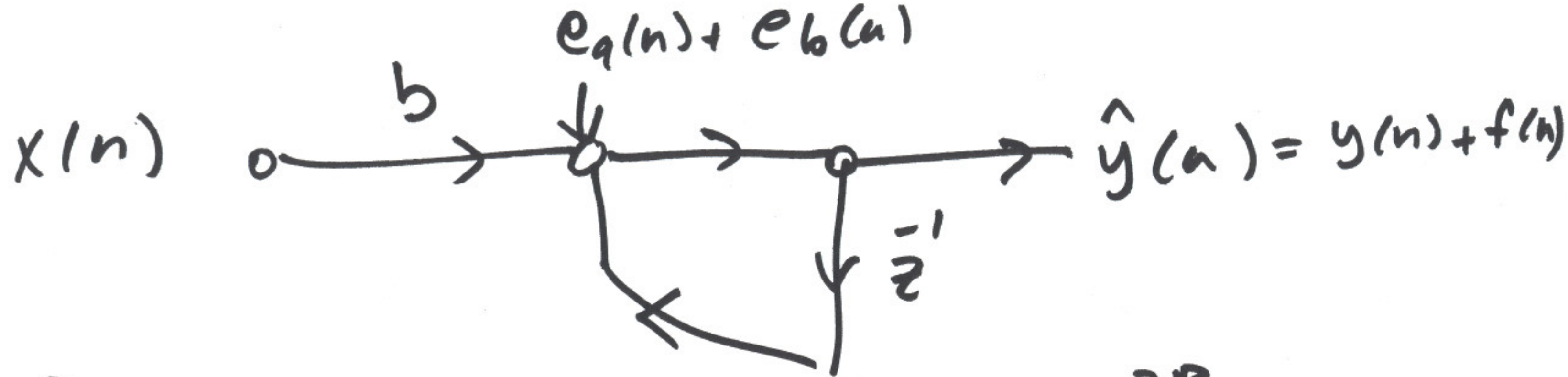
$$\frac{\mathrm{var}[s\,y(n)]}{\mathrm{Var}[\text{noise}]} = \frac{s^2\,\mathrm{Var}(y(n))}{\mathrm{Var}[\text{noise}]}$$

coeff. quant. noise

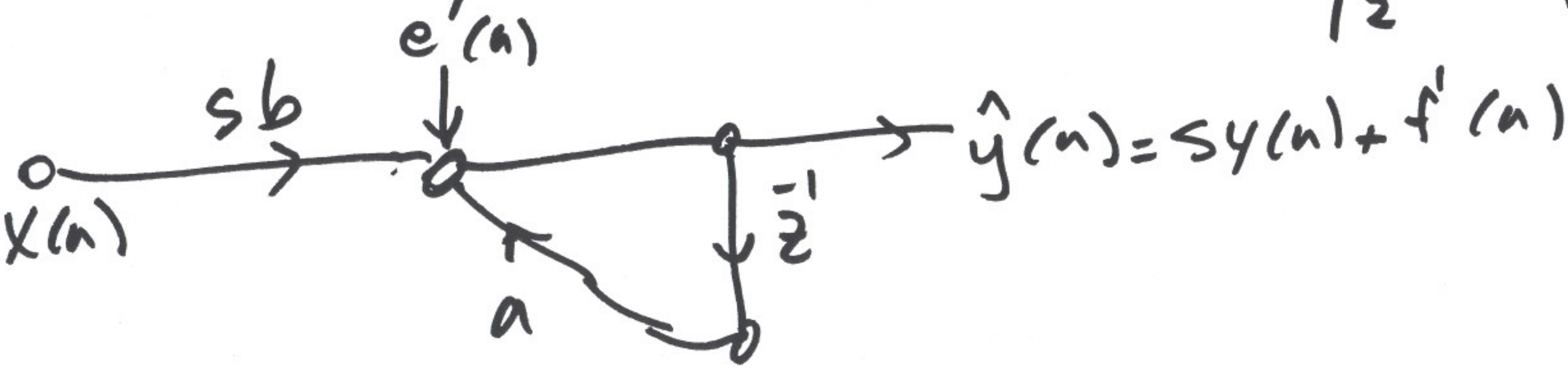
$$s \leq 1$$

$$\Longrightarrow \text{SNR} \downarrow$$

Ex $H(z) = \dfrac{b}{1 - a z^{-1}}$ $\quad |a| < 1$



$$\sigma_f^2 = \text{noise variance at output} = 2 \, \frac{2^{-2B}}{12} \left(\frac{1}{1 - |a|^2} \right)$$



Input : white noise amplitudes uniformly dist. between -1 to $+1$.

Signal variance $= \frac{1}{3}$

choose scale factor $s = \dfrac{1}{\sum_{n=0}^{\infty} |b||a^n|} = \dfrac{1-|a|}{|b|}$

Variance of output $= s^2 \sigma_y^2$

$\sigma_{y'}^2$

$$= s^2 \left(\frac{1}{3}\right) \frac{b^2}{1-a^2} = \text{variance of signal at output}$$

SNR at output $= \dfrac{s^2 \left(\frac{1}{3}\right) \frac{b^2}{1-a^2}}{2\, \frac{2^{-2B}}{12} \left(\frac{1}{1-|a|^2} \right)}$

$$S = \frac{1-|a|}{|b|} \Rightarrow S^2 = \frac{(1-|a|)^2}{|b|^2}$$

$$\text{SNR at output} = \frac{\frac{(1-|a|)^2}{|b|^2} \; \frac{1}{3} \; \cancel{\frac{b^2}{1-a^2}}}{2\,\frac{2^{-2B}}{12} \; \cancel{\frac{1}{1-|a|^2}}}$$

$$\boxed{\text{SNR output} = \frac{2}{2^{-2B}}\left(1-|a|\right)^2 \qquad \text{As } a \to 1 \quad \text{SNR} \downarrow}$$