

FILTRI DIGITALI FIR E IIR

I filtri digitali che studiamo in questo capitolo sono particolari sistemi LTI causali a tempo discreto. Essi possono essere implementati e simulati su macchine digitali (microprocessori o processori specializzati come i DSP); per molti anni addirittura essi sono risultati la più comune applicazione dei DSP.

Il vantaggio sui filtri analogici è duplice:

- essi possono essere riprogrammati via software sullo stesso hardware;
- è possibile modificare in tempo reale i coefficienti dei filtri, ottenendo in tal modo filtri “adattativi”.

I principali tipi di filtri digitali sono i filtri *FIR* (*Finite Impulse Response*) e *IIR* (*Infinite Impulse Response*). Tecnicamente, un filtro FIR è un sistema LTI causale con risposta finita all’impulso; la funzione di trasferimento di un filtro FIR risulta essere un polinomio in z^{-1} . Con filtri IIR invece, intendiamo quella sottoclasse dei sistemi LTI causali con risposta anche infinita all’impulso, dotati di una funzione di trasferimento razionale in z^{-1} .

Nelle applicazioni pratiche, il filtro FIR è da preferire quello IIR dove vi sia richiesta di fase lineare, altrimenti, se la distorsione in fase è tollerabile, si preferisce l’uso di filtri IIR perché comportano un minor numero di parametri, meno memoria e minor complessità computazionale.

Nel primo paragrafo si introducono le definizioni di filtri FIR e IIR, discutendone vantaggi e svantaggi:

- un filtro FIR è sempre stabile e può avere fase lineare;
- un filtro IIR può essere instabile e non ha in generale fase lineare; essi possono tuttavia avere alcune caratteristiche (banda di transizione, attenuazione) migliori rispetto a quelle dei filtri FIR.

Si discute poi la realizzazione di filtri mediante reti i cui elementi costitutivi sono operatori di somma, moltiplicazione per costante e ritardo temporale. Il progetto di queste reti è agevolato da proprietà composizionali quali:

- la funzione di trasferimento del sistema ottenuto dalla “cascata” di due reti è il prodotto delle funzioni di trasferimento dei sistemi realizzati dalle singole reti;
- la funzione di trasferimento del sistema ottenuto dalla “somma” di due reti è la somma delle funzioni di trasferimento dei sistemi realizzati dalle singole reti.

Nel quarto paragrafo vengono introdotte alcune tecniche di progetto di filtri digitali. Vengono analizzati i principali parametri utilizzati nella progettazione aiutata da strumenti CAD; si mostra poi una semplice tecnica di realizzazione di filtri digitali IIR a partire da quelli analogici e si discute in dettaglio la tecnica di progetto di filtri FIR col metodo delle “finestre”.

Il capitolo si conclude con un cenno ai principali tipi di rumore introdotto dal fatto che segnali e sistemi inerentemente analogici vengono trattati con tecniche digitali. In particolare, viene discusso il rumore prodotto dalla quantizzazione del segnale, quello dovuto alla quantizzazione dei coefficienti e quello causato dai troncamenti nel calcolo digitale.

8.1 Filtri FIR e IIR Causali

La causalità, sinonimo di realizzabilità, ha implicazioni importanti sul comportamento in frequenza del filtro che meritano di essere discusse brevemente.

Il progetto di filtri FIR e IIR causali consiste nel determinarne i coefficienti mediante specifiche formulate normalmente nel dominio delle frequenze, espresse in termini di ampiezza e fase della risposta del filtro desiderata. Questa attitudine non garantisce tuttavia che una determinata risposta in frequenza produca una risposta all’impulso causale. Sebbene non vi sia una soluzione generale al problema, si possono individuare condizioni necessarie o almeno sufficienti ponendo restrizioni sulla risposta in frequenza.

Paley e Wiener (1934) hanno dimostrato che, se una sequenza $h(n)$ è nulla per $n < 0$ e ha energia finita, cioè $\sum_{n=0}^{+\infty} |h(n)|^2 < +\infty$, allora la sua trasformata di Fourier verifica:

$$\int_{-\pi}^{\pi} |\ln |H(e^{i\omega})|| d\omega < +\infty.$$

Viceversa, se $|H(e^{i\omega})|$ è quadrato integrabile, cioè

$$\int_{-\pi}^{\pi} |H(e^{i\omega})|^2 d\omega < +\infty$$

e se la condizione di Paley-Wiener è rispettata, allora è sempre possibile associare al modulo di $H(e^{i\omega})$ una risposta in fase $\Phi(\omega)$ in modo tale che il filtro con risposta in frequenza

$$H(e^{i\omega}) = |H(e^{i\omega})| e^{i\Phi(\omega)}$$

sia causale. Un'importante conseguenza di questo risultato è che un filtro causale ha un risposta in frequenza che non può essere nulla su una banda finita, il che conferma la non causalità del filtro ideale.

In aggiunta alla suddetta condizione, la causalità impone anche una relazione forte tra la parte reale $H_R(e^{i\omega})$ e la parte immaginaria $H_I(e^{i\omega})$ di $H(e^{i\omega})$. La spiegazione deriva dalla decomposizione della sequenza $h(n)$ in due sequenze: una pari $h_p(n)$ e una dispari $h_d(n)$, definite come segue:

$$h_p(n) = \frac{1}{2}[h(n) + h(-n)], \quad h_d(n) = \frac{1}{2}[h(n) - h(-n)].$$

È indubbio che un segnale $h(n)$, nullo per $n < 0$, coincide con la sequenza pari $h_p(n)$ associata o alla sequenza dispari $h_d(n)$ associata più $h(0)$. Ne consegue che le due sequenze, quella pari e quella dispari associate a $h(n)$, non sono tra loro indipendenti.

Poiché si può mostrare che le trasformate di $h_p(n)$ e $h_d(n)$ risultano essere rispettivamente:

$$h_p(n) \xleftrightarrow{\mathcal{F}_d} H_R(e^{i\omega}), \quad h_d(n) \xleftrightarrow{\mathcal{F}_d} H_I(e^{i\omega})$$

si deduce che sussiste una relazione di interdipendenza anche tra le due trasformate $H_R(e^{i\omega})$ e $H_I(e^{i\omega})$.

Riassumendo, la causalità ha importanti implicazioni sul progetto e la realizzazione di filtri digitali:

- la risposta in frequenza $H(e^{i\omega})$ non può essere zero eccetto che in un insieme finito di punti;
- l'ampiezza $|H(e^{i\omega})|$ della risposta in frequenza non può essere costante su una banda finita di frequenze; le bande passante e proibita presentano oscillazioni che non possono essere eliminate a causa del troncamento operato su $h(n)$ (fenomeno di Gibbs);
- la parte reale $H_R(e^{i\omega})$ e la parte immaginaria $H_I(e^{i\omega})$ di $H(e^{i\omega})$ sono tra loro interdipendenti, di conseguenza il modulo $|H(e^{i\omega})|$ e la fase $\Phi(e^{i\omega})$ di $H(e^{i\omega})$ non possono essere scelti arbitrariamente.

8.1.1 Filtri FIR e Filtri a Fase Lineare

Richiamiamo che un sistema LTI causale a tempo discreto è detto filtro FIR se la risposta $h(n)$ all'impulso unitario è finita nel senso che $h(n) = 0$ per $n < 0$ e per $n \geq M$ per un opportuno $M > 0$. Il rapporto ingresso-uscita è allora descritto dalla seguente equazione alle differenze finite:

$$y(n) = \sum_{k=0}^{M-1} h(k)x(n-k). \quad (8.1)$$

Passando alle trasformate z e applicando la proprietà della traslazione temporale, si ottiene:

$$Y(z) = H(z)X(z),$$

dove $H(z) = \sum_{k=0}^{M-1} h(k)z^{-k}$ e $X(z)$, $Y(z)$ sono le trasformate z rispettivamente di $x(n)$ e $y(n)$. Si osservi che $H(z)$ è un polinomio in z^{-1} .

Le caratteristiche più interessanti dei filtri FIR sono le seguenti:

1. Un filtro FIR è sempre causale e stabile; ciò può essere rilevato dal fatto che $H(z)$ è un polinomio in z^{-1} , e quindi ha un solo polo in $z = 0$, di fatto interno al cerchio di raggio 1.
2. Un filtro FIR può avere fase lineare: se la funzione $h(n)$ è simmetrica o antisimmetrica rispetto a $(M-1)/2$ (cioè $h(k) = h(M-1-k)$ oppure $h(k) = -h(M-1-k)$), allora la fase $\angle H(e^{i\omega})$ è lineare.

Infatti, se $h(k) = h(M-1-k)$:

$$\begin{aligned}
 H(e^{i\omega}) &= \sum_{k=0}^{M-1} h(k)e^{-ik\omega} \\
 &= \sum_{k=0}^{M-1} h(k) \frac{e^{-ik\omega} + e^{-i(M-1-k)\omega}}{2} \\
 &= e^{-i\frac{M-1}{2}\omega} \sum_{k=0}^{M-1} h(k) \frac{e^{i(\frac{M-1}{2}-k)\omega} + e^{-i(\frac{M-1}{2}-k)\omega}}{2} \\
 &= e^{-i\frac{M-1}{2}\omega} \sum_{k=0}^{M-1} h(k) \cos\left(\frac{M-1}{2} - k\right) \omega.
 \end{aligned}$$

Poichè $\sum_{k=0}^{M-1} h(k) \cos\left(\frac{M-1}{2} - k\right) \omega$ è un numero reale, la fase risulta lineare:

$$\angle H(e^{i\omega}) = -\frac{M-1}{2}\omega.$$

Analogamente, se $h(k) = -h(M-1-k)$, si ottiene:

$$H(e^{i\omega}) = ie^{-i\frac{M-1}{2}\omega} \sum_{k=0}^{M-1} h(k) \sin\left(\frac{M-1}{2} - k\right) \omega.$$

Osservando che $i = e^{i\pi/2}$, si ottiene:

$$\angle H(e^{i\omega}) = -\frac{M-1}{2}\omega + \frac{\pi}{2}.$$

In Figura 8.1 sono rappresentate le risposte all'impulso di due filtri FIR a fase lineare, uno antisimmetrico, l'altro simmetrico.

8.1.2 Filtri IIR

Consideriamo ora un sistema LTI in cui la relazione ingresso-uscita verifica la seguente equazione alle differenze finite:

$$\sum_{k=0}^{L-1} a_k y(n-k) = \sum_{k=0}^{M-1} b_k y(n-k) \quad (a_{L-1} \neq 0).$$

Osserviamo che se $L = 1$ l'equazione precedente definisce un filtro FIR.

Se invece $L > 1$, l'equazione precedente non è in grado di specificare univocamente il sistema: passando infatti alle trasformate z e applicando la proprietà della traslazione temporale, si ottiene:

$$Y(z) = H(z)X(z),$$

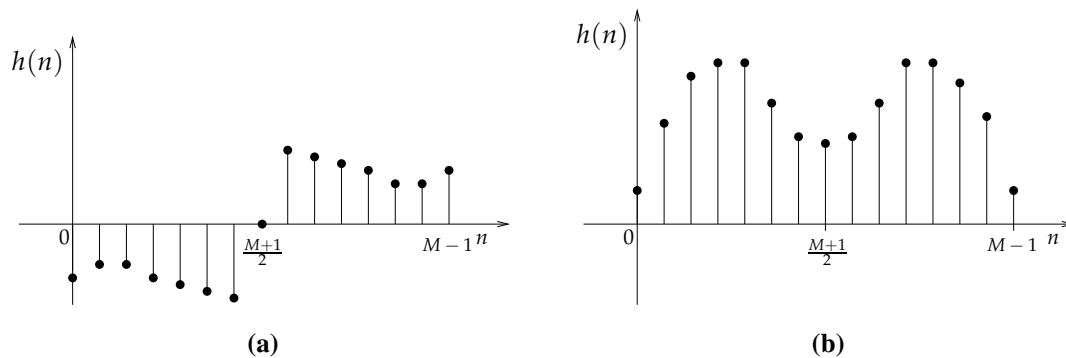


Figura 8.1 (a) Risposta all'impulso di un filtro a fase lineare antisimmetrico. (b) Risposta all'impulso di un filtro a fase lineare simmetrico.

dove $H(z) = \frac{\sum_{k=0}^{M-1} b_k z^{-k}}{\sum_{k=0}^{L-1} a_k z^{-k}}$ e $X(z)$, $Y(z)$ sono le trasformate z rispettivamente di $x(n)$ e $y(n)$. Poiché $H(z)$ è una funzione razionale in z^{-1} dotata di poli distinti da 0, possiamo descrivere più sistemi caratterizzati dalla stessa funzione $H(z)$ ma aventi diverse corone di convergenza. Solo uno di essi tuttavia, e cioè quello contenente ∞ , è causale. Possiamo allora dare la seguente:

Definizione 8.1. Un filtro IIR è un sistema LTI causale tale che la relazione ingresso-uscita verifica l'equazione ricorsiva alle differenze finite:

$$\sum_{k=0}^{L-1} a_k y(n-k) = \sum_{k=0}^{M-1} b_k y(n-k) \quad (a_{L-1} \neq 0, L > 1).$$

Come abbiamo visto sopra, la funzione di trasferimento $H(z)$ di un filtro IIR è $H(z) = \frac{\sum_{k=0}^{M-1} b_k z^{-k}}{\sum_{k=0}^{L-1} a_k z^{-k}}$; la presenza di poli distinti da 0 comporta che, se il filtro è causale, la risposta $h(n)$ all'impulso unitario è nulla per $n < 0$, ma risulta diversa da 0 per infiniti n positivi: la fase di questi filtri non può mai essere lineare.

Esempio 8.1.1.

Il filtro IIR descritto dall'equazione $y(n) + ay(n-1) = x(n)$ ammette la funzione di trasferimento $H(z) = \frac{1}{1+az^{-1}}$ e corona di convergenza $|z| > |a|$, poichè il polo è in $z = -a$. In tale regione vale che $H(z) = \sum_{n=0}^{+\infty} a^n z^{-n}$, quindi la risposta all'impulso unitario del filtro è:

$$h(n) = \begin{cases} 0, & n < 0 \\ a^n, & n \geq 0 \end{cases}.$$

La risposta all'impulso non è finita e il filtro è stabile solo se $|a| < 1$.

L'esempio precedente mostra che, contrariamente a quanto accade per filtri FIR, un filtro IIR può non essere stabile; ulteriormente, mentre esistono filtri FIR con fase lineare, la fase di un filtro IIR non è mai lineare.

Questi svantaggi sono compensati in vari casi dalla maggior semplicità realizzativa dei filtri IIR rispetto ai filtri FIR e dalle migliori caratteristiche di attenuazione a parità di ordine delle equazioni.

8.2 Applicazioni di Filtri FIR e IIR

8.2.1 Zeri di Filtri a Fase Lineare: i Filtri COMB

Un filtro FIR ha funzione di trasferimento del tipo

$$H(z) = \frac{P(z)}{z^M},$$

dove $P(z)$ è un polinomio di grado M a coefficienti reali. A meno di una costante moltiplicativa, il filtro viene allora univocamente determinato dagli zeri di $P(z)$, cioè dalle soluzioni dell'equazione $P(z) = 0$.

Sia z_1 una radice di $P(z)$; essendo i coefficienti di $P(z)$ reali, ricordiamo che o z_1 è reale o z_1 è complesso, ma allora il suo coniugato z_1^* è un'altra radice di $P(z)$.

Il seguente risultato caratterizza i filtri FIR con fase lineare, in termini dei loro zeri:

Fatto 8.1. *Sia dato un filtro FIR con funzione di trasferimento $H(z) = \frac{P(z)}{z^M}$. Allora le due seguenti proposizioni sono equivalenti:*

1. *il filtro ha fase lineare;*
2. *se $re^{i\theta}$ è uno zero di $P(z)$, allora anche $\frac{1}{r}e^{i\theta}$ è uno zero di $P(z)$.*

Ne segue in particolare che filtri i cui zeri sono tutti sulla circonferenza di raggio 1 hanno fase lineare. Un'importante famiglia di filtri che hanno gli zeri sulla circonferenza unitaria è quella dei filtri con funzione di trasferimento del tipo:

$$H(z) = \frac{1}{M}(1 - z^{-M}).$$

In questo caso, gli zeri del filtro sono esattamente le radici dell'unità $z_k = e^{2\pi\frac{k}{M}}$ ($k = 0, \dots, M-1$).

La fase di questi filtri è lineare per ogni M ; il guadagno per $M = 8$, è dato dal grafico in Figura 8.2.

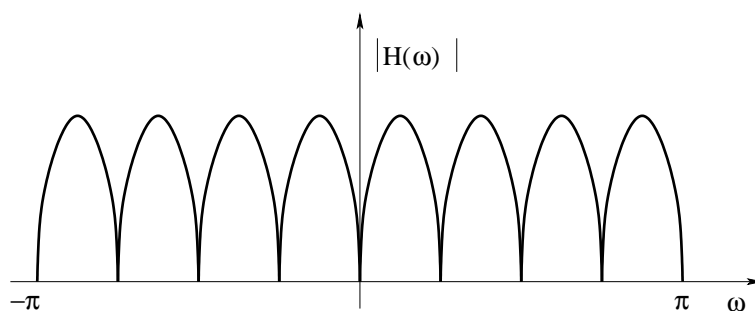


Figura 8.2 Guadagno di un filtro comb di ordine 8.

A causa della forma “a pettine” del grafico precedente, questi filtri sono detti *comb*. Essi sono implementabili in modo veramente efficiente: dette infatti $X(z)$ e $Y(z)$ le trasformate zeta dell'ingresso $x(n)$ e $y(n)$ dell'uscita, vale $Y(z) = \frac{1}{M}(X(z) - z^{-M}X(z))$. Questo implica

$$y(n) = \frac{1}{M}(x(n) - x(n - M)).$$

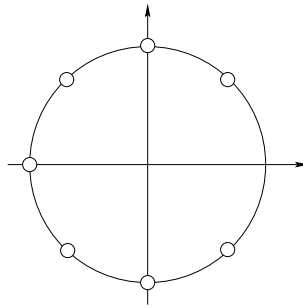


Figura 8.3 Zeri di un filtro comb passa-basso di ordine 8.

È possibile ottenere da un filtro comb da un filtro passa-basso semplicemente eliminando lo zero $z_0 = 1$. Nel caso di filtro comb di ordine 8, gli zeri sono mostrati in Figura 8.3.

La funzione di trasferimento del filtro passa-basso può essere ottenuta dividendo $1 - z^{-M}$ per $1 - z^{-1}$, in modo da eliminare lo zero $z_0 = 0$. Il guadagno della risposta in frequenza, per $M = 8$, è dato in Figura 8.4.

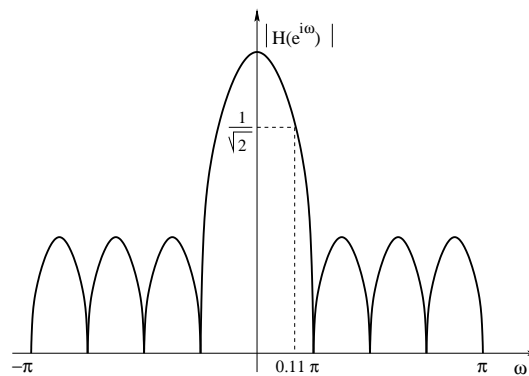


Figura 8.4 Guadagno di un filtro comb passa-basso di ordine 8.

La frequenza di taglio a 3dB, normalizzata, è 0.11π ; le oscillazioni in banda proibita risultano piuttosto alte (≈ 0.3).

La risposta del filtro può essere calcolata in due differenti modi.

1. Ricordando che $\frac{1-z^{-M}}{1-z^{-1}} = 1 + z^{-1} + \dots + z^{-M+1}$, si ottiene:

$$Y(z) = \frac{X(z) + z^{-1}X(z) + \dots + z^{-M+1}X(z)}{M}.$$

Nella rappresentazione in tempo, questo porta al seguente filtro FIR, riconoscibile come il consueto filtro a media mobile:

$$y(n) = \frac{x(n) + x(n-1) + \dots + x(-M+1)}{M}.$$

2. Procedendo direttamente, da si ha:

$$Y(z)(1 - z^{-1}) = \frac{1}{M}X(z)(1 - z^{-M}).$$

Nella rappresentazione in tempo, questo porta al seguente filtro IIR:

$$y(n) = y(n-1) + \frac{1}{M}(x(n) - x(n-M)).$$

Nella realizzazione IIR, viene introdotto un polo in 1 che si sovrappone allo 0: poiché il polo è sulla circonferenza unitaria, il filtro risulta intrinsecamente instabile. Di conseguenza, l'algoritmo che dà la realizzazione IIR deve essere utilizzato solo su brevi intervalli di tempo.

I filtri comb passa-basso hanno prestazioni mediocri; tuttavia, essi sono veloci e di facile realizzazione. Considerando la frequenza di campionamento F_s Hz, la frequenza di taglio del filtro passa-basso comb di ordine 8 risulta $0.11 \frac{F_s}{2}$: per realizzare un filtro comb di data frequenza di taglio basta aggiornare in modo opportuno la frequenza di campionamento.

Esempio 8.2.1.

Si voglia progettare un filtro comb di ordine 8 con frequenza di taglio a 3dB pari a 500 Hz. Ponendo

$$0.11 \frac{F_s}{2} = 500$$

si ottiene $F_s = 9091$ Hz: basta implementare il filtro comb di ordine 8 aggiustando la frequenza di campionamento per ADC e DAC a 9091 Hz.

8.2.2 Filtri Notch (a Intaglio)

I filtri notch hanno l'obiettivo di eliminare una stretta banda di frequenza: il guadagno della risposta in frequenza deve allora essere del tipo in Figura 8.5.

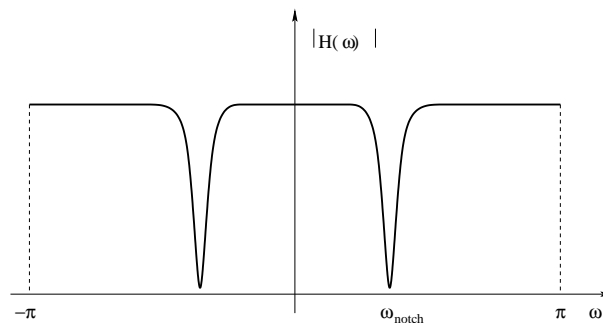


Figura 8.5 Risposta in frequenza di un filtro notch.

Questa caratteristica può essere ottenuta posizionando uno zero $z_0 = e^{i\omega_{\text{notch}}}$ del filtro sulla circonferenza di raggio unitario e posizionando un polo z_p del filtro "vicino" allo zero, ma interno al cerchio di raggio unitario in modo tale che il filtro risulti stabile. Il filtro deve essere completato da un nuovo zero z_0^* e da un nuovo polo z_p^* , coniugati rispettivamente del primo zero e del primo polo, in modo tale che i coefficienti della risposta all'impulso siano numeri reali (vedi Figura 8.6).

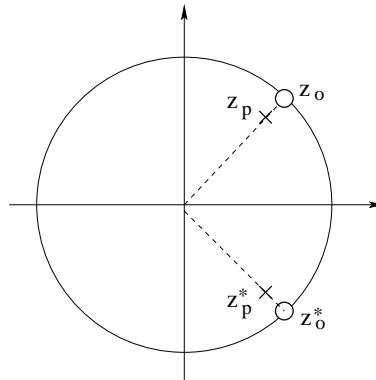


Figura 8.6 Poli e zeri di un filtro notch.

La funzione di trasferimento del filtro specificato sopra è:

$$H(z) = \frac{(z - z_0)(z - z_0^*)}{(z - z_p)(z - z_p^*)}$$

Consideriamo ora il guadagno del filtro, per $0 \leq \omega \leq \pi$. Analizziamo separatamente due casi:

1. ω è "molto vicino" a ω_{notch} , cioè $e^{i\omega} \approx e^{i\omega_{\text{notch}}}$. In questo caso:

$$|H(e^{i\omega})| \approx |H(e^{i\omega_{\text{notch}}})| = 0.$$

2. ω è "molto lontano" da ω_{notch} , cioè $e^{i\omega} - e^{i\omega_{\text{notch}}} \gg 0$. In questo caso, ricordando che $z_0 \approx z_p$ e che $z_0^* \approx z_p^*$, si ha:

$$|H(e^{i\omega})| = \frac{|e^{i\omega} - z_0|}{|e^{i\omega} - z_p|} \cdot \frac{|e^{i\omega} - z_0^*|}{|e^{i\omega} - z_p^*|} \approx 1 \cdot 1 = 1.$$

Di conseguenza, il guadagno del filtro realizzato ha una forma del tipo disegnato in Figura 8.5.

Esempio 8.2.2.

Realizzare un filtro notch a frequenza normalizzata ω_{notch} .

Ponendo ad esempio $z_p = \rho e^{i\omega_{\text{notch}}}$ dove $\rho \approx 1$ ma comunque $\rho < 1$, si ottiene la funzione di trasferimento:

$$H(z) = \frac{1 - z^{-1} \cos \omega_{\text{notch}} + z^{-2}}{1 - z^{-1} 2\rho \cos \omega_{\text{notch}} + \rho^2 z^{-2}}$$

Si ottiene quindi un filtro IIR realizzabile dalla seguente equazione:

$$y(n) = x(n) - 2 \cos \omega_{\text{notch}} x(n-1) + x(n-2) + 2\rho \cos \omega_{\text{notch}} y(n-1) - \rho^2 y(n-2).$$

Una caratteristica importante dei filtri notch è la forma a V dell'incisione: aumentando la distanza tra il polo e lo zero tale V si "apre". È allora conveniente che il polo e lo zero siano più vicini possibile; un limite è dato tuttavia dall'errore di quantizzazione dei coefficienti. L'analisi dell'effetto della quantizzazione è data nel paragrafo 8.4.5 : possiamo qui osservare che la modifica dei coefficienti del filtro in genere si traduce in un cambiamento delle posizioni dei poli e degli zeri, con risultati distruttivi che possono portare alla instabilità del sistema se il polo non risulta più posizionato internamente al cerchio unitario.

8.2.3 Equalizzatore a un Polo

Come abbiamo visto in Sezione 4.4, un convertitore digitale-analogico ZOH ha una risposta in frequenza che diminuisce le ampiezze delle componenti alle frequenze più alte. Molti sistemi contengono allora un equalizzatore che compensa la distorsione del segnale creata dal convertitore. In Figura 8.7 riportiamo sia il guadagno del DAC (linea tratteggiata) sia il guadagno di un equalizzatore ideale.

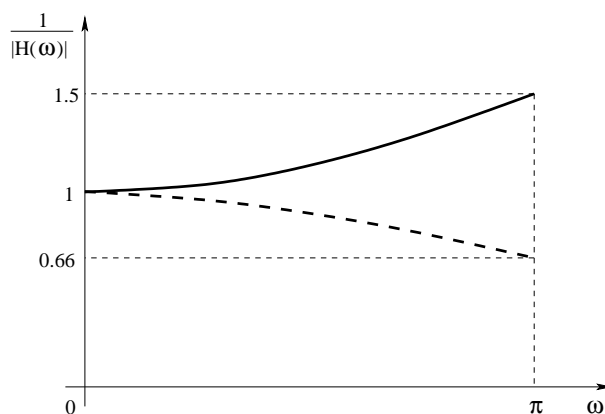


Figura 8.7 Risposta in frequenza ideale di un equalizzatore.

In questo paragrafo costruiamo un semplice equalizzatore, il cui guadagno è una ragionevole approssimazione del guadagno dell'equalizzatore ideale.

Si consideri il filtro con una funzione di trasferimento caratterizzata da un polo in $z_0 = \alpha$, con α reale. Si osservi che, perché il filtro sia stabile, deve essere $-1 < \alpha < 1$.

La funzione di trasferimento, fissata una costante $A > 0$, è:

$$H(z) = \frac{A}{1 - \alpha z^{-1}}.$$

Di conseguenza, il guadagno di questo filtro risulta essere:

$$|H(e^{j\omega})| = \frac{A}{1 + \alpha^2 - 2\alpha \cos \omega}.$$

Posto $\alpha = -0.14$ e $A = 1.13$, la funzione di trasferimento di questo filtro è:

$$H(z) = \frac{1.13}{1 + 0.14z^{-1}}.$$

Tale filtro un filtro risulta essere un filtro IIR causale e stabile, definito dall'equazione:

$$y(n) = 1.13x(n) - 0.14y(n - 1).$$

Il suo guadagno, almeno per $|\omega| < \frac{3}{4}\pi$, è una buona approssimazione del guadagno dell'equalizzatore ideale (con un errore inferiore a 0.01), come si può osservare dalla in Figura 8.8.

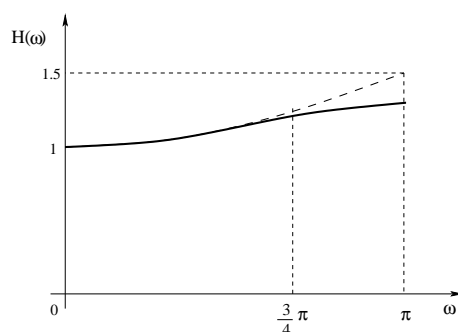


Figura 8.8 Risposta in frequenza di un equalizzatore a un polo.

Il filtro ottenuto risulta allora un buon equalizzatore per il DAC di tipo ZOH.

8.3 Progetto di Filtri Digitali

Un filtro digitale è un sistema LTI a tempo discreto, realizzato con aritmetica a precisione finita. La progettazione di tali filtri richiede dunque tre passi principali:

1. Specificazione delle proprietà desiderate del filtro, ad esempio la frequenza di taglio per un filtro passa-basso; poiché i filtri ideali non sono realizzabili, ci dovremo tuttavia accontentare di una approssimazione e la specifica dovrà riguardare il livello di errore che riteniamo di poter tollerare.
2. Determinazione di un filtro che soddisfa le specifiche stesse; nel caso di un filtro FIR basterà ad esempio determinare i coefficienti che definiscono la sua risposta all'impulso, in un filtro IIR basterà determinare i coefficienti dell'equazione alle differenze finite che lo caratterizzano, oppure determinare i poli e gli zeri della funzione di trasferimento.
3. Realizzazione del sistema con una rete a precisione finita, con eventuale implementazione su un DSP o su un circuito programmabile; è di particolare interesse in questa fase riuscire a controllare gli effetti della quantizzazione dei coefficienti del filtro imposti dall'uso di aritmetica a precisione finita.

Di seguito vengono analizzati più in dettaglio i punti precedenti.

8.3.1 Specifiche di Filtri Digitali

Le proprietà di un filtro digitale sono generalmente ben esprimibili nel dominio delle frequenze. In particolare, per filtri selettivi come filtri passa-basso o passa-banda, le specifiche possono essere date attraverso uno *schema di tolleranza*.

Gli elementi principali di uno schema di tolleranza esemplificati per il filtro passa-basso si riferiscono al guadagno e sono (vedi Figura 8.9):

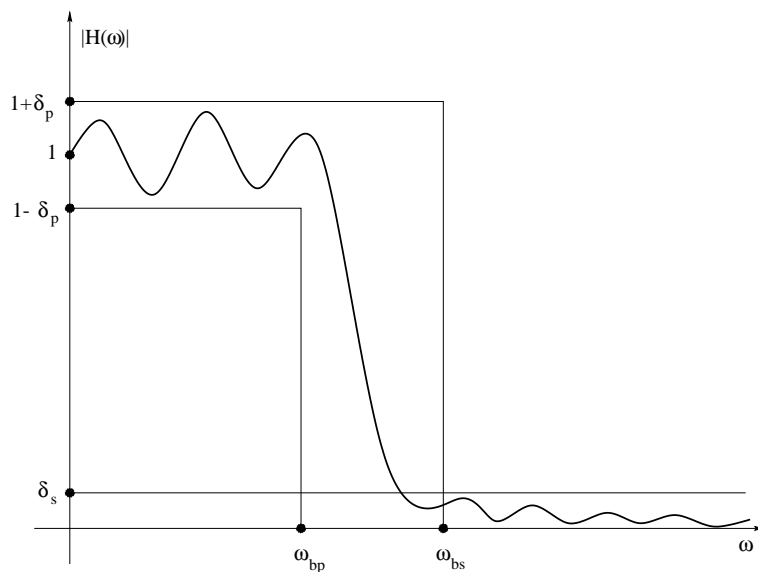


Figura 8.9 Specifiche di un filtro passa-basso.

1. le frequenze ω_{bp} e ω_{bs} che definiscono condizioni di accettabilità per la banda passante e la banda di transizione; queste condizioni possono essere espresse come segue:

$$\omega_{bp} \leq \omega_c < \omega_s \leq \omega_{bs},$$

dove ω_c è la frequenza di taglio e ω_s è la frequenza di stop del filtro. In alternativa, può essere dato un vincolo sulla dimensione $\Delta\omega = \omega_s - \omega_c$ della banda di transizione.

2. le dimensioni massime δ_p e δ_s permesse alle oscillazioni rispettivamente in banda passante e in banda proibita; esse sono usualmente chiamate *deviazioni*. Le deviazioni in banda passante e proibita possono essere espresse anche in decibel:

$$A_p = 20 \log \frac{1 + \delta_p}{1 - \delta_p} \approx 17.4 \delta_p \text{ dB}, \quad A_s = -20 \log \delta_s \text{ dB},$$

dove A_p denota l'oscillazione in banda passante, mentre A_s denota l'attenuazione in banda proibita.

Tutte le frequenze specificate devono essere riferite alla frequenza di campionamento ω_0 ed in particolare devono essere inferiori a $\frac{\omega_0}{2}$ così da soddisfare il criterio di Nyquist. Spesso le frequenze vengono normalizzate rispetto alla frequenza di Nyquist $\frac{\omega_0}{2}$ ed espresse come frazioni di essa ($\frac{\omega}{\omega_0/2}$). Un esempio di schema di tolleranza per un filtro passa-basso è il seguente:

banda passante	ω_c	100 Hz
banda proibita	ω_s	130 Hz
attenuazione (banda proibita)	A_s	40 dB
oscillazione (banda passante)	A_p	0.05 dB
frequenza di campionamento	ω_0	600 Hz

Rispetto al precedente schema, la dimensione della banda di transizione normalizzata è $\Delta\omega = 0.01$ mentre la deviazione in banda proibita è $\delta_s = 10^{-2}$.

Per quanto riguarda la risposta in fase, usualmente sono sufficienti specifiche di tipo qualitativo sulla linearità di tale risposta; criteri quantitativi sono richiesti solo per particolari applicazioni, ad esempio per il progetto di filtri di compensazione della risposta in fase di un sistema.

8.3.2 Progetto di Filtri FIR mediante Finestre

La tecnica di progettazione mediante finestre è basata sull'idea di approssimare un filtro desiderato, eventualmente non causale e con risposta all'impulso $h_d(n)$ di durata infinita, azzerando $h_d(n)$ al di fuori di una finestra temporale di ampiezza N , nella speranza che l'approssimazione sia tanto più buona quanto più la dimensione N della finestra è grande. Più precisamente:

1. si considera il filtro desiderato con risposta all'impulso $h_d(n)$ eventualmente di durata infinita;
2. fissato un intero N , si costruisce il filtro FIR con risposta all'impulso $h_N(n)$ tale che:

$$h_N(n) = \begin{cases} h_d(n), & \text{se } |n| \leq (N-1)/2 \\ 0, & \text{altrimenti.} \end{cases}$$

3. si ottiene infine il filtro FIR causale con risposta all'impulso $h_N(n - N/2)$.

Considerando la *finestra rettangolare* $\text{rett}_N(n)$ data da:

$$\text{rett}_N(n) = \begin{cases} 1, & \text{se } |n| \leq (N-1)/2 \\ 0, & \text{altrimenti,} \end{cases}$$

possiamo scrivere:

$$h_N(n) = \text{rett}_N(n)h_d(n).$$

In altre parole, $h_N(n)$ è ottenuta moltiplicando la risposta all'impulso del filtro che si desidera approssimare per la finestra rettangolare $\text{rett}_N(n)$ di durata finita.

Studiamo ora come il filtro con risposta $h_N(n)$ approssima il filtro con risposta $h_d(n)$, al variare di N . Per semplicità assumiamo che il filtro che desideriamo approssimare sia un filtro passa-basso ideale caratterizzato dalla seguente risposta in frequenza:

$$H_d(e^{i\omega}) = \begin{cases} 1, & \text{se } |\omega| \leq 1 \\ 0, & \text{se } 1 < |\omega| \leq \pi. \end{cases}$$

Denotiamo con $H_N(e^{i\omega})$ la risposta in frequenza del filtro $h_N(n)$. Si possono osservare due fenomeni, di cui in seguito daremo spiegazione:

1. $H_N(e^{i\omega})$ ha una banda di transizione non nulla, la cui ampiezza converge a 0 quando N diverge;
2. $|H_N(e^{i\omega})|$ presenta delle oscillazioni, sia in banda passante sia in banda proibita, la cui ampiezza massima è circa 0.2, indipendentemente dalla dimensione temporale N della finestra.

Il secondo fenomeno è particolarmente negativo, ed è collegato al tipo di convergenza della serie di Fourier (fenomeno di Gibbs); esso tuttavia può essere limitato, anche se non completamente eliminato, scegliendo finestre $w_N(n)$ diverse dalla finestra rettangolare. Tali finestre non dovranno avere discontinuità come quella rettangolare, dovranno essere diverse da 0 nell'intervallo $-(N-1)/2 \leq n \leq (N-1)/2$ e dovranno essere simmetriche rispetto all'origine.

La relazione tra il filtro desiderato e il filtro FIR ottenuto con la finestra $w_N(n)$ è allora:

$$h_N(n) = w_N(n)h_d(n). \quad (8.2)$$

Esempi di finestre comunemente usate oltre alla finestra rettangolare sono elencate di seguito.

Finestra Triangolare (o Bartlett):

$$w_N(n) = \begin{cases} 1 + \frac{2n}{N}, & \text{se } -(N-1)/2 \leq n < 0 \\ 1 - \frac{2n}{N}, & \text{se } 0 \leq n \leq (N-1)/2. \end{cases}$$

Finestra di Hanning:

$$w_N(n) = \frac{1}{2} \left[1 + \cos \frac{2\pi n}{N} \right], \quad |n| \leq (N-1)/2.$$

Finestra di Hamming:

$$w_N(n) = 0.54 + 0.46 \cos \frac{2\pi n}{N}, \quad |n| \leq (N-1)/2.$$

Kaiser ha inoltre proposto una famiglia di finestre $k_N(n, \omega_a)$, parametrizzata da ω_a . La Figura 8.10 illustra il grafico delle finestre sopra elencate.

La progettazione mediante finestre richiede di operare due scelte:

1. il tipo di finestra,
2. la dimensione N dell'intervallo.

Le scelte progettuali sono legate alle seguenti considerazioni:

1. la massima ampiezza delle oscillazioni in banda passante e proibita dipende dal tipo di finestra ma non dipende da N ;
2. l'ampiezza della banda di transizione è inversamente proporzionale ad N , con coefficiente di proporzionalità dipendente dal tipo di finestra.

La Tabella 8.1 riporta l'ampiezza delle oscillazioni in banda passante e proibita e l'ampiezza della banda di transizione (normalizzata alla frequenza di campionamento) per la finestra rettangolare, quella Hanning e quella di Hamming.

Esempio 8.3.1.

Vogliamo determinare col metodo delle finestre i coefficienti di un filtro FIR, che verifica il seguente schema di tolleranza:

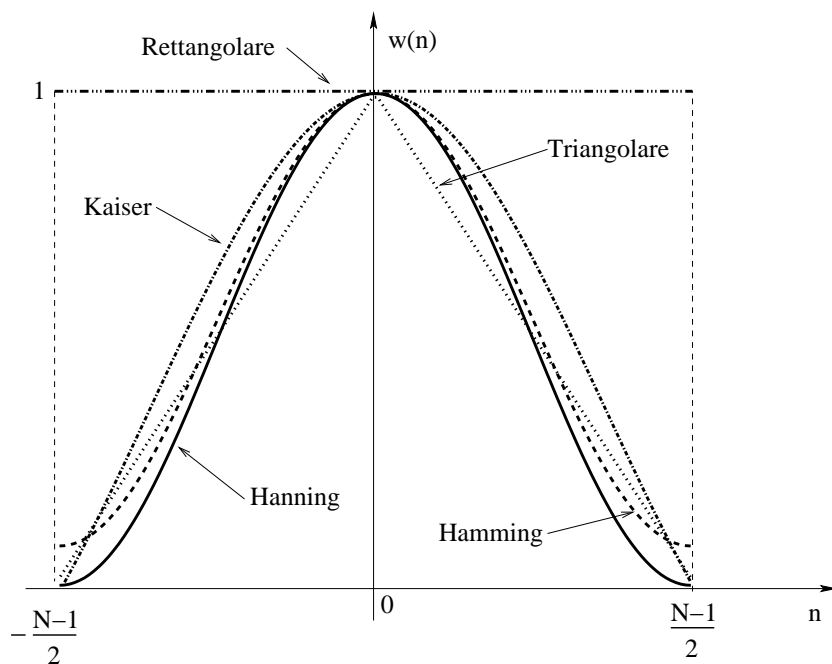


Figura 8.10 Finestre comunemente usate per il progetto di filtri FIR.

Tabella 8.1 Ampiezza delle oscillazioni in banda passante e proibita e l'ampiezza della banda di transizione (normalizzata) di alcune finestre.

Finestra	Osc. A_p (dB)	Att. A_s (dB)	Amp. trans. (Hz)
Rettangolare	0.74	21	$1/N$
Hanning	0.05	44	$3.1/N$
Hamming	0.02	53	$3.3/N$

banda passante	ω_c	500 Hz
banda proibita	ω_s	650 Hz
attenuazione (banda proibita)	A_s	50 dB
oscillazione (banda passante)	A_p	0.05 dB
frequenza di campionamento	ω_0	3 KHz

Dalla Tabella 8.1 si evince che solo la finestra di Hamming permette un'attenuazione di almeno 50 dB.

L'ampiezza della banda di transizione, normalizzata alla frequenza di campionamento, risulta $\Delta\omega = (650 - 500)/3000 = 0.05$; riferendoci alla Tabella 8.1 possiamo concludere che la finestra di Hamming con dimensione $N = 67 > 3.3/0.05$ soddisfa i vincoli dello schema di tolleranza. Tali coefficienti risultano allora ottenibili dalla finestra:

$$w(n) = \begin{cases} 0.54 + 0.46 \cos \frac{2\pi n}{N}, & \text{se } |n| \leq 33 \\ 0, & \text{altrimenti.} \end{cases}$$

Analisi del fenomeno di Gibbs

In questa sottosezione diamo una spiegazione qualitativa dei due fenomeni precedentemente rilevati:

1. $H_N(e^{i\omega})$ ha una banda di transizione non nulla, la cui ampiezza converge a 0 quando N diverge;
2. $|H_N(e^{i\omega})|$ presenta delle oscillazioni, sia in banda passante sia in banda proibita, di ampiezza massima indipendente dalla dimensione temporale N della finestra.

Poiché il filtro $h_N(n)$ ottenuto trattando il filtro desiderato $h_d(n)$ con la finestra $w_N(n)$ è tale che

$$h_N(n) = w_N(n)h_d(n),$$

dal teorema di convoluzione si ottiene

$$H_N(e^{i\omega}) = H_d(e^{i\omega}) * W_N(e^{i\omega}),$$

dove $H_d(e^{i\omega})$ è la risposta in frequenza di un filtro passa-basso desiderato e $W_N(e^{i\omega})$ è la risposta in frequenza della finestra $w_N(n)$.

Per le finestre considerate, la funzione $W_N(e^{i\omega})$ può essere approssimata con funzioni della famiglia $G_{\Delta,a}(\omega)$ (con Δ e a reali positivi), mostrata in Figura 8.11 assieme al grafico della risposta in frequenza del filtro ideale passa-basso:

$$G_{\Delta,a}(\omega) = \begin{cases} \frac{1+a}{\Delta}, & \text{se } |\omega| \leq \Delta \\ -\frac{a}{\Delta}, & \text{se } \Delta < |\omega| \leq 2\Delta \\ 0, & \text{se } |\omega| > 2\Delta. \end{cases}$$

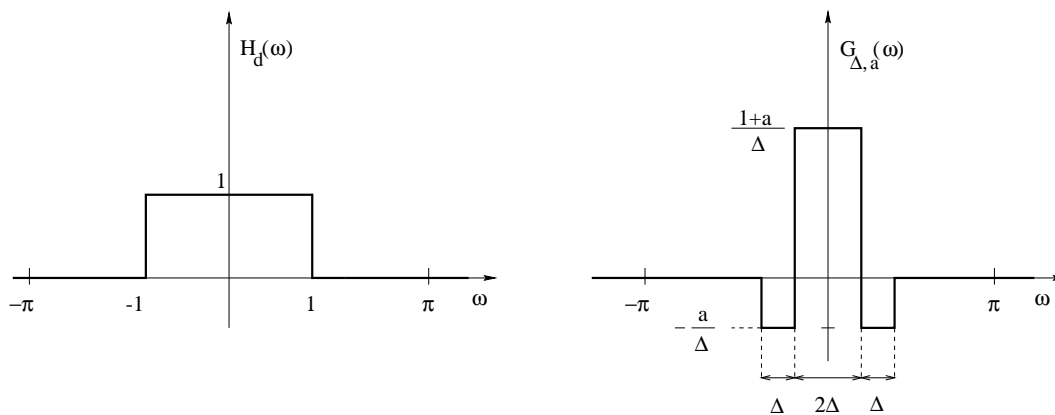


Figura 8.11 Grafico delle funzioni $H_d(\omega)$ e $G_{\Delta,a}(\omega)$.

In altre parole, $G_{\Delta,a}(\omega)$ è composta da un lobo rettangolare centrale di altezza $\frac{1+a}{\Delta}$ e base 2Δ e da due lobi laterali, anch'essi rettangolari, di altezza $\frac{a}{\Delta}$ e base Δ ; siamo qui interessati a studiare il caso $0 < a \ll 1$, per cui vale che $a \approx \frac{\text{altezza lobo laterale}}{\text{altezza lobo centrale}}$.

Esempio 8.3.2.

Per la finestra rettangolare vale:

$$\text{RETT}_N(\omega) = \sum_{n=0}^N e^{i\omega n} = \frac{1 - e^{i\omega(N+1)}}{1 - e^{i\omega}} = e^{-i\omega \frac{N+1}{2}} \frac{\sin(\omega(N+1)/2)}{\sin(\omega/2)}$$

A meno di una variazione di fase, il grafico della funzione $\frac{\sin(\omega N/2)}{\sin(\omega/2)}$ è mostrato in Figura 8.12. Trascurando i lobi laterali ad eccezione dei tre lobi centrali, possiamo

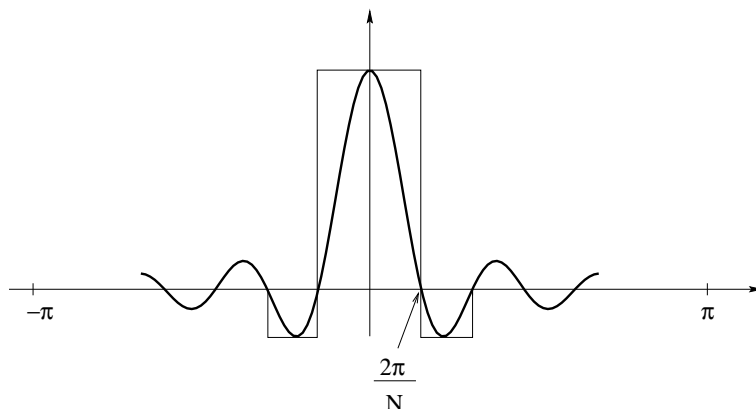


Figura 8.12 Grafico della funzione $\frac{\sin(\omega N/2)}{\sin(\omega/2)}$.

approssimare la funzione $\frac{\sin(\omega N/2)}{\sin(\omega/2)}$ con $G_{\Delta,a}(\omega)$ (come mostrato nella figura), dove $\Delta = \frac{2\pi}{N}$ e $a \approx \frac{2}{3}\pi$, poiché l'altezza del lobo centrale è N e quella dei lobi laterali è $\frac{2}{3}\pi N$.

Le principali caratteristiche delle varie finestre, che permettono di approssimare le risposte in frequenza con una funzione del tipo $G_{\Delta,a}(\omega)$ sono elencate nelle Tabella 8.2

Tabella 8.2 Caratteristiche delle principali finestre.

Finestra	$20 \log a$ (dB)	Larghezza lobo centrale
Rettangolare	-13	$4\pi/N$
Triangolare	-25	$8\pi/N$
Hanning	-37	$8\pi/N$
Hamming	-41	$8\pi/N$

La convoluzione $H_N(e^{i\omega}) = H_d(e^{i\omega}) * G_{\Delta,a}(\omega)$ può essere facilmente calcolata e risulta essere la funzione lineare a pezzi di Figura 8.13. Dal grafico precedente osserviamo che il filtro passa-basso ideale viene approssimato da un filtro che ha le seguenti caratteristiche:

1. la massima ampiezza delle oscillazioni sia in banda passante sia in banda proibita è a ; essa in particolare non dipende da Δ ;

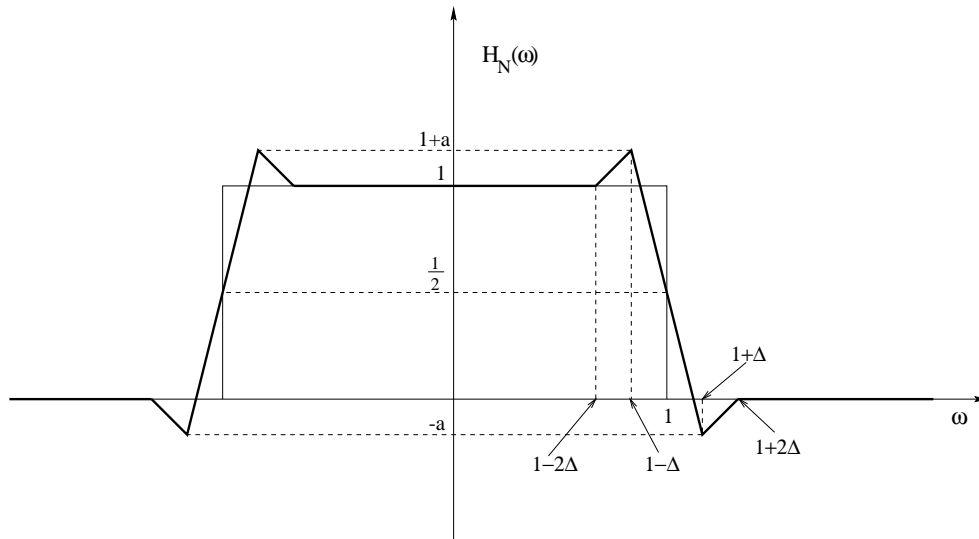


Figura 8.13 Grafico della funzione $H_N(e^{i\omega})$, ottenuta dalla convoluzione di $H_d(e^{i\omega})$ e $G_{\Delta,a}(\omega)$.

2. la dimensione della banda di transizione è proporzionale a Δ .

Possiamo di conseguenza trarre le seguenti conclusioni:

- la dimensione della banda di transizione del filtro approssimato è proporzionale a $1/N$, poiché la larghezza del lobo centrale di una maschera è inversamente proporzionale a N .
- a parità di dimensione della durata temporale delle finestre, la dimensione della banda di transizione del filtro ottenuto dalla finestra rettangolare è metà di quella ottenuta dalle finestre tringolari, di Hanning e di hamming;
- la massima ampiezza delle oscillazioni create dalla finestra rettangolare ($a \approx 0.2$) è molto maggiore di quella delle oscillazioni create dalle altre finestre ($a \approx 0.01$ per la finestra di Hamming). Questo spiega perché la finestra rettangolare, pur più semplice da realizzare, è scarsamente usata.

8.3.3 Progetto di Filtri FIR col Metodo Ottimo

Come abbiamo visto, il metodo della finestra crea filtri con un “picco” nelle oscillazioni in corrispondenza della frequenza di taglio e frequenza di stop, come evidenziato nella Figura 8.14.

Risulta naturale tentare di abbassare l’altezza massima delle oscillazioni “spalmandole” su tutta la banda passante e proibita: questa è l’idea che sta alla base della costruzione del filtro ottimo, nel senso di Chebishev.

Il metodo considera al solito la risposta all’impulso $h_d(n)$ del filtro che si desidera approssimare e che ha una ideale risposta in frequenza $H_d(e^{i\omega})$; fissato N , si prende in considerazione l’insieme F_N tutti i filtri FIR h di ordine N , cioè con risposta all’impulso descrivibile da N valori $h = (h(0), \dots, h(N-1))$. Se $H(e^{i\omega})$ la risposta in frequenza del

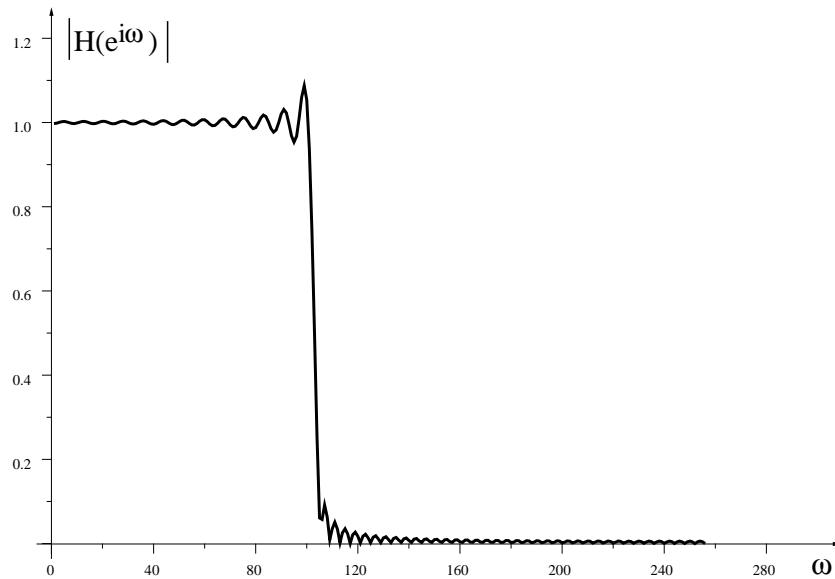


Figura 8.14 Modulo della risposta in frequenza di un filtro passa-basso ottenuto col metodo delle finestre.

generico filtro h , si definisce come errore di approssimazione di h_d con h , la quantità

$$e(h, h_d) = \max_{\omega} |H_d(e^{i\omega}) - H_h(e^{i\omega})| \Psi(\omega),$$

dove $\Psi(\omega)$ è una opportuna funzione peso che consente di trattare il caso in cui i vincoli imposti alle oscillazioni in banda passante e proibita siano diversi.

Il filtro ottimo \hat{h} è il filtro di F_N che minimizza l'errore:

$$\hat{h} = \arg \min_{h \in F_N} e(h, h_d).$$

Si può dimostrare che il filtro ottimo è caratterizzato da "equioscillazioni" in banda passante e banda proibita, come mostrato nella Figura 8.15

Esistono vari algoritmi per risolvere il problema di minimo; i principali sistemi CAD implementano procedure che, avendo in ingresso N e le specifiche del filtro desiderato $h_d(n)$, danno in uscita i coefficienti di \hat{h} .

Poiché in generale il numero N è incognito mentre lo schema di tolleranza dà la dimensione della banda di transizione normalizzata e le deviazioni δ_s e δ_p , è necessario preliminarmente effettuare una stima di N . Per filtri passa-basso N può essere stimato dalla seguente relazione empirica:

$$N \approx \frac{D(\delta_p, \delta_s)}{\Delta_F} - f(\delta_p, \delta_s) \Delta_F + 1$$

dove Δ_F è l'ampiezza della banda di transizione normalizzata alla frequenza di campionamento, δ_p è la deviazione in banda passante e δ_s la deviazione in banda proibita.

Inoltre:

$$D(\delta_p, \delta_s) = \log_{10} \delta_s [a_1 (\log_{10} \delta_p)^2 + a_2 \log_{10} \delta_p + a_3] \\ + [a_4 (\log_{10} \delta_p)^2 + a_5 \log_{10} \delta_p + a_6]$$

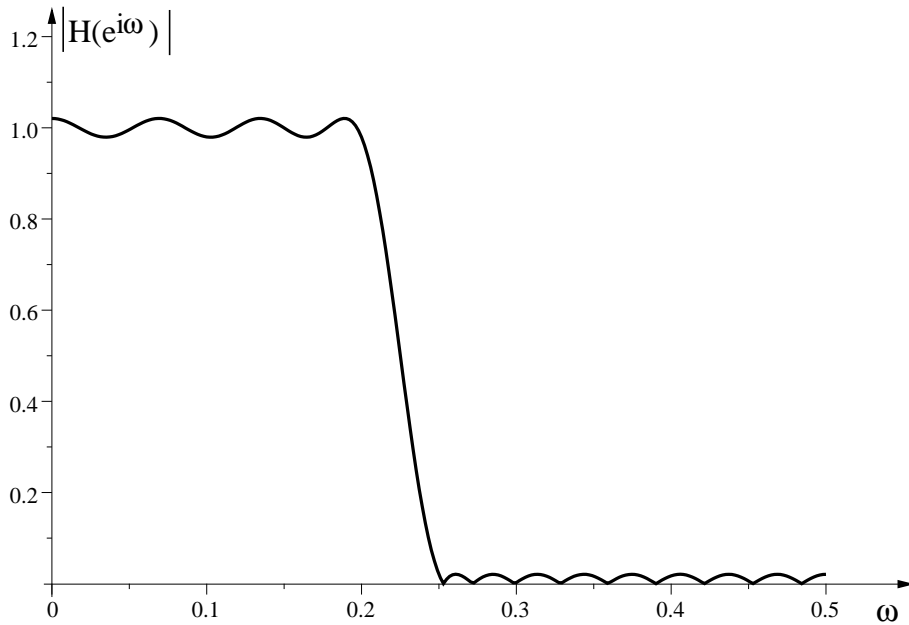


Figura 8.15 Modulo della risposta in frequenza di un filtro passa-basso ottimale di ordine $N = 20$.

e

$$f(\delta_p, \delta_s) = 11.012 + 0.5124(\log_{10} \delta_p - \log_{10} \delta_s),$$

con

$$\begin{aligned} a_1 &= 5.4 \times 10^{-3}, & a_2 &= 7.11 \times 10^{-2}, & a_3 &= -4.76 \times 10^{-1}, \\ a_4 &= 2.66 \times 10^{-3}, & a_5 &= -5.94 \times 10^{-1}, & a_6 &= -4.28 \times 10^{-1}. \end{aligned}$$

Discutiamo qui brevemente come determinare col metodo ottimale i coefficienti di filtri specificati attraverso uno schema di tolleranza, utilizzando il sistema Scilab. In Scilab è implementato l'algoritmo di Remez, che può essere richiamato con un comando che, nella sua forma basilare, è:

$$b = \text{eqfir}(N, F, M, W)$$

dove:

1. N denota il numero di coefficienti del filtro;
2. F denota un vettore di intervalli di frequenze normalizzate (bande);
3. M denota il vettore dei guadagni desiderati in ogni banda;
4. W denota un vettore dei pesi relativi delle ampiezze delle oscillazioni nelle varie bande;
5. b denota il vettore dei coefficienti del filtro ottimo.

Va segnalato che la normalizzazione qui è riferita alla frequenza di campionamento.

Esempio 8.3.3.

Determinare il filtro passa-banda ottimo a 50 coefficienti, quando lo schema di tolleranza è:

banda passante	500-700 Hz
ampiezza bande di transizione	100 Hz
frequenza di campionamento	2000 Hz

La specifica richiede che il guadagno sia 0 nella banda 0-400 Hz, sia 1 nella banda 500-700 Hz, sia nuovamente 0 nella banda 800-1000 Hz. Si ricorda che la frequenza di Nyquist risulta 1000 Hz. Normalizzando rispetto alla frequenza di campionamento, si ottengono i seguenti intervalli: 0-0.2, 0.25-0.35, 0.4-0.5

Un programma scritto in Scilab che calcola i coefficienti del filtro ottimo e dà il grafico del guadagno è:

```

N = 50
F = [0 0.2; 0.25 0.35; 0.4 0.5]
M = [0 1 0]
W = [1 1 1]
b = eqfir(N,F,M,W)
[H,f] = frmag(b,512)
plot(f,abs(H))

```

8.3.4 Progetto di Filtri IIR da Filtri Analogici

Un approccio alla progettazione di filtri IIR, consiste nel trasformare un filtro analogico (per esempio un filtro di Butterworth) in un corrispondente filtro digitale che soddisfi le specifiche.

Supponiamo, ad esempio, che un filtro analogico causale sia descritto dalla seguente equazione differenziale:

$$\sum_{k=0}^{M-1} c_k \frac{d^k}{dt^k} g(t) = \sum_{k=0}^{L-1} d_k \frac{d^k}{dt^k} f(t),$$

dove $f(t)$ è il segnale di ingresso e $g(t)$ quello di uscita. Una semplice idea per ottenere un sistema digitale IIR con caratteristiche simili a quello analogico consiste nell'approssimare l'operazione di derivata con quello di differenza finita, seguendo un approccio tipico dell'analisi numerica. Ci si può aspettare che con alte frequenze di campionamento l'approssimazione risulti accettabile.

A questo riguardo, fissata la frequenza di campionamento $\frac{1}{T}$ Hz, poniamo:

$$x(n) = f(nT), \quad y(n) = g(nT).$$

Approssimando la derivata di una funzione $A(t)$ col rapporto incrementale, si ha che:

$$\left. \frac{d}{dt} A(t) \right|_{t=nT} \approx \frac{A(nT) - A((n-1)T)}{T}.$$

Possiamo allora sostituire $\frac{d}{dt} g(t)$ con $\nabla y(n) = \frac{y(n) - y(n-1)}{T}$ e $\frac{d^k}{dt^k} g(t)$ con $\nabla^k y(n)$, dove $\nabla^0 y(n) = y(n)$ e $\nabla^k y(n) = \nabla(\nabla^{k-1} y(n))$ se $k > 0$. Sostituzioni analoghe possono essere fatte per $\frac{d^k}{dt^k} f(t)$.

Il sistema digitale corrispondente a quello analogico viene allora descritto dalla seguente equazione:

$$\sum_{k=0}^{M-1} c_k \nabla^k y(n) = \sum_{k=0}^{L-1} d_k \nabla^k x(n)$$

Esempio 8.3.4.

Si consideri il sistema analogico descritto dall'equazione differenziale:

$$g(t) = ag'(t) + bg''(t) + f(t).$$

Posto $x(n) = f(nT)$ e $y(n) = g(nT)$, risulta:

$$\begin{aligned} \nabla y(n) &= \frac{y(n) - y(n-1)}{T} \\ \nabla^2 y(n) &= \nabla(\nabla y(n)) = \nabla \frac{y(n) - y(n-1)}{T} = \frac{\nabla y(n) - \nabla y(n-1)}{T} \\ &= \frac{y(n) - 2y(n-1) + y(n-2)}{T^2}. \end{aligned}$$

Da cui si ricava il seguente sistema digitale:

$$y(n) = \frac{T^2}{T^2 - aT - b} \left[\frac{T-2b}{T^2} y(n-1) + \frac{b}{T^2} y(n-2) + x(n) \right].$$

8.4 Realizzazione di Filtri Digitali

In un filtro FIR il valore $y(n)$ dell'uscita al tempo n è dato da:

$$y(n) = \sum_{k=0}^{M-1} b_k x(n-k).$$

Più in generale, in un filtro IIR il valore $y(n)$ dell'uscita al tempo n può essere riscritto nella forma:

$$y(n) = \sum_{k=1}^{L-1} c_k y(n-k) + \sum_{k=0}^{M-1} d_k x(n-k).$$

In ogni caso, per il calcolo di $y(n)$ si richiede:

1. la disponibilità dei valori di uscita ai tempi $n-1, \dots, n-L+1$ e la disponibilità dei valori di ingresso ai tempi $n, \dots, n-M+1$;
2. la disponibilità permanente dei coefficienti moltiplicativi c_1, \dots, c_{L-1} e d_1, \dots, d_{M-1} ;
3. l'effettuazione di moltiplicazioni e somme in accordo alla forma generale.

Queste operazioni possono essere implementate su un calcolatore tradizionale o su hardware specializzato; in quest'ultimo caso è conveniente rappresentare una procedura di calcolo del filtro mediante una rete i cui nodi sono etichettati con gli operatori di somma, moltiplicazione per costante e ritardo temporale, come descritto in Figura 8.16.

In particolare, considereremo reti con un unico nodo di ingresso privo di predecessori, un unico nodo di uscita privo di successori e nodi interni del tipo somma, moltiplicazione per costante e ritardo temporale.

I problemi che affrontiamo sono di due tipi:

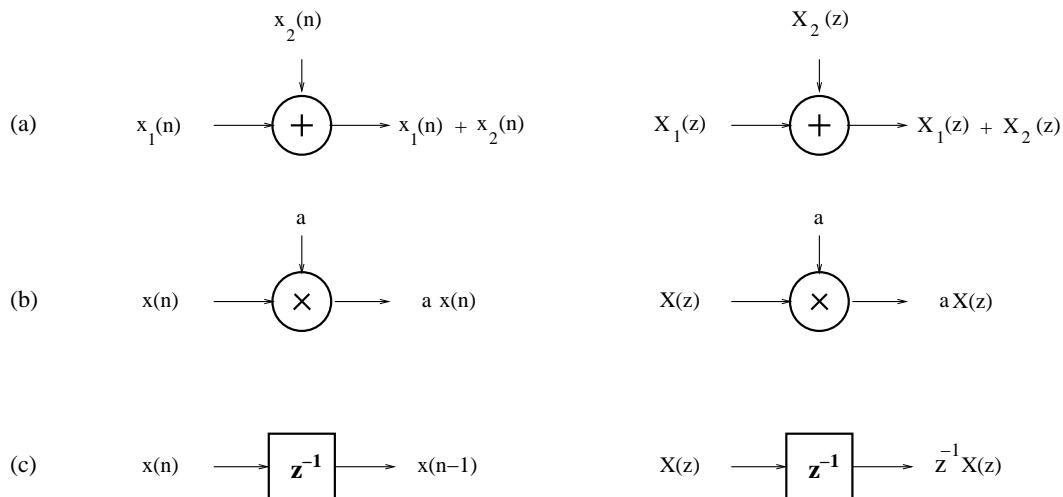


Figura 8.16 Operatore di somma (a), moltiplicazione per costante (b) e ritardo temporale (c), rispettivamente nel dominio del tempo n e nella trasformata z .

Analisi: data una rete, trovare la funzione di trasferimento del filtro digitale realizzato dalla rete stessa.

Sintesi: dato il filtro specificato ad esempio dall'equazione alle differenze o equivalentemente dalla sua funzione di trasferimento, determinare una rete che lo realizzi.

8.4.1 Analisi di Reti

Per quanto riguarda la problematica di analisi, data una rete R , la funzione di trasferimento $H_R(z)$ del filtro realizzato da R può essere ottenuta dalla seguente procedura:

Input: Una rete R con un unico ingresso e un'unica uscita;

1. Associa all'arco di ingresso la funzione $X(z)$, trasformata z del segnale di ingresso $x(n)$, all'arco di uscita la funzione $Y(z)$, trasformata z del segnale di uscita $y(n)$, all'arco di uscita di ogni nodo interno k la funzione ausiliaria $W_k(z)$.
2. Per ogni nodo della rete, ad esclusione del nodo di ingresso, costruisci un'equazione come segue:
 - a. Se il nodo k è un nodo di ritardo temporale e la funzione associata all'arco di ingresso è $W(z)$, poni $W_k(z) = z^{-1}W(z)$.
 - b. Se il nodo k è un nodo di somma e le funzioni associate ai due archi di ingresso sono $W(z)$ e $S(z)$, poni $W_k(z) = W(z) + S(z)$.
 - c. Se il nodo k è un nodo di moltiplicazione per la costante a e la funzione associata all'arco di ingresso è $W(z)$, poni $W_k(z) = aW(z)$.
3. Elimina dalle equazioni ottenute le funzioni ausiliarie $W_k(z)$ associate ai nodi interni, ottenendo una relazione del tipo $Y(z) = H_R(z)X(z)$.

Output: La funzione $H_R(z)$ razionale in z^{-1} .

Esempio 8.4.1.

Si consideri la rete $R1$ indicata in Figura 8.17.

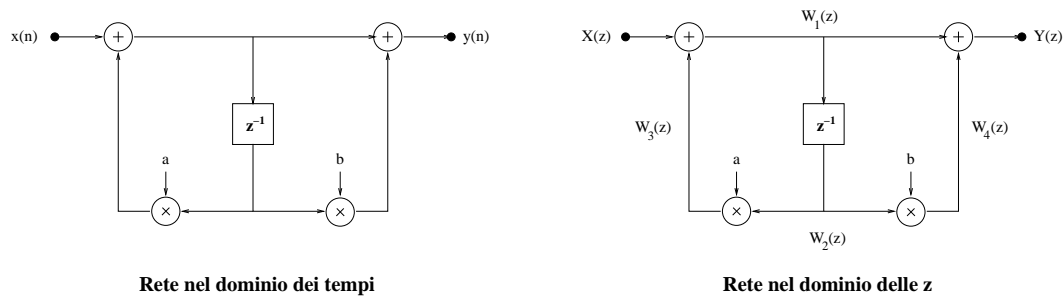


Figura 8.17 Esempio di rete.

Per ottenere la relazione ingresso-uscita $Y(z) = H_{R1}(z)X(z)$ e quindi la funzione di trasferimento $H_{R1}(z)$ del filtro da essa realizzato, è sufficiente scrivere le equazioni ai nodi ed eliminare successivamente le funzioni ausiliarie, come segue.

Le equazioni sono le seguenti:

$$\begin{aligned} W_1(z) &= X(z) + W_3(z) \\ W_2(z) &= z^{-1}W_1(z) \\ W_3(z) &= aW_2(z) \\ W_4(z) &= bW_2(z) \\ Y(z) &= W_1(z) + W_4(z), \end{aligned}$$

da cui si ricava che:

$$H_{R1}(z) = \frac{Y(z)}{X(z)} = \frac{1 + bz^{-1}}{1 - az^{-1}}.$$

Dalla precedente equazione si deriva facilmente:

$$Y(z) = az^{-1}Y(z) + X(z) + bz^{-1}X(z).$$

Antitrasformando, otteniamo l'equazione del filtro IIR realizzato dalla rete:

$$y(n) = ay(n-1) + x(n) + bx(n-1).$$

8.4.2 Reti Modulari a più Ingressi e Uscite

Un rete complessa viene più facilmente analizzata se può essere vista come rete di piccola dimensione, le cui componenti sono a loro volta reti. Questo permette di fattorizzare l'analisi in:

- determinazione della funzione di trasferimento delle varie componenti (moduli);
- determinazione della funzione di trasferimento della rete a partire da quelle delle sue componenti.

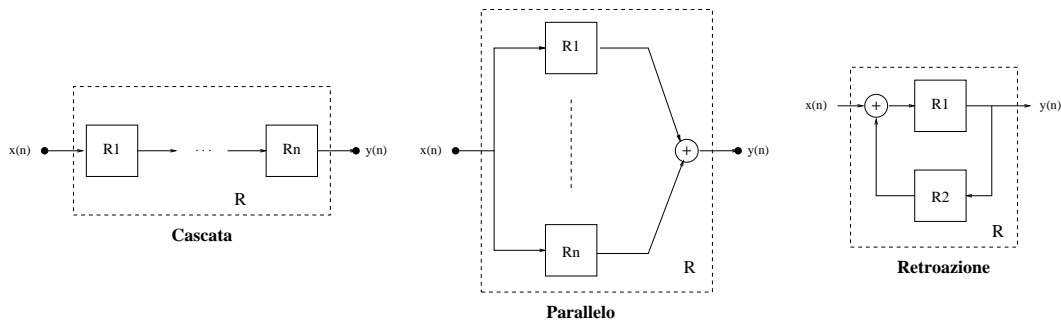


Figura 8.18 Composizione in cascata, parallelo e retroazione.

Reti con caratteristiche di modularità possono essere costruite in modo naturale partendo da reti-base, come la moltiplicazione per una costante o il ritardo, applicando poi semplici operazioni permettono di associare a due o più reti una nuova rete. Alcune di queste operazioni sono illustrate in Figura 8.18.

Composizione sequenziale (o cascata): date m reti R_1, \dots, R_m con funzioni di trasferimento rispettivamente $H_{R_1}(z), \dots, H_{R_m}(z)$, la cascata di esse è la rete R che si ottiene ponendo in ingresso alla rete R_{i+1} l'uscita della rete R_i ($1 \leq i < m$); la rete R ha come funzione di trasferimento $H_R(z) = H_{R_1}(z) \times \dots \times H_{R_m}(z)$.

Composizione parallela: date m reti R_1, \dots, R_m con funzioni di trasferimento rispettivamente $H_{R_1}(z), \dots, H_{R_m}(z)$, la composizione parallela di esse è la rete R che si ottiene ponendo lo stesso ingresso alle reti R_1, \dots, R_m e sommando le uscite; la rete R ha come funzione di trasferimento $H_R(z) = H_{R_1}(z) + \dots + H_{R_m}(z)$.

Retroazione: date due reti R_1 e R_2 con funzioni di trasferimento $H_{R_1}(z)$ e $H_{R_2}(z)$, la retroazione di R_2 su R_1 è la rete R che si ottiene ponendo in ingresso a R_1 la somma dell'input e dell'uscita di R_2 , e ponendo l'uscita di R_1 in ingresso a R_2 ; la rete R ha come funzione di trasferimento $H_R(z) = \frac{H_{R_1}(z)}{1 - H_{R_2}(z)}$.

Esempio 8.4.2.

L'integratore $\sum_{k=0}^n x(k)$ è descritto dalla rete riportata in Figura 8.19.

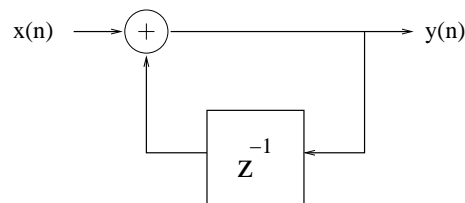


Figura 8.19 Integratore numerico.

Esso risulta dunque la retroazione del ritardo sulla rete identità ed ha come funzione di trasferimento $\frac{1}{1-z^{-1}}$.

Esempio 8.4.3.

La rete in Figura 8.20 è la composizione parallela di un integratore e di un ritardo. La sua funzione di trasferimento è $\frac{1}{1-z^{-1}} + z^{-1}$.

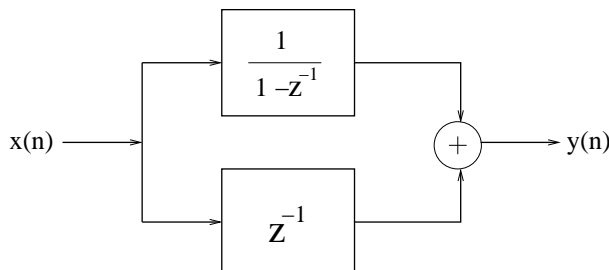


Figura 8.20 Composizione parallela.

Fino ad ora abbiamo considerato reti con un solo ingresso e una sola uscita. Talvolta è necessario prendere in considerazione reti che hanno più di un ingresso o più di una uscita. Un semplice ma importante esempio è il modello di rumore additivo mostrato in Figura 8.21.

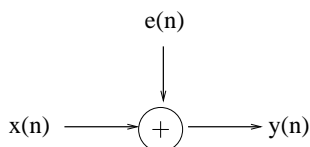


Figura 8.21 Rumore additivo.

Questa rete ha due ingressi, il segnale $x(n)$ e il disturbo $e(n)$. L'uscita $y(n)$ è data dalla somma $x(n) + e(n)$.

Le considerazioni fatte su reti a un ingresso e un'uscita si estendono facilmente a reti arbitrarie. Consideriamo reti con segnali di ingresso $x_1(n), \dots, x_a(n)$ e con segnali di uscita $y_1(n), \dots, y_b(n)$; dette $X_1(z), \dots, X_a(z)$ le trasformate zeta dei segnali di ingresso e $Y_1(z), \dots, Y_b(z)$ le trasformate zeta dei segnali di uscita, si possono facilmente determinare funzioni razionali $R_{jk}(z)$ ($1 \leq j \leq a, 1 \leq k \leq b$) tali che:

$$Y_j(z) = \sum_{k=1}^a R_{jk}(z) X_k(z), \quad (1 \leq j \leq b).$$

La matrice $A(z) = [R_{jk}(z)]$ è detta *matrice di trasferimento* della rete, e gran parte delle considerazioni fatte per le funzioni di trasferimento possono essere estese alle matrici di trasferimento.

Per esempio, la matrice di trasferimento della rete ottenuta mettendo in cascata reti S_1 e S_2 con matrici di trasferimento $A_1(z)$ e $A_2(z)$ è il prodotto $A_1(z)A_2(z)$, mentre quella della rete ottenuta componendo in parallelo reti S_1 e S_2 con matrici di trasferimento $A_1(z)$ e $A_2(z)$ è la somma $A_1(z) + A_2(z)$.

Una notevole differenza è tuttavia legata al fatto che il prodotto di matrici non è in generale commutativo, mentre il prodotto di funzioni di trasferimento lo è: nel caso di reti

a più ingressi e uscite, il risultato di una cascata di reti dipende generalmente dall'ordine, mentre nel caso di reti a un ingresso e un'uscita il risultato non dipende dall'ordine (abbiamo vista una applicazione di questo principio in Sezione 4.4.1).

Come applicazione dei concetti esposti, nella prossima sezione analizziamo una importante classe di filtri digitali: i modulatori sigma-delta.

8.4.3 Analisi del Modulatore Sigma-Delta (SDM)

Il modulatore sigma-delta è un'interessante applicazione alla conversione analogico digitale dei principi del sovracampionamento. Il modulatore sigma-delta del primo ordine è già stato introdotto in Sezione 4.3.2, e può essere descritto dalla rete mostrato in Figura 8.22.

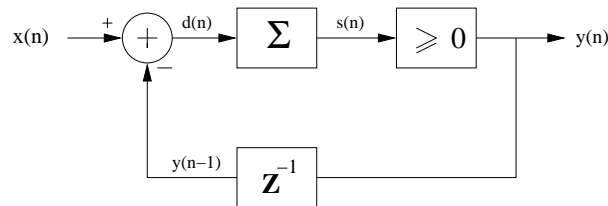


Figura 8.22 SDM del primo ordine.

Il modulo "Σ" è un integratore, così che $s(n) = \sum_{k=-\infty}^n d(k)$, mentre il modulo denotato " ≥ 0 " risulta essere un quantizzatore a 1 bit. Denotiamo con $e(n) = y(n) - s(n)$ l'errore di quantizzazione, così che $y(n) = e(n) + s(n)$. Evidenziando l'errore di quantizzazione come disturbo, il circuito può essere equivalentemente descritto dalla rete in Figura 8.23.

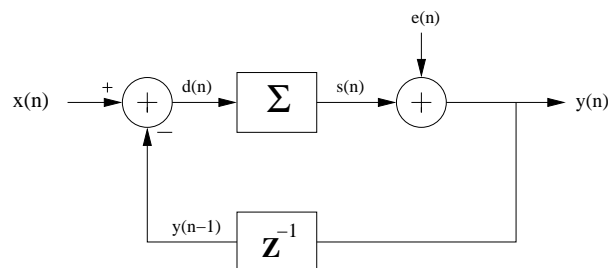


Figura 8.23 SDM con errore di quantizzazione.

Poiché la risposta all'impulso unitario del nodo integratore "Σ" è il gradino $u(n)$, la funzione di trasferimento di "Σ" risulta essere $\sum_{n=0}^{\infty} z^{-n} = \frac{1}{1-z^{-1}}$.

Dette $X(z)$, $Y(z)$, $E(z)$ le trasformate zeta di $x(n)$, $y(n)$, $e(n)$ rispettivamente, la relazione tra $X(z)$, $Y(z)$, $E(z)$ è descritta dalla rete in Figura 8.24.

Vale quindi che $Y(z) = E(z) + \frac{1}{1-z^{-1}}(X(z) - z^{-1}Y(z))$; risolvendo rispetto a $Y(z)$ si ottiene:

$$Y(z) = X(z) + E(z)(1 - z^{-1}).$$

Antitrasformando, l'equazione precedente mostra che l'uscita $y(n)$ è ottenuta aggiungendo all'ingresso $x(n)$ il rumore di quantizzazione $e(n)$ filtrato con un filtro la cui funzione di trasferimento è $1 - z^{-1}$. Come mostrato in Sezione 7.4.1, tale filtro risulta essere un

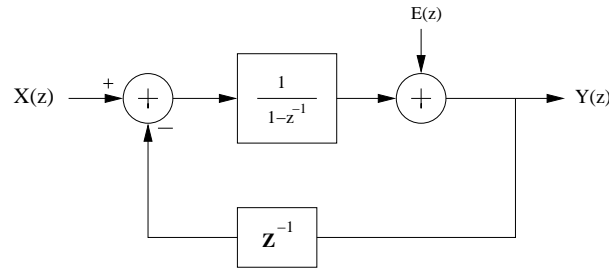


Figura 8.24 SDM in termini di trasformate zeta.

filtro passa-alto il cui guadagno $G(\omega)$, illustrato in Figura 8.25, è:

$$G(\omega) = 4 \sin^2 \frac{\omega}{2},$$

dove $\omega = \frac{2\pi f}{F_s}$ è la frequenza normalizzata alla frequenza di campionamento F_s . L'effetto del filtro è di attenuare il rumore alle basse frequenze, aumentandolo invece alle alte frequenze (noise-shaping).

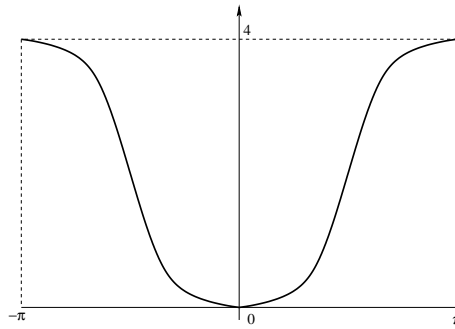


Figura 8.25 Guadagno del filtro passa-alto con funzione di trasferimento $1 - z^{-1}$.

Supponiamo che il modulatore lavori alla frequenza di campionamento F_s processando segnali a banda limitata da f_{\max} , con $f_{\max} \ll F_s$.

Ricordiamo da Sezione 4.3.1 che il rumore di quantizzazione delle componenti armoniche con frequenze comprese tra f e $f + df$ è $\frac{\sigma_e^2}{F_s} df$. Dopo l'applicazione del filtro passa-alto, tale rumore risulta essere $4 \sin^2 \left(\frac{\pi f}{F_s} \right) \frac{\sigma_e^2}{F_s} df$. Se infine applichiamo al segnale $y(n)$ un filtro passa-basso con frequenza di taglio f_{\max} , la potenza complessiva del rumore granulare risulta allora:

$$\int_{-f_{\max}}^{f_{\max}} 4 \sin^2 \left(\frac{\pi f}{F_s} \right) \frac{\sigma_e^2}{F_s} df \approx \frac{\pi^2}{3} \left(\frac{2f_{\max}}{F_s} \right)^3 \sigma_e^2.$$

Il rapporto potenza del segnale-potenza del rumore viene dunque migliorato dal modulatore (e dal filtro passa-basso in uscita), di un fattore $\frac{3}{\pi^2} \left(\frac{F_s}{2f_{\max}} \right)^3$. In termini di decibel, si ottiene:

Fatto 8.2. Il modulatore SDM di ordine 1 produce un miglioramento del rapporto segnale-rumore SQNR di $30 \log_{10} \frac{F_s}{2f_{\max}} - 5$ dB.

Si può ottenere una ulteriore diminuzione del rumore di quantizzazione aumentando l'ordine del modulatore. Per modulatori di ordine N la trasformata zeta $Y(z)$ dell'uscita vale:

$$Y(z) = X(z) + E(z)(1 - z^{-1})^N$$

La riduzione complessiva del rumore di quantizzazione è data dal seguente

Fatto 8.3. Il modulatore SDM di ordine N produce un miglioramento del rapporto segnale-rumore SQNR di $10(2N + 1) \log_{10} \frac{F_s}{2f_{\max}} + 10 \log_{10}(2N + 1) - 20N \log_{10} \pi$ dB.

Indipendentemente dall'ordine del modulatore, si può osservare che il filtro passa-alto abbatte le componenti in frequenza all'interno della banda limitata da f_{\max} ma esalta le componenti esterne a tale banda. E' necessario rimuovere tale rumore con un ulteriore filtro passa-basso; per le alte frequenze in gioco, si preferisce ottenere questo risultato mediante il filtro contenuto in un decimatore, cosa che offre l'ulteriore vantaggio di ridurre la frequenza al tasso di Nyquist $2f_{\max}$. Il sistema ADC complessivo è mostrato in Figura 8.26.

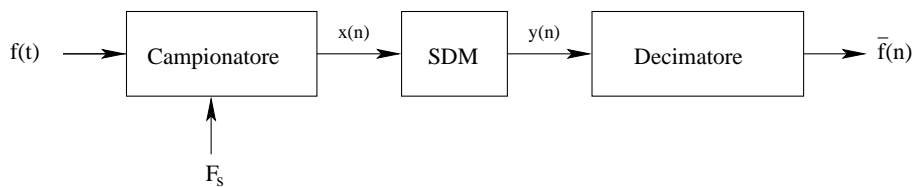


Figura 8.26 Sistema ADC complessivo.

L'operazione di decimazione trasforma il segnale binario $y(n)$ a frequenza F_s in un segnale quantizzato con m bit a frequenza $2f_{\max}$. La effettiva lunghezza di parola m del convertitore è quella equivalente alla risoluzione ottenibile con il miglioramento in SQNR offerto dal modulatore e dalla decimazione.

Esempio 8.4.4.

Un sistema audio per il trattamento di segnali con frequenze 0 – 20 KHz è basato su tecniche di sovracampionamento ed utilizza un SDM del secondo ordine. Il segnale analogico viene trasformato prima in una corrente di bit a una frequenza di 3 MHz e poi, con un processo di decimazione, in un segnale multibit a una frequenza di 48 KHz. Determinare, in bit, la risoluzione del convertitore.

Se il tasso di Nyquist è di 48 KHz, campionando a 3 MHz si ha un tasso di sovracampionamento pari a $\frac{3 \times 10^6}{48 \times 10^3} = 62.5$. Il miglioramento in SQNR offerto da un SDM di ordine 2 è pari a $50 \log_{10} 62.5 - 12 \approx 76$ dB. Un ADC con risoluzione di m bit, lavorando al tasso di Nyquist ha un SQNR pari a $6m + 1.7$; ipotizzando che il miglioramento in SQNR sia dovuto essenzialmente al modulatore, la risoluzione m è ottenuta risolvendo l'equazione $6m + 1.7 = 76$, ciò che comporta $m = 12$ bit.

8.4.4 Sintesi di Reti

Abbiamo visto che ad ogni rete R è associata un'unica funzione $H_R(z)$ razionale in z^{-1} : ogni rete realizza dunque un filtro digitale, generalmente IIR.

Lo stesso filtro può essere realizzato tuttavia con reti diverse. Ad esempio la rete in Figura 8.17 e la rete in Figura 8.27 realizzano lo stesso filtro: la rete specifica dunque non solo il filtro, ma anche il particolare hardware usato per la sua realizzazione.

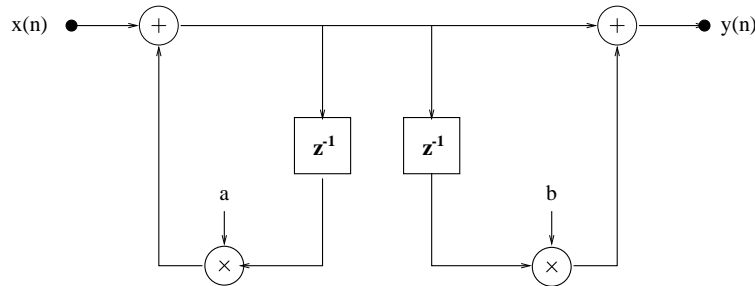


Figura 8.27 Rete che realizza il filtro IIR specificato dall'equazione $y(n) = ay(n-1) + x(n) + bx(n-1)$.

Affrontiamo ora il problema di costruire una rete che realizzi un filtro digitale specificato o dall'equazione alle differenze o, equivalentemente, dalla sua funzione di trasferimento.

Una prima soluzione è data dalle cosiddette *forme dirette I e II*, che estendono al caso generale le reti presentate in Figura 8.17 e in Figura 8.27, introdotte e discusse alla fine del Capitolo 1.

Altre tecniche per la costruzione di reti che realizzano filtri digitali sono basate su regole composizionali presentate in Sezione 8.4.2. A scopo esemplificativo, mostriamo come si possa dare una decomposizione in cascata per filtri FIR e una in parallelo per filtri IIR.

La base matematica su cui poggia la decomposizione in cascata di un filtro FIR è data dal seguente:

Fatto 8.4. *Un qualsiasi polinomio a coefficienti reali può essere decomposto come prodotto di polinomi di primo e di secondo grado a coefficienti reali.*

Dimostrazione. Sia $p(z) = \sum_{k=0}^{L-1} a_k z^k$ un polinomio a variabile complessa con coefficienti a_k reali. Per il teorema fondamentale dell'algebra sappiamo che:

$$p(z) = A \prod_{k=0}^{L-1} (z - z_k),$$

dove z_1, \dots, z_{L-1} sono soluzioni (non necessariamente reali) dell'equazione $p(z) = 0$ e A una costante. Fissato z_k , si hanno due casi:

1. z_k è reale; allora $z - z_k$ è un polinomio di primo grado a coefficienti reali che appare nella decomposizione;
2. z_k è complesso; allora il coniugato z_k^* di z_k è a sua volta una soluzione dell'equazione $p(z) = 0$ poiché

$$0 = p(z_k) = p(z_k)^* = \left(\sum_{j=0}^{L-1} a_j z_k^j \right)^* = \sum_{j=0}^{L-1} a_j^* (z_k^*)^j = \sum_{j=0}^{L-1} a_j (z_k^*)^j = p(z_k^*).$$

Allora $(z - z_k)(z - z_k^*) = z^2 - 2\operatorname{Re}\{z_k\}z + |z_k|^2$ è un polinomio di secondo grado a coefficienti reali che appare nella decomposizione.

□

Dato allora un filtro FIR, caratterizzato da una funzione di trasferimento $H(z)$ che è un polinomio in z^{-1} , per il risultato precedente si può scrivere che:

$$H(z) = AH_1(z) \cdots H_m(z),$$

dove $H_k(z)$ è un polinomio in z^{-1} di primo o di secondo grado a coefficienti reali ($1 \leq k \leq m$). Una rete R che realizza il filtro FIR può essere ottenuta dalla composizione in cascata di reti R_1, \dots, R_m , dove R_k è la rete in forma diretta I o in forma diretta II che realizza il filtro con funzione di trasferimento $H_k(z)$.

Un filtro IIR è invece caratterizzato da una funzione di trasferimento $H(z)$ razionale in z^{-1} . Nell'ipotesi che il grado del numeratore sia minore del grado del denominatore di $H(z)$, possiamo decomporre $H(z)$ in frazioni parziali $H_1(z), \dots, H_m(z)$, con numeratori e denominatori a coefficienti reali e denominatori di grado al più 2:

$$H(z) = AH_1(z) + \cdots + H_m(z).$$

Una rete R che realizza il filtro IIR può essere allora ottenuta dalla composizione in parallelo di reti R_1, \dots, R_m , dove R_k è la rete in forma diretta I o in forma diretta II che realizza il filtro con funzione di trasferimento $H_k(z)$.

8.4.5 Rumore nel Disegno di Filtri Digitali

Il trattamento di segnali analogici mediante sistemi digitali richiede di approssimare numeri reali con numeri che siano rappresentabili con un numero finito di bit: questo fatto forza l'introduzione di errori ineliminabili con cui bisogna imparare a convivere. Presentiamo qui una breve rassegna sui diversi tipi di errori che si vengono a creare nella realizzazione di filtri mediante sistemi digitali; si analizzeranno in particolare gli errori dovuti alla quantizzazione dei coefficienti che entrano nella specifica del filtro.

- *Rumore per quantizzazione del segnale.*

Abbiamo visto che tutti i convertitori analogico-digitale modificano il segnale di ingresso, introducendo quindi un errore di quantizzazione. Nei convertitori analogico-digitale basati su campionamento alla frequenza di Nyquist e quantizzatore a n bit, tale errore può creare un rumore ineliminabile detto *rumore granulare del quantizzatore*: esso può essere misurato in dB di rapporto segnale-rumore, e tale valore è proporzionale al numero n di bit del segnale in uscita ($\approx 6n$). Per convertitori basati su sovra-campionamento e delta-modulazione, il rumore introdotto è legato alle caratteristiche del modulatore e dei filtri che realizzano il convertitore.

- *Rumore per quantizzazione dei coefficienti.*

Un filtro digitale FIR o IIR viene specificato attraverso l'algoritmo (o attraverso la rete) che lo realizza: le operazioni di moltiplicazione per costante richiedono a loro volta l'assegnazione di opportuni coefficienti reali, visti come parametri. L'implementazione dell'algoritmo su un processore con parola di lunghezza fissa (tipicamente 16 o 32 bit), pone dei limiti all'accuratezza con cui possono essere specificati i parametri: viene quindi introdotto un errore, detto *errore di quantizzazione dei coefficienti*, per cui il filtro implementato non coincide in generale con quello specificato. Discuteremo in seguito in maggior dettaglio gli effetti di tale tipo di errore; vogliamo qui ricordare brevemente:

1. Gli effetti della quantizzazione dei coefficienti dipendono dall'aritmetica di macchina del processore su cui il filtro è implementato: la rappresentazione in virgola fissa è generalmente più sensibile a tali errori, a causa della propagazione dell'errore nella moltiplicazione. Su DSP (Digital Signal Processor) a 32 bit o più, con rappresentazione in virgola mobile, questo tipo di errore può invece essere trascurato.
 2. I filtri IIR sono più sensibili dei filtri FIR all'errore di quantizzazione dei coefficienti, a causa della struttura intrinsecamente ricorsiva dei filtri IIR.
 3. Lo stesso filtro può essere realizzato con diverse architetture di rete: la sensibilità all'errore di quantizzazione dei coefficienti è fortemente dipendente dal tipo di rete realizzata. Ad esempio, la sensibilità all'errore di reti che realizzano un filtro in forma diretta è generalmente più alta rispetto alle reti che realizzano lo stesso filtro in forma di cascata o parallelo.
- *Rumore per troncamento.*

L'implementazione di un algoritmo che realizza un filtro digitale richiede l'esecuzione di varie operazioni di somma e prodotto. Ipotizziamo di utilizzare una rappresentazione in virgola fissa. Anche se l'ingresso, l'uscita e i coefficienti del filtro sono numeri rappresentabili con n bit, mantenere nei calcoli questa accuratezza richiede una precisione maggiore poiché, tipicamente, la moltiplicazione di due numeri di n bit produce un numero rappresentabile con $2n$ bit. La necessità di arrotondare i risultati intermedi produce quindi un errore detto *rumore di troncamento*.

Una delle tecniche per controllare l'errore di troncamento è quella di spostare l'operazione di arrotondamento il più possibile nella parte finale del calcolo. A tal riguardo, risulta molto utile implementare un filtro digitale su processori che, lavorando con parole di n bit, hanno registri come l'accumulatore (ACC) o il prodotto (P) di dimensione doppia, cioè $2n$ bit. Ad esempio, un filtro FIR può essere implementato usando la sottoprocedura:

```

Subroutine FILTRO
    ⋮
    P ←  $a_{k-1}x(n - (k - 1))$ 
    ACC ← ACC + P
    P ←  $a_kx(n - k)$ 
    ACC ← ACC + P
    ⋮
fine

```

La precisione a $2n$ bit viene mantenuta durante l'esecuzione della Subroutine, che agisce sui registri ACC e P di $2n$ bit, e solo alla fine il risultato viene troncato per essere memorizzato in n bit. L'errore di troncamento è allora limitato al bit meno significativo del numero memorizzato: il rapporto segnale-rumore di troncamento si mantiene quindi intorno agli n dB.

Analizziamo ora più in dettaglio l'effetto prodotto su un filtro digitale dall'errore di quantizzazione dei coefficienti. Abbiamo visto che un filtro è caratterizzato dai poli e dagli zeri della sua funzione di trasferimento $H(z^{-1})$: la modifica dei coefficienti che specificano

il filtro, causata dall'errore di quantizzazione, provoca a sua volta un cambiamento nella posizione dei poli e degli zeri, che sono responsabili del comportamento del filtro.

La situazione è particolarmente delicata per i filtri IIR: i poli di un filtro IIR, correttamente progettato, possono eventualmente spostarsi al di fuori del cerchio unitario, a causa dell'errore di quantizzazione dei coefficienti, rendendo il filtro instabile.

Studiamo ora in maniera quantitativa questo fenomeno, valutando la sensitività dei poli rispetto al cambiamento dei coefficienti. A questo riguardo, sia $D(z^{-1})$ il denominatore della funzione di trasferimento razionale $H(z^{-1})$ di un filtro IIR: esprimendo $D(z^{-1})$ in forma di polinomio e in forma fattorizzata, si ha:

$$D(z^{-1}) = 1 - \sum_{k=1}^M a_k z^{-k} = \prod_{k=1}^M (1 - z_k z^{-1}),$$

dove a_1, \dots, a_M individuano i coefficienti del polinomio e z_1, \dots, z_M ne sono gli zeri, individuando di conseguenza i poli di $H(z^{-1})$.

Supponiamo ora che la quantizzazione dei coefficienti a_1, \dots, a_M porti ad un nuovo filtro, caratterizzato da nuovi coefficienti $\hat{a}_1, \dots, \hat{a}_M$ tali che:

$$\hat{a}_k = a_k + \Delta a_k \quad (1 \leq k \leq M),$$

dove Δa_k è l'errore di quantizzazione. Il nuovo polinomio avrà nuovi zeri $\hat{z}_1, \dots, \hat{z}_M$, che risulteranno i poli della funzione di trasferimento del nuovo filtro.

Posto $\Delta z_k = \hat{z}_k - z_k$, Δz_k può essere interpretato come errore di localizzazione del polo z_k ; si può facilmente derivare la relazione:

$$\Delta z_k \approx \frac{\sum_{j=1}^M z_k^{M-j} \Delta a_j}{\prod_{j \neq k} (z_k - z_j)}, \quad (1 \leq k \leq M).$$

Questa formula esprime la sensitività dei poli rispetto agli errori di quantizzazione $\Delta a_1, \dots, \Delta a_M$.

Osservando che l'errore di localizzazione Δz_k del polo z_k è tanto più elevato quanto più $\prod_{j \neq k} (z_k - z_j)$ è vicino a 0, concludiamo:

- a parità di ordine M , i filtri i cui poli si raggruppano in poche classi di piccole dimensioni sono i più sensibili agli errori di quantizzazione dei coefficienti;
- filtri stabili di ordine M contengono M poli nel cerchio unitario: "grandi" valori di M forzeranno alcuni poli ad essere necessariamente "vicini". I filtri di ordine elevato risulteranno allora generalmente più sensibili all'errore di quantizzazione dei coefficienti che non i filtri di ordine basso;
- le reti in forma di cascata o in parallelo realizzano separatamente ogni coppia di poli complessi coniugati. Questo è il motivo per cui le forme in cascata o parallelo sono meno sensibili all'errore di quantizzazione dei coefficienti che non le forme dirette.