

01 Brevi cenni di Elaborazione dei Segnali

Si rivedono brevemente i concetti di Teoria dei Segnali e di Elaborazione dei Segnali che verranno utilizzati nel corso. Si rimanda ai relativi corsi per una trattazione approfondita di questi argomenti.

01.01 I segnali

Il segnale è una funzione di variabili indipendenti, come il tempo, la distanza, la posizione. Nel corso considereremo in particolare il segnale audio, che rappresenta la pressione dell'aria in funzione del tempo in un certo punto dello spazio.

I segnali possono essere monodimensionali o multidimensionali, a seconda che siano funzioni di una o più variabili indipendenti, e possono essere monocanali o multicanali, a seconda che siano funzioni scalari o vettoriali. Il segnale audio telefonico è un segnale monocanale, il segnale stereo è invece un segnale a due canali (Left e Right).

Se la variabile indipendente è continua ($t \in \mathbb{R}$) il segnale è detto *a tempo continuo*. Se la variabile indipendente è discreta ($t \in \mathbb{N}$ o $t \in \mathbb{Z}$) il segnale è detto *a tempo discreto*, e in questo caso il segnale è costituito da una sequenza di numeri.

I segnali possono essere ad *ampiezza continua* o ad *ampiezza discreta*. Un segnale ad ampiezza continua può assumere una infinità di valori (in \mathbb{R} o \mathbb{C}). Al contrario i segnali ad ampiezza discreta possono assumere solo un numero finito di valori, che possono essere codificati con un numero finito di cifre binarie.

Un segnale a tempo continuo e ampiezza continua è detto un *segnale analogico*. Un segnale a tempo discreto e ampiezza discreta è detto un *segnale digitale* o un *segnale numerico*. La voce, la musica, i segnali audio in genere, nascono come segnali analogici. Mediante l'operazione di campionamento, quantizzazione e codifica dei campioni, possiamo trasformarli in segnali digitali e come tali elaborarli mediante computer, circuiti digitali, DSP (processori per l'elaborazione del segnale), ...

Nel seguito, indicheremo i segnali analogici con $x_a(t)$ (o semplicemente $x(t)$), funzioni del tempo continuo $t \in \mathbb{R}$, e i segnali digitali con $x(n)$, funzioni dell'indice temporale $n \in \mathbb{Z}$.

01.02 La trasformata di Fourier a tempo continuo

I segnali possono essere rappresentati e trattati nel dominio del tempo, ma possiamo anche rappresentarli e trattarli in un dominio trasformato, il dominio della frequenza.

Lo spettro di un segnale a tempo continuo è dato dalla *CTFT – Continuous Time Fourier Transform*, *Trasformata di Fourier di un segnale a tempo continuo*, o spettro di Fourier,

$$X_a(j\Omega) = \int_{-\infty}^{+\infty} x_a(t) e^{-j\Omega t} dt.$$

Lo spettro $X_a(j\Omega)$ è una funzione complessa della frequenza angolare a tempo continuo $\Omega \in \mathbb{R}$ che si misura in rad/s

La CTFT se esiste è unica ed il segnale originale può essere ricostruito a partire da $X_a(j\Omega)$ mediante trasformazione inversa, ICTFT,

$$x_a(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} X_a(j\Omega) e^{j\Omega t} d\Omega.$$

Possiamo anche esprimere la CTFT in forma polare:

$$X_a(j\Omega) = |X_a(j\Omega)| \cdot e^{j\theta_a(\Omega)},$$

dove $\theta_a(\Omega) = \arg \{X_a(j\Omega)\}$.

$|X_a(j\Omega)|$ è detto *spettro di ampiezza* (magnitude spectrum), e $\theta_a(\Omega)$ è detto *spettro di fase* (phase spectrum). $|X_a(j\Omega)|$ e $\theta_a(\Omega)$ sono funzioni reali della frequenza angolare Ω .

I segnali reali, $x_a(t) \in \mathbb{R}$ hanno spettro a simmetria coniugata,

$$X_a(j\Omega) = X_a^*(-j\Omega),$$

spettro di ampiezza simmetrico, $|X_a(j\Omega)| = |X_a(-j\Omega)|$, spettro di fase antisimmetrico, $\theta_a(\Omega) = -\theta_a(-\Omega)$.

In generale un segnale a tempo continuo $x_a(t)$ ha uno spettro che occupa l'intero range delle frequenze da $-\infty$ a $+\infty$. Un segnale viene detto a *banda limitata* se il suo spettro occupa solo una porzione del range delle frequenze $-\infty < \Omega < +\infty$.

In particolare considereremo segnali a banda limitata passabasso, per i quali lo spettro occupa il range delle frequenze $0 \leq |\Omega| \leq \Omega_p$, dove Ω_p è detta la *larghezza di banda*.

Non è possibile generare un segnale a banda limitata ideale. I segnali reali hanno sempre durata finita e quindi spettro infinito. Nella maggior parte delle applicazioni ci basta garantire che il segnale abbia energia sufficientemente piccola (trascurabile) al di fuori dell'intervallo di interesse $[-\Omega_p, \Omega_p]$.

01.03 La trasformata di Fourier a tempo discreto

La rappresentazione nel dominio della frequenza di un segnale a tempo discreto è data dalla *trasformata di Fourier a tempo discreto* (DTFT – Discrete Time Fourier Transform),

$$X(e^{j\omega}) = \sum_{n=-\infty}^{+\infty} x(n) e^{-j\omega n}.$$

che è una funzione della frequenza angolare normalizzata ω , con $\omega \in \mathbb{R}$.

Se esiste, la DTFT di una sequenza è unica e la sequenza originale può essere ricostruita dalla sua trasformata mediante una trasformazione inversa, IDTFT,

$$x(n) = \frac{1}{2\pi} \int_{-\pi}^{+\pi} X(e^{j\omega}) e^{j\omega n} d\omega.$$

La DTFT di una sequenza è una funzione periodica in ω di periodo 2π ,

$$X(e^{j\omega}) = X(e^{j(\omega+2\pi k)}).$$

In generale, la DTFT di una sequenza è una funzione complessa della variabile reale ω . La DTFT può essere espressa in forma polare,

$$X(e^{j\omega}) = |X(e^{j\omega})| e^{j\theta(\omega)}$$

dove $|X(e^{j\omega})|$ è detto *spettro di ampiezza* (magnitude spectrum), e $\theta(\omega)$ è detto *spettro di fase*. Si noti che c'è una indeterminazione di $2\pi k$, con $k \in \mathbb{Z}$, nella conoscenza di $\theta(\omega)$. Infatti, $\theta(\omega)$ è un angolo misurato in radianti.

Se $x(n) \in \mathbb{R}$, la DTFT $X(e^{j\omega})$ ha simmetria coniugata,

$$X(e^{-j\omega}) = X^*(e^{j\omega}).$$

01.04 Teorema del campionamento

La maggior parte dei segnali che incontriamo nel mondo reale sono dei segnali a tempo continuo (la voce, la musica, i segnali audio in genere). Gli algoritmi di elaborazione numerica del segnale vengono spesso usati per trattare i segnali a tempo continuo. Per farlo i segnali vengono dapprima campionati e codificati con un convertitore A/D (analogico/digitale), vengono elaborati per via numerica e poi riconvertiti in forma analogica con una convertitore D/A (digitale/analogico).

Vediamo quale relazione esiste tra i segnali analogici e i relativi segnali campionati. Sia $x_a(t)$ un segnale analogico e $x(n)$ il segnale digitale ottenuto dal campionamento di $x_a(t)$ con periodo T , ovvero frequenza di campionamento $F_c = \frac{1}{T}$,

$$x(n) = x_a(t) \Big|_{t=nT} = x_a(nT).$$

Se $X(e^{j\omega})$ è lo spettro del segnale campionato e $X_a(j\Omega)$ lo spettro del segnale analogico, si dimostra che

$$X(e^{j\omega}) = \frac{1}{T} \sum_{k=-\infty}^{+\infty} X_a \left[j \left(\frac{\omega + 2\pi k}{T} \right) \right].$$

Lo spettro del segnale campionato è dato dalla ripetizione periodica, con periodo 2π , dello spettro continuo $X_a \left(j \frac{\omega}{T} \right)$.

$X_a \left(j \frac{\omega}{T} \right)$ è identico a $X_a(j\Omega)$, solo che l'asse delle frequenze è stato normalizzato in modo tale che $\omega = 2\pi$ corrisponde alla frequenza angolare di campionamento $\Omega_c = \frac{2\pi}{T}$.

Lo spettro del segnale campionato rappresenterà fedelmente lo spettro del segnale analogico solo se le varie ripetizioni dello spettro non si sovrappongono, ovvero solo se $x_a(t)$ è un segnale a banda limitata Ω_m e lo andiamo a campionare con una frequenza di campionamento sufficientemente elevata, $\Omega_c \geq 2\Omega_m$. Se ci sono delle sovrapposizioni tra gli spettri ripetuti, avremo un errore (una distorsione), detto *errore di aliasing*, che è irrecuperabile.

Il Teorema del Campionamento
(o teorema di Nyquist-Shannon)

Sia $x_a(t)$ un segnale a banda limitata, con $X_a(j\Omega) = 0$ per $|\Omega| > \Omega_m$. $x_a(t)$ può essere fedelmente ricostruito dai suoi campioni $x_a(nT)$, $-\infty < n < +\infty$, se la frequenza angolare di campionamento

$$\Omega_c \geq 2\Omega_m$$

dove $\Omega_c = \frac{2\pi}{T}$.

In definitiva, dato un segnale analogico, per convertirlo in un segnale digitale dobbiamo dapprima filtrarlo con un filtro passa-basso per ottenere un segnale a banda limitata, e solo successivamente lo possiamo campionare con una frequenza almeno pari a due volte la larghezza di banda. I campioni dovranno essere quantizzati e codificati.

Il segnale originale $x_a(t)$ può essere recuperato costruendo un treno di impulsi, equispaziati di T , proporzionali ai campioni $x(n)$,

$$x_p(t) = \sum_{n=-\infty}^{+\infty} x(n)\delta(t - nT),$$

dove $\delta(t)$ è l'impulso di Dirac. Il segnale a tempo-continuo così costituito ha spettro dato dalla ripetizione periodica di $X_a(j\Omega)$ con periodo Ω_c ,

$$X_p(j\Omega) = \frac{1}{T} \sum_{k=-\infty}^{+\infty} X_a [j(\Omega + k\Omega_c)].$$

Filtrando $x_p(t)$ con un filtro passa-basso avente larghezza di banda Ω_m , detto anche filtro anti-imaging, si ottiene il segnale originale.

Nella realtà, i convertitori D/A sostituiscono gli impulsi di Dirac di $x_p(t)$ con impulsi rettangolari, ma rimane la necessità di filtrare il segnale con un filtro anti-imaging.

01.05 La Trasformata Z

La trasformata Z è una generalizzazione della trasformata di Fourier a tempo discreto. La trasformata Z esiste per molte sequenze per le quali la DTFT non esiste. È una funzione della variabile complessa z .

Data una sequenza $x(n)$, la sua trasformata Z è definita da

$$X(z) = \sum_{n=-\infty}^{+\infty} x(n)z^{-n},$$

dove $z \in \mathbb{C}$ è una variabile complessa continua.

Se confrontiamo la definizione di $X(z)$ con la DTFT di $x(n)$:

$$X(e^{j\omega}) = \sum_{n=-\infty}^{+\infty} x(n)e^{-j\omega n}$$

vediamo che

$$X(e^{j\omega}) = X(z) \Big|_{z=e^{j\omega}},$$

purché entrambe queste trasformate esistano. La trasformata Z, valutata sul circolo di raggio unitario coincide con la DTFT.

Il luogo dei punti del piano z per i quali la trasformata Z converge è detto *Regione di Convergenza*.

La trasformata Z non è unica: sequenze diverse ammettono la stessa trasformata, ma in quel caso hanno diversa regione di convergenza. Ovvero, esiste una corrispondenza biunivoca tra le sequenze e le coppie {Trasformata Z, Regione di Convergenza}.

Le sequenze causali, per le quali $x(n) = 0 \forall n < 0$, hanno regione di convergenza esterna a un cerchio centrato nell'origine. Per queste sequenze, esiste una corrispondenza biunivoca tra sequenze e trasformata Z.

Detto \mathcal{C} un contorno chiuso nella regione di convergenza che contenga l'origine del piano complesso (per esempio, \mathcal{C} potrebbe essere un circolo centrato nell'origine del piano complesso e contenuto nella regione di convergenza), l'espressione generale della *Trasformata Z inversa* è data da:

$$g(n) = \frac{1}{2\pi j} \oint_{\mathcal{C}} G(z) z^{n-1} dz,$$

dove $\oint_{\mathcal{C}}$ indica l'integrazione eseguita sul contorno chiuso \mathcal{C} .

La formula è difficile da valutare. Fortunatamente, in genere si considerano solo trasformate Z razionali e sequenze causali. In queste condizioni, esistono dei semplici metodi (e.g., metodo tabellare, metodo dei fratti semplici, metodo della long-division) che consentono il calcolo dell'anti-trasformata Z.

01.06 Trasformata di Fourier discreta

La DFT, Discrete Fourier Transform o Trasformata di Fourier discreta, si applica a sequenze periodiche di periodo N o a sequenze di durata finita N (che possiamo sempre rendere periodiche di periodo N), $x_p(n) = x_p(n + N)$ per tutti gli n . La DFT di $x_p(n)$ è la sequenza:

$$X_p(k) = \sum_{n=0}^{N-1} x_p(n) e^{-j \frac{2\pi}{N} nk}.$$

È facile verificare che $X_p(k)$ è una sequenza complessa periodica di periodo N :

$$X_p(k) = X_p(k + N) \quad \forall k.$$

La Trasformata Discreta di Fourier Inversa (IDFT) è data da:

$$x_p(n) = \frac{1}{N} \sum_{k=0}^{N-1} X_p(k) e^{j \frac{2\pi}{N} nk}.$$

e richiede sole somme e moltiplicazioni complesse ed è quindi facilmente implementabile su calcolatore. Esistono anche degli algoritmi di calcolo veloce della DFT, le FFT - Fast Fourier Transform, particolarmente efficienti quando N è una potenza di 2. Gli algoritmi originali di Cooley e Tukey richiedono un costo computazionale di $\frac{N}{2} \log_2(N)$ moltiplicazioni e $N \log_2(N)$ somme complesse.

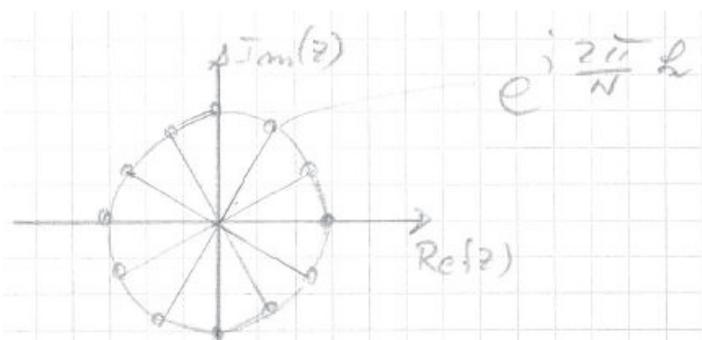
Nel caso di sequenze di durata finita N (ovvero definite diverse da 0 solo nell'intervallo $[0, N - 1]$), la DFT, la DTFT, la trasformata Z, sono equivalenti tra di loro e portano la stessa informazione. Nota una trasformata, esistono formule che consentono di calcolare le altre due trasformate. Confrontando la DFT con la DTFT di una sequenza di durata finita N ,

$$X(e^{j\omega}) = \sum_{n=0}^{N-1} x(n)e^{-j\omega n},$$

è facile verificare che la DFT $X(k)$ coincide con il campionamento di $X(e^{j\omega})$ su N punti equidistanti nell'intervallo $[0, 2\pi]$, i punti $\omega_k = \frac{2\pi}{N}k$. Se consideriamo la trasformata Z,

$$X(z) = \sum_{n=0}^{N-1} x(n)z^{-n},$$

si può verificare in modo analogo che la DFT $X(k)$ coincide con il campionamento della $X(z)$ su N punti equidistanti sul circolo di raggio unitario, i punti $e^{j\frac{2\pi}{N}k}$.



La DFT è il modo più "spiccio" per valutare il contenuto spettrale di una sequenza. Possiamo valutare lo spettro $X(e^{j\omega})$ con il livello di dettaglio da noi desiderato. Anche quando N sia piccolo, possiamo calcolare la DFT su L punti, con $L > N$, semplicemente aggiungendo alla sequenza $L - N$ zeri prima di calcolarne la FFT, ovvero concettualmente rendendola periodica di un periodo $L > N$.

Nel caso di sequenze reali, la DFT è ancora a simmetria coniugata,

$$X(k) = X^*(-k),$$

e in virtù della periodicità di periodo N della DFT (valutata su N punti),

$$X(k) = X^*(N - k).$$

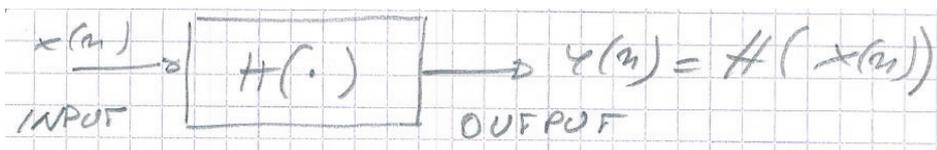
Nel corso, valuteremo lo spettro dei nostri segnali utilizzando gli algoritmi FFT per il calcolo veloce della DFT. Passeremo in ingresso alle routine di calcolo FFT i campioni $x(n)$ per $n = 0, \dots, N - 1$ e otterremo in uscita i campioni $X(k)$ per $k = 0, \dots, N - 1$. Nel caso di segnali reali, è proprio la formula $X(k) = X^*(N - k)$ che consente di trovare i termini coniugati.

01.07 Sistemi e filtri a tempo discreto

Un sistema è un dispositivo, un circuito, un algoritmo implementato su calcolatore o su di un altro processore, che a uno o più segnali di ingresso associa uno o più segnali di uscita.

La funzione di un sistema a tempo discreto è quella di elaborare una o più sequenze, dette *sequenze di ingresso*, per generare una o più sequenze, dette *sequenze di uscita*, che abbiano le proprietà desiderate o che evidenzino certa informazione del segnale o dei segnali di ingresso.

Nella maggior parte dei casi i sistemi hanno un ingresso e una uscita.



L'ingresso viene spesso indicato con $x(n)$ e l'uscita con $y(n)$. Matematicamente, il sistema a tempo discreto è caratterizzato da un operatore $H(\cdot)$ che associa a ciascuna sequenza di ingresso $x(n)$ una sequenza di uscita $y(n)$.

Nei sistemi a tempo discreto di uso pratico tutti i segnali sono segnali digitali (a tempo discreto e ampiezza discreta) e le operazioni su tali segnali portano anche a dei segnali digitali. Questi sistemi sono usualmente chiamati "filtri digitali". Useremo quindi indifferentemente i termini "sistema a tempo discreto" o anche "sistema discreto" o "filtro digitale".

L'origine del termine "filtro" deriva dal fatto che la prima applicazione di questi sistemi è stata il filtraggio dello spettro di un segnale, con il sistema progettato per lasciar passare indenni alcune frequenze del segnale e per filtrare, ovvero per eliminare, le frequenze indesiderate. Il termine oggi viene usato con un'accezione del tutto generale anche per indicare sistemi non bloccanti nella frequenza, come ad esempio i filtri equalizzatori.

Un sistema discreto si definisce *lineare* se soddisfa al principio di sovrapposizione degli effetti, ovvero se per qualunque coppia di segnali $x_1(n)$ e $x_2(n)$, cui corrispondono in uscita i segnali $y_1(n)$ e $y_2(n)$, e per qualunque coppia di costanti a_1 e a_2 , al segnale di ingresso $x(n) = a_1x_1(n) + a_2x_2(n)$ corrisponde in uscita $a_1y_1(n) + a_2y_2(n)$.

$$x_1(n) \longrightarrow y_1(n)$$

$$x_2(n) \longrightarrow y_2(n)$$

$$a_1x_1(n) + a_2x_2(n) \longrightarrow a_1y_1(n) + a_2y_2(n)$$

Un sistema è detto *tempo invariante* o *shift invariante* se, per qualunque segnale di ingresso $x(n)$ cui corrisponde in uscita il segnale $y(n)$, e per qualunque costante $k \in \mathbb{Z}$, al segnale $x(n - k)$ corrisponde in uscita $y(n - k)$:

$$x(n) \longrightarrow y(n)$$

$$x(n - k) \longrightarrow y(n - k)$$

Interessano in modo particolare i sistemi a tempo discreto *lineari e tempo invarianti* (LTI), che soddisfano sia la proprietà di linearità che di tempo-invarianza. Questi sistemi sono facili da analizzare e caratterizzare e di conseguenza sono anche facili da progettare. Sono ampiamente utilizzati nei sistemi per l'elaborazione dei segnali.

01.08 Risposta impulsiva e somma di convoluzione

La *risposta impulsiva* di un sistema LTI è per definizione la risposta del sistema a un impulso unitario in ingresso:

$$x(n) = \delta(n) \longrightarrow y(n) = h(n)$$

dove la sequenza *impulso unitario* $\delta(n)$ è definita da

$$\delta(n) = \begin{cases} 1 & n = 0 \\ 0 & n \neq 0 \end{cases}.$$

La risposta impulsiva caratterizza completamente i sistemi LTI. Nota la risposta impulsiva possiamo calcolare l'uscita del nostro sistema per una qualunque sequenza di ingresso. Infatti,

$$y(n) = \sum_{m=-\infty}^{+\infty} x(m)h(n-m) = x(n) \otimes h(n)$$

La somma viene detta *somma di convoluzione* o *sommatoria di convoluzione*. Diciamo anche che $x(n)$ è *convoluto* con $h(n)$.

La somma di convoluzione gode della proprietà commutativa,

$$x(n) \otimes h(n) = h(n) \otimes x(n),$$

associativa,

$$[x(n) \otimes h_1(n)] \otimes h_2(n) = x(n) \otimes [h_1(n) \otimes h_2(n),]$$

e distributiva,

$$x(n) \otimes [h_1(n) + h_2(n)] = x(n) \otimes h_1(n) + x(n) \otimes h_2(n).$$

Sistemi lineari tempo invarianti causali.

Un sistema discreto si definisce *causale* se l'uscita del sistema a ogni istante di tempo n , $y(n)$, dipende solo dai campioni passati e dal campione presente del segnale di ingresso $x(n)$, $(x(n), x(n-1), x(n-2), \dots)$, mentre non dipende dai campioni futuri del segnale $(x(n+1), x(n+2), x(n+3), \dots)$.

I sistemi che non soddisfano a questa condizione sono detti *non causali*.

Nei sistemi di elaborazione in tempo reale non possiamo osservare i valori futuri del segnale e pertanto i sistemi non causali sono irrealizzabili. Per questo motivo la proprietà di causalità viene anche detta di *realizzabilità*.

Proprietà: Un sistema LTI è causale se e solo se la sua risposta impulsiva è zero per $n < 0$:

$$\text{causalità} \iff h(n) = 0 \quad \forall n < 0.$$

In un sistema causale la somma di convoluzione è data da

$$y(n) = \sum_{m=0}^{+\infty} h(m)x(n-m)$$

Stabilità dei sistemi LTI

Un sistema a tempo discreto si definisce stabile in senso *BIBO* (Bounded Input Bounded Output) se a ogni sequenza di ingresso limitata $x(n)$ corrisponde una sequenza di uscita $y(n)$ limitata.

Proprietà: Un sistema LTI è stabile in senso BIBO se e solo se

$$\sum_{n=-\infty}^{+\infty} |h(n)| < +\infty,$$

ovvero, se e solo se la risposta impulsiva è assolutamente sommabile.

01.09 Risposta in frequenza e funzione di trasferimento

La trasformata DTFT e la trasformata Z hanno la proprietà di trasformare la somma di convoluzione nel prodotto delle relative trasformate:

$$\text{DTFT}\{h(n) \otimes x(n)\} = H(e^{j\omega})X(e^{j\omega})$$

dove

$$H(e^{j\omega}) = \sum_{n=-\infty}^{+\infty} h(n)e^{-j\omega n}.$$

La DTFT della risposta impulsiva è detta *Risposta in frequenza* del sistema LTI. $|H(e^{j\omega})|$ è detta *risposta di ampiezza*, $\arg\{H(e^{j\omega})\}$ è detta *risposta di fase*.

In modo analogo,

$$\mathcal{Z}\{h(n) \otimes x(n)\} = H(z)X(z)$$

dove

$$H(z) = \sum_{n=-\infty}^{+\infty} h(n)z^{-n}.$$

la trasformata Z della risposta impulsiva è detta *funzione di trasferimento* del sistema LTI.

Come la risposta impulsiva, la risposta in frequenza e la funzione di trasferimento caratterizzano completamente il sistema LTI.

01.10 Filtri a risposta impulsiva finita (FIR) e infinita (IIR)

Le due categorie di filtri digitali LTI di maggiore uso nelle applicazioni sono i filtri a risposta impulsiva finita FIR (Finite Impulse Response) e a risposta impulsiva infinita (IIR).

Un sistema LTI a tempo-discreto è detto a *risposta impulsiva finita* se la sua risposta impulsiva ha durata finita. Nel caso di un sistema FIR causale di lunghezza N :

$$h(n) = 0 \quad \forall n < 0 \text{ or } n \geq N,$$

e

$$y(n) = \sum_{m=0}^{N-1} h(m)x(n-m).$$

Il filtro FIR può essere direttamente implementato mediante la sua somma di convoluzione, ed è questa una delle implementazioni più comuni. Il filtro è sempre stabile in senso BIBO. Inoltre, i filtri FIR possono avere risposta di fase lineare (e in tal caso non introducono distorsione di fase). Nei filtri FIR, la funzione di trasferimento è un polinomio in z^{-1} ,

$$H(z) = \sum_{n=0}^{N-1} h(n)z^{-n}$$

per questo i filtri vengono anche detti "a soli zeri".

Quando un sistema LTI ha risposta impulsiva di durata infinita, viene detto *a risposta impulsiva infinita* o IIR. In questo caso, non possiamo implementare il filtro mediante la somma di convoluzione (richiederebbe infinite somme o moltiplicazioni). Nella pratica, si considera una sottoclasse dei filtri LTI, causali, a risposta impulsiva infinita, ovvero la classe di filtri rappresentabili mediante una equazione alle differenze finite, per i quali la relazione ingresso uscita è

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

per tutti gli $n \geq 0$.

Per questa classe di sistemi l'uscita può essere calcolata direttamente da alcuni ultimi campioni del segnale di ingresso e di uscita. Il sistema è causale perché coinvolge solo campioni passati e il campione attuale del segnale di ingresso.

Affiché sia possibile calcolare $y(n)$ dal segnale di ingresso $x(n)$, bisogna conoscere le N condizioni iniziali, $y(-1), y(-2), \dots, y(-N)$. Se le N condizioni iniziali sono tutte zero,

$$y(-1) = y(-2) = \dots = y(-N) = 0,$$

il sistema viene detto *inizialmente a riposo*.

La funzione di trasferimento dei filtri IIR descritti da una equazione alle differenze finite è razionale:

$$H(z) = \frac{\sum_{i=0}^M b_i z^{-i}}{1 + \sum_{j=1}^N a_j z^{-j}}$$

Le radici (in z) del polinomio a numeratore vengono dette *zeri* del sistema LTI. Le radici del polinomio a denominatore vengono dette *poli* del sistema LTI.

Nei filtri IIR, la stabilità in senso BIBO non è garantita a priori, va assicurata in sede di progetto. I filtri IIR non possono avere risposta di fase lineare (introducono sempre una distorsione di fase). Ciononostante, si adattano meglio dei filtri FIR a realizzare le specifiche dei filtri passa-basso, passa-alto, passa-banda, ovvero le realizzano con una minore complessità computazionale. Infatti, i filtri IIR hanno zeri e poli, mentre i filtri FIR hanno solo zeri.