Appunti di
Elaborazione numerica dei segnali
Capitolo 3 (parte I) - **Trasformata Z**

INTRODUZIONE

Data una sequenza x(n) tempo-discreta[1], sappiamo che la sua *trasformata di Fourier* è definita come

$$X(f) = \sum x(n) e^{-j2\pi f n T}$$

dove la sommatoria è estesa al numero di campioni non nulli della sequenza.

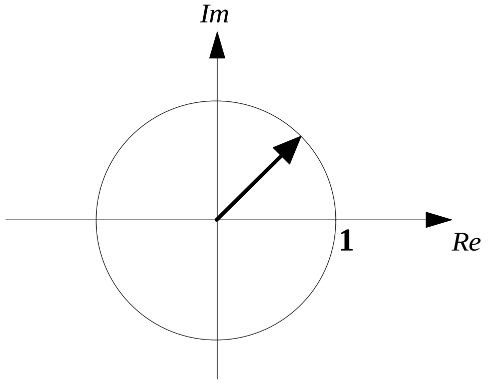
Possiamo porre $z = e^{j2\pi fT}$: con questa posizione, la trasformata assume l'espressione

$$X(z) = \sum x(n) z^{-n}$$

[1] Per semplicità, trascuriamo di indicare il periodo di campionamento T

In questo modo, X(z) assume l'espressione classica di un polinomio nella variabile z^{-1}: i coefficienti del polinomio sono i campioni della sequenza x(n).

Avendo posto $z = e^{j2\pi fT}$, è ovvio che la variabile z assuma solo dei precisi valori al variare della frequenza f: si tratta, in particolare, di quantità complesse di modulo unitario e fase $2\pi fT$, ossia quindi, con riferimento al piano complesso, vettori nell'origine disposti con il vertice libero di muoversi sul cerchio di raggio unitario:



Se però consideriamo z come una variabile complessa assolutamente generica, senza cioè l'imposizione $z = e^{j2\pi fT}$, allora otteniamo qualcosa di simile a ciò che si ottiene, nel dominio continuo, nel passaggio dalla trasformata di Fourier alla trasformata di Laplace: infatti, dato un segnale di partenza f(t), mentre la sua trasformata di Fourier è

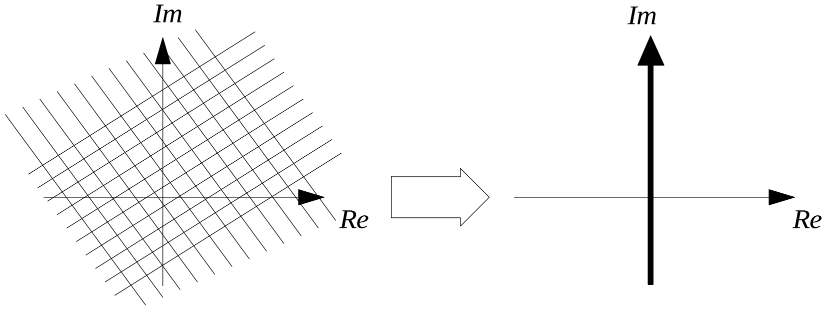
$$\text{Fourier}[f(t)] = \int_{-\infty}^{+\infty} f(t) e^{-j2\pi ft} dt$$

la sua trasformata di Laplace è

$$\text{Laplace}[f(t)] = \int_{-\infty}^{+\infty} f(t) e^{-pt} dt \qquad \text{dove } p = \sigma + j\omega = \sigma + j2\pi f$$

Nella trasformata di Fourier si usa la variabile puramente immaginaria $j\omega$ (dove $\omega=2\pi f$), mentre nella trasformata di Laplace si usa la variabile complessa $p=\sigma+j\omega$. Detto in altre parole, *la trasformata di Fourier è una particolarizzazione della trasformata di Laplace: quest'ultima è definita nel piano complesso*[2]*, mentre la trasformata di Fourier solo sull'asse immaginario* (che è un sottoinsieme del piano complesso):

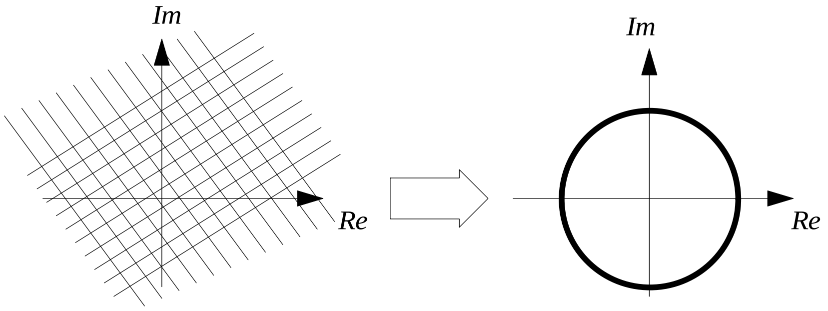
[2] In realtà, sappiamo che la *regione di convergenza* della trasformata di Laplace di un segnale non necessariamente coincide con tutto il piano complesso.



trasformata di Laplace **trasformata di Fourier**

La differenza fondamentale, che deriva da queste considerazioni, è che la trasformata di Fourier aiuta, nello studio dei sistemi lineari tempo-invarianti, solo in condizioni di regime, mentre la trasformata di Laplace, che pure va bene per le condizioni di regime, dà anche informazioni estremamente complete circa il transitorio iniziale (ossia, sostanzialmente, la dipendenza della risposta del sistema alle condizioni iniziali).

Discorso assolutamente analogo vale anche nel dominio discreto: anziché supporre $z = e^{j2\pi fT}$, si lascia che z vari in tutto il piano complesso, il che equivale sostanzialmente a passare dalla DTFT della sequenza x(n) di partenza alla cosiddetta **trasformata zeta** della stessa sequenza. Mentre la DTFT fa riferimento solo al *cerchio unitario*, la trasformata zeta fa riferimento a tutto il piano complesso.



trasformata Zeta **trasformata di Fourier discreta (DTFT)**

E' ovvio, quindi, come del resto sarà chiaro nei prossimi discorsi, che potremo sempre passare dalla trasformata zeta di una sequenza alla trasformata di Fourier (DTFT) della stessa sequenza semplicemente sostituendo $z = e^{j2\pi fT}$.

DEFINIZIONE DI TRASFORMATA ZETA

Abbiamo dunque capito che, data una sequenza x(n) di lunghezza (per il momento) infinita, la sua **trasformata zeta** è definita nel modo seguente:

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

dove z è una variabile complessa: z=σ+jω. Si tratta, quindi, in pratica di *costruire un polinomio in z^{-1} usando come coefficienti i campioni di x(n).* Non solo, ma è evidente che X(z) costituisce, di fatto, una rappresentazione alternativa del segnale considerato:

infatti, considerando il coefficiente del generico termine z^{-n}, esso è il valore del segnale all’istante n. In altre parole, l’esponente di z contiene l’informazione temporale di cui abbiamo bisogno per identificare i campioni del segnale.

Quella relazione è spesso chiamata specificamente **trasformata z diretta**, dato che consente di passare dalla sequenza tempo-discreta x(n) alla sua rappresentazione X(z) nel piano complesso. La procedura inversa, cioè il passaggio da X(z) ad x(n), è detta invece **trasformata z inversa**, per evidenti motivi.

Così come accade per la trasformata di Laplace, è importante definire la **regione di convergenza** di X(z): infatti, essendo X(z) una serie di potenze, essa esiste solo per quei valori di z per i quali la sommatoria converge (ossia assume un valore finito o al più nullo). L’insieme di questi valori di z è detto appunto *regione di convergenza* o, più brevemente, **ROC**. Tutte le volte che parliamo di trasformata zeta, dobbiamo accompagnarla con la sua regione di convergenza, anche perché, come si vedrà, solo le due informazioni insieme consentono di instaurare una relazione biunivoca tra x(n) ed X(z).

Tramite degli esempi anche abbastanza semplici è possibile verificare alcune proprietà della regione di convergenza della trasformata zeta: ad esempio, si verifica facilmente che, *se x(n) è di durata finita, allora la regione di convergenza è l’intero piano complesso, eccetto due soli punti, z=0 e z=¥.* Questo risultato sarà molto importante in seguito.

In molti casi, *possiamo esprimere la sommatoria che definisce la trasformata z in forma chiusa* (ovviamente quando la sommatoria converge), il che ci dà quindi una rappresentazione alternativa, molto compatta, del segnale che stiamo considerando.

Esempio

Consideriamo, per esempio, il segnale

$$x(n) = \frac{1}{2^n} u(n)$$

dove u(n) è il gradino unitario discreto, che impone la causalità del segnale (cioè il fatto che il segnale considerato sia nullo prima dell’istante 0).

Il segnale x(n) consta di un numero infinito di valori diversi da zero (ma decrescenti nel tempo):

$$x(n) = \left\{1 \quad \frac{1}{2} \quad \frac{1}{2^2} \quad \frac{1}{2^3} \quad \dots \quad \frac{1}{2^n} \quad \dots \quad \dots\right\}$$

Calcoliamo la trasformata zeta di questo segnale: applicando la definizione, abbiamo che

$$X(z) = \sum_{n=-\infty}^{+\infty} x(n) z^{-n} = \sum_{n=0}^{+\infty} \frac{1}{2^n} z^{-n} = \sum_{n=0}^{+\infty} \left(\frac{z^{-1}}{2}\right)^n$$

Questa è una classica *serie geometrica*, ad infiniti termini, di *ragione* $z^{-1}/2$. Essa converge solo se la ragione ha modulo minore di 1, ossia $|z| > 1/2$: sotto questa condizione, possiamo esprimere X(z) in forma chiusa come

$$X(z) = \frac{1}{1 - \frac{z^{-1}}{2}}$$

La regione di convergenza è espressa proprio dalla condizione $|z| > 1/2$: ponendo z=σ+jω ed esplicitando il modulo di z, si deduce che la regione di convergenza è all'esterno del cerchio centrato nell'origine e di raggio 1/4.

Regione di convergenza

Possiamo indagare ancora meglio sulla regione di convergenza. Cominciamo col porre la variabile complessa z in forma polare ed a sostituire nell'espressione di Z(z):

$$z = re^{j\theta} \longrightarrow X(z) = \sum_{n=-\infty}^{+\infty} x(n)\left(re^{j\theta}\right)^{-n} = \sum_{n=-\infty}^{+\infty} x(n)r^{-n}e^{-jn\theta}$$

Per definizione, in corrispondenza di ciascun punto della regione di convergenza risulta |X(z)|<∞. Andiamo allora a calcolare il modulo di X(z) espressa in quel modo:

$$|X(z)| = \left|\sum_{n=-\infty}^{+\infty} x(n)r^{-n}e^{-jn\theta}\right| \le \sum_{n=-\infty}^{+\infty}\left|x(n)r^{-n}e^{-jn\theta}\right| = \sum_{n=-\infty}^{+\infty}\left|x(n)r^{-n}\right|$$

dove abbiamo tenuto conto che $e^{-jn\theta}$ è un termine di modulo unitario.

Deduciamo che |X(z)|<∞ se e solo se la sequenza $x(n)r^{-n}$ è *assolutamente sommabile*. In altre parole, la ricerca della regione di convergenza di X(z) equivale alla ricerca della regione in cui la serie $x(n)r^{-n}$ risulta *assolutamente sommabile*. Per fare questa ricerca, facciamo qualche altro passaggio su quella espressione, separando i termini della sommatoria per n negativo da quelli per n positivo:

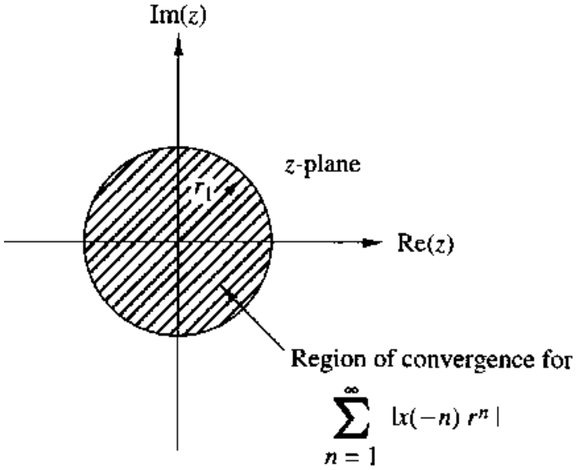
$$|X(z)| \le \sum_{n=-\infty}^{+\infty}\left|x(n)r^{-n}\right| = \sum_{n=-\infty}^{-1}\left|x(n)r^{-n}\right| + \sum_{n=0}^{\infty}\left|x(n)r^{-n}\right| = \sum_{n=-\infty}^{-1}\left|x(n)r^{-n}\right| + \sum_{n=0}^{\infty}\left|\frac{x(n)}{r^n}\right|$$

Possiamo anche fare un semplice cambio di variabile nella prima sommatoria, ponendo m=-n: si ricava evidentemente che

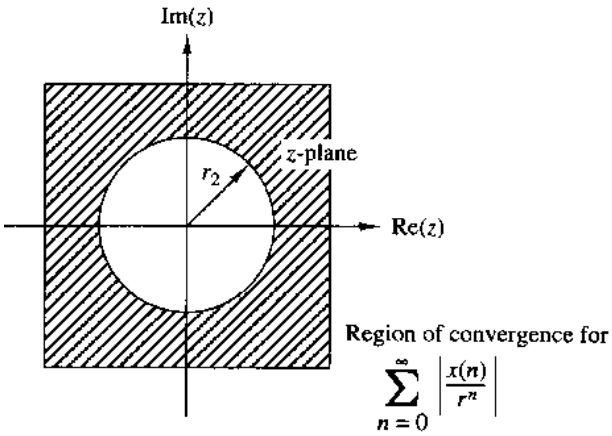
$$|X(z)| \le \sum_{m=1}^{+\infty}\left|x(-m)r^{-m}\right| + \sum_{n=0}^{\infty}\left|\frac{x(n)}{r^n}\right|$$

Se X(z) converge in qualche regione del piano complesso, è ovvio che entrambe le sommatorie a secondo membro devono a loro volta dare un valore finito:

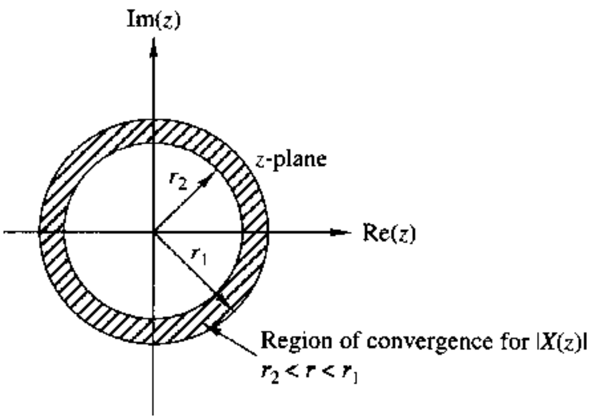
- se la prima sommatoria (m=1,...,∞) converge, significa che esiste un valore di r abbastanza piccolo affinché la sequenza $x(-m)r^{-m}$, per m=1,...,∞ risulti assolutamente sommabile. Quindi, la regione di convergenza per la prima sommatoria consiste in tutti i punti situati all'interno di un cerchio di raggio r_1<∞, come illustrato nella figura seguente:



- se la seconda sommatoria (n=0,...,∞) converge, significa che esiste un valore di r abbastanza grande affinché la sequenza $x(n)r^{-n}$, per n=0,...,∞ risulti assolutamente sommabile. Quindi, la regione di convergenza per la seconda sommatoria consiste in tutti i punti situati all’esterno di un cerchio di raggio $r_2<\infty$, come illustrato nella figura seguente:



A questo punto, dato che è richiesta la convergenza contemporanea delle due sommatorie, deduciamo che la regione di convergenza di X(z) è, in generale, la **regione anulare** data dall’<u>intersezione</u> tra le due regioni indicate prima, così come indicato nella figura seguente:



E’ ovvio, quindi, che tutto dipende dai valori di r_1 ed r_2: ad esempio se $r_2>r_1$, l’intersezione delle due regioni è nulla, per cui X(z) non converge. Un caso particolare si ha quando $r_1=r_2$: in questo caso, la regione di convergenza consiste nei punti della circonferenza di raggio $r_1(=r_2)$.

Esempi

Consideriamo il segnale

$$x(n) = a^n u(n) = \begin{cases} a^n & n \geq 0 \\ 0 & n < 0 \end{cases}$$

Calcoliamo la trasformata zeta di questo segnale: applicando la definizione, abbiamo che

$$X(z) = \sum_{n=-\infty}^{+\infty} x(n) z^{-n} = \sum_{n=0}^{+\infty} a^n z^{-n} = \sum_{n=0}^{+\infty} \left(az^{-1}\right)^n$$

Questa è una classica *serie geometrica*, ad infiniti termini, di *ragione* az^{-1}. Essa converge solo se la ragione ha modulo minore di 1, ossia $|z| > |a|$: sotto questa condizione, possiamo esprimere X(z) in forma chiusa come

$$X(z) = \frac{1}{1 - az^{-1}}$$

La regione di convergenza, espressa dalla condizione $|z| > |a|$, è quella esterna al cerchio centrato nell'origine e di raggio |a|.

Un caso particolare si ha quando a=1:

$$X(z) = \frac{1}{1 - z^{-1}} \qquad |z|>1$$

Adesso consideriamo quest'altro segnale:

$$x(n) = -a^{-n} u(-n-1) = \begin{cases} 0 & n \geq 0 \\ -a^n & n \leq -1 \end{cases}$$

Calcoliamo la trasformata zeta di questo segnale: applicando la definizione, abbiamo che

$$X(z) = \sum_{n=-\infty}^{+\infty} x(n) z^{-n} = \sum_{n=-\infty}^{-1} \left(-a^n\right) z^{-n} = \sum_{n=-\infty}^{-1} \left(-az^{-1}\right)^n$$

Possiamo fare un cambio di variabile nella sommatoria, ponendo m=-n: otteniamo evidentemente che

$$X(z) = \sum_{n=-\infty}^{-1} \left(-az^{-1}\right)^n = -\sum_{m=1}^{\infty} \left(a^{-1}z\right)^m$$

Per esprimere questa sommatoria in forma chiusa, basta ricordare la formula

$$A + A^2 + A^3 + \ldots = A\left(1 + A + A^2 + \ldots\right) = \frac{A}{1-A}$$

valida sempre per |A|<1.

Applicandola al nostro caso, possiamo dunque scrivere che

$$X(z) = -\frac{a^{-1}z}{1-a^{-1}z} = \frac{1}{1-az^{-1}} \qquad |z| < |a|$$

In questo caso, quindi, la regione di convergenza, espressa dalla condizione $|z| < |a|$, è quella interna al cerchio centrato nell’origine e di raggio |a|.

I due esempi appena illustrati sono importanti perché mettono in evidenza due proprietà. La prima riguarda la **non-unicità** della trasformata z. Infatti, se guardiamo le due trasformate ricavate nei due esempi, notiamo che sono uguali, nonostante siano stati considerati due segnali ben diversi: infatti, nonostante il segnale $a^n u(n)$ fosse un segnale causale (nullo prima di 0), mentre il segnale $-a^{-n}u(-n-1)$ fosse anticausale (cioè nullo a partire da 0), è risultato che hanno la stessa trasformata zeta:

$$Z[u(n)a^n] = Z[-a^{-n}u(-n-1)] = \frac{1}{1-az^{-1}}$$

Deduciamo, con questo semplice esempio, che *l’espressione in forma chiusa della trasformata z di un segnale non specifica univocamente il segnale stesso nel dominio del tempo.* L’informazione che interviene a risolvere l’ambiguità, e quindi a garantire l’unicità della X(z), è la regione di convergenza: notiamo infatti che i due segnali dell’esempio hanno sì la stessa trasformata zeta, ma per il primo la convergenza è al di fuori del cerchio di raggio |a|, mentre per il secondo la convergenza è all’interno dello stesso cerchio.

Concludiamo dunque affermando che un segnale tempo-discreto x(n) è univocamente determinato dall’insieme della sua trasformata zeta X(z) e della regione di convergenza di tale trasformata.

Da questo punto di vista, quando nel seguito parleremo di *trasformata zeta* di un dato segnale x(n), implicitamente ci riferiremo sia alla trasformata propriamente detta sia alla sua regione di convergenza (anche se talvolta eviteremo di indicare quest’ultima).

La seconda proprietà messa in evidenza dai due esempi riguarda la struttura della regione di convergenza:

- per il segnale causale $a^n u(n)$, la regione di convergenza è esterna al cerchio di raggio |a|;
- per il segnale anticausale $-a^{-n}u(-n-1)$, la regione di convergenza è interna al cerchio di raggio |a|.

Questa proprietà è assolutamente generale: un segnale causale ha una regione di convergenza all’esterno di un cerchio di un certo raggio r_1, mentre un segnale anticausale ha una regione di convergenza all’interno di un cerchio di un certo raggio r_2.

Esempio

Consideriamo il segnale

$$x(n) = a^n u(n) + b^n u(-n-1)$$

Questo è un segnale esteso da -∞ a +∞, ma è espresso come somma di due segnali, uno casuale (il primo) e uno anticausale (il secondo).
Calcoliamo la trasformata z:

$$X(z) = \sum_{n=-\infty}^{+\infty} x(n) z^{-n} = \sum_{n=0}^{+\infty} a^n z^{-n} + \sum_{n=-\infty}^{-1} b^n z^{-n} = \sum_{n=0}^{+\infty} \left(az^{-1}\right)^n + \sum_{m=1}^{\infty} \left(b^{-1}z\right)^m$$

Abbiamo la somma di due serie di potenze già considerate negli esempi precedenti: la prima sommatoria converge per |z|>|a|, mentre la seconda converge per |z|<|b|. Allora, per stabilire se X(z) converge, dobbiamo necessariamente considerare i due casi possibili:

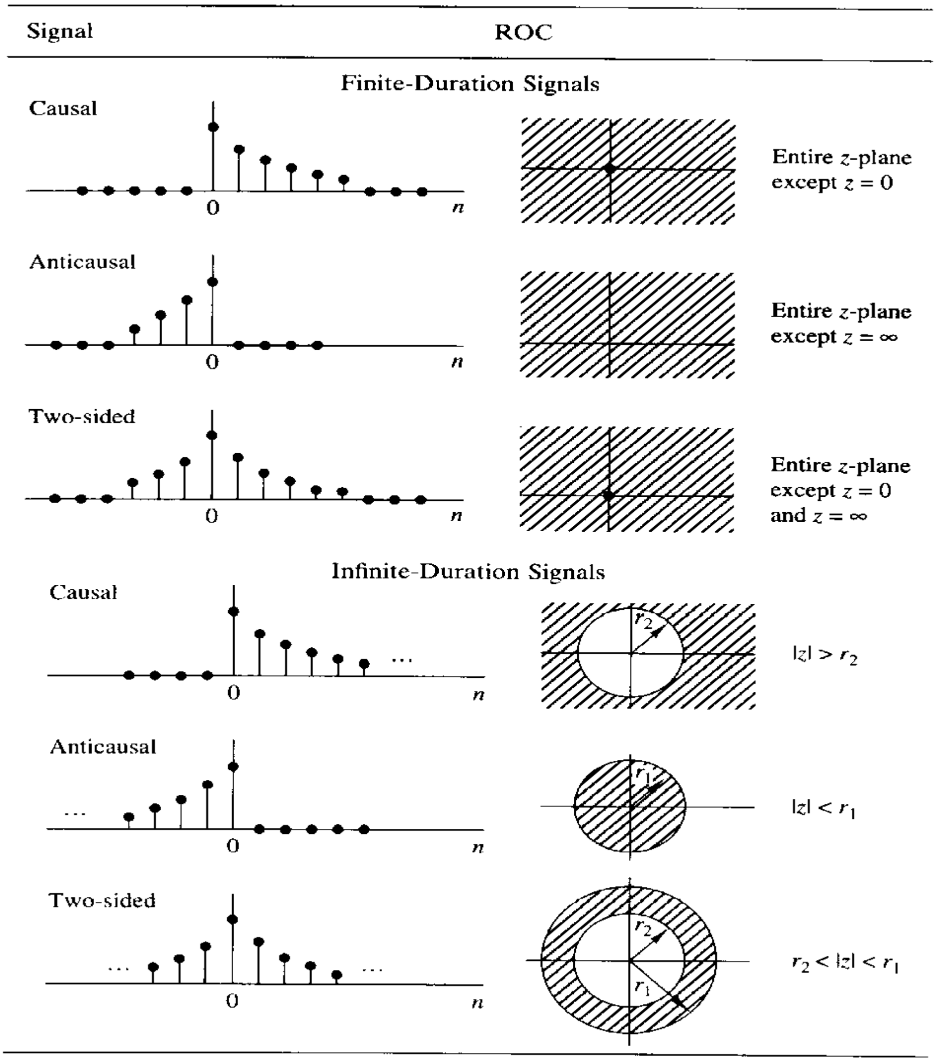
- il primo caso è quello in cui |b|<|a|: in questa situazione, le due regioni di convergenza non hanno alcun punto in comune, per cui X(z) non converge, cioè non esiste;
- il secondo caso è invece quello in cui |b|>|a|: in questa situazione, le due regioni di convergenza hanno una intersezione rappresentata dalla regione anulare di raggio interno |a| e raggio esterno |b|; per tutti i punti di questa regione, la X(z) converge e vale evidentemente

$$X(z) = \sum_{n=0}^{+\infty} \left(az^{-1}\right)^n + \sum_{m=1}^{\infty} \left(b^{-1}z\right)^m = \frac{1}{1-az^{-1}} + \frac{1}{1+bz^{-1}}$$

Sinteticamente, possiamo dire che la regione di convergenza di X(z) è rappresentata dalla condizione $|a| < |z| < |b|$.

L'esempio appena illustrato mostra un risultato molto importante: *dato un segnale tempo-discreto di durata infinita (sia per n<0 sia per n³0) e data la sua trasformata zeta, questa, se converge, lo fa in una regione anulare del piano complesso.* Detto in altre parole, le sequenze di durata infinita hanno una regione di convergenza, per la propria trasformata zeta, che è un anello all'interno del piano complesso, centrato nell'origine.

In generale, gli esempi che abbiamo fatto fino ad ora evidenziano che *la regione di convergenza della trasformata zeta di un segnale tempo-discreto dipende sia dalla durata della sequenza sia dal fatto che il segnale sia causale, anticausale, esteso da -¥ a +¥ e così via.* Nella tabella seguente sono riportate 6 tra le possibilità più comuni:



Il primo caso illustrato è quello di un **segnale causale** (cioè nullo prima di n=0) che decresce a 0 dopo un certo tempo. La regione di convergenza della sua trasformata zeta è in questo caso rappresentata dall'intero piano complesso, eccezion fatta per il l'origine. Tipicamente, questo è il caso della risposta all'impulso h(n) di un **filtro FIR**.

Il secondo caso è opposto al primo, in quanto si considera un **segnale anticausale** (cioè nullo a partire da n=0) che però si estende solo per un certo numero di campioni dall'origine n=0. La regione di convergenza, in questo caso, è l'intero piano complesso, ad eccezione dei punti all'∞.

Il terzo caso mette insieme i due precedenti, dal che consegue che la regione di convergenza è estesa a tutto il piano complesso eccetto z=0 e z=∞.

I successivi 3 casi coinvolgono invece segnali di durata infinita: nel primo caso, il segnale è causale e la regione di convergenza della sua trasformata zeta è al di fuori di un cerchio; nel secondo caso, invece, il segnale è anticausale e la sua regione di convergenza è all'interno di un cerchio. L'ultimo caso mette insieme i due precedenti, dal che consegue che la regione di convergenza è l'intersezione tra due cerchi, cioè, in generale, una regione anulare centrata nell'origine del piano complesso.

PROPRIETÀ DELLA TRASFORMATA ZETA

Le principali proprietà della trasformata zeta ricalcano quelle della trasformata di Laplace nel caso dei segnali tempo-continui.

Linearità

Tra le proprietà più importanti c'è senz'altro la **linearità**, che d'altra parte contraddistingue tutte le trasformate di interesse pratico, dato che consente di scomporre un qualsiasi segnale in una combinazione lineare di altri segnali che risultino più facilmente gestibili.

Analiticamente, nel caso della trasformata zeta, possiamo scrivere che, se **z(n)=ax(n)+by(n)** è una sequenza ottenuta come combinazione lineare delle sequenze x(n) ed y(n), la sua trasformata zeta è **Z(z)=aX(z)+bY(z)**, cioè una combinazione lineare, con gli stessi coefficienti, di X(z) ed Y(z).

La regione di convergenza di Z(z), come è facile dedurre dagli esempi visti in precedenza, è l'intersezione delle regioni di convergenza di X(z) e Y(z): infatti, dovendo combinare X(z) e Y(z), è possibile farlo solo per quei punti in cui entrambe le trasformate sono definite.

Traslazione nel tempo

Se x(n) è una sequenza temporale discreta e X(z) è la sua trasformata, la trasformata della sequenza **y(n)=x(n-k)** ottenuta traslando x(n) di n posizioni verso destra è **$Y(z)=z^{-k}X(z)$**.

E' importante, come sempre, valutare la regione di convergenza di Y(z). E' facile rendersi conto che si tratta della stessa regione di convergenza di X(z), alla quale però bisogna togliere il punto z=0 se k>0 oppure il punto z=∞ se k<0. Il motivo è evidente: Y(z) si ottiene moltiplicando X(z) per la funzione z^{-k}, per cui i punti in cui tale prodotto è definito sono gli stessi di X(z), ai quali però dobbiamo sottrarre i punti in cui non è definita la funzione z^{-k}; quest'ultima non è definita in z=0 quando k>0 (in quanto varrebbe ∞) oppure in z=∞ quando k<0 (anche in questo caso varrebbe ∞).

Proprietà di scala nel dominio trasformato

Sia data la sequenza x(n) e la corrispondente trasformata zeta X(z). Supponiamo di scalare tutti i coefficienti di X(z) di uno stesso fattore A:

$$Y(z) = X\left(\frac{z}{A}\right) = \sum_n \frac{x(n)}{A} z^{-n}$$

Si verificare facilmente, proprio dall'espressione appena riportata, che la sequenza y(n) avente Y(z) come trasformata zeta è

$$y(z) = A^n x(n)$$

In pratica, *un fattore di scala 1/A nel dominio trasformato si trasforma in un fattore di scala A^n nel dominio tempo-discreto.*

Per quanto riguarda la regione di convergenza, è ovvio che, nel passaggio da tra x(n) ed y(n), non cambia la "forma", in quanto abbiamo introdotto solo fattori di scala e non funzioni in z, ma cambia la dimensione. Infatti, ci si può rendere conto facilmente che, se la regione di convergenza di X(z) è $r_1 < |z| < r_2$, la regione di convergenza di Y(z) risulta essere $|a|r_1 < |z| < |a|r_2$.

Ovviamente, il fatto di considerare, per X(z), una regione di convergenza espressa dalla condizione $r_1 < |z| < r_2$, equivale a considerare il caso più generale: infatti, se $r_1=0$, la regione di convergenza è all’interno di un cerchio di raggio r_2, mentre $r_2=\infty$ significa che la regione di convergenza è all’esterno di un cerchio di raggio r_1.

Da notare, infine, che il fattore di scala a può essere sia reale (positivo o negativo) sia complesso.

Inversione nel tempo

Adesso consideriamo una sequenza x(n) e la sua trasformata X(z) e supponiamo di invertire nel tempo x(n). Otteniamo una sequenza **y(n)=x(-n)**, la cui trasformata risulta essere **$Y(z)=X(z^{-1})$**.

La dimostrazione rigorosa di questa proprietà si può fare tramite la definizione. Una prova molto intuitiva può essere invece la seguente: se prendiamo x(n) e la invertiamo nel tempo, il coefficiente di z^{-n}, nella sua trasformata, diventa il coefficiente di z^{n}. Di conseguenza, ribaltare un segnale è equivalente a rimpiazzare z con z^{-1} nella formula della trasformata z.

In parole povere, *un ribaltamento nei tempi equivale ad una inversione nel dominio trasformato.*

Dobbiamo anche indagare circa la regione di convergenza di $Y(z)=X(z^{-1})$, ma la cosa è immediata: se X(z) converge nella regione definita da $r_1 < |z| < r_2$, allora la funzione ottenuta da X(z) sostituendo z con z^{-1} non può che convergere nella regione $\frac{1}{r_1} < |z| < \frac{1}{r_2}$. Anche di questo si può dare una prova molto intuitiva: infatti, se z_0 è un punto che appartiene alla regione di convergenza di X(z), necessariamente il punto $1/z_0$ dovrà appartenere alla regione di appartenenza di $X(z^{-1})$.

Derivazione nel dominio trasformato

Consideriamo sempre la sequenza x(n) e la sua trasformata z. Si verifica facilmente che la funzione $z\frac{dX(z)}{dz}$ risulta essere la trasformata della sequenza **nx(n)** e risulta anche avere la stessa regione di convergenza di X(z).

Convoluzione nel tempo

Vediamo adesso una proprietà fondamentale della trasformata z. Consideriamo due sequenze x(n) ed y(n), che potrebbero essere, ad esempio, rispettivamente l’ingresso e la risposta all’impulso di un sistema lineare tempo-invariante. La convoluzione di tali sequenze dà l’uscita z(n) del sistema considerato. E’ facile verificare che z(n) ha, come trasformata z, il prodotto di X(z) e Y(z):

$$z(n) = x(n) * y(n) \xleftrightarrow{z} Z(z) = X(z)Y(z)$$

Ancora una volta, quindi, *una convoluzione nel dominio dei tempi è equivalente ad un prodotto nel dominio trasformato.*

La verifica di questo proprietà è immediata se si parte ancora una volta dalla definizione.

E’ importante anche indagare circa la regione di convergenza di Z(z). Essendo il prodotto di due funzioni, Z(z) sarà definita almeno nella regione data dall’intersezione tra le regioni di convergenza di tali funzioni. Diciamo *almeno* per il semplice fatto che, nel prodotto, una funzione potrebbe eliminare alcune singolarità dell'altra e viceversa, per cui la regione di convergenza potrebbe anche allargarsi rispetto alla semplice intersezione.

Esempio

Consideriamo le seguenti due sequenze di durata limitata:

$$x(n) = \begin{pmatrix} 1 & -2 & 1 \end{pmatrix}$$

$$h(n) = \begin{cases} 1 & 0 \le n \le 5 \\ 0 & \text{altrove} \end{cases}$$

Calcoliamo la loro convoluzione con il metodo diretto:

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

Dobbiamo dunque mantenere fissa h(n), mentre dobbiamo ribaltare e traslare x(n). Seguendo il noto metodo grafico (nell'applicare il quale supponiamo che le due sequenze siano identicamente nulle al di fuori degli intervalli forniti dalla traccia), si trova che la sequenza risultante è

$$z(n) = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 & -1 & 1 \end{pmatrix}$$

Notiamo che la lunghezza della sequenza risultante è 8 come ci si aspettava: essa è la somma, diminuita di 1, delle lunghezze delle sequenze che sono state convolute[3].

Adesso ripetiamo il calcolo applicando la proprietà della trasformata zeta enunciata poco fa: la costruzione di X(z) ed H(z) è immediata, in quanto basta costruire due polinomi in z^{-1}, di lunghezza rispettivamente 3 e 6, i cui coefficienti corrispondano ai campioni rispettivamente di x(n) ed h(n), per cui abbiamo che

$$X(z) = 1 - 2z^{-1} + z^{-2}$$
$$H(z) = 1 + z^{-1} + z^{-2} + z^{-3} + z^{-4} + z^{-6}$$

Dobbiamo adesso moltiplicare queste due sequenze: tralasciando i calcoli, indichiamo direttamente il risultato finale, ossia

$$Z(z) = X(z)H(z) = 1 - z^{-1} - z^{-6} + z^{-7}$$

Si deduce allora immediatamente che questa è la trasformata zeta della sequenza z(n) ricavata prima con il calcolo diretto: basta infatti considerare i coefficienti dei vari termini in z^{-1} per ottenere la sequenza (1,-1,0,0,0,0,-1,1).

Notiamo che Z(z) può essere ottenuta, in modo molto più rapido, trovando una espressione in forma chiusa sia per X(z) sia per H(z): in effetti, si verifica facilmente che

[3] Notiamo anche che la sequenza ottenuta è significativa in tutti i campioni solo se le sequenze di partenza sono effettivamente nulle al di fuori degli intervalli indicati. Per quanto riguarda h(n), questa garanzia ci viene data, mentre di x(n) ci vengono dati solo 3 campioni. Se non abbiamo modo di sapere quanto valga x(n) al di fuori dell'intervallo specificato, dobbiamo necessariamente ammettere che i primi e gli ultimi campioni di z(n) siano non significativi.

$$\begin{cases} X(z) = \left(1 - z^{-1}\right)^2 \\ H(z) = \dfrac{1 - z^{-6}}{1 - z^{-1}} \end{cases} \longrightarrow X(z)H(z) = \left(1 - z^{-6}\right)\left(1 - z^{-1}\right)$$

ed il prodotto tra queste due funzioni è più facile da eseguire rispetto a prima.

Osservazione: calcolo efficiente di una convoluzione

Nei precedenti capitoli abbiamo più volte fatto cenno all'opportunità di ricercare un *algoritmo efficiente* per il calcolo della convoluzione tra due sequenze temporali. Nel caso si voglia usare la trasformata zeta, l'algoritmo evolve secondo i seguenti passi:

- il primo passo è il calcolo delle trasformate zeta dei due segnali di partenza: non c'è, in realtà, da fare alcun calcolo, dato che, come visto, la trasformata di una sequenza è il polinomio in z^{-1} ottenuto usando, come coefficienti, i campioni della sequenza stessa;
- il secondo passo, l'unico a richiedere effettivamente dei calcoli, è il prodotto delle trasformate;
- il terzo ed ultimo passo, anch'esso privo di calcoli, consiste nel risalire dalla trasformata ottenuta al passo precedente alla corrispondente sequenza.

Questo algoritmo è, nella maggior parte dei casi, molto più efficiente del calcolo diretto (fatto, cioè, applicando la definizione) della convoluzione.

TRASFORMATA ZETA RAZIONALE

Per i nostri scopi, una importante famiglia di trasformate zeta è quella costituita dalle *funzioni X(z) razionali*, esprimibili cioè come rapporto tra polinomi in z^{-1} (o anche in z). Si tratta cioè di funzioni del tipo

$$X(z) = \frac{N(z)}{D(z)} = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2} + \ldots + b_M z^{-M}}{a_0 + a_1 z^{-1} + a_2 z^{-2} + \ldots + a_N z^{-N}} = \frac{\sum_{k=0}^{M} b_k z^{-k}}{\sum_{k=0}^{N} a_k z^{-k}}$$

Gli **zeri** di questa funzione sono le radici del numeratore N(z), mentre i **poli** sono le radici del denominatore D(z). E' possibile esprimere, allora, sia il numeratore sia il denominatore in modo da esplicitare le rispettive radici, ossia appunto poli e zeri di X(z):

$$X(z) = \frac{b_0}{a_0} z^{-M+N} \frac{\prod_{k=1}^{M} (z - z_k)}{\prod_{k=1}^{N} (z - p_k)}$$

In base a questa espressione, la funzione X(z) risulta avere M zeri $(z_1, z_2, \ldots, z_M)$, N poli $(p_1, p_2, \ldots, p_N)$ ed un numero |N-M| di singolarità nell'origine che sono poli se N<M oppure zeri se N>M.

Il modo con cui rappresentiamo, nel piano complesso, la mappa dei poli e degli zeri di X(z) è esattamente lo stesso visto nei precedenti corsi di studio a proposito della trasformata di Laplace: i poli si indicano con delle **croci**, mentre gli zeri si indicano con dei **cerchi**.

E' ovvio, però, che, insieme ai poli ed agli zeri di X(z), dovremo sempre indicare, nel piano complesso, la regione di convergenza di X(z). Evidentemente, gli zeri di X(z) si troveranno sicuramente all'interno di tale regione di convergenza, mentre i poli si troveranno sulla *frontiera* della regione stessa.

La conoscenza della mappa dei poli e degli zeri di una funzione X(z) consente di risalire all'espressione di X(z) stessa, mediante l'ultima formula scritta, a meno del fattore b_0/a_0, che però, essendo semplicemente un fattore di scala, non è comunque essenziale.

I prossimi esempi chiariscono questi concetti.

Esempio

Facciamo un esempio concreto. In primo luogo, consideriamo il seguente segnale:

$$x(n) = a^n u(n) = \begin{cases} a^n & n \geq 0 \\ 0 & n < 0 \end{cases}$$

Si tratta di un segnale causale di durata illimitata. Abbiamo già calcolato la sua trasformata zeta:

$$X(z) = \sum_{n=-\infty}^{+\infty} x(n) z^{-n} = \sum_{n=0}^{+\infty} a^n z^{-n} = \sum_{n=0}^{+\infty} \left(az^{-1}\right)^n = \frac{1}{1-az^{-1}}$$

Questa funzione ha una regione di convergenza rappresentata dalla condizione $|z| > |a|$: si tratta cioè della regione esterna al cerchio di raggio unitario.

Adesso modifichiamo il segnale x(n) di partenza, restringendo la sua durata ad un numero M di campioni:

$$x(n) = \begin{cases} a^n & 0 \leq n \leq M-1 \\ 0 & n < 0 \end{cases}$$

La trasformata zeta risulta essere

$$X(z) = \sum_{n=-\infty}^{+\infty} x(n) z^{-n} = \sum_{n=0}^{M-1} a^n z^{-n} = \sum_{n=0}^{M-1} \left(az^{-1}\right)^n = \frac{1-\left(az^{-1}\right)^M}{1-az^{-1}}$$

Vogliamo individuare poli e zeri di questa funzione.

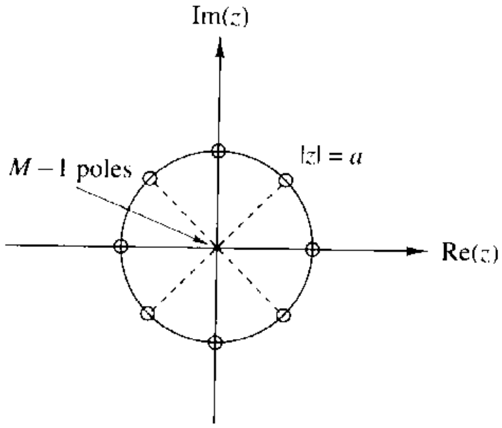
Per quanto riguarda il denominatore, abbiamo una sola radice, rappresentata evidentemente dal punto z=a. Questa radice sarà un polo della funzione X(z) solo se non è radice del numeratore: in caso contrario, essa verrebbe cancellata.

Andiamo allora a studiare il numeratore: possiamo scrivere che

$$X(z) = \frac{1-\left(az^{-1}\right)^M}{1-az^{-1}} = \frac{1-a^M z^{-M}}{1-az^{-1}} = \frac{z^{-M}\left(z^M - a^M\right)}{z^{-1}(z-a)} = \frac{\left(z^M - a^M\right)}{z^{M-1}(z-a)}$$

Questa espressione mostra chiaramente le caratteristiche di X(z): in primo luogo, sia il numeratore sia il denominatore hanno uguale grado, M, e quindi uguale numero di radici; il denominatore, ha come radici z=a (di molteplicità 1) e z=0 (di molteplicità M-1); il numeratore, invece, ha M radici corrispondenti alle soluzioni dell'equazione $z^M-a^M=0$, ossia corrispondenti ai punti $z_k = ae^{j2\pi k/M}$ (k=0,1,...,M-1). Per k=0, si ottiene k_0=a, il che significa che il polo in z=a viene cancellato dallo zero in z=a.

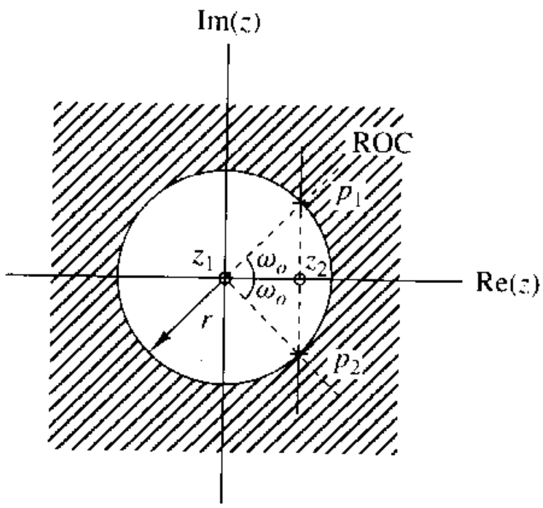
Possiamo allora concludere che X(z) ha M-1 poli (tutti nell'origine) e M-1 zeri, rappresentati da $z_k = ae^{j2\pi k/M}$ (k=1,2,...,M-1). La mappa dei poli e degli zeri, per M=8, è rappresentata nella figura seguente:



Come si nota, tutti gli zeri si trovano sul cerchio di raggio |a| in posizioni molto particolari.

Esempio

Consideriamo la seguente mappa di zeri e poli nel piano complesso:



Vogliamo determinare la trasformata z corrispondente a tale mappa (ed alla indicata regione di convergenza) ed il corrispondente segnale nel tempo (discreto).

Per ispezione della mappa, notiamo la presenza di 2 zeri (uno nell'origine, z_0=0, ed uno sull'asse reale positivo, in corrispondenza di z_1=rcos(ω_0)) e due poli complessi coniugati ($p_1 = re^{j\omega_0}$ e $p_2 = re^{-j\omega_0}$). Possiamo dunque scrivere che

$$X(z)=\frac{b_0}{a_0}z^{-M+N}\frac{\prod_{k=1}^{M}(z-z_k)}{\prod_{k=1}^{N}(z-p_k)}=\frac{b_0}{a_0}\frac{z(z-z_2)}{(z-p_1)(z-p_2)}=\frac{b_0}{a_0}\frac{z(z-r\cos\omega_0)}{(z-re^{j\omega_0})(z-re^{-j\omega_0})}$$

Il termine di guadagno b_0/a_0 resta ovviamente indeterminato. Facendo inoltre qualche semplice passaggio algebrico, si trova che

$$X(z)=\frac{b_0}{a_0}\frac{1-rz^{-1}\cos\omega_0}{1-2rz^{-1}\cos\omega_0+r^2z^{-2}}$$

Questa espressione è molto comoda, in quanto evidenzia che X(z) è una *trasformata notevole*, ossia una di quelle trasformate che si incontrano spesso nella pratica e sono state perciò riportate in apposite tabelle, insieme alla corrispondente antitrasformata: usando una di queste tabella, si deduce immediatamente che X(z) è la trasformata zeta della sequenza

$$x(n)=\frac{b_0}{a_0}r^n u(n)\cos(\omega_0 n)$$

Si tratta dunque di una sequenza causale, data la presenza del gradino unitario u(n), in cui l'unica indeterminazione è nel coefficiente di scala b_0/a_0. Essendo però tale coefficiente applicato a tutti i campioni della sequenza, non risulta comunque essenziale.

E' ovvio che, nota x(n), possiamo ritornare nel dominio trasformato e ricostruire X(z) in termini di polinomio in z^{-1}:

$$X(z)=\sum_{n=-\infty}^{+\infty}x(n)z^{-n}=\sum_{n=0}^{+\infty}\frac{b_0}{a_0}r^n\cos(\omega_0 n)z^{-n}=\frac{b_0}{a_0}\sum_{n=0}^{+\infty}r^n\cos(\omega_0 n)z^{-n}$$

Questa espressione potrebbe risultare più comoda della precedente.

L'esempio appena illustrato mette in evidenza una cosa peraltro già nota: infatti, considerando il denominatore $(z-p_1)(z-p_2)$ della funzione X(z) prima determinata, si osserva che, nonostante i due poli siano complessi e coniugati, il prodotto $(z-p_1)(z-p_2)$ è un polinomio a coefficienti reali. Questa non è altro che una conseguenza del *teorema fondamentale dell'algebra*, in base al quale un polinomio a coefficienti reali può avere radici reali e distinte e/o reali e coincidenti e/o complesse coniugate[4].

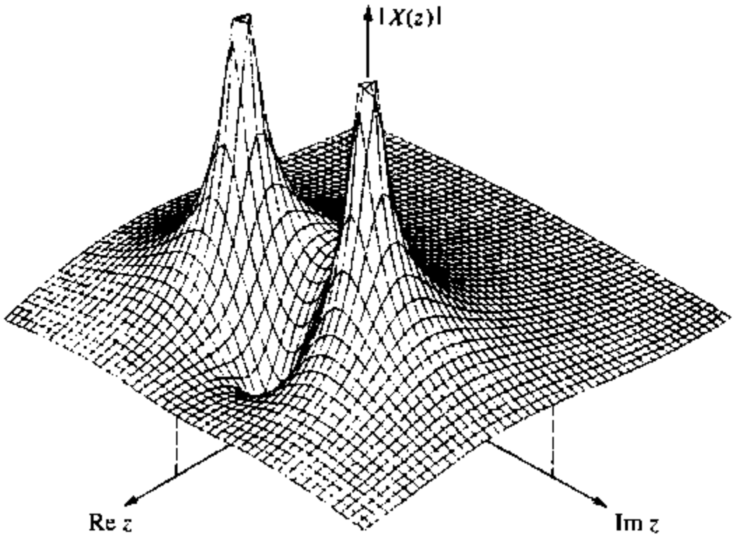
Rappresentazione tridimensionale della trasformata zeta

Per concludere, può essere interessare riportare una comoda rappresentazione di X(z), diagrammata in funzione dei punti del piano complesso. Si tratta cioè di costruire uno spazio tridimensionale, in cui l'asse verticale riporta i valori, ad esempio, di |X(z)| in corrispondenza dei punti del piano complesso (cioè il piano orizzontale del disegno), rappresentati come parte reale e parte immaginaria. Consideriamo, ad esempio, la seguente trasformata zeta:

[4] Il teorema fondamentale dell'algebra dice anche che un polinomio a coefficienti reali può sempre essere scomposto in prodotto di binomi, tanti quante sono le radici del polinomio stesso.

$$X(z) = \frac{z^{-1} - z^{-2}}{1 + 1.2732z^{-1} + 0.81z^{-2}}$$

L'andamento del modulo di questa funzione, in corrispondente dei punti del piano complesso, è il seguente (ottenuto con *Matlab*):



RELAZIONE TRA I POLI DELLA TRASFORMATA Z E DURATA NEL TEMPO

I concetti esposti nei precedenti paragrafi trovano ora una importante applicazione nella ricerca dei legami esistenti tra la *mappa* dei poli e degli zeri della trasformata zeta di una sequenza x(n) e la "*forma*" della stessa sequenza x(n) nel tempo. In particolare, restringiamo la nostra analisi solo a segnali x(n) causali e reali. Vedremo che la "forma" del generico segnale tempo-discreto x(n) dipende dalla posizione dei poli della sua trasformata Z rispetto al cerchio di raggio unitario nel piano complesso.

Cominciamo da un caso semplice, che è quello di un segnale x(n) la cui trasformata Z ha un unico polo. Questo significa che si tratta di una funzione del tipo

$$X(z) = \frac{G}{1 - az^{-1}}$$

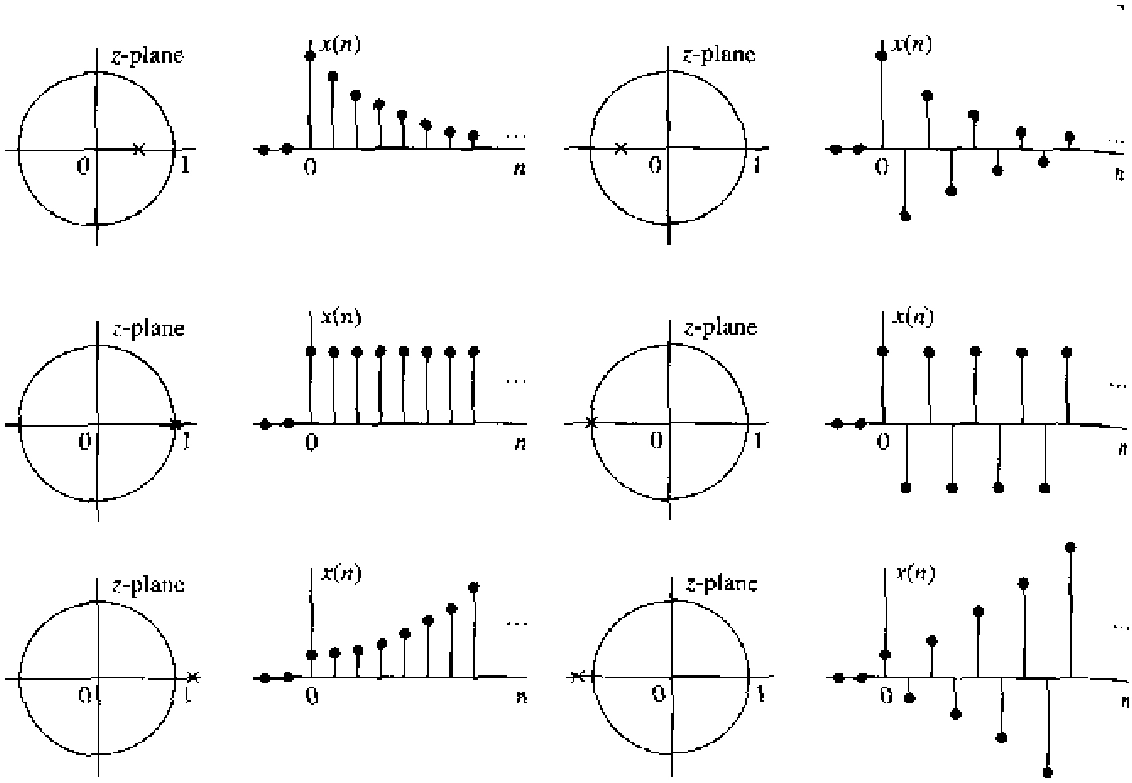
dove G è un guadagno non meglio specificato.

La prima osservazione che possiamo fare riguarda la natura di quel polo: essendo l'unica radice di un polinomio, a coefficienti reali, di grado 1, tale polo non può che essere reale.

L'unico segnale che ha una trasformata di quel tipo è il **segnale esponenziale** (ovviamente tempo-discreto):

$$x(n) = u(n)a^n \xleftrightarrow{Z} X(z) = \frac{1}{1 - az^{-1}} \qquad ROC: |z| > |a|$$

Come abbiamo già osservato, questa funzione X(z) ha un polo (reale) in z=a ed uno zero nell'origine del piano complesso. La posizione del polo, nel piano complesso, determina l'andamento di x(n) (ovviamente per n≥0). I casi possibili sono 6 e sono illustrati nella figura seguente:



Nei primi due casi (in alto a sinistra e in alto a destra), il polo è contenuto nel cerchio di raggio unitario, per cui |a|<1. Se questo accade, come si nota, il segnale decade a 0 dopo un certo tempo. La differenza, tra i due casi, è nei segni dei valori del segnale stesso prima che decada a 0: quando a>0, il segnale è sempre positivo, mentre invece, quando a<0, il segnale alterna campioni positivi e campioni negativi.

Due casi particolari dei precedenti si hanno quando a=1 e a=-1: come si nota, infatti, dalla precedente figura, in questi casi il segnale non decresce mai a 0 (per cui è di durata infinita), con la differenza che è sempre uguale ad 1 per a=1, mentre invece vale alternativamente +1 e -1 quando a=-1.

Gli ultimi due casi si hanno quando il polo z=a è all'esterno del cerchio di raggio unitario: in questo caso, il segnale non solo non decade a zero, ma diverge verso ∞, rimanendo sempre positivo per a>1 e alternando i segni per a<-1.

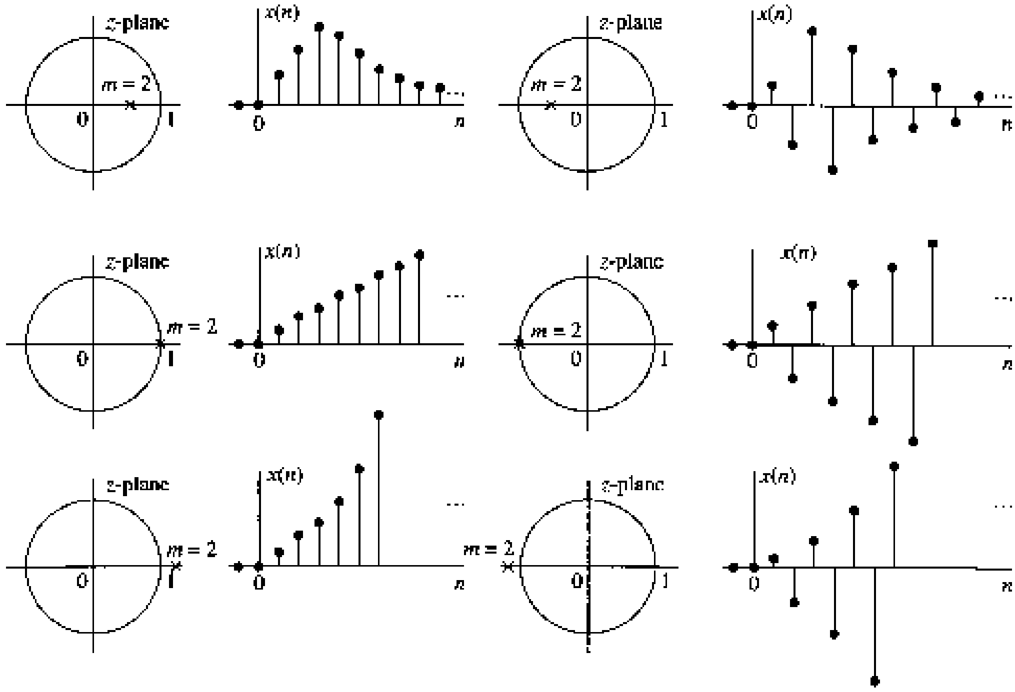
In generale, dunque, possiamo trarre le seguenti conclusioni:

- un <u>segnale reale causale</u>, la cui X(z) ha un solo polo (necessariamente reale), decade a 0 dopo un certo tempo se il polo si trova all'interno del cerchio di raggio unitario, mentre diverge se il polo si trova al di fuori dello stesso cerchio; rimane di ampiezza costante quando il polo si trova sul cerchio unitario;
- in secondo luogo, se il polo è positivo, il segnale è sempre positivo, mentre invece, se il polo è negativo, allora i campioni si alternano in segno.

Per passare da un *segnale a singolo polo* ad un *segnale con un doppio polo*, basta considerare

$$x(n) = u(n)na^n$$

La figura seguente mostra l'andamento di questo segnale, al variare della posizione dei due poli, nell'ipotesi che i due poli stessi siano reali e coincidenti:



Le prime due figure in alto illustrato il caso in cui i due poli si trovano all'interno del cerchio unitario: quando sono positivi, il segnale è sempre positivo e presenta un massimo, dopo il quale decresce a 0; quando sono negativi, si ha ancora un massimo ed una successiva decrescita a 0, con la differenza, rispetto al caso precedente, dell'inversione dei segni nei campioni.

Le due figure intermedie indicano cosa succede quando i due poli sono sul cerchio di raggio unitario, il che significa, sostanzialmente, che hanno entrambi modulo unitario e fase $\pm\pi$. In entrambi i casi si ha divergenza del segnale, il che rappresenta una evidente differenza rispetto al caso del segnale a singolo polo con il polo sul cerchio unitario (in quel caso, infatti, il segnale era di ampiezza costante).

Ancora divergenza, ma più rapida, si ha quando i due poli sono esterni al cerchio unitario.

Nella figura seguente è infine rappresentato quello che accade quando i due poli di X(z) sono complessi e coniugati, con parte immaginaria $\sin\omega_0$ e con parte reale positiva. Si nota quanto segue:

- quando i due poli sono all'interno del cerchio unitario, il segnale x(n) presenta una oscillazione sinusoidale (a frequenza ω_0) decrescente in modo esponenziale (con r^n, dove r<1 è il modulo dei due poli);
- quando i due poli si trovano sul cerchio unitario (per cui hanno modulo r=1), il segnale oscilla indefinitamente (sempre a frequenza ω_0);
- infine, quando i due poli sono al di fuori del cerchio unitario, il segnale ha una oscillazione sinusoidale (sempre a frequenza ω_0) crescente in modo esponenziale (con r^n, dove r>1).

Quindi, *in presenza di due poli complessi e coniugati, la loro distanza dall'origine determina l'inviluppo del segnale sinusoidale, mentre il loro angolo rispetto all'asse reale positivo determina la frequenza dell'oscillazione.*

Senza continuare ad esaminare le diverse situazioni possibili (anche perché quelle appena illustrate sono senz'altro le più ricorrenti), possiamo sintetizzare nel modo seguente: i segnali causali reali aventi poli reali semplici o poli complessi coniugati, disposti all'interno o sul cerchio unitario, sono sempre limitati in ampiezza (in particolare, quando più tali poli sono vicini all'origine, tanto più rapidamente avviene la decadenza del segnale a zero); se i poli sono all'esterno del cerchio unitario, allora i segnali sono divergenti in ampiezza.

Quindi, l'andamento nel tempo di un segnale dipende fortemente dalla posizione relativa dei poli della sua trasformata zeta rispetto al cerchio unitario.

In questo discorso, non sono minimamente subentrati gli eventuali poli che può avere la trasformata zeta del segnale considerato. In effetti, anche gli zeri influenzano l'andamento del segnale, ma non in modo così rigido come i poli. Tanto per fare un esempio, nel caso di un segnale sinusoidale, la presenza e la posizione degli zeri influenza solo la fase del segnale.

SISTEMI LINEARI TEMPO-INVARIANTI E TRASFORMATA ZETA

Sappiamo bene che l'uscita di un sistema tempo-discreto lineare tempo-invariante, in risposta ad un ingresso x(n), è la convoluzione di x(n) con la funzione di risposta all'impulso h(n) del sistema stesso. Interpretando allora tale convoluzione nel dominio della trasformata zeta, possiamo scrivere che *la trasformata dell'uscita è il prodotto delle trasformate dell'ingresso e della risposta all'impulso:*

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

Questa relazione può avere utilizzi diversificati. Quello più semplice consiste nel valutare Y(z), da cui poi risalire ad y(n), note che siano H(z) ed X(z), cioè caratteristiche del sistema ed eccitazione. In alternativa, si può pensare di misurare l'ingresso e l'uscita del sistema, calcolarne le rispettive trasformate e determinare infine la funzione H(z):

$$H(z) = \frac{Y(z)}{X(z)}$$

Sostituendo a ciascuna di queste trasformate le rispettive espressioni esplicite, abbiamo quanto segue:

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

Abbiamo qui considerato sequenze temporali di lunghezza infinita, ma è evidente che basta modificare opportunamente gli indici delle sommatorie per ricondursi a sequenze limitate temporalmente.

La funzione H(z) è dunque rappresentativa del sistema considerato nel dominio trasformato. Essa prende il nome di **funzione di sistema**.

L'espressione $H(z) = \frac{Y(z)}{X(z)}$ è particolarmente utile per ottenere H(z) quando il sistema ha una descrizione in termini di **equazione lineare alle differenze finite a coefficienti costanti**[5]:

[5] Si tratta semplicemente dell'analogo tempo-discreto ad una *equazione differenziale lineare, a coefficienti costanti*, che useremmo nel caso tempo-continuo.

$$\boxed{y(n)+\sum_{k=1}^{N}a_k y(n-k)=\sum_{k=0}^{M}b_k x(n-k)}$$

Questa relazione mostra, in pratica, come l'uscita del sistema, ad un generico istante n, dipenda, in generale, sia dai valori dell'ingresso negli ultimi M+1 istanti (cioè l'istante considerato e gli M precedenti) sia anche dai valori della stessa uscita in un certo numero N di istanti precedenti (definizione di **sistema tempo-discreto recursivo**[6]).

Vediamo allora come possiamo giungere ad H(z) (e quindi anche ad h(n)) partendo da quella equazione.

In primo luogo, se trasformiamo ambo i membri di quella equazione, possiamo applicare le proprietà di linearità e di traslazione nel tempo, per scrivere che

$$\begin{cases} y(n)\xleftrightarrow{Z}Y(z) \\ y(n-k)\xleftrightarrow{Z}z^{-k}Y(z) \\ x(n-k)\xleftrightarrow{Z}z^{-k}X(z) \end{cases} \longrightarrow \quad Y(z)+\sum_{k=1}^{N}a_k z^{-k}Y(z)=\sum_{k=0}^{M}b_k z^{-k}X(z)$$

Le trasformate possono essere tirate fuori dalle sommatorie, per cui si ricava facilmente che

$$Y(z)\left(1+\sum_{k=1}^{N}a_k z^{-k}\right)=X(z)\sum_{k=0}^{M}b_k z^{-k} \longrightarrow \boxed{H(z)=\frac{Y(z)}{X(z)}=\frac{\sum_{k=0}^{M}b_k z^{-k}}{1+\sum_{k=1}^{N}a_k z^{-k}}}$$

La conclusione cui siamo pervenuti (peraltro abbastanza scontata data l'analogia con la trasformata di Laplace nel caso tempo-continuo), è quindi che *un sistema lineare tempo-invariante, descritto da una equazione alle differenze a coefficienti costanti, ha una funzione di sistema di tipo razionale fratto.*

E' da notare il significato dei termini M ed N: per quanto si è visto, M (grado del numeratore) è il numero dei campioni di ingresso arrivati prima dell'istante corrente e che vanno ad influire sull'uscita nell'istante corrente, mentre invece N (grado del denominatore) è il numero dei campioni di uscita precedentemente prodotti e che vanno anch'essi ad influire sull'uscita nell'istante corrente.

Quella ottenuta è la forma più generale per H(z). Ci possono però essere due casi particolari, che in seguito verranno analizzati con grande dettaglio:

- il primo caso è quello in cui i coefficienti a_k sono identicamente nulli: la funzione di sistema si riduce quindi semplicemente a

$$H(z)=\sum_{k=0}^{M}b_k z^{-k}=\frac{1}{z^M}\sum_{k=0}^{M}b_k z^{M-k}$$

  Questa è una funzione che contiene M poli nell'origine[7] ed M zeri generici, i cui valori dipendono evidentemente dai coefficienti b_k. Si parla, in questo caso, di **sistema a soli zeri**.

[6] Un sistema tempo-discreto di tipo recursivo è in pratica l'analogo di un sistema tempo-continuo reazionato (negativamente). L'uscita in un dato istante non dipende solo dallo stato del sistema nell'istante considerato, ma anche da un certo numero di istanti precedenti (al limite infiniti).

[7] Ciò che è lo stesso, un polo nell'origine di molteplicità M

La caratteristica di questo sistema è quella di avere una funzione di risposta all'impulso h(n), antitrasformata di H(z), di durata finita ed è per questo che si parla anche di **sistema FIR**;

- il duale del caso precedente è quello in cui sono i coefficienti b_k ad essere nulli, eccezione fatta per b_0 (in caso contrario si avrebbe una funzione identicamente nulla):

$$H(z) = \frac{b_0}{1+\sum_{k=1}^{N} a_k z^{-k}} = \frac{b_0 z^N}{\sum_{k=0}^{N} a_k z^{N-k}}$$

Questa funzione presenta N zeri nell'origine ed N poli generici, i cui valori dipendono dai coefficienti a_k. Si parla, in questo caso, di **sistema a soli poli** o, equivalentemente, si **sistema IIR**: infatti, la caratteristica di questo sistema è quella di avere una funzione di risposta all'impulso h(n) di durata infinita.

Quando il sistema considerato è caratterizzato da una H(z) che contiene sia poli (in numero N) sia zeri (in numero M), si parla di **sistema poli-zeri**. Un sistema poli-zeri, proprio per la presenza dei poli, è necessariamente un sistema IIR. Possiamo perciò affermare che un sistema a soli poli è un sistema FIR, mentre tutti gli altri sistemi sono IIR.

TRASFORMATA Z INVERSA

In base alla definizione, sappiamo che è immediato il passaggio da una sequenza x(n) alla corrispondente X(z), mentre non necessariamente è immediato il passaggio inverso:

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

Per passare da X(z) ad x(n), ci basta costruire il polinomio in z^{-1} avente, come coefficienti, proprio i campioni di x(n). Viceversa, il passaggio da X(z) ad x(n) è immediato solo se conosciamo X(z) in forma di polinomio. In caso contrario, la situazione è più complicata.

Sicuramente, spesso ci vengono in aiuto delle apposite tabelle in cui vengono riportate alcune trasformate notevoli e le corrispondenti sequenze (nonché le regioni di convergenza). Supponiamo, però, di poter ricorrere a tali tabelle e di non riconoscere in X(z) una trasformata a noi nota. Come procediamo per ricavare la corrispondente x(n)?

Un modo efficiente di procedere è il seguente:

- in primo luogo, osservando come sia fatta la regione di convergenza di X(z) possiamo capire che tipo di segnale dobbiamo aspettarci: se la regione di convergenza è all'esterno di un cerchio, il segnale x(n) sarà causale, mentre invece, se la regione di convergenza è all'interno di un cerchio, siamo sicuri che il segnale x(n) è anticausale;
- in secondo luogo, sempre dalla regione di convergenza possiamo affermare se X(z) è esprimibile come espansione di una serie di potenze di z^{-1} (per i segnali causali) oppure di potenze di z (per i segnali anticausali).

Facciamo allora un esempio significativo, considerando la seguente funzione di z:

$$X(z) = \frac{1}{1 - 1.5z^{-1} + 0.5z^{-2}}$$

Questa funzione, da sola, non è rappresentativa di un unica antitrasformata x(n). Per ottenere l'unicità, abbiamo bisogno di conoscere la sua regione di convergenza.

Supponiamo per esempio che tale regione sia rappresentata dalla condizione |z|>1. Essendo, dunque, all'esterno di un cerchio, deduciamo che il segnale x(n) sarà causale e deduciamo inoltre che possiamo esprimere X(z) come espansione di una serie di potenze nelle potenze negative di z^{-1}:

$$X(z) = a_0 + a_1 z^{-1} + a_2 z^{-2} + \ldots$$

Per ricavare i coefficienti della serie, ci basta effettuare la divisione del numeratore di X(z) per il denominatore: dalla divisione[8] si ottiene

$$X(z) = \frac{1}{1 - \frac{3}{2}z^{-1} + \frac{1}{2}z^{-2}} = 1 + \frac{3}{2}z^{-1} + \frac{7}{4}z^{-2} + \frac{15}{8}z^{-3} + \ldots$$

I coefficienti di questa serie non sono altro, per definizione, che i campioni della corrispondente antitrasformata:

$$x(n) = \begin{pmatrix} \underset{\uparrow}{1} & \frac{3}{2} & \frac{7}{4} & \frac{15}{8} & \ldots. \end{pmatrix}$$

Quella "freccetta" verticale indica il campione in posizione n=0 (si ricordi che il segnale deve essere causale).

Le cose cambiano completamente se, conservando la stessa X(z), cambiamo la regione di convergenza, considerando ad esempio quella rappresentata dalla condizione |z|<0.5. Siamo in questo caso all'interno di un cerchio (di raggio 0.5), per cui il segnale x(n) sarà questa volta anticausale.

L'espressione di X(z) sarà questa volta in termini di potenze positive di z. Ancora una volta, dobbiamo dividere numeratore e denominatore[9], ma dobbiamo aver cura di considerare il denominatore con i termini disposti nell'ordine delle potenze crescenti di z:

$$X(z) = \frac{1}{\frac{1}{2}z^{-2} - \frac{3}{2}z^{-1} + 1} = 2z^2 + 6z^3 + 14z^4 + 30z^5 + \ldots$$

Deduciamo allora che la sequenza avente questa funzione come trasformata zeta è

$$x(n) = \begin{pmatrix} \ldots & 30 & 14 & 6 & 2 & 0 & \underset{\uparrow}{0} \end{pmatrix}$$

E' ovvio, comunque, che il metodo appena esposto diventa troppo noioso e laborioso quando il segnale x(n) ha lunga durata (al limite infinita) e si vuole un numero elevato di campioni per x(n), nel qual caso i passaggi

[8] nella quale, ad ogni passo, va eliminato il termine nella potenze più piccola di z^{-1};

[9] In questo caso, ad ogni passo va eliminato il termine nella potenze più piccola di z;

della divisione diventano davvero troppi. Quindi, è un metodo applicabile tipicamente quando abbiamo bisogno di conoscere solo i primi campioni del segnale.

Sicuramente più convenienti sono i metodi che si basano, così come per la trasformata di Laplace nel caso continuo, nella scomposizione in fratti semplici della X(z), al fine di ricondurci alle trasformate notevole più conosciuti. Ad ogni modo, non scendiamo nei dettagli di tali metodi.

Osservazione

Abbiamo tirato in ballo, nel precedente paragrafo, il *teorema fondamentale dell'algebra*. Tra le altre cose, questo teorema afferma che <u>è sempre possibile, data una sequenza di numeri, rappresentarla come convoluzione di sequenze opportune</u>. Vediamo qualche dettaglio in più.

Il teorema fondamentale dell'algebra dice che ogni polinomio di ordine N, a coefficienti complessi, ha N radici nel campo complesso $z_1,z_2,...,z_N$. Questo polinomio può essere espresso nella duplice forma

$$A(z)=\sum_{n=0}^{N}a_n z^{-n}=a_0\prod_{n=1}^{N}\left(1-z_n z^{-1}\right)$$

Il polinomio è dunque ottenibile come prodotto di N termini del tipo $\left(1-z_n z^{-1}\right)$. Se il polinomio rappresenta la trasformata zeta di una sequenza lunga, in questo caso, N+1 campioni (da n=0 ad n=N), quella relazione dice in pratica che la sequenza è decomponibile nel prodotto di convoluzione che segue:

$$\{a_n\}=a_0*\{1,-z_1\}*\{1,-z_2\}*\{1,....\}*..*\{1,-z_N\}$$

Quindi, ogni sequenza di N+1 campioni complessi, che inizia a partire dall'origine nei tempi (il primo campione è posto nella posizione di indice 0), si può decomporre nella convoluzione di N sequenze elementari costituite, ciascuna, da 2 campioni complessi, di cui il primo è di ampiezza unitaria ed è posizionato nell'origine mentre il secondo, posizionato al posto di indice 1, ha valore $-z_n$ (n=1,...,N). Questi z_n sono gli zeri della trasformata zeta della sequenza.

Vediamo la cosa in modo ancora più concreto, considerando la sequenza $\{a_n\}=\{a_0,a_1,a_2\}$ di lunghezza finita 3. La possiamo ottenere come convoluzione di due sequenze ciascuna di due campioni: siano $\{b_n\}=\{b_0,b_1\}$ e $\{d_n\}=\{d_0,d_1\}$ tali sequenze. Dobbiamo determinare i valori di queste sequenze. Andiamo allora a farne la convoluzione, calcolandola come prodotto delle rispettive trasformate zeta:

$$\begin{cases}\{b_n\}=\{b_0,b_1\}\xrightarrow{\text{trasformando}}B(z)=b_0+b_1z^{-1}\\ \{d_n\}=\{d_0,d_1\}\xrightarrow{\text{trasformando}}D(z)=d_0+d_1z^{-1}\end{cases}\longrightarrow B(z)D(z)=b_0d_0+\left(b_0d_1+d_0b_1\right)z^{-1}+b_1d_1z^{-2}$$

Dobbiamo adesso imporre l'uguaglianza con la trasformata zeta della sequenza $\{a_n\}=\{a_0,a_1,a_2\}$, ossia con il polinomio $A(z)=a_0+a_1z^{-1}+a_2z^{-2}$:

$$a_0+a_1z^{-1}+a_2z^{-2}=b_0d_0+\left(b_0d_1+d_0b_1\right)z^{-1}+b_1d_1z^{-2}\longrightarrow\begin{cases}a_0=b_0d_0\\ a_1=b_0d_1+d_0b_1\\ a_2=b_1d_1\end{cases}$$

Questo è un sistema di 3 equazioni in quattro incognite, che sono appunto i termini delle sequenze $\{b_n\}$ e $\{d_n\}$. Abbiamo quindi un grado di libertà (il sistema ammette ∞^1 soluzioni) e lo impieghiamo, ad esempio, ponendo $d_0=1$. Con questa posizione, si ottiene

$$\begin{cases} b_0 = a_0 \\ a_1 = a_0 d_1 + b_1 \\ a_2 = b_1 d_1 \end{cases}$$

Esplicitando b_1 (risp. d_1) nell'ultima equazione e sostituendo nella seconda, si ottiene una equazione di 2° grado nell'incognita d_1 (risp. b_1) , che quindi ammette due soluzioni per d_1 (risp. b_1):

$$a_1 = a_0 d_1 + \frac{a_2}{d_1} \longrightarrow a_0 d_1^2 - a_1 d_1 + a_2 = 0 \xrightarrow{\text{risolvendo}} (d_1)_{1/2} = \frac{a_1 \pm \sqrt{a_1^2 - 4a_0 a_2}}{2a_0}$$

Ovviamente, dato che b_1 dipende linearmente da d_1, avremo due soluzioni anche per b_1. Abbiamo dunque due possibili sequenze $\{b_n\}$ e $\{d_n\}$ che, convolute, danno la sequenza di partenza $\{a_n\}$.

Legami tra trasformata Z e trasformata di Fourier discreta

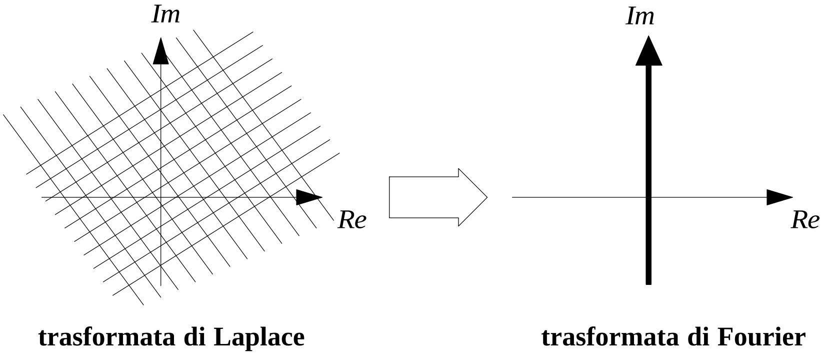
Abbiamo visto che la trasformata zeta di un segnale x(n) è l'equivalente nel discreto della trasformata di Laplace del segnale x(t):

$$x(t) \xleftrightarrow{\text{Laplace}} X(s) = \int s(t) e^{-st} dt$$

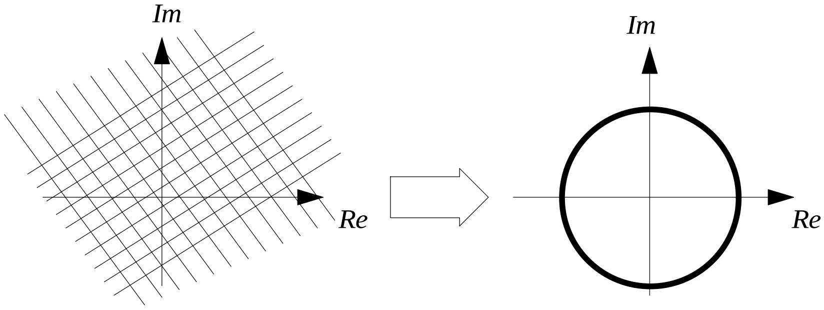
$$x(n) \xleftrightarrow{Z} X(z) = \sum_n x(n) z^{-n}$$

La trasformata di Laplace usa come *funzioni base* (cioè funzioni elementari che, combinate con opportuni coefficienti, forniscono una descrizione alternativa della funzione tempo-continua di partenza) gli esponenziali complessi $\mathbf{e^{st}}$, dove $s=\sigma+j\omega$. Analogamente, la trasformata zeta è una scomposizione delle sequenze tempo-discrete in termini di funzioni base del tipo z^n, ossia potenze di $z=\sigma+j\omega$.

Nel caso continuo, se la regione di convergenza della trasformata di Laplace ha al suo interno l'asse immaginario (cioè tutti i punti $s=j\omega$ puramente immaginari), allora esiste la trasformata di Fourier di x(t), che non è altro che la trasformata di Laplace calcolata sull'asse immaginario:



Una situazione analoga si ha nel caso tempo-discreto: per passare dalla trasformata zeta della sequenza x(n) alla trasformata di Fourier discreta (DTFT), basta porre $z=e^{j2\pi fT}$, ossia basta calcolare la trasformata zeta nei punti del cerchio unitario:



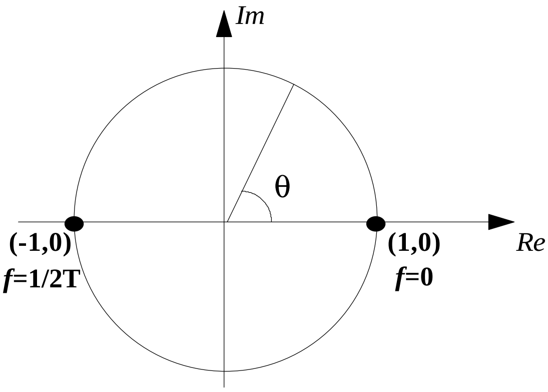
trasformata Zeta **trasformata di Fourier discreta (DTFT)**

Analiticamente, abbiamo perciò che

$$x(t) \xleftrightarrow{\text{Laplace}} X(s) = \int s(t)e^{-st}dt \xrightarrow{s=j\omega} X(f) = \int s(t)e^{-j\omega t}dt$$

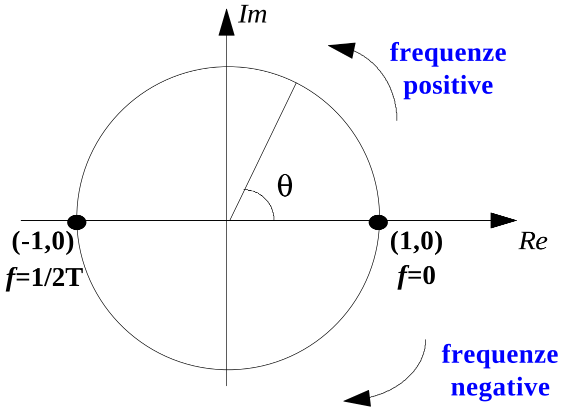
$$x(n) \xleftrightarrow{Z} X(z) = \sum_n x(n)z^{-n} \xrightarrow{z=e^{j2\pi fT}} X(f) = \sum_n x(n)e^{-j2\pi fnT}$$

E' molto importante evidenziare la corrispondenza che esiste tra i punti del piano complesso situati sul cerchio di raggio unitario e le frequenze f che compaiono nella DTFT di x(n). Ad esempio, considerando il termine $z=e^{j2\pi fT}$, se lo calcoliamo per f=0 otteniamo z=1: quindi, la frequenza f=0 corrisponde al punto (1,0) del cerchio unitario. Analogamente, se poniamo f=1/2T, otteniamo z=-1: quindi la frequenza f=1/2T, cioè la *frequenza di Nyquist*, corrisponde al punto (-1,0) del cerchio unitario:



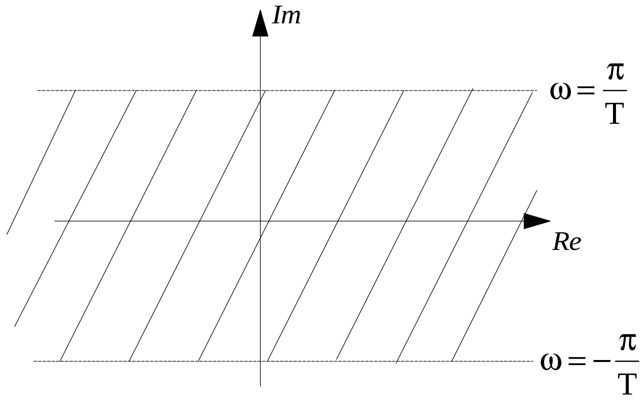
Visto che consideriamo solo punti del cerchio unitario, possiamo individuare ciascun punto tramite un *angolo* θ (cioè una fase), visto che la distanza dall'origine è sempre unitaria. Quindi, ***ad ogni frequenza della DTFT della sequenza x(n) sarà associato un angolo q nel piano complesso:*** avremo ad esempio θ=0 per la frequenza *f*=0 e θ=π per la frequenza di Nyquist *f*=1/2T.

Gli angoli compresi tra θ=0 e θ=π (**semipiano superiore**) sono evidentemente rappresentativi di frequenze comprese tra 0 a 1/2T, ossia sono rappresentativi delle <u>frequenze positive</u>, così come gli angoli compresi tra -π e 0 (**semipiano inferiore**) sono rappresentativi delle frequenze negative[10], cioè quelle comprese tra -1/2T e 0:



In questo modo si esprime, tra le altre cose, la natura periodica della funzione X(f), cioè della DTFT di x(n). Questa funzione spesso viene indicata anche con **X(θ)**, proprio in virtù della citata corrispondenza tra frequenze e angoli nel piano complesso (limitatamente al cerchio unitario).

Continuando i nostri discorsi, vediamo il legame esistente tra il piano complesso di Laplace e quello della trasformata zeta. Per trovare questo legame, però, non possiamo non considerare un campionamento nei tempi anche per il segnale di cui calcoliamo la trasformata di Laplace, la quale quindi dovrà essere periodica. Allora, se c'è questa periodicità, non ha senso considerare l'intero piano complesso, ma basterà una sola porzione, che sarà quella corrispondente a pulsazioni ω comprese tra π/T (pulsazione di Nyquist positiva) e -π/T (pulsazione di Nyquist negativa):



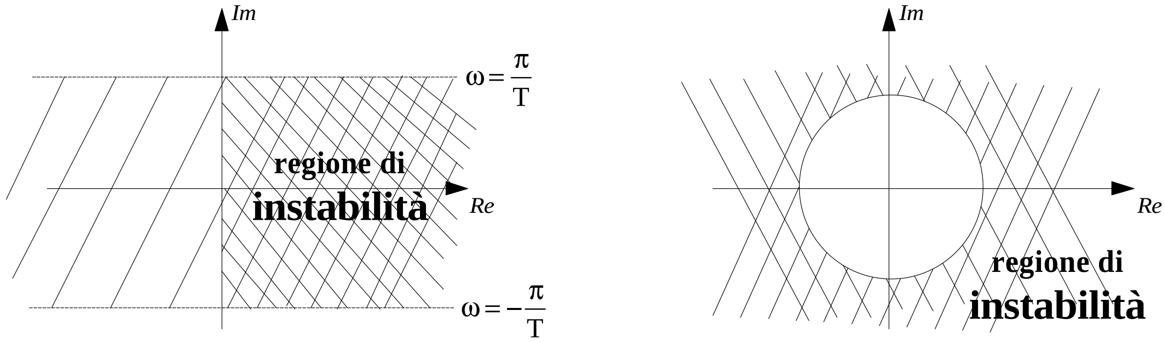
Allora, *il segmento di asse immaginario compreso tra w=-p/T e w=p/T corrisponde al cerchio unitario della trasformata zeta.* Ai punti del semipiano destro corrispondono i punti esterni al cerchio unitario, mentre ai punti del semipiano sinistro corrispondono i punti interni al cerchio unitario.

[10] Ricordiamo che stiamo considerando segnali tempo-continui campionati nei tempi, per cui il loro spettro (la DTFT) si ripete periodicamente, con un periodo pari alla frequenza di campionamento f_C=1/T. Questo è il motivo per cui ha senso considerare solo le frequenze che vanno da -f_C/2 a f_C/2, cioè quelle del cosiddetto periodo fondamentale: in queste frequenze è compreso lo spettro del segnale x(t) di partenza (sempre nell'ipotesi di aver rispettato il teorema del campionamento).

STABILITÀ BIBO

La corrispondenza, esaminata nel paragrafo precedente, tra il piano complesso della trasformata di Laplace e quello della trasformata zeta consente di estendere alla trasformata zeta i concetti di stabilità/instabilità di un sistema visti tramite la trasformata di Laplace:

- così come un polo del sistema H(ω) situato nel semipiano destro corrisponde ad instabilità, analogamente *un polo del sistema H(z) situato all'esterno del cerchio di raggio unitario determinerà l'instabilità del sistema* (il che significa che ad un ingresso di ampiezza limitata corrisponde una uscita di ampiezza non limitata);
- viceversa, così come un polo del sistema H(ω) situato nel semipiano sinistro corrisponde ad asintotica stabilità, analogamente *un polo del sistema H(z) situato all'interno del cerchio di raggio unitario determinerà l'asintotica stabilità del sistema.*



Quindi, affermiamo che la **zona di stabilità**[11], all'interno della quale devono essere contenuti tutti i poli della trasformata zeta di sistemi lineari tempo-invarianti, corrisponde alla parte interna del cerchio unitario. In realtà, c'è da precisare che questo risultato vale solo per i sistemi causali, la cui risposta all'impulso sia cioè identicamente nulla per n<0.

Questa conclusione, ottenuta per via essenzialmente qualitativa, può in realtà essere dedotta in modo più rigoroso. Dimostriamo perciò che *un sistema lineare tempo-invariante causale H(z) è stabile BIBO se e solo se le sue singolarità sono contenute nel cerchio di raggio unitario.*

Per fare questo, ricordiamo intanto un risultato generale ricavato nel corso di *Teoria dei Segnali*: condizione necessaria e sufficiente affinché un sistema lineare tempo-invariante tempo-discreto sia stabile è che la sua risposta all'impulso sia tale che la somma dei moduli dei suoi campioni sia convergente. In formule, deve cioè risultare

$$\sum_n |h(n)| < \infty$$

Supponiamo allora che il nostro sistema sia stabile. La trasformata zeta della funzione di risposta all'impulso è

$$H(z) = \sum_n h(n) z^{-n}$$

[11] Ci riferiamo, in particolare, alla **stabilità BIBO**: *ad ingresso limitato in ampiezza corrisponde una uscita anch'essa limitata in ampiezza.*

Calcoliamo il modulo di entrambi i membri di questa relazione:

$$\left|H(z)\right| = \left|\sum_n h(n)z^{-n}\right| \le \sum_n \left|h(n)z^{-n}\right| = \sum_n \left|h(n)\right|\left|z^{-n}\right|$$

Adesso, calcoliamo |H(z)| sul cerchio unitario, il che equivale a prendere |z|=1:

$$\left|H(z)\right| \le \sum_n \left|h(n)\right| < \infty$$

Dato che $\sum_n \left|h(n)\right| < \infty$, deduciamo che il modulo di H(z) risulta finito per qualsiasi punto venga preso sul cerchio unitario. Questo significa che, se il sistema è stabile (BIBO) per ipotesi, allora i punti del cerchio unitario sono contenuti nella regione di convergenza di H(z).

E’ facile verificare che vale anche il viceversa, da cui deduciamo che *un sistema lineare tempo-invariante è stabile BIBO se e solo se la regione di convergenza della sua funzione di sistema H(z) contiene i punti del cerchio unitario.*

Questo però è solo un risultato parziale, in quanto non abbiamo ancora trovato il legame tra la stabilità e la posizione dei poli rispetto al cerchio unitario.

Per trovare questo legame, dobbiamo riprendere il concetto di *causalità* di un sistema. Ricordiamo infatti che un sistema *causale* è caratterizzato da una funzione di sistema H(z) la cui regione di convergenza è all’esterno di un cerchio di un certo raggio r. Allora, se noi vogliamo che il sistema sia stabile oltre che causale, dobbiamo imporre che la funzione H(z) converga per tutti i punti z tali da soddisfare la relazione

$$\left|z\right| > r < 1$$

In questo modo, infatti, la regione di convergenza include i punti del cerchio unitario e continua ad essere composta dai punti a distanza ≥r dall’origine.

Dato, però, che la regione di convergenza non può contenere alcun polo di H(z), concludiamo che *un sistema causale lineare tempo-invariante è stabile BIBO se e solo se tutti i poli di H(z) sono all’interno del cerchio unitario.*

Autore: **SANDRO PETRIZZELLI**
e-mail: sandry@iol.it
sito personale: http://users.iol.it/sandry
succursale: http://digilander.iol.it/sandry1