

Il metodo della funzione descrittiva

Introduzione

Accade spesso di dover studiare un sistema di controllo in cui sono presenti sottosistemi non lineari. Alcuni di tali sottosistemi sono descritti da equazioni differenziali non lineari, ad esempio i sistemi preda-predatore o i corpi rigidi girevoli intorno a un punto fisso. In questo caso la rappresentazione con lo stato assume la forma:

$$\dot{x} = f(x, u)$$

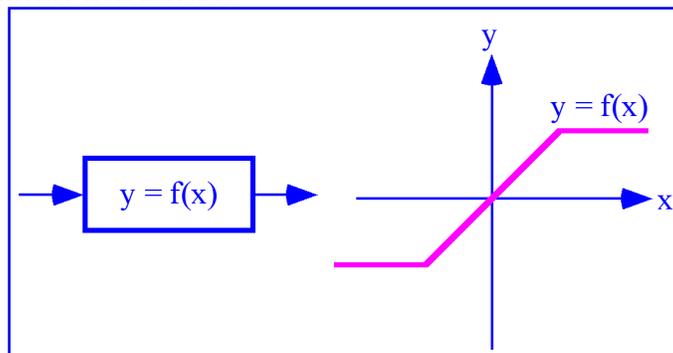
$$y = h(x)$$

e al progettista rimangono due vie: o affrontare il problema con le moderne metodologie non lineari, complesse e tuttora incomplete, oppure ricorrere alla linearizzazione quando questa sia accettabile.

A volte, invece, il comportamento non lineare è presente soltanto in alcuni sottosistemi che non dipendono dal tempo. Tali sottosistemi, cioè, non sono descritti da equazioni differenziali ma da funzioni che ne forniscono l'uscita y , noto l'ingresso x :

$$y = f(x)$$

La f , assegnata in forma analitica o in forma grafica, è la *curva caratteristica* o semplicemente la *caratteristica* dell'elemento non lineare.



Questo caso si presenta frequentemente nella progettazione dei sistemi di controllo, perché nei blocchi che si presumono lineari si hanno in realtà fenomeni di saturazione, di soglia, di isteresi o di gioco che non sempre si possono trascurare. In linea generale si può pensare che tanto maggiore è il livello di potenza a cui opera il singolo sottosistema, tanto più i fenomeni non lineari si fanno evidenti; è dunque lecito aspettarsi che le non linearità siano presenti soprattutto nel processo da controllare.

In taluni casi, poi, si ricorre volutamente a componenti non lineari che presentano vantaggi tecnici o economici; tale è il caso, ad esempio, dei sistemi a relè. Infine è appena il caso di ricordare che anche i più comuni dispositivi elettronici quali il BJT hanno caratteristiche non lineari.

Definizione della funzione descrittiva

Il metodo della funzione descrittiva si applica in modo semplice a non linearità istantanee e simmetriche. Si sceglie come ingresso della non linearità una funzione sinusoidale e si approssima la risposta, che è ovviamente periodica, con una sinusoide della stessa frequenza, di ampiezza e fase opportune, trascurando quindi la distorsione provocata dalla non linearità.

Quest'approssimazione risulta sufficientemente giustificata (ammesso che siano assenti o trascurabili le subarmoniche) quando la non linearità si trova a far parte di un sistema in una posizione tale che gli elementi che la seguono hanno un comportamento del tipo filtro passa-basso; ciò avviene in quasi tutti (ma non tutti) i sistemi di controllo a controreazione. In pratica, questa *ipotesi dell'azione filtrante* può essere accettata quando l'ampiezza di ogni armonica trascurata è inferiore al 10% di quella della prima armonica.

Per definire con precisione la funzione descrittiva si consideri una non linearità istantanea e simmetrica e si ponga per l'ingresso:

$$x(t) = X \sin \omega t$$

La $y(t)$ è periodica con lo stesso periodo di x e dunque si può espandere in serie di Fourier:

$$y(t) = \sum_{n=1}^{\infty} A_n \sin(n\omega t) + \sum_{n=1}^{\infty} B_n \cos(n\omega t)$$

Il valore medio è uguale a zero perché la non linearità è simmetrica. Trascurando le armoniche superiori, si approssima la $y(t)$ con la sola prima armonica:

$$\tilde{y}(t) = A_1 \sin \omega t + B_1 \cos \omega t$$

Al solito si può anche scrivere:

$$\tilde{y}(t) = Y \sin(\omega t + \varphi)$$

con

$$Y = \sqrt{A_1^2 + B_1^2} \quad \varphi = \arctg \frac{B_1}{A_1}$$

I coefficienti dello sviluppo in serie sono, come è noto:

$$A_1 = \frac{1}{\pi} \int_0^{2\pi} y(t) \sin \omega t \cdot d\omega t$$

$$B_1 = \frac{1}{\pi} \int_0^{2\pi} y(t) \cos \omega t \cdot d\omega t$$

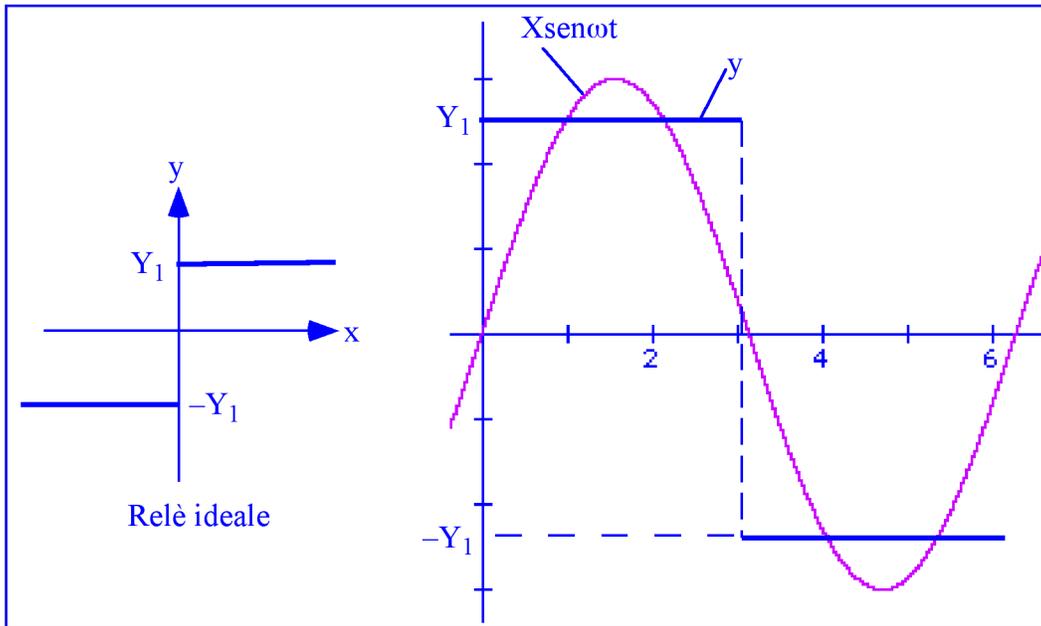
Si può ora definire la funzione descrittiva D da associare alla non linearità attraverso il rapporto delle espressioni simboliche dell'uscita approssimata e dell'ingresso. Si ha:

$$D(X) = \frac{A_1 + jB_1}{X} = \frac{Y e^{j\varphi}}{X}$$

Grazie ad una nota proprietà dello sviluppo in serie di Fourier, la D così definita rende minimo l'errore quadratico medio che si commette nell'approssimare l'uscita con una funzione sinusoidale.

Calcolo della funzione descrittiva

A titolo di esempio si vuole ora calcolare la funzione descrittiva di un relè ideale, la cui caratteristica è mostrata in figura:



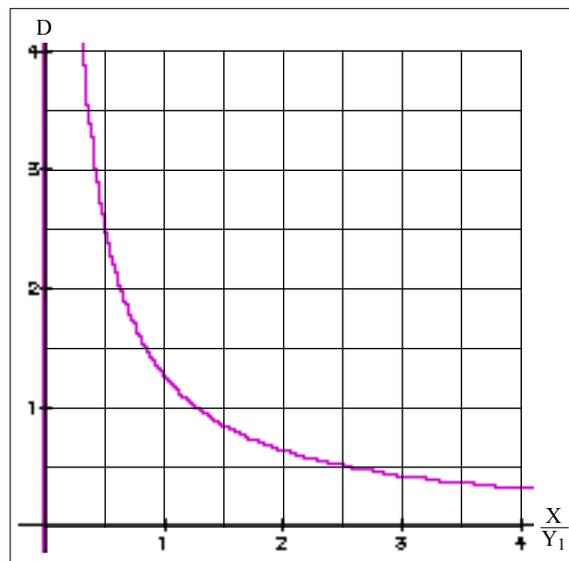
La figura mostra anche che l'uscita della non linearità è un'onda quadra di ampiezza Y_1 . Per trovare la funzione descrittiva, si applica la definizione calcolando i coefficienti A_1 e B_1 ; è chiaro che B_1 è nullo, mentre per A_1 si ha:

$$A_1 = \frac{1}{\pi} \int_0^{2\pi} y(t) \sin \omega t \cdot d\omega t = \frac{2}{\pi} \int_0^{\pi} Y_1 \sin \omega t \cdot d\omega t = \frac{4Y_1}{\pi}$$

Pertanto la funzione descrittiva è:

$$D(X) = \frac{4Y_1}{\pi X}$$

il cui andamento è, ovviamente, un ramo d'iperbole che nel grafico seguente mostra le ascisse normalizzate rispetto a Y_1 .

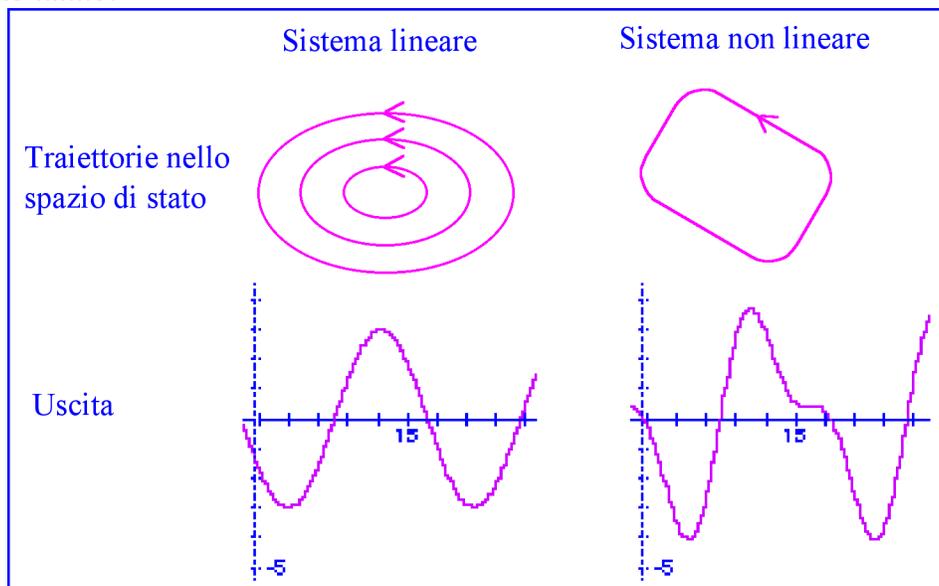


Esistenza dei cicli limite

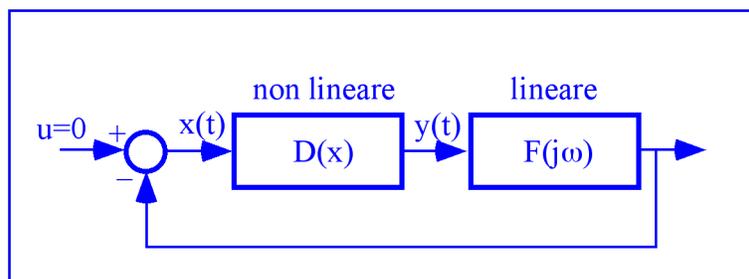
I sistemi di controllo debbono avere proprietà di stabilità adeguate, che nella sintesi classica vengono garantite imponendo dei margini di stabilità, cioè una sufficiente «distanza» dal limite di stabilità. Se per variazioni parametriche o altre cause un sistema lineare raggiunge il limite di stabilità, si presentano modi naturali costanti (autovalore nullo) o periodici (autovalori immaginari). Quest'ultimo caso è il più frequente, e quindi si suole dire che «un sistema al limite di stabilità oscilla».

In tali condizioni, l'oscillazione di un sistema lineare è sinusoidale con ampiezza che dipende dalle condizioni iniziali e può assumere qualsiasi valore. Se, per un'ulteriore variazione parametrica, gli autovalori immaginari assumono una pur piccolissima parte reale positiva, l'oscillazione diverge e tende all'infinito.

I sistemi non lineari che stiamo esaminando presentano caratteristiche diverse. Le oscillazioni non hanno in genere forma sinusoidale e contengono armoniche superiori; la presenza di saturazioni ne limita l'ampiezza, per cui un sistema non lineare instabile può assestarsi su oscillazioni periodiche ma non sinusoidali, di ampiezza costante e talora indipendente dalle condizioni iniziali, corrispondenti a curve chiuse nello spazio di stato. Tale fenomeno prende il nome di *ciclo limite*.



Per scoprire se un sistema non lineare ammette cicli limite in evoluzione libera, si consideri il sistema in figura, al quale si può sempre ricondurre un sistema con una non linearità in catena diretta:



Se esiste un ciclo limite, l'uscita y del blocco non lineare è periodica; sotto l'ipotesi dell'azione filtrante, il segnale x che esce dal blocco lineare può essere considerato sinusoidale e quindi l'ingresso della non linearità è:

$$x(t) = X \sin \omega t$$

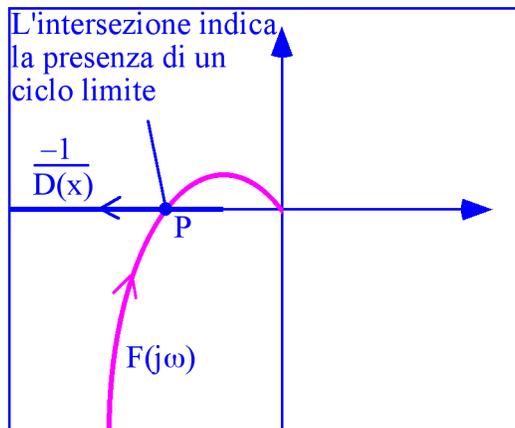
In forma simbolica questo segnale, moltiplicato per $D(x)$ e per $F(j\omega)$ e cambiato di segno dal sommatore, deve ripresentarsi identico all'ingresso della non linearità; si deve dunque avere

$$-D(X) \cdot F(j\omega) = 1$$

Questa equazione esprime dunque, nei limiti dell'approssimazione propria della funzione descrittiva, la condizione di esistenza di un ciclo limite. Essa viene riscritta nella forma

$$F(j\omega) = -\frac{1}{D(X)}$$

A questa scrittura si dà un'immediata interpretazione grafica: l'esistenza di un ciclo limite corrisponde sul piano di Nyquist all'intersezione P tra le curve che rappresentano la $F(j\omega)$ e l'inverso della $D(X)$ cambiato di segno. L'ampiezza della prima armonica del ciclo limite, detta per brevità ampiezza del ciclo limite, è il valore di X per cui si ha l'intersezione e si legge su quest'ultima curva; la pulsazione del ciclo limite è il valore di ω per cui si ha l'intersezione e si legge sul grafico della F .



Le due curve sono chiamate, per evidenti motivi, *luogo delle ampiezze* e *luogo delle frequenze*; un ciclo limite è presente se il luogo delle ampiezze interseca il luogo delle frequenze.

Stabilità dei cicli limite

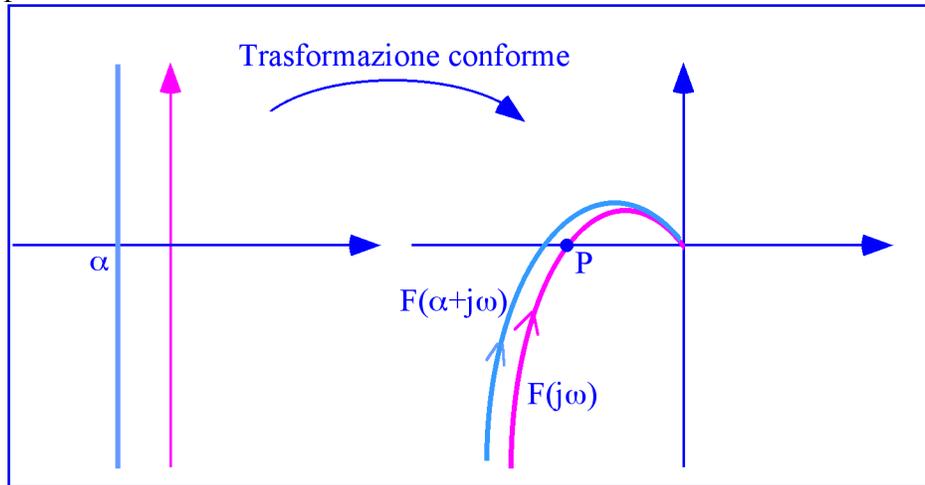
Per appurare se un ciclo limite è stabile, si può cercare di capire che cosa accade in presenza di una piccola perturbazione. Se l'ampiezza del ciclo limite tende lentamente a variare, al blocco lineare viene applicato un segnale periodico di ampiezza lentamente variabile che, nello spirito delle approssimazioni fatte sinora, può essere considerato quasi sinusoidale e scritto almeno qualitativamente nella forma:

$$y(t) \cong e^{\alpha t} \sin \omega t$$

dove α è una costante molto piccola, positiva se il segnale tende a crescere e negativa in caso opposto.

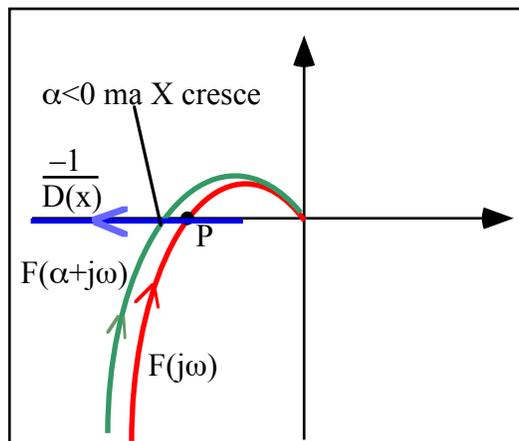
A tale segnale il blocco lineare dà una risposta a regime permanente individuata dal modulo e dalla fase della funzione $F(\alpha + j\omega)$. Come è noto, la F è una trasformazione conforme dal piano s al piano di Nyquist e pertanto mantiene le orientazioni. Se allora si suppone $\alpha < 0$, la retta $\text{Re}(s) = \alpha$

percorsa quando ω varia da $-\infty$ a $+\infty$ lascia alla sua destra l'asse immaginario; corrispondentemente, la funzione $F(\alpha+j\omega)$ deve lasciare alla sua destra la $F(j\omega)$. Analogo ragionamento vale ovviamente per $\alpha>0$.

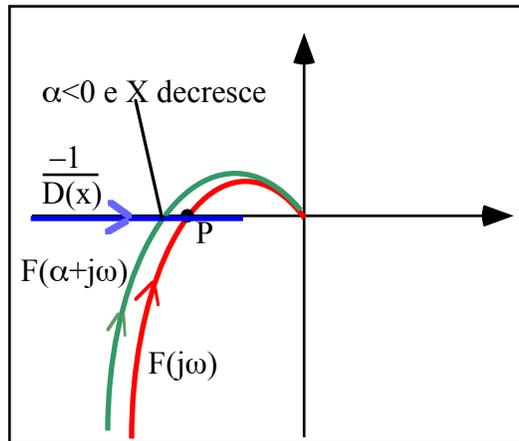


Osservando ora il punto d'intersezione P che individua il ciclo limite, si possono presentare i due casi seguenti.

1. Per $\alpha<0$, il punto P si sposta su un valore X maggiore. In questo caso ci si trova di fronte ad una contraddizione: partendo dall'ipotesi che l'ampiezza del ciclo limite sia diminuita, si conclude che invece è aumentata. Dunque l'ipotesi è falsa e l'ampiezza non può essere diminuita; poiché ragionando in modo simmetrico si dimostra che tale ampiezza non può neppure essere aumentata, si conclude che il ciclo limite è *stabile*.

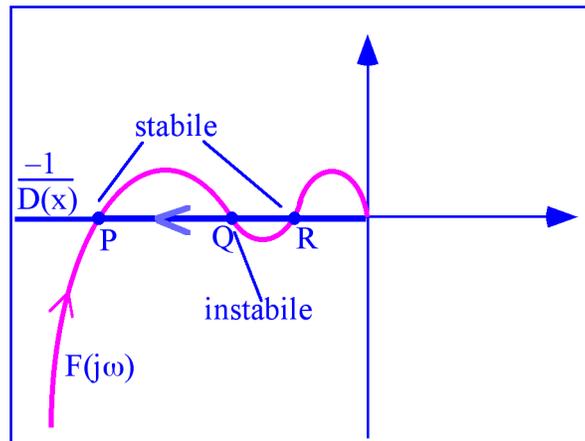


2. Per $\alpha<0$, il punto P si sposta su un valore X minore. In questo caso ci si trova di fronte ad una conferma: partendo dall'ipotesi che l'ampiezza del ciclo limite sia diminuita, si conclude che effettivamente è diminuita. Poiché ragionando in modo simmetrico si dimostra che tale ampiezza può anche essere aumentata, si conclude che il ciclo limite è *instabile*, nel senso che una perturbazione porta il sistema ad allontanarsene.



Da questo esame si trae allora il seguente criterio: un ciclo limite è stabile se, percorrendo la F nel verso delle ω crescenti, all'intersezione il luogo delle ampiezze appare orientato in senso antiorario quando è percorso nel verso delle X crescenti.

A titolo di esempio, la figura mostra il caso di un relè ideale con un luogo delle frequenze che determina tre intersezioni P , Q , R e quindi tre cicli limite aventi ampiezze e frequenze diverse; quelli corrispondenti ai punti P ed R sono stabili, mentre quello corrispondente a Q è instabile. Se dunque il sistema dovesse trovarsi sul ciclo Q , una minima perturbazione potrebbe portarlo in P o in R .



Occorre tuttavia ricordare che il ragionamento fatto è puramente euristico e quindi, in caso di dubbio, il risultato va verificato con una simulazione.

Regolatori autosintonizzanti

Il relè ideale ha un luogo delle ampiezze che coincide con il semiasse reale negativo; pertanto si avrà un ciclo limite con qualunque sistema lineare la cui fase scenda sotto $-\pi$.

Questa ipotesi è verificata nella maggior parte dei processi da controllare; pertanto chiudendo un processo in un anello di reazione con un relè ideale si otterrà un'oscillazione.

Da questa osservazione nasce l'idea dei regolatori ad autosintonia. Il regolatore comprende al suo interno un relè quasi ideale e un deviatore che permette di inserire in ciclo il relè o la funzione di regolazione. Inserendo il relè, s'innescano un'oscillazione di cui vengono misurati la pulsazione ω^* e l'ampiezza K^* in ingresso al relè.

Questa oscillazione permette di ricavare i valori di T_{cr} e K_{cr} richiesti da Ziegler e Nichols per la taratura dei regolatori. Tale metodo, come è noto, consiste nell'aumentare il guadagno della catena diretta fino a innescare un'oscillazione e misurare il valore di tale guadagno K_{cr} e il valore del periodo T_{cr} .

Nel sistema che qui si esamina, il punto d'intersezione tra i luoghi delle ampiezze e delle frequenze è caratterizzato dal fatto che la fase di $P(j\omega)$ vale $-\pi$; pertanto la pulsazione ω^* dà proprio il periodo critico richiesto da Ziegler e Nichols per la taratura dei regolatori. Sempre nel punto d'intersezione risulta inoltre:

$$|G(j\omega^*)| = \frac{\pi K^*}{4Y_1}$$

da cui

$$\frac{1}{|G(j\omega^*)|} = \frac{4Y_1}{\pi K^*}$$

Ma questo è proprio il guadagno critico necessario a portare il sistema all'oscillazione secondo la tecnica di Ziegler e Nichols; si sono così ottenuti tutti i parametri richiesti:

$$T_{cr} = \frac{2\pi}{\omega^*} \quad K_{cr} = \frac{4Y_1}{\pi K^*}$$

A questo punto, ricordando che il regolatore PID ha funzione di trasferimento

$$G(s) = K_p \left(1 + \frac{1}{\tau_I} s + \tau_D s \right)$$

i valori dei parametri da tarare si ottengono dalla seguente tabella:

regolatore	K_p	τ_I	τ_D
P	$0,5K_{cr}$		
PI	$0,45K_{cr}$	$0,8T_{cr}$	
PID	$0,6K_{cr}$	$0,5T_{cr}$	$0,125T_{cr}$

Questo modo di procedere presenta due vantaggi:

- 1 l'ampiezza dell'oscillazione è direttamente proporzionale a Y_1 e quindi può essere resa piccola quanto si vuole alimentando opportunamente il relè. Con il metodo tradizionale, invece, si porta il processo a un'oscillazione di cui non è possibile prevedere e controllare l'ampiezza; ciò può essere pericoloso o anche inaccettabile.
- 2 il metodo può essere facilmente automatizzato: il regolatore, controllato da un microprocessore, inserisce il relè, rileva i valori di ω^* e di K^* e predispose i suoi

parametri automaticamente. Nasce così l'idea di un *regolatore ad autosintonia* (self tuning); oggi molti regolatori commerciali sono di questo tipo, anche se a volte usano algoritmi più complessi di quello qui descritto.



Riferimenti

- A. Ruberti: *Introduzione allo studio dei sistemi di controllo a controreazione non lineari*, CNR 1959
- G. Marro: *Controlli automatici*, quarta edizione, Zanichelli, in particolare il cap. 7
- P. Bolzern, R. Scattolini, N. Schiavoni: *Fondamenti di controlli automatici*, McGraw-Hill, in particolare il cap. 16 da cui è tratto il secondo esempio
- P. Atherton: *The describing function method*, in, *The Control Handbook*, W.S. Levine Editor, CRC-IEEE Press, cap. 19