Applicazione di un algoritmo PID su sistemi a microcontrollore.

ing. Aldo OMICCIOLI

revisione del 30-04-2003

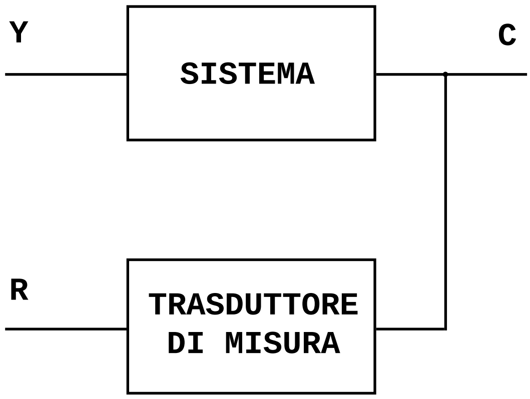
Introduzione

Il presente documento illustra in modo semplice ed intuitivo alcuni concetti di base dei sistemi di regolazione ad anello chiuso dotati di rete correttrice di tipo PID. Particolare riferimento sarà fatto ai sistemi di termoregolazione, ma i concetti riportati sono comunque validi anche per il controllo di grandezze fisiche diverse dalla temperatura.

Lo scopo principale del documento è quello di fornire gli strumenti necessari allo sviluppo software di un controllo PID su sistemi a microprocessore. Inoltre sono fornite le indicazioni necessarie alla determinazione software dei parametri caratteristici (costanti P, I, D) di lavoro di un sistema reale. E' noto infatti che i parametri in questione caratterizzano lo specifico sistema controllato e non sono sempre facilmente individuabili. Varie tecniche sono state sviluppate a questo proposito, molte delle quali richiedono costosi strumenti di misura e registrazione nonché molto tempo e pazienza da parte del progettista. Demandare il calcolo dei parametri allo stesso sistema di gestione del controllo PID permette di scavalcare completamente tutte queste difficoltà e di poter ricalcolare automaticamente, in campo, i loro valori ogni qual volta lo si ritiene opportuno (per esempio se le caratteristiche fisiche del sistema sono per un qualche motivo variate). L'algoritmo di autocalcolo costituisce anche un valido strumento per caratterizzare il sistema controllato, determinandone i parametri tipici direttamente nelle unità di misura opportune.

La regolazione PID

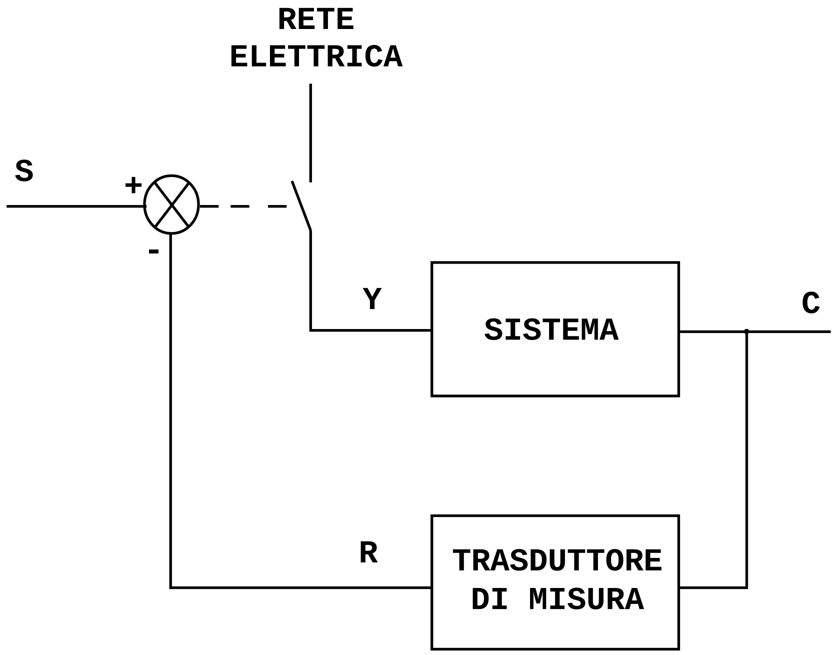
Consideriamo un generico SISTEMA del quale ci interessa controllare un suo particolare stato fisico. La VARIABILE CONTROLLATA del sistema, indicata con C, verrà misurata mediante un'opportuno TRASDUTTORE DI MISURA, il quale consentirà di conoscere istante per istante il VALORE REALE della grandezza in questione, indicato con R. Ovviamente per costringere il sistema a variare a nostro piacimento la variabile controllata C, occorre fornire al sistema stesso un input che chiamiamo GRANDEZZA MANIPOLABILE ed indichiamo con Y:



Per fare un esempio concreto, supponiamo che il sistema sia un forno realizzato mediante una resistenza elettrica di potenza; la grandezza C che vogliamo controllare sia la temperatura del forno stesso. Utilizziamo quindi un termometro elettronico capace di fornire una tensione di uscita proporzionale alla temperatura misurata nel forno; questa informazione di tensione ci permetterà di capire se la temperatura è pari al valore che ci proponiamo. Nel caso in questione, il sistema è costituito dal forno, la grandezza controllata C è la sua temperatura, il valore reale R la tensione fornita dal termometro e la grandezza manipolabile Y la potenza elettrica che forniamo alla resistenza del forno stesso.

Supponiamo ora di voler controllare il sistema affinché la sua temperatura rimanga nell'intorno di un valore voluto, che chiamiamo SET POINT ed indichiamo con S; questo valore lo esprimeremo con il valore di una tensione elettrica così come per il valore della temperatura misurata dal termometro elettronico. In questo modo sarà possibile confrontare le due grandezze omogenee del valore reale R e del set point S e stabilire se la temperatura del forno è più o meno vicina a quella voluta.

Un metodo molto semplice ed economico per realizzare tale controllo è quello di comparare il valore reale R con il valore del set point S e di interrompere l'erogazione di energia al forno nel caso che la temperatura reale superi il valore del set point noi impostato. Questo è quello che succede con un dispositivo molto utilizzato nel controllo della temperatura detto TERMOSTATO. Ricollegandosi allo schema precedente, un sistema di regolazione di questo tipo si può così schematizzare:



Il blocco componente indicato con un cerchio barrato individua un particolare circuito di comparazione delle due tensioni corrispondenti al set point S ed al valore reale R. Tale comparatore sarà realizzato in modo da aprire il contatto di potenza alla resistenza riscaldatrice del forno ogni qual volta il valore di tensione R supererà il valore di tensione S. In questo caso, in mancanza di energia fornita, il sistema tenderà a raffreddarsi naturalmente, disperdendo il suo calore nell'ambiente circostante; conseguentemente il valore reale R, corrispondente alla temperatura, comincerà a scendere. Quando il valore di R diventerà inferiore al quello di S, il comparatore commuterà di nuovo il suo stato fornendo l'alimentazione alla resistenza del forno. Il processo appena descritto continuerà all'infinito con un periodo di ripetizione dipendente dal sistema controllato (in particolare dalla sua inerzia termica) ed anche dalla prontezza di risposta del trasduttore di misura e del comparatore. Proprio per il suo modo caratteristico di collegare l'intera potenza al sistema e di interromperla, tale regolazione viene chiamata di tipo "ON-OFF".

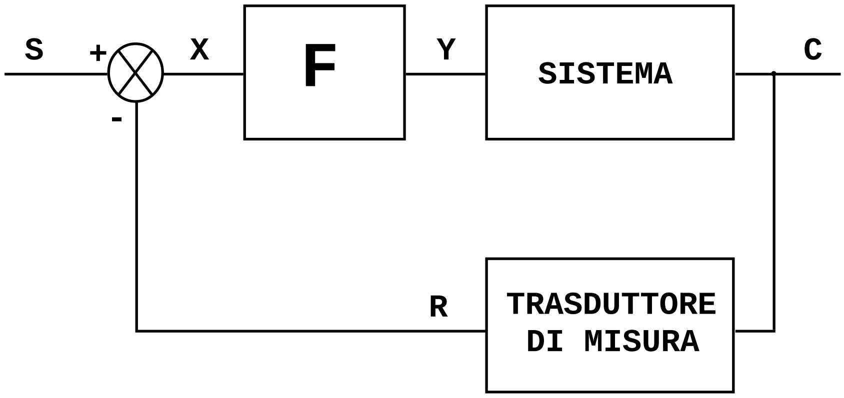
Un sistema di regolazione del genere è molto semplice da realizzare ed in certi casi permette di ottenere risultati soddisfacenti. Molti sistemi tuttavia hanno funzioni di risposta della grandezza controllata C, rispetto alla grandezza manipolabile Y, tali da non permettere al valore reale R di rimanere sufficientemente prossimo al set point S. Nel caso di un forno elettrico la difficoltà nella regolazione di temperatura può aumentare, per esempio, se la potenza elettrica della resistenza è eccessiva per la capacità termica del forno; se supponiamo anche che la sonda di temperatura, utilizzata per la misura del valore reale R, sia collocata in una posizione lontana dalla resistenza di riscaldamento, si intuisce che sarà molto difficile evitare, con un sistema di regolazione ON-OFF, che la temperatura superi eccessivamente il set point. Questo perché, quando la resistenza è alimentata, la temperatura del forno cresce molto velocemente mentre il termometro si accorge in ritardo di tale aumento e quindi il comparatore non è in grado di interrompere prontamente l'erogazione di energia.

Un sistema di regolazione ON-OFF ha dunque lo svantaggio di controllare il sistema potendo agire solo su due possibilità, corrispondenti all'applicazione di tutta la potenza oppure allo spegnimento totale dell'energia. In più si aggiunge il fatto che una variazione tra questi due stati è possibile solo in corrispondenza del raggiungimento del valore reale R di una prefissata soglia.

Per rimuovere tali limitazioni, occorre dunque sostituire la grandezza manipolabile Y con una grandezza non più digitale (ON-OFF) ma analogica, capace di assumere un certo insieme di valori intermedi alle due situazioni estreme. Si potrà quindi fornire al sistema una potenza anche intermedia, come per esempio il 25% oppure il 70% e così via; in questo modo si avrà un maggior grado di libertà nel controllare la potenza fornita al sistema.

Anche le variabili di set point e di valore reale andranno trattate in modo diverso: il comparatore non fornirà più alla sua uscita un valore a soli due stati (ON-OFF) dipendente dal fatto che una delle grandezze supera l'altra, ma un valore analogico contenente anche l'informazione di quanto le due grandezze differiscono. La grandezza analogica che si viene così ad introdurre verrà chiamata ERRORE e sarà nel seguito indicata con X; più precisamente la variabile X rappresenterà, istante per istante, la differenza tra le due grandezze omogenee di set point S e di valore reale misurato R.

A questo punto viene naturale introdurre un blocco funzionale F nel sistema di controllo capace di generare un opportuno valore della grandezza manipolabile Y in funzione esclusivamente del valore dell'errore X:



Il blocco funzionale F ha l'incarico di pilotare il sistema mediante la sua uscita Y (grandezza manipolabile) avendo come unica informazione il valore della variabile errore X. La funzione F può avere diverse forme realizzative, corrispondenti a specifiche rappresentazioni matematiche; la più semplice funzione matematica che ci può venire in mente è quella lineare:

$$Y = F(X) = K_1 X$$

Un blocco funzionale del genere permette di fornire potenza al sistema in modo proporzionale al valore corrente dell'errore; questo significa che se il valore reale misurato si discosta molto dal valore del set point desiderato, la potenza fornita al sistema è prossima ad uno dei due valori limite (0% oppure 100%) a seconda del segno della variabile X. Più precisamente, nel dimensionare le grandezze in gioco nel sistema fisico reale, si stabilisce una banda attorno al set point, corrispondente alla variazione analogica della potenza fornita al sistema; ciò equivale all'inserimento a valle del blocco F di un blocco limitatore capace di bloccare il valore di Y ai valori limiti permessi.

Un sistema di controllo in retroazione, utilizzante un blocco F di questo tipo, viene chiamato "regolatore di tipo P" poiché il suo comportamento è basato su un'azione di tipo proporzionale. Un regolatore di tipo P lo si impiega in quei sistemi che permettono un'elevata costante di guadagno d'anello senza incorrere in problemi di stabilità.

La funzione F non deve necessariamente determinare il valore corrente della grandezza manipolabile Y basandosi esclusivamente sul valore allo stesso istante della variabile errore X. Infatti notevoli miglioramenti nelle prestazioni del controllo si ottengono utilizzando funzioni F che tengano conto anche dei valori assunti dall'errore in precedenza.

Basti pensare che, conoscendo l'andamento nel tempo della variabile errore, è possibile stabilire la velocità della variabile stessa corrispondente ad un'informazione sulle capacità del sistema controllato di rispondere alle sollecitazioni della grandezza manipolabile Y. Ciò consente, in un certo senso, di prevedere la reazione del sistema al controllo fornito mediante la grandezza Y e di agire di conseguenza.

In particolare alla funzione F si potrebbe aggiungere un ulteriore termine costituito da una funzione integrale della variabile errore:

$$Y = F(X) = K_1 X + K_2 \int X \, dt$$

Una regolazione con tale funzione F viene chiamata "regolazione di tipo PI", in quanto la sua azione ha un effetto di tipo integrativo oltre che proporzionale; questa funzione tiene memoria della storia passata della variabile X, fornendo un'azione correttiva capace di aggiungere o togliere la potenza in modo graduale al fine di far avvicinare il valore reale R al set point S.

Infine ulteriori possibilità sono offerte da una funzione F comprendente anche un termine derivativo:

$$Y = F(X) = K_1 X + K_2 \int X \, dt + K_3 \frac{dX}{dt}$$

L'azione derivativa, indicata generalmente con D, permette di ottenere maggior prontezza di risposta nel controllo del sistema, in quanto la correzione è tanto più elevata quanto rapida è la variazione dell'errore.

La funzione alla quale siamo giunti prende il nome di funzione PID in quanto comprende tutti e tre gli effetti correttivi. La funzione PID è certamente la funzione correttrice più utilizzata nei sistemi di controllo ad anello chiuso; essa infatti, per la sua generalità, permette di adattarsi a gran parte dei sistemi da controllare. Variando il peso dei tre termini è possibile configurare il sistema di controllo alle più svariate esigenze ed è per questo che un regolatore di tipo PID è anche chiamato regolatore STANDARD.

Implementazione software della funzione PID

La realizzazione pratica di un sistema di regolazione ad anello chiuso con rete correttrice di tipo PID richiede normalmente una componentistica di tipo analogico; per esempio un effetto di tipo integrativo lo si ottiene inserendo nell'anello una rete ritardatrice costituita da un partitore RC (R in serie e C verso massa) capace di integrare nel tempo il valore dell'errore. Un effetto derivativo richiede invece una rete anticipatrice costituita da un partitore RC (C in serie e R verso massa) capace di lasciar passare le rapide variazioni dell'errore.

Disponendo di un sistema a microprocessore, capace di elaborare solo grandezze discrete, si richiede la conseguente discretizzazione della funzione PID; la funzione analogica viene aggiornata in modo digitalizzato e quindi simulata tramite software.

Derivando rispetto al tempo ambo i membri della precedente equazione, si ottiene la seguente forma discreta utilizzata per il calcolo della funzione PID:

$$\frac{dY}{dt} = K_1 \frac{dX}{dt} + K_2 X + K_3 \frac{d^2X}{dt^2}$$

$$\frac{Y_n - Y_{n-1}}{T} = K_1 \frac{X_n - X_{n-1}}{T} + K_2 X_n + K_3 \frac{(X_n - X_{n-1}) - (X_{n-1} - X_{n-2})}{T^2} \quad \text{da cui}$$

$$Y_n = Y_{n-1} + P(X_n - X_{n-1}) + IX_n + D(X_n - 2X_{n-1} + X_{n-2})$$

$$\text{dove} \quad P = K_1 \qquad I = K_2 T \qquad D = \frac{K_3}{T}$$

Essendo una funzione discretizzata, è stato introdotto il tempo T di aggiornamento dell'equazione; questo è l'intervallo che intercorre tra due successivi aggiornamenti del valore dell'uscita Y della funzione. Conseguentemente i valori delle variabili X ed Y sono stati considerati solo in tali istanti di campionamento e per questo indicati con n, n-1, n-2; il valore n si riferisce all'ultimo campionamento effettuato, n-1 al penultimo ed n-2 al terzultimo.

L'utilizzo di tale formula discretizzata consente di approssimare con sufficiente precisione l'equazione matematica continua riportata nel precedente paragrafo; è importante sottolineare che l'accuratezza della funzione è tanto più elevata quanto più i campionamenti sono frequenti e quindi quanto più il tempo T di ripetizione dell'aggiornamento è piccolo. La discretizzazione si traduce infatti in un'approssimazione a gradini dell'andamento reale della variabile continua analogica; per la precisione, il valore medio della variabile discretizzata a gradini è ritardato di T/2 rispetto al vero andamento della variabile continua.

Una limitazione al desiderio di abbassare il tempo T, con lo scopo di ottenere le migliori prestazioni possibili, è dovuta all'impegno che l'aggiornamento dell'equazione richiede al microprocessore, in quanto essa richiede una certa mole di calcoli matematici. Si tratta quindi di stabilire un tempo T tale da non costringere il microprocessore ad aggiornare l'equazione più frequentemente di quanto serva (in funzione della più o meno lentezza di variazione della variabile di ingresso X) ma, al tempo stesso, capace di far reagire prontamente la variabile Y alla variazione del segnale di errore X.

Il valore di tempo T, assieme ai tre parametri caratteristici P, I, D della funzione, costituiscono le uniche costanti da determinare per definire completamente l'algoritmo di calcolo: il valore di questi parametri dipende fortemente dal sistema controllato e esprime le caratteristiche fisiche proprie dello stesso.

I parametri PID si possono ridefinire diversamente in funzione di grandezze più adatte ad individuare le caratteristiche fisiche di un sistema:

$$P = K \qquad I = K \frac{T}{T_I} \qquad D = K \frac{T_D}{T}$$

Dalle formule precedenti si deduce che per tale ridefinizione si è considerato:

$$K = K_1 \qquad T_I = \frac{K_1}{K_2} \qquad T_D = \frac{K_3}{K_1}$$

Il valore K si dice costante di proporzionalità o guadagno proporzionale, il tempo T_I tempo integrativo ed il tempo T_D tempo derivativo. la costante K è espressa nel rapporto delle unità di misura di Y su X, mentre i tempi T_I e T_D sono normalmente espressi in secondi così come il tempo T di ripetizione dell'aggiornamento.

Con le nuove grandezze definite la formula per il calcolo discretizzato della funzione PID può essere così riscritta:

$$Y_n = Y_{n-1} + K(X_n - X_{n-1}) + \frac{KT}{T_I} X_n + \frac{KT_D}{T} (X_n - 2X_{n-1} + X_{n-2})$$

Si consideri che su un sistema a microprocessore può essere vantaggioso utilizzare unità di misura diverse da quelle più comuni allo scopo di eseguire operazioni solo su grandezze di tipo intero. Per esempio le costanti di tempo possono essere espresse in decimi di secondo, mentre la costante K può essere premoltiplicata per 1000 per considerarne tre cifre decimali pur utilizzando un valore intero. In tal caso occorrerà dividere tutta la parte incrementale (a destra del termine Y_{n-1}) della formula per lo stesso fattore 1000 per aggiustarla alla nuova unità di misura di K. Si noti che l'utilizzo di unità di misura delle costanti di tempo diverse ma tra loro omogenee non richiede aggiustaggi della formula in quanto i termini contengono sempre rapporti di valori temporali. Infine per quanto riguarda i valori analogici di ingresso X e di uscita Y della formula si possono, con opportune scelte, considerare ancora solo numeri interi. Per esempio se la lettura della temperatura reale R è effettuata mediante un convertitore analogico/digitale a 8 bits dove lo stadio di amplificazione e offset del sensore sono stati tarati per fornire il valore 0-255 in corrispondenza a 0-100 gradi ed inoltre il set point S è impostabile da 0 a 100 gradi mediante il valore 0-255 di un secondo ingresso analogico connesso ad un potenziometro, non occorre convertire in gradi la lettura dei convertitori ma è possibile considerare le temperature espresse in questa nuova unità di misura intera. Analogamente l'uscita Y normalmente corrisponde al valore 0-255 di una uscita analogica del microprocessore o di un circuito PWM a 8 bits.

I valori delle costanti considerate sono funzione del sistema controllato e quindi non definibili a priori; queste costanti costituiscono spesso un notevole ostacolo a chi si accinge a sviluppare un controllo di tipo PID. Esistono infatti diversi metodi per il loro calcolo, alcuni strettamente empirici, altri più teorici basati su alcune formule che richiedono tuttavia la determinazione di dati di partenza mediante metodi sperimentali.

Un metodo frequentemente utilizzato è quello di misurare la risposta del sistema ad anello aperto in seguito ad una sollecitazione a gradino della variabile manipolabile. Tracciando un grafico dell'andamento temporale della grandezza controllata dopo l'applicazione del gradino è possibile determinare il tempo di risposta del sistema e da tale valore è possibile dedurre i parametri ottimali del PID. Il metodo della misurazione sperimentale della funzione di risposta al gradino risulta abbastanza laboriosa e richiede una strumentazione adeguata.

Un metodo alternativo, dal quale discende l'algoritmo utilizzato per l'autocalcolo dei parametri, è quello della determinazione della "banda proporzionale di pendolazione". Abbiamo visto che, in assenza delle azioni integrale e derivativa, l'equazione PID si riduce a:

$$Y = K X$$

dove la costante K è più propriamente detta "sensibilità proporzionale". Si definisce "banda proporzionale" l'inversa della sensibilità proporzionale e corrisponde alla variazione del valore reale R necessaria per provocare una variazione unitaria della variabile manipolabile Y.

Il metodo della banda proporzionale di pendolazione consiste nel porre il sistema retroazionato in una condizione di stabilità limite, corrispondente ad una situazione di oscillazione permanente: per far ciò occorre momentaneamente escludere dall'anello le azioni integrale e derivativa. In tale situazione di oscillazione permanente il valore reale R oscilla in modo periodico attorno al valore del set point S, con un'ampiezza ed un periodo dipendenti dalle caratteristiche fisiche del sistema; misurando il valore della banda proporzionale $1/K_0$ del sistema così ottenuto e del periodo T_0 di oscillazione è possibile calcolare il parametri ottimali della funzione PID applicando le seguenti formule di Ziegler e Nichols:

$$K = 0.66\ K_0 \qquad T_I = 0.5\ T_0 \qquad T_D = 0.125\ T_0$$

Vediamo ora come applicare in pratica tale metodo e in particolare come sviluppare un algoritmo capace di calcolare in modo del tutto automatico tali parametri.

Per portare il sistema in una condizione di oscillazione permanente nella quale l'azione integrale e derivativa non hanno effetto, occorre considerare una forte prevalenza del fattore proporzionale: una simile situazione la si ha per un sistema di controllo in retroazione del tipo ON-OFF. Infatti se si considera un fattore di sensibilità K elevato, delle piccole entità di errore X provocano degli elevati valori della grandezza manipolabile Y, con conseguente saturazione ai due valori limite estremi; questi due valori limite, nel caso in questione della termoregolazione mediante potenza elettrica applicata ad una resistenza, corrispondono alla situazione di potenza 100% e di potenza 0% e quindi riconducibile ad un ON-OFF.

Occorre dunque impostare una regolazione di tipo ON-OFF, attendere il termine del primo ciclo di oscillazione attorno al set point per la stabilizzazione dell'oscillazione ed infine calcolare il corrispondente valore di banda proporzionale $1/K_0$ e di periodo T_0 di oscillazione. Applicando le precedenti formule si ricavano i parametri PID. Infine sostituiamo la regolazione di tipo ON-OFF con la regolazione di tipo PID utilizzante i parametri così ricavati.

La banda proporzionale, secondo la sua definizione, corrisponde alla variazione del valore reale R necessaria per provocare una variazione unitaria della variabile manipolabile Y. Nel caso della regolazione ON-OFF abbiamo una variazione della grandezza manipolabile Y tra i suoi due estremi (0% e 100%) e quindi pari al valore di fondo scala (indicato con F.S.); inoltre se misuriamo la variazione complessiva, indicata con A, del valore reale R in un ciclo di pendolazione permanente (valore picco-picco dell'oscillazione), possiamo calcolare la banda proporzionale di pendolazione mediante la sua stessa definizione:

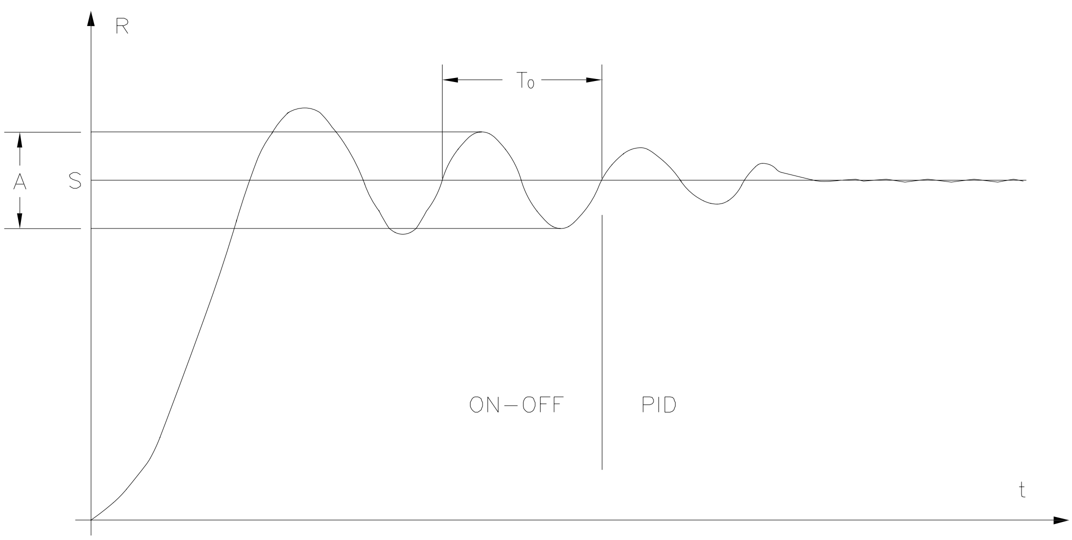
$$1/K_0 = \frac{A}{F.S.}$$

Per quanto riguarda la scelta del tempo T di aggiornamento dell'equazione in genere si utilizza un valore pari a 1/20 del tempo T_0 allo scopo di consentire una pronta reazione della funzione PID alle variazioni delle condizioni fisiche del sistema controllato. Si noti che dei valori di T più bassi del necessario portano ad un elevato valore della costante derivativa D con probabili saturazioni della variabile Y. Valori di T più alti del necessario portano invece a bassi valori della costante integrativa I e nel caso si eseguano calcoli solo tra interi si può incorrere in troncamenti dei decimali tali da annullare tale azione correttiva anche in presenza di una costante K premoltiplicata per per una certa potenza del 10.

E' importante infine far seguire il blocco funzione PID da un blocco di limitazione del valore Y tra un valore Y_{min} ed un valore Y_{max} corrispondenti al range di variazione utile ed effettiva dell'uscita. Infatti, a causa di improvvise e consistenti variazioni del setpoint, la formula può produrre valori di Y_n valori molto al di fuori dei limiti suddetti e che non comportano ulteriori variazioni della variabile controllata. La variabile Y_n potrà, a causa della stessa natura della formula PID, impiegare molto tempo per ritornare entro i limiti consentiti con conseguente permanenza dell'uscita in uno stato di saturazione.

Con l'introduzione dei suddetti limiti dell'uscita Y, il precedente valore di fondo scala F.S. viene a corrispondere alla differenza (Y_{max} - Y_{min}). Da ciò segue che il valore di K_0 può essere ottenuto dal rapporto tra questa differenza ed il valore picco-picco della variabile d'errore X.

Di seguito è riportato un tipico andamento nel tempo del valore reale R durante la fase preliminare di regolazione ON-OFF (nella quale vengono calcolati i parametri PID) e successivamente durante la fase di regolazione PID.



Determinazione dei parametri del sistema dalla regolazione ON-OFF

Le regolazioni P e PI

La funzione di regolazione ad anello chiuso comprendente i 3 termini (proporzionale, integrativo, derivativo) rappresenta un regolatore dalle caratteristiche universali. Tuttavia certi sistemi possono essere ben controllati anche tramite forme più semplificate della funzione. Infatti, procedendo in modo analogo a quello già descritto, possono essere definiti i regolatori di tipo P (dotati del solo termine proporzionale) ed i regolatori di tipo PI (dotati dei termini proporzionale ed integrativo).

Le relative equazioni discretizzate si ricavano immediatamente da quella del regolatore PID eliminando i termini non previsti. Per quanto riguarda il calcolo dei parametri si possono applicare le adeguate formule di Ziegler e Nichols:

REGOLATORE P: $K = 0.5\ K_0$

REGOLATORE PI: $K = 0.45\ K_0$ $T_I = 0.833\ T_0$

Per quanto riguarda il tempo T di aggiornamento dell'equazione discretizzata valgono le stesse considerazioni del paragrafo precedente. Un regolatore di tipo PD è invece poco conveniente; ciò si deduce dall'analisi, condotta nel prossimo paragrafo, del significato fisico delle azioni svolte dai 3 possibili termini.

Considerazioni pratiche sui parametri PID

Il termine proporzionale P corregge il valore dell'uscita Y aggiungendo o sottraendo ad esso, ad ogni ciclo di aggiornamento della funzione (tempo T), una quantità pari a K volte la variazione dell'errore rispetto al ciclo precedente, a seconda che tale variazione sia rispettivamente positiva o negativa. Per esempio in caso di temperatura misurata R costante e set point S fisso risulta un errore X costante: in tal caso non viene apportata alcuna azione correttiva e così anche la Y rimane costante. Se ad un certo punto la temperatura comincia a diminuire ecco che si crea una variazione dell'errore con un relativo incremento a scalini (di ampiezza K volte la variazione) dell'uscita Y. Se a questo punto la temperatura si ristabilizza costante, di nuovo terminano i contributi di tale termine. Infine se la temperatura riaumenta verranno invece tolte delle quantità (di ampiezza K volte la variazione) all'uscita Y ad ogni tempo di aggiornamento T. Per questo motivo una funzione solo con il termine proporzionale può soffrire di difficile convergenza del valore reale R verso il set point S, ossia portare ad errori a regime costanti proprio perchè, come già detto, errori costanti non causano azioni correttive.

Un rimedio all'errore a regime è invece fornito dall'azione del termine integrativo. Infatti il termine integrativo I corregge il valore dell'uscita Y aggiungendo o sottraendo a questa (a seconda del segno di X), ad ogni ciclo di aggiornamento della funzione (tempo T), una quantità pari a $(K * T) / T_I$ volte il valore corrente dell'errore X in tale ciclo.

Si intuisce come un errore costante presente per tutto il tempo integrativo T_I comporta l'applicazione continua e progressiva dei suddetti piccoli incrementi portando dopo il tempo di integrazione T_I un contributo totale alla Y pari a K volte l'errore stesso.

In questo modo il termine integrativo è capace di annullare gli errori a regime facendo convergere la variabile reale R con il set point S. Questa convergenza spesso richiede un certo tempo, dipendente ovviamente dal tempo di integrazione T_I. Verrebbe spontaneo quindi utilizzare dei valori bassi di tale tempo; tuttavia come si deduce dalle equazioni, essendo T_I al denominatore, ciò porterebbe ad eccessivi contributi alla Y ad ogni ciclo di aggiornamento dell'equazione con conseguente instabilità del sistema e saturazione dell'uscita Y.

Infine il termine derivativo D corregge il valore dell'uscita Y aggiungendo o sottraendo ad essa, ad ogni ciclo di aggiornamento della funzione (tempo T), una quantità pari a $(K * T_D) / T$ volte il valore corrente della variazione della variazione dell'errore rispetto al ciclo precedente, a seconda che tale variazione sia rispettivamente positiva o negativa. Per esempio in caso di temperatura misurata R costante e set point S fisso risulta un errore X costante; in tal caso non viene apportata alcuna azione correttiva e così anche la Y rimane costante.

Se ad un certo punto la temperatura comincia rapidamente a cambiare in modo tale che tra due cicli successivi la variazione dell'errore è addirittura aumentata o diminuita, allora viene fornito uno scalino di variazione della Y pari al termine suddetto. Se invece la temperatura varia ma in modo costante il termine derivativo non contribuisce a modificare il valore dell'uscita. Lo scopo di tale termine è quindi quello di rispondere immediatamente a rapide variazioni dell'errore che comportano una sorta di accelerazione o decelerazione di questo. Il termine derivativo opera in pratica come "dumper" nel senso che è capace di smorzare eventuali rapide variazioni dovute a rumore a frequenza alta o ad oscillazioni in corrispondenza a brusche variazioni dell'errore. Un valore elevato del tempo derivativo, essendo T_D al numeratore dell'espressione, porta a notevoli contributi alla Y in un solo ciclo di aggiornamento con conseguente saturazione dell'uscita.

Per concludere le considerazioni di tipo pratico possiamo dire che un termine proporzionale elevato sarebbe auspicabile per una pronta risposta del sistema e per un errore a regime basso; un esempio estremo è la regolazione ON-OFF che è veloce in salita (piena potenza), veloce in discesa (potenza nulla) e commuta a cavallo del set point. Tuttavia una regolazione fortemente proporzionale rischia di avere elevate pendolazioni attorno al set point e eccessivi overshoots proprio come la regolazione ON-OFF. Una regolazione ON-OFF corrisponde ad un proporzionale con un guadagno elevatissimo e quindi con una buona eliminazione dell'errore medio a regime se pur con elevate pendolazioni attorno al set point. Diminuendo il guadagno proporzionale l'ampiezza picco-picco delle pendolazioni diminuisce ma il valore medio della temperatura tende a discostarsi dal set point causando un certo errore a regime. In sistemi dove sia accettabile un compromesso tra le pendolazioni e l'errore a regime, ottenuto con un certo valore di K, può essere sufficiente un regolatore di tipo P.

Per eliminare l'errore a regime si può passare al regolatore PI in quanto il termine integrativo è capace di ridurre il valor medio dell'errore. Tuttavia forti presenze di tale termine (si ricordi che il termine contribuisce in modo inversamente proporzionale al tempo T_I) possono aumentare le oscillazioni, gli overshoots e il tempo di salita e discesa della variabile controllata fino ad arrivare all'instabilità.

Per porre rimedio agli aumentati overshoots, come conseguenza dell'introduzione del termine integrativo, si può infine inserire il termine derivativo che si comporta da "damper" ossia da smorzatore delle oscillazioni. Tuttavia aumentando in modo indiscriminato il fattore derivativo (aumentando il tempo T_D) si aumenta il tempo di salita e discesa e si può anche causare instabilità al sistema. Il termine derivativo invece non influisce minimamente sull'errore a regime.