

Controllori PID e tecniche "anti wind-up"

Giovanni Testolin

Indice

1	Introduzione ai controllori PID	4
1.1	Azione proporzionale	6
1.2	Azione integrale	9
1.3	Azione derivativa	11
1.4	Il PID nel suo insieme	13
2	Tecniche "anti wind-up"	15
2.1	Desaturazione dell'azione integrale	16
2.2	Ricalcolo del termine integrale	20
2.3	IMC (internal model control)	23

Introduzione

L'argomento trattato all'interno di questa tesi riguarda i controllori PID nel loro insieme ed in particolare le tecniche anti wind-up.

Nel primo capitolo vengono analizzate dettagliatamente le parti fondamentali del controllore PID, documentate con schemi funzionali e con simulazioni dei vari processi.

Nel secondo capitolo viene esposto il problema del wind-up, dovuto alle saturazioni degli attuatori e alla presenza dell'azione integrale nei sopracitati controllori PID; vengono inoltre presentate delle possibili soluzioni (attraverso schemi anti wind-up) per poter migliorare la dinamica del sistema.

Capitolo 1

Introduzione ai controllori PID

Il controllore PID è un algoritmo di controllo dotato di una struttura predefinita, che viene tarato modificando il valore di alcuni parametri. Grazie alla sua semplicità di utilizzo, combinata ad una apprezzabile efficacia in diversi campi di impiego, è l'algoritmo di controllo di gran lunga più usato nelle applicazioni industriali.

Dal punto di vista matematico, il regolatore PID è un sistema dinamico che elabora un segnale in ingresso, detto "errore" (ottenuto come differenza fra il riferimento e la variabile controllata $e(t) = r(t) - y(t)$) dando poi in uscita un segnale di controllo $u(t)$ (Figura 1.1).

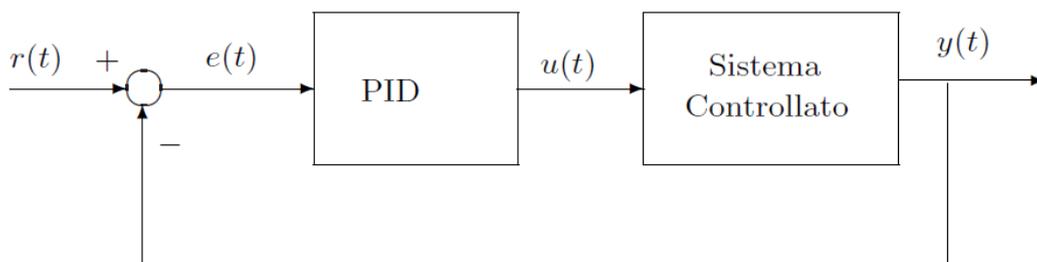


Figura 1.1: Schema di un sistema di controllo in retroazione.

Il successo dei PID è dovuto principalmente a:

- Notevole efficacia nella regolazione di un'ampia gamma di processi industriali (processi termici, meccanici, ecc.);
- Possibilità di essere realizzati con tecniche diverse (meccaniche, idrauliche, elettroniche sia analogiche che digitali, ecc.);

- Relativa semplicità di taratura dei parametri di lavoro in quanto il loro utilizzo non necessita da parte dell'operatore una conoscenza dettagliata del processo da controllare;
- Buon rapporto efficacia/costo perchè, in presenza di inevitabili condizioni di rumore o di inopportune calibrazioni errate, l'efficacia della legge di controllo decade, quindi l'utilizzo di altri sistemi di controllo più raffinati sarebbe inutile, in quanto le loro prestazioni diverrebbero equiparabili a quelle dei PID;
- Duttilità nel loro utilizzo; difatti spesso vengono impiegati per ottenere configurazioni più articolate per rispondere a esigenze complesse.

Il regolatore PID nella sua forma base, ha una struttura comprendete la somma di tre termini di controllo

$$u(t) = K_P \cdot e(t) + K_I \cdot \int_0^t e(\tau) d\tau + K_D \cdot \frac{de(t)}{dt}$$

I parametri che identificano il PID dunque sono K_P , K_I e K_D ed essi sono chiamati anche gradi di libertà del controllore.

Passiamo ora a osservare nel dettaglio le parti che compongono la struttura del controllore PID.

1.1 Azione proporzionale

La prima azione è quella chiamata proporzionale in quanto lega algebricamente l'ingresso $e(t)$ e l'uscita $u(t)$ secondo una costante K_P chiamata guadagno proporzionale.

Guardando la formula base del PID espressa in precedenza, possiamo capire quale significato assume il guadagno proporzionale K_P quando lavora in condizioni di regolatore puramente proporzionale ($K_I = K_D = 0$).

In queste condizioni il regolatore provoca l'attuazione di una correzione proporzionale (con costante di proporzione K_P) all'errore $e(t)$. Al crescere di K_P la velocità di risposta del sistema migliora ma, diminuisce la stabilità aumentando le oscillazioni.

Il parametro caratteristico dell'azione proporzionale, K_P , ha una definizione prettamente teorica, infatti nella realtà industriale al posto di quest'ultimo si fa invece riferimento ad un'altra grandezza chiamata Banda Proporzionale e indicata con il simbolo B_P , legata a K_P e definita nel seguente modo:

B_P è la variazione dell'ingresso $e(t)$ (espressa in percentuale) che porta l'uscita $u(t)$ al suo valore di fondo scala.

Di seguito sono evidenziati due esempi nei quali si ha un campo di valori uguale per ingresso ed uscita del PID e si può vedere come nel primo caso la percentuale di variazione dell'ingresso $e(t)$ che provoca il raggiungimento da parte dell'uscita $u(t)$ del suo valore di fondo scala è del 50% mentre nel secondo caso avviene la stessa cosa ma con una variazione del 25%.

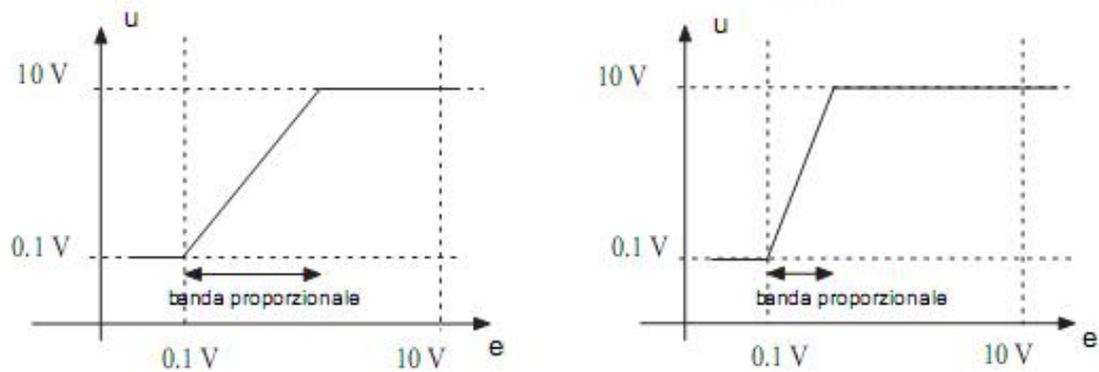


Figura 1.2:

Quando K_P è espresso come rapporto tra i segnali normalizzati ai rispettivi valori di fondo scala, cioè

$$K_P = \frac{u(t)/u_f}{e(t)/e_f}$$

dove u_f è il valore di fondo scala di $u(t)$ ed e_f è il valore di fondo scala di $e(t)$, si può usare la seguente relazione tra B_P e K_P

$$K_P = \frac{100}{B_P}$$

dalla quale si evince che riducendo la banda proporzionale si ottiene una maggiore precisione dello strumento in quanto è sufficiente una minore variazione dell'errore in ingresso perchè questa porti ad un significativo cambiamento in uscita.

Dunque riducendo la banda proporzionale il guadagno aumenta e la precisione migliora (aumenta la velocità di risposta) ma non è tutto, infatti a un aumento del guadagno corrisponde anche un aumento del valore raggiunto all'uscita del blocco

P, che si discosterà sempre di meno dal valore originale del segnale $r(t)$ senza però raggiungerlo a causa dell'errore a regime.

All'aumento del guadagno c'è però un limite dovuto al fatto che oltre un certo valore il sistema entra in uno stato di auto oscillazione, cioè di instabilità.

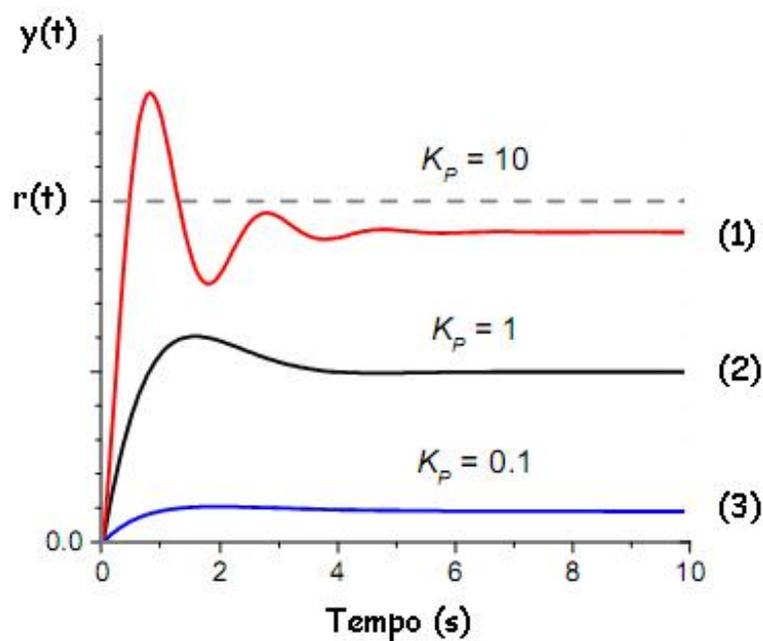


Figura 1.3:

Nella Figura 1.3 sono riportati dei grafici che raffigurano gli effetti di una banda proporzionale eccessivamente piccola (1), adeguata (2) ed eccessivamente grande (3).

1.2 Azione integrale

La seconda azione è quella integrale il cui contributo è proporzionale all'integrale dell'errore $e(t)$ (quindi proporzionale al suo valor medio) e in cui la costante K_I è data dall'inverso della costante di tempo integrale T_I (chiamata anche tempo di reset) moltiplicata per il guadagno proporzionale

$$K_I = \frac{K_P}{T_I}$$

Questa azione ha come caratteristica principale quella di eliminare l'errore a regime causato dal blocco proporzionale, che si presenta qualora si abbia a che fare con un gradino di riferimento.

Infatti l'uscita di questo blocco è proporzionale all'integrale nel tempo dell'errore in ingresso $e(t)$, e nel caso di un segnale di riferimento $r(t)$ a gradino, l'integrale crescerà sempre più come vediamo in Figura 1.4 .

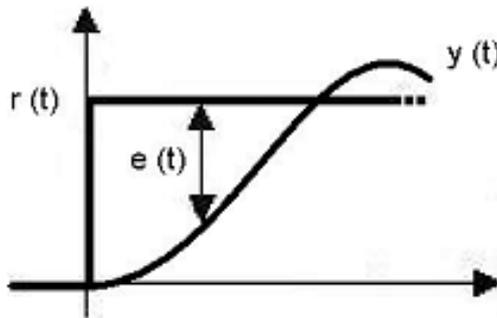


Figura 1.4:

Questo significa che, grazie alla retroazione, viene riportato in ingresso e mitiga l'influenza del segnale di riferimento $r(t)$ sul sistema; infatti l'integrale dell'errore aumenta finché esso è positivo e diminuisce quando $e(t)$ assume valori negativi.

Per questo blocco, il termine variabile da definire in base al sistema con cui si ha a che fare, è la costante di tempo di integrazione indicata da T_I ; da essa infatti dipende l'effetto di integrazione che è tanto maggiore quanto più T_I è ridotta. Infatti riducendo la costante di tempo, l'integrale salirà più velocemente verso il valore $e(t)$ in ingresso al blocco, pagando però questa velocità di salita con delle forti oscillazioni che richiedono tempo prima di stabilizzarsi. Viceversa aumentando T_I e facendolo tendere ad infinito si ha come conseguenza l'eliminazione dell'effetto dell'integrale.

Per chiarire meglio il concetto si riporta in Figura 1.5 il grafico dell'andamento di un controllore formato dai blocchi proporzionale ed integrale e si forniscono i risultati dell'uscita per diversi valori di T_I .

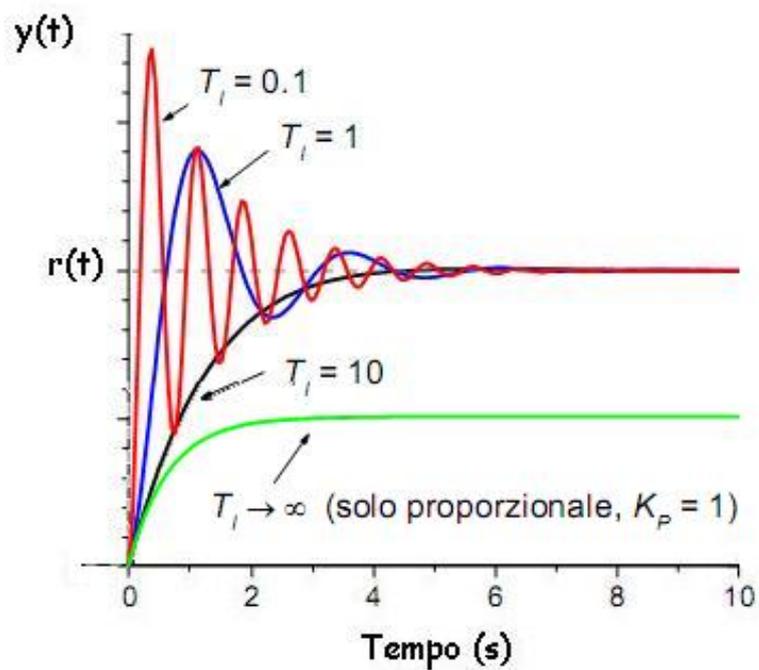


Figura 1.5:

1.3 Azione derivativa

La terza azione è quella derivativa che fornisce in uscita la derivata rispetto al tempo dell'errore $e(t)$, dove la costante K_D è data dal prodotto tra il guadagno proporzionale e la costante di tempo derivativa T_D

$$K_D = K_P \cdot T_D$$

La sua funzione consiste nel derivare il segnale che si trova al suo ingresso, perciò tiene conto delle rapide variazioni dell'errore e cerca, in qualche modo, di anticipare la futura azione correttiva.

In pratica, non potendo essere utilizzata come azione singola per inserire un riferimento costante a causa dello zero che introduce nell'origine, la sua funzione si concretizza in una riduzione delle oscillazioni del segnale attorno al valore d'uscita, mentre per quanto riguarda l'offset non lo influenza in alcun modo.

Infatti le proprietà della derivata sono quelle di aumentare in modo proporzionale alla velocità di variazione della grandezza a cui fa riferimento, perciò nel caso in cui il segnale $r(t)$ assuma una forma a gradino, l'azione derivativa agirà in modo da inseguire l'errore e correggerlo velocemente.

Il parametro che governa questo blocco è la costante di tempo derivativa indicata dal simbolo T_D , il cui valore determina la velocità di salita del segnale.

Di seguito, in Figura 1.6, si può apprezzare l'azione anticipatrice del blocco derivatore e la sua correzione sulle oscillazioni attorno al valore desiderato, inoltre si conferma la non incidenza sull'offset tra $y(t)$ ed $r(t)$.

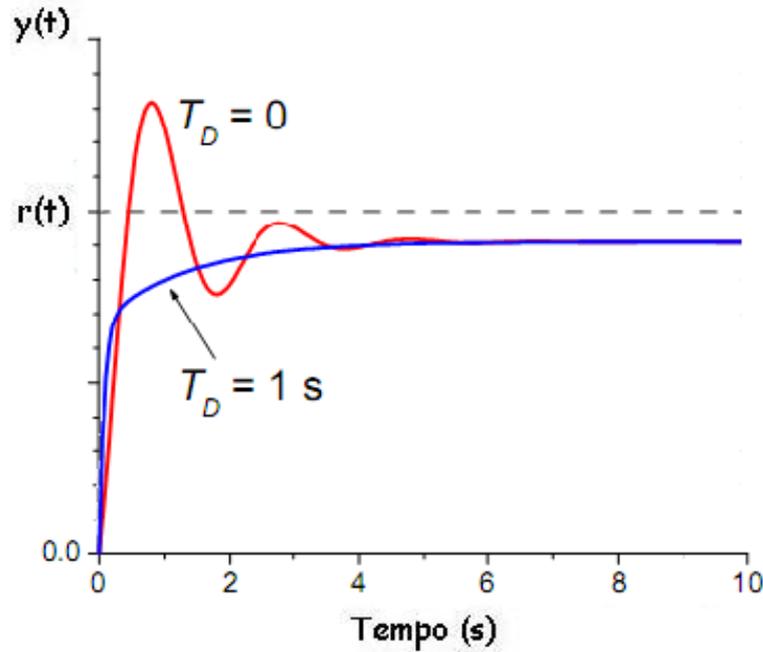


Figura 1.6:

Il termine derivativo purtroppo introduce l'inconveniente di amplificare i segnali con contenuto armonico a frequenze elevate.

Siccome in genere questo tipo di segnali corrispondono al rumore elettromagnetico sovrapposto al segnale utile, l'utilizzo del termine derivativo deve essere valutato con cautela.

Per capire meglio come il rumore di misura venga amplificato dal termine derivativo, consideriamo un rumore approssimabile da un segnale sinusoidale in ingresso al blocco derivatore

$$n(t) = A \sin(\omega t)$$

Tale segnale all'uscita del blocco derivatore diventerà:

$$u_n(t) = T_D \frac{dn(t)}{dt} = AT_D \omega \cos(\omega t)$$

quindi segnali di rumore aventi un contributo sensibile alle alte frequenze, e quindi una ω molto alta, vengono amplificati notevolmente dal termine derivativo del PID.

1.4 Il PID nel suo insieme

Oltre alla configurazione PID, come accennato prima, si trovano industrialmente utilizzate anche altre configurazioni ognuna con peculiarità proprie e specifiche per dati processi.

- **Configurazione P** : viene utilizzato nei sistemi nei quali sono consentiti scostamenti tra il valore a regime della grandezza controllata e quello desiderato, come ad esempio negli impianti di riscaldamento degli edifici. Infatti esso introduce un errore a regime in funzione di K_P , aumenta la velocità di risposta del sistema e, con guadagno elevato, diminuisce la stabilità aumentando le oscillazioni.
- **Configurazione PI** : l'unione di questi due moduli permette di ottenere una maggiore precisione senza peggiorare il grado di stabilità del sistema (sempre che le costanti siano opportunamente scelte), unita ad una maggiore velocità di risposta. Vengono utilizzati quando è richiesto un errore a regime di modesta entità unito ad una buona velocità di risposta alle variazioni della sollecitazione; pertanto vengono inseriti soprattutto nei sistemi in cui le variazioni di carico avvengono lentamente: infatti cambiamenti improvvisi di carico possono portare il sistema verso l'instabilità, quando il coefficiente K_I dell'azione integrale non è scelto in modo opportuno. In ambito industriale oltre il 90% dei controllori sono di questo tipo.
- **Configurazione PD** : per le sue caratteristiche il regolatore PD tende ad anticipare l'errore e pertanto può essere utilizzato nei sistemi in cui si hanno improvvise variazioni di carico come in alcuni sistemi di controllo per servomotori, oppure anche in sistemi che non presentano problemi di stabilità ma che, invece, richiedono una buona velocità di risposta.
- **Configurazione PID** : è il controllore utilizzato in quei processi lenti, che non hanno bisogno di risposte immediate alla sollecitazione, dove però la presenza di oscillazioni crea problemi.

Da notare come i controllori ID o D non vengano menzionati, in quanto mai utilizzati. Il primo a causa del fatto che tale regolatore presenta un'anti-risonanza ad una determinata pulsazione, il secondo in quanto non in grado di controllare l'errore a regime.

Le tre azioni viste finora, combinate assieme per formare la configurazione PID, danno origine alla formula completa scritta in precedenza ma con degli accorgimenti

$$u(t) = K_P \left(e(t) + \frac{1}{T_I} \cdot \int_0^t e(\tau) d\tau + T_D \cdot \frac{de(t)}{dt} \right)$$

Passando alla rappresentazione tramite trasformate di Laplace delle equazioni del PID otteniamo la successiva equazione dove $U(s)$ ed $E(s)$ sono le trasformate dei segnali d'uscita e d'ingresso del PID

$$U(s) = K_P \left(1 + \frac{1}{T_I \cdot s} + s \cdot T_D \right) \cdot E(s)$$

Capitolo 2

Tecniche "anti wind-up"

Gli attuatori utilizzati nei sistemi di controllo hanno dei vincoli sull'ampiezza delle uscite, che non possono superare dei valori massimi e minimi. Quando si utilizza un regolatore con azione integrale, è possibile che l'uscita del controllore raggiunga i suddetti limiti; in tal caso l'azione dell'attuatore non può crescere, anche se l'errore di regolazione $e(t)$ non è nullo.

La situazione reale che spesso si incontra è quella riportata in Figura 2.1 (attuatore in saturazione), in cui:

$$m(t) = \begin{cases} -U_M & , \text{ se } u(t) < -U_M \\ u(t) & , \text{ se } |u(t)| \leq U_M \\ U_M & , \text{ se } u(t) > U_M \end{cases}$$

Il fenomeno del wind-up

In presenza di saturazione, come detto in precedenza, può verificarsi che l'uscita dell'attuatore non cresca, pur rimanendo l'errore di regolazione $e(t)$ non nullo. Conseguentemente, il termine integrale continua a crescere, ma tale incremento non produce alcun effetto sulla variabile di comando dell'impianto. Tale situazione, oltre a non far

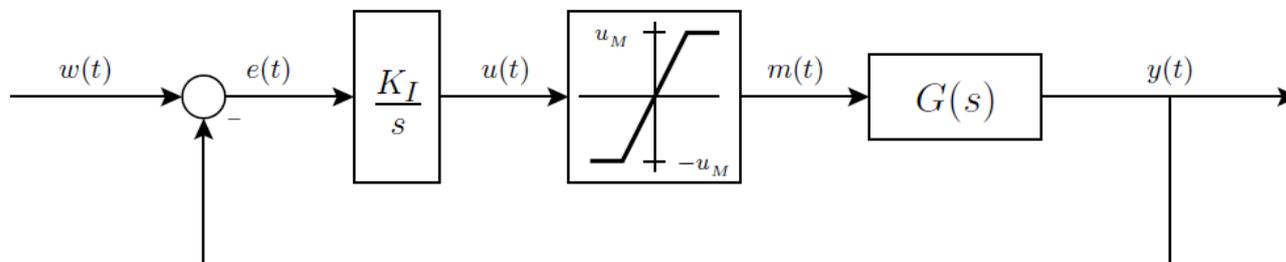


Figura 2.1: Schema in cui l'attuatore presenta una saturazione.

funzionare correttamente il regolatore, rende inattivo il regolatore anche quando l'errore diminuisce o si inverte di segno; infatti, il sistema di regolazione può riattivarsi solo allorché il segnale $u(t)$ rientra nella zona di linearità della caratteristica dell'attuatore (scarica del termine integrale). Questo fenomeno si chiama comunemente carica integrale o integral wind-up.

2.1 Desaturazione dell'azione integrale

Sia data la seguente funzione di trasferimento:

$$G(s) = \frac{4}{s + 2}$$

Ipotizziamo di utilizzare un controllore integrale con $K_I=1$.

Supponiamo adesso che l'attuatore presenti una saturazione con $U_M=0.53$. I risultati della simulazione per un ingresso a gradino sono mostrati in Figura 2.2.

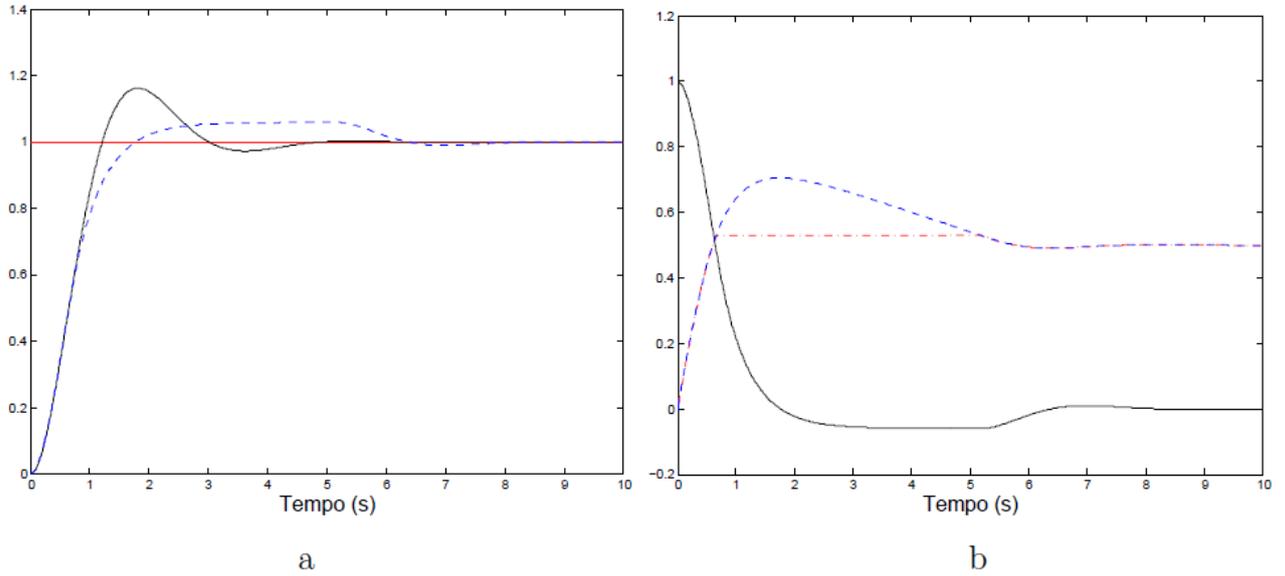


Figura 2.2: (a) Risposta di un sistema in assenza (—) e in presenza (- -) di saturazione.(b) Andamento dei segnali $e(t)$ (—), $u(t)$ (- -) ed $m(t)$ (-.) in presenza di saturazione.

Schema di desaturazione

Il problema del wind-up può essere evitato interrompendo l'azione integrale non appena l'uscita del controllore raggiunge il livello di saturazione dell'attuatore.

Una possibile soluzione è riportata in Figura 2.3

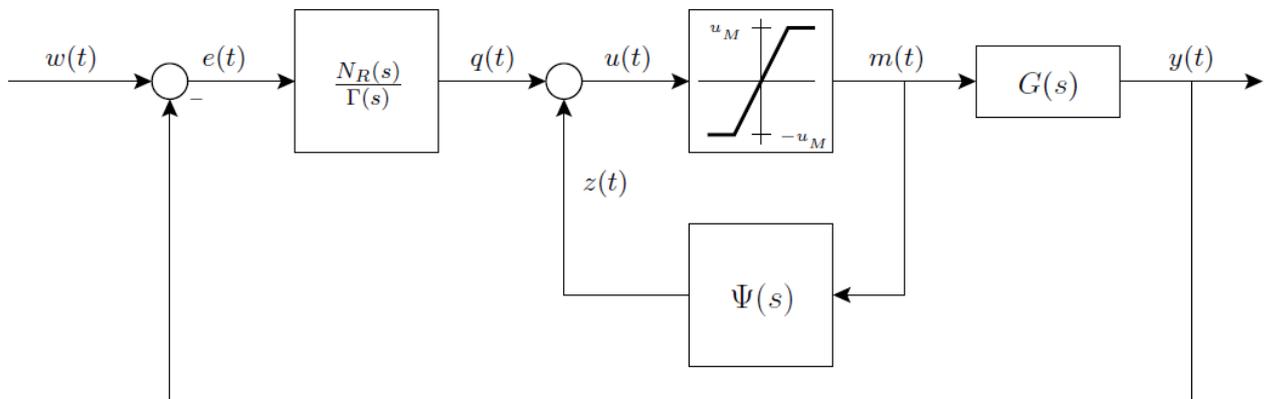


Figura 2.3: Schema di realizzazione di un controllore PID con dispositivo anti wind-up

Supponiamo che il controllore PID che deve essere realizzato sia della forma generale

$$R_{PID}(s) = \frac{N_R(s)}{D_R(s)}, \quad \text{con } D_R(0) = 0,$$

in cui la condizione $D_R(0) = 0$ implica la presenza di un blocco integrale nel controllore. Supponiamo che sia $N_R(0) > 0$. Allora, riferendosi allo schema riportante in Figura 2.3, si sceglie il polinomio $\Gamma(s)$ in modo tale che la funzione di trasferimento

$$\Psi(s) := \frac{\Gamma(s) - D_R(s)}{\Gamma(s)}$$

sia bipo-stabile, strettamente propria e con guadagno unitario ($\Psi(0) = 1$).

Si può allora osservare che:

- Se l'attuatore opera in regione di linearità, la funzione di trasferimento fra $e(t)$ e $m(t)$ coincide con la $R_{PID}(s)$ desiderata.

Dimostrazione

$$Q(s) = E(s) \cdot R_{PID}(s)$$

$$U(s) = Q(s) + M(s) \cdot \Psi(s), \quad \text{ma siccome l'attuatore non è in saturazione}$$

$$U(s) = M(s), \quad \text{quindi } Q(s) = M(s) \cdot (1 - \Psi(s)).$$

Concludendo, la f.d.t. tra $M(s)$ e $E(s)$ risulta:

$$\frac{M(s)}{E(s)} = \frac{N_R(s)}{\Gamma(s) \cdot (1 - \Psi(s))} = \frac{N_R(s)}{\Gamma(s) - \Gamma(s) \cdot \frac{\Gamma(s) - D_R(s)}{\Gamma(s)}} = \frac{N_R(s)}{D_R(s)}$$

che risulta essere esattamente la f.d.t. del controllore.

- Se il segnale errore $e(t)$ permane dello stesso segno per un tempo elevato, allora anche $q(t)$, in funzione della dinamica di $\Gamma(s)$, assumerà lo stesso segno; $m(t)$ satura al valore massimo U_M dell'attuatore. Dato che $\Psi(0) = 1$, anche $z(t)$ si assesterà al valore U_M , sempre con una dinamica che dipende da $\Gamma(s)$. Se $e(t)$ cambia di segno, anche $q(t)$ cambia segno e quindi il segnale $u(t) = q(t) + z(t)$ scende subito sotto il valore di saturazione U_M , attivando il comportamento lineare dell'attuatore. Le prestazioni del sistema di desaturazione dipendono dalla scelta del polinomio $\Gamma(s)$, che costituisce il nucleo del progetto del desaturatore.

Assumendo $\Gamma(s) = s + 8 \implies \Psi(s) = \frac{8}{s+8}$

la risposta del sistema al gradino è rappresentata in Figura 2.4.

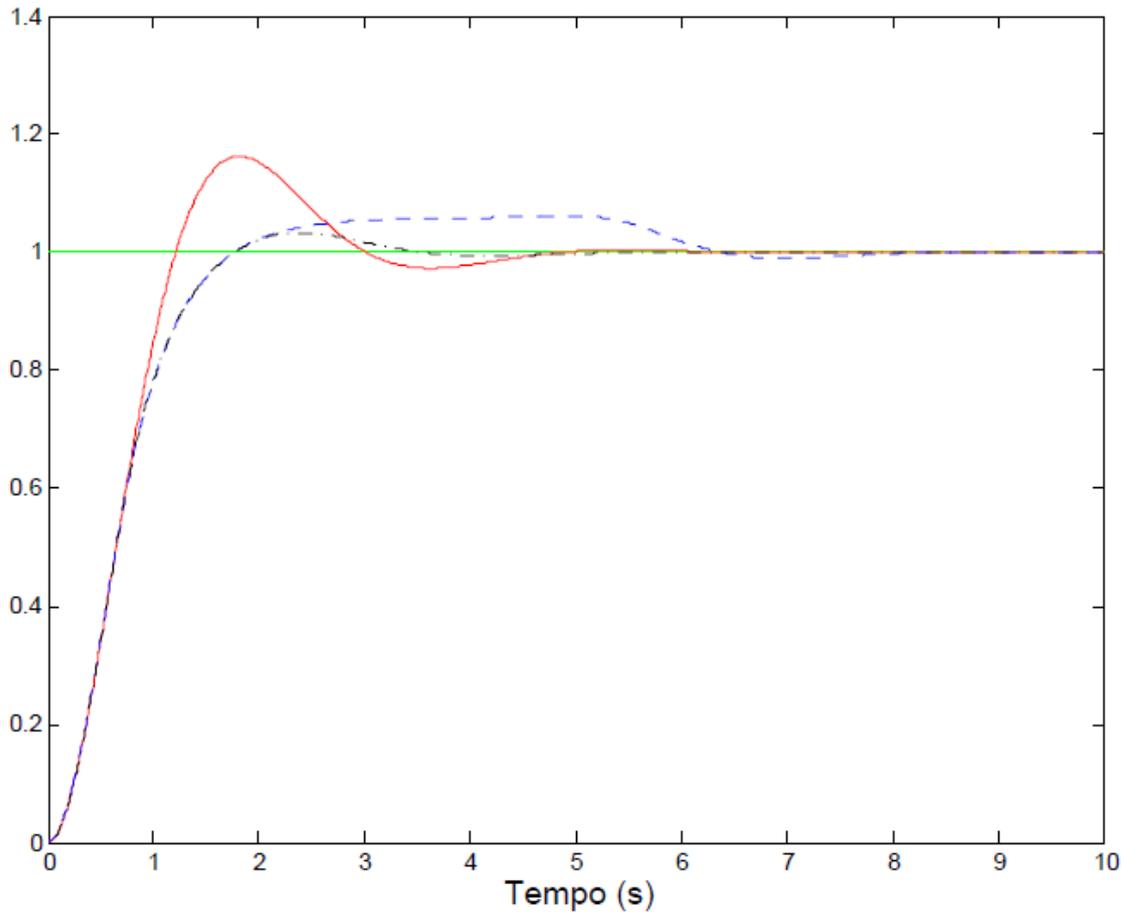


Figura 2.4: Risposta del sistema senza saturazione(—), con saturazione (- -) e con saturazione e anti wind-up (-.)

2.2 Ricalcolo del termine integrale

Una seconda tecnica anti wind-up introduce un termine $e_s(t) = v(t) - u(t)$ che viene riportato all'ingresso del blocco integrale e sottratto al valore dell'errore $e(t)$.

Questo segnale viene generato prendendo il valore della grandezza di interesse misurata all'uscita dell'attuatore e confrontandolo con il valore che la stessa grandezza presenta all'ingresso.

Così facendo si otterrà un valore nullo (che quindi non influirà in alcun modo nel controllo) nel caso in cui l'attuatore stia lavorando entro i limiti di saturazione, mentre quando questi limiti verranno superati si genererà un valore che aiuterà a smorzare la salita del termine integrale.

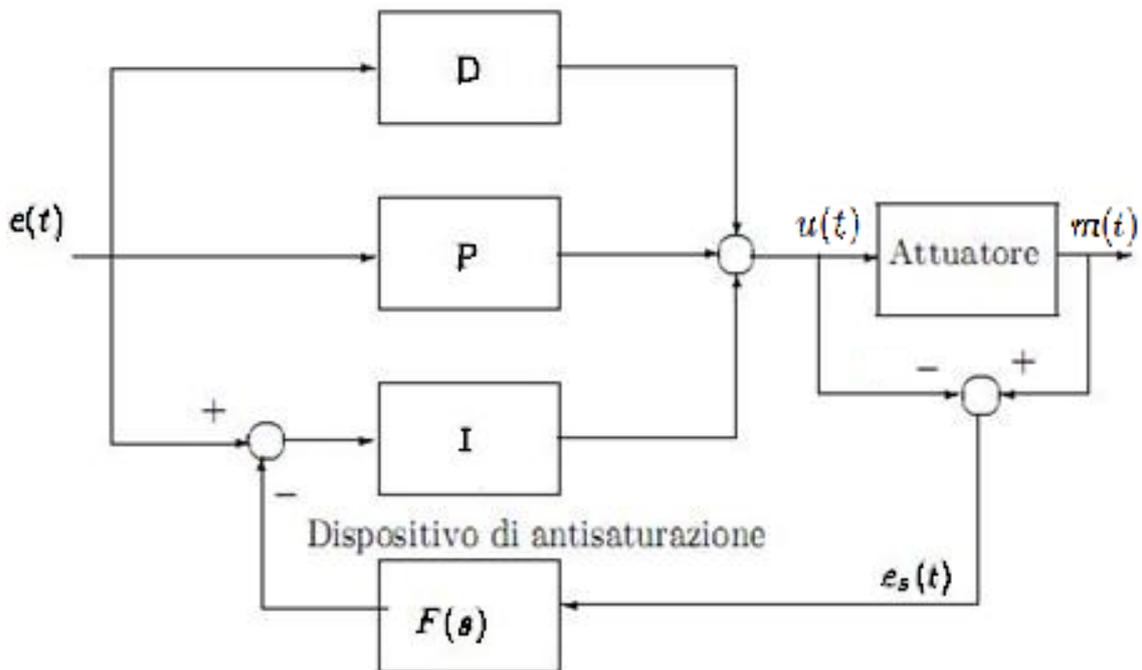


Figura 2.5: Schema di principio per la costruzione della variabile di attuazione.

Da questo schema si nota inoltre come sia possibile inserire prima del ritorno del segnale di compensazione al blocco integratore, un altro blocco chiamato dispositivo di antisaturazione, la cui funzione di trasferimento è solitamente pari a $1/T_A$ (motivata dal fatto che il termine di compensazione andrà poi in ingresso ad un blocco integratore) e il cui valore controlla la velocità dell'azione di desaturazione.

Infatti, in presenza di valori della costante di tempo T_A ridotti, si avrà un'azione di desaturazione più rapida ma molto più facilmente soggetta a disturbi; in caso contrario invece, se il valore assegnato alla costante di tempo risulterà elevato, l'azione di desaturazione sarà meno influenzata da disturbi ma anche più lenta.

Come nel caso precedente facciamo riferimento allo stesso sistema da controllare ma utilizzando un controllore PI.

$$G(s) = \frac{4}{s+2} \quad , \quad R_{PI} = 1 + \frac{5}{s}$$

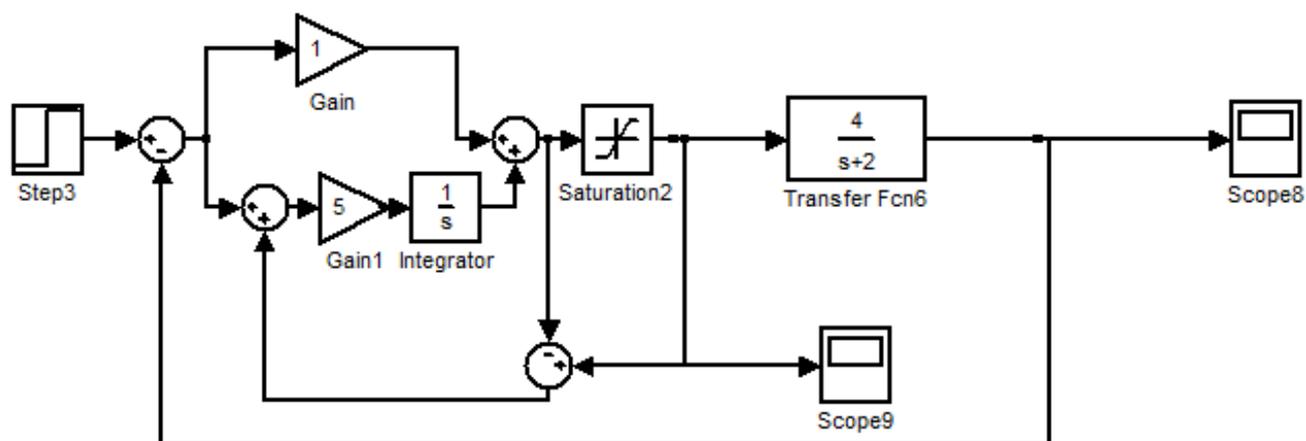


Figura 2.6: Schema di desaturazione realizzato con Simulink.

Il sistema in ingresso presenta una segnale "tagliato" in quanto l'attuatore è in saturazione, Figura 2.7.

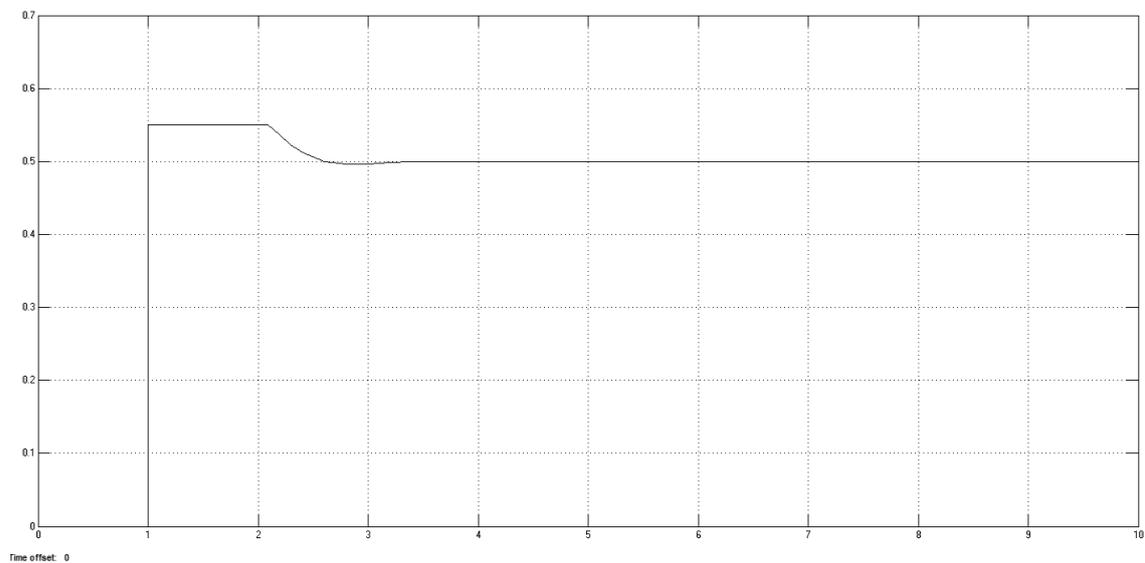


Figura 2.7:

In uscita analizziamo i due casi, con sistema anti wind-up e non. Nel primo notiamo come l'uscita va regime con un tempo elevato rispetto la seconda, dovuta al fatto che il termine integrale ci mette più tempo a scaricarsi, Figura 2.8.

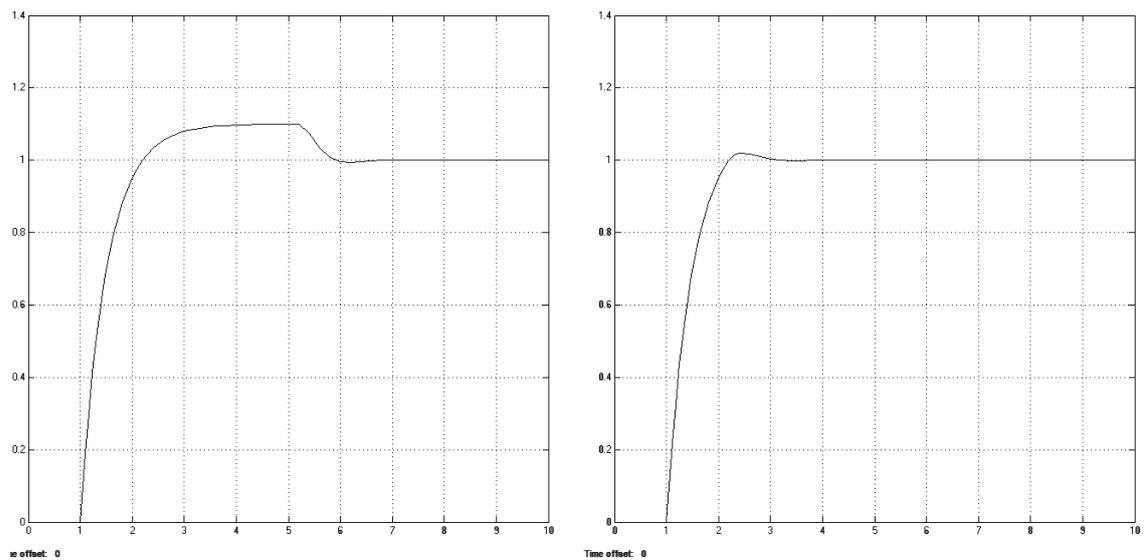


Figura 2.8:

2.3 IMC (internal model control)

Un'altra tecnica per eliminare il problema del wind-up, è la cosiddetta tecnica IMC il cui acronimo significa:

- **Internal:** il modello del sistema è interno al controllore;
- **Model:** è infatti richiesto un modello del processo da controllare;
- **Control:** il controllore usa il modello per migliorare le prestazioni.

Essa basa il suo funzionamento sull'ipotesi di conoscere il processo che dobbiamo controllare, in modo da poterne ottenere una copia e integrarla nel controllore, come illustrato in Figura 2.9.

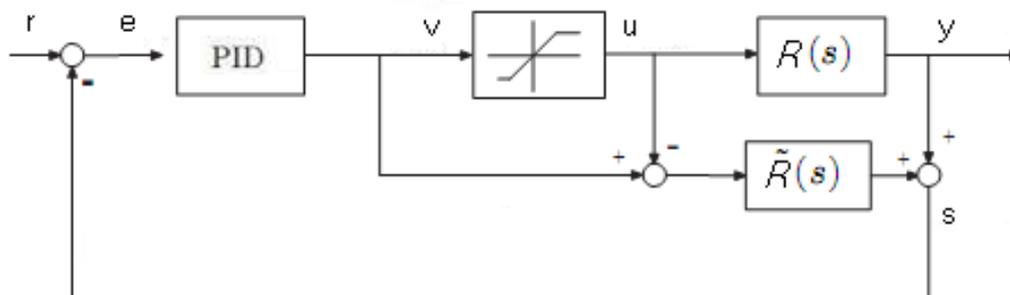


Figura 2.9:

In questa Figura si è definita $R(s)$ la funzione di trasferimento del sistema controllato e $\tilde{R}(s)$ quella della copia del sistema.

Il principio di funzionamento di questa tecnica viene descritto matematicamente di seguito:

per prima cosa si calcola l'errore $e = r - s \longrightarrow e = r - [y + \tilde{R} \cdot (v - u)]$

$$e = r - Ru - \tilde{R}v + \tilde{R}u$$

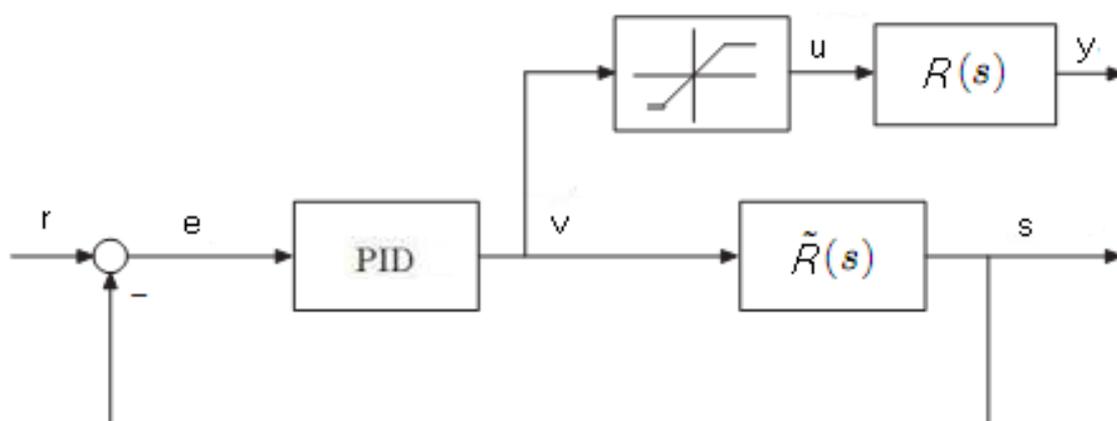
considerando poi che all'inizio si è imposto che la copia del sistema controllato rispecchiasse esattamente il sistema originale si può dire che

$$R = \tilde{R}$$

e dunque la formula si riduce a

$$e = r - Rv$$

di conseguenza adottando questa soluzione, e quindi costruendo un controllore che abbia al suo interno una copia del sistema, il tutto si può ricondurre a uno schema di questo tipo



nel quale l'elemento non lineare è stato spostato al di fuori dall'anello di controllo.

Conclusione

Abbiamo dunque visto come attraverso dei semplici e convenienti schemi funzionali si può risolvere il problema del wind-up, migliorando la risposta del sistema a tal punto da essere superiore a quello in funzionamento lineare.

Bibliografia

1. G. Marro, "Controlli automatici", V edizione, Zanichelli Editore, Bologna.
2. Bolzern, Scattolini, Schiavoni, Fondamenti di controlli automatici, Mc Graw-Hill, 2008