Introduction to control theory

WEEK 5

PI controllers

We have seen in WEEK 4 that, in order to eliminate the steady-state error, C should have a pole at $s=0$.
A PI (proportional-integral) controller has the transfer function

$$C(s) = K_p + \frac{K_i}{s} ,$$

where K_p is the proportional term, and K_i/s is the integral term. The constants K_p and K_i must be tuned according to the plant and according to the control objectives. In time domain, if e denotes the error signal and v denotes the controller output, then

$$v(t) = K_p e(t) + K_i \int_0^t e(\sigma)\, d\sigma .$$

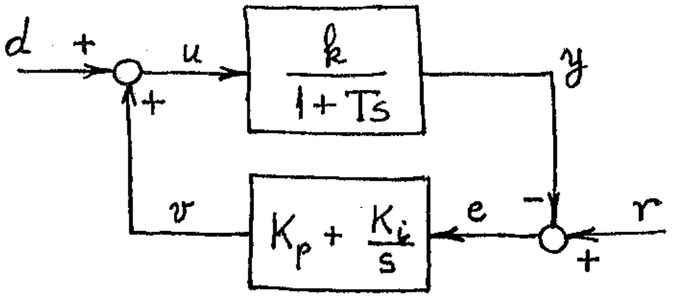
Such controllers are used to combine the advantages of proportional and of integral controllers. Since there are two parameters to tune, we have more freedom of choice and can achieve better performance. Since C has a pole at zero, if the control system is stable, the steady state error corresponding to step reference and disturbance signals is zero.

In the sequel we assume that the plant is a first order system,

$$P(s) = \frac{k}{1+Ts} .$$

Thus, the loop gain is

$$P(s)\,C(s) = \frac{k(K_i + K_p s)}{s(1+Ts)} .$$



The sensitivity $S = (1+PC)^{-1}$ is easily computed:

$$S(s) = \frac{s(1+Ts)}{Ts^2 + (1+kK_p)s + kK_i} .$$

As expected, we have $S(0)=0$, so that the steady state error is zero. Note that, compared to integral control, the difference is the appearance of the term kK_p in the coefficient of s in the denominator.

We discuss stability. If K_p and K_i are of the same sign (usually they are both positive), then there can not be an unstable pole-zero cancellation in $P \cdot C$. Thus, the feedback system is stable if S is stable, which is the case if

$$\boxed{\frac{1+kK_p}{T} > 0, \qquad \frac{kK_i}{T} > 0.} \quad \text{Stability conditions}$$

Possible also for $T<0$.

We introduce the parameters

$$\omega_n^2 = \frac{kK_i}{T}, \qquad \zeta = \frac{1+kK_p}{2\sqrt{kK_iT}},$$

Note that we can easily achieve $\zeta = 1$ and large ω_n, we have more freedom than with an integral controller.

then

$$S(s) = \frac{s^2 + \frac{1}{T}s}{s^2 + 2\zeta\omega_n s + \omega_n^2}$$

The transfer function from r to y is

$$G(s) = 1 - S(s).$$

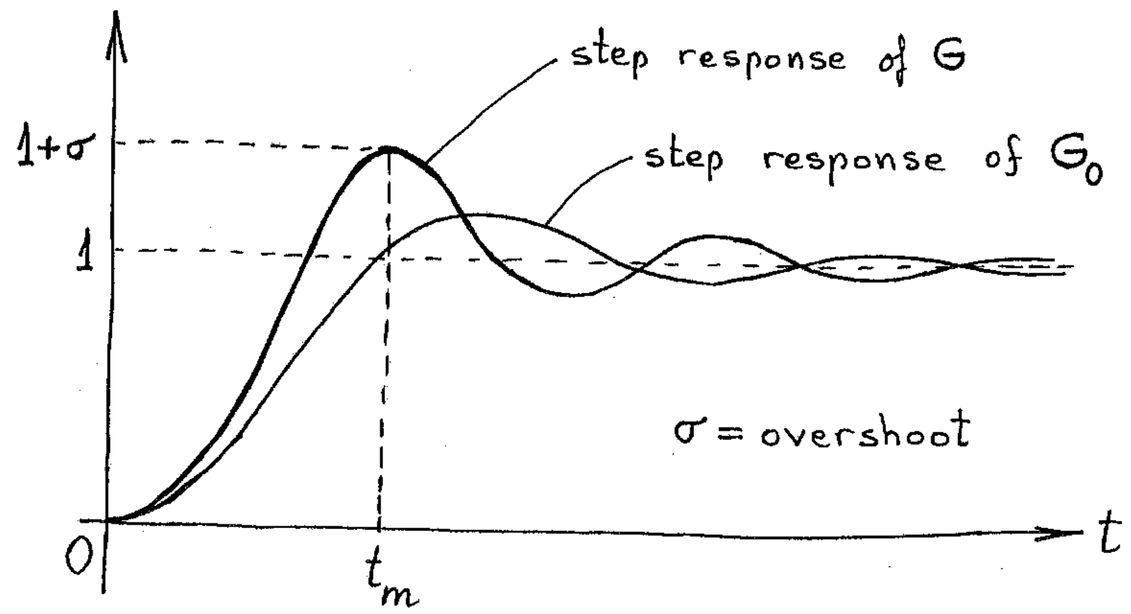
Denoting $z = \frac{K_i}{K_p}$, this can be written (check!)

$$G(s) = \frac{s+z}{s^2 + 2\xi\omega_n s + \omega_n^2} \cdot \frac{\omega_n^2}{z} \, .$$

Note that $G(0)=1$, and G has a zero at $-z$. The presence of this zero makes the formula for the step response more complicated, so we omit to write it. Compared to the step response which we would have if there were no zero, i.e., compared to

$$G_0(s) = \frac{\omega_n^2}{s^2 + 2\xi\omega_n s + \omega_n^2}$$

(which corresponds to $z \to \infty$, $K_p \to 0$, integral control) the step response raises quicker, and has a higher overshoot, as shown below:



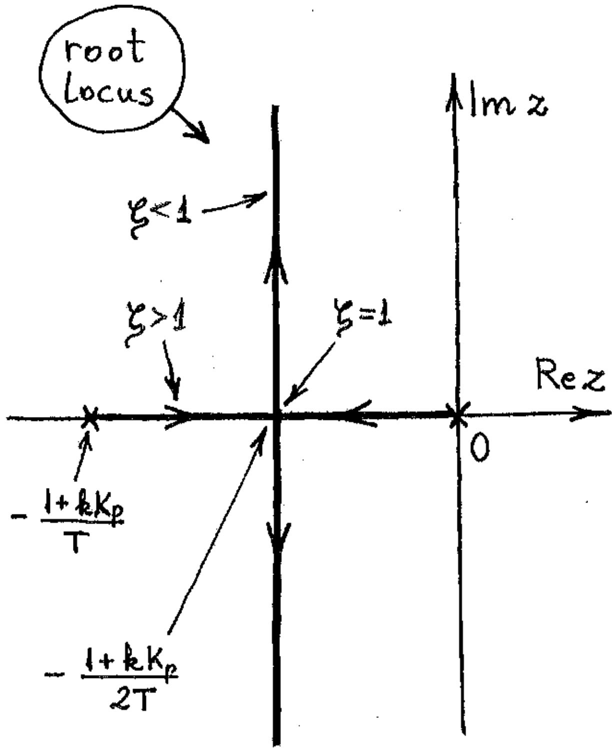
This picture was drawn under the assumption that ξ and ω_n for the two step responses is the same, and $\xi < 1$ (otherwise there would be no oscillations). There will be overshoot also for $\xi = 1$.

We want to understand how the poles of S (same as the poles of G) depend on K_p and K_i. From the formula for S we see that, denoting the poles by p_1 and p_2,

$$p_1 + p_2 = -\frac{1 + kK_p}{T}, \quad p_1 p_2 = \frac{kK_i}{T}.$$

Thus, the average of the poles is always $-\frac{1+kK_p}{2T}$, and p_1 and p_2 can be either symmetrically left and right of this average (if $\xi > 1$), they can be both equal to the average (if $\xi = 1$) or they can be symmetrically above and below it (if $\xi < 1$). If we keep K_p fixed then for $K_i = 0$ one pole is at zero and the other at $-(1+kK_p)/T$ (this corresponds to $\xi = \infty$, unstable system). As we increase K_i, the poles move closer, until they meet for

$$K_i = \frac{(1 + kK_p)^2}{4kT} \quad \left(\text{because then } \xi = 1\right)$$



The centerpoint of this root-locus diagram, the average $-(1+kK_p)/2T$, can be moved to the left by increasing K_p. However, we cannot increase K_p arbitrarily much, because for a step reference signal

$$u(0) = C(\infty)e(0) = K_p$$

(assuming $d=0$) and we cannot impose too large $u(0)$ (this was discussed also at proportional control). After choosing K_p, we choose K_i such that $\xi \approx 1$.

The Ziegler-Nichols empirical tuning rules for PI controllers

Most controllers used in industry are PI. This is because PI controllers are easy to understand and they ensure $e_{ss}=0$ for constant r and d, which is a common requirement. Often, the model of the plant is not known to the engineer who designs the control system. For this situation (which requires experimental tuning), J. Ziegler and N. Nichols have proposed (in 1942) the following empirical tuning method:

1. Take a proportional controller. The plant is assumed to be stable and with positive DC-gain. For small controller gain, the feedback system will be stable. Increase the controller gain until instability is reached at the gain K_c (called the critical gain).
2. At the critical gain ($C(s)=K_c$) the feedback system is marginally stable, i.e., it has poles on the imaginary axis, hence the signals u, y, e will oscillate. Measure the period of the oscillations, T_p.
3. Use the PI controller $C(s)=K_p\left(1+\frac{1}{T_i s}\right)$, where
$$K_p = 0.45\,K_c\,, \qquad T_i = 0.833\,T_p\,.$$

There is no guarantee that this will work, but it works and has reasonable performance in most cases.

PI controllers with anti-windup

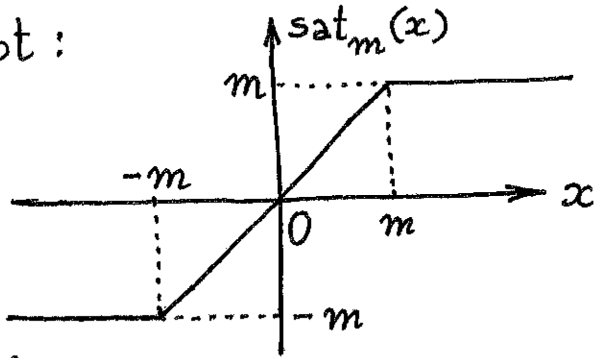
If the plant in a control system with PI controller stops working for a while (because of a technical failure or operator command), then the integrator in the controller may reach a very high value. When the plant becomes operational again, the high integrator value will cause a very high control input, which may cause malfunction or damage. Even if no damage is caused, it will take some time until the control system returns to normal operation. This phenomenon is called <u>windup</u> of the integrator (probably inspired by winding up the spring in a mechanical clock). The "technical failure" that causes windup is often a saturation effect within the plant, not really a failure in the usual sense, but saturation causes the plant model to change.

To avoid windup, various techniques have been proposed, some very sophisticated (see for example Fen Wu and Bei Lu, Systems and Control Letters, Vol. 52, 2004). Industrial PI controllers often have a circuit (or a part of the program, in the case of digital controllers) to prevent windup. We shall present a simple block diagram for anti-windup further below.

Define the <u>saturation function</u> sat_m as follows:

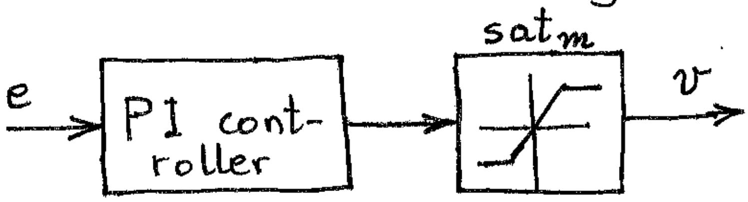
$$\mathrm{sat}_m(x) = \begin{cases} m & \text{if } x > m, \\ x & \text{if } |x| \le m, \\ -m & \text{if } x < -m, \end{cases}$$

where $m > 0$ is the saturation value, see the plot:

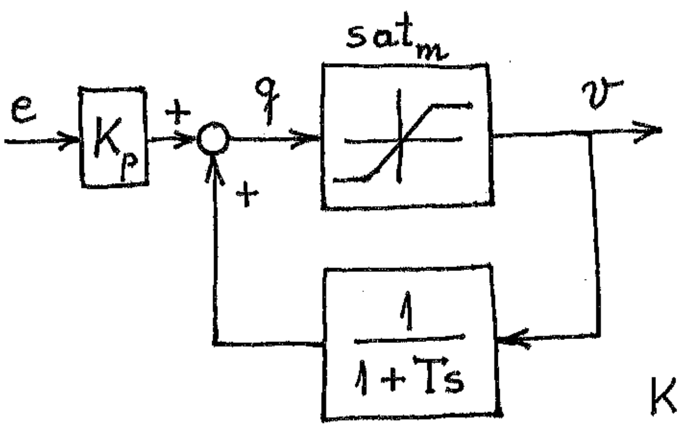


In practice, all the sensors and actuators saturate, so that a sufficiently precise mathematical model of them should contain a saturation block.

Here is a <u>bad</u> way to do anti-windup:



This system limits v, but it does not prevent the integrator from winding up. Now here is a good PI controller with anti-windup:



As long as $|q| \le m$, we have $v = q$ and the transfer function from e to v is

$$K_p \cdot \frac{1}{1 - \frac{1}{1+Ts}} = K_p\left(1 + \frac{1}{Ts}\right).$$

Clearly $|v| \le m$ (always), so the state of the low pass filter remains bounded.

Eliminating the steady-state error for ramp and for sinusoidal inputs

Sometimes it is required to achieve

$$e_{ss} = \lim_{t\to\infty} e(t) = 0$$

when r and d are of the form

$$r(t) = at + b .$$

For this, it will be enough to solve the problem corresponding to $r(t) = t$ (this is called a ramp input) and $d = 0$. (Check that indeed it is enough to consider $r(t) = t$ and $d = 0$, as an exercise.)

$\hat{r}(s) = \frac{1}{s^2}$

By the final value theorem,

$$e_{ss} = \lim_{s\to 0} s\hat{e}(s) = \lim_{s\to 0} s\, S(s) \frac{1}{s^2}$$

To obtain $e_{ss} = 0$, S should have a double zero at $s=0$. Assuming that $P(0) \neq 0$, this is possible if C has a double pole at $s=0$. Thus, <u>to eliminate the steady-state error for a ramp input, C should have a double (or higher) pole at $s=0$.</u> If $P(0)=0$, then the steady state error for ramp input cannot be eliminated (check in detail why not).

Now consider the problem to achieve $e_{ss}=0$ when r and d are of the form

$$r(t) = R\cos(\omega_0 t + \psi)$$

where $\omega_0 > 0$ is fixed and R, ψ can be any real numbers, not known in advance. For this, it will be enough to solve the problem corresponding to $r(t) = \cos\omega_0 t$ and $d=0$ (check that it is indeed sufficient to consider this particular case). According to what we said on p.7 of WEEK 1, we have

$$e(t) = A_{\omega_0}\cos(\omega_0 t + \varphi_{\omega_0}) + \varepsilon(t),$$

where $\lim_{t\to\infty}\varepsilon(t)=0$ (this transient component was denoted by e in WEEK 1), and

$$A_{\omega_0} = |S(i\omega_0)|, \qquad \varphi_{\omega_0} = \arg S(i\omega_0).$$

Hence, to obtain $\lim_{t\to\infty} e(t) = 0$, we need $S(i\omega_0)=0$. For this, C should have a pole at $i\omega_0$. By symmetry, C must have a pole also at $-i\omega_0$. It will not work if $P(i\omega_0)=0$.

For study-group

Example. If $P(s) = \dfrac{k}{1+Ts}$ and $C(s) = \dfrac{Ks}{s^2+\omega_0^2}$ with k, T, K, ω_0 all positive, then the feedback system of P and C is stable (check using p.5 of WEEK 1) and $e_{ss}=0$ for r, d sinusoidal signals of angular frequency ω_0. Check the details!