Chapter 4

SISO IMC DESIGN FOR STABLE SYSTEMS

The IMC design procedure consists of two steps

STEP 1: Nominal Performance

$\tilde{q}$ is selected to yield a "good" system response for the input(s) of interest, without regard for constraints and model uncertainty.

STEP 2: Robust Stability and Performance

$\tilde{q}$ is augmented by a low-pass filter f ($q = \tilde{q}f$) to achieve robust stability and robust performance.

We will proceed with a detailed discussion of these two steps.

4.1 Nominal Performance

For SISO systems we will generally choose $\tilde{q}$ such that it is H_2 optimal for a particular input v where $v = d$ or $v = -r$. Thus $\tilde{q}$ has to solve

$$\min_{\tilde{q}} \|e\|_2 = \min_{\tilde{q}} \|(1 - \tilde{p}\tilde{q})v\|_2 \qquad (4.1-1)$$

subject to the constraint that $\tilde{q}$ is stable and causal. Problem (4.1–1) reaches its absolute minimum (zero) for

$$\tilde{q} = \frac{1}{\tilde{p}} \qquad (4.1-2)$$

However, the model inverse is an acceptable solution only for MP systems. For NMP systems the exact inverse (4.1–2) is unstable and/or noncausal. The objective function cannot be made zero and an "approximate inverse" of $\tilde{p}$ has to be found such that the weighted 2-norm of the sensitivity function is minimized. Note that for NMP systems the optimal solution depends on the weight (the input v), while for MP systems (4.1–2) is optimal *independent* of the weight.

4.1.1 H_2-Optimal Control

In the next chapter we will derive a fairly simple general analytic procedure to solve the problem (4.1–1). In this chapter we will just state the solution and demonstrate the results for some important special cases.

Theorem 4.1-1. *Assume that $\tilde{p}$ is stable. Factor $\tilde{p}$ into an allpass portion $\tilde{p}_A$ and a MP portion $\tilde{p}_M$.*

$$\tilde{p} = \tilde{p}_A \tilde{p}_M \tag{4.1-3}$$

so that $\tilde{p}_A$ includes all the RHP zeros and delays of $\tilde{p}$ and

$$|\tilde{p}_A(i\omega)| = 1 \qquad \forall \omega \tag{4.1-4}$$

In general $\tilde{p}_A$ has the form

$$\tilde{p}_A(s) = e^{-s\theta} \prod_i \frac{-s + \zeta_i}{s + \zeta_i^H} \qquad \mathrm{Re}(\zeta_i),\ \theta > 0 \tag{4.1-5}$$

where the superscript H denotes complex conjugate.

Factor the input v similarly

$$v = v_A v_M \tag{4.1-6}$$

The controller $\tilde{q}$ which solves (4.1–1) is given by

$$\tilde{q} = (\tilde{p}_M v_M)^{-1} \left\{ \tilde{p}_A^{-1} v_M \right\}_* \tag{4.1-7}$$

where the operator $\{\cdot\}_*$ denotes that after a partial fraction expansion of the operand all terms involving the poles of $\tilde{p}_A^{-1}$ are omitted.

In general the optimal controller $\tilde{q}$ is not proper. This is of no concern at this point because it is to be augmented by a low-pass filter in Step 2 of the IMC design procedure.

Note that the objective function (4.1–1) and the optimal controller (4.1–7) are unaffected by NMP elements in v. For example, for the design of $\tilde{q}$ and the optimal closed-loop performance it makes no difference if a step disturbance enters directly at the plant output or after passing through a time delay. We will learn in Chap.6 that this is exactly the situation when feedforward control can be used effectively.

The analytic expression (4.1–7) allows us to make some general observations regarding the pole-zero configuration of the closed-loop system with the H_2-optimal controller $\tilde{q}$.

Theorem 4.1-2. *The H_2-optimal complementary sensitivity function*

$$\bar{\eta} = \tilde{p}\tilde{q} = \tilde{p}_A v_M^{-1} \left\{ \tilde{p}_A^{-1} v_M \right\}_* \tag{4.1-8}$$

has the following properties:

1. *The poles of $\bar{\eta}$ are at the mirror images of the plant RHP zeros and at the zeros of v_M.*

2. *$\bar{\eta}$ has RHP zeros at the plant RHP zeros. If $\bar{\eta}$ has any additional RHP zeros they are at mirror images of zeros of v_M.*

3. *If v_M has no finite zeros, then $\bar{\eta}$ has no RHP zeros except those of the plant. This implies that $\tilde{q}$ is MP.*

Proof.

1. The potential poles of $\bar{\eta}$ are the poles of $\tilde{p}_A$, the zeros of v_M and the poles of $\{\tilde{p}_A^{-1} v_M\}_*$, which are the poles of v_M. However, the poles of v_M are cancelled by zeros of v_M^{-1} premultiplying $\{\cdot\}_*$.

2. Assume that $\tilde{q}_1$ is stable and minimizes

$$\|(1 - \tilde{p}\tilde{q}_1)v\|_2 \qquad (4.1-9)$$

The RHP zeros of $\tilde{p}$ appear in $\bar{\eta} = \tilde{p}\tilde{q}_1$ in the form of an allpass. Define a new plant $\hat{p} = \tilde{p}\tilde{q}_1$. The stable $\hat{q}_2$ which minimizes

$$\|(1 - \hat{p}\hat{q}_2)v\|_2 \qquad (4.1-10)$$

is unity (if it were not, $\tilde{q}_1$ could not be optimal). The RHP zeros of $\hat{p}$ have to appear in $\hat{\eta} = \hat{p}\hat{q}_2 = \hat{p} = \bar{\eta}$ in the form of an allpass. Because of 1. all RHP zeros in addition to the plant RHP zeros have to be located at the mirror images of the zeros of v_M.

3. Follows directly from 2. □

RHP zeros are generally considered as "bad" for performance. However, 2. implies that depending on the input v_M the optimal controller $\tilde{q}$ can sometimes *add* RHP zeros.

Example 4.1-1. $\tilde{p} = \tilde{p}_A = \frac{-s+2}{s+2}, v_M = \frac{s+1}{s}$

$$\bar{q} = \frac{s}{s+1}\left\{\frac{s+2}{-s+2}\frac{s+1}{s}\right\}_* = \frac{s}{s+1}\left\{-1 + \frac{1}{s} + \frac{6}{-s+2}\right\}_* = \frac{-s+1}{s+1}$$

$$\tilde{\eta} = \frac{(-s+2)(-s+1)}{(s+2)(s+1)} \qquad (4.1-11)$$

The controller $\tilde{q}$ added a RHP zero at (+1,0). As predicted by Thm. 4.1-2 it is at the mirror image of the disturbance zero at (-1,0). □

Table 4.1-1. H_2–optimal controller $\tilde{q}$ for some typical input forms ($p = \tilde{p}$).

Input v_M	Controller $\tilde{q}$	Compl. Sensitivity $\bar{\eta}$
$\frac{1}{s}$	p_M^{-1}	p_A
$\frac{1}{\tau s+1}$	$p_M^{-1} \cdot p_A^{-1}\vert_{s=-\frac{1}{\tau}}$	$p_A \cdot p_A^{-1}\vert_{s=-\frac{1}{\tau}}$
$\frac{1}{s(\tau s+1)}$	$p_M^{-1}\left[1+(1-p_A^{-1}\vert_{s=-\frac{1}{\tau}})\tau s\right]$	$p_A\left[1+(1-p_A^{-1}\vert_{s=-\frac{1}{\tau}})\tau s\right]$
$\frac{1}{s^2}$	$p_M^{-1}(1-s\cdot\frac{dp_A}{ds}\vert_{s=0})$	$p_A(1-s\cdot\frac{dp_A}{ds}\vert_{s=0})$

4.1.2 Design for Specific Input Forms

The evaluation of (4.1–7) for specific inputs v_M yields the results shown in Table 4.1-1.

Consider, for example, the special case of a step input $v_M = s^{-1}$

$$\tilde{q} = p_M^{-1} s \left\{p_A^{-1} s^{-1}\right\}_* = p_M^{-1} s p_A^{-1}(0) s^{-1} = p_M^{-1} \qquad (4.1-12)$$

which agrees with the corresponding entry in Table 4.1-1. Consider next the input $v_M = (\tau s+1)^{-1}$

$$\tilde{q} = p_M^{-1}(\tau s+1)\left\{p_A^{-1}(\tau s+1)^{-1}\right\}_* = p_M^{-1}(\tau s+1)\cdot p_A^{-1}\vert_{s=-\frac{1}{\tau}} \cdot (\tau s+1)^{-1}$$

$$= \left[p_M \cdot p_A\vert_{s=-\frac{1}{\tau}}\right]^{-1} \qquad (4.1-13)$$

This result also agrees with Table 4.1-1. The derivation of $\tilde{q}$ for the other inputs is left as an exercise.

Tables 4.1-2 and 4.1-3 illustrate the application of the general formulas to the specific cases of a system with delay and a system with a single RHP zero. The following observations can be made:

1. For the ISE to be bounded, the error has to vanish as $t \to \infty$. This implies that the optimal controllers for A and C have to generate a Type 1 system. The optimal controller for D has to lead to a Type 2 system. Indeed we find for A and C.

Type 1:

$$\lim_{s\to 0} p\tilde{q} = 1$$

and for D

Type 2:

$$\lim_{s\to 0} p\tilde{q} = 1 \quad \text{and} \quad \lim_{s\to 0} \frac{d}{ds}(p\tilde{q}) = 0$$

2. From the plots of $|\bar{e}|$ it is evident that the performance of the optimal controllers can be quite bad for inputs other than the one they are designed for. In particular, high frequency inputs tend to be amplified. For cases A and B the amplification is less or equal to 2, for cases C and D it increases with frequency.

3. For "sluggish" inputs (cases C and D) the high performance is paid for with poor robustness as the large values of $|\bar{\eta}|$ demonstrate.

4.1.3 Minimum Error Norm for Step Inputs

We learned that for MP systems the H_2-norm of the error can be made arbitrarily small. For stable NMP systems subjected to steps, a simple formula can be derived to compute the minimum error norm. First, we will show that the contributions of different NMP elements to the ISE are additive. Then, we will compute the ISE arising from a single RHP zero.

Theorem 4.1-3. *Let the plant be given by $p = p_1 p_2 p_M$ where both p_1 and p_2 are allpass. Then for step inputs*

$$\min_{\tilde{q}} \|e\|_2^2 = \|(1 - p_1 p_2)s^{-1}\|_2^2 = \|(1 - p_1)s^{-1}\|_2^2 + \|(1 - p_2)s^{-1}\|_2^2 \qquad (4.1-14)$$

The proof of this theorem can be found in Sec. 5.2.3.

If $p_A = e^{-s\theta}$ then the effect of any control action is not felt before $t = \theta$. The optimal $\tilde{q}$ reduces the error to zero for $t > \theta$. Thus, for a unit step input

$$\min_{\tilde{q}} \|e\|_2^2 = \theta \qquad (4.1-15)$$

For RHP zeros the minimum error norm can be found via the calculus of residues.

Theorem 4.1-4. *Let $p_A = \frac{-s+\zeta}{s+\zeta^H}$ then for a unit step input*

$$\min_{\tilde{q}} \|e\|_2^2 = \frac{2\mathrm{Re}(\zeta)}{|\zeta|^2} \qquad (4.1-16)$$

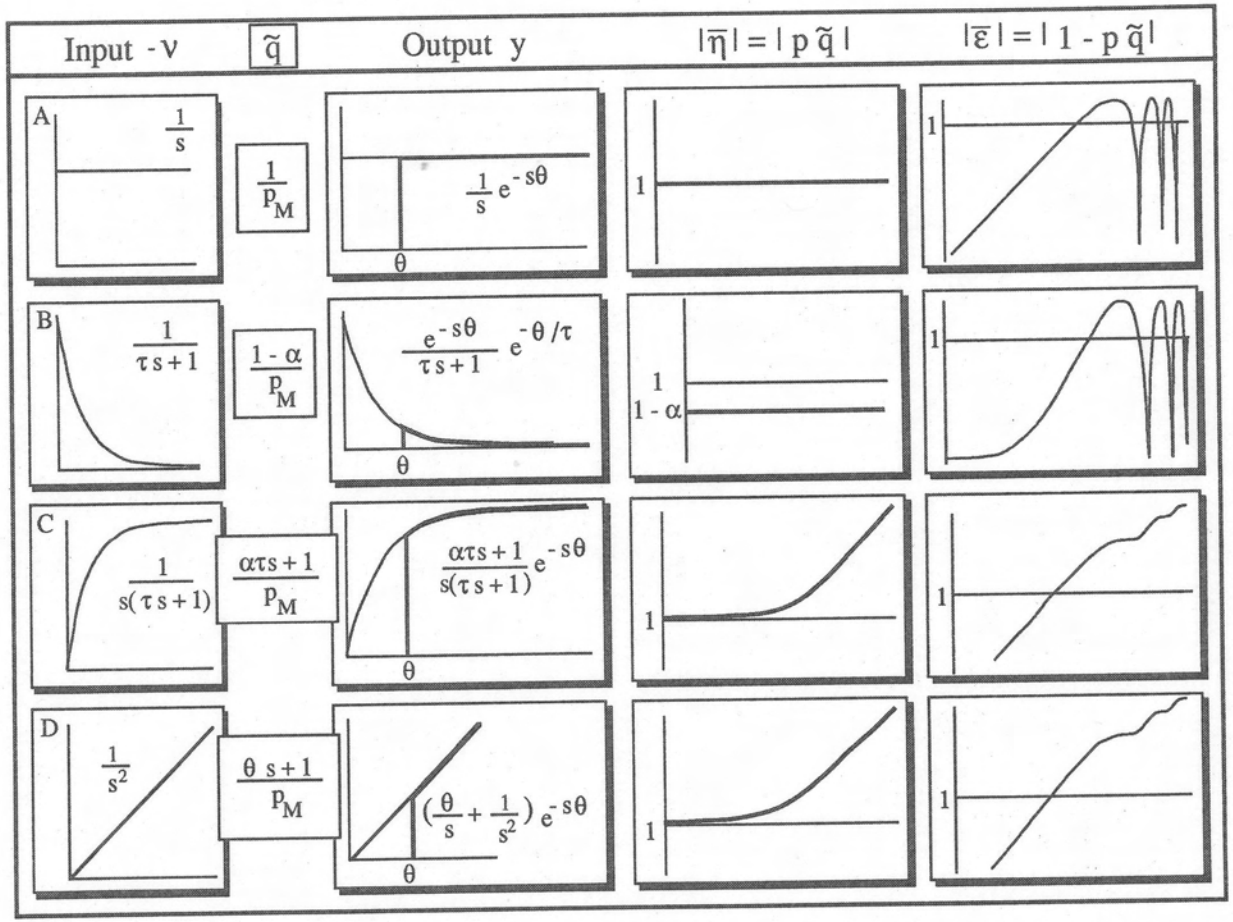


Table 4.1-2. Optimal controller $\tilde{q}$, output y, sensitivity $\bar{\varepsilon}$ and complementary sensitivity $\bar{\eta}$ for different inputs v. The system is $p = \tilde{p} = p_M e^{-s\theta}$. $\alpha = 1 - e^{-\theta/\tau}$.

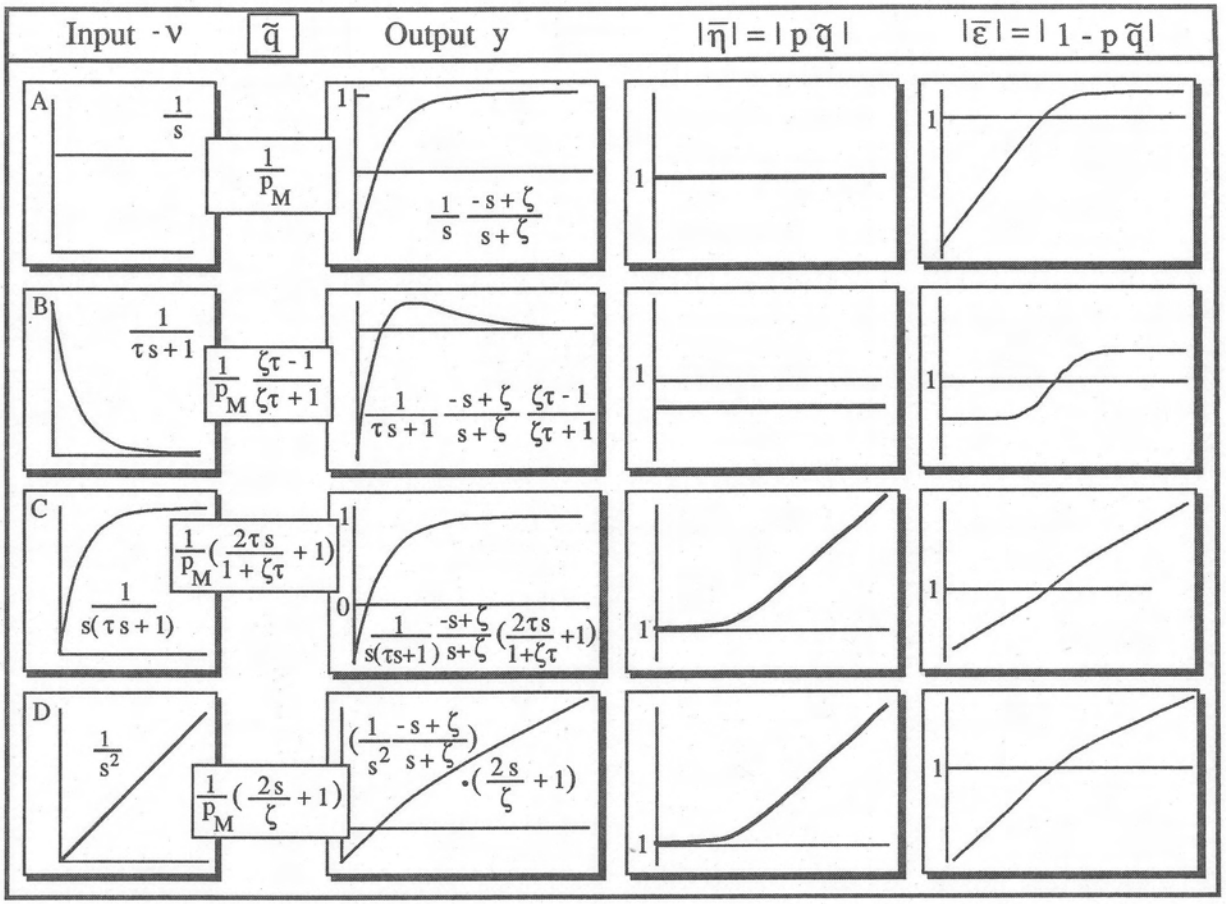


Table 4.1-3. Optimal controller $\tilde{q}$, output y, sensitivity $\bar{\varepsilon}$ and complementary sensitivity $\bar{\eta}$ for different inputs v. The system is $\tilde{p} = p = p_M \frac{-s+\zeta}{s+\zeta}$.

Proof.

$$\|e\|_2^2 = \frac{1}{2\pi i}\int_{-i\infty}^{+i\infty} |(1-p_A)s^{-1}|^2\, ds = \frac{1}{2\pi i}\int_{-i\infty}^{+i\infty} \left|1-\frac{-s+\zeta}{s+\zeta^H}\right|^2 |s^{-1}|^2 ds$$

$$= \frac{1}{2\pi i}\int_{-i\infty}^{+i\infty} \left[2-\frac{-s+\zeta}{s+\zeta^H}-\frac{s+\zeta^H}{-s+\zeta}\right] |s^{-1}|^2 ds$$

$$= \text{Res}\left[\frac{s+\zeta^H}{-s+\zeta}\frac{1}{s^2}\right]\Big|_{s=\zeta} = \frac{2\text{Re}(\zeta)}{|\zeta|^2}$$

□

Theorems 4.1-3 and 4.1-4 together with (4.1–15) allow one to easily compute the minimum ISE ($\min_{\tilde{q}} \|e\|_2^2$) for any stable system subjected to step inputs.

It is interesting to note that the detrimental effect of RHP zeros is inversely proportional to their distance from the origin. Zeros close to the origin are very bad for closed-loop performance, zeros which are far away from the origin have a negligible effect.

4.2 The IMC Filter

4.2.1 Filter Form

In general, the complementary sensitivity functions $\bar{\eta}$ obtained for the H_2-optimal controller $\tilde{q}$ exhibit undesirable high frequency behavior (see Tables 4.1-2 and 4.1-3). For robustness $\tilde{q}$ has to be augmented by a low-pass filter f ($q = \tilde{q}f$). In principle both the structure and the parameters of f should be determined such that an optimal compromise between performance and robustness is reached. To simplify the design task we fix the filter structure and search over a small number of filter parameters (usually just one) to obtain desired robustness characteristics.

It is logical to choose f such that the closed-loop system retains its asymptotic tracking properties (Type m). For systems of Type m, f has to satisfy (3.3–10).

$$\lim_{s\to 0}\frac{d^k}{ds^k}(1-\tilde{p}\tilde{q}f) = 0 \quad 0 \le k < m \qquad (4.2-1)$$

If $\tilde{q}$ is designed to satisfy (4.2–1) for $f = 1$ then the conditions on f for (4.2–1) to be satisfied are

$$\lim_{s\to 0}\frac{d^k}{ds^k}(1-f) = 0 \quad 0 \le k < m \qquad (4.2-2)$$

Thus

Type 1:

$$f(0) = 1 \qquad (4.2-3)$$

Type 2:

$$f(0) = 1 \quad \text{and} \quad \lim_{s \to 0} \frac{df}{ds} = 0 \tag{4.2-4}$$

Typically we will use one-parameter filters with unity steady-state gain of the form

$$f = (\beta_{m-1} s^{m-1} + \ldots \beta_1 s + 1) \frac{1}{(\lambda s + 1)^n} \tag{4.2-5}$$

where λ is the adjustable filter parameter, n is selected large enough to make q proper and β_i is chosen to satisfy (4.2–2). The simplest filters of the form (4.2–5) which satisfy (4.2–2) are

Type 1:

$$f(s) = \frac{1}{(\lambda s + 1)^n} \tag{4.2-6}$$

Type 2:

$$f(s) = \frac{n\lambda s + 1}{(\lambda s + 1)^n} \tag{4.2-7}$$

4.2.2 A Qualitative Interpretation of the Function of the Filter

With the filter f and the H_2-optimal $\tilde{q}$ the complementary sensitivity function becomes

$$\tilde{\eta} = \frac{y}{r} = \tilde{p}\tilde{q}f \tag{4.2-8}$$

Examples of $\bar{\eta} = \tilde{p}\tilde{q}$ for typical inputs were shown without the appropriate filter in Table 4.1-1. Note that in *all* cases the poles of $\tilde{\eta}$, which are the closed-loop system poles, are the poles of the system allpass p_A and the poles of the filter at $-\lambda^{-1}$. Adjusting λ is equivalent to adjusting the speed of the closed-loop response. When $\lambda = 0$ the system response is H_2 optimal. When $\lambda \ll \zeta_i^{-1}$ then the poles $-\zeta_i$ of the allpass will dominate the response and the filter will have little effect. When $\lambda \gg \zeta_i^{-1}$ the filter dominates the response and the closed-loop time constant becomes equal to λ. This is illustrated in Fig. 4.2-1.

Next we will describe techniques to adjust λ on the basis of model uncertainty information. However, because of its directness and intuitive appeal it is conceivable that λ is left for on-line adjustment by the operating personnel. With a good model a high speed of response can be demanded from the system. When the knowledge about the system — i.e., the model, is poor then the speed of response has to be decreased.

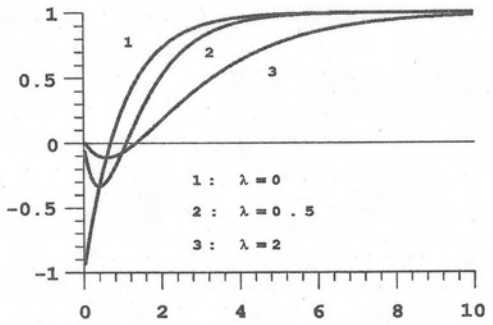


Figure 4.2-1. Step response of $\tilde{p}\tilde{q}f = (-s+1)(s+1)^{-1}(\lambda s+1)^{-1}$.

4.3 Robust Stability

4.3.1 Norm-Bounded Uncertainty Regions

Theorem 2.5-1 states that the closed-loop system is robustly stable if and only if

$$\|\tilde{\eta}\bar{\ell}_m\|_\infty = \sup_\omega |\tilde{\eta}\bar{\ell}_m(\omega)| < 1 \qquad (4.3-1)$$

For IMC $\tilde{\eta} = \tilde{p}q = \tilde{p}\tilde{q}f$ and (4.3–1) becomes

$$|\tilde{p}\tilde{q}f\bar{\ell}_m(\omega)| < 1 \qquad \forall\omega \qquad (4.3-2)$$

Corollary 4.3-1. *Assume that the family of stable plants* Π *is described by*

$$\Pi = \left\{p : \left|\frac{p(i\omega)-\tilde{p}(i\omega)}{\tilde{p}(i\omega)}\right| < \bar{\ell}_m(\omega)\right\} \qquad (4.3-3)$$

and that $\tilde{q}$ *is stable. Then the closed-loop system is robustly stable if and only if the IMC filter satisfies*

$$|f| < \frac{1}{|\tilde{p}\tilde{q}\bar{\ell}_m|} \qquad \forall\omega \qquad (4.3-4)$$

An f can always be found such that (4.3–4) is satisfied. However $|f|$ small implies $|\tilde{\eta}|$ small and thus poor performance. If performance requirements have to be met, the uncertainty expressed through $\bar{\ell}_m$ has to be limited for robust stability to be possible. As a simple performance specification let us require the closed-loop system to be of Type 1 — i.e., $\tilde{p}(0)\tilde{q}(0) = f(0) = 1$. Then we find directly from (4.3–4) the following corollary.

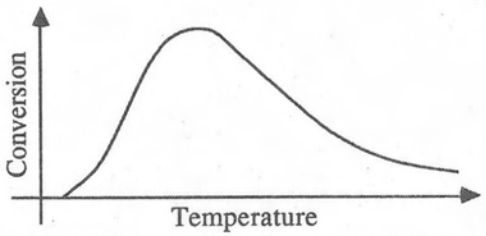


Figure 4.3-1. Conversion vs. Temperature for CSTR with exothermic reversible reaction.

Corollary 4.3-2. *Assume $\bar{\ell}_m(\omega)$ is continuous. There exists a filter f such that the closed-loop system is Type 1 and robustly stable for the family Π described by (4.3-3) if and only if $\bar{\ell}_m(0) < 1$.*

Corollary 4.3-2 implies that it is possible to design a feedback controller which guarantees a zero tracking error for steps if and only if the process steady-state gain uncertainty satisfies

$$\left|\frac{p(0) - \tilde{p}(0)}{\tilde{p}(0)}\right| < 1 \qquad (4.3-5)$$

i.e., if the gain error does not exceed 100% or equivalently if the gain does not change sign. When the gain changes sign a positive feedback loop results and the system becomes unstable.

Example 4.3-1. Consider an exothermic continuous flow stirred tank reactor (CSTR) where conversion is related to reactor temperature as shown in Fig. 4.3-1. Assume that conversion is controlled by adjusting the reactor temperature. The steady-state gain of the linearized reactor model is equal to the slope of the curve in Fig. 4.3-1 and depends on the operating point. It is positive for low temperatures and negative for high temperatures. According to Cor. 4.3-2 it is not possible to design a controller for Type 1 performance which is stable over the whole operating range. □

The uncertainty bound $\bar{\ell}_m$ generally increases with frequency. How small $|f|$ has to be made to satisfy (4.3–4) for robust stability depends on the H_2-optimal complementary sensitivity function for the nominal system $|\tilde{p}\tilde{q}|$ and the uncertainty bound $\bar{\ell}_m$. Tables 4.1-2 and 4.1-3 show plots of $|\tilde{p}\tilde{q}|$ with the optimal controllers $\tilde{q}$ designed for different inputs. For the same uncertainty, cases C and D require more drastic filter action than A and B because $|\bar{\eta}|$ increases with frequency.

From another point of view we can say that designing controllers specifically

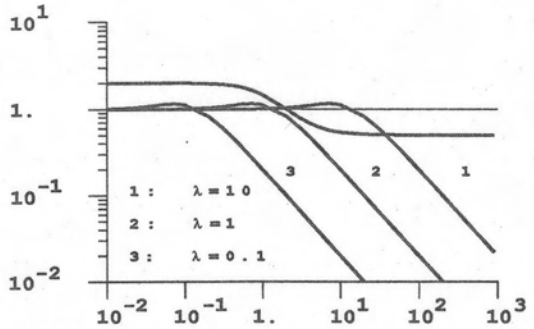


Figure 4.3-2. Robust stability bound (4.3–4) and Type 2 filter f for different values of the filter time constant λ.

for slowly rising inputs (C and D) might not bring any performance advantages over designing controllers for steps (A) unless the model uncertainty is sufficiently small such that large values of $|\tilde{\eta}|$ can be tolerated at high frequencies.

The filter designer chooses first the appropriate form of the filter for Type 1 (4.2–6) or Type 2 (4.2–7) behavior and then increases the filter time constant until (4.3–4) is satisfied (Fig. 4.3-2). In some cases it can be prudent to adjust the filter order n to meet the constraint (4.3–4) with the least conservatism.

4.3.2 General Uncertainty Regions

The controller c is to be designed to guarantee the stability of a family Π of plants surrounding a nominal plant $\tilde{p}$. All members of the family Π are assumed to be open-loop stable. At each frequency ω the magnitude and phase of the plant lies in a region $\pi(\omega)$ on the Nyquist plane. The union of the regions forms a *Nyquist band* and constitutes the family Π.

At each frequency ω the controller $c(i\omega)$ is a complex number. The multiplication of a region with a number $(\pi(\omega)c(i\omega))$ stretches and rotates the region (Fig. 4.3-3). These operations can be carried out easily numerically and the resulting band $\pi(\omega)c(i\omega)$ can be displayed on a graphics screen.

Theorem 4.3-1. *The closed-loop system is robustly stable for the family* Π *if and only if the Nyquist band* $\pi(\omega)c(i\omega)$ *does not encircle or cover (-1,0).*

For the controller c the IMC parametrization is employed

$$c = \frac{\tilde{q}f}{1 - \tilde{p}\tilde{q}f} \qquad (4.3-6)$$

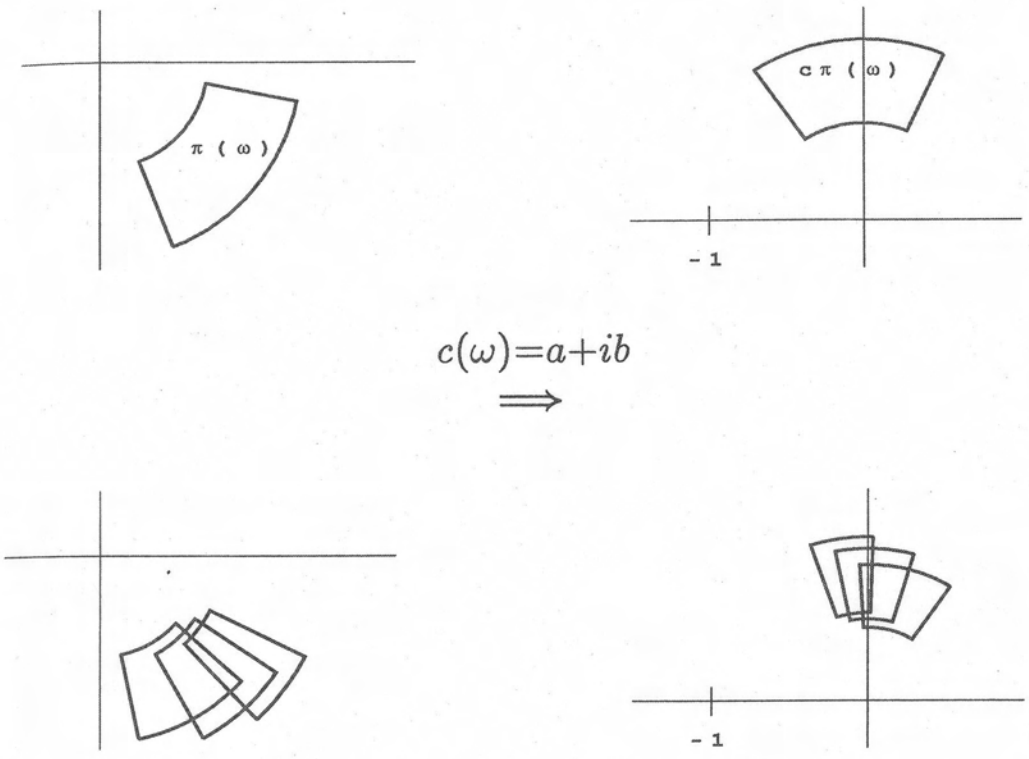


Figure 4.3-3. Transformation of region $\pi(\omega)$ by multiplication with complex number $c(i\omega)$.

It is intuitively obvious and it also follows from our discussions in the previous section that there always exists some filter (e.g., the trivial case of $f = 0$) for which robust stability can be achieved. In order for the closed-loop system to be robustly stable and of Type 1 the family of steady-state gains $\pi(0)$ has to be restricted. The following corollary follows directly from Cor. 4.3-2 and the subsequent discussion.

Corollary 4.3-3. *A controller c exists such that the closed loop system is Type 1 and robustly stable for the family Π if and only if all the plant gains in $\pi(0)$ have the same sign.*

Because the nominal plant $\tilde{p}$ appears in the controller c it has a great effect on the closed stability and performance of the family Π. In Sec. 4.3.1. $\tilde{p}$ was naturally defined as the center of the disk shaped uncertainty regions. For the general uncertainty region considered in this section there is no natural center and $\tilde{p}$ has to be selected judiciously. For example, $\tilde{p}(0)$ should be chosen as the center of the segment $\pi(0)$ of the real line. If the conditions of Cor. 4.3-3 are satisfied and if the IMC parametrization (4.3–6) is used for this $\tilde{p}$ then the existence of a Type 1 filter (4.2–6) is guaranteed for which the system will be robustly stable. This follows directly from Cor. 4.3-2.

The advantage of using general uncertainty regions is that tighter stability bounds can be established than when the regions are approximated by disks. The disadvantage is the need for a graphical rather than an analytic/numerical procedure.

4.4 Robust Performance

4.4.1 Norm-Bounded Uncertainty Regions

Theorem 2.6-1 states the following necessary and sufficient condition for robust performance

$$|\tilde{\eta}\bar{\ell}_m| + |\tilde{\epsilon}w| < 1 \quad \forall\omega \tag{4.4-1}$$

For IMC $\tilde{\eta} = \tilde{p}\tilde{q}f$ and the following corollary results.

Corollary 4.4-1. *Assume that the family Π of stable plants is described by*

$$\Pi = \left\{ p : \left| \frac{p(i\omega) - \tilde{p}(i\omega)}{\tilde{p}(i\omega)} \right| < \bar{\ell}_m(\omega) \right\} \tag{4.4-2}$$

Then the closed loop system will meet the performance specification

$$\|\epsilon w\|_\infty = \sup_\omega |\epsilon w| < 1 \qquad \forall p \in \Pi \tag{4.4-3}$$

if and only if

$$|\tilde{p}\tilde{q}f\bar{\ell}_m| + |(1 - \tilde{p}\tilde{q}f)w| < 1 \qquad \forall\omega \qquad (4.4-4)$$

Increasing the filter time constant λ will tend to decrease the first term and increase the second term in (4.4–4). Thus, depending on $\bar{\ell}_m$ and w there might not be any λ for which (4.4–4) is satisfied. On the other hand, assume that the controller/filter is selected such that the system is of Type 1 or higher ($\tilde{p}\tilde{q}f(0) = 1$) and that $\bar{\ell}_m(0) < 1$. Then "robust performance at $\omega = 0$" can be achieved for any w.

Corollary 4.4-2. *There exists a filter f such that the closed-loop system exhibits "robust performance at $\omega = 0$" for any weight w if and only if $\bar{\ell}_m(0) < 1$.*

Corollary 4.4-2 simply implies that if the system is robustly stable with a controller c with integral action (Type 1) then the steady-state performance is perfect even when there is modelling error.

For IMC the shapes of $\tilde{\epsilon}$ and $\tilde{\eta}$ are essentially fixed by the H_2-optimal controller $\tilde{q}$ and the filter f which has only a single adjustable parameter. With only one degree of freedom available in the controller it is not always meaningful to define a very complex performance weight which has to be satisfied. Thus, we often select w to be a constant. This implies that the filter parameter has to be selected such that the maximum peak of the sensitivity function does not exceed w^{-1} for all plants in the uncertainty set. Because this objective is generally satisfied for the trivial choice $\lambda = \infty$ — i.e., open loop — the following simple filter design procedure is proposed.

1. Check that (4.4–4) can be satisfied at $\omega = 0$ for $f(0) = 1$ and $w = 0$. This guarantees the existence of a filter for Type 1 performance and robust stability.

2. Select a constant performance weight w and adjust the filter time constant so that (4.4–4) is met as an equality. (In the "worst case" the maximum peak of ϵ is exactly equal to w^{-1}.) Some typical iterations in the design procedure are illustrated in Fig. 4.4-1.

For general nonconstant weights w it is possible that (4.4–4) cannot be satisfied for any λ. Then either the filter structure is too simplistic to meet the performance specifications or the specifications are too tight for the uncertainty which is present. After the filter parameter λ has been determined a bound on the error norm can be established (see Sec. 2.6) for all plants in the family Π subjected to a specific input v.

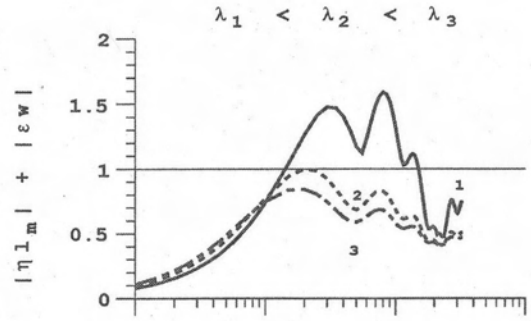


Figure 4.4-1. Robust performance bound (4.4–4) for different values of the filter time constant λ.

Though in the context of process control it is rarely justified to use anything more complicated than a one-parameter filter (4.2–6 or 4.2–7) our design procedure is in no way limited to this case. More complex filters have the potential of satisfying the robust stability constraint with a smaller loss of performance but at the expense of a more complicated search procedure for the appropriate filter parameters.

Also, in general, w can be selected in a manner which reflects the performance specifications more accurately than is possible with a constant w. It is often meaningful to base w on the H_2-optimal $\bar{\epsilon}$ obtained for the typical system input:

$$w = \frac{\alpha}{|\bar{\epsilon}|} \qquad \alpha > 0 \tag{4.4-5}$$

The parameter α is selected by the designer and can be interpreted as follows. If the robust performance objective (4.4–3) with the weight (4.4–5) is met then

$$|\epsilon| < \frac{|\bar{\epsilon}|}{\alpha} \qquad \forall\omega, \quad \forall p \in \Pi \tag{4.4-6}$$

and for a specific input v

$$\|e\|_2^2 = \frac{1}{2\pi}\int_{-\infty}^{\infty} |\epsilon|^2|v|^2 d\omega$$

$$\leq \frac{1}{\alpha^2}\frac{1}{2\pi}\int_{-\infty}^{\infty} |\bar{\epsilon}|^2|v|^2 d\omega \qquad \forall p \in \Pi \tag{4.4-7}$$

Equation (4.4–7) implies that for all plants $p \in \Pi$ the error norm exceeds the minimum error norm by not more than a factor α^{-1}.

The choice (4.4–5) is not always practical. For example, for a system with a delay, w would be unbounded at nonzero frequencies (see Table 4.1-2) and the

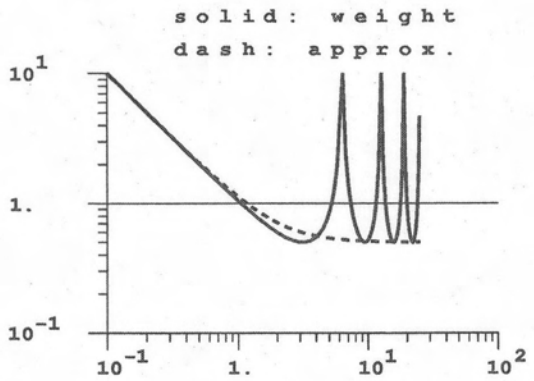


Figure 4.4-2. Weight and approximate weight for a system with time delay subject to a step input.

robust performance specifications could not be met by any controller. In this case w can be chosen as a reasonable approximation of (4.4–5). Let us assume that we are interested in a system with time delay subjected to step inputs — i.e., $\bar{\epsilon} = 1 - e^{-\theta s}$. Then a meaningful weight can be derived from (4.4–5) when a Padé approximation is used for the time delay in $\bar{\epsilon}$.

$$w = \alpha \left| 1 - \frac{-\theta s + 2}{\theta s + 2} \right|^{-1} = \alpha \left| \frac{\theta s + 2}{2\theta s} \right| \qquad (4.4-8)$$

The effect of the approximation is illustrated in Fig. 4.4-2.

4.4.2 General Uncertainty Regions

The controller c is to be designed to guarantee robust stability and robust performance for the family Π described by the Nyquist band discussed in Sec. 4.3.2. For robust performance it is required that

$$|\epsilon w| < 1 \qquad \forall \omega, \quad \forall p \in \Pi \qquad (4.4-9)$$

or

$$|1 + \pi c| > w \qquad \forall \omega \qquad (4.4-10)$$

where $\pi c(i\omega)$ denotes the *Nyquist region* at frequency ω which was constructed in Sec. 4.3.2 by multiplying the region $\pi(\omega)$ with the complex number $c(i\omega)$. For robust stability the Nyquist band πc must not encircle or cover (-1,0), for robust performance the distance of (all points in) πc from the point (-1,0), i.e., $|1 + \pi c|$, has to exceed the specified minimum w.

Theorem 4.4-1. *The closed-loop system satisfies the robust performance condition*

$$|\epsilon w| < 1 \qquad \forall \omega, \quad \forall p \in \Pi \qquad (4.4-9)$$

if and only if it is robustly stable for the family Π *and the Nyquist band* $\pi(\omega)c(\omega)$ *retains a distance of at least* $w(\omega)$ *from the point (-1,0).*

The controller c is parametrized by the IMC filter (4.3–6). As a design procedure w is generally selected to be constant and the regions $\pi(\omega)c(\omega)$ are displayed on a graphics screen for different values of the filter time constant λ. Then λ is increased just enough to make the minimum distance from (-1,0) equal to w.

4.5 Summary of IMC Design Procedure

Required Information:

1. Process model $\tilde{p}$
2. Type of inputs (setpoints and disturbances) v affecting the process output — e.g., steps, ramps, steps entering through first-order lag, and so on.
3. Performance specifications:
   (i) Closed-loop system Type (1 or 2)
   (ii) Frequency dependent performance weight or in the case of a one-parameter filter simply the maximum allowed peak height (w^{-1}) of the sensitivity function ϵ (typically $0.3 < w < 0.9$).
4. Uncertainty information $\bar{\ell}_m(\omega)$.
   Family of plants considered for robust design

$$\Pi = \left\{ p : \left| \frac{p(i\omega) - \tilde{p}(i\omega)}{\tilde{p}(i\omega)} \right| \leq \bar{\ell}_m(\omega) \right\}$$

Design Procedure:

STEP 1: Nominal Performance
The H_2-optimal controller $\tilde{q}$ is determined which minimizes

$$\|(1 - \tilde{p}\tilde{q})v\|_2$$

for the specified input v. If the input v is of a standard type, $\tilde{q}$ can be obtained directly from Table 4.1-1, 4.1-2, or 4.1-3. Otherwise Thm. 4.1-1 can be used.

$$\tilde{q} = (\tilde{p}_M d_M)^{-1} \left\{ \tilde{p}_A^{-1} d_M \right\}_* \qquad (4.1-7)$$

STEP 2: Robust Stability and Robust Performance

The controller $\tilde{q}$ is augmented by the IMC filter f

$$q = \tilde{q}f$$

where for asymptotically constant inputs

Type 1:

$$f = \frac{1}{(\lambda s + 1)^n} \qquad (4.2-6)$$

and for asymptotically ramp-like inputs

Type 2:

$$f = \frac{n\lambda s + 1}{(\lambda s + 1)^n} \qquad (4.2-7)$$

Here n is selected large enough for q to be proper. For MP systems and in the absence of modeling error λ is the closed-loop time constant. For NMP systems λ becomes the dominant closed-loop time constant when it is made large enough. In general, increasing λ slows down the system and makes it more robust.

Robust Stability: Check if

$$|\tilde{p}\tilde{q}f\bar{\ell}_m| < 1 \qquad \text{for } \omega = 0 \qquad (4.5-1)$$

This condition is necessary and sufficient for a filter time constant $\lambda > 0$ to exist for which the system is robustly stable.

Robust Performance: Increase λ just enough to meet the condition

$$|\tilde{p}\tilde{q}f\bar{\ell}_m| + |(1 - \tilde{p}\tilde{q}f)w| < 1 \qquad \forall\omega \qquad (4.4-4)$$

i.e., choose λ to make (4.4–4) an equality for some specific value(s) of ω.

4.6 Application: IMC Design for a First-Order Deadtime System

This particular example system was chosen because of its relevance for process control. We will demonstrate the control system design for both the exact uncertainty regions and for norm-bounded uncertainties — i.e., disk-shaped approximations of the exact uncertainty regions.

4.6.1 Deadtime Uncertainty

We will follow exactly the summary of the IMC Design Procedure in Sec. 4.5.

Information:

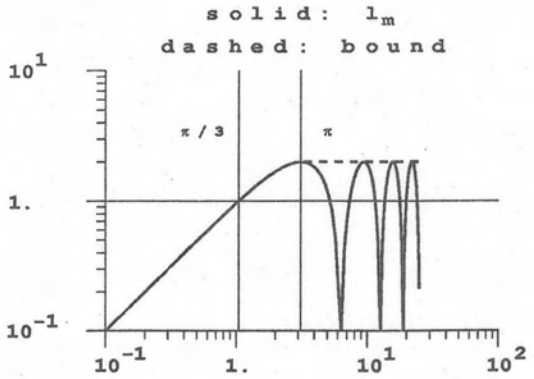


Figure 4.6-1. Multiplicative uncertainty ℓ_m for $\delta = \bar{\delta}$ and bound $\bar{\ell}_m$ versus $\omega\bar{\delta}$.

1. Process model $\tilde{p} = \frac{ke^{-s\theta}}{\tau s+1}$
2. Type of input: step
3. Performance specification
   (i) Type 1 (no offset for steps)
   (ii) Maximum peak of ϵ: $w^{-1} = 2.5$
4. Uncertainty information: Only deadtime uncertainty

$$\theta = \tilde{\theta} + \delta \quad \text{where} \quad |\delta| \leq \bar{\delta} \tag{4.6-1}$$

$$\frac{p - \tilde{p}}{\tilde{p}} = e^{-s\delta} - 1 \tag{4.6-2}$$

$$\ell_m(i\omega) = e^{-i\omega\delta} - 1 \qquad |\delta| \leq \bar{\delta} \tag{4.6-3}$$

Figure 4.6-1 shows the plot of $|\ell_m(i\omega)|$ for $\delta = \bar{\delta}$. It is easily seen that

$$\bar{\ell}_m(\omega) = |e^{-i\omega\bar{\delta}} - 1| \qquad \omega\bar{\delta} \leq \pi \tag{4.6-4a}$$

$$\bar{\ell}_m(\omega) = 2 \qquad \omega\bar{\delta} \geq \pi \tag{4.6-4b}$$

A plot of the bound $\bar{\ell}_m(\omega)$ is also shown in Fig. 4.6-1. Note the characteristic increase with frequency.

Design Procedure

STEP 1: Nominal Performance

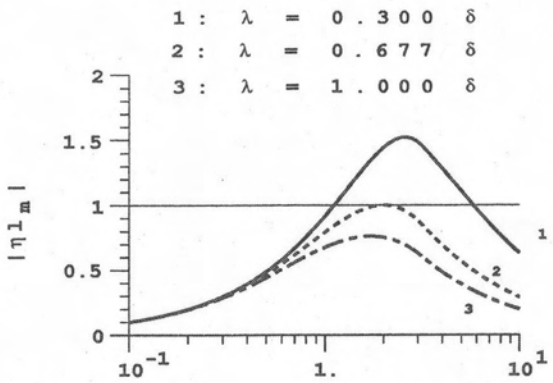


Figure 4.6-2. Robust stability bound $|\tilde{\eta}\bar{\ell}_m|$ for different values of the filter parameter λ versus $\omega\bar{\delta}$.

For step inputs the H_2-optimal controller is provided by Thm. 4.1-1 or case A in Tab. 4.1-2.

$$\tilde{p}_A = e^{-s\tilde{\theta}} \tag{4.6-5}$$

$$\tilde{q} = \tilde{p}_M^{-1} = k^{-1}(\tau s + 1) \tag{4.6-6}$$

$$\tilde{p}\tilde{q} = e^{-s\theta} \tag{4.6-7}$$

STEP 2: Robust Stability and Robust Performance

Type 1 filter:

$$f = \frac{1}{\lambda s + 1} \tag{4.6-8}$$

A first-order filter is sufficient to make q proper.

Robust Stability:

$$|\tilde{p}\tilde{q}f\bar{\ell}_m| = |f\bar{\ell}_m| < 1 \tag{4.6-9}$$

Figure 4.6-1 immediately suggests a filter with a corner frequency $\omega = \pi/3\bar{\delta} \cong 1/\bar{\delta}$ or $\lambda = \bar{\delta}$. A more careful search reveals $\lambda > 0.67\bar{\delta}$ for robust stability (Fig. 4.6-2)

Robust Performance:

The constraint

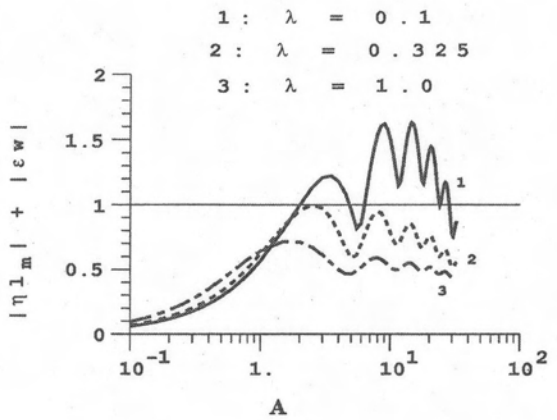
$$|\tilde{p}\tilde{q}f\bar{\ell}_m| + |(1 - \tilde{p}\tilde{q}f)w| < 1$$

or

$$|f\bar{\ell}_m| + |(1 - e^{-i\omega\tilde{\theta}}f)w| < 1 \tag{4.6-10}$$

is shown in Fig. 4.6-3A and B for a 15% and 30% maximum deadtime error respectively ($\tilde{\theta} = 1$, $\bar{\delta} = 0.15$ and 0.3).

Note from (4.6–9) and (4.6–4) that λ for robust stability depends only on the deadtime error $\bar{\delta}$ but not on the absolute value of the deadtime $\tilde{\theta}$. For robust



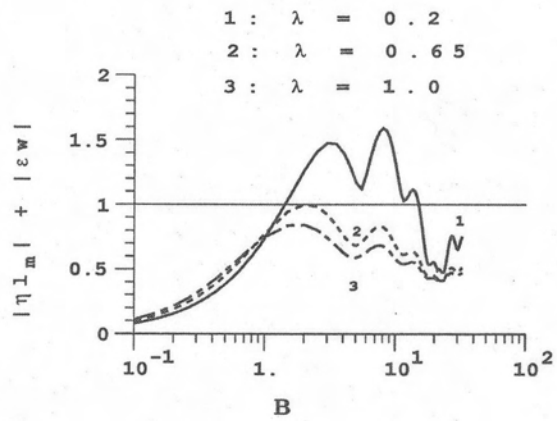


Figure 4.6-3. Robust performance test for 15% (A) and 30% (B) deadtime uncertainty for different filter parameters.

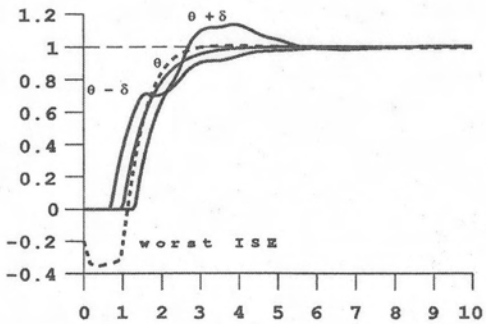


Figure 4.6-4. System response for $\lambda = 0.65$ and various deadtimes. "Worst-ISE" response is also shown.

performance (4.6–10) both $\bar{\delta}$ and $\tilde{\theta}$ have to be taken into consideration. The parameters which are known exactly, k and τ, do not affect the filter design. Typical closed-loop responses with $\lambda = 0.65$ chosen for robust performance are shown in Fig. 4.6-4. Also shown is the response of the plant contained in the set bounded by (4.6–4) which gives rise to the largest error norm. According to the arguments in Sec. 2.6 the frequency response of this plant is determined by locating at each frequency the point in the uncertainty disk around $\tilde{p}c$ which is closest to (-1,0) — i.e., the point p which satisfies

$$|1 + pc| = |1 + \tilde{p}c| - |\tilde{p}\bar{\ell}_m| \qquad (4.6-11)$$

It is intuitively clear that with pure deadtime uncertainty the worst error should occur for the plant with the maximum deadtime. The "worst-ISE" response shown in Fig. 4.6-4 illustrates the conservativeness arising from the norm-bound approximation of the deadtime uncertainty.

Next we will compute the filter parameters for robust stability and robust performance when we account for the deadtime uncertainty exactly. Figure 4.6-5A shows the Nyquist band where each region $\pi(\omega)$ is just a line for the case of pure deadtime uncertainty. Figures 4.6-5B and C show the transformed band $\pi(\omega)c(i\omega)$ with

$$c = \frac{\tilde{q}f}{1 - \tilde{p}\tilde{q}f} = \frac{s+1}{\lambda s + 1 - e^{-s}} \qquad (4.6-12)$$

Figure 4.6-5B shows the Nyquist band for the minimum λ required for robust stability. In Fig. 4.6-5C λ is chosen for robust performance. In Fig. 4.6-6 some simulations for different values of the uncertain parameter θ with $\lambda = 0.5$ are shown. Table 4.6-1 compares the values of λ for robust stability and performance found by the two procedures.

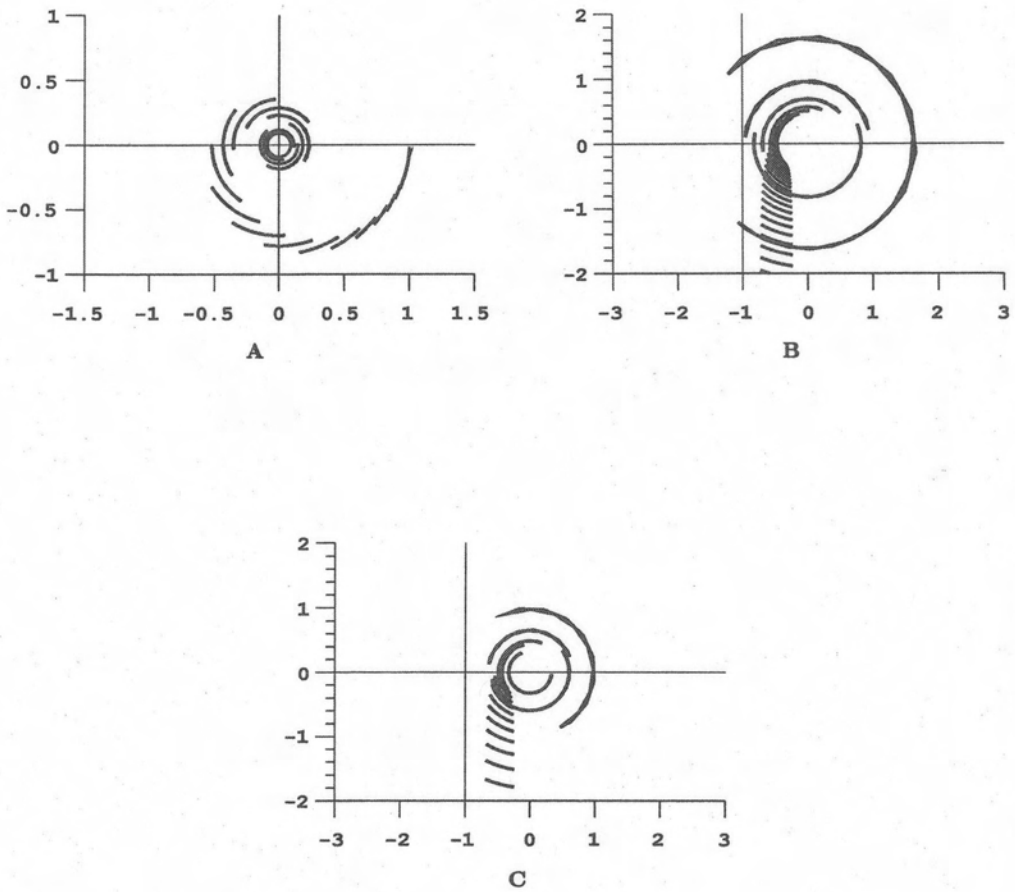


Figure 4.6-5. Nyquist bands for system (A), system with controller tuned for robust stability (B) and system with controller tuned for robust performance (C).

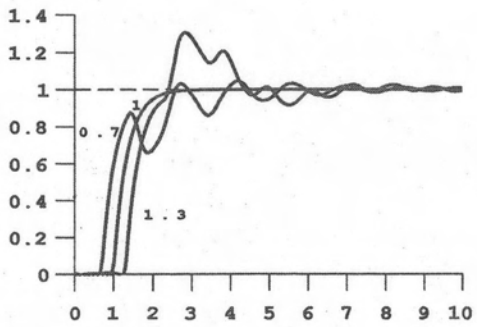


Figure 4.6-6. Response of closed loop system tuned for robust performance for various deadtimes.

Table 4.6-1. Filter parameter λ for robust stability and robust performance when the deadtime uncertainty ($\bar{\delta} = 0.3$) is treated exactly and when it is approximated by a norm-bound.

	Stability	Performance $w^{-1} = 2.5$	Performance $w^{-1} = 2.0$
Exact	0.20	0.34	0.50
Approximate	0.20	0.65	0.99

From this and other examples it can be concluded that the choice $\lambda = \bar{\delta}$ is a good rule of thumb which leads to good performance in practice.

4.6.2 Three-Parameter Uncertainty

The problem is identical with the one treated in the last section except that $w^{-1} = 2.0$ and all three parameters in the model

$$\tilde{p} = \frac{\tilde{k}e^{-s\tilde{\theta}}}{\tilde{\tau}s + 1}$$

are uncertain

$$0.8 \leq k \leq 1.2$$
$$0.7 \leq \tau \leq 1.3 \qquad (4.6-13)$$

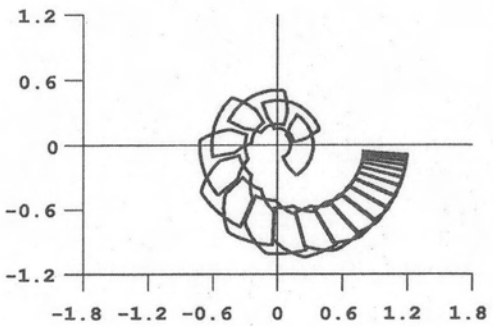


Figure 4.6-7. Nyquist band for system (4.6–13).

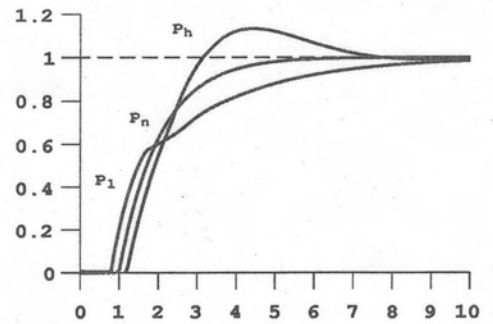


Figure 4.6-8. Response of closed-loop system tuned for robust performance for various system parameters.

$$0.8 \leq \theta \leq 1.2$$

The Nyquist band can be seen in Fig. 4.6-7. With the controller (4.6–12) the filter time constants for robust stability and performance are 0.21 and 1.03, respectively. Simulations with representative sample parameters from the set (4.6–13) are shown in Fig. 4.6-8.

Alternatively we can approximate the regions in Fig. 4.6-7 by disks centered at the nominal plant $\tilde{p}$ defined by the mean parameter values $\tilde{k}, \tilde{\tau}$ and $\tilde{\theta}$. The disk radii translate into the multiplicative uncertainty bound $\bar{\ell}_m(\omega)$ in Fig. 4.6-9. Going through a procedure equivalent to that demonstrated in the previous section, we find that for robust stability λ has to exceed 0.23 and for robust performance λ is 1.51. These values can be seen to be slightly more conservative than those obtained with the exact uncertainty description (Tab. 4.6-2).

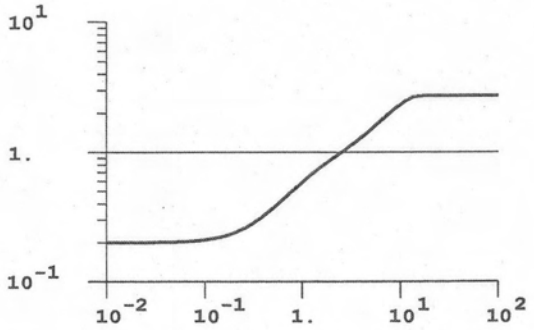


Figure 4.6-9. Multiplicative uncertainty bound $\bar{\ell}_m$ generated from Nyquist band in Fig. 4.6-7.

Table 4.6-2. Filter parameter λ for robust stability and robust performance when the uncertainty is treated exactly and when it is approximated by a norm bound.

	Stability	Performance
Exact	0.21	1.03
Approximate	0.23	1.51

4.7 References

4.1.1. The expression (4.1–7) for the H_2-optimal controller was derived by Frank (1974) who follows the work by Newton, Gould & Kaiser (1957).

4.1.3. Theorem 4.1-2 was obtained from Frank (1974).

4.3.2. Several researchers have proposed controller-design techniques using arbitrarily shaped regions $\pi(\omega)$ to represent model uncertainty. The procedure described here was developed by Laughlin et al. (1986) and is very similar to that outlined by Chen (1984). Laughlin et al. review other techniques.

All example calculations in this chapter were carried out by D. Laughlin and C. Scali.

Chapter 5

SISO IMC DESIGN FOR UNSTABLE SYSTEMS

There are two main challenges in extending IMC to open-loop unstable systems. First, the parametrization of all controllers c, for which the closed-loop system is stable, in terms of the IMC controller q is not straightforward: stability of q does not guarantee closed-loop stability when p is unstable. Second, designing the controller for robust stability is significantly more difficult. When p is stable we can simply detune the controller (even to the degree of almost turning it off) to make it arbitrarily robust. When p is unstable there is a limit to how much we can detune the controller; some minimal controller gain is needed just to stabilize the system.

5.1 Parametrization of All Stabilizing Controllers

5.1.1 Conditions for Internal Stability

For stable systems the IMC parametrization offers two advantages. The search for the controller which yields good closed-loop behavior is greatly simplified: if the search is carried out over all possible stable and unstable feedback controllers c, closed-loop stability has to be checked for each choice of c. On the other hand, any stable IMC controller q automatically yields a feedback controller c for which the closed-loop system is stable. Furthermore, if the IMC structure is also used for the controller implementation, closed-loop stability is preserved in the presence of constraints and controller nonlinearities, contrary to what happens for the classic feedback structure.

As we move on to unstable systems we have to abandon the IMC structure for the control system *implementation*. This is apparent from Thm. 3.2-1 where we established that stability of both p and q is required for internal stability. It is also obvious from Fig. 3.2-1. If $p = \tilde{p}$, then the input signal u_2 does not generate any feedback signal and, therefore, no control action u. The signals $\tilde{y}$ and $y = pu_2$

grow without bound for an unstable plant.

On the other hand even for unstable systems we can use the IMC parametrization for the *design* of the feedback controller c. Let us establish the conditions q must satisfy for the closed-loop system with

$$c = \frac{q}{1 - \tilde{p}q} \qquad (5.1-1)$$

to be internally stable. Substituting (5.1–1) into (2.3–1) we find for $p = \tilde{p}$

$$\begin{pmatrix} y \\ u \end{pmatrix} = \begin{pmatrix} pq & (1-pq)p \\ q & -pq \end{pmatrix} \begin{pmatrix} r \\ u' \end{pmatrix} \qquad (5.1-2)$$

For internal stability all four individual transfer functions in (5.1–2) have to be stable. This implies:

(i) q stable.

(ii) pq stable $\Rightarrow$ q has to cancel all RHP poles of p or equivalently: q has to have zeros wherever p has unstable poles.

(iii) $(1-pq)p$ stable $\Rightarrow$ $(1-pq)$ has to cancel all RHP poles of p or equivalently: $(1-pq)$ has to have zeros wherever p has unstable poles.

Now assume that (iii) holds but (ii) does not. Then pq is unbounded at the unstable poles of p. If pq is unbounded, $(1-pq)$ is unbounded at the unstable poles. This contradicts the assumption that $(1-pq) = 0$ at the unstable poles of p (iii). Thus (iii) implies (ii). These arguments are summarized in the following theorem.

Theorem 5.1-1. *Assume that the model is perfect ($p = \tilde{p}$) and that p has k unstable poles at $\pi_1, \ldots, \pi_k$. The feedback system with the controller $c = q(1-pq)^{-1}$ is internally stable if and only if*

(i) q is stable

(ii) $(1-pq)$ has k RHP zeros at $\pi_i, \ldots, \pi_k$.

Note that Thm. 5.1-1 reduces to Thm. 3.2-1 when p is stable. Condition (ii) is quite instructive: recall that $\tilde{\eta} = \tilde{p}q$ is the complementary sensitivity function of the system which should be small for robustness. For unstable systems the sensitivity function cannot be chosen arbitrarily, but it has to satisfy (ii) – i.e., it must have a value of unity at all RHP plant poles; the achievable robustness is *constrained* by the RHP plant poles. We will elaborate further on this issue later in this chapter.

5.1.2 Controller parametrization

IMC controllers q which satisfy condition (ii) of Thm. 5.1-1 are not particularly easy to find. Thus, at this point the IMC parametrization does not offer any advantages. In this section we will derive a convenient expression for *all* q's satisfying (ii).

Define the allpass

$$b_p(s) = \prod_{i=1}^{k} \frac{-s + \pi_i}{s + \pi_i^H} \qquad (5.1-3)$$

where π_i are the poles of $\tilde{p}$ in the open RHP.

Theorem 5.1-2. *Assume that $p = \tilde{p}$ has k unstable poles at $\pi_1, \ldots, \pi_k$ in the open RHP and ℓ poles at the origin. Assume furthermore that there exists a q_0 such that $c = q_0(1 - pq_0)^{-1}$ stabilizes p. Then all controllers which stabilize p are parametrized by*

$$c = q(1 - pq)^{-1} \qquad (5.1-1)$$

$$q = q_0 + b_p^2 s^{2\ell} q_1 \qquad (5.1-4)$$

where q_1 is any arbitrary stable transfer function.

Proof.$\Rightarrow$ Assume that q and q_0 stabilize p. Then Thm. 5.1-1 implies that q, q_0, $(1 - pq)b_p^{-1}s^{-\ell}$, and $(1 - pq_0)b_p^{-1}s^{-\ell}$ are stable. Thus the differences $q - q_0$ and $p(q - q_0)b_p^{-1}s^{-\ell}$ are stable. Define

$$pb_p s^{\ell}(q - q_0)b_p^{-2}s^{-2\ell} \triangleq pb_p s^{\ell} q_1 \qquad (5.1-5)$$

By construction (5.1–5) is stable. Therefore any RHP poles of q_1 are a subset of the RHP zeros of p. Because $q - q_0$ is stable, any RHP poles of $q_1 = (q - q_0)b_p^{-2}s^{-2\ell}$ must be a subset of the RHP poles of p. Because p cannot have RHP poles and zeros at the same location, q_1 is stable.

$\Leftarrow$ Assume q_0 stabilizes p and q_1 is stable. Then by Thm. 5.1-1, $q_0, q_1, (1 - pq_0)b_p^{-1}s^{-\ell}$ are stable. Also

$$q = q_0 + b_p^2 s^{2\ell} q_1$$

is stable. Then

$$(1 - pq)b_p^{-1}s^{-\ell} = (1 - pq_0)b_p^{-1}s^{-\ell} - pb_p^2 s^{2\ell} q_1 b_p^{-1}s^{-\ell} = (1 - pq_0)b_p^{-1}s^{-\ell} - pb_p s^{\ell} q_1$$

is stable. Therefore q is stabilizing. □

Note that Thm. 5.1-2 assumes the existence of a stabilizing $q_0(s)$. The construction of the H_2-optimal controller in Sec. 5.2.1 will serve as a proof that such a controller always exists.

Any q constructed from (5.1–4) automatically satisfies the conditions of Thm. 5.1-1. After an initial stabilizing controller q_0 has been found the search for the appropriate controller q which yields desirable closed-loop performance characteristics is again very simple. All stable transfer functions q_1 yield controllers for which the closed-loop system is stable. There are also no disadvantages to using (5.1–4): there is *no* controller c which stabilizes p and which is not parametrized by a stable q_1.

In the special case of an open-loop stable system $b_p = 1$ and $\ell = 0$. The controller $q_0 = 0$ is stabilizing and (5.1–4) yields $q = q_1$, the IMC-parametrization for stable systems, as expected.

Note that q_1 in (5.1–4) is not limited to be proper and q can be improper in general. Because in Step 2 of the IMC design procedure a filter is always added for properness and robustness we are not concerned with properness at this point and hence improper controllers are permitted.

5.2 Nominal Performance

The controller design philosophy for unstable systems is the same as for stable ones. First a controller $\tilde{q}$ is selected to minimize the ISE or equivalently the 2-norm of the sensitivity function weighted by the input of interest. Then $\tilde{q}$ is augmented by a low-pass filter to achieve robust stability and performance. In this section we will derive the formulas for the H_2-optimal $\tilde{q}$.

5.2.1 H_2-Optimal Controller

Theorem 5.2-1. *Let $p = \tilde{p}$ have k poles at $\pi_1, \ldots, \pi_k$ in the open RHP and a pole of multiplicity ℓ at the origin. Define*

$$b_p = \prod_{i=1}^{k} \frac{-s + \pi_i}{s + \pi_i^H} \qquad (5.2-1)$$

and factor the plant into an allpass portion p_A and a MP portion p_M such that

$$p = p_A p_M \qquad (5.2-2)$$

Factor the input v similarly

$$v = v_A v_M \tag{5.2-3}$$

Assume without loss of generality that the open RHP poles of the input v are the first k' poles π_i of the plant in the open RHP[1] *and define accordingly*

$$b_v = \prod_{i=1}^{k'} \frac{-s + \pi_i}{s + \pi_i^H} \tag{5.2-4}$$

Assume further that v has at least ℓ poles at the origin.[2] *The controller $\tilde{q}$ which minimizes the objective*

$$\|e\|_2^2 = \int_0^\infty e^2 dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |e|^2 d\omega = \|(1 - p\tilde{q})v\|_2^2 \tag{5.2-5}$$

is given by

$$\tilde{q} = b_p (p_M b_v v_M)^{-1} \left\{ (b_p p_A)^{-1} b_v v_M \right\}_* \tag{5.2-6}$$

where the operator $\{\cdot\}_*$ denotes that after a partial fraction expansion of the operand all terms involving the poles of p_A^{-1} are omitted.

The following proof of (5.2–6) is somewhat involved and can be skipped by result-oriented readers without loss of continuity.

Proof. Before proceeding with the proof we have to establish some preliminary definitions and results. Let $L_2(i\mathcal{R})$ be the Hilbert space of scalar-valued functions on $i\mathcal{R}$ with inner product

$$< f, g >= \frac{1}{2\pi} \int_{-\infty}^{\infty} f(i\omega)^H g(i\omega) d\omega \tag{5.2-7}$$

for which $< f, f >$ is finite.

Let $H_2(i\mathcal{R})$ be the subspace of functions $f(s)$ analytic in $\mathrm{Re}(s) > 0$. The orthogonal complement of $H_2(i\mathcal{R})$, denoted by $H_2^\perp(i\mathcal{R})$, is the subspace of functions analytic for $\mathrm{Re}(s) < 0$. Thus, if $f_1 \in H_2(i\mathcal{R})$ and $f_2 \in H_2^\perp(i\mathcal{R})$ then

$$< f_1, f_2 >= \frac{1}{2\pi} \int_{-\infty}^{\infty} f_1(i\omega)^H f_2(i\omega) d\omega = 0 \tag{5.2-8}$$

(Equation (5.2–8) can be easily established via the calculus of residues).

Any $f(s) \in L_2$ is analytic on the imaginary axis. Thus any $f(s)$ can be decomposed uniquely into two orthogonal functions $\{f\}_+ \in H_2$ and $\{f\}_- \in H_2^\perp$

$$f(s) = \{f\}_+ + \{f\}_- \tag{5.2-9}$$

[1] If this assumption were not made the problem would not be meaningful: Unbounded controller action would be necessary for the error to vanish as $t \to \infty$. See also the discussion at the end of Sec. 2.2.3.

[2] This assumption is made so that disturbances occurring at the plant *input* can be rejected with vanishing error as $t \to \infty$.

where $\{\cdot\}_+(\{\cdot\}_-)$ denotes that after a partial fraction expansion only the terms with poles in the LHP (RHP) are retained. The 2-norm of f induced by the inner product (5.2–7) can be computed as

$$\|f\|_2^2 =< f, f >=< \{f\}_- + \{f\}_+, \{f\}_- + \{f\}_+ >=< \{f\}_-, \{f\}_- >$$
$$+ < \{f\}_+, \{f\}_+ >= \|\{f\}_-\|_2^2 + \|\{f\}_+\|_2^2 \qquad (5.2-10)$$

where the cross terms $< \{f\}_+, \{f\}_- >=< \{f\}_-, \{f\}_+ >= 0$ because of orthogonality.

According to Thm. 5.1-2 all stabilizing controllers of p are parametrized by

$$\tilde{q} = q_0 + b_p^2 s^{2\ell} q_1 \qquad (5.2-11)$$

where q_1 is an arbitrary stable transfer function and q_0 is a stabilizing controller — i.e., it is chosen to be stable and such that $(1 - pq_0)$ has zeros at the unstable poles of the plant p. We will also assume that q_0 is such that $(1 - pq_0)v$ is stable. Not every stabilizing q_0 may have this property, since v may have more poles at the origin than p.

These properties of q_0 can be summarized by assuming that q_0 is chosen stable and such that $(1 - pq_0)$ has zeros at the unstable poles of $b_p^{-1} b_v v_M$ (here b_v has been included to avoid duplication of the k' unstable poles which are present in both p and v). The final construction of the H_2-optimal q serves as proof for the existence of a q_0 with such properties.

By substituting (5.2–11) into (5.2–5) the objective function which is to be minimized by the appropriate choice of a stable q_1 becomes

$$\|(1 - p\tilde{q})v\|_2^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} |1 - p(q_0 + b_p^2 s^{2\ell} q_1)|^2 |v|^2 d\omega$$

The integrand can be multiplied by the allpass $|b_p p_A b_v^{-1} v_A|^{-2} = 1$ without changing the value of the integral:

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} |b_p^{-1} p_A^{-1} (1 - pq_0) b_v v_M - p_M b_p s^{2\ell} q_1 b_v v_M|^2 d\omega$$
$$\triangleq \frac{1}{2\pi} \int_{-\infty}^{\infty} |f_1 - f_2 q_1|^2 d\omega = \|f_1 - f_2 q_1\|_2^2 \qquad (5.2-12)$$

The optimal q_1 should yield a finite integral (5.2–12), which means that the integrand, $f_1 - f_2 q_1$, must be strictly proper and not have any poles at the origin. By inspection, if $f_2 q_1$ has any unstable poles, they are at the origin. Since all unstable poles of f_1 (the poles of p_A^{-1}) are in the *open* RHP, it cannot cancel any poles at the origin of $f_2 q_1$. Thus $f_2 q_1$ must be stable in order for q_1 to be optimal.

We will assume that q_1 has this property and verify later that our solution has indeed the property.

Let us now consider the requirement that the optimal q_1 must make $f_1 - f_2q_1$ strictly proper. One can easily see that a constant term may be present in the partial fraction expansion (PFE) of f_1, while both constant and improper terms may be present in the PFE of f_2q_1. The constant term in the PFE of f_1 is equal to $\lim_{s\to\infty} f_1(s) \triangleq f_1(\infty)$. Then in order for q_1 to make $f_1 - f_2q_1$ strictly proper, q_1 must make f_2q_1 proper and satisfy

$$\lim_{s\to\infty}(f_2q_1) \triangleq (f_2q_1)(\infty) = f_1(\infty) \qquad (5.2-13)$$

We can now write (5.2-12) as

$$\|(1-p\tilde{q})v\|_2^2 = \|\hat{f}_1 - (f_2q_1 - f_1(\infty))\|_2^2 \qquad (5.2-14)$$

where $\hat{f}_1(s) \triangleq f_1(s) - f_1(\infty)$ is an L_2 function, which can be decomposed according to (5.2–9)

$$\hat{f}_1 = \{\hat{f}_1\}_+ + \{\hat{f}_1\}_-$$

Since the optimal q_1 makes f_2q_1 proper and satisfies (5.2–13), $(f_2q_1 - f_1(\infty))$ is in H_2. Use of this fact and the decomposition of $\hat{f}_1$ yields after application of (5.2–10) to (5.2–14):

$$\|(1-p\tilde{q}v\|_2^2 = \|\{\hat{f}_1\}_-\|_2^2 + \|\{\hat{f}_1\}_+ - (f_2q_1 - f_1(\infty))\|_2^2$$

Since $\{\hat{f}_1\}_-$ is independent of q_1, the obvious solution q_1 which minimizes $\|(1 - p\tilde{q})v\|_2^2$ can be obtained from the condition

$$\|\{\hat{f}_1\}_+ - (f_2q_1 - f_1(\infty))\|_2 = 0$$

or

$$f_2q_1 = \{\hat{f}_1\}_+ + f_1(\infty) \triangleq \{f_1\}_{\infty+}$$

or

$$q_1 = (p_M b_p s^{2\ell} b_v v_M)^{-1}\{b_p^{-1}p_A^{-1}(1-pq_0)b_v v_M\}_{\infty+}$$

where the notation $\{\cdot\}_{\infty+}$ is used to indicate that after a PFE is taken, all strictly proper stable terms as well as the constant term are kept.

Now we have to check that this obvious solution is permissible. Because, by assumption, v_M has at least ℓ poles at the origin, q_1 is stable as required. Also, simple substitution shows that f_2q_1 is proper and stable and that q_1 satisfies (5.2–13). Thus the above obvious solution is indeed the optimal solution. Substituting the expression for q_1 into (5.2–11) we find

$$\tilde{q} = q_0 + b_p(p_M b_v v_M)^{-1}\left\{b_p^{-1}p_A^{-1}(1-pq_0)b_v v_M\right\}_{\infty+}$$

$$= b_p(p_M b_v v_M)^{-1}[b_p^{-1}p_A^{-1}pq_0 b_v v_M - \left\{b_p^{-1}p_A^{-1}pq_0 b_v v_M\right\}_{\infty+} + \left\{b_p^{-1}p_A^{-1}b_v v_M\right\}_{\infty+}]$$

$$= b_p(p_M b_v v_M)^{-1}[\left\{b_p^{-1}p_A^{-1}pq_0 b_v v_M\right\}_{0-} + \left\{b_p^{-1}p_A^{-1}b_v v_M\right\}_{\infty+}] \qquad (5.2-15)$$

where $\{\cdot\}_{0-}$ indicates that in the partial fraction expansion all poles in the closed RHP are retained. For (5.2–15) these poles are the poles of $b_p^{-1}b_v v_M$ in the closed RHP; $p_A^{-1}pq_0 = p_M q_0$ is stable because q_0 is a stabilizing controller. Indeed, q_0 is chosen such that $pq_0 = 1$ at all unstable poles of $b_p^{-1}b_v v_M$. Thus (5.2–15) simplifies to

$$\tilde{q} = b_p(p_M b_v v_M)^{-1}\{b_p^{-1}p_A^{-1}b_v v_M\}_* \qquad (5.2-6)$$

□

Consider the optimal complementary sensitivity function

$$\tilde{\eta} = p\tilde{q} = p_A b_p(b_v v_M)^{-1}\{b_p^{-1}p_A^{-1}b_v v_M\}_* \qquad (5.2-16)$$

We note that the optimal closed-loop pole configuration depends in general on both the plant and the input characteristics. The poles of $\{b_p^{-1}p_A^{-1}b_v v_M\}_*$ are the poles of $b_p^{-1}b_v v_M$. These poles are cancelled by the zeros of $b_p(b_v v_M)^{-1}$ premultiplying $\{\cdot\}_*$ in (5.2–16). Therefore, unless the system is MP, the finite optimal closed-loop poles are the poles of $p_A b_p(b_v v_M)^{-1}$. Thus, some of the optimal closed-loop poles (the poles of $p_A b_p$) are at all the mirror images of the system RHP zeros and poles. The location of the other finite closed-loop poles depends on the zeros of the inputs.

These observations illustrate that without some fundamental insight of the kind attempted here, it is very difficult to state a priori the closed-loop pole locations which will lead to good nominal performance characteristics. We will see in the next section that, independent of RHP poles and input type, the optimal controller pushes all closed-loop poles of MP systems toward infinity. On the other hand, for nontrivial systems with several RHP zeros and poles a trial and error search for reasonable closed-loop pole locations can be a futile task.

5.2.2 Design for Common Input Forms

In this section we will interpret the general optimal solution (5.2–6) for specific systems and inputs.

i) *MP System:* $p_A = 1, p_M = p$

$$\left\{b_p^{-1} p_A^{-1} b_v v_M\right\}_* = b_p^{-1} b_v v_M$$

$$\tilde{q} = p^{-1} \qquad (5.2-17)$$

For MP systems the optimal controller $\tilde{q}$ is the system inverse regardless of whether the open-loop system is stable or not. Perfect control ($\epsilon = 0$) can be achieved for any MP system in the limit as improper controllers are used. The RHP poles impose no performance limitations for MP systems.

ii) *Stable System:* $b_p = b_v = 1$

$$\tilde{q} = (p_M v_M)^{-1} \left\{p_A^{-1} v_M\right\}_* \qquad (5.2-18)$$

This formula was stated in Thm. 4.1-1.

iii) *Integrator:* $p_M = s^{-1}$

$$\tilde{q} = (p_M v_M)^{-1} \left\{p_A^{-1} v_M\right\}_* \qquad (5.2-19)$$

This formula is the same as that for stable systems (5.2–18). Thus Tables 4.1-1, 2 and 3 can be used as well for plants which have one pole at the origin. However, some care is required in defining the appropriate input v_M. Recall that according to our convention v ($v = d$ or $v = -r$) acts at the process output (see Fig. 2.1-1B). If the external input acts at the process input (for example, u' in Fig. 2.3-1) then the equivalent disturbance acting at the process output has to include the plant pole at the origin. If there is a step disturbance (s^{-1}) at the process input and the process is a pure integrator, then $v_M = s^{-2}$, a ramp. If the process model is $p = (\tau s + 1)^{-1} s^{-1}$, then $v_M = (\tau s + 1)^{-1} s^{-2}$. Let us compute the H_2-optimal $\tilde{q}$ for this input.

$$\left\{\frac{p_A^{-1}}{s^2(\tau s+1)}\right\}_* = \frac{p_A^{-1}|_{s=0}}{s^2} + \frac{\tau^2 \cdot p_A^{-1}|_{s=-\frac{1}{\tau}}}{\tau s+1} - \left[\left.\frac{dp_A}{ds}\right|_{s=0} + \tau\right] \cdot \frac{1}{s}$$

$$\tilde{q} = p_M^{-1}\left[(\tau s+1) + \tau^2 s^2 p_A^{-1}|_{s=-\frac{1}{\tau}} - s(\tau s+1)\left[\left.\frac{dp_A}{ds}\right|_{s=0} + \tau\right)\right]$$

$$= p_M^{-1}\left[s^2\tau^2 (p_A^{-1}|_{s=-\frac{1}{\tau}} - \frac{1}{\tau} \cdot \left.\frac{dp_A}{ds}\right|_{s=0} - 1) - s \cdot \left.\frac{dp_A}{ds}\right|_{s=0} + 1\right] \qquad (5.2-20)$$

Note that for $\tau = 0$ (5.2–20) reduces to the corresponding entry in the last row of Table 4.1-1.

5.2.3 Minimum Error Norm for Step Inputs to Stable Systems

Theorem 4.1-3 states that the contributions of different NMP elements to the ISE are additive.

Theorem 4.1-3. *Let the plant be given by $p = p_1 p_2 p_M$, where both p_1 and p_2 are allpass and $p_1(0) = p_2(0) = 1$. Then for step inputs*

$$\min_{\tilde{q}} \|e\|_2^2 = \|(1 - p_1 p_2)s^{-1}\|_2^2 = \|(1 - p_1)s^{-1}\|_2^2 + \|(1 - p_2)s^{-1}\|_2^2 \qquad (5.2 - 21)$$

The proof uses similar arguments as the proof of Thm. 5.2-1.

Proof. We know from the last section that the optimal $\tilde{q} = p_M^{-1}$. Thus

$$\min_{\tilde{q}} \|e\|_2^2 = \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} |1 - p_1 p_2|^2 |s|^{-2} ds$$

$$= \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} |p_1^{-1} - p_2|^2 |s|^{-2} ds$$

$$= \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} |(p_1^{-1} - 1)s^{-1} + (1 - p_2)s^{-1}|^2 ds$$

The first term is strictly unstable, the second term strictly stable. Because of the orthogonality we can rewrite the last expression as

$$\frac{1}{2\pi i} \left[\int_{-i\infty}^{i\infty} |(p_1^{-1} - 1)s^{-1}|^2 ds + \int_{-i\infty}^{i\infty} |(1 - p_2)s^{-1}|^2 ds \right]$$

$$= \frac{1}{2\pi i} \left[\int_{-i\infty}^{i\infty} |(1 - p_1)s^{-1}|^2 ds + \int_{-i\infty}^{+i\infty} |(1 - p_2)s^{-1}|^2 ds \right]$$

□

5.2.4 Two-Degree-of-Freedom Controller

In Sec. 3.3.2 we derived for the two-degree-of-freedom controller in Fig. 3.3-1 the closed-loop relationship

$$e = (1 - pq_d)d - (1 - pq_r)r \qquad (3.3 - 7)$$

when there is no modelling error ($p = \tilde{p}$). In the case of open-loop unstable plants, the structure in Fig. 3.3-1 cannot be implemented. The feedback structure of Fig. 2.4-2 has to be used instead. Equations (3.3–6) and (3.3–7), however, still describe the relation between d, r, and e provided that

$$q_d = \frac{c_1 c_3}{1 + \tilde{p} c_1 c_3} \qquad (5.2-22)$$

$$q_r = \frac{c_1 c_2}{1 + \tilde{p} c_1 c_3} \qquad (5.2-23)$$

In this section we want to determine what limitations plant RHP poles place on disturbance rejection and setpoint tracking performance.

We found earlier (Sec. 5.2.2) that for MP systems the optimal controller is

$$\tilde{q} = p^{-1} \qquad (5.2-17)$$

which implies that perfect control ($\epsilon = 0$) can be achieved for *any* MP system (stable or unstable) in the limit as improper controllers are used.

For NMP systems the controller choice (5.2–17) is not allowed because p^{-1} is unstable. In addition to stability, q_d has to satisfy condition ii) of Thm. 5.1-1. That is, q_d has to be chosen such that $(1 - pq_d)$ has zeros at the unstable plant poles. *The unstable plant poles restrict the choice of q_d and thus they impose limitations on the disturbance rejection performance which can be achieved.*

Let us first design $\tilde{q}_d$ via (5.2–6) for $v = d$. Then we can use (5.2–22) to obtain $\hat{c} \triangleq c_1 c_3$

$$\hat{c} = \frac{\tilde{q}_d}{1 - \tilde{p}\tilde{q}_d} \qquad (5.2-24)$$

where the individual factors c_1 and c_3 are yet to be determined.

Next, let us consider the design of c_2 for setpoint tracking.

$$e = y - r = -(1 - \frac{pc_1}{1 + pc_1 c_3} c_2) r \qquad (5.2-25)$$

If we define a new plant

$$\hat{p} = \frac{pc_1}{1 + pc_1 c_3} \qquad (5.2-26)$$

and a new controller

$$\hat{q} = c_2 \qquad (5.2-27)$$

we can rewrite (5.2–25) in the familiar form

$$e = -(1 - \hat{p}\hat{q}) r \qquad (5.2-28)$$

The RHP zeros of $\hat{p}$ are the RHP zeros of p and c_1 and the RHP poles of c_3. In order to minimize the effect of RHP zeros on e, c_1 should be selected MP and

c_3 stable. Then the only RHP zeros of $\hat{p}$ are those of p. Obviously, for internal stability there must not be any RHP pole-zero cancellations between c_1 and c_3. Generally, the decomposition of $\hat{c}$ into c_1 and c_3 satisfying these requirements is not unique.

With these choices of c_1 and c_3 we can find the H_2-optimal $\hat{q}$ from (5.2–6) with $\hat{p} = \hat{p}_A \hat{p}_M$, $\hat{p}_A = p_A$ and $\hat{b}_p = b_v = 1$:

$$c_2 = \hat{q} = (\hat{p}_M r_M)^{-1} \left\{ p_A^{-1} r_M \right\}_* \qquad (5.2-29)$$

and for $d = 0$

$$y = \hat{p}\hat{q}r = p_A r_M^{-1} \left\{ p_A^{-1} r_M \right\}_* r \qquad (5.2-30)$$

which shows explicitly that the setpoint response is affected by the RHP zeros of p only. *In the two-degree-of-freedom control structure unstable plant poles impose no limitations on the nominal setpoint tracking performance.*

For special cases the number of control blocks can be reduced to two without affecting internal stability. For example, when c_3 is MP it can be combined with c_1 and c_2. Also, when c_1 is stable it can be incorporated into c_2 and c_3.

Of course, when *robust* performance is the objective the designs of q_r and q_d are not independent and the RHP plant poles affect the achievable performance for both disturbance rejection and setpoint tracking (see Secs. 3.4.2 and 5.5).

5.3 The IMC Filter

The philosophy behind the IMC filter is the same as for stable systems. The objective is robust stability and robust performance. The filter structure is fixed a priori and a few parameters (usually just one) are adjusted to meet the robustness objectives.

5.3.1 Filter Form

The IMC filter $f(s)$ is chosen to be a rational function that satisfies the following requirements:

(i) Pole-zero excess. The controller $q = \tilde{q}f$ must be proper. Assume that the designer has specified a pole-zero excess of m for the filter $f(s)$.

(ii) Internal stability, i.e., $\tilde{q}f$ and $(1 - p\tilde{q}f)p$ have to be stable.

(iii) Asymptotic tracking of inputs — i.e., $(1 - p\tilde{q}f)v$ has to be stable.

We assume that p and v have k distinct poles in the closed RHP at locations $\pi_1, \ldots, \pi_k$ (counting poles which appear in both p and v only once) and that $\tilde{q}$ has been designed such that

$$1 - p\tilde{q}(s) = 0 \quad \text{at} \quad s = \pi_1, \ldots, \pi_k \tag{5.3-1}$$

Then, for simple poles, requirements (ii) and (iii) translate to

$$f(s) = 1 \quad \text{at} \quad s = \pi_1, \ldots, \pi_k \tag{5.3-2}$$

A filter of the form

$$f(s) = \frac{a_{k-1}s^{k-1} + \ldots + a_1 s + a_0}{(\lambda s + 1)^{m+k-1}} \tag{5.3-3}$$

automatically satisfies requirement (i) as well as (with the k coefficients $a_{k-1}, \ldots, a_0$ properly chosen) requirements (ii) and (iii). Furthermore, the closed-loop system will be internally stable and will track inputs asymptotically as time $\to \infty$ for all $\lambda > 0$. This will allow the adjustable filter time constant λ to be selected freely without any concern for nominal stability.

In the following a procedure is presented to obtain the coefficients of the filter numerator. The k equalities (5.3–2) expand to the following system of linear equations:

$$\begin{pmatrix} \pi_1^{k-1} & \ldots & \pi_1 & 1 \\ \vdots & & \vdots & \vdots \\ \pi_k^{k-1} & \ldots & \pi_k & 1 \end{pmatrix} \begin{pmatrix} a_{k-1} \\ \vdots \\ a_0 \end{pmatrix} = \begin{pmatrix} (\lambda\pi_1 + 1)^{m+k-1} \\ \vdots \\ (\lambda\pi_k + 1)^{m+k-1} \end{pmatrix} \tag{5.3-4}$$

The matrix in (5.3–4) is of the Vandermonde form. Using a Lagrange-type interpolation formula it is quite simple to develop the expression for the filter explicitly. Define

$$b_j(s) = \frac{1}{(\lambda s + 1)^n} \prod_{\substack{i=1 \\ i \neq j}}^{k} (s - \pi_i), \qquad j = 1, \ldots, k \tag{5.3-5}$$

where $n = m + k - 1$ and note that

$$b_j(\pi_i) = 0 \qquad i = 1, \ldots k; \quad i \neq j \tag{5.3-6}$$

$$b_j(\pi_j) = \frac{1}{(\lambda\pi_j + 1)^n} \prod_{\substack{i=1 \\ i \neq j}}^{k} (\pi_j - \pi_i) \tag{5.3-7}$$

The stabilizing filter is easily constructed by additively combining b_j's (5.3-5) with the appropriate scaling factors (5.3–7):

$$f(s) = \frac{1}{(\lambda s + 1)^n} \sum_{j=1}^{k} (\lambda\pi_j + 1)^n \prod_{\substack{i=1 \\ i \neq j}}^{k} \frac{s - \pi_i}{\pi_j - \pi_i} \tag{5.3-8}$$

This formula is not valid for repeated poles. For a pole of multiplicity ℓ at π_i the filter has to satisfy

$$f(\pi_i) = 1 \qquad (5.3-9)$$

and

$$\frac{d^j f(s)}{ds^j}|_{s=\pi_i} = 0 \qquad j = 1, \ldots, \ell - 1 \qquad (5.3-10)$$

For each multiple pole the system of equations (5.3–4) has to be augmented by the linear equalities (5.3–10) to determine the coefficients $a_0, \ldots, a_{k-1}$. Most common are repeated poles at the origin. For k distinct poles at $\pi_1, \ldots, \pi_k$ in the open RHP and a double pole at the origin the formula (5.3–8) is modified as follows

$$f(s) = \frac{1}{(\lambda s+1)^n}\left[\sum_{j=1}^{k}(\lambda\pi_j+1)^n \frac{s^2}{\pi_j^2}\prod_{\substack{i=1\\ i\neq j}}^{k}\frac{s-\pi_i}{\pi_j-\pi_i} + (\alpha s+1)\prod_{i=1}^{k}\frac{s-\pi_i}{-\pi_i}\right] \qquad (5.3-11)$$

where

$$\alpha = \lambda n + \sum_{i=1}^{k}\pi_i^{-1} \qquad (5.3-12)$$

Example 5.3-1. Assume that p and v have a common RHP pole at $s = \pi(k = 1)$. Then the filter form for $m = 2$ is

$$f(s) = \frac{a_0}{(\lambda s+1)^2} \qquad (5.3-13)$$

and (5.3–2), or equivalently (5.3–4), becomes

$$a_0 = (\lambda\pi + 1)^2$$

resulting in the IMC filter

$$f(s) = \frac{(\lambda\pi+1)^2}{(\lambda s+1)^2} \qquad (5.3-14)$$

If $\pi = 0$, (5.3–14) reduces to the standard filter for stable systems with Type 1 performance: $f = (\lambda s + 1)^{-2}$.

Note that the filter (5.3–14) was not designed to satisfy any specific low-frequency performance requirement. This could be accomplished by imposing the additional interpolation constraint $f(0) = 1$. □

Example 5.3-2. Assume that the same pole-zero excess requirements as in Ex. 5.3-1 apply but that p and v have a double pole at the origin. The filter form is

$$f(s) = \frac{a_1 s + a_0}{(\lambda s+1)^3} \qquad (5.3-15)$$

(5.3–2) and (5.3–10) become respectively

$$\frac{a_1 \cdot 0 + a_0}{(\lambda \cdot 0 + 1)^3} = 1 \Rightarrow a_0 = 1 \qquad (5.3-16)$$

$$\frac{a_1 \cdot (\lambda \cdot 0 + 1)^3 - (a_1 \cdot 0 + a_0) 3\lambda (\lambda \cdot 0 + 1)^2}{(\lambda \cdot 0 + 1)^6} = 0 \Rightarrow a_1 = 3\lambda \qquad (5.3-17)$$

and 5.3–14 yields

$$f(s) = \frac{3\lambda s + 1}{(\lambda s + 1)^3} \qquad (5.3-18)$$

This is the filter we proposed for stable systems for Type 2 performance. □

5.3.2 Qualitative Interpretation of the Filter Function

With the filter f and the H_2-optimal controller $\tilde{q}$ the complementary sensitivity function becomes

$$\tilde{\eta} = p\tilde{q}f \qquad (5.3-19)$$

In Sec. 5.2.2. we noted that $p\tilde{q} = 1$ for MP systems. Thus for MP systems the finite closed-loop poles are the poles of the filter f. Adjusting the filter time constant λ is equivalent to adjusting the speed of the closed-loop response.

The filter is always designed such that the closed-loop system is stable for all values of $\lambda > 0$. For stable systems the filter form is unconstrained except at $s = 0$ where Type 1 performance requires $f(0) = 1$ and Type 2 requires, in addition, $\frac{df}{ds}|_{s=0} = 0$. For unstable systems the filter has to be unity at all unstable system poles. This constraint limits the range of filter time constants λ which can be chosen for reasonable performance as we will show next.

Consider the MP system

$$p = p_M = \frac{1}{-s + \beta} \qquad \beta > 0 \qquad (5.3-20)$$

and assume that a step disturbance acts at the process input

$$v = v_M = \frac{1}{s(-s + \beta)} \qquad (5.3-21)$$

The H_2-optimal controller is $\tilde{q} = p^{-1}$. For internal stability and asymptotically error-free disturbance compensation we have to require that the filter satisfies

$$f(0) = f(\beta) = 1 \qquad (5.3-22)$$

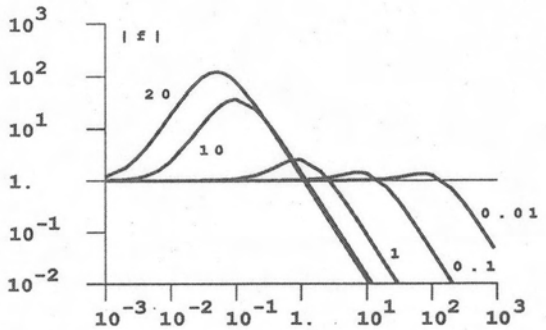


Figure 5.3-1. Filter magnitude for different values of λ; system has one unstable pole at (1,0).

We will assume that the filter is second order ($n = 2$) and use (5.3–8) to find its transfer function

$$f(s) = \frac{1}{(\lambda s+1)^2}\left[\frac{-s+\beta}{\beta} + \frac{(\lambda\beta+1)^2 s}{\beta}\right] = \frac{1}{(\lambda s+1)^2}\left[\lambda s(\lambda\beta+2)+1\right] \tag{5.3-23}$$

The following approximate expressions are valid for large and small filter time constants respectively

$$\lambda\beta >> 2 \qquad f(s) \cong \frac{\lambda^2\beta s+1}{(\lambda s+1)^2} \tag{5.3-24}$$

$$\lambda\beta << 2 \qquad f(s) \cong \frac{2\lambda s+1}{(\lambda s+1)^2} \tag{5.3-25}$$

For large λ's the zero $(-\lambda^2\beta)^{-1}$ of (5.3–24) will occur at much smaller frequencies than the pole $-\lambda^{-1}$. This will result in a large peak of the complementary sensitivity function η and is therefore undesirable. For small λ's the zero $(-2\lambda)^{-1}$ is close to the pole $-\lambda^{-1}$ and therefore the peak in η will be moderate.

In conclusion, for unstable systems *it is undesirable from a performance and robustness point of view to make the filter time constant much larger than the inverse of the unstable pole* β^{-1} (Fig. 5.3-1). The closed loop response cannot be made arbitrarily slow, otherwise the performance deteriorates drastically. This observation is very reasonable from a physical point of view: If a system has a very fast unstable pole — i.e., the pole is far out in the RHP, rapid control action is necessary to get acceptable performance.

For NMP systems, in general, the optimal $\tilde{\eta}$ (5.3–19) includes as poles the mirror images of the plant RHP zeros and poles. Increasing the filter time constant —

i.e., moving the filter pole closer to the origin will affect the closed-loop response noticeably only when the filter poles become of the same order of magnitude as the poles of $p\tilde{q}$.

Difficulties arise when the unstable pole and the unstable zero are located close to each other. Because $\epsilon(\zeta) = 1$ at the zero ζ, and $\epsilon(\pi) = 0$ at the pole π, the sensitivity function has to vary rapidly which leads to high peaks in ϵ as we will demonstrate in Sec. 5.7.2. These control difficulties are inherent in the plant itself and little can be done via sophisticated control system design.

5.4 Robust Stability

5.4.1 Norm-Bounded Uncertainty Regions

For IMC ($\tilde{\eta} = \tilde{p}\tilde{q}f$) Thm. 2.5-1 becomes Cor. 5.4-1.

Corollary 5.4-1 (Robust Stability). *Assume that all plants p in the family Π*

$$\Pi = \left\{ p : \left| \frac{p(i\omega) - \tilde{p}(i\omega)}{\tilde{p}(i\omega)} \right| < \bar{\ell}_m(\omega) \right\} \qquad (5.4-1)$$

have the same number of RHP poles. Then the system is robustly stable if and only if the IMC filter satisfies

$$|f| < \frac{1}{|\tilde{p}\tilde{q}\bar{\ell}_m|} \qquad \forall\omega \qquad (5.4-2)$$

where $\tilde{q}$ is a stabilizing controller for the nominal plant $\tilde{p}$.

For stable systems there are no constraints on f. Therefore there always exists a filter f which satisfies (5.4–2) regardless how large the uncertainty $\bar{\ell}_m$. For unstable systems f is constrained to be unity at the RHP system poles. Thus, depending on $\bar{\ell}_m$, there may not exist any filter time constant λ for which the constraint (5.4–2) is met. Indeed, there may not exist any filter — however complicated — which satisfies (5.4–2). A minimum amount of information is necessary or equivalently a maximum amount of uncertainty is allowed to stabilize an unstable system. The necessary information at $\omega = 0$ can be characterized easily.

Corollary 5.4-2. *Assume that a filter f is to be designed for a system with poles at the origin — i.e., $f(0) = 1$. There exists an f such that the closed-loop system is robustly stable for the family Π described by (5.4-1) only if $\bar{\ell}_m(0) < 1$.*

Note that contrary to Cor. 4.3-2, Cor. 5.4-2 is only necessary. For unstable systems, in general, the filter has to satisfy other constraints in addition to the one at $s = 0$.

The filter design procedure is again to increase the filter time constant until (5.4–2) is satisfied. In Sec. 5.3.2. we found that making the filter poles slower than the mirror images of the RHP system poles leads to peaks in the complementary sensitivity function and consequently to increased robustness problems. Thus the range over which the time constant λ can be varied to improve robustness is limited.

5.4.2 General Uncertainty Regions

The problem formulation is identical to the one for stable systems presented in Sec. 4.3.2. The next corollary follows directly from Thm. 4.3-1.

Corollary 5.4-3. *Assume that all plants in the family* Π *have n poles in the open RHP. The closed-loop system is robustly stable for* Π *if and only if the Nyquist band* $\pi(\omega)c(i\omega)$ *makes n net counterclockwise encirclements of the point (-1,0) without covering it.*

The discussion of Sec. 4.3.2 carries over to the case of unstable systems with minor obvious modifications.

5.5 Robust Performance

The key results, Cor. 4.4-1 and Thm. 4.4-1, hold for unstable systems if it is assumed that all plants in the family Π have the same number of RHP poles and if the controller $\tilde{q}$ and filter f are stabilizing for the nominal plant $\tilde{p}$. The remarks of Sec. 4.4. can be interpreted in the context of open-loop unstable systems with minor self evident changes. They are not repeated here for brevity.

For the two-degree-of-freedom structure, it is necessary to use the Structured Singular Value (used in the chapters on MIMO systems) in order to obtain non-conservative conditions for robust performance. An alternative is to use condition (3.4–20) (which is only sufficient) and the following procedure. First design f_d ($q_d = \tilde{q}_d f_d$) according to the one-degree-of-freedom design method. Then design f_r ($q_r = \tilde{q}_r f_r$) to satisfy (3.4–20). When this is accomplished, the control system performance will be robust for both disturbance rejection and setpoint tracking. However, the fact that (3.4–20) is only sufficient may make the task of designing a satisfactory f_r impossible. Clearly, the above procedure emphasizes robustness of disturbance rejection rather than of setpoint tracking as is appropriate for process control applications.

After q_d and q_r have been designed, the controllers for the feedback structure in Fig. 3.3-1 can be obtained through

$$c_1 c_3 = \frac{q_d}{1 - \tilde{p} q_d} \qquad (5.5-1)$$

$$c_2 = q_r \frac{1 + c_1 c_3 \tilde{p}}{c_1} \qquad (5.5-2)$$

where c_1 is MP, c_3 is stable and there are no RHP pole-zero cancellations between c_1 and c_3. Otherwise c_1 and c_3 are arbitrary.

5.6 Summary of the IMC Design Procedure

The required information for the IMC design is the same as that for stable systems: process model, input type, performance specifications and uncertainty information. The input specification requires some care. If the physical disturbance enters at the plant input, the input v used in the design procedure has to include the unstable system poles for the resulting controller to yield offset free performance.

Design Procedure

Step 1: Nominal Performance

The stabilizing H_2-optimal controller $\tilde{q}$ is determined which minimizes

$$\|(1 - \tilde{p}\tilde{q})\|_2$$

for the specified input v. After the unstable plant and input poles and zeros have been determined the optimal controller $\tilde{q}$ can be found explicitly from:

$$\tilde{q} = b_p (p_M b_v v_M)^{-1} \left\{ (b_p p_A)^{-1} b_v v_M \right\}_* \qquad (5.2-6)$$

If all the unstable plant/input poles are at the origin, the entries in Table 4.1-1, 2 and 3 can be used to find $\tilde{q}$ for typical inputs.

Step 2: Robust Stability and Robust Performance

The controller $\tilde{q}$ is augmented by the IMC filter f

$$q = \tilde{q} f$$

In order for q to be stabilizing, f has to be unity at all unstable plant/input poles $\pi_1, \ldots, \pi_k$. For distinct poles the one-parameter filter has the form

$$f(s) = \frac{1}{(\lambda s + 1)^n} \sum_{j=1}^{k} (\lambda \pi_j + 1)^n \prod_{\substack{i=1 \\ i \neq j}}^{k} \frac{s - \pi_i}{\pi_j - \pi_i} \qquad (5.3-8)$$

For repeated unstable poles, in general, a system of linear equations has to be solved to determine the numerator coefficients of the filter (Sec. 5.3.1). The order

n is selected large enough for q to be proper. The filter poles are a subset of the system closed-loop poles. If they are made much slower than the mirror images of the system RHP poles undesirable performance and robustness properties result. For MP systems and in the absence of modelling error, λ is the closed-loop time constant. For NMP systems λ becomes the dominant closed-loop time constant when it is made large enough.

Robust Stability. Check if

$$|\tilde{p}\tilde{q}f\bar{\ell}_m| < 1 \quad \text{for} \quad \omega = 0 \tag{5.6-1}$$

This condition is necessary for a filter time constant $\lambda > 0$ to exist for which the system is robustly stable.

Robust Performance. Increase λ just enough to meet the condition

$$|\tilde{p}\tilde{q}f\bar{\ell}_m| + |(1 - \tilde{p}\tilde{q}f)w| < 1 \qquad \forall\omega \tag{4.4-4}$$

i.e., choose λ to make (4.4–4) (almost) an equality for some specific value(s) of ω.

If the two-degree-of-freedom structure is used, then, at least for the nominal case, q_d for disturbance rejection and $\hat{q}$ for setpoint tracking can be designed sequentially (Sec. 5.2.4). Recall that with the two-degree-of-freedom structure the RHP system poles impose limitations on disturbance rejection only but not on setpoint tracking performance.

5.7 Applications

5.7.1 Distillation Column Base Level Control

When the steam input to a distillation column is increased, the liquid level in the column base can exhibit inverse response behavior. The following type of model

$$\frac{y}{u} = \frac{1}{s}(1 - 2e^{-s\theta}) \tag{5.7-1}$$

was found to describe the behavior of many industrial columns adequately. The response of y to a step change in u is shown in Fig. 5.7-1. The deadtime θ is equal to the sum of the hydraulic time constants τ of the n individual trays ($\theta = n\tau$). Actually, the time delay in (5.7–1) is used to approximate a large number of first order lags modelling the tray behavior

$$e^{-sn\tau} \cong \frac{1}{(\tau s + 1)^n} \tag{5.7-2}$$

The following numbers would be typical: $\tau = 5$ sec, $n = 18$ and $\theta = 1.5$ min.

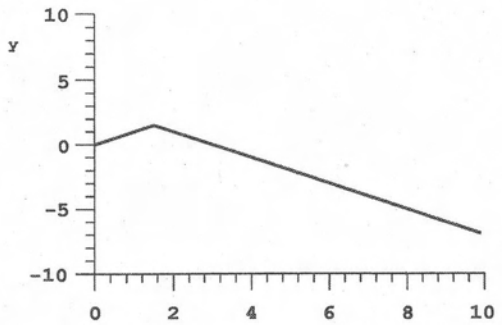


Figure 5.7-1. Reboiler level response to a step change in steam input.

For the purpose of this example we will assume that the "real" column behaves like a sequence of lags. The model uses the time delay approximation (5.7–2). The absolute value of the multiplicative error

$$\ell_m = \frac{p - \tilde{p}}{\tilde{p}} = \frac{e^{-1.5s} - (0.083s + 1)^{-18}}{0.5 - e^{-1.5s}} \qquad (5.7 - 3)$$

as well as the smooth bound $\bar{\ell}_m = |2.6s(1.3s + 1)^{-1}|$ are shown in Fig. 5.7-2 as a function of frequency. Because here we are concerned only with *one* model and *one* plant it is clearly conservative to define a *set* of plants bounded by $\bar{\ell}_m(\omega)$. Nevertheless, we will design the control system to meet the performance specifications for all plants in this set, assuming that some of the unknown model errors are captured with this uncertainty description.

All major disturbances are step-like and enter at the plant input. Thus for the control system design we will assume $v = s^{-2}$. As a robust performance specification we will require the maximum peak of ϵ not to exceed $w^{-1} = 2.5$.

Design Procedure

Step 1: Nominal Performance

For (5.7–1) $b_p = b_v = 1, v_M = s^{-2}$. Because (5.7–1) has an infinite number of RHP zeros

$$\zeta_k = \frac{1}{\theta}(\ln 2 + 2k\pi i) \quad -\infty < k \text{ (integer)} < +\infty \qquad (5.7 - 4)$$

it is somewhat more complicated to determine the allpass portion p_A of the model. The allpass must have the poles at the mirror images of the zeros

$$\pi_k = \frac{1}{\theta}(-\ln 2 + 2k\pi i) \quad -\infty < k \text{ (integer)} < +\infty \qquad (5.7 - 5)$$

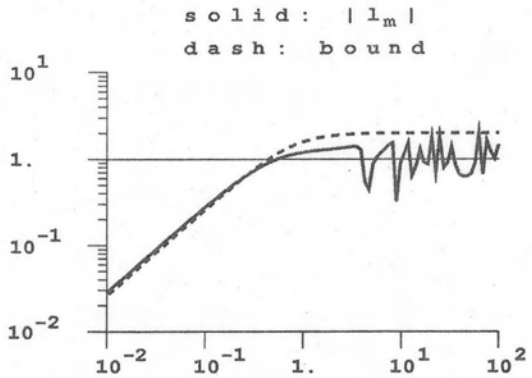


Figure 5.7-2. Model error caused by delay approximation.

and therefore

$$p_A = \frac{1 - 2e^{-s\theta}}{-2 + e^{-s\theta}} \tag{5.7-6}$$

$$p_M = \frac{1}{s}(-2 + e^{-s\theta}) \tag{5.7-7}$$

The H_2-optimal controller is

$$\tilde{q} = (-2 + e^{-s\theta})^{-1} s^3 \left\{ \frac{-2 + e^{-s\theta}}{1 - 2e^{-s\theta}} \frac{1}{s^2} \right\}_* = (-2 + e^{-s\theta})^{-1} s^3 \left[\frac{1}{s^2} + \frac{3\theta}{s} \right]$$

$$= (-2 + e^{-s\theta})^{-1} s(3\theta s + 1) \tag{5.7-8}$$

$$\bar{\eta} = \tilde{p}\tilde{q} = \frac{1 - 2e^{-s\theta}}{-2 + e^{-s\theta}}(3\theta s + 1) \tag{5.7-9}$$

The error resulting from a ramp disturbance is shown in Fig. 5.7-3. $\bar{\eta}$ and $\bar{\epsilon}$ can be seen in Fig. 5.7-4. Note that because of the NMP components the achievable performance measured either in terms of time domain behavior (Fig. 5.7-3) or bandwidth (Fig. 5.7-4B) is quite poor.

Step 2: Robust Stability and Robust Performance

Because the only two unstable poles are at the origin, a Type 2 filter can be used to improve the systems robustness. The order has to be three to make q proper.

$$f(s) = \frac{3\lambda s + 1}{(\lambda s + 1)^3} \tag{5.7-10}$$

The effect of λ on robustness ($|\tilde{\eta}|$) and performance ($|\tilde{\epsilon}|$) is shown in Fig. 5.7-4. Robust stability for the set of plants bounded by $\bar{\ell}_m$ requires $\lambda > 5$.

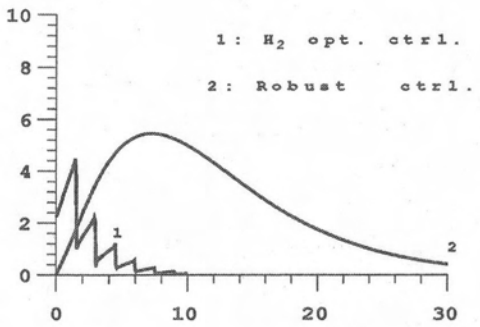


Figure 5.7-3. Closed-loop response for ramp disturbance; ISE-optimal controller (1) and "robust" controller (2).

The robust performance constraints for different λ's are seen in Fig. 5.7-5. A time constant $\lambda = 7$ is necessary for robust performance when $w^{-1} = 2.5$. The response of the "real" plant for this controller is compared to the ideal response in Fig. 5.7-3.

Because the open-loop system is unstable, the controller must be implemented via the classic feedback structure. The optimal controller is calculated as

$$c = \frac{\tilde{q}}{1 - \tilde{p}\tilde{q}} = \frac{(3\theta s + 1)s}{3\theta s(-1 + 2e^{-s\theta}) - 3(1 - e^{-s\theta})} \qquad (5.7 - 11)$$

Note that c has a double pole at the origin. One is canceled by the zero at the origin, the other one provides the integral action required for the error free tracking of ramp inputs.

Mainly because of the delay, the expression (5.7–11) for the controller is quite complex. It becomes even more complex when the Type 2 filter is added. In the next chapter we will show that approximating c by a PID controller brings about only a minor loss of performance.

5.7.2 NMP Unstable Systems

In this section we will demonstrate the design difficulties arising when the plant contains both RHP poles and zeros. Consider the general example

$$p = \frac{-s + \alpha}{-s + \beta} \qquad \alpha, \beta > 0 \qquad (5.7 - 12)$$

so that

$$p_A = \frac{-s + \alpha}{s + \alpha} \qquad (5.7 - 13)$$

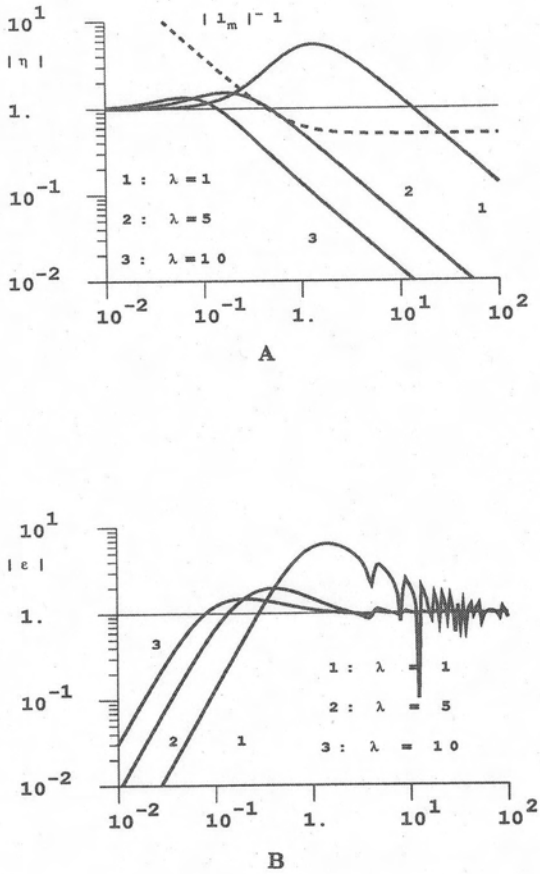


Figure 5.7-4. $|\tilde{\eta}|$ and $|\tilde{\epsilon}|$ for different filter parameters; $\lambda > 5$ is required for robust stability.

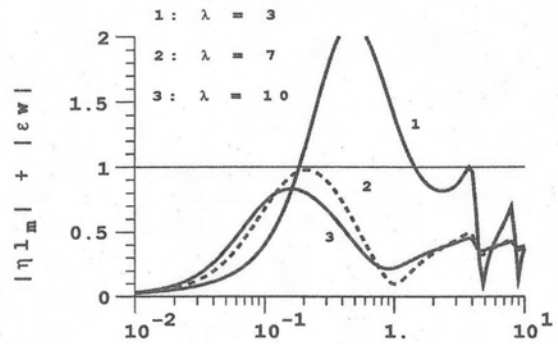


Figure 5.7-5. Robust performance bound for different filter parameters.

$$p_M = \frac{s+\alpha}{-s+\beta} \qquad (5.7-14)$$

$$b_p = \frac{-s+\beta}{s+\beta} \qquad (5.7-15)$$

A step disturbance is assumed to pass through the unstable system portion:

$$v_M = \frac{1}{s(-s+\beta)} \qquad (5.7-16)$$

$$b_v = \frac{-s+\beta}{s+\beta} \qquad (5.7-17)$$

Substituting (5.7–13 to 17) into (5.2–6) yields after some algebra

$$\tilde{q} = \frac{(-s+\beta)(2s+\alpha-\beta)}{(\alpha-\beta)(s+\alpha)} \qquad (5.7-18)$$

The appropriate filter satisfying (5.3–2) for $\pi_2 = 0$ and $\pi_2 = \beta$ is

$$f(s) = \frac{\beta^{-1}\left[(\lambda\beta+1)^3-1\right]s+1}{(\lambda s+1)^3} \qquad (5.7-19)$$

The sensitivity and complementary sensitivity functions are shown in Fig. 5.7-6 A and B for $\alpha = 1, \beta = 10$. One notices from the behavior of ϵ as a function of λ that the bandwidth is limited to about 1 rad min^{-1} because of the RHP zero at $s = 1$. Also, the complementary sensitivity cannot be forced to drop below one at frequencies below 10 rad min^{-1} because of the RHP pole at $s = 10$. This pole

is also the cause of the high peaks in ϵ and η when λ is decreased much below 0.05 min.

When the pole moves closer to the zero ($\alpha = 1, \beta = 2$) the maximum peak in both ϵ and η cannot be reduced much below 13 regardless of how λ is chosen (Fig. 5.7-7). This reflects the control difficulties inherent in systems with RHP poles and zeros close to each other.

5.8 References

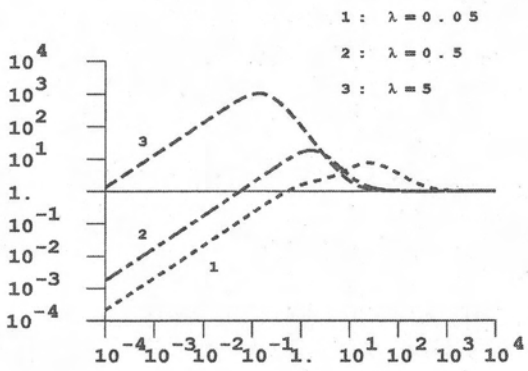
5.1.2. A similar controller parametrization was proposed by Zames & Francis (1983). It was modified here to allow poles at the origin. The parametrization is equivalent to that proposed by Youla et al. (1976).

5.2.3. The additivity is pointed out in the book by Frank (1974).

5.2.4. The limitations imposed by RHP zeros and poles in the context of the two-degree-of-freedom structure are discussed by Vidyasagar (1985).

5.3.1. An efficient and robust algorithm for solving systems of linear equations in the Vandermonde form is reported by Golub & VanLoan (1983; pp. 119-124). When the conditions (5.3–10) are added, the system is of a confluent Vandermonde form. Efficient solution algorithms for related systems are studied by Bjork and Elfring (1973).

5.7.1. This control problem was proposed to the authors by J. Shunta of DuPont. The model is derived in the book by Buckley, Luyben & Shunta (1985) where other control schemes for this system are also discussed.



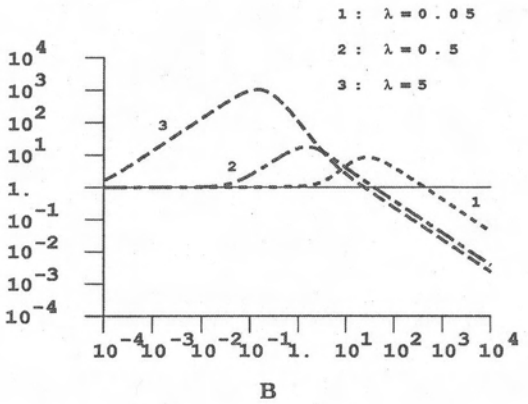


Figure 5.7-6. Sensitivity (A) and complementary sensitivity (B) for $\alpha = 1, \beta = 10$ for different values of the filter parameter λ.

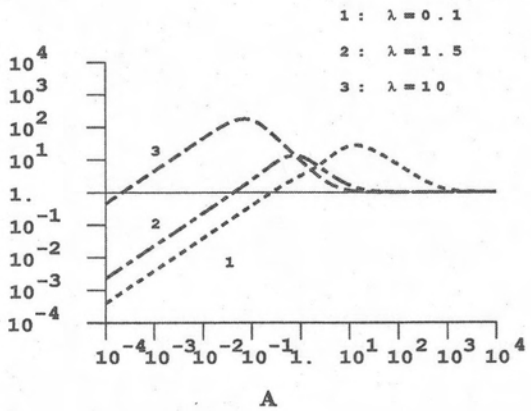


Figure 5.7-7. Sensitivity (A) and complementary sensitivity (B) for $\alpha = 1, \beta = 2$ for different values of the filter parameter λ.

Chapter 6

ISSUES IN SISO IMC DESIGN

6.1 Implications of IMC for Classic Feedback Controllers

By far the most widely used controllers in the process industries are the two-term PI and the three-term PID controller. Though IMC is clearly more general and therefore more powerful it is worthwhile to explore the relationships between IMC, PI and PID in order to gain insight into the tuning of these simpler controllers, their performance, robustness and limitations.

6.1.1 General Relationships

In Chap. 3 we derived the general relationship between the classic feedback controller c and the IMC controller q when the process model is $\tilde{p}$:

$$c = \frac{q}{1 - \tilde{p}q} \qquad (6.1-1)$$

In Chaps. 4 and 5 we proposed to choose as the IMC controller q the controller $\tilde{q}$, which is H_2 optimal for the system $\tilde{p}$ for a particular input v, and to augment it by a lowpass filter f.

$$q = \tilde{q}f \qquad (6.1-2)$$

The order of q obtained in this fashion is generally higher than the order of the model $\tilde{p}$. The complexity of q and similarly, the complexity of the equivalent classic controller c is determined by the complexity of the model. Simple models give rise to simple controllers. In particular, for a large number of simple models used commonly in the process industries the IMC design procedure yields classic PI and PID controllers as we will show in the next sections. We will compare these IMC-tuning rules with the rules proposed by Ziegler & Nichols and Cohen & Coon in terms of performance and robustness.

6.1.2 PID Settings for Simple Models

The IMC controllers shown in Table 6.1-1 were designed via the standard procedure developed in Chaps. 4 and 5. We find that IMC leads to PID controllers for virtually all models common in industrial practice. Note that the table includes systems with pure integrators and RHP zeros. Occasionally, the PID controllers are augmented by a first-order lag with time constant τ_F.

For Case O the IMC design procedure would yield a controller of too high an order to be interpreted as a PID controller with lag. Therefore a slightly modified procedure was used resulting in a controller with the same attractive properties as the standard IMC controller. A few remarks regarding Table 6.1-1 are appropriate.

- The controllers have a single adjustable parameter λ. For MP systems λ is the closed-loop time constant. For NMP systems λ is the dominant closed-loop time constant when $\lambda > \beta$, where β^{-1} is the location of the zero in the RHP.

- When the PID controller of the specified form is applied to the model $\tilde{p}$ the closed-loop system is stable for all values of $\lambda \geq 0$.

- For the specified inputs the controllers yield offset free performance for all $\lambda \geq 0$ and are H_2 optimal for $\lambda = 0$.

- For MP plants the controllers with $\lambda = 0$ are H_2 optimal for *all* inputs which behave asymptotically ($t \rightarrow \infty$) like the specified v_M.

- For Cases A, B, C, G, and I the controller gain is inversely proportional to λ, thus demonstrating that on-line PID controller adjustment is effectively achieved by manipulating k_c. These are MP models, for which the model itself imposes no limitations on the bandwidth.

  For Cases D, E, F, K, L, and N the controller gain k_c and the filter constant τ_F are the only parameters dependent on λ, but because of the presence of a RHP zero, there is a maximum gain which cannot be surpassed no matter how small λ is.

  For Cases H, J, M, and O, λ appears in *all* the parameters of the classic feedback controller. It is not surprising that for such processes trial and error tuning of PID controllers is notoriously difficult. However, the IMC parametrization shows how all the controller parameters may be adjusted *simultaneously* in an effective manner.

- In practice the ideal PID controller does not exist. An additional lag is always present in the controller to provide roll-off at high frequencies. Cases D, E, F, and K through O yield "practical" PID controllers with "optimal" roll-off elements $(\tau_F s + 1)^{-1}$.

- No systems with LHP zeros are listed in Table 6.1-1. LHP zeros translate into lags in the feedback controller structure when the IMC design procedure is used. Therefore, for models with LHP zeros, the PID controller from Table 6.1-1 should be augmented with the corresponding lags.

- Controller complexity, as stated in the introduction, depends on the model and the control system objectives. Consider the cases G (a pure integrator), A (a first-order model) and B (a second-order noninteracting model), for which the input is a step. Only a proportional controller is necessary for G, a PI-controller must be used for A, while a PID controller is needed for B. Likewise, consider Cases N and O where the process model is the same: as the inputs become more demanding (ramps instead of steps) the complexity of the controller increases.

- Table 6.1-1 can also be used for systems with delays by approximating the deadtime with a Padé element; then the entry for the rational approximate model provides the controller parameters. This design approach leads to simpler controllers than when the time delay is considered exactly. The use of the Padé approximation, however, introduces modelling error, which consequently limits the achievable bandwidth and the minimum value for λ. This loss of performance is usually insignificant. Corollary 4.3-1 provides the condition

$$\bar{\ell}_m < \frac{1}{|\tilde{\eta}|} = \frac{1}{|\tilde{p}\tilde{q}f|} \qquad (6.1-3)$$

  which guarantees robust stability when the multiplicative model error is bounded by $\bar{\ell}_m$. The Padé approximation does not generate a family of plants but generates a specific error ℓ_m. If we set $\bar{\ell}_m = |\ell_m|$, (6.1–3) is only sufficient and therefore conservative. Nevertheless it is useful for estimating a lower bound on λ which assures closed-loop stability.

Table 6.1-1. IMC controllers for simple models interpreted as PID controllers with filter:

$$c = k_c(1 + \tau_D s + \frac{1}{\tau_I s})\frac{1}{\tau_F s + 1}; \quad \gamma = \frac{2\beta\tau_2}{\beta + \tau_2}$$

	Model $\tilde{p}$	Input v_M	$\tilde{\eta} = \tilde{p}q = \tilde{p}\tilde{q}f$	$k_c k$	τ_I	τ_D	τ_F
A	$\frac{k}{\tau s+1}$	$\frac{1}{s}$	$\frac{1}{\lambda s+1}$	$\frac{\tau}{\lambda}$	τ	–	–
B	$\frac{k}{(\tau_1 s+1)(\tau_2 s+1)}$	$\frac{1}{s}$	$\frac{1}{\lambda s+1}$	$\frac{\tau_1+\tau_2}{\lambda}$	$\tau_1+\tau_2$	$\frac{\tau_1\tau_2}{\tau_1+\tau_2}$	–
C	$\frac{k}{\tau^2 s^2+2\zeta\tau s+1}$	$\frac{1}{s}$	$\frac{1}{\lambda s+1}$	$\frac{2\zeta\tau}{\lambda}$	$2\zeta\tau$	$\frac{\tau}{2\zeta}$	–
D	$k\frac{-\beta s+1}{\tau s+1}$	$\frac{1}{s}$	$\frac{-\beta s+1}{(\beta s+1)(\lambda s+1)}$	$\frac{\tau}{2\beta+\lambda}$	τ	–	$\frac{\beta\lambda}{2\beta+\lambda}$
E	$k\frac{-\beta s+1}{\tau_1 s+1}$	$\frac{1}{\tau_2 s+1}\frac{1}{s}$	$\frac{(-\beta s+1)(\gamma s+1)}{(\beta s+1)(\lambda s+1)}$	$\frac{\gamma+\tau_1}{2\beta-\gamma+\lambda}$	$\gamma+\tau_1$	$\frac{\gamma\tau_1}{\gamma+\tau_1}$	$\frac{\beta(\gamma+\lambda)}{2\beta-\gamma+\lambda}$
F	$k\frac{-\beta s+1}{\tau^2 s^2+2\zeta\tau s+1}$	$\frac{1}{s}$	$\frac{-\beta s+1}{(\beta s+1)(\lambda s+1)}$	$\frac{2\zeta\tau}{2\beta+\lambda}$	$2\zeta\tau$	$\frac{\tau}{2\zeta}$	$\frac{\beta\lambda}{2\beta+\lambda}$
G	$\frac{k}{s}$	$\frac{1}{s}$	$\frac{1}{\lambda s+1}$	$\frac{1}{\lambda}$	–	–	–

H	$\frac{k}{s}$	$\frac{1}{s^2}$	$\frac{2\lambda s+1}{(\lambda s+1)^2}$	$\frac{2}{\lambda}$	2λ	–	–
I	$\frac{k}{s(\tau s+1)}$	$\frac{1}{s}$	$\frac{1}{\lambda s+1}$	$\frac{1}{\lambda}$	–	τ	–
J	$\frac{k}{s(\tau s+1)}$	$\frac{1}{s^2}$	$\frac{2\lambda s+1}{(\lambda s+1)^2}$	$\frac{2\lambda+\tau}{\lambda^2}$	$2\lambda+\tau$	$\frac{2\lambda\tau}{2\lambda+\tau}$	–
K	$k\frac{-\beta s+1}{s}$	$\frac{1}{s}$	$\frac{-\beta s+1}{(\beta s+1)(\lambda s+1)}$	$\frac{1}{2\beta+\lambda}$	–	–	$\frac{\beta\lambda}{2\beta+\lambda}$
L	$k\frac{-\beta s+1}{s}$	$\frac{1}{\tau_2 s+1}\frac{1}{s}$	$\frac{(-\beta s+1)(\gamma s+1)}{(\beta s+1)(\lambda s+1)}$	$\frac{1}{2\beta-\gamma+\lambda}$	–	γ	$\frac{\beta(\gamma+\lambda)}{2\beta-\gamma+\lambda}$
M	$k\frac{-\beta s+1}{s}$	$\frac{1}{s^2}$	$\frac{(-\beta s+1)(2\beta s+1)(2\lambda s+1)}{(\beta s+1)(\lambda s+1)^2}$	$\frac{2(\beta+\lambda)}{2\beta^2+\lambda^2}$	$2(\beta+\lambda)$	$\frac{2\beta\lambda}{\beta+\lambda}$	$\frac{\beta\lambda^2+4\beta^2\lambda}{2\beta^2+\lambda^2}$
N	$k\frac{-\beta s+1}{s(\tau s+1)}$	$\frac{1}{s}$	$\frac{-\beta s+1}{(\beta s+1)(\lambda s+1)}$	$\frac{1}{2\beta+\lambda}$	–	τ	$\frac{\beta\lambda}{2\beta+\lambda}$
O	$k\frac{-\beta s+1}{s(\tau s+1)}$	$\frac{1}{s^2}$	$\frac{(-\beta s+1)(2(\beta+\lambda)s+1)}{(\beta s+1)(\lambda s+1)^2}$	$\frac{2(\beta+\lambda)+\tau}{2\beta^2+4\beta\lambda+\lambda^2}$	$2(\beta+\lambda)+\tau$	$\frac{2\tau(\beta+\lambda)}{2(\beta+\lambda)+\tau}$	$\frac{\beta\lambda^2}{2\beta^2+4\beta\lambda+\lambda^2}$

Example 6.1-1. The system

$$p(s) = \frac{1 - k_1 e^{-\theta s}}{s} \qquad (6.1-4)$$

was studied in Sec. 5.7.1. Using a first-order Padé approximation we find

$$\tilde{p}(s) = \frac{1 - k_1 + \frac{\theta}{2}(1 + k_1)s}{s(\frac{\theta}{2}s + 1)} \qquad (6.1-5)$$

If $k_1 > 1$ then (6.1–5) has a RHP zero and a controller from Case N or O can be selected. If $k_1 < 1$ the resulting LHP zero should be removed by a simple lag, as explained above. PID parameters can then be obtained from Case I or J.

For the specific distillation column model studied in Sec. 5.7.1, (6.1–5) becomes

$$\tilde{p}(s) = \frac{-1 + 2.25s}{s(0.75s + 1)} \qquad (6.1-6)$$

Case O provides the PID parameters

$$k = -\frac{5.25 + 2\lambda}{10.125 + 9\lambda + \lambda^2} \qquad (6.1-7)$$

$$\tau_I = 5.25 + 2\lambda \qquad (6.1-8)$$

$$\tau_D = \frac{3.375 + 1.5\lambda}{5.25 + 2\lambda} \qquad (6.1-9)$$

$$\tau_F = \frac{2.25\lambda^2}{10.125 + 9\lambda + \lambda^2} \qquad (6.1-10)$$

With this controller

$$|\tilde{\eta}| = \left|\frac{2(2.25 + \lambda)s + 1}{(\lambda s + 1)^2}\right| \qquad (6.1-11)$$

In Fig. 6.1-1 the norm of the model error

$$\ell_m = \frac{p - \tilde{p}}{\tilde{p}} = \frac{-e^{-\theta s} + \frac{-\theta s + 2}{\theta s + 2}}{\frac{1}{k_1} - \frac{-\theta s + 2}{\theta s + 2}} \qquad (6.1-12)$$

$$= \frac{2 - 1.5s - (1.5s + 2)e^{-1.5s}}{-1 + 2.25s} \qquad (6.1-13)$$

is compared with $|\tilde{\eta}|^{-1}$ for $\lambda = 2$. Condition (6.1–3) is satisfied and stability is assured for all $\lambda \geq 2$.

The error resulting from a ramp input (s^{-2}) is shown in Fig. 6.1-2 for $\lambda = 2$. It should be compared with the error obtained with the H_2-optimal controller. With regard to the maximum deviation from steady state there is little incentive

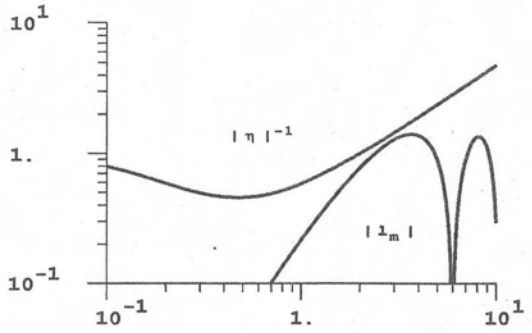


Figure 6.1-1. Error ℓ_m caused by Padé approximation and complementary sensitivity $\tilde{\eta}$ for $\lambda = 2$.

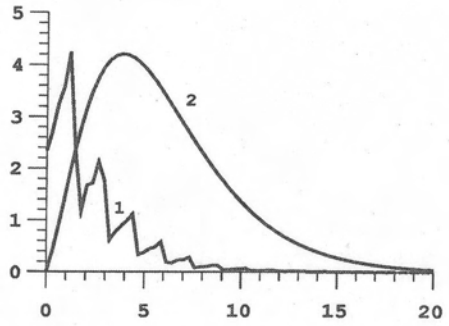


Figure 6.1-2. Error resulting from ramp input for ISE-optimal controller (1) and PID controller (2).

for using anything more sophisticated than a PID controller. If settling time or ISE is critical, the H_2-optimal controller is preferred. Note however, that the H_2-optimal controller used in the simulation is improper and would have to be detuned in practice for robustness reasons. Then its advantages over a properly tuned PID controller might become quite small. □

Example 6.1-2. Most commonly in process control the time delay appears as a multiplicative term in the transfer function

$$p = p_M e^{-\theta s} \qquad (6.1-14)$$

For the zeroth-order Padé approximation ($e^{-\theta s} \cong 1$)

$$|\ell_m| = |1 - e^{-\theta s}| \qquad (6.1-15)$$

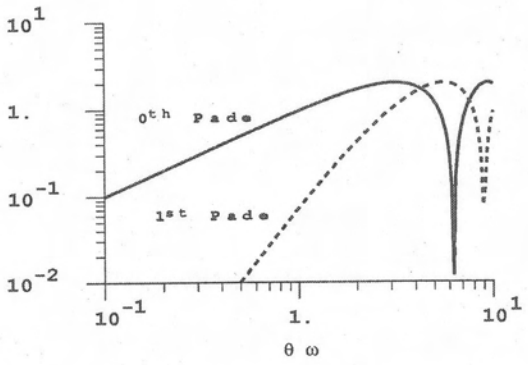


Figure 6.1-3. Error ℓ_m caused by zeroth and first order Padé approximation.

and for the first-order Padé approximation

$$|\ell_m| = \left| \frac{-\frac{\theta}{2}s + 1}{\frac{\theta}{2}s + 1} - e^{-\theta s} \right| \qquad (6.1-16)$$

Both multiplicative errors are plotted in Fig. 6.1-3. For the first-order Padé approximation, $|\ell_m| = 1$ at $\omega \cong 3/\theta$ and thus a sufficient condition for stability is to choose $\lambda > \theta/3$. For the zeroth-order approximation $|\ell_m| = 1$ at $\omega \cong 1/\theta$ and therefore $\lambda > \theta$ is required. Because the bandwidth is inherently limited by the NMP nature of the system, one can expect that using the first-order Padé approximation will yield designs very close to optimal (i.e., as if no approximation were present). The zeroth-order approximation will be adequate, however, when small bandwidths and low frequency inputs are involved.

For the popular first-order model with deadtime

$$p(s) = \frac{ke^{-\theta s}}{\tau s + 1} \qquad (6.1-17)$$

the zeroth-order Padé approximation yields

$$\tilde{p}(s) = \frac{k}{\tau s + 1} \qquad (6.1-18)$$

Entry A in Table 6.1-1 provides a PI controller for this structure. A first-order Padé approximation yields

$$\tilde{p}(s) = \frac{k(-\frac{\theta}{2}s + 1)}{(\tau s + 1)(\frac{\theta}{2}s + 1)} \qquad (6.1-19)$$

Entry F (PID controller with first-order lag) is applicable here. This problem is discussed in more detail in the next subsection. □

Table 6.1-2. IMC-based "real" PID parameters for $p(s) = ke^{-\theta s}(\tau s+1)^{-1}$ and practical recommendations for λ.

Controller	kk_c	τ_I	τ_D	τ_F	Recommended λ/θ ($\lambda > 0.2\tau$ always)
PID	$\frac{2\tau+\theta}{2(\lambda+\theta)}$	$\tau+\frac{\theta}{2}$	$\frac{\tau\theta}{2\tau+\theta}$	$\frac{\lambda\theta}{2(\lambda+\theta)}$	> 0.25
PI	$\frac{\tau}{\lambda}$	τ	–	–	> 1.7
Improved PI	$\frac{2\tau+\theta}{2\lambda}$	$\tau+\frac{\theta}{2}$	–	–	> 1.7

6.1.3 PID Settings for a First-Order System with Deadtime

The important role of the first-order lag/deadtime model in process control mandates a more detailed discussion of Ex. 6.1-2. Our attention is directed to a further understanding of the PI and PID rules generated by cases A and F.

The IMC-based controllers obtained with first- and zeroth-order Padé approximations for the time delay are shown in the first two rows of Table 6.1-2. Inspection of the closed-loop transfer functions for system (6.1–17) with these controllers indicates a number of advantages:

- The closed-loop response is independent of the system time constant τ (the process lag $(1+\tau s)^{-1}$ is cancelled by the controller).
- Time is scaled by θ.
- The shape of the response depends on λ/θ only.

In other words, specifying one value of λ/θ for any first-order-lag-with-deadtime model results in an identical response when time is scaled by θ, regardless of k, θ,

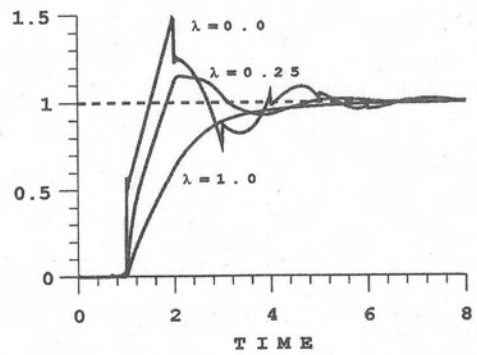


Figure 6.1-4. Step responses with IMC-PID controller for different values of λ.

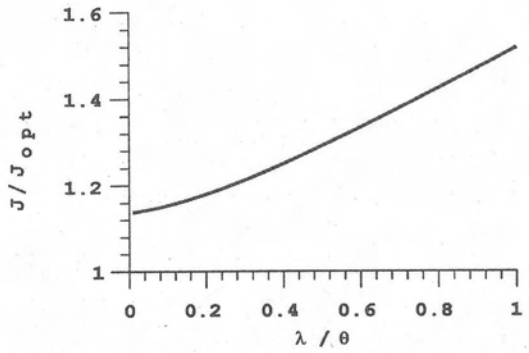


Figure 6.1-5. Normalized ISE for IMC-PID controller as function of λ/θ.

and τ. The choice of the "best" ratio λ/θ must be based on performance and robustness considerations.

PID Controller

For the PID controller Fig. 6.1-4 demonstrates the dependence of the step response on λ/θ. A ratio of $\lambda/\theta = 0.25$ offers a good compromise between rise time and overshoot. Because of the scaling properties of the closed-loop transfer function a convenient design plot can be made (Fig. 6.1-5). The performance measure J, the ISE for a step input change, has been plotted as a function of λ/θ. J is normalized by J_{opt}, the ISE corresponding to the optimum response $y/r = e^{-\theta s}$. In theory, a Smith predictor with infinite gain ($k_c = \infty$) accomplishes this response (see Sec. 6.2). Very small values of λ/θ lead to robustness problems. Based on Figs. 6.1-4 and 6.1-5, $\lambda/\theta = 0.25$ is recommended.

Figure 6.1-5 also confirms that the first-order Padé approximation leads to

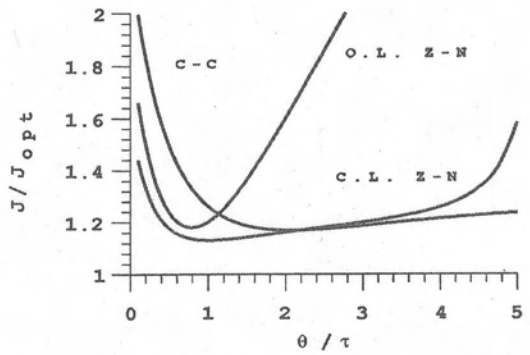


Figure 6.1-6. Comparison of normalized ISE for PID controllers tuned according to the closed-loop and open-loop Ziegler-Nichols rules and according to Cohen-Coon.

relatively little performance deterioration. For $\lambda/\theta = 0.25$ the result is a PID controller that performs with only 20% greater ISE than the *optimal* Smith predictor, while retaining favorable robustness characteristics.

The recommendation $\lambda/\theta = 0.25$ is made under the assumption of no plant uncertainty apart from the "uncertainty" introduced by the Padé approximation. Significant plant uncertainty within the bandwidth of the controller will require the designer to select a larger value of λ. This consideration is of particular concern when $\lambda/\tau << 1$. Because for the process industries the closed-loop bandwidth can rarely exceed ten times the open-loop bandwidth $(10/\tau)$, a practical requirement is to always select $\lambda > \tau/10$. For the IMC-PID parameters this inequality is dominant for $\theta/\tau < 0.4$.

Next we compare IMC-PID with "real" PID controllers

$$c(s) = k_c\left[1 + \frac{1}{\tau_I s} + \frac{\tau_D s}{0.1\tau_D s + 1}\right] \qquad (6.1-20)$$

tuned according to the classic Ziegler-Nichols and Cohen-Coon tuning rules (Fig. 6.1-6). The first notable difference between these rules and those from IMC is that J depends strongly on θ/τ, while for the IMC rules the performance is independent of $\theta/\tau(J/J_{\text{opt}} = 1.2)$. For small and large values of θ/τ the IMC rules clearly lead to superior performance. For $\theta/\tau > 6$ the closed-loop system is even *unstable* with the closed-loop Ziegler-Nichols parameters. Figures 6.1-7A to C show that in the range of $\theta/\tau \approx 1$ the smaller ISE obtained with the non-IMC settings is obtained at the cost of excessive overshoot: 50% are rarely acceptable in practice.

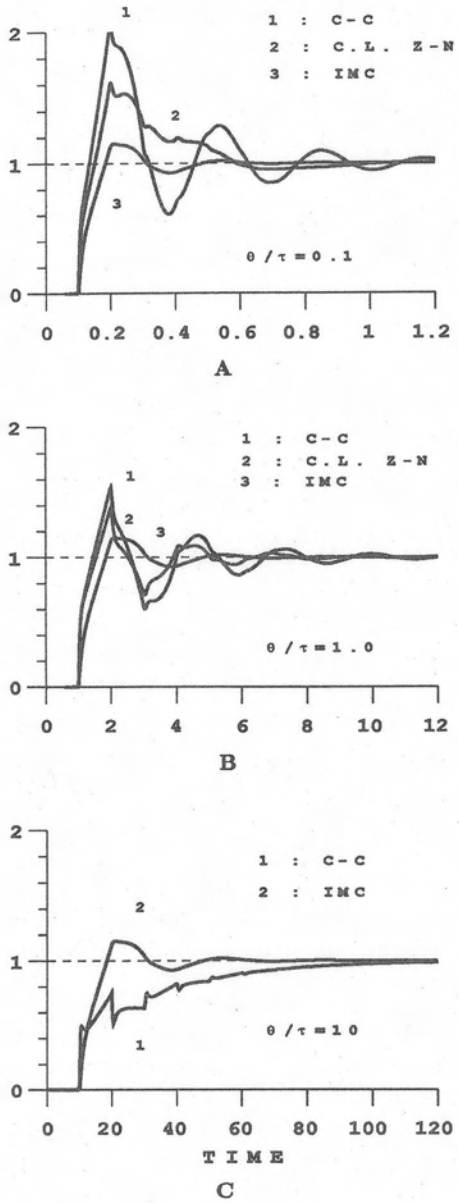


Figure 6.1-7. Step responses with PID controllers tuned according to IMC ($\lambda/\theta = 0.25$), Ziegler-Nichols and Cohen-Coon for different ratios θ/τ. A: $\theta/\tau = 0.1$, B: $\theta/\tau = 1.0$, C: $\theta/\tau = 10$.

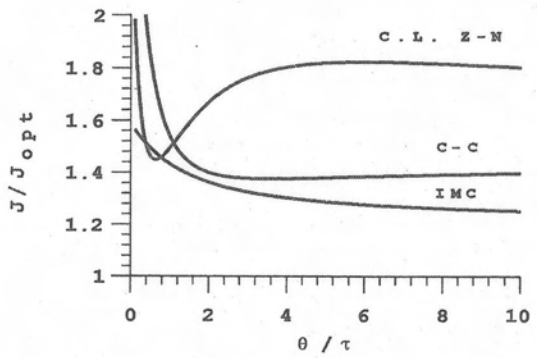


Figure 6.1-8. Normalized ISE for PI controllers tuned according to IMC, Ziegler-Nichols and Cohen-Coon.

PI Controller

Overall the IMC-PI rule is not superior to the Ziegler-Nichols and Cohen-Coon expressions because the zeroth-order Padé approximation neglects deadtime altogether. This can be remedied by incorporating the deadtime in the internal model through other means and leads to the "Improved PI rule" shown in Table 6.1-2. The performance of the IMC-PI controller is now dependent on θ/τ, but for $\lambda/\theta = 1.7$ it is superior to Ziegler-Nichols and Cohen-Coon for almost all values of θ/τ (Fig. 6.1-8).

Robustness to Deadtime Errors

It is important to emphasize that the improved performance obtained through the IMC settings has not been obtained at the cost of robustness. It can be verified that both the PID and the PI controller allow at least 140% deadtime error for all values of θ/τ before the closed-loop system becomes unstable.

6.1.4 Summary

The IMC design procedure was used to derive PID controller parameters for a variety of models commonly used in the process industries (Tab. 6.1-1). The adjustable parameter λ corresponds approximately to the closed-loop time constant and is selected by the designer to achieve the appropriate compromise between performance and robustness and to keep the action of the manipulated variable within bounds. The closed-loop system is stable for all $\lambda \geq 0$. The performance is ISE optimal for $\lambda = 0$.

It is recommended that for time delays a first-order Padé approximation is used ($\theta \cong (-\theta s+2)/(\theta s+2)$). The PID parameters for the rational approximate model can then be found again in Table 6.1-1. Because of the approximation ("model

error") there is a lower bound on λ for stability. When the delay appears in the transfer function as a factor, $\lambda > \theta/3$ is recommended. For the specific case of a first order process with delay the results are summarized in Table 6.1-2.

6.2 IMC Interpretation of Smith Predictor Controller

The "Smith predictor" was proposed in the 1950s to improve the closed-loop performance for systems with time delay. The opinions on its merits expressed in dozens of papers which have appeared in the last three decades vary widely. Our discussion aims to clarify some of the misconceptions and to make some concrete statements regarding the robustness properties.

6.2.1 General Relationships

The feedback control structure with the Smith predictor in place is shown in Fig. 6.2.1A. Here

$$p = p^* e^{-s\theta} \qquad (6.2-1)$$

where p^* does not contain any delay terms. The block marked with double lines is called "predictor." The reason for the name "predictor" can be understood from the following arguments. Without the predictor the feedback signal z is equal to the output y

without predictor:

$$y = pu \qquad (6.2-2)$$

with predictor:

$$z = p^* u \qquad (6.2-3)$$

From (6.2–2) and (6.2–3) the feedback signal z with the predictor in place is related to the output y by

$$z = e^{+s\theta} y \qquad (6.2-4)$$

Thus the block $(p - p^*)$ *predicts the effect of the manipulated variable* u *on the output* y *and modifies the feedback signal* z *accordingly.*

A simple rearrangement of the block diagram in Fig. 6.2-1A which leaves all input-output relationships unaffected makes the similiarity with the IMC structure (Fig. 6.2-1B) apparent.

$$q = \frac{c}{1 + p^* c} \qquad (6.2-5)$$

The IMC structure in Fig. 6.2-1B is internally stable if and only if p and q are stable (Thm. 3.2-1). This implies first that the Smith-predictor controller implemented in the way indicated in Fig. 6.2-1B can be used for *stable* plants

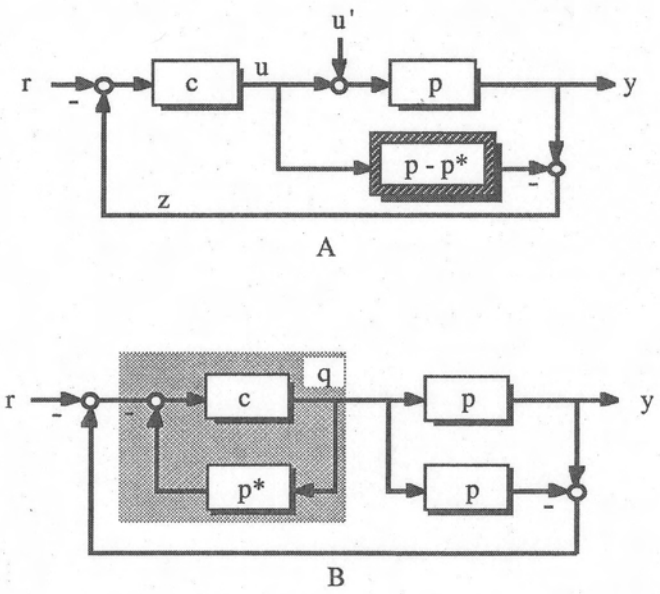


Figure 6.2-1. Smith predictor controller in standard (A) and IMC (B) form.

only. The alternative implementation indicated in Fig. 6.2-1A is also internally unstable for unstable plants. To prove this consider the effect of the external input u' on the output y

$$\frac{y}{u'} = \frac{p^* e^{-s\theta}}{1 + p^* c}[1 - (e^{-s\theta} - 1)p^* c] \qquad (6.2-6)$$

The transfer function (6.2–6) has RHP poles for all choices of the controller c: If c cancels the RHP poles of p^* the first part of expression (6.2–6) is unstable. If it does not, the term in brackets is unstable.

The second implication of Thm. 3.2-1 is that q defined by (6.2–5) has to be stable for internal stability of the Smith-predictor control structure. q is stable if the closed-loop system with the controller c applied to the *delay-free* plant p^* is stable. This is one of the main motivations for using the Smith-predictor: stability of the *nominal* system can be studied in terms of the delay-free plant.

While the Smith-predictor structure is in many aspects similar to the IMC structure, it does not utilize the same controller parametrization and thus loses some of the main advantages of IMC: (1) The set of c's which yields closed-loop stable systems is difficult to define. With the IMC parametrization any stable IMC controller q guarantees closed loop stability. (2) Performance measured in terms of the sensitivity function is a nonlinear function of c while it is a linear function of q.

6.2.2 Some Myths about the Tuning of Smith-Predictor Controllers

The literature abounds with myths regarding the properties, advantages, and drawbacks of Smith-predictor controllers. We hope to diffuse some of them in the following.

Myth #1. With the predictor in place the controller c can be tuned just like for a system p^* without time delay. If stability is the only issue then the statement is correct. If the designer is also concerned about *performance* then it becomes a myth which is clearly not true. For Fig. 6.2-1 the sensitivity function which is indicative of performance is

$$\tilde{\epsilon} = 1 - pq = 1 - p\frac{c}{1 + p^* c} \qquad (6.2-7)$$

It depends on the plant *with* delay and this dependence has to be taken into consideration when adjusting c. Let us assume for example that p^* is MP. If we design c for p^* then it would be optimal (with respect to any objective function) to increase the gain of the controller c to infinity leading to

$$q = \frac{1}{p^*} \qquad (6.2-8)$$

and

$$\tilde{\epsilon} = 1 - e^{-s\theta} = 1 - p_A \qquad (6.2-9)$$

Tables 4.1-1 and 4.1-2 indicate that the sensitivity function (6.2–9) is H_2-optimal for step inputs *only*. For any other input, tuning the controller c for good performance on the delay-free plant p^* will *not* lead to good performance when c is applied to the real plant p (with delay) in the context of the Smith-predictor structure.

Myth #2. Smith-predictor controllers work well for setpoint changes but not for disturbances. Implicit in this statement is usually the assumption that both the setpoint changes and the disturbances are steps but that the disturbances go through a lag (e.g., the plant itself) before acting on the plant output. Myth #1 indicates that unknowingly c is usually tuned for an H_2-optimal response to step inputs. It is indeed true that these controller parameters will not result in good performance for slowly changing disturbances. However, this is not a fault of the Smith-predictor structure but rather a consequence of incorrect tuning of the controller c. Using the results of Chaps. 4 and 5, q and thus c can be found easily for optimal disturbance rejection. However, these parameters again will generally not yield a good setpoint step response. This is the basic trade-off problem the designer faces when tuning a controller for two very different inputs and it has nothing to do with the Smith predictor. It can be easily resolved by the two-degree-of-freedom structure discussed on several occasions in this book.

Myth #3. Smith-predictor controllers are sensitive to model errors. Let the norm of the multiplicative model error be bounded by $\bar{\ell}_m(\omega)$. Then we know from Thm. 2.5-1 that the system is robustly stable if the complementary sensitivity function of the model satisfies

$$|\tilde{\eta}| < \frac{1}{\bar{\ell}_m} \qquad (6.2-10)$$

Because $|p| = |p^*|$, (6.2–10) becomes for the Smith-predictor structure

$$\left|\frac{p^*c}{1+p^*c}\right| < \frac{1}{\bar{\ell}_m} \qquad (6.2-11)$$

The same robust stability condition would have to be satisfied if the model did not include a time delay in the first place. The Smith-predictor does not in any way increase the sensitivity to modelling error. Problems arise in practice because c tends to be adjusted without attention to the limitations imposed by model uncertainty. The ideas presented in Chap. 4 can be easily adapted to design high performance robust controllers for Smith-predictor control systems as is demonstrated in the next subsection.

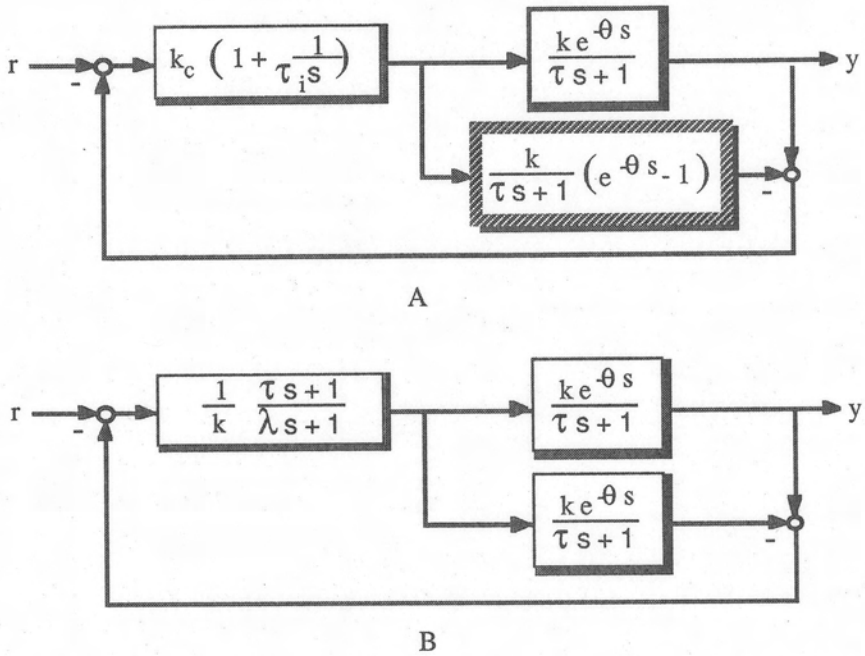


Figure 6.2-2. Smith-predictor/PI controller for first-order deadtime system in standard (A) and IMC (B) form.

6.2.3 Robust Tuning of Smith-Predictor Controller for First-Order System with Deadtime

For a first-order system with deadtime a Smith predictor with a PI controller (Fig. 6.2-2A) is equivalent to IMC (Fig. 6.2-2B) if

$$k_c = \frac{\tau}{\lambda k} \qquad (6.2-12)$$

$$\tau_I = \tau \qquad (6.2-13)$$

If the integral time is set equal to the system time constant, adjusting the PI controller gain is equivalent to adjusting the IMC filter parameter λ. Because IMC is stable for all $\lambda \geq 0$ the Smith predictor with PI controller is stable for all $k_c \geq 0, \tau_I = \tau$. Model uncertainty imposes an upper (lower) limit on $k_c(\lambda)$.

Deadtime Uncertainty. Let us first consider uncertainty in the process deadtime only. This was discussed in detail in Sec. 4.6.1. We found that the rule of

thumb $\lambda = \bar{\delta}$ or

$$k_c = \frac{\tau}{\bar{\delta} k} \qquad (6.2-14)$$

yields acceptable robust performance for a maximum deadtime error of $\bar{\delta}$.

General Uncertainty. When all three model parameters have error bounds associated with them the complete robust design procedure demonstrated in Sec. 4.6.2 has to be carried out to determine the appropriate choice of k_c. The integral time is always set equal to the model time constant regardless of model uncertainty.

6.2.4 Summary

We related the Smith-predictor control structure to the IMC structure (Fig. 6.2-1) and concluded that IMC has the same predictor capabilities. For the design of the feedback control the IMC design procedure is preferred because it deals optimally with different input forms (steps, ramps, etc.) and addresses robustness explicitly. If designed properly — e.g., via the IMC design procedure, the Smith-predictor controller can offer performance advantages over classic feedback control. These advantages are small (or negligible) if model uncertainty requires significant detuning.

6.3 Feedforward Control

6.3.1 Objectives and Structure

Feedforward control can be used to compensate for measurable disturbances. It is discussed here for stable systems only. The combined classic feedforward-feedback structure is shown in Fig. 6.3-1.

The effect of the disturbance on the output is described by

$$y = \frac{p_d - c_f p}{1 + pc} d' \qquad (6.3-1)$$

Ideally one should select

$$c_f = \frac{p_d}{p} \qquad (6.3-2)$$

which leads to perfect disturbance compensation. If (6.3–2) is noncausal or unstable then perfect feedforward control is not possible, and c_f should be chosen to minimize the effect of d' on y. It is apparent from (6.3–1) that the optimal choice of c_f depends both on the feedback controller c and the nature of the

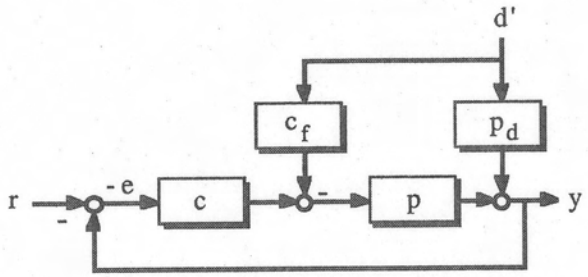


Figure 6.3-1. Classic feedforward/feedback control scheme.

disturbance d'. In the context of IMC the structure shown in Fig. 6.3-2 is recommended. Through some block diagram manipulations it can be shown that the IMC controllers are related to the classic controllers by

$$c = \frac{q}{1 - \tilde{p}q} \tag{6.3 - 3}$$

$$c_f = \frac{q_f - \tilde{p}_d q}{1 - \tilde{p}q} \tag{6.3 - 4}$$

The advantage of the IMC parametrization is that it separates the effect of the feedforward controller q_f and the feedback controller q and allows them to be designed independently. When there is no model error ($p = \tilde{p}, p_d = \tilde{p}_d$)

$$e = (p_d - pq_f)d' - (1 - pq)r \tag{6.3 - 5}$$

If

$$q_f = \frac{p_d}{p} \tag{6.3 - 6}$$

is causal and stable it can be implemented for perfect disturbance compensation. If it is not then it is natural to select q_f such that

$$\|(p_d - pq_f)d'\|_2 \tag{6.3 - 7}$$

is minimized.

6.3.2 Design

Theorem 6.3-1. *Assume that p and p_d are stable and that d' is bounded. The H_2-optimal feedforward controller $\tilde{q}_f$ minimizes*

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} |(p_d - p\tilde{q}_f)d'|^2 d\omega \tag{6.3 - 8}$$

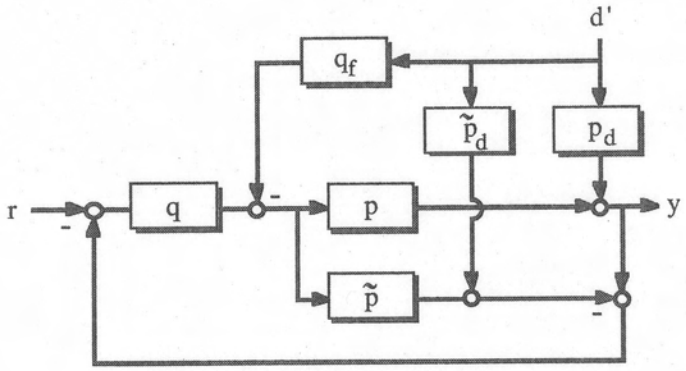


Figure 6.3-2. IMC feedforward/feedback scheme.

and is given by

$$\tilde{q}_f = (p_M d'_M)^{-1} \left\{ p_A^{-1} p_d d'_M \right\}_* \qquad (6.3-9)$$

where the operator $\{\ \}_*$ *denotes that after a partial fraction expansion of the operand all terms involving the poles of* p_A^{-1} *are omitted.*

Proof. The proof follows the same arguments as the proof of Thm. 5.2-1.

It is instructive to compare the feedforward design problem which aims to minimize (6.3–7) with the standard design problem discussed in Chap. 5 which has as its objective to minimize

$$\|(1 - p\tilde{q})p_d d'\|_2^2 \qquad (6.3-10)$$

In the latter case $p\tilde{q}$ is to approximate identity, in the former case $p\tilde{q}_f$ is to approximate p_d. If p_d and p happen to have the same NMP elements then $\tilde{q}_f$ can be chosen to make the approximation perfect (see 6.3–6). On the other hand, perfect control of NMP systems is not feasible through feedback control. *Thus, feedforward control can only be more effective than feedback if both* p *and* p_d *are NMP.* This is seen more clearly by comparing the error norms obtained with the two optimal controllers (6.3–9) and (4.1–7).

Feedforward:

$$\|p_d d'_M - p_A \left\{ p_A^{-1} p_d d'_M \right\}_* \|_2 \qquad (6.3-11)$$

Feedback:

$$\|p_{dM} d'_M - p_A \left\{ p_A^{-1} p_{dM} d'_M \right\}_* \|_2 \qquad (6.3-12)$$

If the process is MP ($p_A = 1$) then both (6.3–11) and (6.3–12) are zero. If the disturbance transfer function p_d is MP ($p_d = p_{dM}$) then (6.3–11) and (6.3–12) are identical and as far as nominal performance is concerned there is no incentive to use feedforward. Evaluation of (6.3–11) and (6.3–12) allows one to determine the potential improvements via feedforward control in the general case.

Example 6.3-1.

$$p = p_A = e^{-2s}, \quad p_d = e^{-s}, \quad d' = \frac{1}{s}$$

Feedforward:

$$q_f = s\{e^{2s} \cdot e^{-s} \cdot s^{-1}\}_*$$
$$= s\{e^{s} \cdot s^{-1}\}_* = s \cdot s^{-1} = 1$$
$$\|e\|_2 = \|e^{-s}s^{-1} - e^{-2s}s^{-1}\|_2 = \|(1 - e^{-s})s^{-1}\|_2 = 1$$

Feedback:

$$\|e\|_2 = \|(1 - e^{-2s})s^{-1}\|_2 = 2$$

□

Example 6.3-2.

$$p = p_A = \frac{-s+1}{s+1}, \quad p_d = e^{-2s}, \quad d' = \frac{1}{s}$$

Feedforward:

$$q_f = s\left\{\frac{(s+1)}{(-s+1)}e^{-2s}\frac{1}{s}\right\}_* = s\left[\frac{(s+1)e^{-2s}}{(-s+1)s} - \frac{2e^{-2}}{-s+1}\right] = \frac{s(e^{-2s} - 2e^{-2}) + e^{-2s}}{-s+1}$$

$$\|e\|_2 = \left\|e^{-2s}\frac{1}{s} - \frac{-s+1}{(s+1)s}\left[\frac{s(e^{-2s} - 2e^{-2}) + e^{-2s}}{-s+1}\right]\right\|_2 = \left\|\frac{2e^{-2}}{s+1}\right\|_2 = 2e^{-2} \cdot \frac{1}{\sqrt{2}} = 0.19$$

Feedback: From Thm. 4.1-4

$$\|e\|_2 = \sqrt{2} = 1.41$$

□

All the discussions about the advantages of feedforward over feedback in this section neglect model error. On one hand, model error forces the feedback controller to be detuned for robustness and nominal performance to be sacrificed. On the other hand, the performance of a feedforward controller is more sensitive to model mismatch than that of a feedback controller. A comparison of feedforward and feedback in the presence of model mismatch is a topic of current research.

6.3.3 Summary

If perfect feedforward compensation is possible then the IMC feedforward/feedback structure does not offer any specific advantage over the classic structure. If perfect compensation is not possible then the feedforward and the feedback designs "interact" in the classic structure

$$\frac{y}{d'} = \frac{p_d - c_f p}{1 + pc} \qquad (6.3-1)$$

but are *independent* for IMC. The H_2-optimal IMC feedforward controller is computed from

$$\tilde{q}_f = (p_M d_M')^{-1} \{p_A^{-1} p_d d_M'\}_* \qquad (6.3-9)$$

The achievable performance with feedforward is superior to that with feedback only if *both* the plant and the disturbance transfer function are NMP.

6.4 Cascade Control

6.4.1 Objectives, Structure, and Design

Cascade control (Fig. 6.4-1) is a specific form of a two-degree-of-freedom control structure which can be used if in addition to the output to be controlled (y_2) another process output (y_1) can be measured. It is popular to design c_1 first with the outer loop open and then to design c_2 for the new "plant" consisting of the closed inner loop combined with p_2. Four questions arise:

1. Consider the error e_2 in response to the inputs r_2, d_1 and d_2. Under what conditions is the performance which is achievable by the 2-loop structure superior to that which can be obtained with one controller alone — i.e., with the inner loop open?

2. Does the two step-design procedure, where the inner loop is designed first, limit the performance which is achievable with the outer loop?

3. How should c_1 and c_2 be designed?

4. How does model uncertainty affect design and performance?

1. Advantages of the Two-Controller Structure

If we assume that only a single loop (the outer one) is used, then the three inputs affect the output in the following manner

$$e_2 = (1 - p_1 p_2 q)(d_2 + p_2 d_1 - r_2) \qquad (6.4-1)$$

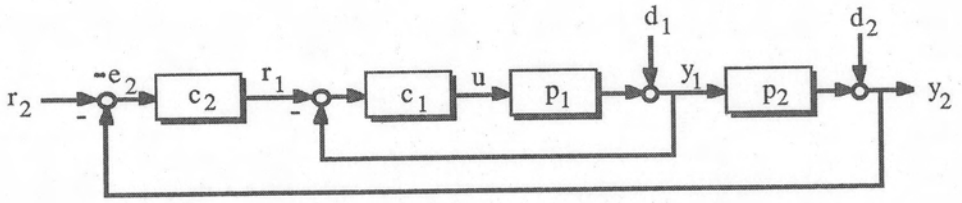


Figure 6.4-1. Cascade control system.

Here we have employed the IMC parametrization for the controller. If two controllers are employed then (6.4–1) becomes

$$e_2 = (1 - p_1q_1p_2q_2)(d_2 - r_2) + (1 - p_1q_1)p_2d_1 \qquad (6.4-2)$$

If we set $q_1q_2 = q$ and compare (6.4–1) and (6.4–2) we note that tracking of r_2 and suppression of d_2 are equally effective with a one and a two-controller structure. We rewrite the response to d_1

$$1-\text{loop}: \quad e_2 = (1 - p_1p_2q)p_2d_1 \qquad (6.4-3)$$

$$2-\text{loop}: \quad e_2 = (1 - p_1q_1)p_2d_1 \qquad (6.4-4)$$

The suppression of d_1 with the 2-loop structure is superior if q_1 can be designed such that

$$\min_{q_1} \|(1 - p_1q_1)p_2d_1\|_2 << \min_{q} \|(1 - p_1p_2q)p_2d_1\|_2 \qquad (6.4-5)$$

i.e., if the inner loop alone compensates for the disturbance p_2d_1 much better than the outer loop by itself ever could. It should be clear from our discussions of H_2-optimal control that (6.4–5) is an equality if p_2 is MP and stable. *Thus, a cascade controller is only useful if p_2 has RHP zeros or a time delay.*

2. Performance Limitations Imposed on Outer Loop

If the inner loop is effective ((6.4–5) holds), then the outer loop should be designed primarily for r_2 and d_2. The question is whether performance of the outer loop has been sacrificed by the design of the inner loop (q_1). Let us assume that the controller for the inner loop has been implemented in the form shown in Fig. 6.4-2 where

$$c_1c_3 = \frac{q_1}{1 - p_1q_1} \qquad (6.4-6)$$

We require that c_1 is MP and c_3 stable and that there are no RHP pole-zero cancellations between c_1 and c_3. Otherwise c_1 and c_3 are arbitrary. We showed in

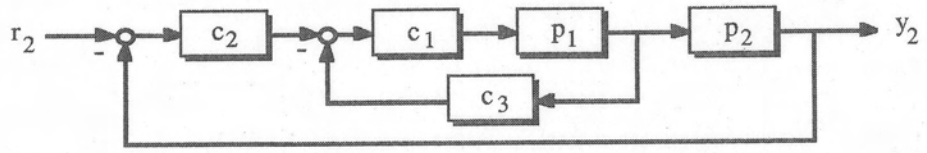


Figure 6.4-2. Cascade control system with additional controller block c_3.

Sec. 5.2.4 that with this choice the inner loop imposes no performance limitations on the outer loop. Furthermore the outer loop can be designed without regard for any RHP poles in p_1. When c_3 is MP or c_1 is stable then the number of control blocks can be reduced from three to two (Fig. 6.4-1) without effect on closed-loop performance or internal stability.

3. Design

Usually only two controller blocks (Fig. 6.4-1) are used for the implementation. Let us assume that (6.4–5) is satisfied as a strong inequality or in other words, that p_2 has significant NMP character. Then a two-step design procedure is reasonable where q_1 is designed first to solve

$$\min_{q_1} \|(1 - p_1 q_1) p_2 d_1\|_2 \qquad (6.4-7)$$

Then q_2 is determined from

$$\min_{q} \|(1 - p_1 p_2 q) d_2\|_2 \qquad (6.4-8)$$

$$q_2 = \frac{q}{q_1} \qquad (6.4-9)$$

(This assumes that q_1 is MP, otherwise more than the two blocks in Fig. 6.4-1 are necessary.) If $p_1 q_1 \cong 1$ over the expected bandwidth of the outer loop then it is even justifiable to assume for the outer loop design that the inner loop is under perfect control ($p_1 q_1 = 1$) and to obtain q_2 from

$$\min_{q_2} ||(1 - p_2 q_2) d_2||_2 \qquad (6.4-10)$$

4. Uncertainty

The preceding discussions dealt with the cases when model uncertainty was negligible. To date we have not explored cascade control under uncertainty in any depth. It is without question that very often the inner loop is introduced primarily to remove model uncertainty. For example, a flow control loop can correct for uncertain and possibly nonlinear valve characteristics.

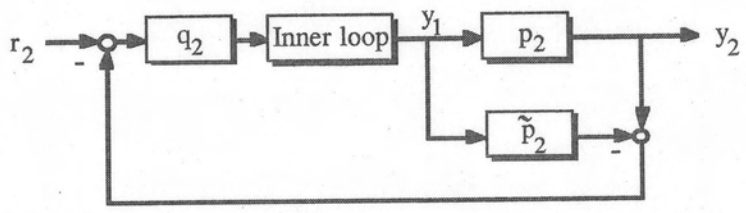


Figure 6.4-3. IMC implementation of cascade control system to avoid reset windup.

6.4.2 Implementation

The cascade controller implementation shown in Fig. 6.4-1 or 6.4-2 can lead to serious reset windup problems. Reset windup can be avoided by informing the controller when its output, the manipulated variable, is saturating. Upon saturation the controller turns off the integral action as an "anti-reset windup" measure.

Difficulties arise because the output of controller c_2 is the setpoint r_1 of the inner loop rather than the true manipulated variable u. Reset windup occurs when r_1 continues to change after u has saturated.

The IMC implementation (Fig. 6.4-3) avoids this windup problem. The inner loop can be implemented in the IMC or the classic form. By feeding the true output y_1 of the inner loop to the model $\tilde{p}_2$, the outer loop will recognize any saturation of u as it is reflected in y_1 and windup does not occur (compare the discussion on classic feedback via IMC implementation in Sec. 3.2.3).

6.4.3 Summary

Cascade control in its basic form shown in Fig. 6.4-1 is useful under two circumstances:

(i) Disturbance d_1 is significant and p_2 is NMP.

(ii) Plant p_1 has much uncertainty associated with it — e.g., poorly known nonlinear valve characteristics — and the inner loop serves to remove this uncertainty.

Disregarding model uncertainty we arrived at the following conclusions:

(i) For the extended structure in Fig. 6.4-2 the achievable performance (regarding control of d_2 and r_2) with the outer loop is unaffected by the inner loop. If $c_3 = q_1$ is MP the same is true for the simple structure in Fig. 6.4-1.

(ii) When p_2 has significant NMP character then a simple two step design procedure is recommended:

(1) q_1 is designed to solve

$$\min_{q_1} \|(1 - p_1 q_1) p_2 d_1\|_2 \qquad (6.4-7)$$

(2) q_2 is determined from

$$\min_{q_2} \|(1 - p_2 q_2) d_2\|_2 \qquad (6.4-10)$$

(This assumes that $p_1 q_1 \cong 1$ over the expected bandwidth of the outer loop).

6.5 References

6.1. The ideas in this section were expressed originally by Morari, Skogestad & Rivera (1984). The presentation here follows closely that by Rivera, Skogestad & Morari (1986). They also studied the performance of the ideal (improper) PID controller for a first-order system with deadtime.

6.1.3. Similar settings for the PID controller were proposed by Smith & Corripio (1985). It should be noted that on commercial hardware different PID implementations can be found (see Smith & Corripio, 1985). The PID parameters proposed here have to be adapted accordingly. The form (6.1–20) was mentioned in the book by Åström & Wittenmark (1984). The tuning rules by Ziegler & Nichols were published in 1942. The rules by Cohen & Coon (1953) are limited to first-order deadtime models.

In case the process model is not of a form which leads naturally to PI or PID controller, model reduction has to be applied first or simultaneously with the controller design procedure (Rivera, 1987). This technique has been implemented in the ROBEX software (Lewin et al., 1987).

6.2. The predictor was first proposed by Smith (1957). Palmor & Shinnar (1981) discuss its optimality and robustness. They show that gain margin and phase margin are useless robustness measure for this case. Åström (1977) points

out that the Smith predictor is an extreme form of lead compensator. The insights presented in a paper by Horowitz (1983) are similar to the ones in this section. The tuning rule (6.2–14) was mentioned first by Brosilow (1979).