on-line tuning we can distinguish between three main approaches to controller design:

1. **Shaping of transfer functions.** In this approach the designer specifies the *magnitude* of some transfer function(s) as a function of frequency, and then finds a controller which gives the desired shape(s).
   (a) **Loop shaping.** This is the classical approach in which the magnitude of the open-loop transfer function, $L(j\omega)$, is shaped. Usually no optimization is involved and the designer aims to obtain $|L(j\omega)|$ with desired bandwidth, slopes etc. We will look at this approach in detail later in this chapter. However, classical loop shaping is difficult to apply for complicated systems, and one may then instead use the Glover-McFarlane $\mathcal{H}_\infty$ loop-shaping design presented in Chapter 9. The method consists of a second step where optimization is used to make an initial loop-shaping design more robust.
   (b) **Shaping of closed-loop transfer functions, such as S, T and KS.** Optimization is usually used, resulting in various $\mathcal{H}_\infty$ optimal control problems such as mixed weighted sensitivity; more on this later.
2. **The signal-based approach.** This involves time domain problem formulations resulting in the minimization of a norm of a transfer function. Here one considers a particular disturbance or reference change and then one tries to optimize the closed-loop response. The "modern" state-space methods from the 1960's, such as Linear Quadratic Gaussian (LQG) control, are based on this signal-oriented approach. In LQG the input signals are assumed to be stochastic (or alternatively impulses in a deterministic setting) and the expected value of the output variance (or the 2-norm) is minimized. These methods may be generalized to include frequency dependent weights on the signals leading to what is called the Wiener-Hopf (or $\mathcal{H}_2$-norm) design method.
   By considering sinusoidal signals, frequency-by-frequency, a signal-based $\mathcal{H}_\infty$ optimal control methodology can be derived in which the $\mathcal{H}_\infty$ norm of a combination of closed-loop transfer functions is minimized. This approach has attracted significant interest, and may be combined with model uncertainty representations, to yield quite complex robust performance problems requiring μ-synthesis; an important topic which will be addressed in later chapters.
3. **Numerical optimization.** This often involves multi-objective optimization where one attempts to optimize directly the true objectives, such as rise times, stability margins, etc. Computationally, such optimization problems may be difficult to solve, especially if one does not have convexity. Also, by effectively including performance evaluation and controller design in a single step procedure, the problem formulation is far more critical than in iterative two-step approaches. The numerical optimization approach may also be performed on-line, which might be useful when dealing with cases with constraints on the inputs and outputs. On-line optimization approaches such as model predictive control are likely to become more popular as faster computers and more efficient and reliable computational algorithms are developed.

2.6 Loop shaping

In the classical loop-shaping approach to controller design, "loop shape" refers to the magnitude of the loop transfer function $L = GK$ as a function of frequency. An understanding of how K can be selected to shape this loop gain provides invaluable insight into the multivariable techniques and concepts which will presented later in the book, and so we will discuss loop shaping in some detail in the next two sections.

2.6.1 Trade-offs in terms of L

Recall equation (2.19), which yields the closed-loop response in terms of the control error $e = y - r$:

$$e = -\underbrace{(I+L)^{-1}}_{S} r + \underbrace{(I+L)^{-1}}_{S} G_d d - \underbrace{(I+L)^{-1}L}_{T} n \tag{2.48}$$

For "perfect control" we want $e = y - r = 0$; that is, we would like

$$e \approx 0 \cdot d + 0 \cdot r + 0 \cdot n \tag{2.49}$$

The first two requirements in this equation, namely disturbance rejection and command tracking, are obtained with $S \approx 0$, or equivalently, $T \approx I$. Since $S = (I+L)^{-1}$, this implies that the loop transfer function L must be large in magnitude. On the other hand, the requirement for zero noise transmission implies that $T \approx 0$, or equivalently, $S \approx I$, which is obtained with $L \approx 0$. This illustrates the fundamental nature of feedback design which always involves a trade-off between conflicting objectives; in this case between large loop gains for disturbance rejection and tracking, and small loop gains to reduce the effect of noise.

It is also important to consider the magnitude of the control action u (which is the input to the plant). We want u small because this causes less wear and saves input energy, and also because u is often a disturbance to other parts of the system (e.g. consider opening a window in your office to adjust your body temperature and the undesirable disturbance this will impose on the air conditioning system for the building). In particular, we usually want to avoid fast changes in u. The control action is given by $u = K(r - y_m)$ and we find as expected that a small u corresponds to small controller gains and a small $L = GK$.

The most important design objectives which necessitate trade-offs in feedback control are summarized below:

1. Performance, good disturbance rejection: needs large controller gains, i.e. L large.
2. Performance, good command following: L large.
3. Stabilization of unstable plant: L large.
4. Mitigation of measurement noise on plant outputs: L small.
5. Small magnitude of input signals: K small and L small.
6. Physical controller must be strictly proper: $K \to 0$ at high frequencies.
7. Nominal stability (stable plant): L small (because of RHP-zeros and time delays).
8. Robust stability (stable plant): L small (because of uncertain or neglected dynamics).

Fortunately, the conflicting design objectives mentioned above are generally in different frequency ranges, and we can meet most of the objectives by using a large loop gain ($|L| > 1$) at low frequencies below crossover, and a small gain ($|L| < 1$) at high frequencies above crossover.

2.6.2 Fundamentals of loop-shaping design

By *loop shaping* one usually means a design procedure that involves explicitly shaping the magnitude of the loop transfer function, $|L(j\omega)|$. Here $L(s) = G(s)K(s)$ where $K(s)$ is the feedback controller to be designed and $G(s)$ is the product of all other transfer functions around the loop, including the plant, the actuator and the measurement device. Essentially, to get the benefits of feedback control we want the loop gain, $|L(j\omega)|$, to be as large as possible within the bandwidth region. However, due to time delays, RHP-zeros, unmodelled high-frequency dynamics and limitations on the allowed manipulated inputs, the loop gain has to drop below one at and above some frequency which we call the crossover frequency ω_c. Thus, disregarding stability for the moment, it is desirable that $|L(j\omega)|$ falls sharply with frequency. To measure how $|L|$ falls with frequency we consider the logarithmic slope $N = d\ln|L|/d\ln\omega$. For example, a slope $N = -1$ implies that $|L|$ drops by a factor of 10 when ω increases by a factor of 10. If the gain is measured in decibels (dB) then a slope of $N = -1$ corresponds to -20 dB/ decade. The value of $-N$ at higher frequencies is often called the *roll-off* rate.

The design of $L(s)$ is most crucial and difficult in the crossover region between ω_c (where $|L| = 1$) and ω_{180} (where $\angle L = -180°$). For stability, we at least need the loop gain to be less than 1 at frequency ω_{180}, i.e., $|L(j\omega_{180})| < 1$.

Thus, to get a high bandwidth (fast response) we want ω_{180} large, that is, we want the phase lag in L to be small. Unfortunately, this is not consistent with the desire that $|L(j\omega)|$ should fall sharply. For example, the loop transfer function $L = 1/s^n$ (which has a slope $N = -n$ on a log-log plot) has a phase $\angle L = -n \cdot 90°$. Thus, to have a phase margin of 45° we need $\angle L > -135°$, and the slope of $|L|$ cannot exceed $N = -1.5$.

In addition, if the slope is made steeper at lower or higher frequencies, then this will add unwanted phase lag at intermediate frequencies. As an example, consider $L_1(s)$ given in (2.13) with the Bode plot shown in Figure 2.3. Here the slope of the asymptote of $|L|$ is -1 at the gain crossover frequency (where $|L_1(j\omega_c)| = 1$), which by itself gives $-90°$ phase lag. However, due to the influence of the steeper slopes of -2 at lower and higher frequencies, there is a "penalty" of about $-35°$ at crossover, so the actual phase of L_1 at ω_c is approximately $-125°$.

The situation becomes even worse for cases with delays or RHP-zeros in $L(s)$ which add undesirable phase lag to L without contributing to a desirable negative slope in L. At the gain crossover frequency ω_c, the additional phase lag from delays and RHP-zeros may in practice be $-30°$ or more.

In summary, a desired loop shape for $|L(j\omega)|$ typically has a slope of about -1 in the crossover region, and a slope of -2 or higher beyond this frequency, that is, the roll-off is 2 or larger. Also, with a proper controller, which is required for any real system, we must have that $L = GK$ rolls off at least as fast as G. At low frequencies, the desired shape of $|L|$ depends on what disturbances and references we are designing for. For example, if we are considering step changes in the references or disturbances which affect the outputs as steps, then a slope for $|L|$ of -1 at low frequencies is acceptable. If the references or disturbances require the outputs to change in a ramp-like fashion then a slope of -2 is required. In practice, integrators are included in the controller to get the desired low-frequency performance, and for offset-free reference tracking the rule is that

- *$L(s)$ must contain at least one integrator for each integrator in $r(s)$.*

To see this, let $L(s) = \widehat{L}(s)/s^{n_I}$ where $\widehat{L}(0)$ is nonzero and finite and n_I is the number of integrators in $L(s)$ — sometimes n_I is called the *system type*. Consider a reference signal of the form $r(s) = 1/s^{n_r}$. For example, if $r(t)$ is a unit step then $r(s) = 1/s$ ($n_r = 1$), and if $r(t)$ is a ramp then $r(s) = 1/s^2$ ($n_r = 2$). The final value theorem for Laplace transforms is

$$\lim_{t\to\infty} e(t) = \lim_{s\to 0} se(s) \tag{2.50}$$

In our case, the control error is

$$e(s) = -\frac{1}{1+L(s)}r(s) = -\frac{s^{n_I-n_r}}{s^{n_I}+\widehat{L}(s)}$$

and to get zero offset (i.e. $e(t \to \infty) = 0$) we must from (2.50) require $n_I \geq n_r$, and the rule follows. In Section 2.6.4, we discuss how to specify the loop shape when disturbance rejection is the primary objective of control.

In conclusion, one can define the desired loop transfer function in terms of the following specifications:

1. The gain crossover frequency, ω_c, where $|L(j\omega_c)| = 1$.
2. The shape of $L(j\omega)$, e.g., in terms of the slope of $|L(j\omega)|$ in certain frequency ranges. Typically, we desire a slope of about $N = -1$ around crossover, and a larger roll-off at higher frequencies. The desired slope at lower frequencies depends on the nature of the disturbance or reference signal.
3. The system type, defined as the number of pure integrators in $L(s)$.

Loop-shaping design is typically an iterative procedure where the designer shapes and reshapes $|L(j\omega)|$ after computing the phase and gain margins, the peaks of closed-loop frequency responses (M_T and M_S), selected closed-loop time responses, the magnitude of the input signal, etc. The procedure is illustrated next by an example.

Example 2.6 Loop-shaping design for the inverse response process.

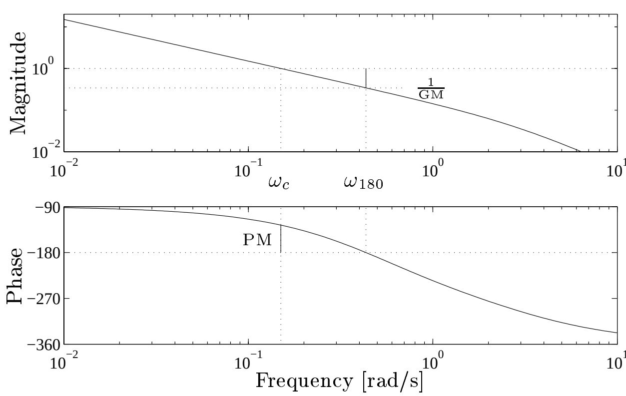


Figure 2.16: Frequency response of $L(s)$ in (2.51) for loop-shaping design with $K_c = 0.05$. (GM= 2.92, PM= 54°, $\omega_c = 0.15$, $\omega_{180} = 0.43$, $M_S = 1.75$, $M_T = 1.11$)

We will now design a loop-shaping controller for the example process in (2.27) which has a RHP-zero at $s = 0.5$. The RHP-zero limits the achievable bandwidth and so the crossover region (defined as the frequencies between ω_c and ω_{180}) will be at about 0.5 rad/s. We only require the system to have one integrator (type 1 system), and therefore a reasonable approach is to let the loop transfer function have a slope of -1 at low frequencies, and then to roll off with a higher slope at frequencies beyond 0.5 rad/s. We choose the following loop-shape

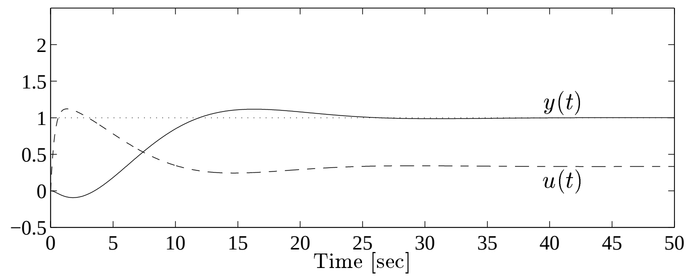


Figure 2.17: Response to step in reference for loop-shaping design.

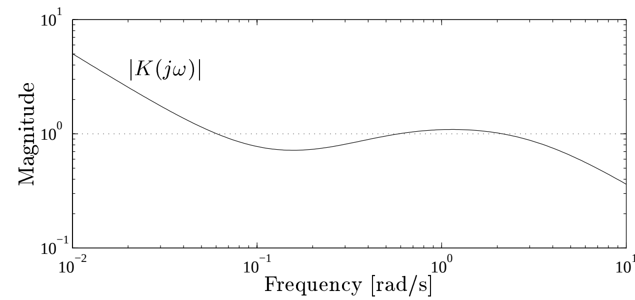


Figure 2.18: Magnitude Bode plot of controller (2.52) for loop-shaping design.

$$L(s) = 3K_c \frac{(-2s+1)}{s(2s+1)(0.33s+1)} \tag{2.51}$$

The frequency response (Bode plots) of L is shown in Figure 2.16. The asymptotic slope of $|L|$ is -1 up to 3 rad/s where it changes to -2. The controller corresponding to the loop-shape in (2.51) is

$$K(s) = K_c \frac{(10s+1)(5s+1)}{s(2s+1)(0.33s+1)}, \quad K_c = 0.05 \tag{2.52}$$

The controller has zeros at the locations of the plant poles. This is desired in this case because we do not want the slope of the loop shape to drop at the break frequencies $1/10 = 0.1$ [rad/s] and $1/5 = 0.2$ [rad/s] just before crossover. The controller gain K_c was selected to get a reasonable trade-off between speed of response and the robustness margin to instability. The phase of L is $-90°$ at low frequency, and *at $\omega = 0.5$ [rad/s] the additional contribution from the term $\frac{-2s+1}{2s+1}$ in (2.51) is -90°, so for stability we need $\omega_c < 0.5$ [rad/s]. The selection of $K_c = 0.05$ yields $\omega_c = 0.15$ [rad/s] corresponding to GM$= 2.92$ and PM$=54^\circ$. The corresponding time response is shown in Figure 2.17. It is seen to be much better than the responses with either the simple PI-controller in Figure 2.7 or with the P-controller in Figure 2.5. Figure 2.17 also shows that the magnitude of the input signal is reasonable (assuming the signals have been scaled such that we want the input to be less than about 1 in magnitude). This means that the controller gain is not too large at high frequencies. The magnitude Bode plot for the controller (2.52) is shown in Figure 2.18. It is interesting to note that in the crossover region around $\omega = 0.5$ [rad/s] the controller gain is quite constant, around 1 in magnitude, which is similar to the "best" gain found using a P-controller (see Figure 2.5).*

Limitations imposed by RHP-zeros and time delays.

Based on the above loop-shaping arguments we can now examine how the presence of delays and RHP-zeros limit the achievable control performance. We have already argued that if we want the loop shape to have a slope of -1 around crossover (ω_c), with preferably a steeper slope before and after crossover, then the phase lag of L at ω_c will necessarily be at least -90°, even when there are no RHP-zeros or delays. Therefore, if we assume that for performance and robustness we want a phase margin of about 35° or more, then the additional phase contribution from any delays and RHP-zeros at frequency ω_c cannot exceed about -55°.

First consider a time delay θ. It yields an additional phase contribution of $-\theta\omega$, which at frequency $\omega = 1/\theta$ is -1 rad $= -57^\circ$ (which is more than -55°). Thus, for acceptable control performance we need $\omega_c < 1/\theta$, approximately.

Next consider a real RHP-zero at $s = z$. To avoid an increase in slope caused by this zero we place a pole at $s = -z$ such that the loop transfer function contains the term $\frac{-s+z}{s+z}$, the form of which is referred to as all-pass since its magnitude equals 1 at all frequencies. The phase contribution from the all-pass term at $\omega = z/2$ is $-2\arctan(0.5) = -53^\circ$ (which is close to -55°), so for acceptable control performance we need $\omega_c < z/2$, approximately.

2.6.3 Inverse-based controller design

In Example 2.6, we made sure that $L(s)$ contained the RHP-zero of $G(s)$, but otherwise the specified $L(s)$ was independent of $G(s)$. This suggests the following possible approach for a minimum-phase plant (i.e, one with no RHP-zeros or time delays). Select a loop shape which has a slope of -1 throughout the frequency range, namely

$$L(s) = \frac{\omega_c}{s} \tag{2.53}$$

where ω_c is the desired gain crossover frequency. This loop shape yields a phase margin of 90° and an infinite gain margin since the phase of $L(j\omega)$ never reaches -180°. The controller corresponding to (2.53) is

$$K(s) = \frac{\omega_c}{s} G^{-1}(s) \tag{2.54}$$

That is, the controller inverts the plant and adds an integrator ($1/s$). This is an old idea, and is also the essential part of the IMC (Internal Model Control) design procedure of Morari (Morari and Zafiriou, 1989) which has proved successful in many applications. However, there are at least two good reasons for why this controller may not be a good choice:

1. The controller will not be realizable if $G(s)$ has more poles than zeros, and may in any case yield large input signals. These problems may be partly fixed by adding high-frequency dynamics to the controller.
2. The loop shape resulting from (2.53) is *not* generally desirable, unless the references and disturbances affect the outputs as steps. This is illustrated by the following example.

Example 2.7 Disturbance process. *We now introduce our second main example process and control problem in which disturbance rejection is an important objective in addition to command tracking. We assume that the plant has been appropriately scaled as outlined in Section 1.4.*

Problem formulation. *Consider the disturbance process described by*

$$\boxed{G(s) = \frac{200}{10s+1}\frac{1}{(0.05s+1)^2}, \quad G_d(s) = \frac{100}{10s+1}} \tag{2.55}$$

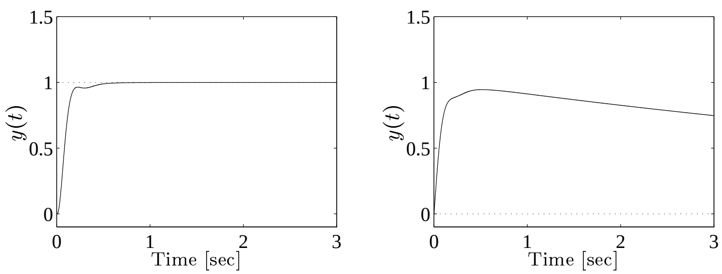
with time in seconds. A block diagram is shown in Figure 2.20. The control objectives are:

1. *Command tracking: The rise time (to reach 90% of the final value) should be less than 0.3 [s] and the overshoot should be less than 5%.*
2. *Disturbance rejection: The output in response to a unit step disturbance should remain within the range $[-1, 1]$ at all times, and it should return to 0 as quickly as possible ($|y(t)|$ should at least be less than 0.1 after 3 s).*
3. *Input constraints: $u(t)$ should remain within the range $[-1, 1]$ at all times to avoid input saturation (this is easily satisfied for most designs).*

Analysis. *Since $G_d(0) = 100$ we have that without control the output response to a unit disturbance ($d = 1$) will be 100 times larger than what is deemed to be acceptable. The magnitude $|G_d(j\omega)|$ is lower at higher frequencies, but it remains larger than 1 up to $\omega_d \approx 10$ [rad/s] (where $|G_d(j\omega_d)| = 1$). Thus, feedback control is needed up to frequency ω_d, so we need ω_c to be approximately equal to 10 rad/s for disturbance rejection. On the other hand, we do not want ω_c to be larger than necessary because of sensitivity to noise and stability problems associated with high gain feedback.* We will thus aim at a design with $\omega_c \approx 10$ [rad/s].

Inverse-based controller design. *We will consider the "inverse-based" design as given by (2.53) and (2.54) with* $\omega_c = 10$. *This yields an unrealizable controller and therefore we choose to approximate the plant term* $(0.05s+1)^2$ *by* $(0.1s+1)$ *and then in the controller we let this term be effective over one decade, i.e., we use* $(0.1s+1)/(0.01s+1)$ *to give the realizable design*

$$K_0(s) = \frac{\omega_c}{s}\frac{10s+1}{200}\frac{0.1s+1}{0.01s+1}, \quad L_0(s) = \frac{\omega_c}{s}\frac{0.1s+1}{(0.05s+1)^2(0.01s+1)}, \quad \omega_c = 10 \tag{2.56}$$



(a) Tracking response. (b) Disturbance response.

Figure 2.19: Responses with "inverse-based" controller $K_0(s)$ for disturbance process.

The response to a step reference is excellent as shown in Figure 2.19 (a). The rise time is about 0.16 *s and there is no overshoot so the specifications are more than satisfied. However, the response to a step disturbance (Figure 2.19 (b)) is much too sluggish. Although the output stays within the range* $[-1, 1]$, *it is still* 0.75 *at* $t = 3$ *s (whereas it should be less than* 0.1*). Because of the integral action the output does eventually return to zero, but it does not drop below* 0.1 *until after* 23 *s.*

The above example illustrates that the simple "inverse-based" design method where L has a slope of about $N = -1$ at all frequencies, does not always yield satisfactory designs. The objective of the next section is to understand why the disturbance response was so poor, and to propose a more desirable loop shape for disturbance rejection.

2.6.4 Loop shaping for disturbance rejection

At the outset we assume that the disturbance has been scaled such that at each frequency $|d(\omega)| \le 1$, and the main control objective is to achieve $|e(\omega)| < 1$. With feedback control we have $e = y = SG_d d$, so to achieve $|e(\omega)| \le 1$ for $|d(\omega)| = 1$ (the worst-case disturbance) we require $|SG_d(j\omega)| < 1, \forall\omega$, or equivalently,

$$|1+L| \ge |G_d| \quad \forall\omega \tag{2.57}$$

At frequencies where $|G_d| > 1$, this is approximately the same as requiring $|L| > |G_d|$. However, in order to minimize the input signals, thereby reducing the sensitivity to noise and avoiding stability problems, we do not want to use larger loop gains than necessary (at least at frequencies around crossover). A reasonable initial loop shape $L_{\min}(s)$ is then one that just satisfies the condition

$$|L_{\min}| \approx |G_d| \tag{2.58}$$

where the subscript *min* signifies that $L_{\min}$ is the smallest loop gain to satisfy $|e(\omega)| \le 1$. Since $L = GK$ the corresponding controller with the minimum gain satisfies

$$|K_{\min}| \approx |G^{-1}G_d| \tag{2.59}$$

In addition, to improve low-frequency performance (e.g. to get zero steady-state offset), we often add integral action at low frequencies, and use

$$|K| = |\frac{s+\omega_I}{s}||G^{-1}G_d| \tag{2.60}$$

This can be summarized as follows:

- For disturbance rejection a good choice for the controller is one which contains the dynamics (G_d) of the disturbance and inverts the dynamics (G) of the inputs (at least at frequencies just before crossover).
- For disturbances entering directly at the plant output, $G_d = 1$, and we get $|K_{\min}| = |G^{-1}|$, so an inverse-based design provides the best trade-off between performance (disturbance rejection) and minimum use of feedback.
- For disturbances entering directly at the plant input (which is a common situation in practice – often referred to as a load disturbance), we have $G_d = G$ and we get $|K_{\min}| = 1$, so a simple proportional controller with unit gain yields a good trade-off between output performance and input usage.
- Notice that a reference change may be viewed as a disturbance directly affecting the output. This follows from (1.17), from which we get that a maximum reference change $r = R$ may be viewed as a disturbance $d = 1$ with $G_d(s) = -R$ where R is usually a constant. This explains why selecting K to be like G^{-1} (an inverse-based controller) yields good responses to step changes in the reference.

In addition to satisfying $|L| \approx |G_d|$ (eq. 2.58) at frequencies around crossover, the desired loop-shape $L(s)$ may be modified as follows:

1. Around crossover make the slope N of $|L|$ to be about -1. This is to achieve good transient behaviour with acceptable gain and phase margins.
2. Increase the loop gain at low frequencies as illustrated in (2.60) to improve the settling time and to reduce the steady-state offset. Adding an integrator yields zero steady-state offset to a step disturbance.

3. Let $L(s)$ roll off faster at higher frequencies (beyond the bandwidth) in order to reduce the use of manipulated inputs, to make the controller realizable and to reduce the effects of noise.

The above requirements are concerned with the magnitude, $|L(j\omega)|$. In addition, the dynamics (phase) of $L(s)$ must be selected such that the closed-loop system is stable. When selecting $L(s)$ to satisfy $|L| \approx |G_d|$ one should replace $G_d(s)$ by the corresponding minimum-phase transfer function with the same magnitude, that is, time delays and RHP-zeros in $G_d(s)$ should not be included in $L(s)$ as this will impose undesirable limitations on feedback. On the other hand, any time delays or RHP-zeros in $G(s)$ must be included in $L = GK$ because RHP pole-zero cancellations between $G(s)$ and $K(s)$ yield internal instability, see Chapter 4.

Remark. The idea of including a disturbance model in the controller is well known and is more rigorously presented in, for example, research on the internal model principle (Wonham, 1974), or the internal model control design for disturbances (Morari and Zafiriou, 1989). However, our development is simple, and sufficient for gaining the insight needed for later chapters.

Example 2.8 Loop-shaping design for the disturbance process

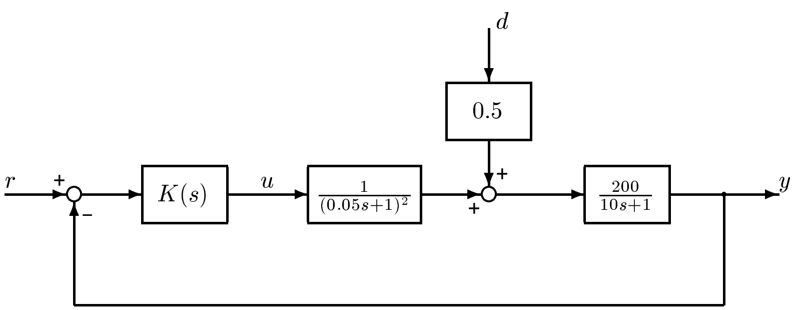


Figure 2.20: Block diagram representation of the disturbance process in (2.55)

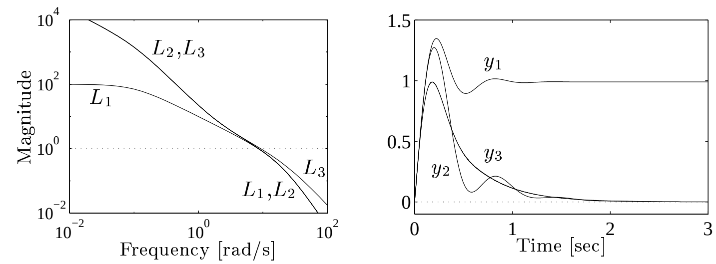
Consider again the plant described by (2.55). The plant can be represented by the block diagram in Figure 2.20, and we see that the disturbance enters at the plant input in the sense that G and G_d share the same dominating dynamics as represented by the term $200/(10s+1)$.

Step 1. Initial design. *From (2.58) we know that a good initial loop shape looks like $|L_{\min}| = |G_d| = \left|\frac{100}{10s+1}\right|$ at frequencies up to crossover. The corresponding controller is $K(s) = G^{-1}L_{\min} = 0.5(0.05s+1)^2$. This controller is not proper (i.e, it has more zeros than poles), but since the term $(0.05s+1)^2$ only comes into effect at $1/0.05 = 20$*

[rad/s], which is beyond the desired gain crossover frequency $\omega_c = 10$ [rad/s], we may replace it by a constant gain of 1 resulting in a proportional controller

$$K_1(s) = 0.5 \tag{2.61}$$

The magnitude of the corresponding loop transfer function, $|L_1(j\omega)|$, and the response ($y_1(t)$) to a step change in the disturbance are shown in Figure 2.21. This simple controller works surprisingly well, and for $t < 3s$ the response to a step change in the disturbance response is not much different from that with the more complicated inverse-based controller $K_0(s)$ of (2.56) as shown earlier in Figure 2.19. However, there is no integral action and $y_1(t) \to 1$ as $t \to \infty$.



(a) Loop gains. (b) Disturbance responses.

Figure 2.21: Loop shapes and disturbance responses for controllers K_1, K_2 and K_3 for the disturbance process.

Step 2. More gain at low frequency. *To get integral action we multiply the controller by the term $\frac{s+\omega_I}{s}$, where ω_I is the frequency up to which the term is effective (the asymptotic value of the term is 1 for $\omega > \omega_I$). For performance we want large gains at low frequencies, so we want ω_I to be large, but in order to maintain an acceptable phase margin (which is 44.7° for controller K_1) the term should not add too much negative phase at frequency ω_c, so ω_I should not be too large. A reasonable value is $\omega_I = 0.2\omega_c$ for which the phase contribution from $\frac{s+\omega_I}{s}$ is $\arctan(1/0.2) - 90^\circ = -11^\circ$ at ω_c. In our case $\omega_c \approx 10$ [rad/s], so we select the following controller*

$$K_2(s) = 0.5\frac{s+2}{s} \tag{2.62}$$

The resulting disturbance response (y_2) in shown in Figure 2.21 satisfies the requirement that $|y(t)| < 0.1$ at time $t = 3$ s, but $y(t)$ exceeds 1 for a short time. Also, the response is slightly oscillatory as might be expected since the phase margin is only 31° and the peak values for $|S|$ and $|T|$ are $M_S = 2.28$ and $M_T = 1.89$ (see Table 2.2).

Step 3. High-frequency correction. *To increase the phase margin and improve the transient response we supplement the controller with "derivative action" by multiplying $K_2(s)$ by a lead-lag term which is effective over one decade starting* *at 20 rad/s:*

$$K_3(s) = 0.5\frac{s+2}{s}\,\frac{0.05s+1}{0.005s+1} \tag{2.63}$$

The corresponding disturbance response (y_3) is seen to be faster initially and $y_3(t)$ stays below 1.

Table 2.2: Alternative loop-shaping designs for the disturbance process

	GM	PM	ω_c	M_S	M_T	Reference		Disturbance	
						t_r	$y_{\max}$	$y_{\max}$	$y(t=3)$
Spec.→			≈ 10			$\leq .3$	≤ 1.05	≤ 1	≤ 0.1
K_0	9.95	72.9°	11.4	1.34	1	.16	1.00	0.95	.75
K_1	4.04	44.7°	8.48	1.83	1.33	.21	1.24	1.35	.99
K_2	3.24	30.9°	8.65	2.28	1.89	.19	1.51	1.27	.001
K_3	19.7	50.9°	9.27	1.43	1.23	.16	1.24	0.99	.001

Table 2.2 summarizes the results for the four loop-shaping designs; the inverse-based design K_0 for reference tracking and the three designs K_1, K_2 and K_3 for disturbance rejection. Although controller K_3 satisfies the requirements for disturbance rejection, it is not *satisfactory for reference tracking; the overshoot is 24% which is significantly higher than the maximum value of 5%. On the other hand, the inverse-based controller K_0 inverts the term $1/(10s+1)$ which is also in the disturbance model, and therefore yields a very sluggish response to disturbances (the output is still 0.75 at $t = 3$ s whereas it should be less than 0.1).*

2.6.5 Two degrees-of-freedom design

For the disturbance process example we see from Table 2.2 that none of the controller designs meet all the objectives for both reference tracking and disturbance rejection. The problem is that for reference tracking we typically want the controller to look like $\frac{1}{s}G^{-1}$ see (2.54), whereas for disturbance rejection we want the controller to look like $\frac{1}{s}G^{-1}G_d$, see (2.60). We cannot achieve both of these simultaneously with a single (feedback) controller.

The solution is to use a two degrees-of-freedom controller where the reference signal r and output measurement y_m are independently treated by the controller, rather than operating on their difference $r - y_m$. There exist several alternative implementations of a two degrees-of-freedom controller. The most general form is shown in Figure 1.3(b) on page 12 where the controller has two inputs (r and y_m) and one output (u). However, the controller is often split into two separate blocks as shown in Figure 2.22 where K_y denotes the feedback part of the controller and K_r a reference prefilter. The feedback controller K_y is used to reduce the effect of uncertainty (disturbances and model error) whereas the prefilter K_r shapes the commands to improve performance. In general, it is optimal to design the combined two degrees-of-

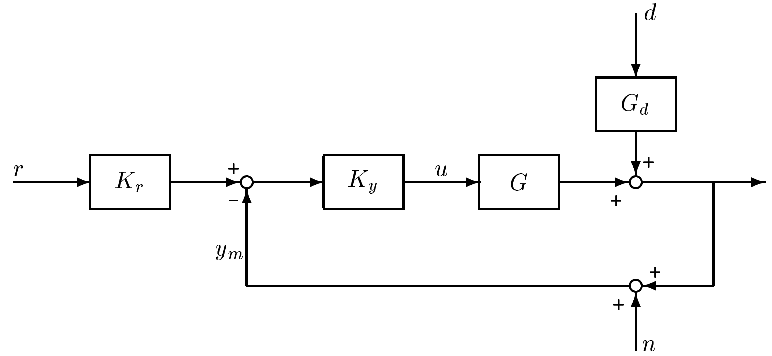


Figure 2.22: Two degrees-of-freedom controller.

freedom controller K in one step. However, in practice K_y is often designed first for disturbance rejection, and then K_r is designed to improve reference tracking. This is the approach taken here.

Let $T = L(1+L)^{-1}$ (with $L = GK_y$) denote the complementary sensitivity function for the feedback system. Then for a one degree-of-freedom controller $y = Tr$, whereas for a two degrees-of-freedom controller $y = TK_r r$. If the desired transfer function for reference tracking (often denoted the reference model) is T_{ref}, then the corresponding ideal reference prefilter K_r satisfies $TK_r = T_{\text{ref}}$, or

$$K_r(s) = T^{-1}(s)T_{\text{ref}}(s) \tag{2.64}$$

Thus, in theory we may design $K_r(s)$ to get any desired tracking response $T_{\text{ref}}(s)$. However, in practice it is not so simple because the resulting $K_r(s)$ may be unstable (if $G(s)$ has RHP-zeros) or unrealizable, and relatively uncertain if $T(s)$ is not known exactly. A convenient practical choice of prefilter is the lead-lag network

$$K_r(s) = \frac{\tau_{\text{lead}}s+1}{\tau_{\text{lag}}s+1} \tag{2.65}$$

Here we select $\tau_{\text{lead}} > \tau_{\text{lag}}$ if we want to speed up the response, and $\tau_{\text{lead}} < \tau_{\text{lag}}$ if we want to slow down the response. If one does not require fast reference tracking, which is the case in many process control applications, a simple lag is often used (with $\tau_{\text{lead}} = 0$).

Example 2.9 Two degrees-of-freedom design for the disturbance process

In Example 2.8 we designed a loop-shaping controller $K_3(s)$ for the plant in (2.55) which gave good performance with respect to disturbances. However, the command