1997 European Control Conference (ECC)
1-4 July 1997, Brussels, Belgium

LIMITATIONS ON CONTROL SYSTEM PERFORMANCE[†]

Karl J. Åström

Department of Automatic Control, Lund University
Box 118, S-221 00 Lund, Sweden
Email: kja@Control.LTH.SE, Fax +46-46-138118

Keywords Control design, RHP poles and zeros, time delays, measurement noise, bandwidth.

Abstract

This paper deals with limitations on control system performance due to load disturbances, measurement noise, actuator saturation and system dynamics. Simple problems which capture the essence of the problem are formulated and solved. The results make it possible to quickly explore a design problem for preliminary assessment before attempting to do massive computations. They are also useful ingredients in basic courses on control.

1. Introduction

Much research has been devoted to developing methods for design of control system. A common approach has been to strive for optimality, see [7]. Optimization methods give the the best performance with the specified criteria and constraints if the goals are achievable. They do not tell what to do if the goals cannot be attanined and they seldom give good insight into the mechanisms that cause the limitations. It is therefore desirable to have techniques that give insight into the factors that fundamentally limit the achievable performance of a control system. Such methods are also very useful in systems with higher automation levels where performance assessment is done automatically, see [1]. Fundamental problems were investigated in early work of Bode, see [3], but it has not received much interest after that. The situation is very different in information theory, where Shannon's results on limitations on information transmission is one of the corner stones. Lately there has however been a renewed interest in these questions starting with [5] and followed by [8], [4], [2], [9]. The usefulness of investigating fundamental limitations was emphasized very clearly in [10].

[†] This work was supported by the Swedish Research Council for Engineering Sciences (TFR) under contract TFR 95-759.

2. Minimum Phase Systems

Minimum phase systems are easy to control. Since the process poles and zeros are in the left half plane they can be canceled freely to obtain a desired loop transfer function L. A simple design procedure is to choose a desired complementary sensitivity function T. The corresponding loop transfer function is then $L = T/(1-T)$ and the controller becomes $C = L/P$. Such a controller is said to have a high control authority because the properties of the process have no influence on the closed loop response. If disturbances and measurement noise are neglected it is possible to obtain a closed loop system with arbitrarily high bandwidth. This is, however, not true in the presence of measurement noise and actuator saturation. Measurement noise is injected into the system and they may result in large control signals that saturate the actuators. Measurement noise and actuator saturations are thus factors that limit the performance for a minimum phase system.

The key problem is to determine how the controller gain is influenced by the specifications. Consider a process whose transfer function is minimum phase. Assume that a gain crossover frequency ω_{gc} is desired. The phase of the process transfer function at ω_{gc} is $\arg G(i\omega_{gc})$. A phase lead $\varphi = \varphi_m - \arg G(i\omega_{gc}) - \pi$ is thus required. It follows from Bode's phase area formula [3] that the gain required is given by $N = e^{2\gamma\varphi}$ where γ is close to 1, see [2]. Assuming that the lead compensator is symmetrical around ω_{gc} it follows that the compensator gain at crossover is $\sqrt{N}$. The gain of the controller must also be adjusted so that the loop gain at ω_{gc} is one. This implies that

$$|G(i\omega_{gc})|K\sqrt{N} = 1, \tag{1}$$

where K is the controller gain. The maximum controller gain K_c is then

$$K_c = NK = \frac{\sqrt{N}}{|G(i\omega_{gc})|}.$$

Equation (1) then gives

$$K_c = \frac{1}{|G(i\omega_{gc})|} e^{\gamma(\varphi_m - \arg G(i\omega_{gc}) - \pi)}. \tag{2}$$

This equation gives an estimate of the controller gain required to obtain a specified bandwidth. The particular value of γ depends on the details of the design method, a reasonable value is $\gamma = 1$. We will illustrate the result with and example.

EXAMPLE 1
Consider a system with the transfer function $G(s) = 1/(s+1)^n$. Equation (2) gives

$$K_c = \left(1 + \omega_{gc}^2\right)^{n/2} e^{\gamma(\varphi_m + n \arctan \omega_{gc} - \pi)}$$

With $\gamma = 1$, $\varphi_m = 45^\circ$ and $n = 2$ we get the numerical values in Table 1. With a K_c restricted to 1000 we find that the maximum crossover frequency is about 22.3 rad/s. Notice that the main contribution is due to the process gain. The phase advance required is not more than 40°.

The corresponding results for a system with $n = 8$ are shown in Table 2. If the maximum controller gain is $K_c = 1000$ the crossover frequency must be less than 1 rad/s. A substantial portion of the gain is due to the phase advance required. Notice that the gain increases very rapidly with the crossover frequency because of the exponential term in Equation (2). □

Table 1 Controller gain and phase advance required to obtain a given crossover frequency for the process $G(s) = 1/(s+1)^2$

.

Crossover frequency	10	20	50
Controller gain	182	796	5270
Phase lead	34	39	43

Table 2 Controller gain and phase advance required to obtain a given crossover frequency for the process $G(s) = 1/(s+1)^8$

.

Crossover frequency	0.5	0.75	1	1.5
Controller gain	9.4	97	812	27470
Phase lead	78	160	225	315

3. Non-Minimum Phase Systems

For systems with poles and zeros in the right half plane there are additional restrictions that are due to the process dynamics. Consider a system with transfer function $G(s)$, and factor it as

$$G(s) = G_{mp}(s)G_{nmp}(s), \tag{3}$$

where G_{mp} is the minimum phase part and G_{nmp} is the non-minimum phase part. It is assumed that the factorization is normalized so that $|G_{nmp}(0)| = 1$ and the sign is chosen to satisfy the requirement on encirclement by the Nyquist criterion. We choose to characterize the bandwidth by the gain crossover frequency ω_{gc}.

The Design Inequality

We will now derive an inequality for the gain crossover frequency. The loop transfer function is $L(s) = G(s)C(s)$. Requiring a phase margin φ_m we get

$$\begin{aligned} \arg L(i\omega_{gc}) &= \arg G_{nmp}(i\omega_{gc}) + \arg G_{mp}(i\omega_{gc}) \\ &\quad + \arg C(i\omega_{gc}) \geq \varphi_m - \pi, \end{aligned} \tag{4}$$

recall that $C(s)$ is the controller transfer function. To proceed it is necessary to make some assumptions about the controller transfer function. Let the controller be chosen so that $G_{mp}C$ corresponds to Bode's ideal loop transfer function [3], $L(s) = (s/\omega_{gc})^n$, then

$$\arg G_{mp}(i\omega) + \arg C(i\omega) = n\frac{\pi}{2} \tag{5}$$

where n is the slope of the compensated minimum phase part of the system. This can always be achieved with arbitrary accuracy by loop shaping. Equation (5) is actually a good approximation for most controllers if n is chosen as the slope n_{gc} at the crossover frequency. This follows from Bode's relations between amplitude and phase [3].

If there is some uncertainty ΔG in the process transfer function G the inequality (4) becomes

$$\arg G_{nmp}(i\omega_{gc}) + \arg \Delta G(i\omega_{gc}) \geq \varphi_m - (1 + \frac{n_{gc}}{2})\pi. \tag{6}$$

This equation which we call *the design inequality* gives bounds on the crossover frequency ω_{gc}. Since the magnitude of the loop transfer function must decrease at ω_{gc}, it must be required that n_{gc} is negative. For minimum-phase systems the inequality reduces to

$$n_{gc} = -2 + 2\frac{\varphi_m}{\pi}$$

. Typical values are $-1.7 < n_{gc} < -1.3$ for minimum phase systems. Larger values are used for non-minimum phase systems, typically $n_{gc} = -0.5$.

Systems with RHP Zeros

We will now discuss limitations imposed by right half plane zeros. First consider systems with one zero in the right half plane. The non-minimum phase part of the transfer function is

$$G_{nmp}(s) = \frac{a-s}{a+s}. \tag{7}$$

The low frequency gain must be positive to satisfy the encirclement condition of the Nyquist stability criterion. We have

$$\arg G_{nmp}(i\omega) = -2\arctan\frac{\omega}{a}$$

. Since the phase of G_{nmp} decreases with frequency the inequality (6) gives an upper bound on the crossover frequency. Neglecting process uncertainties it follows from the design inequality, Equation (6) that

$$n_{gc}\frac{\pi}{2} - 2\arctan\frac{\omega_{gc}}{a} \geq -\pi + \varphi_m$$

. Hence

$$\frac{\omega_{gc}}{a} \leq \tan(\frac{\pi}{2} - \frac{\varphi_m}{2} + n_{gc}\frac{\pi}{4}). \tag{8}$$

EXAMPLE 2—RHP ZERO AT $s = a$

Consider a process with a right half plane zero at $s = a$. Let the phase margin be $\varphi_m = \pi/4$ and the slope $n_{gc} = -1/2$, then $\omega_{gc} < a$. □

A right half plane zero thus gives an upper bound to the achievable bandwidth. The bandwidth decreases with decreasing frequency of the zero. It is thus more difficult to control systems with slow zeros.

Systems with Time Delays

System with dead time will now be investigated. The transfer function for such systems has an essential singularity at infinity. The non-minimum phase part of the transfer function of the process is

$$G_{nmp}(s) = e^{-s\tau}, \tag{9}$$

hence, $\arg G_{nmp}(i\omega) = -\omega\tau$. Neglecting process uncertainties it follows from the design inequality (6) that

$$\omega_{gc}\tau \leq \pi - \varphi_m + n_{gc}\frac{\pi}{2}. \tag{10}$$

EXAMPLE 3—TIME DELAYS

Consider a system with time delay τ. Let the phase margin be $\varphi_m = \pi/4$ and the slope $n_{gc} = -1/2$, then $\omega_{gc}\tau \leq \frac{\pi}{2}$. □

Time delays thus give an upper bound on the achievable bandwidth.

Systems with RHP Poles

The limitations imposed by right half plane poles will now be investigated. Consider a system with one pole in the right half plane. Thus, the non-minimum phase part of the transfer function is

$$G_{nmp}(s) = \frac{s+b}{s-b}. \tag{11}$$

To have a stable closed loop system it follows from Nyquist's stability criterion that the Nyquist curve must make one encirclement of the critical point. The transfer function G_{nmp} has been parameterized to preserve this condition. We have

$$\arg G_{nmp}(i\omega) = -2\arctan\frac{b}{\omega}$$

. Neglecting process uncertainties it follows from the design inequality, Equation (6), that

$$n_{gc}\frac{\pi}{2} - 2\arctan\frac{b}{\omega_{gc}} \geq -\pi + \varphi_m$$

. Hence

$$\omega_{gc} \geq \frac{b}{\tan(\pi/2 - \varphi_m/2 + n_{gc}\pi/4)}. \tag{12}$$

EXAMPLE 4—RHP POLE AT $s = b$

Consider a system with a right half plane pole at $s = b$. Let the phase margin be $\varphi_m = \pi/4$ and the slope $n_{gc} = -1/2$, then $\omega_{gc} \geq b$. □

Unstable poles thus give a lower bound on the crossover frequency.

For systems with right half plane poles the bandwidth must thus be sufficiently large. By computing the phase lag of the minimum phase part of the system at ω_{gc} we can determine the phase lead required to compensate the minimum phase part. Equation (2) then gives the gain required to achieve the phase lead. Knowledge of the measurement noise and the saturation levels of the control signal then indicates the feasibility of stabilizing the system.

Systems with fast unstable poles are more difficult to control than systems with slow unstable poles. The opposite is true for unstable zeros.

Systems with RHP Poles and Zeros

The calculations can be extended to cases with both poles and zeros in the right half plane. We will give the results for a pole-zero pair. The non-minimum phase part of the transfer function is then

$$G_{nmp}(s) = \frac{(a-s)(s+b)}{(a+s)(s-b)} \tag{13}$$

For $a > b$ we have

$$\begin{aligned}\arg G_{nmp}(i\omega) &= -2\arctan\frac{\omega}{a} - 2\arctan\frac{b}{\omega} \\ &= -2\arctan\frac{\omega/a + b/\omega}{1 - b/a}.\end{aligned}$$

Neglecting process uncertainties it follows from the design inequality, Equation (6), that

$$\frac{\omega_{gc}}{a} + \frac{b}{\omega_{gc}} \leq (1 - \frac{b}{a})\alpha$$

where

$$\alpha = \tan(\frac{\pi}{2} - \frac{\varphi_m}{2} + n_{gc}\frac{\pi}{4}). \tag{14}$$

The left hand side has its smallest value for $\omega_{gc} = \sqrt{ab}$, i.e. the geometric mean of the unstable pole and zero. Hence

$$2\sqrt{\frac{b}{a}} \leq \alpha(1 - \frac{b}{a})$$

. This implies

$$\frac{a}{b} \geq 1 + \frac{2 + 2\sqrt{1+\alpha^2}}{\alpha^2}. \tag{15}$$

Solving Equation (14 for α we get

$$\alpha = \frac{\sqrt{a/b}}{a/b - 1}.$$

Inserting this expression into (14) and solving for φ_m gives the following relation between the ratio a/b and the phase margin.

$$\varphi_m = \pi + n_{gc}\frac{\pi}{2} - 2\arctan\frac{\sqrt{a/b}}{a/b - 1}. \tag{16}$$

The unstable zero should thus be faster than the unstable pole and that the ratio between them should be sufficiently large. If this is not the case the desired phase margin cannot be obtained.

EXAMPLE 5—RHP POLE ZERO PAIR

Consider a system with a right half plane zero at $s = a$ and a right half plane pole at $s = b$. Let the phase margin be $\varphi_m = \pi/4$ and the slope $n_{gc} = -1/2$, then $a \geq 5.83b$. □

EXAMPLE 6—THE X-29

Considerable design effort has been devoted to the design of the flight control system for the X-29 aircraft. One of the design criteria was that the phase margin should be greater than 45° for all flight conditions. At one flight condition the model has the following non-minimum phase component

$$G_{nmp}(s) = \frac{s - 26}{s - 6}$$

. Since $a = 4.33b$, it follows from Example 5 that a phase margin of 45° cannot be achieved. It is interesting to note that many design methods were used in a futile attempt to reach the design goal. A simple calculation of the type given in this paper would have given much insight. □

EXAMPLE 7—KLEIN'S UNRIDABLE BICYCLE

Some interesting bicycles have been constructed by Professor Klein at University of Illinois, see [6]. One of the bicycles, which has rear wheel steering, has proven impossible to ride. This is easily understood from the theory we have developed.

Bicycle dynamics can be described by the following transfer function which relates tilt angle θ to steering angle δ:

$$\frac{\theta(s)}{\delta(s)} = \frac{mhV_0}{J_p c}\frac{V_0 - bs}{s^2 - mgh/J_p}.$$

The variable V_0 is the forward velocity, m is the mass of the system and J is the moment of inertia, h is the height of the center of mass, x_0 is the coordinate of the center of mass measured from the front wheel contact point, and c is the distance between the contact points of the wheels.

The system has a RHP zero at $s = V_0/x_0$, and a RHP pole at $s = \sqrt{mgh/(J+mh^2)}$. Hence $a/b = V_0/x_0\sqrt{(J+mh^2)/(mgh)}$. The values $m = 70$ kg, $h = 1.2$ m, $x_0 = 0.7$, $J = 20$ kgm^2 and $V_0 = 5$ m/s, give $a/b = 2.7$. It then follows from Equation (16) that with $n_{gc} = -0.5$ the phase margin is less than $\varphi_m = 10.4$. This implies that it is impossible to get a system with reasonable stability margins with any controller.

The ratio a/b can be increased by increasing the height h and decreasing x_0. Klein has constructed a bicycle, with rear wheel steering, where h is large and x_0 is small. This bicycle is is indeed ridable. □

Generalization

The calculations can be extended to systems having time delays and many poles and zeros in the right half plane. For a system with a time delay τ, n real right half plane zeros, a_i, and m real right half plane poles, b_i, the design inequality (6) becomes

$$\sum_{i=1}^{n} \arctan\frac{\omega_{gc}}{a_i} + \sum_{i=1}^{k} \arctan\frac{b_i}{\omega_{gc}} + \frac{1}{2}\omega_{gc}\tau \le \frac{\pi}{2} - \frac{\varphi_m}{2} + n_{gc}\frac{\pi}{4}. \quad (17)$$

4. A Paradox

It is well known that a system that is controllable and observable can be controlled in such a way that the closed loop system has arbitrary closed loop poles. The poles can for example be chosen arbitrarily fast. At a first sight this seems to contradict the result that a process zero close to the origin limits the achievable bandwidth. We will use an example to reconcile the apparent paradox.

EXAMPLE 8—FAST SYSTEM WITH LOW BANDWIDTH
Consider a process with the transfer function

$$G(s) = \frac{\beta s + 1}{s(s+1)}. \quad (18)$$

The zero is unstable if $\beta < 0$. For example if $\beta = -10$ it follows from the design inequality (6) that it is difficult to obtain crossover frequencies higher than 0.1 rad/s.

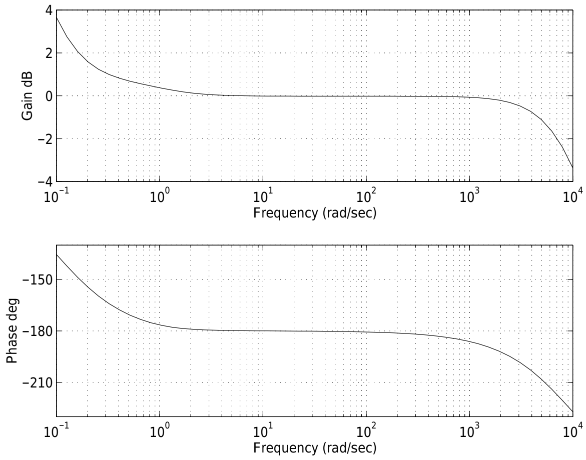

Figure 1. Bode diagram for the loop transfer function for a system with a slow unstable zero, $\beta = -10$. The specifications are, $\omega_0 = 10$ and $\alpha = 1$.

To carry out a pole-placement design assume that it is desirable to have a closed loop system with the characteristic polynomial

$$(s + \alpha\omega_0)(s^2 + \omega_0 s + \omega_0^2).$$

This can be accomplished with a lead-lag compensation with the transfer function

$$C(s) = \frac{s_0 s + s_1}{s + r}.$$

Straight forward calculations give

$$\begin{aligned} s_0 &= \alpha\omega_0^3 + \frac{(1-\alpha\omega_0)(\omega_0^2 - \omega_0 + 1)}{1-\beta}, \\ s_1 &= \alpha\omega_0^3, \\ r &= -\beta s_0 - 1 + \omega_0(1+\alpha). \end{aligned}$$

To obtain a fast closed loop system we choose $\omega_0 = 10$ and $\alpha = 1$. For $\beta = -10$ we get $r = 9274$, $s_0 = 925.5$ and $s_1 = 1000$, and the loop transfer function becomes

$$L(s) = \frac{(\beta s + 1)(s_0 s + s_1)}{s(s+1)(s+r)}.$$

For large ω_0 and β we get

$$L(s) \approx \frac{\alpha\omega_0^3(\beta s + 1)}{s(s - \alpha\beta\omega_0^3)}.$$

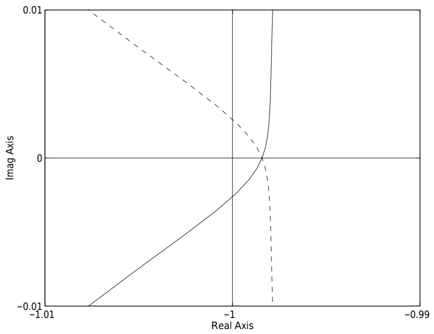

Figure 2. Nyquist curve for the loop transfer function for a system with a slow unstable zero, $\beta = -10$. The specifications are, $\omega_0 = 10$ and $\alpha = 1$.

The process pole at $s = -1$ is thus almost canceled by the zero at $s = -s_1/s_0$. For small s we have

$$L(s) \approx -\frac{\beta s + 1}{\beta s} = -1 - \frac{1}{\beta s}$$

The loop transfer function has a low frequency asymptote that intersects the axis $|L| = 0$ at $s = -1/\beta$, i.e. at the slow unstable zero. The loop gain is close to one at that point and remains close to one until the break point at $s\omega = -r \approx \alpha\beta\omega_0^3$. The phase is close to -180° which means that the stability margin is very poor.

A Bode diagram of the loop transfer function is shown in Figure 1. The slope of the gain curve at crossover is very small which indicates that the system has very small stability margins. This is also clearly seen in the Nyquist curve shown in Figure 2. The bandwidth of the system is small even if the closed loop poles are fast -10, $-10 \pm \sqrt{75}$. The loop gain is practically one from $\omega = 0.1$, which means that the feedback has little effect for frequencies above 0.1 rad/s. □

The example shows that even if a system is designed for specifications that give fast closed loop poles the bandwidth is still small because of the limitations imposed by the slow zero in the right half plane. The results also indicate that the feedback has little effect for frequencies larger than z_i where z_i is the zero in the right half plane.

The importance of being aware of the limitations when designing control systems are clearly illustrated by the example. The design gives a very poor closed loop system. By considering the limitations and selecting a smaller value of ω_0 it is possible to obtain good designs although with much smaller bandwidths.

5. Conclusions

Estimates of limitations on control system performance imposed by measurement noise, actuator saturation, and singularities of the transfer function in the right half plane are presented. The results give useful insight into control design and they are useful complements to design methods. The key limitations for minimum-phase systems arise from the fact that a compensator with high gain is required to obtain a large phase lead. For non-minimum phase systems the limitations are due to system dynamics. Time delays and zeros in the right half plane give upper bounds to the bandwidth of the closed loop system and poles in the right half plane give lower bounds. The results are elementary and can be taught in the first course in feedback control.

6. References

[1] K. J. ÅSTRÖM. "Autonomous controllers." *Control Eng. Practice*, **1:2**, pp. 227–232, 1993.

[2] K. J. ÅSTRÖM. "Fundamental limitations of control system performance." In PAULRAJ *et al.*, Eds., *Mathematical Engineering: A Kailath Festschrift*. Kluwer Academic Publishers, 1996.

[3] H. W. BODE. *Network Analysis and Feedback Amplifier Design*. Van Nostrand, New York, 1945.

[4] S. P. BOYD AND C. H. BARRATT. *Linear Controller Design – Limits of Performance*. Prentice Hall Inc., Englewood Cliffs, New Jersey, 1991.

[5] J. S. FREUDENBERG AND D. P. LOOZE. "Right half plane pole and zeros and design tradeoffs in feedback systems." *IEEE Transactions on Automatic Control*, **AC-30**, pp. 555–565, 1985.

[6] R. E. KLEIN. "Using bicycles to teach system dynamics." *IEEE Control Systems Magazine*, **9:3**, pp. 4–9, 1989.

[7] J. M. MACIEJOWSKI. *Multivariable Feedback Design*. Addison-Wesley, Reading, Massachusetts, 1989.

[8] R. H. MIDDLETON. "Trade-offs in linear control system design." *Automatica*, **27**, pp. 281–292, 1991.

[9] M. M. SERON, J. H. BRASLAVSKY, AND G. C. GOODWIN. *Fundamental Limitations in Filtering and Control*. Springer, Berlin, 1945.

[10] G. STEIN. "Respect the unstable - the 1995 bode lecture." In *30th IEEE Conference on Decision and Control*, Honolulu, Hawai, December 1990.