Lecture 7

Fundamental Limitations*

7.1 Introduction

Maybe the cardinal mistake in control engineering would be to consider the process as fixed once and for all. In fact, the control system specifications could very well be impossible to meet once the process is constructed and fixed. A striking example of this is given in the following citation from F. R. Whitt and D. G. Wilson (MIT Press, 1974), Bicycling Science - Ergonomics and Mechanics:

> *"Many people have seen theoretical advantages in the fact that front-drive, rear-steered recumbent bicycles would have simpler transmissions than rear-driven recumbents and could have the center of mass nearer the front wheel than the rear. The U.S. Department of Transportation commissioned the construction of a safe motorcycle with this configuration. It turned out to be safe in an unexpected way: No one could ride it."*

This lecture is devoted to the fundamental limitations that are inherited from properties of the controlled plant and will address questions like the following two:

- Why are some bicycles impossible to ride?
- How short inverted pendulums can be balanced by hand?

One of the practically most important restrictions to control performance is the fact that actuators have limited capacity and may saturate. However, saturation is a non-linear effect and will not be studied further here. Instead, the focus will be on limitations caused by unstable zeros, unstable poles and time-delays.

An unstable pole p means that the response to a disturbance grows exponentially as e^{pt}. It is intuitively clear that in order to stabilize such a system, the feedback loop must be "faster" than the time constant $1/p$. A formal argument for this will be given in this lecture.

In case the unstable system also includes a time-delay, the control problem could become "impossible". A time-delay T means that control action at time t does not have any effect until time $t + T$. Hence, it is intuitively clear that an unstable pole can not be stabilized unless T is small compared to $1/p$. The argument applies to the question about pendulum balancing above, since the human feedback system through the eyes always involves a time delay.

A more intricate performance limitation is imposed by unstable zeros. It is well known that an unstable zero results in a step response that initially goes in the "wrong" direction. In fact,

*Written by A. Rantzer with contributions by K.J. Åström

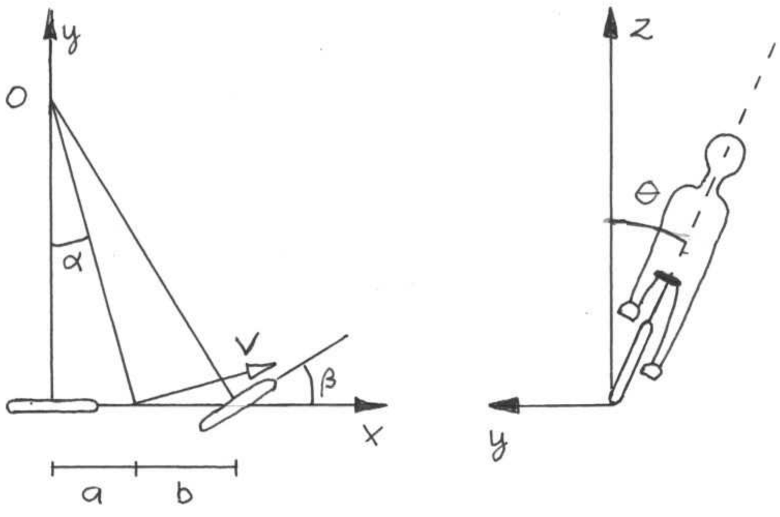

Figure 7.1 Schematic picture of a bicycle. The top view is shown on the left and the rear view on the right.

this can be seen directly in the expression for the Laplace transform $Y(s)$ of the step response $y(t)$, where the zero at z means that

$$0 = Y(z) = \int_0^\infty y(t)e^{-zt}dt$$

Clearly the integral cannot be zero unless $y(t)$ takes both positive and negative values. The time duration of such dynamics is approximately $1/z$ and limits the achievable rate of control. Hence, while an unstable pole requires a fast feedback loop, an unstable zero gives an upper bound on how fast it can be. A combination of the two phenomena can make the system impossible to control.

EXAMPLE 7.1
A tourque balance for a bicycle can be written as

$$J\frac{d^2\theta}{dt^2} = mg\ell\theta + \frac{mV_0\ell}{b}\left(V_0\beta + a\frac{d\beta}{dt}\right)$$

where the physical parameters have typical values as follows:

Mass:	$m = 70$ kg
Distance rear-to-center:	$a = 0.3$m
Height over ground:	$\ell = 1.2$ m
Distance center-to-front:	$b = 0.7$ m
Moment of inertia:	$J = 120$ kgm^2
Speed:	$V_0 = 5$ ms^{-1}
Acceleration of gravity:	$g = 9.81$ ms^{-2}

The transfer function from β to θ is

$$P(s) = \frac{mV_0\ell}{b}\frac{as + V_0}{Js^2 - mg\ell}$$

The system has an unstable pole p with time-constant

$$p^{-1} = \sqrt{\frac{J}{mg\ell}} \approx 0.4 \text{ s}$$

The closed loop system must be at least as fast as this. Moreover, the transfer function has a zero z with

$$z^{-1} = -\frac{a}{V_0} \approx 0.06\text{s}$$

Riding the bicycle at this speed, the zero is not really an obstacle for control. However, with a rear-wheel-steered bicycle, the speed gets a negative sign and the zero becomes unstable. In particular, for slow speed (≈ 0.7m/s) there is an unstable pole-zero cancellation, which is impossible to stabilize. □

7.2 The Maximum Modulus Theorem

More formal arguments about fundamental limitations can be obtained using theory for analytic functions. It is natural that analytic functions appear, since we have a seen that a controller is stabilizing if and only if the closed loop transfer function is analytic in the right half plane (all poles in the left half plane). The main mathematical theorem to be used is the following:

THEOREM 7.1—THE MAXIMUM MODULUS THEOREM
Suppose that the function f is analytic in a set containing the unit disc. Then

$$\max_{|z|\leq 1} |f(z)| = \max_{|z|=1} |f(z)|$$

□

In Laplace transform applications, the stability boundary will be the imaginary axis. It is therefore convenient to note that for every stable rational transfer function $G(s)$, analytic in the right half plane, the function

$$f(z) = G\left(\frac{1+z}{1-z}\right)$$

is analytic in the unit disc. Hence the Maximum Modulus Theorem can be applied to give the following corollary:

COROLLARY 7.1
Suppose that all poles of the rational function $G(s)$ have negative real part. Then

$$\max_{\text{Re } s\geq 0} |G(s)| = \max_{\omega\in\mathbf{R}} |G(i\omega)|$$

□

7.3 Sensitivity bounds from unstable zeros and poles

It is easy to see that the sensitivity function must be equal to one at an unstable zero of the transfer function:

$$P(z) = 0 \qquad \Rightarrow \qquad S(z) := \frac{1}{1 + C(z)P(z)} = 1$$

Notice that the unstable zero in the plant can not be cancelled by an unstable pole in the controller, since this would give an unstable transfer function $C/(1+CP)$ from measurement noise to control input.

Similarly, the complimentary sensitivity must be one at an unstable pole:

$$P(p) = \infty \qquad \Rightarrow \qquad T(p) := \frac{C(p)P(p)}{1 + C(p)P(p)} = 1$$

In this case, cancellation by an unstable zero in the controller would give an unstable transfer function $P/(1+CP)$ from input disturbance to plant output.

Combining the first constraint $S(z) = 1$ with Corollary 7.1 immediately gives a lower bound on the norm of the sensitivity function:

$$\max_{\omega \in \mathbf{R}} |S(i\omega)| = \max_{\text{Re } s \geq 0} |S(s)| \geq |S(z)| = 1 \qquad \Rightarrow \qquad \|S\|_\infty \geq 1$$

This bound is however not particularly interesting, since usually $S(i\omega) \approx 1$ for high frequencies anyway. A much more interesting conclusion will next be obtained by using a weighting function.

Recall that disturbance rejection requires small sensitivity for small frequencies. One way to formalize this condition is to define

$$W_a(s) = \frac{s + a}{2s}$$

and require that

$$\sup_\omega |W_a(i\omega)S(i\omega)| \leq 1 \tag{7.1}$$

for some value of a. See Figure 7.2, left. Satisfying (7.1) with a high value of a means fast disturbance rejection.

The specification requires that $S(s)$ has a zero in the origin. This is often obtained by an integrator in the controller. Moreover, Corollary 7.1 implies that

$$\sup_\omega |W_a(i\omega)S(i\omega)| = \sup_{\text{Re } s \geq 0} |W_a(s)S(s)| \geq |W_a(z_i)|$$

for every unstable zero z_i of the plant P. In particular, the specification (7.1) *is impossible to satisfy* unless $|W_a(z_i)| \leq 1$, or in other words $a \leq z_i$, for every unstable zero z_i. Hence the unstable zeros give an upper bound on the achievable bandwidth. In the following theorem, this discussion is summarized together with a corresponding argument for unstable poles:

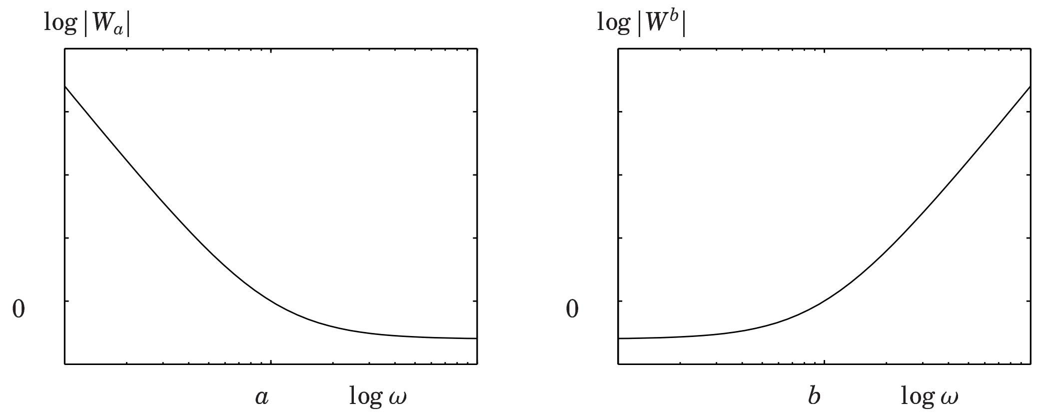

Figure 7.2 Amplitude plots for weighting functions. The left weighting function is used to bound the sensitivity at small frequencies, while the right function is used to boud the complementary sensitivity at high frequencies

THEOREM 7.2
Suppose that the plant $P(s)$ has unstable zeros z_i and unstable poles p_j. Then the specifications

$$\sup_\omega |W_a(i\omega)S(i\omega)| \leq 1 \qquad \sup_\omega \left|W^b(i\omega)T(i\omega)\right| \leq 1$$

are impossible to meet with a stabilizing controller unless $\|W_a(z_i)\| \leq 1$ for every unstable zero z_i and $\|W^b(p_j)\| \leq 1$ for every unstable pole p_j.

In particular, if $W_a = (s+a)/(2s)$ and $W^b(s) = (s+b)/(2b)$, it is necessary that $a \leq \min_i z_i$ and $b \geq \max_j p_j$. □

Proof. The statement about the sensitivity function was proved above, and the statment about the complementary sensitivity function is analogous. □

EXAMPLE 7.2
Let us see what Theorem 7.2 has to say about the bicycle example. The unstable pole gives a bound $\|W^b T\| \leq 1$ for $b \geq \sqrt{mg\ell/J}$. This shows that the closed loop transfer function from measurement noise to process output can not be forced small for frequencies below $\sqrt{mg\ell/J}$. A loose interpretation is that it is impossible to ride the bicycle and keep the eyes shut except for a sample every second. This applies for the bicycle with normal steering regardless of speed.

For a rear wheel steering bike, there is the second complication of an unstable zero at V_0/a, which gives a bound on how fast disturbances one can reject. For low speed, only slow disturbances can be rejected.

The special difficulties corresponding to a combination of an unstable pole and an unstable zero nearby are however not apparent in Theorem 7.2. Such problems will be treated next. □

The following theorem gives simple expressions for the limitations caused by an unstable pole/zero pair.

THEOREM 7.3
If $P(s)$ has an unstable pole p and an unstable zero z, then

$$\left\|\frac{1}{1+CP}\right\|_\infty \geq \left|\frac{z+p}{z-p}\right|$$

for every stabilizing $C(s)$. □

Note that if S is very large, then the same is true for T, since $S + T \equiv 1$. Hence, if $|(z+p)/(z-p)|$ is significantly larger than one, the system is impossible to control because of poor robustness to model errors and amplification of measurement noise.

Proof. Assume that $P(s) = (s-z)(s-p)^{-1}\widehat{P}(s)$, with $\widehat{P}$ proper and $\widehat{P}(p) \neq 0$. Then the sensitivity function satisfies

$$
\begin{aligned}
\|S\|_\infty = \sup_\omega \left| \frac{1}{1+CP} \right| &= \sup_\omega \left| \frac{1}{1+C\widehat{P}(i\omega-z)(i\omega-p)^{-1}} \right| \\
= \sup_\omega \left| \frac{i\omega-p}{i\omega-p+C\widehat{P}(i\omega-z)} \right| &= \sup_\omega \left| \frac{i\omega+p}{i\omega-p+C\widehat{P}(i\omega-z)} \right| \\
= \sup_{\operatorname{Re} s\geq 0} \left| \frac{s+p}{s-p+C\widehat{P}(s-z)} \right| &\geq \left| \frac{z+p}{z-p} \right|
\end{aligned}
$$

The fourth inequality uses that $|i\omega - p| = \sqrt{\omega^2+p^2} = |i\omega+p|$ and the fifth inequality is Corollary 7.1. □

A similar argument can be applied to a system involving a time delay but application of the maximum modulus theorem is less straightforward in this case.

7.4 Bode's integral formula

Another striking performance limitation, known as Bode's integral formula, shows that the effort to make the sensitivity function small is always a trade-off between different frequency regions:

THEOREM 7.4
If $P(s)$, $C(s)$ and $S(s) = [1+C(s)P(s)]^{-1}$ are stable and $s^2C(s)P(s)$ is bounded, then

$$\int_0^\infty \log|S(i\omega)|\,d\omega = 0$$

□

Proof. Proof sketch. From the theory of analytic functions, recall that Cauchy's formula states that

$$\int_\gamma f(z)dz = 0$$

for every closed path γ in the region where the function f is analytic. Bode's integral formula follows by application of Cauchy's formula to

$$f(z) = \log S(z)$$

The stability of C and P guarantee that f is well-defined and analytic in the whole right half plane. Integration along the imaginary axis can be extended to integration along a closed path by adding a large half-circle in the right half plane. The condition that $s^2C(s)P(s)$ is bounded is needed to make sure that the contribution from the half-circle vanishes as the radius tends to infinity. □

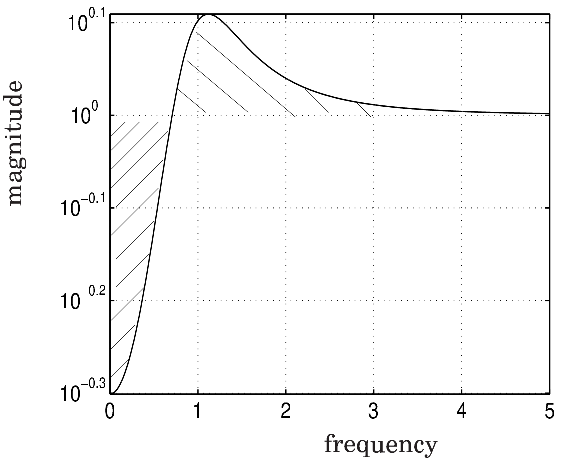

Figure 7.3 The amplitude curve of the sensitivity function always enclose the same area below the level $|S| = 1$ as above.

The invariance of Bode's integral is sometimes referred to as the "water-bed" effect: If the designer tries to push the magnitude of the sensitivity function down at some point, it will inevitably pop up somewhere else!

The assumptions behind Bode's integral formula deserve some discussion. The expression $s^2C(s)P(s)$ is always bounded whenever $C(s)$ and $P(s)$ correspond to real sensor/actuator interconnections, since direct terms are not physically implementable. With unstable poles in $C(s)P(s)$ the integral formula changes into

$$\int_0^\infty \log|S(i\omega)|\,d\omega = \pi \sum_i \mathrm{Re}\, p_i$$

which makes it even harder to push down the sensitivty magnitude! The faster unstable modes, the harder it is. In fact, this can be used as an argument why unstable controllers should in general be avoided.