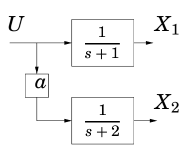
Lecture 7: Fundamental Limitations

- System limitations – examples
- Motivation from loop shaping
- The concepts of minimum and non-minimum phase
  - How magnitude and phase are coupled
- Intuitive arguments for limitations
- the Bode integral theorem
- Bicycle example
- The Maximum Modulus Theorem
- Limitations imposed by
  - unstable poles
  - nonminimum-phase zeros (zeros in RHP)

Based on material from K.J Åström and A. Rantzer
See lecture notes and [G&L Ch. 7]

Limitations: Controllability [from lec 6]



System $\dot{x} = \begin{bmatrix} -1 & 0 \\ 0 & -2 \end{bmatrix} x + \begin{bmatrix} 1 \\ a \end{bmatrix} u$

State x_2 is *uncontrollable* for $a = 0$ and "hard to control" for small values of a.

Controllability matrix

$W_c = \begin{bmatrix} B & AB \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ a & -2a \end{bmatrix}$

Controllability gramian S

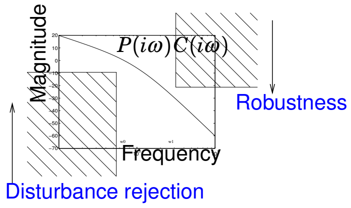
$AS + SA^T + BB^T = 0 \Longrightarrow$

$S = ... = \begin{bmatrix} \frac{1}{2} & \frac{1}{3}a \\ \frac{1}{3}a & \frac{1}{4}a^2 \end{bmatrix}$

Plot of $\begin{bmatrix} x_1 & x_2 \end{bmatrix} \cdot S^{-1} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 1$ corresponds to what states we can reach by $\int_0^{t_1} |u(t)|^2 dt = 1$.

Frequency specs

We typically have low frequency specs for disturbance rejection and high frequency specs for measurement noise rejection and robustness



We will have the cross-over frequency w_c with stability margins A_m and ϕ_m in the range between “low” and “high fq”.

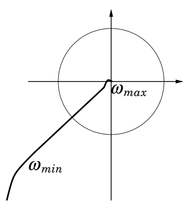
Loop Shaping Methods

- Shape the loop frequency range by frequency range
- Lag compensation (performance, disturbance attenuation, AK)
- Lead compensation (robustness, phase margin, AK)
- High frequency roll-off (robustness, unmodeled high frequency, AK?)
- Notch filters (reduce gain at certain frequencies)
- Slope n_{gc} at crossover frequency ω_{gc} and phase margin $\underbrace{\varphi_m \approx 180° + n_{gc} 90°}_{approximation}$ (minimum phase systems)

Bodes Ideal Loop Transfer Function

The repeater problem. Large gain variations in vacuum tube amplifiers. What should a transfer function look like to be independent of gain?

$$L(s) = \left(\frac{s}{\omega_{gc}} \right)^n$$

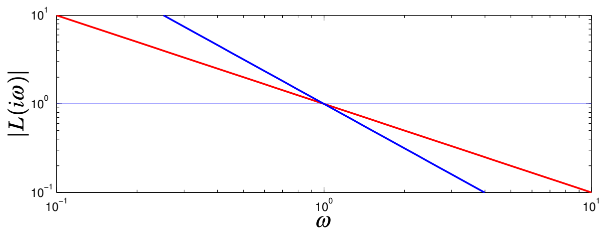


The approximate version of Bodes relations is exact for $L(s)$. Phase margin invariant with loop gain.

The slope $n = -1.5$ gives the phase margin $\varphi_m = 45°$.

Q: How do we get slope -1.5?

Trade-offs



- Blue curve slope $n = -5/3$ phase margin $\varphi_m = 30°$
- Red curve slope $n = -1$ phase margin $\varphi_m = 90°$
- Making the curve steeper reduces the frequency range but also the phase margin

Magnitude and phase relations

Q: How “fast / steep” can we have the magnitude of the loop gain PC to go between these areas and still have positive phase margin ϕ_m?

Bode’s relation (See Th 7.1 and 7.2 in [G&L]) describes the relation between the *magnitude curve* and the *phase curve*.

Remember the basic rules for sketching Bode diagrams how poles and zeros relate to the slope of the magnitude curve ('breakpoints and asymptotes') and how the corresponding phase curve will look like.

Non-minimum Phase Systems

Dynamics poses a severe limitation on achievable performance for systems with

- Right half plane poles
- Right half plane zeros
- Time delays

Bode introduced the concept *non-minimum phase* to capture this. A system is *minimum phase* system if all its poles and zeros are in the left half plane.

How should dynamics limitations be captured quantitatively?

Bode's Relations

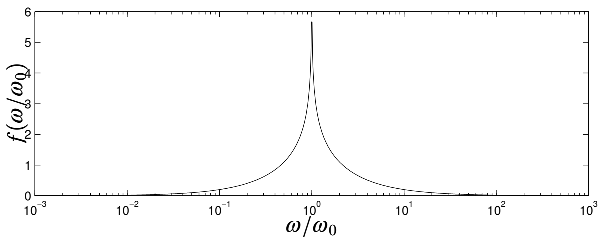
For minimum phase systems the phase curve is given by the gain curve and vice versa. The exact relations are given by Bodes relation(s). (only in one direction here)

$$\begin{aligned}\arg G(i\omega_0) &= \frac{2\omega_0}{\pi}\int_0^\infty \frac{\log|G(i\omega)| - \log|G(i\omega_0)|}{\omega^2-\omega_0^2}d\omega \\ &= \frac{1}{\pi}\int_0^\infty \frac{d\log|G(i\omega)|}{d\log\omega}\log\left|\frac{\omega+\omega_0}{\omega-\omega_0}\right| d\log\omega \\ &\approx \underbrace{\frac{\pi}{2}\frac{d\log|G(i\omega)|}{d\log\omega}_{|\omega=\omega_0}}_{'slope\ n'} = \frac{\pi}{2}n\end{aligned}$$

Proven by contour integration

The Weighting Function

$$f\left(\frac{\omega}{\omega_0}\right) = \frac{2}{\pi^2}\log\frac{|\omega+\omega_0|}{|\omega-\omega_0|}$$



Almost like impulse – cuts out at $\omega = \omega_0$

Non-minimum Phase Systems

Factor process transfer function as $P(s) = P_{mp}(s)P_{nmp}(s)$ such that $|P_{nmp}(i\omega)| = 1$ and P_{nmp} has negative phase. Requiring a phase margin φ_m we get

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

Approximate $\arg\left(P_{mp}(i\omega_{gc})C(i\omega_{gc})\right) \approx n\pi/2$ gives

$$\arg P_{nmp}(i\omega_{gc}) \geq -\pi + \varphi_m - n\frac{\pi}{2}$$

This equation called, the phase crossover inequality. Equality holds exactly if $P_{nmp}C$ is Bode's ideal loop transfer function, the expression is an approximation for other designs if n is the slope at the crossover frequency.

The Crossover Frequency Inequality

The inequality

$$\arg P_{nmp}(i\omega_{gc}) \geq -\pi + \varphi_m - n_{gc}\frac{\pi}{2}$$

says that the phase lag of the non-minimum phase component must not be too large at the crossover frequency!

Simple rules of thumb:

- $\varphi_m = 45°$ and $n_{gc} = -1$ gives

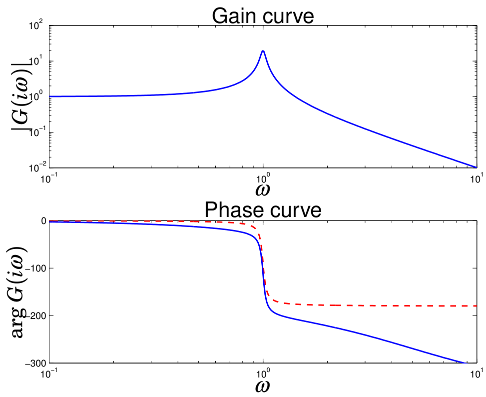
$$\arg P_{nmp}(i\omega_{gc}) \geq -\frac{\pi}{4} = -0.8\ [rad], \quad 45°$$

- $\varphi_m = 45°$ and $n_{gc} = -0.5$ gives

$$\arg P_{nmp}(i\omega_{gc}) \geq -\frac{\pi}{2} = -1.6\ [rad], \quad 90°$$

Non-minimum phase components may only have a phase-lag of at most $45° - 90°$ at the gain cross over frequency!

Bode Plots Should Look Like This



Unstable poles - “intuitive reasoning”

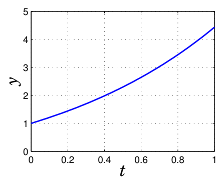
An unstable pole p makes the output signal for a bounded input grow exponentially as $\sim e^{pt}$. To stabilize this system, one require measurements witha frequency content up to $\sim 1/p$.

One can only tolerate time-delays $T \ll \frac{1}{p}$

System with RHP Pole – bandwidth constraints

NMP part of transfer function

$$P_{nmp}(s) = \frac{s+p}{s-p}$$



Notice normalization $P(0) < 0$!
Cross over frequency inequality

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

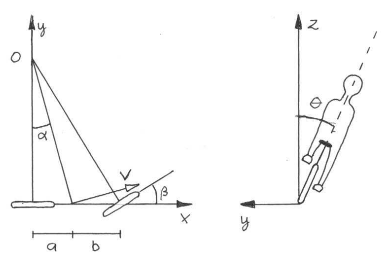
Hence

$$\omega_{gc} \geq \frac{p}{\tan\left(\frac{\pi}{2} - \frac{\varphi_m}{2} + n_{gc}\frac{\pi}{4}\right)}$$

The simple rule of thumb ($\varphi_{lagnmp} = \pi/4$) gives $\omega_{gc} \geq 2.4p$

Bike example

A (linearized) torque balance for a bicycle can be approximated as



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

Bike example, cont'd

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

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

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

Bike example, cont'd

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

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

For the back-wheel steered bike we have the same poles but different sign of V_0 and the zero will thus the be in the RHP!

Systems with time-delay – "intuitive"

Assume that the plant contains a time-delay T. This means e.g. that a load disturbance is not visible in the output signals until after at least T time unit. This corresponds roughly to a frequency of $1/T$ and it is thus unrealistic to try to achieve a higher bandwidth or cross-over frequency.

How do we see this inherent constraint in the analysis?

System with Time Delay

NMP part of transfer function

$$P_{nmp}(s) = e^{-sT}$$

Cross over frequency inequality

$$\omega_{gc}T \le \pi - \varphi_m + n_{gc}\frac{\pi}{2}$$

The simple rule of ($\varphi_{lagnmp} = \pi/4$) gives

$$\omega_{gc}T \le \frac{\pi}{4} = 0.8$$

Notice $e^{-sT} \approx \frac{1-sT/2}{1+sT/2}$, zero at $s = 2/T$, rule for RHP zero ($\omega_{gc} < 0.4z$) gives same result.

Nonminimum-phase Zeros - "intuitive"

The step response of a system with a process *zero in the right half plane* (i.e, with positive real part) goes initially in the "wrong direction".

The time constant for this dynamics is $\sim 1/z$ and puts an upper limit for how fast control can be made.

The Laplace transform of the system output signal $G(s)U(s)$ will be 0 if we evaluate it at $s = z$ where z is a process zero. If we in particular look at the step response, call it $y(t)$, and its Laplace transform we get

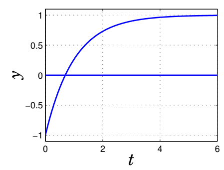
$$0 = Y(s)_{s=z} = Y(z) = \int_0^\infty y(t)\underbrace{e^{-zt}}_{>0}dt$$

To satisfy the equation we must have that $y(t)$ takes both positive and negative values!

System with RHP Zero – bandwidth

NMP part of transfer function

$$P_{nmp}(s) = \frac{z-s}{z+s}$$



Notice normalization $P(0) > 0$!
Cross over frequency inequality

$$\arg P_{nmp}(i\omega_{gc}) = -2\arctan\frac{\omega_{gc}}{z} \ge -\pi + \varphi_m - n_{gc}\frac{\pi}{2}$$

Hence

$$\frac{\omega_{gc}}{z} \le \tan\left(\frac{\pi}{2} - \frac{\varphi_m}{2} + n_{gc}\frac{\pi}{4}\right)$$

The simple rule of thumb ($\varphi_{lagnmp} = \pi/4$) gives

$$\omega_{gc} < 0.4z$$

In-phase and Reverse-phase Systems

Example

$$P(s) = \frac{1}{s^2} - \frac{k}{s^2+1} = \frac{(1-k)s^2+1}{s^2(s^2+1)}$$

If $k > 1$ the system has a RHP zero at

$$z = \frac{1}{\sqrt{k-1}}$$

and the simple rule ($\varphi_{lagnmp} = \pi/4$) limits the gain crossover frequency to

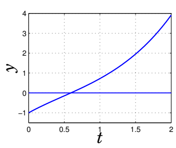
$$\omega_{gc} < z = 0.4\frac{1}{\sqrt{k-1}}$$

k	> 1	1.01	1.1	1.25	2	10	100
ω_{gc}	∞	100	3.2	2	1	0.33	0.10

System with RHP Pole and Zero Pair

NMP part of transfer function

$$P_{nmp}(s) = \frac{(z-s)(s+p)}{(z+s)(s-p)}$$



Notice normalization!
For $z > p$ the cross over frequency inequality becomes

$$\varphi_m < \pi + n_{gc}\frac{\pi}{2} - 2\arctan\frac{2\sqrt{p/z}}{1-p/z}$$

With $n_{gc} = -0.5$ we get

z/p	2	2.24	4.11	5	5.83	8.68	10	20
φ_m	-6.0	0	30	38.6	45	60	64.8	84.6

The sensitivity function

$$S = \frac{1}{1+PC}$$

and the complementary sensitivity function

$$T = \frac{PC}{1+PC}$$

satisfy

$$S + T = 1$$

Q: Can we make $|S(i\omega)| \leq 1$ for all ω?

Bode's Integral theorem ("The water bed effect")

Bode's Integral theorem

See [G&L Theorem 7.3] for details/asumptions.

For a system with loop gain $L = PC$ which has a relative degree ≥ 2 the following *conservation law* for the sensitivity function $S = \dfrac{1}{1+L}$ holds.

$$\int_0^{+\infty} log|S(i\omega)|d\omega = 0 + \pi \sum_{i=1}^{M} Re(p_i)$$

where $\{p_i\}_1^M$ are the M unstable poles in the loop gain L.

Note: We want to keep the sensitivity function low. One can see that if there are unstable poles in L these increase the average level of S.

(G. Stein: –"Conservation of "dirt!"")

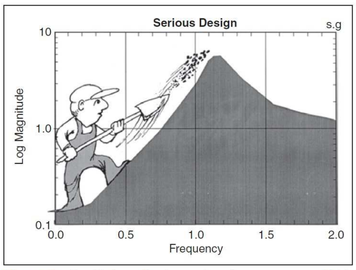


Figure 3. *Sensitivity reduction at low frequency unavoidably leads to sensitivity increase at higher frequencies.*

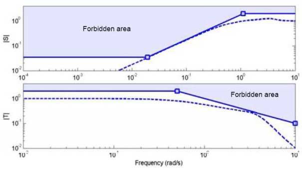
Picture from Gunter Steins Bode Lecture (1985) "Respect the unstable". Reprint in [IEEE Control Systems Magazine (Aug 2003)]

Constraints on S and T

In the lecture on loop shaping [lecture 4] we used specifications of the form

$$|S(i\omega)| \leq |W_S^{-1}(i\omega)|$$

$$|T(i\omega)| \leq |W_T^{-1}(i\omega)|$$



We will see that if there are zeros in the right half plane (so called *nonminimum phase zeros*) and/or unstable poles in the loop gain $L = PC$, this will put constraints on what specifications W_S and W_T we can satisfy.

To prove some of the following constraints on S and T we will use "The Maximum Modulus Theorem" which is introduced on the next slides.

The Maximum Modulus Theorem

Theorem (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 (see next slide):

The Maximum Modulus Theorem, cont'd

Corollary

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

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

OK, let's continue...

Sensitivity bounds from nonmin phase zeros

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

$$P(s = z_u) = 0 \qquad \Rightarrow \qquad S(z_u) := \frac{1}{1 + \underbrace{P(z_u)}_{0} C(z_u)} = 1$$

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 $C/(1 + PC)$ from measurement noise to control input.

Sensitivity bounds from unstable poles

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

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

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

Spec. for disturbance rejection

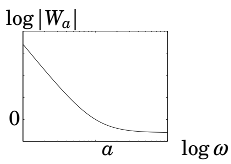
Recall that disturbance rejection requires small sensitivity for small frequencies. One way to formalize this condition is to define a weighting function

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

and require that

$$\sup_\omega |W_a(i\omega)S(i\omega)| \le 1 \qquad (1)$$

for some value of a (see figure below). Satisfying (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 1 implies that

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

for every unstable zero z_i of the plant P.

In particular, the specification (1) *is impossible to satisfy* unless $|W_a(z_i)| \le 1$, or in other words $a \le z_i$, for every RHP zero z_i.

Hence the RHP zeros give an upper bound on the achievable bandwidth.

Spec. for noise rejection/robustness

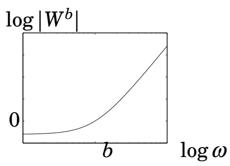
Suppose that the plant $P(s)$ has unstable poles p_j. Define the weighting function $W^b(s) = (s + b)/(2b)$. Then the specification

$$\sup_\omega \left| W^b(i\omega)T(i\omega) \right| \le 1$$

is impossible to meet with a stabilizing controller unless

$$b \ge \max_j p_j$$

Hence unstable poles give a lower bound on the needed bandwidth.



Apply what you have learnt to save time and effort!

Example - The X-29

Advanced experimental aircraft. Much design effort was done with many methods and much cost. Specifications $\varphi_m = 45°$ could not be reached. Here is why!

Non-minimum phase part of the transfer function

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

The zero pole ratio is $z/p = 4.33$ with $n_{gc} = -1/2$ we get $\varphi_m = 32°$

Not possible to get a phase margin of $45°$!

See more in [G. Stein: Respectct the unstable]

Klein's Bicycle with Rear Wheel Steering

Richard Klein at UIUC has built several UnRidable Bicycles (URBs). We have versions in Lund

Transfer function

$$P(s) = \frac{am\ell V_0}{bJ} \frac{-s + \dfrac{V_0}{a}}{s^2 - \dfrac{mg\ell}{J}}$$

Pole at $p = \sqrt{\dfrac{mg\ell}{J}} \approx 3 \;\; \text{rad/s}$

RHP zero at $z = \dfrac{V_0}{a}$

Pole independent of velocity but zero proportional to velocity. There is a velocity such that $z = p$ and the system is uncontrollable. The system is difficult to control robustly if z/p is in the range of 0.25 to 4.

UCSB Version

Bonus material: Time Delay and RHP Pole

NMP part of process transfer function

$$P_{nmp}(s) = \frac{s+p}{s-p} e^{-sT}.$$

$$\arg P_{nmp}(i\omega_{gc}) = -2\arctan \frac{p}{\omega_{gc}} - \omega_{gc}T > -\pi + \varphi_m - n_{gc}\frac{\pi}{2}$$

Hence

$$2\arctan \frac{pT}{\sqrt{2pT-(pT)^2}} + pT\sqrt{2pT-(pT)^2} < \pi - \varphi_m + n_{gc}\frac{\pi}{2}$$

Stability condition $pT < 2$, Why? The simple rules:

$n_{gc} = -0.5$, $\varphi_m = \pi/4$ gives $pT < 0.33$.
$n_{gc} = -1$, $\varphi_m = \pi/4$ gives $pT < 0.07$.

Stabilizing an Inverted Pendulum with Delay

Right half plane pole at

$$p = \sqrt{\frac{g}{\ell}}$$

With a neural lag of 0.07 s and $pT < 0.33$ we find $\ell > 0.45$.

A vision based system with sampling rate of 50 Hz gives a time delay of 0.02 s, this gives $\ell > 0.036$.

Summary of Limitations - Part 1

Developed by assuming Bode's ideal loop transfer function

- A RHP zero z gives an upper bound to bandwidth

$$\frac{\omega_{gc}}{z} \leq \begin{cases} 0.5 & \text{for } M_s, M_t < 2 \\ 0.2 & \text{for } M_s, M_t < 1.4. \end{cases}$$

- A time delay T gives an upper bound to bandwidth

$$\omega_{gc}T \leq \begin{cases} 0.7 & \text{for } M_s, M_t < 2 \\ 0.37 & \text{for } M_s, M_t < 1.4. \end{cases}$$

- A RHP pole p gives a lower bound to bandwidth

$$\frac{\omega_{gc}}{p} \geq \begin{cases} 2 & \text{for } M_s, M_t < 2 \\ 5 & \text{for } M_s, M_t < 1.4. \end{cases}$$

Summary of Limitations - Part 2

- RHP poles and zeros must be sufficiently separated

$$\frac{z}{p} \geq \begin{cases} 6.5 & \text{for } M_s, M_t < 2 \\ 14.4 & \text{for } M_s, M_t < 1.4. \end{cases}$$

- RHP poles and zeros must be sufficiently separated

$$\frac{p}{z} \geq \begin{cases} 6.5 & \text{for } M_s, M_t < 2 \\ 14.4 & \text{for } M_s, M_t < 1.4 \end{cases}$$

- The product of a RHP pole and a time delay cannot be too large

$$pT \leq \begin{cases} 0.16 & \text{for } M_s, M_t < 2 \\ 0.05 & \text{for } M_s, M_t < 1.4. \end{cases}$$

To Read More

- Skogestad Postleithwaite Multivariable Feedback Control - Analysis and Design Wiley 2005
- Åström, Limitations on Control System Performance. European Journal of Control, (6:1) 2000 2–20. (ast00a)
- Gunter Stein, "Respect the unstable", reprint of GS's Bode lecture 1989 in IEEE Control Systems Magazine (aug, 2003).
- "Åström, Klein and Lennartsson, "Bicycle Dynamics and Control", IEEE Control Systems Magazine, vol 25(4), pp 26–47, 2005.