WHY FREQUENCY RESPONSE?

BY M.T. Tham
Dept. of Chemical and Process Engineering
University of Newcastle upon Tyne
(APRIL 1999)

INTRODUCTION

Thus far, all analysis of system dynamics and the performance of process control systems have been carried out in either the time or Laplace domains. Laplace Transforms are used to expedite analysis of dynamical systems. Familiarity with Laplace Transform techniques means that we do not need to solve the underlying ordinary differential equations. The use of Laplace Transforms has been very useful and in many cases, analysis within the Laplace domain is more than satisfactory. However, there are other cases where this approach is inflexible and leads to intractable problems. In fact, frequency response techniques can be very useful in a whole spectrum of process control activities, e.g.

- system identification
- controller tuning
- stability analysis
- robustness analysis
- design of noise filters

SYSTEM IDENTIFICATION

After studying the frequency responses of various systems, it should be clear that the characteristics of the responses are determined by parameters of the systems. Not only do the shapes and features of frequency response plots enable quick categorisation of system types and structures, they also enable more accurate determination of system parameters. Take the Bode diagram shown in Fig. 1 as an example. From the AR plot, we observe that there are 3 asymptotes, one LFA and two HFA. The largest gradient of the high frequency asymptote (HFA2) is -40dB/decade and these combine to indicate that the system is a second order system. The gain of the system is K_p and because there is no overshoot in the AR plot, it is most likely that the system is an overdamped one. Since it is an overdamped second order system, it should have two time constants, and these are given by the inverse of the frequencies at which the asymptotes intersect, i.e. τ_1 and τ_2. Now, if the system is purely second order, then the maximum phase shift should be -180^o. However, the phase plot shows that the system's phase shift tends towards infinity, which is an indication that the system possesses a time delay. To determine the value of this time delay, all we need to do is to calculate the phase shift of the system determined from the AR plot, namely $\frac{K_p}{(1+\tau_1)(1+\tau_2)}$, at a particular frequency, and compare that with the value obtained from the phase-shift plot. The difference must be due to the time delay, which can then be back calculated.

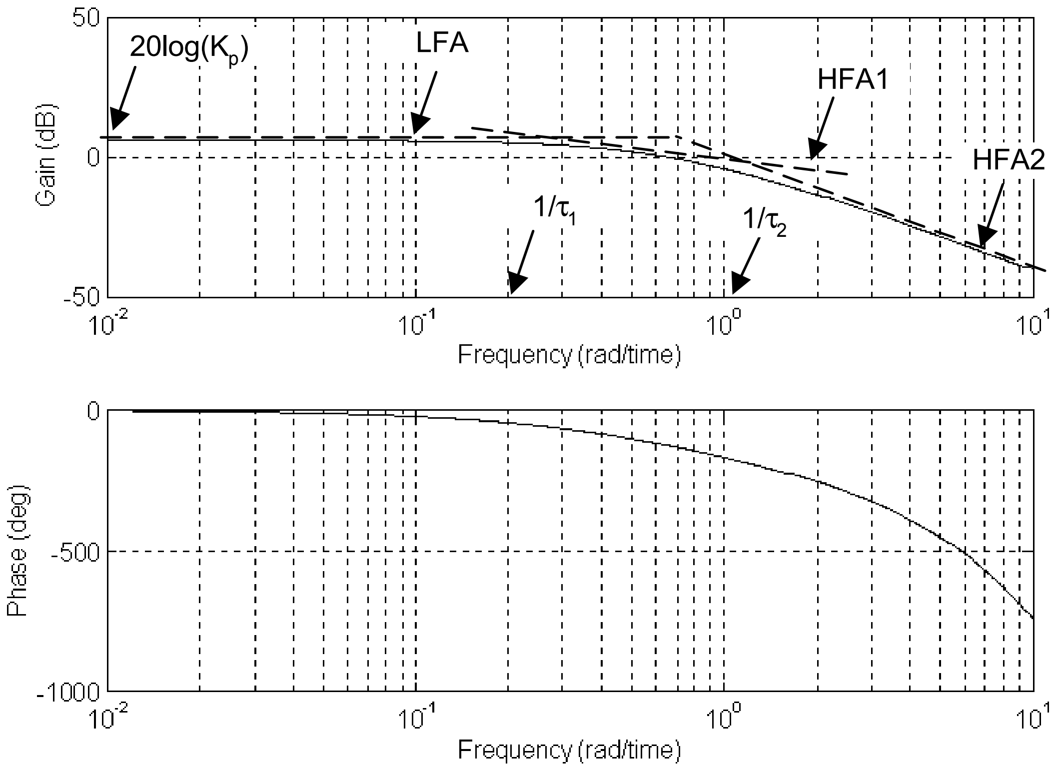

Figure 1. Using Bode diagrams in system identification

STABILITY ANALYSIS

Laplace Domain

One of the most important tasks that is facilitated by working in the frequency domain is closed loop stability analysis. With Laplace transfer function based methods, we need to obtain the closed loop expression, extract the characteristic polynomial, and examine the nature of the roots of the characteristic equation. One of the problems associated with this approach is the presence of a time delay. For example, consider the proportional control of a first-order plus time delay system. The closed loop transfer function is:

$$\frac{Y(s)}{R(s)} = \frac{K_c.\dfrac{K_p\exp(-\theta s)}{1+\tau s}}{1+K_c.\dfrac{K_p\exp(-\theta s)}{1+\tau s}}$$

$Y(s)$ is the output; $R(s)$ is the set-point; K_c is the proportional gain; K_p is process gain; θ the time delay, and τ the time constant. Simplification gives

$$\frac{Y(s)}{R(s)} = \frac{K_c K_p \exp(-\theta s)}{1 + \tau s + K_c K_p \exp(-\theta s)}$$

The closed loop characteristic equation is therefore:

$$1 + \tau s + K_c K_p \exp(-\theta s) = 0$$

and the roots of this equation determines whether the closed loop is stable. The presence of the delay term therefore poses a difficulty, as it has to be approximated by a polynomial, to enable solution for the roots. Even discounting the accuracy of approximation, all that the roots tell us is whether the closed loop is stable or otherwise (absolute stability), not how near the system is to becoming unstable (relative stability).

Frequency Domain

By working in the frequency domain, the problem imposed by time delays in the transfer function approach is avoided. Stability analysis in the frequency domain also tells us how close the system is to instability. Finally, the frequency domain allows the stability of the closed loop system to be evaluated from the frequency response of the open loop system. Let us first examine this last aspect.

Let the Laplace transfer functions of the controller and the process be $G_c(s)$ and $G_p(s)$ respectively. The closed loop characteristic equation is therefore:

$$1 + G_c(s)G_p(s) = 0$$

Any root that satisfy this equation also has to satisfy

$$G_c(s)G_p(s) = -1$$

In particular, for strict stability, the closed loop poles, p_i must be such that

$$G_c(s)G_p(s)|_{s=p_i} < -1$$

Transforming this alternative form of the characteristic equation into the frequency domain, we get

$$G_c(j\omega)G_p(j\omega) = -1$$

Expressing this in polar form, we obtain

$$G_c(j\omega)G_p(j\omega) = \left|G_c(j\omega)G_p(j\omega)\right| \angle G_c(j\omega)G_p(j\omega) = |-1| \angle -1$$

which implies that any root satisfying the characteristic equation must also satisfy the following 2 equations,

$$\left|G_c(j\omega)G_p(j\omega)\right| = |-1| = 1 \qquad \text{and} \qquad \angle G_c(j\omega)G_p(j\omega) = \angle -1 = -180^\circ$$

For strict stability, this means that over the entire frequency range,

$$\left|G_c(j\omega)G_p(j\omega)\right| < 1 \qquad \textbf{and} \qquad \angle G_c(j\omega)G_p(j\omega) > -180^\circ$$

Since $G_c(j\omega)G_p(j\omega)$ is the open-loop case, this means that we can determine the stability of the closed loop by examining the frequency response of the open loop system. Moreover, the differences between $\left|G_c(j\omega)G_p(j\omega)\right|$ and 1; and between $\angle G_c(j\omega)G_p(j\omega)$ and -180° will provide measures of how far the closed loop system is from stable/unstable conditions. These differences are known as the "gain margin" and "phase margins" respectively.

The gain margin (G_m) may be defined as

> *"the amount of gain that can be introduced into the system before the closed loop becomes unstable"*

Similarly, the phase margin (ϕ_m) is defined as

> *"the amount of extra phase shift that the system can tolerate before the closed loop becomes unstable"*

The gain and phase margins can be easily determined from both polar and Bode plots.

Gain and Phase Margins from Polar Plots

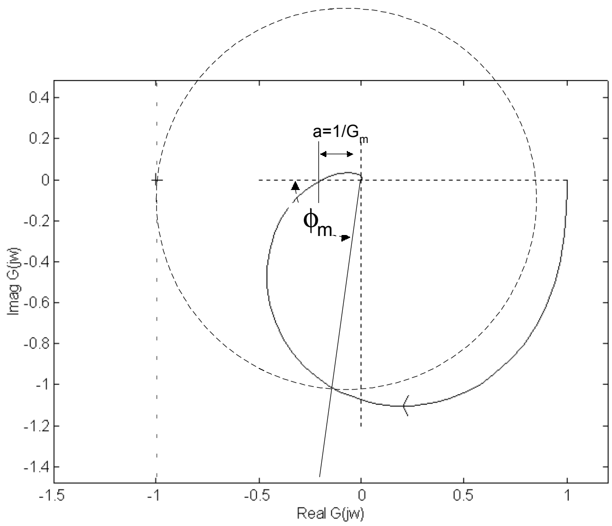

Figure 2. Gain and phase margins from polar plots

Consider the polar plot above. Suppose the distance between the origin and the point where the frequency response intersects the negative real axis is a. The value of the gain that will bring this to the -1 point is therefore $1/a$ (since any additional gain is multiplicative). Thus the gain margin is given by the inverse of the distance between the origin and the point where the frequency response intersects the negative real axis.

To measure the phase margin from a polar plot, a circle of unit radius, and centred at the origin is drawn. A line is then drawn from the origin to the point where this circle intersects the frequency response. The angle between this line and the negative real axis will be the amount of extra phase shift that the system can tolerate before it becomes closed loop unstable, i.e. the phase margin, ϕ_m.

If all the open loop components of a system is grouped as G_{OL}, then the determination of gain and phase margins from polar plots can be expressed mathematically as:

$$G_m = \frac{1}{|G_{OL}|} \quad \text{when} \quad \angle G_{OL} = -180°$$

$$\phi_m = 180° + \angle G_{OL} \quad \text{when} \quad |G_{OL}| = 1$$

Thus for the system to be stable, the gain margin must be greater than 1 and the phase margin must be positive. This also means that if the frequency response crosses the (-1,0) point, then the closed loop system will be unstable.

Gain and Phase Margins from Bode Diagrams

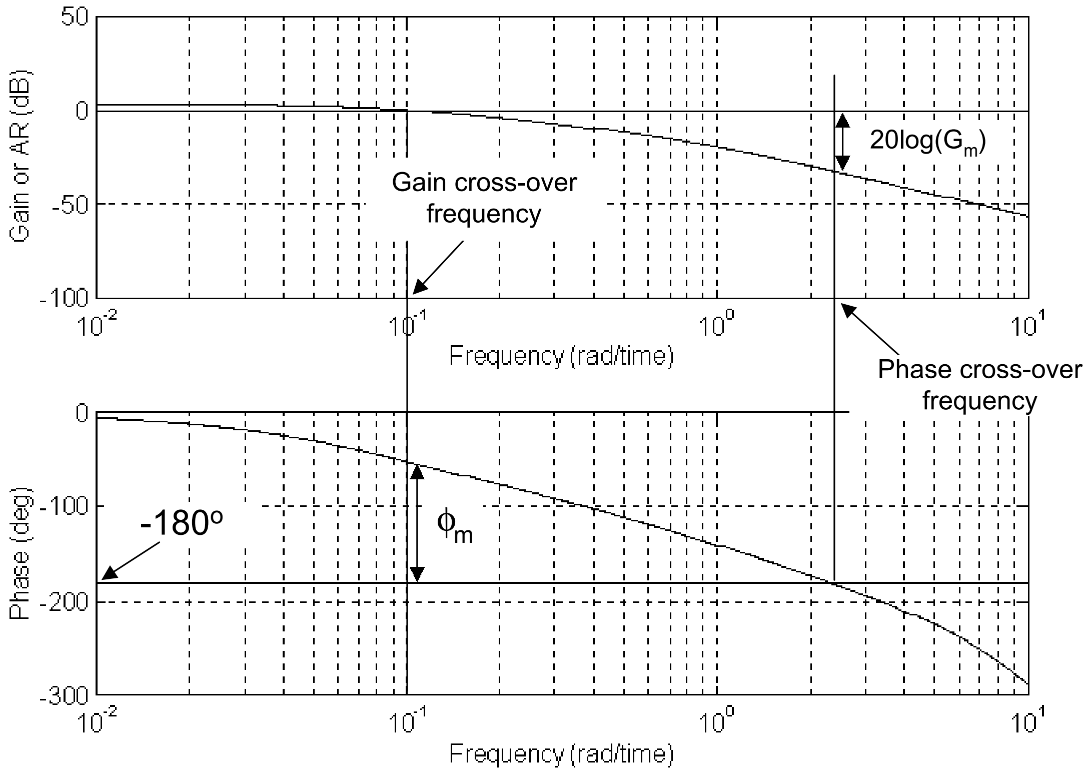

Figure 3. Gain and phase margins from Bode diagrams

Determining the gain and phase margins from Bode diagrams is just as simple. Consider the following Bode diagram. To obtain the gain margin from a Bode plot, first determine where the phase plot cuts the -180^o line. The frequency at which this occurs is known as the "phase cross-over frequency". The difference between the AR (dB) plot and the 0 dB line is the gain margin.

To obtain the phase margin from a Bode plot, determine where the AR (dB) plot intersects the 0 dB line. This occurs at the so-called "gain cross-over frequency". The difference between the phase shift at this frequency and -180^o is the phase margin.

Again, if all the open loop components of a system is grouped as G_{OL}, then the determination of gain and phase margins from Bode diagrams can be expressed mathematically as:

$$20\log(|G_m|) = -20\log(|G_{OL}|) \quad \text{when} \quad \angle G_{OL} = -180^\circ$$

$$\phi_m = 180^\circ + \angle G_{OL} \quad \text{when} \quad 20\log(|G_{OL}|) = 0$$

Again for the closed loop system to be stable, the phase margin and the gain margin (expressed in dB) must both be positive.

Summary

In summary, frequency response provides a powerful stability analysis tool. For a closed loop system to be stable, the frequency response of its open loop components must have a positive phase margin and the gain margin must be greater than 1 (0 dB). It should also be clear that stable delay free first and second order systems under proportional control will always remain stable in the closed loop. This fact can best be visualised using the polar plot: the frequency response will never cross the negative real axis.

CONTROLLER TUNING

Offline Z-N tuning

Controllers can also be tuned using frequency response techniques. Perhaps the best known method is the Ziegler-Nichols off-line tuning approach. Instead of attempting to drive the process to constant amplitude oscillation under proportional control, the offline approach uses a model of the open loop process to determines its gain margin. Recall that the gain margin is the amount of extra gain that can be introduced into the system before the closed loop becomes unstable. Thus, we know that when we impose this additional gain on the system, the closed loop will become marginally unstable, i.e. respond with constant amplitude oscillation. Thus, the gain margin is equivalent to the "ultimate gain" determined experimentally. The marginally stable system will oscillate with a frequency equal to the gain cross-over frequency. Thus, we can also determine the "ultimate period",

Ultimate gain: $K_u = G_m$

Ultimate period: $P_u = \dfrac{2\pi}{\omega_G}$ where ω_G = gain cross-over frequency

which is the inverse of the gain cross-over frequency. Having determined the "ultimate gain" and "ultimate period", these can be substituted into the Z-N tuning formulae below to obtain the appropriate P, PI or PID settings for the process.

	P	PI	PID
K_c	$K_u/2$	$K_u/2.2$	$K_u/1.7$
T_i	-	$P_u/1.2$	$P_u/2$
T_d	-	-	$P_u/8$

Ziegler-Nichols Tuning Formulae

Specifications based on gain and phase margins

Controllers have also been designed based on gain and phase margin specifications. Typical requirements are to aim for a phase margin between 45^o to 60^o, or a gain margin of between 2 and 3. Say we wish to control a process using a PI controller $G_c(s)$. Given the process model $G_p(s)$, we can then plot the frequency response of $G_c(j\omega)G_p(j\omega)$ and adjust the parameters K_c and T_i to meet the frequency domain specifications. Note that K_c will only affect the amplitude ratio and not the phase characteristics. Also, controller design in the frequency domain takes into account in a natural manner, both the phase shift and gain requirements of the control system. Laplace domain designs, on the other hand, tend to consider only the gain.

Online Tuning using Relay feedback

In both the above tuning techniques, it is assumed that we have a model of the process from which to generate the open loop frequency responses for control system design. Unfortunately, sometimes a process model is not available. Nevertheless, it is still possible to tune controllers based on frequency domain considerations, but we have to do it online, much like the online Z-N tuning method. The relay feedback tuning scheme for a PI controller is illustrated in the schematic below:

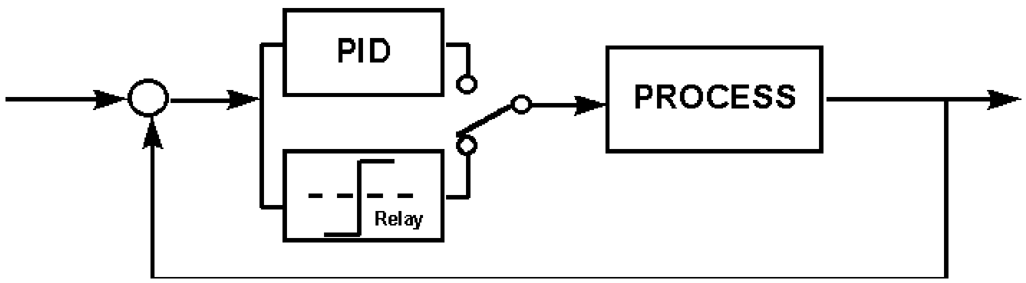

Figure 4. Schematic of relay feedback tuning scheme

A relay is set up in parallel with the PI controller. During tuning, it is switched in to perturb the process in the close loop. The output of the relay is essentially a series of ON-OFF signals of pre-determined amplitude, and this will drive the process output to sustained oscillations, i.e. a limit cycle, as shown in the following diagram.

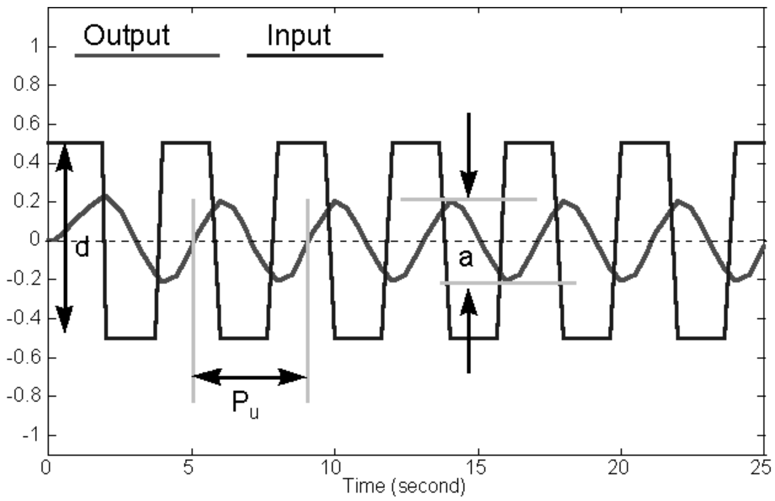

Figure 5. Result of relay feedback on output signal

The ultimate gain is determined from, $K_u = 4d / \pi a$ and the ultimate period is given by the period of the output limit cycle, P_u. Once these parameters are found, they can be used in the Z-N formulae to get appropriate P, PI or PID settings.

An advantage of online frequency domain based tuning techniques is that we do not need to know the structure of the process, i.e. whether it is first or second order, and whether it possesses a time delay. Unlike the online Z-N tuning method though, the relay feedback tuning strategy will always maintain the process around its nominal set-point. The scheme presented here is also the basis for many commercial auto-tuning controllers. These commercial devices use slightly more sophisticated approaches to mitigate the effects of process noise and disturbances, and variations of the Z-N tuning formulae. A discussion of how they function is however, beyond the scope of this course.

ROBUSTNESS ANALYSIS

In the design of a process control system, linear models of the plant are often used. The advantages are that:

- linear models can readily be developed from plant operating data
- there is a wealth of knowledge on the design of linear control systems.

However, process plants are usually very complex systems, i.e.

- they are affected by many inputs and outputs, some measurable some not.
- they suffer from varying time delays
- they possess high order-dynamics
- they often exhibit non-linear behaviour.

Therefore, linear models may not be sufficiently accurate. Since the control system has been designed based on a linear model, a couple of questions immediately arise:

- will the control system still perform?
- will the closed loop system remain stable?

when the real plant characteristics are different from those condensed in the model used for control system design. Given that we have knowledge only of Laplace domain procedures, how would we answer these questions?

An Illustrative Example

Consider the simple example of the servo control of a first order process with time delay. The Laplace transfer function model of the process is therefore:

$$G_p(s) = \frac{Y(s)}{U(s)} = \frac{K_p \exp(-\theta s)}{1+\tau s}$$

where K_p is the process gain

θ is the process delay

τ is the time constant of the process

Y(s) is the output

U(s) is the input

Suppose a P+I controller has been designed based on this model using one of the numerous recipes for conventional controller design, such as Cohen-Coon, 3-C etc., and this has the transfer function

$$G_c(s) = \frac{U(s)}{E(s)} = K_c\left[1+\frac{1}{T_i s}\right]$$

where K_c is the proportional gain

T_i is the integral time constant

and *E(s)* is the error between set-point and output

We have to ensure that

- the controller $G_c(s)$ provides stable control

Department of Chemical and Process Engineering
University of Newcastle upon Tyne

- the closed loop performance is acceptable (e.g. is the overshoot reasonable? is the rise time fast enough?)

To investigate stability, we normally develop the closed loop expression, i.e.

$$\frac{Y(s)}{R(s)} = \frac{G_c(s)G_p(s)}{1+G_c(s)G_p(s)}$$

where *R(s)* is the desired output level or set-point

- Closed loop stability is ensured if the roots of the characteristic polynomial, $(1+G_c(s)G_p(s)$, have negative real parts

- Given that the closed loop will be stable, closed loop performance can be easily assessed by applying the controller to a simulation of the linear model.

If we are not happy with either of the above results, then we have to re-design the P+I controller and repeat the analysis.

Suppose you are satisfied with the results of both investigations …. it is now time to apply the controller to the real process!

- What are the factors that will compromise the integrity of the closed loop?

  The controller has been designed based on a linear (assumed accurate) model of the process to be controlled.

  If the model is inaccurate or if the characteristics of the process change, then the designed control system will no longer be appropriate.

  This is the problem of process-model mismatch

- In the presence of process-model mismatch, how confident are you that the closed loop will be stable and will perform as required?

- Can we account for these factors in the design of the controller prior to implementation?

Robustness Analysis in the Laplace Domain

Working within the Laplace domain, there are two possible ways of tackling the problem of process-model mismatch in control systems design.

- Evaluate the designed controller against a variety of process models via simulation

This will enable us to establish some measure of the conditions under which the controller will or will not perform or maintain closed loop stability.

However, this can be time consuming and some significant conditions may be omitted.

- Stipulate some degree of uncertainty in the process model mathematically, and analyse the performance and stability of the controller

Here the formulated problem may not be tractable. To give you an indication of this difficulty, suppose we say that there is an error of ΔK_p associated with the gain in our process model. To analyse the effects of this error on closed loop stability say, the resulting closed loop transfer function would be:

$$\frac{Y(s)}{R(s)} = \frac{G_c(s)G_p(s)}{1+G_c(s)G_p(s)} = \frac{G_c(s)\dfrac{\left(K_p \pm \Delta K_p\right)\exp(-\theta s)}{1+\tau s}}{1+G_c(s)\dfrac{\left(K_p \pm \Delta K_p\right)\exp(-\theta s)}{1+\tau s}}$$

We would now have to assess the stability of the system using the closed loop expression:

The closed loop equation is then

$$1+G_c(s)\frac{\left(K_p \pm \Delta K_p\right)\exp(-\theta s)}{1+\tau s} = 0$$

i.e. check the roots of the following 2 expressions:

$$1+G_c(s)\frac{\left(K_p + \Delta K_p\right)\exp(-\theta s)}{1+\tau s} = 0$$

and

$$1+G_c(s)\frac{\left(K_p - \Delta K_p\right)\exp(-\theta s)}{1+\tau s} = 0$$

This is actually not that difficult, but when we include possible errors in the time-delay and time constant terms, then the closed loop expression becomes more complicated, i.e.

Department of Chemical and Process Engineering
University of Newcastle upon Tyne

$$1 + G_c(s)\frac{\left(K_p \pm \Delta K_p\right)\exp(-[\theta \pm \Delta\theta]s)}{1+(\tau \pm \Delta\tau)s} = 0$$

In this case, we would have to solve for the roots of the closed loop equation for various combinations of $\pm\Delta K_p$, $\pm\Delta\theta$ and $\pm\Delta\tau$!!

Moreover, we would have to ensure that the magnitudes of ΔK_p, $\Delta\theta$ and $\Delta\tau$ are physically compatible.

Robustness Analysis in the Frequency Domain

Therefore, while Laplace Transforms have simplified the analysis of linear systems in many ways, its use becomes unwieldy when we want to investigate system properties under 'real' conditions. The frequency domain offers a much simpler approach because the parameters of a process in the frequency domain are condensed into two parameters, the amplitude ratio and the phase-shift. This is regardless of the order of the process and whether the process possesses dead-time. Thus, unlike analysis in the Laplace domain, analysis of the effects of process model mismatch in linear systems reduces to studying the effects on these two parameters over a range of frequency values.

In the frequency domain, the frequency response of the system is simply a locus of frequency dependent points plotted on the complex plane. For example, the frequency response of the nominal model is plotted as the continuous line in the figure below. Process model mismatch would lead to a different locus of points.

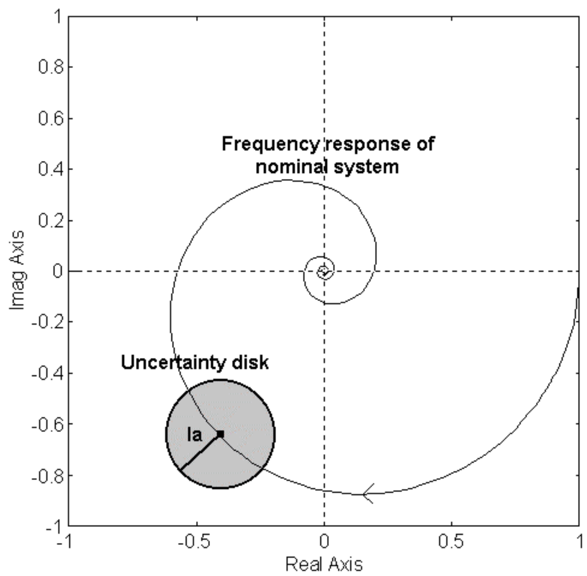

Figure 6. Describing model uncertainties in the frequency domain

To represent process-model mismatch, all that has to be done is to quantify a frequency dependent region that would include all possible mismatch scenarios at a particular frequency. Due to its simple geometry, the uncertainty region is usually specified as a circle. Therefore, instead of postulating errors for individual parameters as in the case when performing Laplace domain analysis, in the frequency domain, all we need to do is to specify a radius for the uncertainty disk . The radius of this disk is chosen to encapsulate all the possible variations in amplitude ratio and phase shift at a particular frequency.

DESIGN OF NOISE SUPPRESSION FILTERS

Another aspect of process control where frequency domain techniques has been traditionally useful, is in the design of noise suppression filters. Process measurements are usually corrupted by high frequency disturbances from different sources. The nature of the process itself may contribute to high frequency transients. These are collectively termed "noise" which if excessive, can be detrimental to the control system. A well designed controllers should not react to noise. There are two ways to achieve this using frequency response techniques:

a) choose controller parameters such that it effects control mainly in the low frequency region

b) design filters to attenuate noise from measurements used for feedback control

The two approaches are related. The first integrates noise suppression function within the controller, thus accounting simultaneously for the additional phase lag that normally occurs in noise filtering. The second separates the noise suppression task from the control task, but in this case, the controller may have to be redesigned to accommodate the extra phase shifts due to the implementation of noise filters. The degree of noise attenuation is best visualised from Bode diagrams. The diagram below is the frequency response of a first order lag with unit gain and time constant of 10.

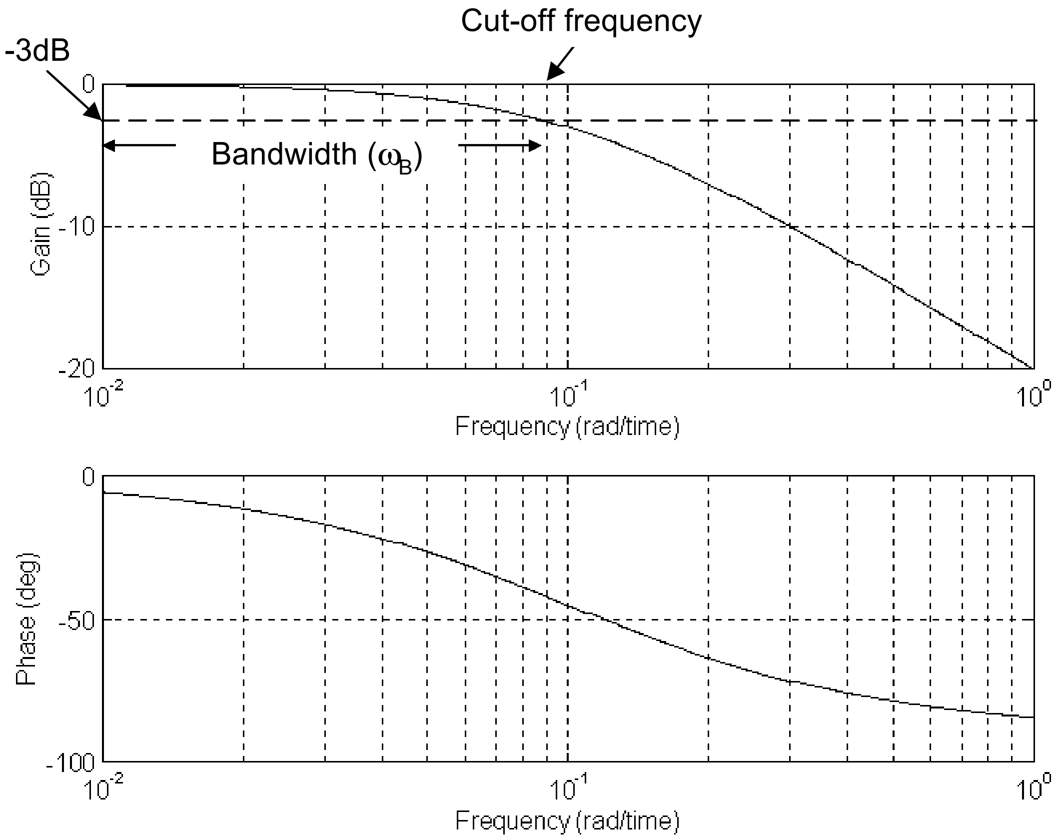

Figure 7. Bandwidth and cut-off frequency

From the plot, we can observe that the amplitude ration decreases with increasing frequency. In other words, this first order lag has the ability to attenuate signals in the high frequency range. The degree of attenuation is usually measure by the “bandwidth”, which is defined as the frequency range over which the amplitude ratio plot remains above -3dB. The upper frequency value of the bandwidth is known as the “cut-off frequency” and defines the frequency above which the signal is considered to be negligible. Note that the cut-off frequency is identical to the “corner frequency” for first order lags. Thus, we can deduce that a first order lag with larger time constants will have a smaller bandwidth, hence better filtering properties. This is shown below in the amplitude ratio plots of 3 first order lags, with time constants 10, 20 and 30. Here we can observe clearly that the bandwidth becomes smaller with increasing time constants.

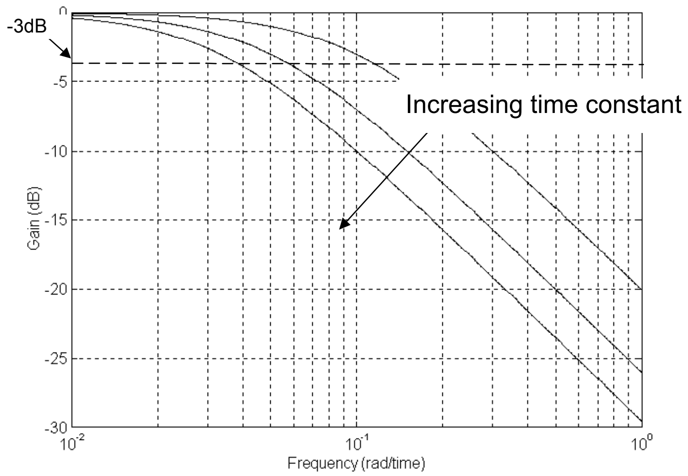

Figure 8. Decreasing bandwidths with increasing time constants

Another factor that affects the efficiency of noise suppression is the rate of signal attenuation. The rate of attenuation is given by the slope of the amplitude ratio plot. Again, from this, we can further deduce that higher order lags will have sharper cut-offs. This feature is illustrated by the Bode diagram below which shows the amplitude ratios of $\frac{1}{(1+10j\omega)^n}$, with n = 1,2 and 3.

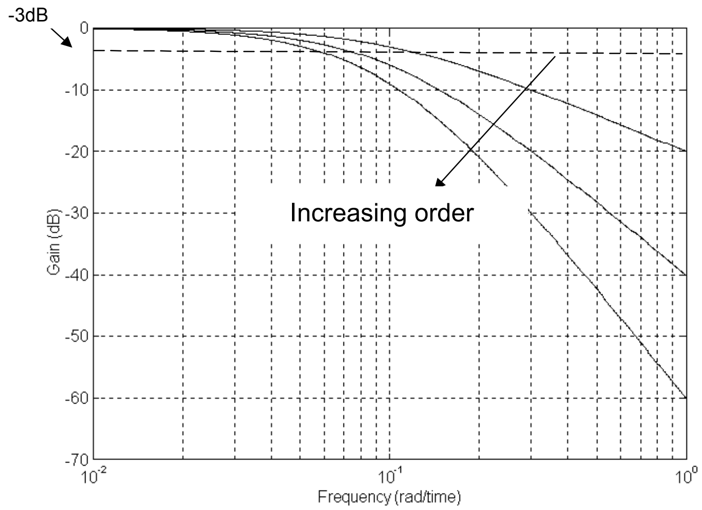

Figure 9. Decreasing bandwidths and higher attenuation rates with increasing order

This plot shows that the rate of attenuation increases with increasing order. Additionally, the bandwidths are also reduced with increasing order. However, high order filters come at a price; the higher order the filter, the greater the phase shift, and this can severely affect the performance of the closed loop system as discussed previously. The design of noise filters is therefore an exercise in compromise between maximum noise attenuation and minimising the amount of phase shift that is incurred. Clearly, frequency domain analysis has a significant role to play in helping to solve this problem.