Clarification of doubts: Bode plots

Q1 *If all poles and zeroes are on RHP, then also shouldn't it be a minimum phase system? Is it because so RHP poles that we don't consider these?*

Answer: When you consider all (or even one pole for that matter) to be on the RHP, remember that the magnitude and phase plots that you will be dealing with do *not* convey any information about steady state information, which a Bode plot is supposed to do. This is because an RHP pole would make the system unstable, thereby ruling out any *steady state*. This is not to say that the functions $M(\omega)$ and $\phi(\omega)$ cease to exist in such cases. You can still plot those functions, but they are not meaningful in steady state analysis.

Since the terms "minimum phase" or "non-minimum phase" appear in the context of Bode plots, using which a steady state behaviour is to be analyzed, we do not consider the existence of RHP poles while defining "minimum phase" or "non-minimum phase" systems.

Q2 *Can you please explain the Errata: 1 posted in the comments? I think there is no mistake in the video!*

Answer: I had said (and also written) that "...all n p_is are in lhp". While my statement was correct till I maintained that $p_i > 0$, the equivalent statement should have been "...all n $-p_i$s are in lhp". This is because I wrote the factors as $(s + p_i)$, thereby making $-p_i$ the pole rather than p_i. So, for stability, we need $-p_i$ to be in the lhp, or p_i in the rhp. This mistake necessitated the comment in Errata 1.

Q3 *At 55:10 What if $|n - m + 2r| < |n - m|$ (ex: n=2,m=4,r=1) ? Then the system with 1 RHP zero has lesser variation in net phase than the system with all LHP poles and zeroes. The condition that all zeroes and poles should be in LHP seems wrong here.*

Answer: Once again, in the context of control systems, we deal with proper/biproper systems (this significance will be clearer when we study state space analysis; improper transfer functions have no state space descriptions! However, that is a story for another day...) where the number of poles, n, is always greater than or equal to the number of zeros, m.

Even from Bode plots, you can infer that if $m > n$ for any system, the eventual high frequency asymptote will rise at the rate of $20(m - n)$dB/decade. This will imply amplification of high frequency signals. Note that noises are also of high frequency and need to be rejected, instead of amplified. So from practical considerations, you do not want to design controllers that lead to such improper transfer functions.

For these reasons, we always consider $n > m$ as a prerequisite in our analyses (I think I had mentioned in passing that $n > m$, during the lecture, though there is a chance that I may have missed it.).

N.B.: You may argue that a standalone PID controller (of the form $K_p + K_i/s + K_d s$) is improper, and you would be correct. But while implementing a PID control in practice, one always uses something called a roll-off pole (leading to a transfer function such as $\frac{K_p + K_i/s + K_d s}{1 + s\tau_{\mathrm{r}f}}$) which flattens the magnitude plot beyond a certain frequency ($1/\tau_{\mathrm{r}f}$ being a sufficiently large number).

Q4 *Why are you saying that we have to shift the graph towards the right when we increase the gain? Are we not going to consider HPF or BPF transfer functions? Shouldn't it be technically moving the graph up and not right?*

Answer: I do not recall if I had phrased it exactly the way you put it. Nevertheless, what I meant was that an increase in gain lifts the magnitude plot up and pushes the gain crossover frequency (and results in larger bandwidth) to the right (I think that this is trivial to see and you should try to convince yourself that it is so). While an increase in gain is desirable to achieve lower rise time (larger bandwidth) requirements, it is detrimental to stability (generally leads to lower gain and phase margins). Therefore, we have to tread cautiously here while designing suitable value of gain K.

In fact, I have provided a detailed explanation about the line of reasoning employed here, in response to one of your classmates' (Aniruddh) queries. Please go through the same on moodle. If many of you find it hard to follow that explanation, I will consider including it in a separate video.

As transfer functions of most physical systems are generally of a low pass nature, I did not consider HPF and BPF. You are, however, free to do so. The important point to remember, though, is that even for such systems you can see the analogy between root locus method and frequency domain methods. That was the main purpose of my discussion.

In fact, you can encounter all sorts of systems whose magnitude plots may even attain 0 dB or -3 dB values at multiple values of frequency, and/or the phase plot may attain a value of $180°$ at multiple values of frequency. In such cases the bandwidth is taken to be the lowest value of frequency at which -3dB is crossed while the phase margin is given by the lowest value of $180° + \phi(\omega_{\mathrm{gcf}_i})$. Similarly, the gain margin is given by the least value of $-20 \log |G(j\omega_{\mathrm{pcf}_i})|$ (i, j denote the indices of the multiple gain and/or phase crossover frequencies, respectively). I consciously chose not to deal with such 'rogue' systems to start with. Since this is an introductory course, it is more important to grasp the main concepts and to be able to relate the features extracted from Bode plot with those obtained from the root locus. If that much is clear, you can then go on to study other types of systems and draw the analogies between the concepts by yourself.

Q5 *At 1:16:00, what is k_0? It should be 1 in order to match the 2 different GM definition given in lecture: $20 \log \gamma_{crit}$ and $-20 \log |G(j\omega_{pcf})|$.*

Answer: I think I see what caused the confusion here. Thanks for pointing this out! Consider k_0 to be the gain that is inherent to the plant (I think I called it nominal gain in the video) and let γ be akin to some controller gain (or maybe changes in the nominal gain, k_0, due to ageing, or wear and tear of components). So k_0 can have any constant value, as it comes from the plant's transfer function. I considered γ and k_0 together as an ensemble block with the plant, say $KG(j\omega)$, where $K = \gamma$ (not $K = \gamma k_0$, as mentioned in the video) can be tuned (or captures the changes in k_0 due to wear and tear/faults), and showed the resultant effect of this tuning/change on the Bode plot (i.e., raising or sinking the function $20 \log |KG(j\omega_{\mathrm{pc}f})|$). Therefore, in evaluating $|G(j\omega_{\mathrm{pc}f})|$, we have already accounted for k_0. If, on the other hand, you do not consider k_0 as part of $G(j\omega)$, then your expression for gain margin will read $20 \log |\gamma_{crit} k_0|$ with $K = \gamma k_0$ being a part of the controller.

Q6 *At 1:21:30, if GM, as defined just above, is positive, then we can only increase the gain till $|G(j\omega_{pcf})|$ becomes 1. I think it should be - we can keep decreasing the gain without encountering instability.*

Answer: I do not see any conflict in the two statements you have made, in the context of the example we had considered. Indeed, there is no lower bound on the value of gain that would result in closed loop poles migrating to the rhp.

However, as I had explained earlier, reducing the gain does not serve the purpose of reducing the rise time (increasing bandwidth). At the heart of any controller synthesis problem is striking the right balance between obtaining satisfactory performance, while also having a sufficient margin for stability. We cannot be so deeply concerned with ensuring stability (for which a low value of gain is good enough) as to completely overlook the closed loop performance requirements. At the same time, we cannot afford to design a controller that leads to closed loop instability either. So, we are not interested in statements or results that guarantee stability at lower values of gain, mainly because we are interested in the stability question for larger values of gain, where bandwidth is likely to be satisfactory.

Q7 *1:28:50 Why is a negative phase margin desirable? Is it because we actually want to increase k without encountering instability?*

Answer: You have deduced correctly. As I explained in response to **Q6**, we want to increase the gain to obtain satisfactory bandwidth for the closed loop system (which is close to the open loop bandwidth; see my response to Aniruddh's query on moodle for details). A negative phase margin (when defined as $PM = -180^\circ - \phi(\omega)$) or a positive phase margin (when defined as $PM = \phi(\omega) + 180^\circ$) helps us to choose higher values of gain that facilitate this.

By the way, the second definition of phase margin is more common, so you might as well use that and consider a positive phase margin as desirable.

Q8 *Why do we care about this thing where changing gain may result in an unstable system? If we just tweak the gain a little, we won't have the instability!*

Response: The question of gain margin assumes significance both from an analysis and a synthesis perspective. While synthesizing a controller, you might say that choosing a gain that is slightly below the critical value will ensure stability and also meet rise time specifications satisfactorily. However, what if you are required to meet more performance requirements, such as certain bounds on resonant peak, M_r, or some phase margin? In such cases just evaluating the critical value of gain and choosing k to be slightly lower than that critical value may not satisfy all the deign requirements.

In fact, designers often make a heuristic approximation given by $\zeta \approx \mathrm{PM}/10$ (PM is in radians and this relation holds for some restricted values of ζ only), which relates the frequency domain quantity, phase margin, with the time domain quantity ζ, and specify certain requirements on percentage overshoot of the closed loop system. In such cases, it is important to know what the phase margin is. Similarly, gain margin too has significance from a synthesis point-of-view.

However, for analyzing a system, the roles of gain margin/phase margin cannot be overemphasized. Suppose you are faced with the following question: Given a plant and a controller, how much change is acceptable in the dc gain of the plant till the closed loop system becomes unstable?

This question is one that provides information about the robust stabilty of the closed loop system or the robustifying property of the controller. So, you may have designed a controller (say, a static gain) for the nominal plant many years back. But over several years of operation, wear and tear may have led to changes in the dc gain, leading to changes in both gain and phase margins. The question is, whether the controller you had designed all those years ago is still good enough to result in a stable closed loop behaviour. If the answer is in the affirmative, then you would naturally want to know, as part of safety assessment, how much more wear and tear (changes in dc gain) can the plant undergo, before the closed loop system becomes unstable. Here, phase margin and gain margin provide the answer. Now, if you had designed a controller that provided sufficient gain and phase margins, obviously your plant would be able to withstand further wear and tear before the closed loop system becomes unstable.

By the same token, if your plant's poles and zeros shift due to wear and tear, the $\phi(\omega)$ plot will also be altered along with the $20 \log |M(\omega)|$ plot, thus leading to changes in both the phase and gain crossover frequencies, and consequently in both gain and phase margins. Intuitively, it stands to reason that if your designed controller had initially resulted in a large difference between $\omega_{\mathrm{pc}f}$ and $\omega_{\mathrm{gc}f}$, then the controller is more robust to changes/perturbations in the plant's parameters. These and more such key insights can be gleaned from quantities such as gain and phase margins.

Once again, let me emphasize that EE302 being an introductory course, we do not have the scope of exploring these and several other such avenues in greater detail. These will come within the purview of more advanced courses on controls. So long as you start developing a basic appreciation for and working knowledge of these time-domain and frequency-domain methods, and are able to see the connections between these approaches, the purpose of this course would have been served.

Nevertheless, I would encourage all of you to keep asking questions and even if some questions are not addressed during regular lectures, mostly due to time-constraints, you may use the office hours, upon resumption of regular classes, to clarify your doubts.