UNDERGRADUATE PROCESS CONTROL

Clarification of Some Concepts

R. RAVI*
Indian Institute of Technology Kanpur • Kanpur 208 016, India

Teaching undergraduate process control can be an enjoyable experience for an instructor given the wide range of quality chemical engineering textbooks that are now available.[1-6] After teaching the course a couple of times, however, I felt there was still a need for clarification of some fundamental concepts, especially in the areas of frequency response and stability. In this article I hope to achieve such a clarification through some simple, yet illustrative, examples.

FREQUENCY RESPONSE: ONLY FOR STABLE SYSTEMS?

In the context of process control, the frequency response is usually associated with the response of a linear, time invariant (constant coefficient) system to a sinusoidal input. In the most common way of "deriving" the frequency response result, the output response is shown to be a sinusoidal function of the same frequency (ω) as the input, *once the transients have died out.* Further, the ratio of the amplitude of the output to that of the input, called the amplitude ratio (AR), is shown to be equal to $|G(j\omega)|$, while the phase difference (ϕ) between the output and the input is shown to be $\arg[G(j\omega)]$, where G(s) is the transfer function representation of the system of interest and $j=\sqrt{-1}$.

Thus, the frequency response calculation is reduced to the calculation of the magnitude and phase of the complex number, $G(j\omega)$, as a function of the frequency. This information is usually represented in the form of a Bode diagram or a Nyquist plot.

The key point of our discussion is the condition

"once the transients have died out."

Clearly, this happens if the system is stable, *i.e.,* if all the poles of the transfer function G(s) lie in the left half (of the complex) plane (LHP). Thus it might appear that frequency response makes sense only for stable systems. But we do find Bode diagrams and Nyquist plots for the pure capacity (G(s)=A/s) and the PI controller, $G(s)=[K_c(\tau_I s+1)]/\tau_I s$, both of which are (open-loop) unstable.

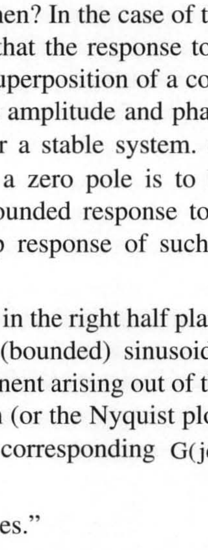
Do these diagrams mean anything then? In the case of the pure capacity system, one can show that the response to a sinusoidal input is bounded and is a superposition of a constant and a sinusoidal function whose amplitude and phase are in fact provided by $G(j\omega)$, as for a stable system. (It should be noted that a system with a zero pole is to be regarded as unstable in spite of a bounded response to a sinusoidal input. Recall that the step response of such a system grows with time.)

But what about a system with a pole in the right half plane (RHP) for which the response to a (bounded) sinusoidal input will have a time-growing component arising out of the unstable pole? Does the Bode diagram (or the Nyquist plot) for such a system obtained from the corresponding $G(j\omega)$ have any meaning?

The answer to the last question is "yes."

The common way of deriving the frequency response result is only a method of measuring the frequency response for stable systems and does not constitute a fundamental

***R. Ravi** obtained his BTech from the Indian Institute of Technology, Madras, in 1984, and his PhD from Purdue University in 1991. His research interests are in applied statistical mechanics and process control. For the past few years, his abiding passion has been the understanding of the origins of thermodynamics and fluid mechanics.*



* *Present address: Indian Institute of Technology Madras, Madras 600028, India*

definition of it. The fundamental definition is provided by a basic result of linear systems theory.[7] There exists a periodic solution for a linear time invariant system subjected to a periodic forcing; this periodic solution has the same frequency as that of the input forcing, and its amplitude and phase at the particular frequency are determined (as explained above) from the complex number $G(j\omega)$. This result holds whether the system is stable or not.

In general, the response of a linear system to a periodic forcing will be the superposition of the periodic solution and a non-periodic component, and the frequency response is defined with respect to the periodic component. Thus, the Bode diagram for an unstable system makes sense in that it represents the same relationship between the periodic component of the (output) response and the input periodic forcing as it does for a stable system.

This point is not of minor significance as it gives universal status to Bode diagrams or Nyquist plots as signatures of systems they represent, be they stable or unstable. The open-loop method of measuring the frequency response (after waiting for the transient to die out) will not work for unstable systems (pure capacity being an exception).

In the next section, we point out two possible methods of measuring the frequency response of unstable systems—one an open-loop method and the other a closed-loop method. Although both methods are valid in principle, the latter is more practicable. The reasons are outlined below.

Frequency Response of Unstable Systems

We illustrate the procedures through a simple system with one unstable pole

$$G_o(s)=\frac{K}{(s-a)} \qquad (1)$$

Open-Loop Method

For the Open-Loop Method we consider a sinusoidal input

$$u(t)=A_u \sin(\omega t+\phi_u) \qquad (2)$$

The response of the system to this input can be shown (for instance, by a straightforward Laplace inversion) to be

After teaching [undergraduate process control] a couple of times, I felt there was a need for clarification of some fundamental concepts, especially in the areas of frequency response and stability. In this article I hope to achieve such a clarification through some simple, yet illustrative, examples.

$$y(t)=\frac{KA_u(\omega\cos\phi_u+a\sin\phi_u)e^{at}}{a^2+\omega^2}$$
$$+A_u\,|G_o(j\omega)|\sin\{\omega t+\varphi_u+\arg[G_o(j\omega)]\} \qquad (3)$$

This suggests a way of "stabilizing" the response by choosing ϕ_u such that

$$\omega\cos\phi_u+a\sin\phi_u=0 \qquad (4)$$

so that only the *stable* periodic component of the solution remains, enabling the determination of its amplitude and phase. In practice, thus, one is left to choose a unique value of ϕ_u (between 0 and 2π) for each ω; this can be problematic given that the value of the unstable pole, a, is not known *a priori*. Hence, we discuss a more practicable method involving closed-loop stabilization.

Closed-Loop Method

We consider the same first-order unstable system. It is easy to show that the system can be stabilized in a feedback loop using a proportional controller of gain K_c greater than a/K (Figure 1 illustrates the scheme). In fact

$$\frac{y(s)}{r(s)}=\frac{K_cK}{s+b}\equiv G_{CL}(s) \qquad (5)$$

where $b=K_cK-a>0$.

If a sinusoidal variation is given in the reference signal, r, *i.e.*,

$$r(t)=A_r\sin\omega t \qquad (6)$$

then we can show that (by Laplace inversion, for instance)

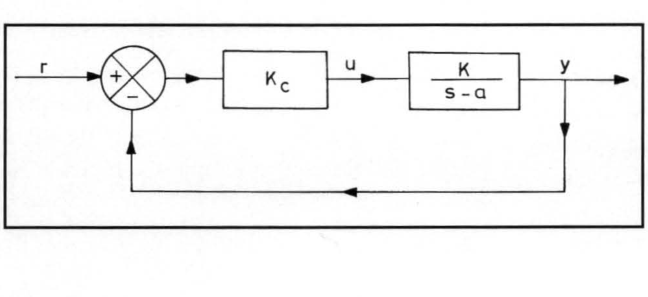


Figure 1.
An open-loop unstable system in a feedback loop with a proportional controller.

$$y(t) = C_1 e^{-bt} + A_y \sin(\omega t + \phi_y) \quad (7)$$

where

$$C_1 = \frac{K_c K A_r \omega}{\omega^2 + b^2}; \; A_y = A_r |G_{CL}(j\omega)|; \; \phi_y = \arg[G_{CL}(j\omega)] \quad (8)$$

The signal u(t) = K_c[r(t)-y(t)] can be expressed as

$$u(t) = -K_c C_1 e^{-bt} + A_u \sin(\omega t + \phi_u) \quad (9)$$

It is possible to show that

$$\frac{A_y}{A_u} = |G_o(j\omega)| \quad \text{and} \quad \phi_y - \phi_u = \arg([G_o(j\omega)]) \quad (10)$$

i.e., the amplitudes and the phases of the "input" and the "output" signals of the unstable system, $G_o(s)$, are related as before by the complex number $G_o(j\omega)$. The stabilization effect is noted in the e^{-bt} term (note: b > 0) in both y and u in contrast to the open-loop case where we get the time-growing term, e^{at}, in the output (for the same input $A_r \sin \omega t$). For concreteness and simplicity, we illustrate the above result with a numerical example.[8] We choose

$$G_o(s) = \frac{2}{s-1} \quad (11)$$

It is easy to see that a unity gain (K_c = 1) proportional controller stabilizes the above system in a feedback loop. In fact

$$\frac{y(s)}{r(s)} = \frac{2}{s+1} \quad (12)$$

If we choose the input to be

$$r(t) = 0.5 \sin 2t \quad (13)$$

then we can show that

$$y(t) = \frac{2}{5} e^{-t} + (0.2)^{1/2} \sin[2t - 1.11(\text{rad})] \quad (14)$$

Further

$$u(t) = r(t) - y(t) = \frac{-2}{5} e^{-t} + 0.5 \sin[2t + 0.93(\text{rad})] \quad (15)$$

and

$$|G_o(2j)| = \frac{2}{\sqrt{5}} \quad \text{and} \quad \arg[G_o(2j)] = -2.04 \text{ rad} \quad (16)$$

Thus, we see that

$$|G_o(2j)| = \frac{A_y}{A_u} \quad \text{and} \quad \arg[G_o(2j)] = \phi_y - \phi_u \quad (17)$$

Of course, the above analysis is based on a given system transfer function. This is not known *a priori* and, in fact, the purpose of the frequency response experiment is to determine the transfer function. But what one has to do is to tune the proportional controller to obtain a stable system. Then, for a known sinusoidal input, r(t), at various frequencies, one would have to measure the amplitude and phase of both u(t) and y(t) (after the transients die out) to construct the transfer function, $G_o(s)$.

FREQUENCY RESPONSE AND STABILITY CRITERIA

We now turn to another aspect of frequency response and stability, the famous Nyquist stability criterion. The Nyquist criterion helps one to infer the stability of a feedback control system from the Nyquist (polar) plot of the loop transfer function, $G_L(s)$, which is the product of the transfer functions of all the elements in the control loop. The advantage of stability criteria based on frequency response is their ability to deal with non-polynomial $G_L(s)$ that the Routh-Hurwitz criteria cannot treat rigorously. This advantage is particularly relevant to chemical engineering systems that often contain a time-delay element.

Most chemical engineering textbooks on process control do not give as much prominence to the Nyquist criterion as they do to the Bode stability criterion, which is easier to use. An exception is the Luyben[2] book where a detailed discussion with illustrative examples can be found. It is to be noted that the Bode criterion is not general and specifically cannot be applied in cases where the Bode diagram for $G_L(s)$ is not monotonically decreasing. It is our objective here to highlight the potential sources of error in the application of the Nyquist criterion. It is not uncommon to find special statements of the criterion that might work in many cases but fail to yield the correct result for at least some systems. Often, these special statements are not accompanied by the conditions under which they hold. Thus it is desirable to always use the general form of the criterion that is given below.

Let N be the number of net rotations of the Nyquist plot of $G_L(s)$ $(-\infty < \omega < \infty)$ about the point (-1,0). This is the net angle traced out by the line segment from (-1,0) to the Nyquist plot as the frequency changes from $-\infty$ to ∞. The sign convention is a positive value for N if the net rotation is in the counterclockwise direction and negative if it is in the clockwise direction. Let P_R be the number of poles of $1+G_L(s)$ (note that this is the same as the number of poles of $G_L(s)$) in the RHP. Then

$$Z_R = P_R - N \quad (18)$$

where Z_R is the number of zeros of $1+G_L(s)$ in the RHP. Hence, Z_R is the number of roots of the characteristic equation $1+G_L(s)=0$ that lie in the RHP. Clearly, Z_R must be zero for a stable system.

It is not our objective here to give a proof of the above statement (see, for instance, Ref. 9), but we illustrate its proper use through a simple example. In our opinion, the following points are crucial:

- While the portion of the Nyquist plot from $-\infty$ to 0 is

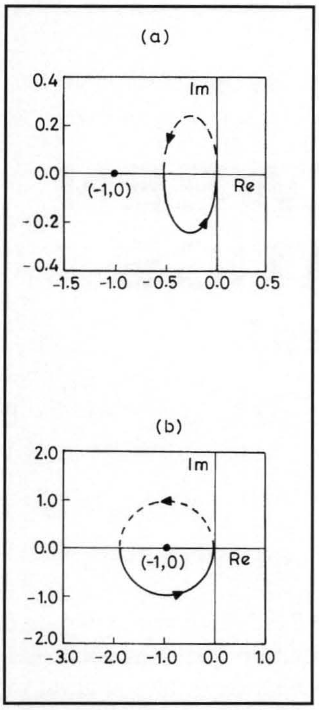


Figure 2. *Nyquist plots for a)* $G_{L1}(s)=\frac{1}{2(s-1)}$ *and b)* $G_{L2}(s)=\frac{2}{(s-1)}$ *The dotted (---) portion is for* $-\infty<\omega\le 0$ *while the solid (—) portion is for* $0\le\omega<\infty$. *The direction of the arrow is in the direction of increasing* ω.

simply the mirror image (about the real axis) of the portion from 0 to ∞, not using the full plot can lead to erroneous conclusions.

- The precise meaning of the commonly used notion of "encirclement" about the (-1,0) point must be understood. It is not uncommon[6,9] to have cases where the (-1,0) point is entirely within Nyquist plot and hence appears "encircled," but the net encirclement is, in fact, zero. Further, the direction of encirclement is crucial. Encirclement in itself does not necessarily mean that the closed-loop system is unstable.
- The number of RHP poles of $G_L(s)$ must be known.

We demonstrate the above points by choosing a simple system—the same one we chose in the previous section

$$G_o(s)=\frac{2}{s-1} \tag{19}$$

in a feedback loop with a proportional controller of gain $K_{c1}=1/4$ and $K_{c2}=1$. It is easy to see that the first control system is unstable, while the second is stable, by considering the characteristic equations $1+G_{L1}(s)=0$ and $1+G_{L2}(s)=0$, respectively. But our objective here is in the application of the Nyquist criterion.

Figure 2a shows the Nyquist plot of

$$G_{L1}(s)=\frac{2K_{C1}}{s-1}=\frac{1}{2(s-1)} \tag{20}$$

The figure clearly shows that N=0 as the net angle traced out by the full Nyquist plot (with reference to the (-1,0) point) is zero. Since $P_R=1$, we get

$$Z_R=P_R-N=1-0=1 \tag{21}$$

Thus the closed-loop system is unstable with one root of the characteristic equation in the RHP. Note here that even though the Nyquist plot does not encircle the (-1,0) point, the closed-loop system is unstable.

Figure 2b shows the Nyquist plot for

$$G_{L2}(s)=\frac{2}{s-1}K_{C2}=\frac{2}{s-1} \tag{22}$$

Here the Nyquist plot encircles (-1,0) once. Note that the net angle traced is 2π, but this is in the counterclockwise direction, implying that N=1. Again, since $P_R=1$, we obtain

$$Z_R=P_R-N=0 \tag{23}$$

Thus the closed-loop system is stable, even though the Nyquist plot encircles the (-1,0) point. Note further that if we restrict ourselves to the 0 to ∞ segment, we will not see any encirclement.

Thus, we have highlighted the aspects we set out to illustrate—the importance of considering the entire frequency range (-∞ to ∞), the importance of the direction of encirclement, and the necessity of knowing the number of unstable poles of $G_L(s)$.

CONCLUSIONS

We have clarified the concept of frequency response for linear time-invariant systems, demonstrating its validity for unstable systems as well. We have also highlighted some pitfalls in the use of the Nyquist criterion and pointed out how to avoid them.

REFERENCES

1. Coughanowr, D.R., *Process Systems Analysis and Control,* McGraw-Hill Book Company, New York, NY (1991)
2. Luyben, W.L., *Process Modeling, Simulation, and Control for Chemical Engineers,* McGraw-Hill Book Company, New York, NY (1990)
3. Marlin, T.E., *Process Control: Designing Processes and Control Systems for Dynamic Performance,* McGraw-Hill Book Company, New York, NY (1995)
4. Ogunnaike, B.A., and W.H. Ray, *Process Dynamics, Modelling, and Control,* Oxford University Press, New York, NY (1994)
5. Seborg, D.E., T.F. Edgar, and D.A. Mellichamp, *Process Dynamics and Control,* John Wiley & Sons Inc., New York, NY (1989)
6. Stephanopoulos, G., *Chemical Process Control,* Prentice-Hall, Englewood Cliffs, NJ (1984)
7. Brockett, R.W., *Finite Dimensional Linear Systems,* John Wiley & Sons, Inc., New York, NY (1970)
8. Wolovich, W.A., *Automatic Control Systems: Basic Analysis and Design,* Harcourt Brace (1994)
9. D'Azzo, J.J., and C.H. Houpis, *Feedback Control System Analysis and Synthesis,* McGraw Hill Book Company, New York, NY (1966) ❒