A Student's Guide to Classical Control

by

Dennis S. Bernstein[1]

Automatic control is a "miracle" technology that exploits **feedback** to improve the performance of a tremendous range of technological systems from the steam engine to the space station. Watt's governor tamed the steam engine and made the Industrial Revolution possible. However, feedback is used by everyone: if the shower water is too hot, open up the cold water faucet. Today, feedback control is used in radios (amplifiers), CD players (laser tracking), automobiles (cruise control, engines and suspensions), flight control (autopilots and stability augmentation), spacecraft (attitude control and guidance), machine tools, robots, power plants, materials processing, and many other applications. In many cases automatic control is an *enabling technology* since the system cannot operate without it. Feedback is also used to regulate virtually every system in the human body and is constantly at work in ecological systems (but let's not go too far afield here). At first sight, control engineering looks simple and straightforward. It is not. While automatic control is a powerful technology, the subject is amazingly subtle and intricate in both theory and practice. This subtlety is mainly because changes cannot be effected instantaneously in a dynamical system--when the shower is too hot to the touch, it is already too late to shut down the hot water. An otherwise correct control decision applied at the wrong time could result in catastrophe. Accounting for this and numerous other effects is what control engineering is about.

Who should read on?

Because automatic control is such an intricate subject it is extremely easy for students to miss the forest for the trees. This guide is intended to be of help to both undergraduate and graduate students. Undergraduates may skim this guide at the beginning of a first course on control to get a rough road map of the subject and terminology and later refer to it periodically during the course. At the end of the course these notes can be useful for reviewing the course while studying for the final exam. For graduate students embarking on a course in modern control, these notes can provide a quick review at the beginning of the course to help place their prior knowledge in perspective.

Lessons

1. Feedback is pervasive.

The interaction of any pair of systems almost always involves feedback. System 1 reacts to System 2 and vice versa. It is **cascade** (one way interaction) that is the exceptional case. Newton's third law is a statement of feedback: The force applied to A by B is counteracted by the force applied to B by A. When two electrical circuits are wired together, each affects the other, and Kirchoff's laws determine what that interaction is.

2. Block diagrams are not circuit diagrams.

Block diagrams are helpful for analyzing feedback. However, block diagrams should not be confused with circuit diagrams. It is useful to be able to translate circuit diagrams into block diagrams. This translation requires representing impedances and admittances as systems with inputs and outputs that are voltages and currents. Remember that the lines connecting the boxes are not wires. Analogously, it is useful to be able to represent mechanical systems by block diagrams with forces, positions, velocities, and accelerations as inputs and outputs and with the blocks representing kinematics and dynamics.

---

[1] Aerospace Engineering Department, The University of Michigan, Ann Arbor, MI 48109, dsbaero@umich.edu.

3. Determine the equilibrium points and linearize.

An **equilibrium point** is a state in which the system would remain if it were unperturbed by external disturbances. An equilibrium point can be **unstable** (an egg standing on its end), **neutrally stable** (a mass connected to a spring), or **stable** (a book lying on a table). For a system under feedback control, an equilibrium point is called an **operating point.**

Real systems are **nonlinear**. However, a **linearized model** can be used to approximate a nonlinear system near an equilibrium point of the nonlinear system by a procedure called **linearization**. The resulting **linear system** has an equilibrium point at zero that corresponds to the equilibrium point of the nonlinear system. While linearized models are only an approximation to the nonlinear system, they are convenient to analyze and they give considerable insight into the behavior of the nonlinear system near the equilibrium point. For example, if the zero equilibrium point of the linear system is stable, then the equilibrium of the nonlinear system is **locally stable**. Unless stated otherwise, we will consider only the approximate linearized model of the system.

4. Check stability first.

Once a linearized model has been obtained, stability must be checked. Stability can be asked of any system, whether the system is uncontrolled (with the control turned off) or the system is in closed-loop operation (with the control determined by feedback). If the system is represented by a **transfer function**, then the roots of the denominator polynomial (called the **poles** of the system) determine whether the equilibrium or **operating point** is stable. Stability can be tested *explicitly* by computing the roots of this polynomial or *implicitly* by using the **Routh criterion.** If the system is in **state space form,** then the eigenvalues of the **dynamics matrix** can be computed as an explicit test of stability. Stability can also be tested by forming the characteristic polynomial of the dynamics matrix and applying the Routh criterion. However, this approach is inconvenient if the system is of high order.

5. Stable systems have a frequency response.

Much of control-system analysis involves the **frequency response** of linear systems. The meaning of the frequency response can be understood by keeping in mind the **fundamental theorem of linear systems:** Suppose that a sinusoidal (or **harmonic**) input (such as a forcing) with frequency ω is applied to a stable linear system $G(s)$. Then the output of the system approaches a sinusoidal motion whose frequency is the same as the frequency of the input. This limiting sinusoidal motion is called the **harmonic steady-state response.** Note, however, that the output of the system does not have a limit in the usual mathematical sense since the harmonic steady-state response does not approach a constant value. The transient behavior of the system before its response reaches harmonic steady state depends on the poles and **zeros** of the system as well as on the initial conditions of the internal states. The ratio of the amplitude of the harmonic steady-state response to the amplitude of the sinusoidal input is equal to the absolute value or **gain** of the transfer function evaluated at the input frequency ω, that is, $|G(j\omega)|$, while the phase shift or **phase** of the harmonic steady-state response relative to the phase of the input is given by the phase angle of the transfer function evaluated at the input frequency, that is, $\tan^{-1}(\mathrm{Im}\ G(j\omega)/\mathrm{Re}\ G(j\omega))$. **Bode plots** show the gain and phase of $G(j\omega)$ for each frequency ω in a range of frequencies. Although the fundamental theorem of linear systems does not apply to neutrally stable and unstable systems, Bode plots can be drawn for these systems as well, although the gain may be infinite at certain frequencies. For convenience we refer to these plots as the "frequency response" even when the system is not stable.

6. Remember the loop transfer function.

Breaking the loop of a closed-loop system reveals the **loop transfer function**, which is the product of all of the transfer functions in the loop, including the plant, the sensor, the controller and the actuator. Always be aware of whether the loop transfer function of a closed-loop system is stable or not, and be sure to note whether the closed-loop system involves positive or negative feedback. The gain and phase of the loop transfer function at a given frequency are called the **loop gain** and **loop phase**, respectively, and these quantities are defined *whether or not the loop transfer function is stable*. Note that if the loop transfer function is stable then it does not necessarily follow that the corresponding closed-loop system is stable, and that if the loop transfer function is unstable then it does not necessarily follow that the corresponding closed-loop system is unstable. Stability of a closed-loop system can be tested *directly* by applying the Routh criterion to the closed-loop system or by computing the poles of the closed-loop transfer function. In the case of negative feedback, the closed-loop transfer function involves the **sensitivity function** which is *one over one plus the loop transfer function.*

While the characteristics of the loop transfer function (such as its frequency response and its poles and zeros) may be well understood, the closed-loop system is relatively complicated. Thus it is much more convenient to analyze stability *indirectly* in terms of the loop transfer function. The **Nyquist criterion** and **root locus procedure** allow you to do just that.

While it is intuitively clear from the form of the sensitivity function that the frequency response of the loop transfer function of a stable system must never be equal to -1 if the closed-loop system is to be stable, the **Nyquist criterion** shows precisely how the frequency response of the loop transfer function must avoid this value. Specifically, the Nyquist criterion says that the closed-loop system with negative feedback is stable if and only if the polar plot of the loop transfer function, with its argument following a contour that includes the imaginary axis, avoids poles of the loop transfer function on the imaginary axis, and encloses the right half plane, encircles the **critical point -1+j0** as many times counterclockwise as there are unstable (open right-half plane) poles in the loop transfer function, whether these unstable poles arise from the plant or from the controller. Note that the Nyquist plot can only encircle the critical point -1+j0 if the loop gain is greater than unity in some frequency range. Thus stabilization imposes a minimal requirement on the loop gain in certain frequency ranges. This requirement thus imposes a constraint on the gain of the controller that depends upon the gain of the plant.

The **root locus procedure** shows the location of the closed-loop poles for each value of a *constant feedback gain.* The closed-loop poles are located near the poles of the loop transfer function for small values of the feedback gain, and, as the feedback gain goes to infinity, converge to the locations of the loop transfer function zeros and move toward infinity along asymptotes. Hence a **nonminimum phase system**, that is, a system with at least one zero in the open right half plane, can be destabilized by large constant feedback gains. Furthermore, it can be seen from the form of the asymptotes that a system that has **relative degree** (pole-zero excess) of at least three can also be destabilized by large constant feedback gains.

7. After stability, performance is everything.

Once stability is settled, **performance** is everything. In fact, for plants that are open-loop stable, performance is the *only* reason for using feedback control. Basic criteria for performance include the ability to track or reject signals such as steps, ramps, sinusoids, and noise (random signals). Good performance generally requires high loop gain which corresponds to small gain of the sensitivity function and thus tracking error reduction or disturbance attenuation. Unfortunately, **Bode's integral theorem** tells us that for any real control system the sensitivity function cannot have gain less than unity at all frequencies.

8. Perfect performance is asymptotically possible.

In the extreme case, asymptotically perfect tracking or disturbance rejection can be obtained by means of infinite loop gain and thus zero sensitivity at the disturbance frequencies. This is the case when an **integral controller** is used to provide infinite gain at DC (zero frequency) and thus asymptotically track a step command or reject a step disturbance. The same idea can be used to asymptotically track or reject a sinusoidal signal by using infinite controller gain at the disturbance frequency (which we are assuming is known). Since the controller incorporates a model of the command or disturbance, its operation is a special case of the **internal model principle.** Don't forget that closed-loop stability must still be checked for this infinite-gain controller.

9. Assure nominal stability.

If the system is uncertain, as least assure nominal stability. Everything that has been said so far with regard to stability and performance is based upon the assumption that the linearized model of the plant is an accurate representation of the system near the equilibrium. While all real systems are nonlinear, it is convenient to view the linear approximation itself as possessing **modeling uncertainty**, so that the model we have been discussing is merely a **nominal model** of the plant. All of the stability tests, whether direct or indirect, can be applied to the nominal model to guarantee **nominal stability.**

10. What guarantees robust stability?

Robust stability refers to stability for all linear models in the range of the modeling uncertainty. Robust stability can be guaranteed in principle by applying any stability test to all possible models of the uncertain system. Unfortunately, this is hard work. However, the Nyquist criterion can be used to determine robust stability with a **frequency domain uncertainty model**, which measures the level of modeling uncertainty at each frequency. In general, uncertainty tends to be greater at higher frequencies, and loop gain is generally known better than loop phase.

11. The Nyquist criterion can determine robust stability.

The number of encirclements of the critical point -1+j0 by the Nyquist plot of the loop transfer function is the crucial frequency domain test for nominal stability. Once the number of encirclements is correct for nominal stability, the distance from the Nyquist plot of the loop transfer function to the critical point -1+j0 determines robust stability in terms of a frequency domain uncertainty model. This distance is the reciprocal of the gain of the sensitivity. Thus, small sensitivity at a given frequency corresponds to large distance from the Nyquist plot of the loop transfer function to the critical point -1+j0 and hence robust stability at that frequency. The **gain margin** measures robust stability for frequencies at which the phase of the loop transfer function is 180 degrees, which quantifies robust stability for a *pure gain perturbation* of the loop transfer function, while the **phase margin** quantifies robust stability for frequencies at which the loop gain is unity (these are the **crossover frequencies**), which quantifies robust stability for a *pure phase perturbation* of the loop transfer function. The Nyquist criterion and the root locus procedure can both be used to determine gain and phase margins. Note that a closed-loop system with nonminimum phase loop transfer function has limited gain margin (this was seen from the root locus procedure), while a closed-loop system with loop gain greater than unity at some frequency has limited phase margin (this follows from the Nyquist criterion and Lesson 15 below). Never forget that the distance from the Nyquist plot to the critical point -1+j0 is irrelevant if the number of encirclements is wrong, that is, if nominal stability does not hold. That is why indiscriminate use of large controller gains is not a viable control strategy.

12. Always conserve phase.

With the critical point sitting at -1+j0, the Nyquist criterion shows that any closed-loop system is never more than 180 degrees from instability at every crossover frequency. And 180 degrees is not a whole lot, especially at high frequencies where the plant is more difficult to accurately model and identify and thus the loop phase is more likely to be uncertain. Every degree of loop phase at crossover frequencies is precious and must be maintained by careful engineering. This issue especially arises in **digital control** where analog-to-digital (sampling) and digital-to-analog (zero-order-hold) devices can cause phase lag.

13. Beware of lightly damped poles.

Don't forget that every lightly damped pair of complex conjugate poles in the plant (or controller) entails high loop gain near the **resonance frequency** as well as a whopping 180 degrees of phase shift over a small frequency band. A **notch** in the controller can reduce the loop gain at the resonance frequency to help avoid an inadvertent change in the number of Nyquist encirclements due to modeling uncertainty. However, this technique will only work if you actually know the frequency of the resonance. Note that a notch in the controller entails some performance loss in the frequency band of the notch. An **antinotch** can be used to increase the loop gain and thus improve performance in a frequency band where the plant is well known.

14. High controller gain has lots of benefits.

If the controller gain is high and the plant gain is not too small, then the loop gain will be high, which implies small sensitivity and thus good tracking or good disturbance rejection. Also, assuming that nominal stability holds, high loop gain means large distance from the Nyquist plot of the loop transfer function to the critical point -1+j0 and thus some measure of robust stability. Thus it would seem that high gain gives both good performance *and* robust stability. However, there are (at least) three hidden catches to having the best of both worlds, namely, rolloff, saturation, and noise.

15. Practice safe rolloff.

Don't forget (and we can't stress this enough) that high loop gain is useless if the number of Nyquist encirclements is wrong for nominal stability. Thus, as already noted, indiscriminate use of high controller gain is not recommended. If nominal stability *does* hold, then there is still the problem of **rolloff**, since the gain of a real system, and thus the loop gain, *always goes to zero asymptotically at high frequency.* As the loop gain drops below unity at the crossover frequency, gain and phase margins must be maintained to provide adequate distance from the critical point -1+j0 in order to assure robust stability against modeling uncertainty. In general, as the frequency increases you will have to roll off the loop gain more quickly than the plant uncertainty grows. Also remember that achieving good rolloff isn't as easy as adding poles to the controller to roll off the loop gain since, as the loop gain rolls off, the loop phase **lags** (that is, the loop phase decreases) with 90 degrees of lag incurred at high frequency for each pole. Hence good rolloff requires that the loop gain decrease adequately *without* accumulating excessive loop-phase lag. Another Bode integral theorem shows that most of the loop-phase lag is due to the rolloff rate at the crossover frequency with steeper rolloff causing greater phase lag. **Lead-lag compensators** are useful for **shaping** the loop gain and loop phase to achieve high gain and safe rolloff.

16. Saturation can rob you.

As if that's not all the trouble high loop gain can get you into, high loop gain is useless if the actuators cannot deliver the specified control signal. If the controller asks for 4 Newtons and the actuator can deliver only 2 Newtons, then you have a serious problem. You can think of the Nyquist plot as being "squashed down" due to this **saturation** effect. Margins will be reduced, and, even worse, the Nyquist encirclement count can change and the closed-loop system will be unstable. It is extremely important to remember that the inability of the actuators to deliver the specified control signal is not just the fault of the controller gain being too high, but rather is also due to both the size of the plant gain *and the amplitude of the disturbance signal.* If the disturbance signal has large amplitude, then the actuator may saturate and you will have no choice but to reduce the gain of the controller and thus sacrifice some of the performance you want (and which was the reason for using feedback control in the first place!). Therefore, saturation can rob you of both stability and performance. Although we discussed saturation in terms of linear stability analysis, saturation is actually a special kind of nonlinearity.

17. High gain amplifies noise.

If your high gain controller survives the rolloff dragon and the saturation beast, it may yet succumb to the noise monster. While an integrator ("1/s") tends to smooth and attenuate noise, a differentiator ("s") tends to amplify noise. Every pole in a transfer function is "like" an integrator, while every zero is "like" a differentiator. As the plant gain rolls off, you may wish to include zeros in your controller in order to increase the loop gain for better performance while adding phase lead to the loop transfer function in order to increase your phase margin for robust stability. Zeros will do both of these things quite nicely for you. However, the resulting high controller gain will now amplify noise in the measurements and this amplification may outweigh the performance and stability benefits of the high loop gain and loop-phase lead. This amplification of noise was to be expected since the zeros, after all, act like differentiators.

18. Time delays can be deadly.

A **time delay** can be deadly--think of the Trojan horse or the AIDS virus. A time delay in the feedback loop corresponds to a transfer function that has unit magnitude at all frequencies as well as phase lag that increases linearly with frequency. This phase lag will warp the Nyquist plot, especially at high frequencies. Thus, if a closed-loop system has an unmodeled time delay in the loop, then the number of Nyquist encirclements by the loop transfer function of the actual system may be different from what you expect, and the closed-loop system may be unstable. If you know about the delay in advance, then you may be able to counteract the effect of the delay through careful loop-shaping design. However, although you can (at least in theory) design stabilizing controllers in the presence of even large time delays, the closed-loop performance will most likely be poor. Imagine trying to make decisions using old information.

19. "Respect the unstable."

As Gunter Stein discussed in his classic 1989 Bode lecture (watch the video!), controlling *unstable* systems can be a dangerous undertaking. Real unstable plants (except for rigid body motion) are always nonlinear, and the ability to stabilize them requires a minimal amount of actuator bandwidth and stroke. A disturbance may perturb the state of a nonlinear system outside of its **domain of attraction** so that the actuators are unable to move the state back to the equilibrium and thus recover from the disturbance. In addition, actuator failure can lead to disaster and thus cannot be tolerated.

20. Multi-loop control is nontrivial.

Multi-loop control is much more challenging than single-loop control. Everything that has been said so far applies to **single-loop** control. Multiple control loops are needed whenever a plant has multiple sensors or actuators. In this case the interaction of every feedback loop with every other feedback loop must be accounted for. While many single-loop concepts hold in principle in the multi-loop case, the technicalities are much more involved. The performance benefits of multi-loop control, however, are often far more than one would expect from a collection of single-loop controllers.

21. Nonlinearities are always present.

Almost everything that has been said so far applies to linearized plant models. Real systems, however, have all kinds of *nonlinearities*: deadband, backlash, Coulomb friction, hysteresis, quantization, saturation, kinematic nonlinearities, and many others. Thus a controller designed for a linear plant model to satisfy performance specifications may perform poorly when applied to the actual plant! The tradeoff here is between mathematical tractability of the linearized model and greater validity of a nonlinear model.

22. People's lives may be at stake.

Control-system engineers must account for all of these issues in designing and analyzing control systems that work. They must also specify, design, analyze, build, program, test, and maintain the electromechanical hardware, processors, and software needed to implement control systems. Real control systems must be extremely reliable, especially if people's lives depend on them. These are challenging and rewarding engineering tasks that will keep lots of control-system engineers busy for a long time to come.

What's next?

Now that you have completed the basic course in *classical control*, you are ready to enter the wonderland of **modern control.** The linear control you have learned, which has been limited to single-input, single-output systems, can now be expanded to multiple-input, multiple-output plants. **Multivariable control** is often studied with state space (differential equation) models and transfer function matrices. State space models provide the means for designing control systems that are optimal with respect to certain design criteria. **Optimal control** encompasses LQG and H control theory where explicit formulas are used to synthesize multivariable feedback controllers. **Robust control** seeks controllers that provide robust stability and performance for uncertain plants. The computer implementation of these controllers is the subject of **digital control.** While the jump from single-variable control to optimal multivariable control is a major and important step, **nonlinear control** takes nonlinearities into account and shows how to design nonlinear controllers to obtain improvements over linear control. The next, and most exciting, leap is into the subject of **adaptive control**, where controllers learn and adapt in response to changes in the plant and disturbances.

With amazing advances in sensing, actuation, and processing as well as a better understanding of learning and adaptation (for example, using neural computers), automatic control will become *the* technology of the next millennium. You are definitely in the right place at the right time!

Bibliography

You can read about the fascinating history of automatic control in

O. Mayr, *The Origins of Feedback Control*, MIT Press, Cambridge, 1970.

S. Bennett, *A History of Control Engineering 1800-1930*, Peter Peregrinus Ltd., London, 1979.

S. Bennett, *A History of Control Engineering 1930-1955*, Peter Peregrinus Ltd., London, 1993.

To get a glimpse of the future and what hath Watt wrought, see the exciting book

K. Kelly, *Out of Control*, Addison-Wesley, Reading, 1994.

Acknowledgments

I would like to thank an anonymous reviewer and Steve Yurkovich for numerous helpful suggestions. I especially thank all of my students, graduate and undergraduate, who surely taught me far more than I taught them. This guide is their legacy.

About the Author

I discovered the field of control engineering entirely by accident when I was an undergraduate at Brown University majoring in applied mathematics. I came across a sale-table copy of a book entitled *Some Vistas of Modern Mathematics* by Richard Bellman (University of Kentucky Press, 1968) (one of the great control theorists) and I became instantly fascinated by the subject. To pursue this interest, I attended the University of Michigan, which had an interdisciplinary program combining systems theory with all branches of engineering. I was fortunate to do my Ph.D. research under Elmer Gilbert, one of the pioneers of modern control. I eventually got a job in the aerospace industry at Harris Corporation where the goal of my research group was to use control technology to reduce vibrations in large deployable space structures, such as dish antennas with surfaces knitted out of wire mesh.

Currently I teach and do research in an aerospace engineering department. I have found that the most satisfying aspect of control is the interaction with other disciplines such as structures, fluids, combustion and propulsion, where I can apply and try out new control methods. All of the graduate students who work with me do theoretical research and apply their results to hardware experiments. Every experiment is exciting to build and operate, and, by implementing controllers on real hardware, we learn from Mother Nature (the ultimate judge) what does and does not work. For more about this story, see "Four and a Half Control Experiments and What I learned from Them" in the Proceedings of the 1997 American Control Conference.

I am completely bullish on controls as a key 21st century technology that has barely begun to reach its full potential. With new sensors and actuators and with much faster processors, the prospects for applying innovative control techniques are unlimited. As for control theory, I believe that the future lies in adaptive control. By weaning control from modeling we can fulfill the original promise of feedback, namely, to use our eyes, mind, and hands to make up for our uncertainty about the real world. First we must teach controllers (and the machines, systems, and processes they are part of) how to learn as we humans routinely do.