

What Makes Some Control Problems Hard?

By Dennis S. Bernstein

Success or Failure?

Control engineering has application to virtually all other branches of engineering, yet it is independent of any particular technology. There are numerous control engineering successes, including applications in electrical, mechanical, chemical, and aerospace engineering, in which control technology has significantly enhanced the value of products, vehicles, and processes. In the most outstanding cases, such as high-performance flight control, the contribution of control technology is of such critical importance that the system simply cannot operate without it. In other cases, such as in the process industries, control technology enhances the performance of the system, rendering it more competitive and profitable.

On the other hand, there are emerging applications of control technology for which successful implementation to the point of feasibility and profitability has not yet occurred (see, for example, [1]-[3]). With these and other applications in mind, the purpose of this article is to explore those aspects of control engineering that stand as impediments to successful control application. In a nutshell, I will ask: What makes some control problems hard?

I explore this question in five phases: *control strategy*, *control physics*, *control architecture*, *control hardware*, and *control tuning*. The reader will quickly surmise that these phases are deeply linked and that aspects of each one can have a strong impact on the others. In practice, these phases are visited iteratively. Nevertheless, I submit that in each phase there are reasonably distinct issues that can render a control problem hard and that are worthy of careful and individual consideration.

The goal of this article is to suggest a framework for viewing control applications that will help the control practitioner understand and articulate the nature of the engineering challenge. My hope is that this framework will provide a useful guide to approaching new control applications while increasing the chances of success.

A related goal is to provide terminology and concepts to help the control engineer communicate and collaborate with specialists from other disciplines. Control engineers recognize that the success of control technology often depends on system-level concepts such as dimensionality, nonlinearity, authority, and uncertainty. The ability to communicate these concepts to engineers who specialize in domain-specific technology can be valuable, if not essential, for ensuring the success of an engineering project.

Some background for this article is provided by [4]-[6], which this article complements and extends by providing a broader perspective on control engineering.

Control Strategy

Control strategy is the initial phase of a control engineering project in which the need for control is assessed. The use of control technology in many engineering projects must be justified in terms of cost, performance, and risk. Therefore, it is first necessary to exhaust other approaches that provide alternatives to the use of control. For example, it might suffice to upgrade the system in terms of component specifications and tolerances. However, these upgrades may not be technically feasible, or they may simply be too expensive.

When system upgrade is not feasible, system redesign may provide a solution. Such a redesign may entail mechanical and electrical modifications, which again may be infeasible or expensive. If the engineering design has been done with care, such modifications will have been exhausted in the preliminary project stages.

A special case of system redesign involves *passive control*, which entails mechanical or electrical subsystems that effectively perform a control function. In passive control, all sensing and actuation functions are integrated within the subsystem, and independent energy sources are usually not needed. Classical examples of passive control include the gain-setting circuit for an amplifier (Fig. 1) and the tuned mass absorber (Fig. 2). These implementations can be viewed and analyzed as feedback controllers (Figs. 3 and 4),

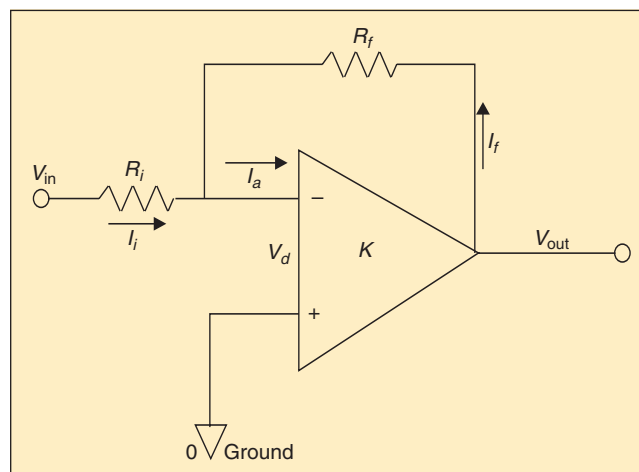


Figure 1. The feedback circuit for an operational amplifier is a classical example of a passive feedback controller. The feedback resistance R_f is used to set the amplifier gain.

The author (dsbaero@umich.edu) is with the Aerospace Engineering Department, University of Michigan, Ann Arbor, MI, U.S.A.

Although I have spent virtually my entire professional career thinking about control engineering, I continue to be amazed at the subtlety and complexity of the subject. This article is an attempt to bring together diverse, interrelated ideas that underlie control engineering challenges. However, I admit that I have barely scratched the surface of this intricate subject. In any event, it is useful to keep in mind that feedback is the bidirectional interaction of interacting components, and it is this interaction that has powerful implications. This should not be surprising. If you dismantled your car piece by piece and put all the parts in a pile, it wouldn't function as a vehicle. But if you assemble its thousands of parts just right, it will run fine. The same applies to you as a person, with even more interacting components. The study of control is the quintessential systems discipline, where our goal is to understand how systems composed of interacting components are able to function. This is a subject without boundaries, with much to offer to virtually all other scientific and engineering disciplines.

— DSB

although they do not possess distinct components for sensing and actuation.

Beyond passive control we can consider control strategies that require distinct components, including *actuators* and, in the case of feedback control, *sensors*. We now turn our attention to the physical basis for sensing and actuation.

Control Physics

A critical aspect of control engineering applications is the *control physics*, which refers to the physical phenomena that determine the ability to influence the plant dynamics. The control physics provides the physical basis for the design and construction of actuators for applying control inputs [7]. For example, forces and moments can be produced by means of an electric motor, in which a current interacts with a magnetic field to produce a torque (Fig. 5(a)), and the internal combustion engine, in which the burning of fuel inside a cylinder produces an increase in pressure (Fig. 5(b)). As devices for producing forces and moments, the motor and internal combustion engine can be viewed as actuators. These devices are based on *conventional control physics*.

Since the effectiveness of a control strategy depends on the control physics, it is important to determine limiting factors. One such factor is the physically allowable range of variables. In practice, there are always constraints on stroke, maximum current, and virtually all physical vari-

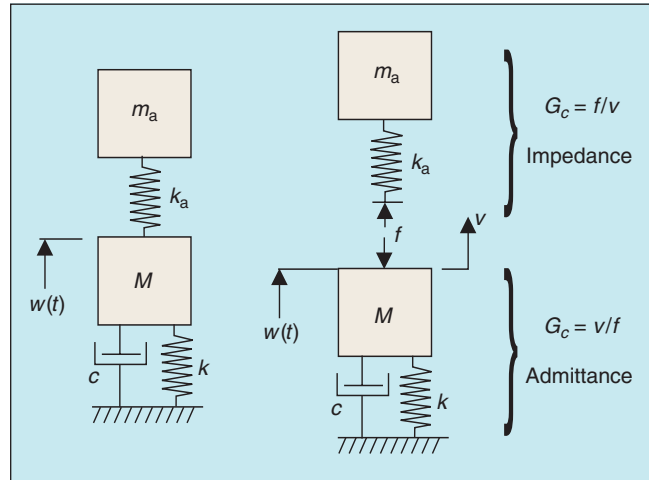


Figure 2. The tuned mass absorber is a passive feedback controller with integrated sensing and actuation. This controller is often implemented as part of plant redesign for vibration suppression.

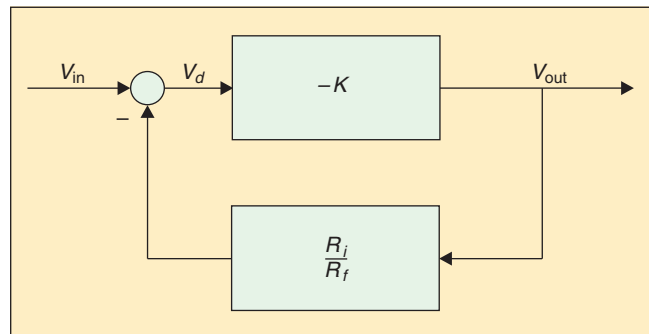


Figure 3. This block diagram representation of the amplifier circuit in Fig. 1 shows its function as a feedback controller. Since $V_{out} / V_{in} \gg R_f / R_i$ when $R_f \ll KR_i$, the feedback resistance R_f can be chosen to set the amplifier gain to a high degree of precision despite uncertainty in the gain K of the operational amplifier.

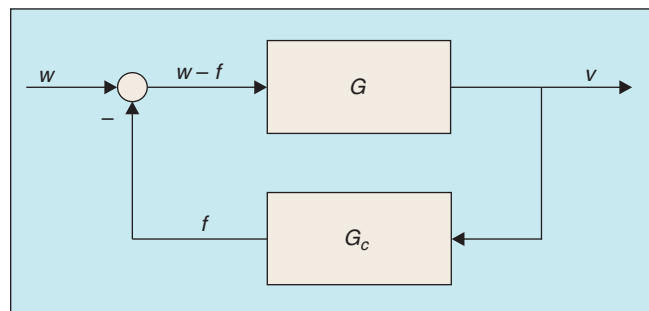


Figure 4. This block diagram representation of the tuned mass absorber in Fig. 2 illustrates its function as a feedback controller. Specifically, the tuned mass absorber provides narrow-band disturbance rejection by means of the feedback controller $G_c(s) = 1 / (m_a s^2 + k_a)$. By appropriately choosing the mass and stiffness values m_a and k_a , this internal model controller can asymptotically reject a sinusoidal disturbance of known frequency $w_d = \sqrt{k_a / m_a}$ but unknown amplitude and phase.

ables. These constraints arise from inherent physics as well as design details.

The laws of physics, such as conservation of energy and momentum, as well as the laws of thermodynamics, represent limitations to all control physics. Because of conservation of energy, every physical device must satisfy an energy flow constraint, which means that the power (rate of energy) output of any device cannot exceed the total rate of energy input available from all internal and external energy sources. Note that this power constraint does not entail separate constraints on force and velocity, which (ignoring the

range constraints mentioned earlier) can independently be made arbitrarily large (for example, by gearing) without violating the laws of physics; it is their product that is constrained by the available power. Analogous remarks apply to current and voltage.

In addition to input-output power constraints, the achievable power output of any device is limited by its efficiency, which reflects the loss of energy due to friction and resistance. Energy is also lost when it is converted from one form to another, such as when a current interacts with a magnetic field to drive a motor. This *transduction efficiency* is a material property. All energy loss is ultimately thermal, taking the form of heat dissipation and governed by the laws of thermodynamics.

Exotic control physics, which refers to control physics that is not as widely used in commercial applications as conventional control physics, is often based on material transduction. Examples include piezoelectric materials, which exhibit strain in response to an electric field (Fig. 5(c)), as well as shape memory alloys, which undergo large strain at critical transition temperatures (Fig. 5(d)).

Control technology is most successful when it can exploit *high-leverage control physics*. This refers to applications where the control system takes advantage of existing conditions to effect large amplitude control inputs with minimal cost and effort. The classical example is the triode vacuum tube [8], which is used to increase the amplitude of an information-laden, time-dependent signal by extracting and converting energy from an informationless dc source (Fig. 5(e)). Another example is aircraft control systems, where the engine produces thrust for forward velocity and, by means of the wings, lift. Stability and attitude control are obtained inexpensively by the small motion of aerodynamic surfaces, which produce moments by interaction with the airflow (Fig. 5(f)).

In many applications, such as servo control of mechanical systems, the leverage of the control physics is not a driving issue. What is most important in these applications is the performance of the system, and expensive, high-power devices can usually be justified to achieve performance objectives. On the other hand, in other applications, such as flow control for drag reduction, the leverage of the control physics is crucial; in fact, if a drag reduction flow control system is “only” 100% efficient, then it is effectively worthless, since the applied power could presumably be applied equally well to the vehicle’s engines. Consequently, it is the leverage of the control physics that often determines the viability of new control technology and applications.

Control Architecture

Control architecture refers to the spatial arrangement of sensors and actuators as well as their interconnection with performance, disturbance, and command signals through the plant and controller. To understand the implications of vari-

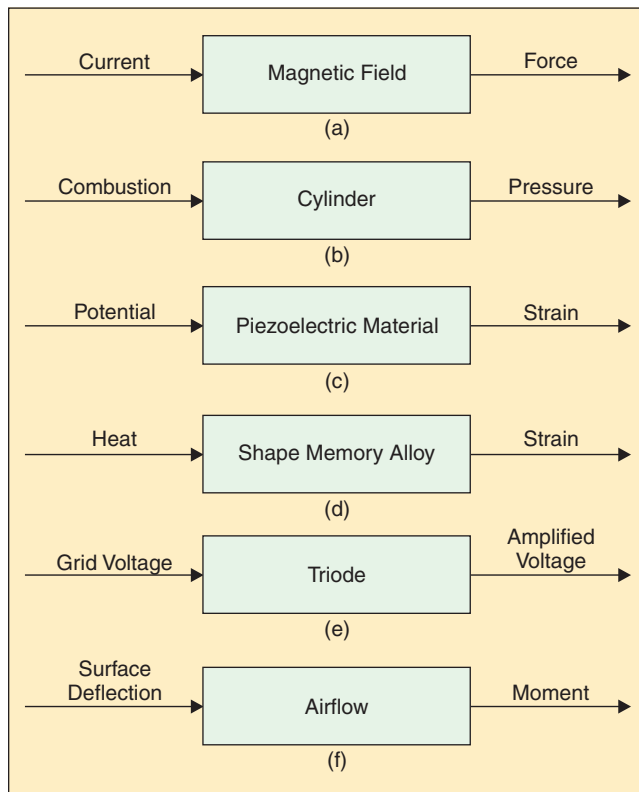


Figure 5. (a) An electric motor is the most widespread example of conventional control physics. In this case, the interaction of a current with a magnetic field results in a force or torque. (b) Another example of conventional control physics is the internal combustion engine, which applies force to a vehicle by burning fuel inside a cylinder; which causes a gas to expand with a resulting pressure increase. (c) As an example of exotic control physics, a piezoelectric material undergoes strain when subjected to an electric field. (d) Another example of exotic control physics is a shape memory alloy, which undergoes large strain upon reaching critical temperatures. (e) The triode is a classical example of high-leverage control physics. By converting energy from a dc source (such as a battery) to a time-dependent signal (such as a demodulated radio signal), it forms the basis for the feedback amplifier. (f) The aerodynamic surface is another example of high-leverage control physics. Control moments are obtained inexpensively by the motion of aerodynamic surfaces, which produce moments by interaction with the ambient airflow produced by an engine whose primary function is to provide thrust for lift.

ous control architectures, it is helpful to consider the *general control architecture* shown in Fig. 6. The general control architecture provides the framework for the *standard problem*; see [9].

The general control architecture involves four (possibly vector) signals of fundamental significance in any control problem (see Fig. 6). These are the *exogenous input* w , which can represent disturbances or commands, the *controlled input* u , which is provided by actuators, the *performance* z , which determines the performance of the system, and the *measurement* y , which provides the input to the feedback controller. Note that z need not be physically measured, although there may be overlap between the signals that comprise z and the measured signals that comprise y so that z may be partly or fully measured.

Using transfer function notation (although these transfer functions can be replaced by nonlinear systems or operators), the input-output equations for the general control architecture have the form

$$\begin{aligned} z &= G_{zw}w + G_{zu}u, \\ y &= G_{yw}w + G_{yu}u, \\ u &= G_c y. \end{aligned}$$

These signals may be scalar or vector, continuous or discrete time, and the relationships that link them may be linear or nonlinear, time invariant or time varying. In addition, the feedback control algorithm may be linear or nonlinear, fixed gain or adaptive. The general control architecture includes all of the signal types that occur in control problems, and it can be used to represent stabilization, command following, tracking (command following with advance command signal knowledge), and disturbance rejection problems.

In control problems involving input commands, it is sometimes sufficient to consider an *open-loop* control architecture in which the command or disturbance measurement is processed by means of *feedforward* control before plant actuation and without employing feedback sensors (Fig. 7 with $G_{fb} = 0$). Feedforward control is often effective in command-following and tracking problems where shaping filters are used to modify or deadbeat the plant input [10], [11]; it is also commonly used in disturbance rejection problems when a measurement of the disturbance is available [12]. The feedforward control architecture is a special case of the general control architecture, as shown in Fig. 8 with $G_{fb} = 0$.

In many applications, open-loop control is an adequate architecture. Open-loop control requires actuators, but it does not require sensors. When the system possesses a high level of uncertainty in terms of its dynamics and disturbances, it may be necessary to use a *feedback* or *closed-loop* control architecture. Feedback control is significantly more sophisticated than open-loop control since it requires sensors, actuators, and processing components in closed-loop

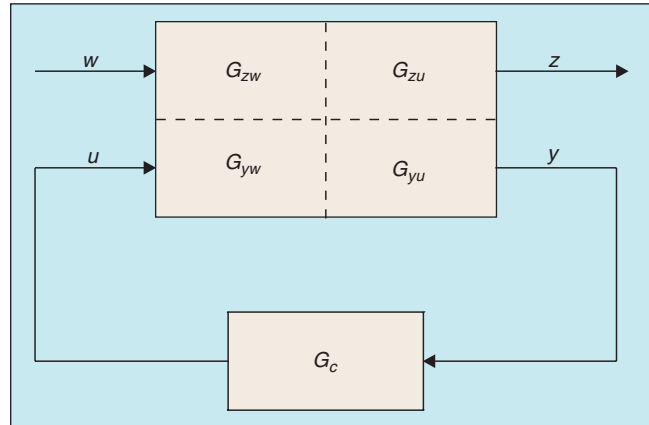


Figure 6. The general control architecture involves four different kinds of signals: exogenous input w , controlled input u , performance z , and measurement y . The exogenous input w can represent a command or disturbance signal. Special cases of the general control architecture include virtually all control problems of interest.

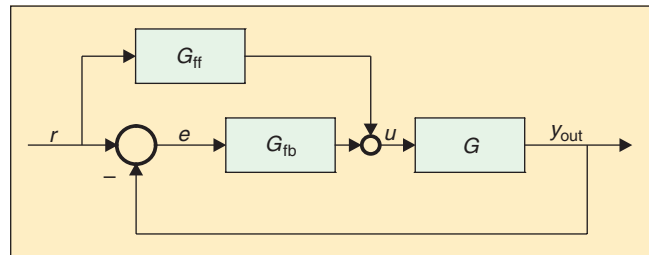


Figure 7. Feedback control requires a more sophisticated, costly, and risky control architecture than open-loop control involving sensing, actuation, and processing components in closed-loop operation. In practice, feedback control must be justified over open-loop control, and this justification is usually based on the presence of uncertainty in plant dynamics or exogenous signals. This block-diagram representation of the classical command-following problem involves feedback and feedforward gains G_{ff} and G_{fb} . Note that $e = (1 + GG_{fb})^{-1}(1 - GG_{ff})r$. When $G_{fb} = 0$, this control architecture specializes to the open-loop control architecture for feedforward control.

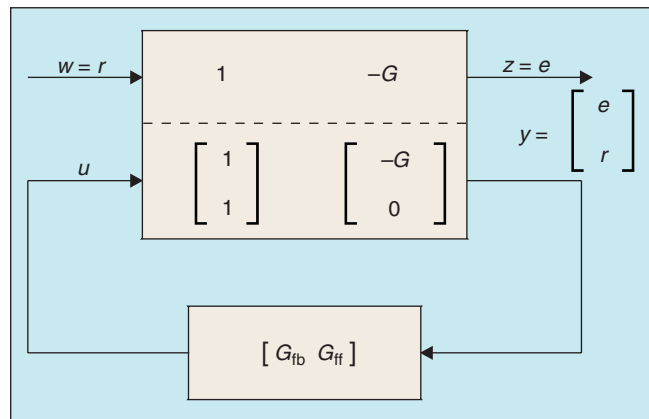


Figure 8. This is the classical command-following architecture shown in Fig. 7 recast in the general control architecture framework.

operation. In practice, feedback must be justified in view of the cost of the necessary components and the risk due to potential component failure. The classical command-following problem is shown in Fig. 7 and recast in the general control architecture in Fig. 8. Note that, although passive control involves feedback, distinct sensors and actuators are not needed, and thus the implementation of passive controllers is generally less expensive and more reliable.

The general control architecture can also be used to capture adaptive control algorithms. Adaptive controllers often

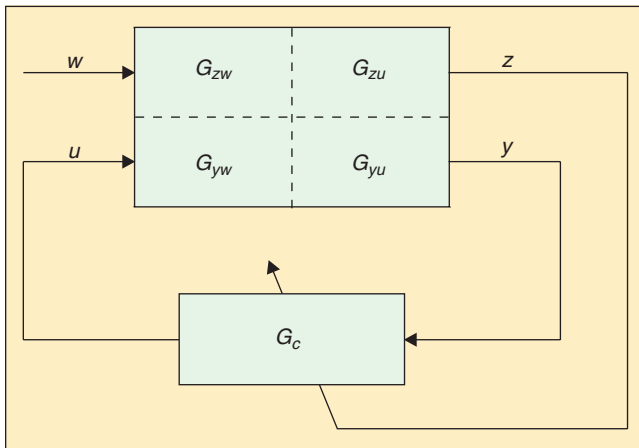


Figure 9. This adaptive control architecture involves an instantaneously (frozen-time) linear feedback controller adapted by using measurements of the performance z . This architecture can be recast in terms of the general control architecture shown in Fig. 6.

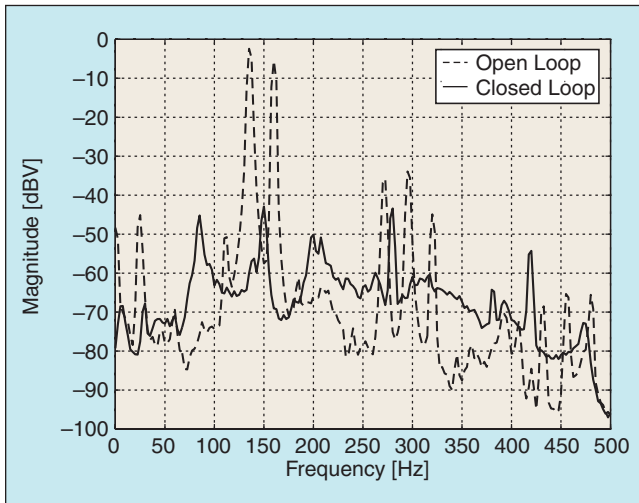


Figure 10. This frequency response plot compares the open- and closed-loop performance of an active noise control system with a dual-tone disturbance. The control algorithm is based on the adaptive controller architecture shown in Fig. 9, in which an instantaneously (frozen-time) linear feedback controller is updated by using a measurement of the performance z . The adaptive control algorithm has knowledge of only the zeros of the transfer function G_{zu} and has no knowledge of the disturbance spectrum. For details, see [13].

require that the performance z be measured so that the measured signals that comprise y include those that comprise z . A special class of adaptive controllers involves instantaneously (frozen-time) linear controllers that are periodically updated (that is, adapted) based on measurements of the performance. This adaptive control architecture is a special case of the general control architecture, which is illustrated by means of the modified general control architecture shown in Fig. 9. This adaptive architecture is used for adaptive disturbance rejection in applications such as active noise control (see [12], [13], and Fig. 10).

Control Architecture and Achievable Performance

The general control architecture can be helpful in understanding the effect of sensor and actuator placement on achievable performance. Consider, for example, a traditional home heating system, as shown in Fig. 11(a), where w denotes the outside temperature and weather conditions, u denotes the heat input from the furnace, z represents the performance variable (for example, temperature at a central location), and y represents a temperature measurement. To appreciate the advantages of the traditional heating system architecture, it is helpful to consider alternative architectures. For example, consider the case in which the control input u is moved outside of the house (see Fig. 11(b)). In this case, the heat output u of the furnace can directly counteract the cold and windy weather w , but the control physics has extremely low leverage since attempting to heat the outdoors is a hopeless task (global warming notwithstanding).

Next, consider an alternative heating system architecture in which the measurement y is colocated with the disturbance w outside the house (see Fig. 11(c)). In this case, there is an advantage in having a direct measurement of the disturbance, and it can be shown that achievable performance is improved. However, no measurement is available at the location where the performance is determined (recall that z is not a measurement per se); hence, this architecture may require a model of G_{zw} and G_{zu} and thus a more complex, possibly model-based (and thus model-dependent), control algorithm. In contrast to the alternative heating system architecture in Fig. 11(c), the traditional architecture in Fig. 11(a) employs colocated y and z , which permits model-free control with virtually no control tuning. Including z as an additional measurement, $y_2 = z$ in Fig. 11(d) increases the hardware cost but reduces model dependence. Hence the control architecture has implications for the leverage of the control physics, modeling requirements, control algorithm complexity, and achievable performance.

To further illustrate the implications of the control architecture, consider the classical command-following problem given by Fig. 7 with $G_{ff} = 0$. The error e is given by

$$e = Sr,$$

where

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

is the *sensitivity function* and

$$L = GG_{fb}$$

is the *loop transfer function*. Note that open-loop operation with $G_{fb} = 0$ always yields $e = r$. On the other hand, if $G_{fb} \neq 0$ and $|S(j\omega)| > 1$ holds for some value of ω , then $|e| > |r|$ and thus amplification occurs relative to open loop. Similarly, if $G_{fb} \neq 0$ and $|S(j\omega)| < 1$ holds for some ω , then $|e| < |r|$ and thus attenuation occurs relative to open loop. When amplification occurs, the system experiences *spillover*; that is, undesirable amplification relative to open-loop operation.

To understand when spillover can occur, it is useful to recall an important property of the sensitivity function, namely, that if S is stable and L has relative degree 2 or greater, then the *Bode sensitivity constraint* (see [14]-[17]) is given by

$$\int_0^\infty \log|S(j\omega)|d\omega = p \cdot \text{sum of the right-half-plane poles of } L.$$

To illustrate this result, consider the stable loop transfer function

$$L(s) = \frac{1}{s^2 + 2s + 3},$$

which yields the stable sensitivity function $S(s) = (s^2 + 2s + 3)/(s^2 + 2s + 4)$. Hence

$$\int_0^\infty \log|S(j\omega)|d\omega = 0,$$

which shows that the log sensitivity curve provides equal areas of amplification and attenuation (see Fig. 12).

As another example, consider the loop transfer function

$$L(s) = \frac{4}{(s - 1)(s + 2)},$$

which yields the stable sensitivity function $S(s) = 4(s^2 + s - 2)/(s^2 + s + 2)$. In this case

$$\int_0^\infty \log|S(j\omega)|d\omega = p,$$

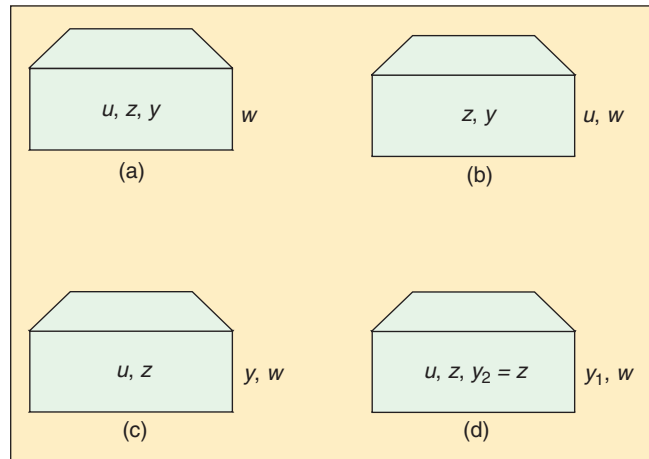


Figure 11. (a) In the traditional home heating system architecture, the control input u (heat), measurement y (temperature), and performance z (temperature) are colocated, and these signals are separated from the exogenous input w , which represents the outside weather disturbance. (b) In this alternative heating system architecture, the control input u is colocated with the disturbance w , resulting in extremely low-leverage physics. (c) In this alternative heating system architecture, the measurement y is colocated with the disturbance w , resulting in better achievable performance than the traditional architecture in (a). In this case, the indoor temperature z is not measured, necessitating greater reliance on plant modeling. (d) In this architecture, both the indoor temperature y_2 and the outdoor temperature y_1 are measured, providing measurements of both the performance variable z and disturbance w . This architecture yields better performance with reduced reliance on plant modeling as compared to (c), although at greater hardware expense.

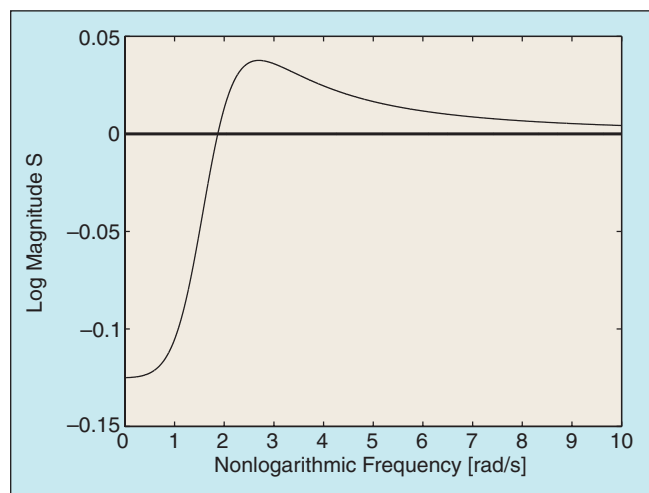


Figure 12. This plot of the sensitivity function for the stable loop transfer function $L(s) = 1/(s^2 + 2s + 3)$ shows attenuation at low frequency and amplification (spillover) above 2 rad/s. In accordance with the Bode integral constraint, the integral of $\log|S(j\omega)|$ is zero, and thus the log sensitivity function provides equal areas of attenuation and amplification.

which shows that the log sensitivity function provides a greater area of amplification than attenuation. In fact, for this loop transfer function, the sensitivity function indicates spillover at all frequencies (Fig. 13). It can also be shown that the peak magnitude of the sensitivity increases (so that the performance degrades) when the loop transfer function has right-half-plane zeros (see [18]).

In place of the classical command-following architecture, let us now return to the general control architecture and use the Bode integral constraint to investigate the effect of each of the four transfer functions on robustness and achievable performance. For simplicity, we assume that the system is linear time invariant and all signals are scalar. In this case, the performance z is given by

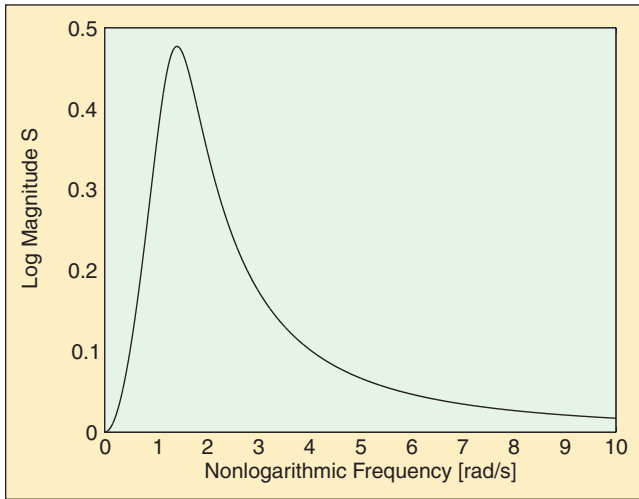


Figure 13. This plot of the sensitivity function for the unstable loop transfer function $L(s) = 4 / [(s - 1)(s + 2)]$ shows amplification (spillover) at all frequencies. In accordance with the Bode integral constraint, the integral of $\log |S(j\omega)|$ is π .

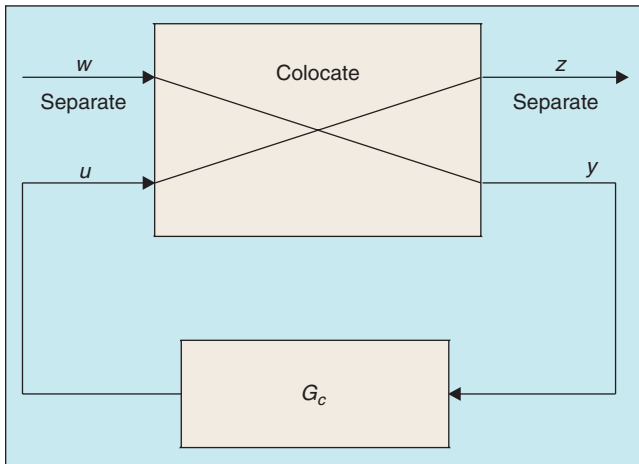


Figure 14. In the ideal control architecture for passive systems, the signal pairs (y, w) and (z, u) are colocated (as indicated by the diagonal lines), and the pairs (z, y) and (u, w) are separated.

$$z = \tilde{G}_{zw}w,$$

where the closed-loop transfer function \tilde{G}_{zw} is given by

$$\tilde{G}_{zw} = F(G_c)S,$$

the architecture function $F(G_c)$ is given by

$$F(G_c) = G_{zw} + (G_{zu}G_{yw} - G_{zw}G_{yu})G_c,$$

and the sensitivity function S now has the form

$$S = \frac{1}{1 - G_{yu}G_c}.$$

For the general control architecture, the Bode integral constraint provides insight into the achievable performance resulting from sensor and actuator placement. For example, suppose z and y are colocated, that is, $z = y$, which occurs in the command-following problem with $G_{if} = 0$. It then follows that

$$F(G_c) = G_{zw},$$

and thus

$$\tilde{G}_{zw} = G_{zw}S.$$

However, since $|S(j\omega)| > 1$ for some ω , it follows that

$$|\tilde{G}_{zw}(j\omega)| > |G_{zw}(j\omega)|$$

for some ω . Therefore, collocation of y and z causes spillover. Likewise, it can be shown that collocation of u and w causes spillover. Hence, the signal pairs (z, y) and (u, w) must be separated to avoid spillover. For further details, see [19]. Note that colocating y and z has the advantage that z is automatically measured.

As discussed earlier, the availability of a measurement of z reduces the dependence on models of G_{zw} and G_{zu} for controller tuning. On the other hand, if y and z are separated, then a measurement of z is available only when an additional sensor is implemented. For example, the home heating architecture in Fig. 11(d) requires two temperature sensors since $z = y_2$ is measured. While the use of a second sensor incurs greater cost, the achievable performance is enhanced relative to the traditional architecture in Fig. 11(a).

In addition to spillover ramifications, it follows from linear quadratic Gaussian (LQG) theory that if G_{zu} is minimum phase, then the regulation cost can be suppressed; likewise, if G_{yw} is minimum phase, then the observation cost can be suppressed; for details, see [20]-[22]. For passive systems

(that is, stable systems without energy-generating sources), a transfer function involving colocated input and output signals is positive real and thus minimum phase (see [19]). Note that colocating z and u entails physically placing the control input u at the location of the performance signal z , whereas colocating y and w is achieved by physically placing the measurement sensor y at the location of the disturbance signal w . If collocation is not possible, then it is desirable to place the sensors and actuators so that G_{zu} and G_{yw} are minimum phase.

Furthermore, placing y and u to avoid right-half-plane zeros in G_{yu} (for example, by colocating y and u) yields high gain margins. In fact, if G_{yu} is minimum phase, y is noise free, and there is no disturbance noise, then perfect state reconstruction is feasible, which thereby permits full-state feedback control and its associated advantage of unconditional stabilization, which is discussed later. However, if full-state feedback is not feasible but G_{yu} is minimum phase and its relative degree is not greater than two, with root locus asymptotes in the open left-half plane, then G_{yu} can be unconditionally (high-gain) stabilized; that is, the closed-loop system will have infinite gain margin. This property holds for multi-input, multi-output (MIMO) positive real systems, which are minimum phase and have relative degree 1, with strictly positive real feedback. In contrast, plants having either right-half-plane zeros or relative degree 3 or higher cannot be unconditionally stabilized. Hence, placement of u and y has implications for robustness.

These observations suggest that collocation of the pairs (y,w) and (z,u) along with separation of the pairs (z,y) and (u,w) represents the ideal control architecture for passive systems (see Fig. 14). This placement of sensors and actuators is typical for active noise and vibration control applications (Fig. 15). Experimental results (see [23] and Fig. 16) confirm that this arrangement of sensors and actuators can be effective in avoiding spillover.

Control Hardware

Now let's take a closer look at the effect of sensor and actuator hardware on achievable system performance. Achievable performance is determined by sensor and actuator specifications such as bandwidth and authority, as well as sensor and actuator placement.

Sensor authority includes input range and other specifications [24], [25], whereas, for actuators, authority includes the achievable amplitude and slew rate of the control input. Amplitude saturation is unavoidable due to stroke, current, force, and power constraints, which limit the recoverable region for unstable plants [26]. Similarly, rate saturation limits the bandwidth of the controller by introducing premature rolloff and additional phase lag.

To understand the implications of sensor and actuator placement, consider a linear time-invariant system in state-space form

$$\begin{aligned}\dot{x} &= Ax + Bu, \\ y &= Cx\end{aligned}$$

with corresponding transfer function

$$G(s) = C(sI - A)^{-1}B.$$

The poles of G and the associated time constants, modal frequencies, and damping ratios depend on the plant dynamics matrix A but are independent of sensor and actuator placement, which determine the input and output matrices B and C . The achievable performance depends on both the number of sensors and actuators and their placement relative to the plant dynamics.

To demonstrate the effect of B and C on achievable performance, note that

$$G(s) = \frac{1}{\det(sI - A)} H(s),$$

where the polynomial matrix H , which is a polynomial in the case of a single-input, single-output (SISO) system, is given by

$$H(s) = \text{Badj}(sI - A)C$$

and $\text{adj}(sI - A) = [\det(sI - A)](sI - A)^{-1}$ is the adjugate of $sI - A$. Consequently, the zeros of G are determined by B and C , and thus by the placement of the sensors and actuators. The type of sensors and actuators (force, velocity, position, etc.) affects the zeros as well, whereas the order of each entry of G is determined by pole-zero cancellation. In addition, for MIMO systems, the placement of sensors and actuators determines the coupling between inputs and outputs.

To determine the relative degree of the entries of G , note that, for large $|s|$,

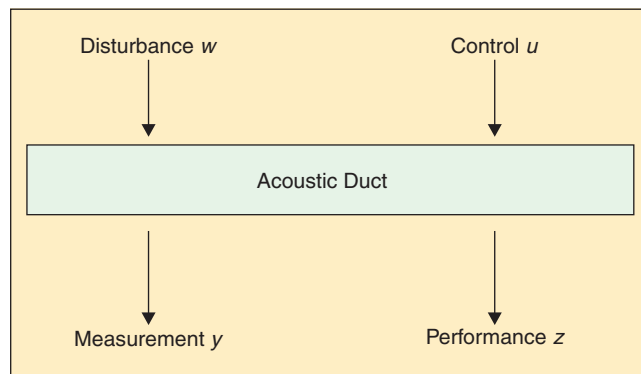


Figure 15. The arrangement of disturbance, control input, measurement, and performance signals for this acoustic duct disturbance rejection problem corresponds to the ideal architecture shown in Fig. 14.

$$G(s) = \frac{1}{s} C_E I - \frac{1}{s} A_{\pm}^{-1} B \gg \frac{1}{s} CB + O(s^{-2}),$$

which shows that the nonzero entries of CB correspond to the entries of G that have relative degree 1. As already noted, a minimum-phase SISO transfer function with relative degree 2 or less and open left-half-plane asymptotes can be unconditionally stabilized. Unconditional stabilizability in the MIMO case is more complex; however, in certain cases discussed below, it can readily be guaranteed.

Accessibility refers to the extent to which sensors and actuators are able to effect control over the plant dynamics. There are two extreme cases of interest. In terms of the state space model, *full-state sensing* occurs when the number of sensors is equal to the number of states and C is nonsingular, which implies direct sensing of each state. This is equivalent to the assumption $C = I$ in a suitable basis.

On the other hand, *full-state actuation* occurs when the number of control inputs is equal to the number of states and B is nonsingular, which implies direct control of each state. This is equivalent to the assumption $B = I$ in a suitable basis. Full-state actuation occurs in fully actuated force-to-velocity control. A nonlinear example is Euler's equation for spacecraft angular velocity with three-axis torque inputs given by

$$J\dot{w} = w \cdot Jw + u.$$

An important result in modern control theory is that every linear system with full-state sensing can be unconditionally stabilized. This property is achieved by the linear quadratic regulator (LQR), which has infinite upward gain margin. This property is nontrivial since it is not obvious how to achieve

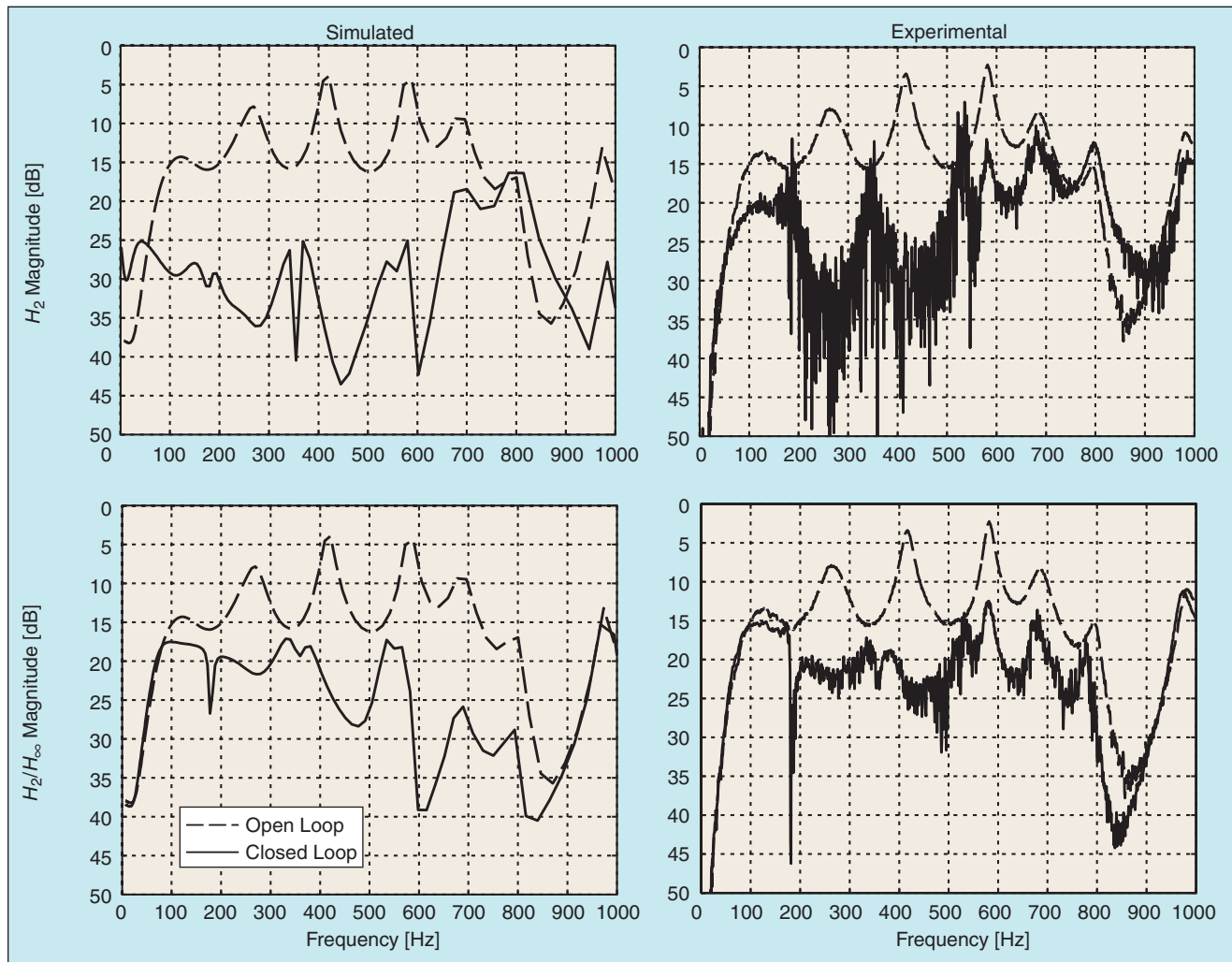


Figure 16. These plots show simulated and experimental data for an acoustic duct active noise suppression experiment. The data for the lower right plot were obtained for a controller designed using a α -domain extension of H_2/H_∞ control methods (see [23]). The absence of spillover is due to the ability of this technique to account for uncertainty arising from the identification fit error, as well as placement of the sensor, actuator, disturbance, and performance, which were arranged in accordance with the ideal architecture shown in Figs. 14 and 15.

this property using pole placement techniques. In contrast, SISO plants with relative degree 3 or greater or right-half-plane zeros cannot be unconditionally stabilized.

On the other hand, suppose that a plant is fully actuated with $B = I$ but not fully sensed. Then an unconditionally stabilizing controller can be obtained by applying LQR synthesis to the dual plant (A^T, C^T) and implementing the resulting LQR gain K in the control law $u = K^T y = K^T Cx$ to obtain the closed-loop dynamics matrix $A + K^T C$, which is asymptotically stable since it has the same eigenvalues as the asymptotically stable matrix $A^T + C^T K$. Although the solution obtained by this dual procedure is not optimal, it does represent an unconditionally stabilizing controller for the full-state-actuation output feedback problem.

When full-state actuation is not possible, we can consider an *accessible degree of freedom*, which has the form

$$m\ddot{q} + c(q)\dot{q} + k(q)q = u$$

corresponding to force-to-position control. This plant may be unstable and nonlinear with position-dependent damping and stiffness. A classical example is the van der Pol oscillator, which has limit cycle dynamics. An accessible degree of freedom has low dimensionality and limited phase variation; in the linear case with position measurement, it has relative degree 2.

An accessible degree of freedom is fundamentally easy to control under full-state (position and velocity) sensing by implementing high-gain SISO position and rate loops. Adaptive PD or PID controllers are highly effective, even in the nonlinear case, which shows the power of full-state sensing (but not necessarily PID control per se) (see [27]-[29] and Fig. 17). Consequently, it may be desirable to render each degree of freedom accessible through sensor/actuator placement or through control algorithm decoupling. However, both of these approaches depend on accurate modeling.

Multiple sensors and actuators provide greater accessibility to the plant dynamics. In fact, the extreme case of SISO control provides the smallest number of sensors and actuators relative to the plant dynamics, which is precisely the case considered in classical control. From an achievable performance point of view, SISO control is the most challenging, whereas multiple sensors and actuators provide the *potential* for improved achievable performance. The difficulty of MIMO control (except in the extreme cases of full-state sensing and full-state actuation) lies in synthesizing high-performance yet nonconservatively robust multiloop controllers, an often challenging problem in practice despite 50 years of linear state-space control theory.

Control Tuning

There are numerous impediments to *control tuning*, which refers to the choice of a suitable algorithm and tuning parameters to achieve desired performance and robustness.

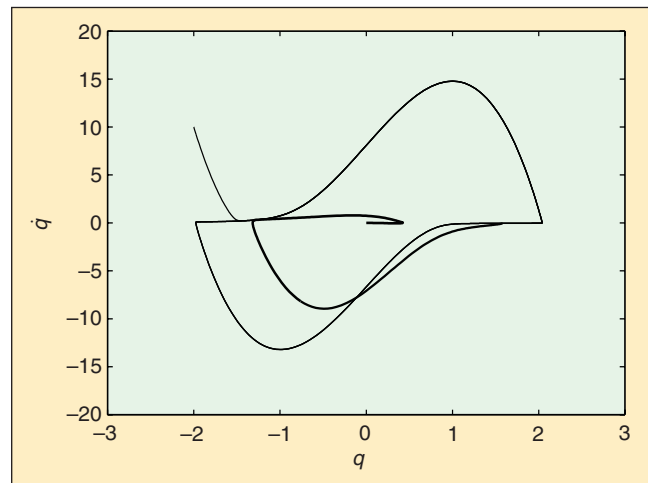


Figure 17. The van der Pol oscillator is an unstable, nonlinear, single-degree-of-freedom plant with position-dependent damping. Under full-state (position and velocity) sensing, this plant can be adaptively stabilized without knowledge of the damping coefficient. For details, see [29].

Accessibility and authority issues have a severe impact on tunability through the presence of phase variation and saturation. In addition, open-loop instability, modeling uncertainty, and nonlinearity also play a critical role.

Open-loop instability exacerbates virtually every aspect of control tuning. For example, empirical modeling, that is, identification, requires extrapolation from stable regimes so that control tuning must rely to a greater extent on analytical modeling. In addition, stabilization itself is impeded by excessive phase variation, plant dimensionality, zeros, relative degree, delays, and authority limitations.

Modeling uncertainty impedes control tuning by reducing the ability to perform model-based synthesis. In analytical modeling, uncertainty arises from unknown physics, high sensitivities, and unmodeled subsystem interaction. In empirical modeling, uncertainty results from lack of repeatability, ambient disturbances, unknown model structure, and risk and cost impediments to detailed testing.

It is important to stress that analytical modeling is valuable for the development of a suitable control architecture and associated control hardware. These phases of control engineering often occur before the system hardware is available for component-wise or end-to-end empirical modeling. On the other hand, empirical modeling is desirable for control tuning, which may depend on modeling details that are difficult to obtain from first principles. Most importantly, in each phase of control engineering, it is necessary to determine which modeling information is needed to achieve performance specifications and robustness guarantees.

Nonlinearities impede control tuning when they are difficult to model and identify, and, even when they are well modeled, they may be difficult to account for in control syn-

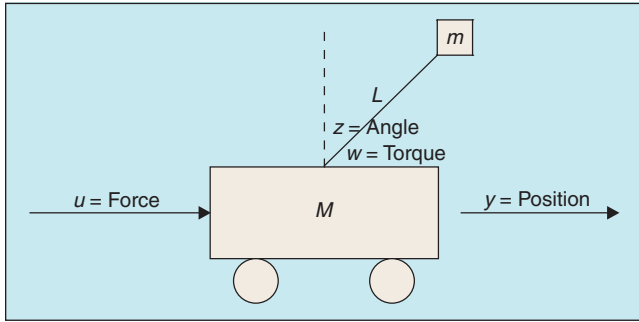


Figure 18. The inverted pendulum on a cart with cart control forcing, cart position measurement, pendulum torque disturbance, and pendulum angle performance involves linearized transfer functions that are both unstable and nonminimum phase, rendering the system difficult to control.

thesis. Although some uncontrolled plants have reasonably linear dynamics, sensors and actuators invariably introduce nonlinearities. For example, saturation is unavoidable. *Global nonlinearities* are generally smooth (linearizable) and have an increasing effect over a large range of motion. These nonlinearities are often geometric and kinematic in origin, involving trigonometric and polynomial functions. On the other hand, *local nonlinearities* are often nonsmooth (not linearizable) and have a predominant effect over a small range of motion. Typical local nonlinearities include dead zone, quantization, and backlash.

A classical example illustrating difficulties in control tuning and performance limitations is the inverted pendulum on a cart (see Fig. 18); for details, see [17]. This plant is especially difficult to control when the control input u is the cart forcing, the measurement y is the cart position, the disturbance w is the pendulum torque, and the performance z is the pendulum angle. In this case, the linearized transfer functions from w to z , from w to y , from u to z , and from u to y are given by

$$G_{zw} = \frac{m + M}{mML^2(s - p_0)(s + p_0)}$$

$$G_{yw} = G_{zu} = \frac{-1}{ML(s - p_0)(s + p_0)}$$

$$G_{yu} = \frac{(s - z_0)(s + z_0)}{Ms^2(s - p_0)(s + p_0)},$$

where the pole p_0 and zero z_0 are given by

$$p_0 = \sqrt{\frac{g}{L} + \frac{mg}{ML}}, \quad z_0 = \sqrt{\frac{g}{L}}.$$

It is shown in [17] that, because of the presence of the right-half-plane pole and zero in G_{yu} , only small stability margins are achievable under linear time-invariant control. The achievability of limited stability margins implies that model-

ing uncertainty is an impediment to control tuning for robust stability. Although alternative sensor/actuator arrangements can alleviate this difficulty, this example shows that the placement of sensors and actuators has implications for model accuracy requirements. The additional presence of nonlinearities and authority limitations further exacerbates the difficulty of control tuning. Therefore, it is the presence of multiple factors involving accessibility, authority, instability, uncertainty, and nonlinearity that render the control tuning problem difficult.

An additional impediment to control tuning is sensor noise. As discussed earlier, there is a crucial distinction between full-state feedback and output feedback in terms of achievable performance and robust stability. However, full-state sensing per se is not as strong a requirement when sensor and disturbance noise are absent. This point has already been made in terms of singular estimation theory [20]-[22]. Alternatively, consider the discrete-time system

$$x(k+1) = Ax(k) + Bu(k)$$

$$y(k) = Cx(k).$$

Assuming (A, C) is observable, it is possible to solve exactly for $x(k)$ using a window of past measurements $y(k), \dots, y(k-n)$ and control inputs $u(k-1), \dots, u(k-n)$. With these data, the state $x(k)$ can be reconstructed exactly, and full-state feedback control can be used just as in the case of full-state sensing; that is, $C = I$. The technique of solving for $x(k)$ in terms of a window of measurements is equivalent to implementing a deadbeat observer. Thus, the availability of noise-free measurements is ultimately equivalent to a full-state-feedback control architecture. However, noise-free measurements are not available in practice. The analogous approach in continuous time requires differentiating y multiple times, which is not feasible in the presence of plant and sensor noise. These observations imply that noise is a major impediment to achievable performance.

So What Makes Some Control Problems Hard?

The above discussion shows that each phase of control engineering presents impediments to the effectiveness of control technology and suggests that success depends on multiple critical aspects. In the control strategy phase, it is important to assess the need for control in terms of performance, cost, and risk. In the control physics phase, it is essential to exploit high-leverage control physics. In the control architecture phase, it is important to design a sensor/actuator/disturbance/performance architecture that balances robustness, performance, and hardware requirements. In the control hardware phase, it is essential to provide adequate accessibility and authority. And, finally, in the control tuning phase, it is necessary to account for instability, nonlinearity, control-loop coupling, and uncertainty, as

well as their mutual interaction. In addition to these five phases, there are many important subissues, such as software engineering, fault tolerance, and hardware maintainability, which I have not addressed here.

The control engineer must be aware of engineering trade-offs throughout all of these phases. In contrast, academically oriented research papers typically focus on a limited range of issues. The ability of the research community to address the full range of issues that arise in control engineering is crucial to making theoretical control research more relevant to technology, thereby closing the gap between theory and practice.

So, what makes some control problems hard? Our holistic point of view is that a control problem is hard when multiple impediments occur simultaneously. Constraints on physics, architecture, accessibility, authority, nonlinearity, instability, dimensionality, uncertainty, and noise can often be overcome without much difficulty when they are effectively the *only* operative constraint. However, when multiple constraints are present, the control problem suddenly becomes vastly more difficult. Fortunately, engineers are often able to circumvent this situation by designing out the difficulties through plant redesign, improved hardware, benign architecture, and more detailed modeling. In other cases, however, the control problem is intrinsically difficult, and no amount of redesign or expenditure of effort can make the difficulties disappear. It is in these cases that we look toward innovative basic control research to extend the capabilities of systems and control theory to overcome the challenges of truly difficult control problems.

Acknowledgments

I would like to thank Scott Erwin, Andy Sparks, and Kris Hollot for helpful comments.

References

[1] S.J. Elliott, "Down with noise," *IEEE Spectrum*, vol. 36, pp. 54-61, June 1999.
 [2] B. de Jager, "Rotating stall and surge: A survey," in *Proc. IEEE Conf. Decision and Control*, New Orleans, LA, Dec. 1995, pp. 1857-1862.
 [3] S. Ashley, "Warp drive underwater," *Sci. Amer.*, pp. 70-79, May 2001.
 [4] D.S. Bernstein, "A student's guide to classical control," *IEEE Contr. Syst. Mag.*, vol. 17, pp. 96-100, Aug. 1997.
 [5] D.S. Bernstein, "On bridging the theory/practice gap," *IEEE Contr. Syst. Mag.*, vol. 19, pp. 64-70, Dec. 1999.
 [6] D.S. Bernstein, "A plant taxonomy for designing control experiments," *IEEE Contr. Syst. Mag.*, vol. 21, pp. 25-30, June 2001.
 [7] I.J. Busch-Vishniac, *Electromechanical Sensors and Actuators*. New York: Springer, 1998.
 [8] P.J. Nahin, *The Science of Radio*. Woodbury, NY: American Institute of Physics, 1996.
 [9] B.A. Francis, *A Course in H_∞ Control Theory*. New York: Springer, 1987.
 [10] G.H. Tallman and O.J.M. Smith, "Analog study of dead-beat posicast control," *IRE Trans. Automat. Contr.*, vol. 3, pp. 14-21, 1958.

[11] N.C. Singer and W.P. Seering, "Preshaping command inputs to reduce system vibration," *Trans. ASME J. Dyn., Meas., Contr.*, vol. 112, pp. 76-82, 1990.
 [12] S.M. Kuo and D.R. Morgan, *Active Noise Control Systems*. New York: Wiley, 1996.
 [13] R. Venugopal and D.S. Bernstein, "Adaptive disturbance rejection using ARMARKOV system representations," *IEEE Trans. Contr. Syst. Technol.*, vol. 8, pp. 257-269, 2000.
 [14] H.W. Bode, *Network Analysis and Feedback Amplifier Design*. New York: Van Nostrand, 1945.
 [15] R.H. Middleton, "Trade-offs in linear control system design," *Automatica*, vol. 27, pp. 281-292, 1991.
 [16] J. Chen, "Sensitivity integral relations and design tradeoffs in linear multivariable feedback systems," in *Proc. Amer. Control Conf.*, 1993, pp. 3160-3164.
 [17] M.M. Seron, J.H. Braslavsky, and G.C. Goodwin, *Fundamental Limitations in Filtering and Control*. New York: Springer, 1997.
 [18] J.C. Doyle, B.A. Francis, and A.R. Tannenbaum, *Feedback Control Theory*. New York: Macmillan, 1992.
 [19] J. Hong and D.S. Bernstein, "Bode integral constraints, collocation, and spillover in active noise and vibration control," *IEEE Trans. Contr. Syst. Technol.*, vol. 6, pp. 111-120, 1998.
 [20] H. Kwakernaak and R. Sivan, "The maximally achievable accuracy of linear optimal regulators and linear optimal filters," *IEEE Trans. Automat. Contr.*, vol. 17, pp. 79-86, 1972.
 [21] Y. Halevi and Z.J. Palmor, "Admissible MIMO singular observation LQG designs," *Automatica*, vol. 24, pp. 43-51, 1988.
 [22] E. Soroka and U. Shaked, "The LQG optimal regulation problem for systems with perfect measurements: Explicit solution, properties, and application to practical designs," *IEEE Trans. Automat. Contr.*, vol. 33, pp. 941-944, 1988.
 [23] R.S. Erwin and D.S. Bernstein, "Discrete-time H_2 / H_∞ control of an acoustic duct: α -domain design and experimental results," in *Proc. IEEE Conf. Decision and Control*, San Diego, CA, Dec. 1997, pp. 281-282.
 [24] D.S. Bernstein, "Sensor performance specifications," *IEEE Contr. Syst. Mag.*, vol. 21, pp. 9-18, Aug. 2001.
 [25] R. Koplon, M.L.J. Hautus, and E.D. Sontag, "Observability of linear systems with saturated outputs," *Lin. Alg. Appl.*, vol. 205-206, pp. 909-936, 1994.
 [26] D.S. Bernstein and A.N. Michel, "A chronological bibliography on saturating actuators," *Int. J. Robust Nonlinear Contr.*, vol. 5, pp. 375-380, 1995.
 [27] J. Hong and D.S. Bernstein, "Experimental application of direct adaptive control laws for adaptive stabilization and command following," in *Proc. Conf. Decision and Control*, Phoenix, AZ, Dec. 1999, pp. 779-783.
 [28] J. Hong and D.S. Bernstein, "Adaptive stabilization of nonlinear oscillators using direct adaptive control," *Int. J. Contr.*, vol. 74, pp. 432-444, 2001.
 [29] A.V. Roup and D.S. Bernstein, "Stabilization of a class of nonlinear systems using direct adaptive control," *Proc. Amer. Contr. Conf.*, Chicago, IL, June 2000, pp. 3148-3152; also, *IEEE Trans. Autom. Contr.*, to appear.

Dennis S. Bernstein obtained his Ph.D. degree from the University of Michigan in 1982, where he is currently a faculty member in the Aerospace Engineering Department. His research involves both theory and experiments relating to aerospace engineering, especially noise, vibration, and motion control. In short, he is interested in controlling anything that moves. He is currently focusing on identification and adaptive control, which is motivated by the fundamental question: "What do you really need to know about a system to control it?"