Fundamental Characteristics of Feedback Mechanisms

DRAFT

Technical Report of the ISIS Group
at the University of Notre Dame
ISIS-09-006
September, 2009

Panos J. Antsaklis
Department of Electrical Engineering
University of Notre Dame
Notre Dame, IN 46556

Interdisciplinary Studies in Intelligent Systems

Fundamental Characteristics of Feedback Mechanisms

1 Introduction and Summary of Results

Feedback is everywhere. Feedback is ubiquitous. Feedback is all around us and inside us.

It would be truly interesting to find out how old feedback is. How far back in time can we trace feedback mechanisms? Since the beginning of life? Certainly even single cell creatures react to sensory inputs, they change direction, are attracted to light. The purpose of having sensors is to use the information and act upon it to feed, to avoid danger, to find shelter. This is feedback control. Single cell creatures in addition to using feedback to react to external stimuli they also have feedback to regulate automatically internal functions. Is there life without feedback? In my opinion it is doubtful! Life functions and feedback go hand in hand! But even before the beginning of life, one could imagine feedback playing a central role in physical phenomena helping settle processes to equilibrium points. But this discussion is probably for another time and place.

In the area of Systems and Control theory, the emphasis has been on designing feedback controllers given a model of the process to be controlled. Many powerful methodologies have been introduced in the past half century to design controllers, decision mechanisms, that stabilize and achieve desired performance in a robust way, being tolerant to certain class of plant parameter variations and external disturbances. Feedback or closed loop control is used, instead of feedforward or open loop control, because of uncertainties in the plant and its environment. Methods that optimize performance (LQR/LQG, Hinf) have also been used successfully for certain classes of systems. The models are typically ordinary differential or difference equations mostly linear and time-invariant but also time-varying, and nonlinear. Less often the behavior of interest is described by partial differential equations. Discrete event systems such as manufacturing systems are typically described by automata and Petri nets.

Significantly less effort has been spent in the past half century on understanding exactly how and why feedback works so well not only in the control of engineered systems, but in natural systems as well. What are the fundamental principles, the fundamental mechanisms which make feedback control so powerfully effective? These fundamental mechanisms should be independent of the particular type of mathematical models used, that is the system may be described by differential equations, by automata, by logic expressions, by natural language since we do know that feedback is ubiquitous and works! What are these fundamental properties that are present everywhere? Deeper understanding would make it possible to understand better the mechanisms at work in natural systems and would lead to designing better controllers.

So the question is whether there are intrinsic properties of feedback that transcend particular applications and models and are present in electrical, mechanical, physical, biological, social, economic systems. *Is there a fundamental, ever-present, feedback property?*

1.1 Feedback

Feedback is a mechanism, ubiquitous in nature that drastically and dramatically changes the behavior of the system (of the plant, the process to be controlled). By behavior we mean the observed response of the system to external stimulus such as an input or initial condition. Examples of feedback abound. Here a familiar situation every car driver has experienced is described.

When driving and the slope of the road starts increasing, the car speed starts decreasing. Typically the driver detects this by looking at the speedometer and presses the gas pedal a bit more to increase the fuel rate and bring the speed up again to the previous level. The driver detects-via the speedometer-the difference between the desired and actual speeds (the error). When the error is positive-meaning that the actual speed is less than the desired-the driver increases the fuel to the engine; if negative-that is the actual speed is higher than the desired speed (going downhill for example)-the driver decreases the fuel input and the car slows down. Cruise (speed) controlers in cars do the same thing but automatically.

In what way does feedback alter the behavior of a system? Consider the system consisting of the car and the control fuel input set at certain level corresponding to the desired speed (here the output) when the road is horizontal. When there is a positive road incline and no corrective action is applied the car normally will start slowing down, as its normal behavior dictates. Consider now having as input the desired speed with its corresponding fuel rate and adding to this an appropriate additional positive fuel rate when the incline is positive (and the error is positive). Now the system can be seen as having the same reference input (desired speed) as before but with feedback it exhibits a different behavior since now the car does not slow down. So with the same desired speed as input, feedback makes it possible for the car to have a different dynamic behavior!

Can this be done without feedback? If we do know the details of the incline and have an accurate model of the response of the car when the fuel rate is increased, then the driver, or a machine, can apply just the right additional fuel to do the job. However this implies knowledge we do not have. How for example can we have such accurate knowledge so to tell exactly-without using sensor information-when the incline starts and the car slows down? How do we know that there will not be a sudden gust of headwind, a disturbance, that will slow us down? In fact both such uncertainties in the plant model and in disturbances are rather common in practice and so open loop control typically does not work except in special cases.

The amazing thing is that with feedback the change of behavior is automatic. When feedback information is available the driver maintains the desired speed, without intimate knowledge of the slope of the incline or of the engine of the car, by just observing the speedometer and adding fuel when the error is positive, and reducing fuel when the error is negative. (Note that a more sophisticated controller may consider not only the error in speed but also the rate of change of the actual speed so to react faster).

1.2 Feedback's Fundamental Properties

In view of the above discusion, it appears that *feedback is a way to change behavior as if we were changing the plant itself but without actually doing so.* How is feedback changing the plant behavior?

What are feedback's most fundamental intrinsic properties? Is it its ability to reduce sensitivity of the behavior to uncertainties in the plant parameters and external disturbances? This is appealing because the reasons for using feedback instead of open loop control are these uncertainties. Unfortunately this is not so. Low sensitivity depends on the particular choice for the controller and the choice may decrease or increase sensitivity.

For example the sensitivity S of a plant $G = \frac{1}{s+1}$ in a unity (negative) feedback configuration with a static controller $G_c = k$ is $S = (1+GG_c)^{-1} = \frac{s+1}{s+1+k}$. Note that the plant is stable for $1+k > 0$ or $-1 < k$. For $-1 < k < 0$ the sensitivity to parameter variations is greater than 1 that is the sensitivity of the closed loop is worse than that of an open loop (see also Appendix).

So reducing sensitivity cannot be a property that is present independently of the particular controller used. Is then stabilization the most fundamental feedback property? Similarly the choice of controller may stabilize or destabilize the system and so *stabilization cannot be an intrinsic feedback property.* So what is it?

Automatically Changing the Dynamics

A key feedback property that transcends all applications and all choices for the controller appears to be the ability of feedback to completely alter the plant dynamic behavior no matter what the particular plant dynamics are. This property is independent and distinct and separate from the ability to assign new dynamics by selecting the controller appropriately for stabilization or performance. Note that in the open loop non-feedback case we also have the ability to easily assign new dynamics, the problem being that to cancel existing dynamics is not always possible.

Even small feedback gains can change the dynamic behavior. Consider for example the root-locus of a LTI SISO plant. As the gain k increases from 0, even by a very small amount, the closed loop poles are not the open loop poles any longer-the open loop poles seem to vanish (similarly when k decreases from 0).

For example consider the plant $G(s) = \frac{1}{s(s+1)(s+2)}$ and its root locus for $k \geq 0$ (see Appendix 11). For $k = 0$ the closed loop poles are at the open loop pole locations and for small positive k the closed loop poles are different from the open loop pole locations.

As k goes towards infinity the closed loop poles move towards the finite zeros of the plant and to points at infinity along the asymptotes. For very large gains k the plant dynamics seem to cancel out completely. These are well known phenomena in the controls literature and provide the clues for the fundamental mechanisms of feedback.

Automatically Changing the Gains

A second fundamental characteristic of feedback is change in gain. It should be noted however that in contrast to the previous property of changing the dynamics, this property is dependent on the selection of the feedback gain and strictly speaking it may not qualify as a fundamental property. Never the less, as it will be shown, it is directly related to a fundamental property of feedback, namely the ability to reduce the sensitivity to parameter variations in the plant. As it is discussed below, large gain variations inside the loop typically can only cause small gain variations outside the loop and so the effect of uncertainties may be much reduced. The automatic change in the plant dynamics when closing the loop is essential in the resulting ability of appropriately changing the feedback gains to stabilize the system in a robust way. In an analogous fashion, the automatic change in the gain is essential to the ability of appropriately choosing the feedback gain to reduce the dependence of the system to uncertainties. *So, in feedback, the automatic change of the plant dynamics is related to stabilization, while* *the automatic change in the plant gain is related to reduction of sensitivity to uncertainties.*

Changing the dynamics is more complicated and so the next sections will be devoted to explaining the mechanisms involved. The mechanism of changing the gain can be easily seen from a simple feedback loop involving only static gains. Let the plant be an amplifier of gain A, i.e. $G(s) = A$, and consider a unity feedback control configuration with controller $G_c = k$, a static gain (see Fig. 1.6).

Here the output of the closed loop system is given by

$$Y = \frac{kA}{1+kA}R = TR \tag{1}$$

with

$$U = \frac{k}{1+kA}R, \quad E = \frac{1}{1+kA}R \tag{2}$$

while the open loop gain is

$$Y = AU. \tag{3}$$

Sensitivity to plant parameter variations can be studied in detail using the sensitivity function $S = \frac{1}{1+kA}$ and the relation

$$\frac{\Delta T}{T} \approx S\frac{\Delta G}{G} \tag{4}$$

where G is the plant and T is the closed loop transfer function.

The R to Y gain remains the same, equal to the plant's gain A only when $k = \frac{1}{1-A}$; in general the closed loop gain will be different from A. The absolute value of the closed loop gain will be less than 1 for $-\frac{1}{2} < kA$ and will be greater than 1 for $kA < -\frac{1}{2}$. The sensitivity function $S = \frac{1}{1+kA}$ will have absolute value $|S| < 1$ for $kA > 0$ and for $kA < -2$; it will have $|S| > 1$ only for $-2 < kA < 0$. That is, the open loop gain kA maps the sensitivity S to $-1...1$ range and the closed loop gain T to $0...1$ range for all $kA > 0$. A consequence of this is that variations in A (uncertainties) will not affect as much the overall R to Y gain, that is feedback reduces the sensitivity to parameter variations at the expense of reducing the gain. As a specific example, consider $k = 1$ and $A = 10000$. In this case the R to Y gain is

$$\frac{A}{1+A} = \frac{10000}{1+10000} \approx 1 \tag{5}$$

which represents a great reduction in gain. The benefit is that if A changes say by 20% then

$$\frac{A+.2A}{1+A+.2A} = \frac{1.2A}{1+1.2A} = \frac{1.2}{\frac{1}{A}+1.2} \approx 1. \tag{6}$$

That is, the overall gain is insensitive to variations in the open loop gain A.

Similar results may be seen in the more general case.

Here

$$Y = \frac{GG_c}{1+GG_c}R = TR \tag{7}$$

with

$$U = \frac{G_c}{1+GG_c}R, \quad E = \frac{1}{1+GG_c}R \tag{8}$$

For specific frequencies, when $GG_c > -\frac{1}{2}$ the closed loop gain has absolute value less than 1. When $GG_c > 0$ or $GG_c < -2$, the absolute value of the sensitivity S is less than 1. Feedback reduces automatically the sensitivity function to less than 1 (for a large class of feedback gains). For example, here for any $GG_c > 0$ $|S| < 1$. Not that also $|T| < 1$. So the sensitivity function S maps all positive loop gains to the 0...1 range for $|S|$ at the expense of reducing the closed loop gain $|T|$ to the 0...1 range as well. This reduction of the overall R to Y gain of the compensated system is the price to pay for low sensitivity.

The Return Difference

The return difference in a feedback loop is the difference between the transmitted (measured) and returned signals at the output of the plant; see section 1.10. Both feedback fundamental characteristics, automatic change of dynamcis (poles) and gains, are caused by the feedback interconnection, which can be expressed in terms of the return difference. Specifically, in the unity feedback configuration, the output $Y = -GG_cY + GG_cR$ or $(1 + GG_C)Y = GG_cR$ where $(1 + GG_c)$ is the return difference; see Section 1.10.

Sidebar: *The reason for not having a clear explanation of the feedback mechanisms at work after many decades of impressive developments in the mathematical theory of control may perhaps be due to the fact that modern systems and control theories typically consider feedback to be already part of the setup and study the behavior of the whole system. So the actual feedback mechanisms have not been explored nearly as well as the effects of feedback on the compensated systems, where selection of appropriate feedback gains are of importance and of main interest. In the earlier era of classical control where control specialists typically were closer to applications, it was clearly seen that the control law is there to manipulate the input u and produce the desired effect. The plant dynamics cannot be changed. So understanding exactly how u acts on the plant would have been of great interest, but the understanding brought forth by internal system descriptions was not readily available then. Today we can look at this problem having the benefit of the* *insights developed over many years since the classical era of control in the 1950s.*

In the following sections the focus will be on simple LTI plants under open and closed loop control. We focus on the automatic changes of dynamics when closing the loop. Some basic concepts will be reviewed and presented in a way that sheds light into the basic fundamental mechanisms of feedback control. The Appendix contain discussion on several related topics including pole/zero cancellations, sampled data, and nonlinear systems. The effect of high gains is discussed in Section 2, open loop control in Section 3, and closed loop control in Sections 4 and 5. In Section 6, state space representations are discussed, and the two degrees of freedom configurations are discussed in Sections 7 and 8. In section 9, open and closed loop control are compared and in Section 10 the role of the return difference is discussed.

2 High Gains in the Feedback Loop—A First Glimpse at the Feedback Mechanism

It is well known that feedback control can be seen as a mechanism that approximately inverts the plant dynamics, producing an "approximate" inverse of the plant at its control input. This can be seen for example using the simple error feedback control systems in Figure 1. When $H = 1$, the transfer functions are

$$\frac{Y}{R} = T = \frac{GG_c}{1 + GG_c} \quad \text{and} \quad \frac{U}{R} = M = \frac{G_c}{1 + GG_c}. \tag{9}$$

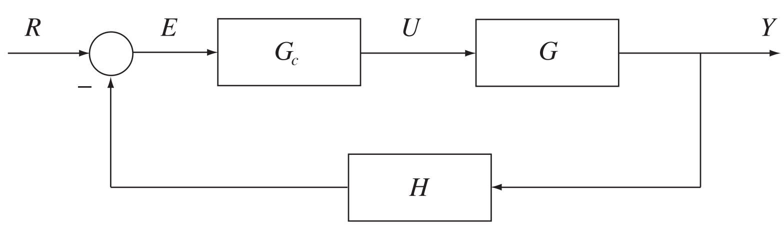

Figure 1:

If $|GG_c| >> 1$ then

$$\frac{Y}{R} = T \cong 1 \quad \text{and} \quad \frac{U}{R} = M \cong \frac{1}{G}. \tag{10}$$

That is, at the control input u of the plant, the external input r acts through an inverse of the plant G $(U \cong \frac{1}{G}R)$, so to cancel all plant dynamics and produce an output y, which is approximately equal to the reference input r ($Y = GU \cong G\frac{1}{G}R \cong R$). In other words, the input to the plant u, generated by the external input r, is such that when applied to the plant G causes the plant output y to be (approximately) equal to the externally applied input r.

Note that in the case when $G_c = k$, a real gain, this effect can also be seen from the Root Locus where as the gain increases the closed-loop poles go towards the open-loop finite and infinite zeros along the asymptotes and so for high gain, pole-zero cancellations do occur and the overall transfer function is approximately 1.

If in addition there is a controller H ($\neq 1$) in the feedback path, then again the plant and its inverse cancel, however the overall gain in this case is (approximately) independent of the plant and equals $\frac{1}{H}$:

$$\frac{Y}{R} = T \cong \frac{1}{H} \quad \text{and} \quad \frac{U}{R} = M \cong \frac{1}{GH}. \tag{11}$$

H is selected to have a precise value (typically less than 1) so the compensated system has the desirable gain, while it remains robust to parameter variations in G_cG.

Similar results can be shown in the nonlinear case (see for example [1]–pp. 29-36 and Appendix A.6). Again in this case, when high gains are applied the external input acts through an inverse of the plant on the plant's input.

Remark: High gains of course can have undesirable effects such as amplification of measurement noise and even worse, can cause instability. The latter can be easily seen, for example, via the Root-Locus in the case of non-minimum phase plants where the closed-loop poles approach right-half plane zeros for high gains and so the closed-loop system becomes unstable.

Discussion: Can it then be said that the feedback mechanism always acts by generating a plant "inverse" and canceling somehow the plant dynamics, as it was shown to be true in the case of high-loop gains?

Although our intuition based conjecture is basically correct, it is not very exact. Explaining exactly how feedback acts on the plant is the goal of the present work. Clear understanding of the feedback mechanism is very important especially today when feedback is identified and used to explain the development of a variety of processes found in diverse areas from biology, to physics, to finance.

If we understand clearly how the feedback control produces all these wonderful results, it will perhaps be easier to understand the cases when feedback information is not readily available as it is the case, for example, in networked control systems, where the plant may have to operate often in an open-loop configuration.

In view of this, the closely related open-loop control versus closed-loop control topic is discussed in detail in the folowing. Furthermore, as it was discussed above, at the plant input u, for high gains the external input r acts through an inverse of the plant that can be seen as an open-loop "equivalent" to feedback configuration, and these mechanisms will also be discussed below. Note that in the Appendix A.1 a review of pole/zero cancellation mechanisms (in both frequency and time domains) are given for completeness.

3 Open-loop Control (Feed-forward Control)—A Simple Example

We are interested in obtaining a desired transfer function (desired response to any allowed input) from a given plant, the input of which is controlled by a controller in series with the plant. We shall start with a simple case which nevertheless contains the important salient features.

Consider a plant to be controlled described by the first-order differential equation $a(dy/dt) + y = u$ with initial condition $y(0)$. If $Y(s)$ and $U(s)$ are the Laplace transforms of the output and the input respectively, the transfer function is

$$Y(s)/U(s) = G(s) = \frac{1}{as+1}. \tag{12}$$

Let the output disturbance be $d(t)$, so that

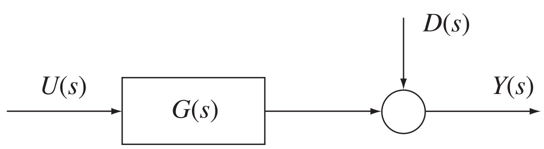

Figure 2:

From the differential equation it is easy to see that $a(sY_p(s) - y(0)) + Y_p(s) = U(s)$ from which

$$Y(s) = \frac{a}{as+1}y(0) + \frac{1}{as+1}U(s) + D(s). \tag{13}$$

If we consider the open-loop controller in Figure 1.3,

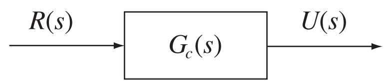

Figure 3:

where

$$U(s)/R(s) = G_c(s) = \frac{bs+1}{cs+1}(c(du/dt) + u = b(dr/dt) + r), \qquad (14)$$

it can be shown that

$$U(s) = \frac{cu(0) - br(0)}{cs+1} + \frac{bs+1}{cs+1}R(s). \qquad (15)$$

The output of the compensated system, in Figure 1.4,

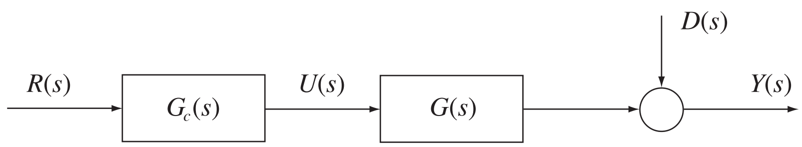

Figure 4:

is then

$$Y(s) = \frac{(cs+1)ay(0) + cu(0) - br(0)}{(as+1)(cs+1)} + \frac{bs+1}{(as+1)(cs+1)}R(s) + D(s). \qquad (16)$$

The behavior may be changed by canceling undesirable plant dynamics via pole-zero cancellations. As it can be seen in Appendix 1, in the time domain, these cancellations correspond to making the coefficients of the corresponding modes equal to zero (or making the modes uncontrollable and/or unobservable).

Consider the part of the output that is due solely to the input $R(s)$,

$$Y_R(s) = GG_c(s) = \frac{bs+1}{(as+1)(cs+1)}R(s). \qquad (17)$$

If we select $b = a$ then the pole of the plant G at $-1/a$ cancels out with the zero of the controller G_c and the overall pole dynamics are completely characterized by the pole of the controller at $-1/c$ which can be chosen to our liking. To illustrate, let $r(t) = 1(t)$ the unit step; then $R(s) = \frac{1}{s}$ and (let $a \neq c$ for simplicity)

$$\begin{aligned} y_R(t) &= L^{-1}\left\{\tfrac{bs+1}{(as+1)(cs+1)}\tfrac{1}{s}\right\} \\ &= L^{-1}\left\{\tfrac{-a\frac{a-b}{a-c}}{as+1} + \tfrac{-c\frac{c-b}{c-a}}{cs+1} + \tfrac{1}{s}\right\} \\ &= \left[\tfrac{b-a}{a-c}e^{-\frac{t}{a}} + \tfrac{c-b}{a-c}e^{-\frac{t}{c}} + 1\right]1(t). \end{aligned} \tag{18}$$

When $b = a$ then

$$y_R(t) = \left[1 - e^{-\frac{t}{c}}\right]1(t). \tag{19}$$

If however, a is not known exactly (i.e. the exact location of the pole of the plant is not known exactly) and b is not taken to be exactly equal to a, but $b = a + \varepsilon$, then

$$y_R(t) = \left[\frac{\varepsilon}{a-c}e^{-\frac{t}{a}} + \left[1 - \left(1 + \frac{\varepsilon}{a-c}\right)e^{-\frac{t}{c}}\right]\right]1(t) \tag{20}$$

where it can be seen that the plant pole at $-1/a$ has not been cancelled. If $-1/a$ is positive (unstable pole) then the corresponding mode will grow with time and the system will be unstable.

The part of the response due to initial condition is $Y_I(s)$, and in the time domain is (let $a \neq c$ for simplicity)

$$\begin{aligned} y_I(t) &= L^{-1}\left\{\frac{a}{as+1}y(0) + \frac{cu(0)-br(0)}{a-c}\left(\frac{a}{as+1} - \frac{c}{cs+1}\right)\right\} \\ &= \left[\left[y(0) + \frac{cu(0)-br(0)}{a-c}\right]e^{-\frac{t}{a}} - \frac{cu(0)-br(0)}{a-c}e^{-\frac{t}{c}}\right]1(t). \end{aligned} \tag{21}$$

When b, c, $u(0)$, and $r(0)$ are such that the coefficient of $e^{-\frac{t}{a}}$ is zero, then the plant dynamics are suppressed from the response. This happens when

$$(a-c)y(0) + cu(0) - br(0) = 0 \tag{22}$$

which is exactly the condition in $Y_I(s)$ for the factor $as+1$ in the denominator to cancel with the numerator (the numerator should be zero for $s = -1/a$ which is the pole to be cancelled). Again, as in the $y_R(t)$ case above, if a and $y(0)$ are not exactly known then the unstable mode will not be eliminated from $y_I(t)$.

For completeness let us also find the expression for the input to the plant. The control input $u(t)$ to the plant when $R(s) = 1/s$ is

$$\begin{aligned} u(t) &= L^{-1}\{U(s)\} = L^{-1}\left\{\frac{cu(0) - br(0)}{cs+1} + \frac{bs+1}{cs+1}\frac{1}{s}\right\} \\ &= \left[(u(0) - \frac{b}{c}r(0))e^{-\frac{t}{c}} + (b-c)e^{-\frac{t}{c}} + 1\right] 1(t). \end{aligned} \tag{23}$$

Note that when $b = a$ and $u(0)$, $r(0)$ are chosen to satisfy 22, then

$$u(t) = \left[-\frac{y(0)}{c}(a-c)e^{-\frac{t}{c}} + (a-c)e^{-\frac{t}{c}} + 1\right] 1(t) \tag{24}$$

In view of the above analysis it is clear that in open loop control, in order to change the plant poles and therefore the plant dynamic behavior one needs exact knowledge of the pole location $(-1/a)$ *and of the initial condition* $(y(0))$*. Furthermore, the disturbance* $d(t)$ *can only be suppressed if it is measured directly. Uncertainties in the plant model and the environment are part of almost every design and so* $-1/a$, $y(0)$, *and* $d(t)$ *are typically not known exactly. So it is impossible to stabilize an unstable system using open-loop control.*

Note: The open loop control has high "fragility." It is very sensitive to the location of the poles and of the initial conditions. The effect of such errors in the location of unstable poles can be catastrophic.

4 Closed-loop Control (Feedback Control)—A Simple Example

Consider now the unity (error) feedback configuration, in Figure 1.5, where the plant is described again by $a(dy/dt) + y = u$ with initial condition $y(0)$. The transfer function is $G(s) = 1/(as+1)$ as before.
Here again

$$Y(s) = \frac{a}{as+1}y(0) + \frac{1}{as+1}U(s) + D(s) \tag{25}$$

The control input U is now generated via a feedback mechanism. Specifically

$$U(s) = k(R(s) - Y(s)) = kR(s) - \frac{ka}{as+1}y(0) - \frac{k}{as+1}U(s) - kD(s) \tag{26}$$

from which

$$U(s) = -\frac{kay(0)}{as+1+k} + \frac{k(as+1)}{as+1+k}R(s) - \frac{k(as+1)}{as+1+k}D(s) \tag{27}$$

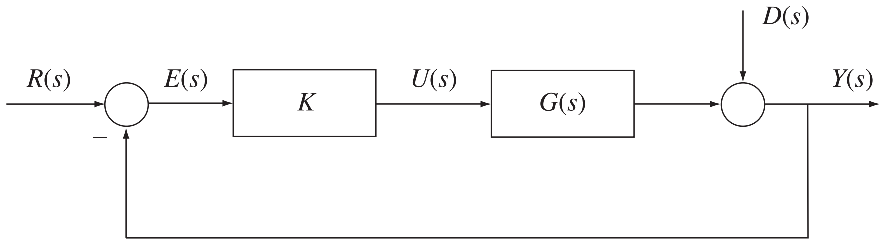

Figure 5:

When this input is applied to the plant

$$
\begin{aligned}
Y(s) &= \left[\frac{a}{as+1}y(0) - \frac{kay(0)}{(as+1)(as+1+k)}\right] + \frac{k(as+1)}{(as+1)(as+1+k)}R(s) \\
&\quad - \frac{k(as+1)}{(as+1)(as+1+k)}D(s) + D(s) \\
&= \frac{a(as+1)y(0)}{(as+1)(as+1+k)} + \frac{k(as+1)}{(as+1)(as+1+k)}R(s) \\
&\quad \frac{(as+1)^2}{(as+1)(as+1+k)}D(s)
\end{aligned}
$$

or

$$
Y(s) = \frac{ay(0)}{as+1+k} + \frac{k}{as+1+k}R(s) + \frac{as+1}{as+1+k}D(s) \tag{28}
$$

Observe that the factor $(as+1)$ that corresponds to the undesirable open-loop dynamics was cancelled in all the terms. The denominator $(as+1+k)$ that represents the desirable dynamics appears in all the terms. Note that the system is stable for all k such that $\frac{1+k}{a} > 0$. (When $a > 0$ (stable plant) for $k > -1$ the closed-loop is stable; when $a < 0$ (unstable plant) the closed-loop is stable for $k < -1$.

The range of the acceptable values for the gain k for stability or the stability robustness of the system is remarkable and it is achieved with feedback. In the case of the unstable plant for example, the gain k can be selected within a very wide range $(-\infty < k < -1)$ and the system will be stable even when the exact pole location $(-\frac{1}{a})$ and the initial condition $(y(0))$ are not known. This is not the case when open-loop control is used as it was shown above.

The above examples suggest that feedback acts in two distinct steps. In the first step the plant dynamics are cancelled automatically. In the second step new desirable dynamics are assigned by appropriately choosing the *feedback control law. In the following, these two fundamental feedback actions are discussed at length with the cancellation of plant dynamics shown initially for a more general case and for the general two degrees of freedom controllers in the next section. The exact feedback mechanism that cancels the plant dynamics is shown.*

5 Open and Closed-loop Control—A More General Analysis

Similar results can be derived in the more general case when $G(s) = \frac{n(s)}{d(s)}$ and $G_c(s) = \frac{n_c(s)}{d_c(s)}$, where $n(s)$ and $d(s)$ are polynomials with real coefficients and G, G_c are rational proper transfer functions. Consider first the open-loop control case of Figure 4. Here the plant output is

$$Y = \frac{n_o}{d} + \frac{n}{d}U + D \tag{29}$$

where n_o is a polynomial term involving the initial conditions of the plant; when the initial conditions are zero, $n_o = 0$. Similarly, the controller output is

$$U = \frac{n_{co}}{d_c} + \frac{n_c}{d_c}R \tag{30}$$

Therefore the overall system output in the open loop control case (in Figure 4) is

$$Y = \frac{n_o d_c + n n_{co}}{d d_c} + \frac{n n_c}{d d_c}R + D. \tag{31}$$

To change the plant behavior, all undesirable plant dynamics in the plant denominator d must be cancelled via pole/zero cancellations. This can be accomplished by selecting n_c and also the initial conditions in n_{co} (for cancellation between d and $n_o d_c + n n_{co}$). It is clear that when there are uncertainties in the undesirable plant pole locations and initial conditions, it is not possible to select the open loop controller to cancel the undesirable plant dynamics. So similar results as in the previous section are derived for the open loop control case as expected.

Consider now the feedback case. Again, let $G = \frac{n}{d}$ and $G_c = \frac{n_c}{d_c}$ and consider the feedback interconnection of Figure 6, where D is the disturbance (in the Laplace transform domain).
Then

$$Y = \frac{n_o}{d} + \frac{n}{d}U + D \tag{32}$$

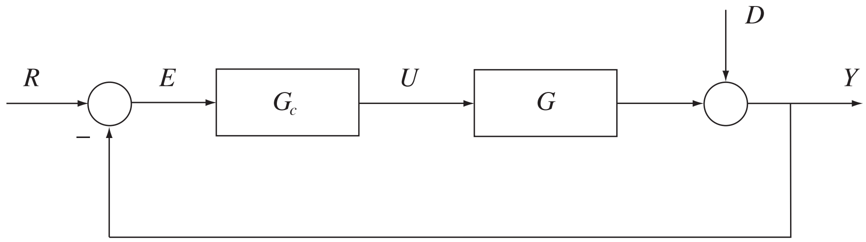

Figure 6:

and $U = \frac{n_{co}}{d_c} + G_c(R - Y) = \frac{n_{co}}{d_c} + \frac{n_c}{d_c}R - \frac{n_c}{d_c}\left(\frac{n_o}{d} + \frac{n}{d}U + D\right)$ from which $\left(1 + \frac{n_c n}{d_c d}\right)U = \frac{n_{co}d - n_c n_o}{d_c d} + \frac{n_c}{d_c}R - \frac{n_c}{d_c}D$ or

$$U = \frac{n_{co}d - n_c n_o}{dd_c + nn_c} + \frac{n_c d}{dd_c + nn_c}R - \frac{n_c d}{dd_c + nn_c}D. \tag{33}$$

Also

$$\begin{aligned} Y &= \left[\frac{n_o}{d} + \frac{n(n_{co}d - n_c n_o)}{d(dd_c + nn_c)}\right] + \frac{n_c dn}{d(dd_c + nn_c)}R + \left[1 - \frac{n_c dn}{d(dd_c + nn_c)}\right]D \\ &= \frac{d(n_o d_c + nn_{co})}{d(dd_c + nn_c)} + \frac{dn_c n}{d(dd_c + nn_c)}R + \frac{ddd_c}{d(dd_c + nn_c)}D \end{aligned}$$

or

$$Y = \frac{n_o d_c + nn_{co}}{d_k} + \frac{n_c n}{d_k}R + \frac{dd_c}{d_k}D \tag{34}$$

where

$$d_k \triangleq dd_c + nn_c. \tag{35}$$

Here d, the denominator of the plant, was cancelled in all three terms. Note that for internal stability d_k must be a Hurwitz polynomial (all roots must have strictly negative real parts). If d_k^{-1} is stable, stability is guaranteed independently of the initial conditions. Selecting n_c and d_c (with $G_c = n_c/d_c$ proper) to assign the closed-loop poles is straightforward. See the formulas that characterize all solutions of the Diophantine equation (See for example [2] section 7.2E).

Again here it is seen that all the plant poles in d are automatically cancelled when the loop is closed. That is, when feedback is applied and the loop is closed, the input to the plant u is such that all the plant modes change automatically. The closed loop characteristic polynomial has roots (closed loop eigenvalues) that are different from the poles (eigenvalues) of the plant

G for almost any G_c (unless poles of G or G_c cancel in the loop gain GG_c in which case there are uncontrollable and/or unobservable modes that cannot be altered via output feedback.)

Remarks:

i From (34) it can also be seen how to compensate for disturbances such as step disturbances $D(s) = \frac{1}{s}$ while preserving internal stability. Select $dd_c = s(\cdot)$ for the numerator in the D term with d_k remaining Hurwitz. Clearly n should not have an s and this is a condition on the plant for regulation with internal stability.

ii The error

$$E = R - Y = -\frac{n_o d_c + n n_{co}}{d_k} + \frac{dd_c}{d_k} R - \frac{dd_c}{d_k} D$$

For zero steady-state error to a step input we must have $dd_c = s^k(\cdot)$ $k \geq 1$. (That is the system Type should be 1 or greater, a well known result.)

iii If the disturbance D enters at the plant input instead of plant output, then it can be seen that the disturbance term in (34) will be $(nd_c/d_k)D$. Here, again, if $D(s) = \frac{1}{s}$ then the numerator should be chosen as $nd_c = s(\cdot)$ with d_k remaining Hurwitz.

iv Also present is the corresponding analogous property that all the controller poles in G_c are automatically cancelled when the loop is closed. This can be seen from the expression for U in (33) where d_c was cancelled in similar fashion as d was cancelled in the expression for Y in (34).

6 State Variable Representations

In this section state variable representations are used and similar results are shown. In particular, it is shown that the plant dynamics cancel out automatically when linear state feedback is used.

Consider the plant,

$$\dot{x} = Ax + Bu. \tag{36}$$

Let the linear state feedback control law be given by

$$u = -Kx + r. \tag{37}$$

The closed loop system is given by

$$\dot{x} = (A - BK)x + Br. \tag{38}$$

In the Laplace Transform domain, this becomes

$$X(s) = (sI - (A - BK))^{-1}x_0 + (sI - (A - BK))^{-1}BR(s) \tag{39}$$

where x_0 is the initial state.

The control input to the plant can be shown to be [2] (p. 327)

$$U(s) = -K(sI - (A - BK))^{-1}x_0 + (I + K(sI - A)^{-1}B)^{-1}R(s). \tag{40}$$

In order to show how the control acts on the plant, consider the open loop plant given by

$$X(s) = (sI - A)^{-1}x_0 + (sI - A)^{-1}BU(s) \tag{41}$$

Substituting the value of $U(s)$ from (40) in the above equation, after some manipulation we obtain

$$\begin{aligned} X(s) = & \ (sI - A)^{-1}\left[(sI - A)(sI - (A - BK))^{-1}\right]x_0 \\ & + \left[(sI - (A - BK))^{-1}(sI - A)\right](sI - A)^{-1}BR(s) \\ = & \ (sI - (A - BK))^{-1}x_0 + (sI - (A - BK))^{-1}BR(s) \end{aligned} \tag{42}$$

(which is exactly the result in 39). This derivation shows that the open loop dynamics included in $(sI - A)$ cancel when feedback is applied. In particular, at the input $U(s)$, the factor $(sI - A)$ is introduced and cancels with $(sI - A)^{-1}$ when $X(s)$ is generated.

The same result can be shown quite easily using polynomial matrix descriptions. In particular, consider the plant $Dz = u$ where z is the "partial state", and $u = Fz + r$ is the linear state feedback control law (here $Fz = -Kx$; D and F are polynomial matrices). The control input in the Laplace domain is

$$U = D(D - F)^{-1}R = DD_F^{-1}R. \tag{43}$$

When this control input is applied to the system, we obtain

$$Z = D^{-1}U = D^{-1}\left[D(D - F)^{-1}R\right] = (D - F)^{-1}R = D_F^{-1}R. \tag{44}$$

D, which represents the open loop dynamics, cancels out. The control input U always contains the factor D which cancels with D^{-1} of the plant. D is the inverse of the map from input U to partial state Z given by $Z = D^{-1}U$.

Remark: The linear state feedback gain may be chosen to satisfy additional requirements beyond stabilization. Such requirements may impose the restrictions that certain open loop eigenvalues should become unobservable, by canceling them with zeros (as for example is the case in the disturbance decoupling problem). In this case, some of the closed loop eigenvalues are equal to the open loop ones and so they are fixed.

Here, again, U is given by (43) and

$$Y = ND^{-1}U = (ND^{-1})(DD_F^{-1}R) = ND_F^{-1}R. \tag{45}$$

Then if $D_F = \hat{D}_F N_g$ where $N = \hat{N}N_g$,

$$Y = \hat{N}N_g(\hat{D}_F N_g)^{-1}R = \hat{N}\hat{D_F}^{-1}R. \tag{46}$$

That is, the eigenvalues in N_g (in D_F) are unobservable and cancel out in the transfer function.

7 Two Degrees of Freedom Feedback Control

Consider a general 2-degrees of freedom feedback controller

$$u = [C_y, C_r]\begin{bmatrix} y \\ r \end{bmatrix} = C\begin{bmatrix} y \\ r \end{bmatrix}$$

and the diagram in Figure 7 (see [2]-pp. 626). Note that in this and following sections lower case symbols are used for the Laplace transformed variables; this is clear from the context and hopefully no confusion arises.

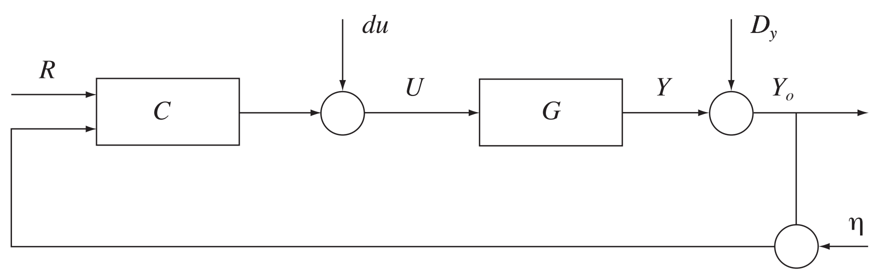

Figure 7:

Here

$$u = [C_y, C_r]\begin{bmatrix} y + d_y + \eta \\ r \end{bmatrix} + d_u. \tag{47}$$

Then

$$y = G(I - C_y G)^{-1} [C_r r + C_y d_y + C_y \eta + d_u] \tag{48}$$

$$u = (I - C_y G)^{-1} [C_r r + C_y d_y + C_y \eta + d_u] \tag{49}$$

or

$$y = Tr + (S_o - I)d_y + GQ\eta + GS_i d_u \tag{50}$$

$$u = Mr + Qd_y + Q\eta + S_i d_u \tag{51}$$

where, for $G = ND^{-1}$

$$\begin{array}{rcl} T &=& G(I - C_y G)^{-1} C_r = GM = NX \\ M &=& (I - C_y G)^{-1} C_r = DX \\ Q &=& (I - C_y G)^{-1} C_y = DL \\ S_o &=& (I - GC_y)^{-1} = I + GQ \\ S_i &=& (I - C_y G)^{-1} = I + QG \end{array} \tag{52}$$

S_i and S_o are the input and output comparison sensitivity matrices. S_i is the transfer function between u and d_u, and S_o is the transfer function between y_o and d_y as it can be seen from

$$y_o = y + d_y = Tr + S_o d_y + GQ\eta + GS_i d_u. \tag{53}$$

M is the transfer function between u and r; Q is the transfer function between u and d_y or η.

M and Q can be seen as design parameters. M is chosen primarily to satisfy response requirements between r and y, while Q is selected to satisfy feedback properties such as low sensitivity to parameter uncertainties, disturbance attenuation, etc. In the 2-degrees of freedom controller, M and Q maybe selected independently. This is not the case in more restricted configuration (see [2]–pp. 629–632) when M and Q are related. For example, in the unity feedback configuration $M = Q$.

M and Q can always be written as $M = DX$ and $Q = DL$, where D is the denominator of the plant ($G = ND^{-1}$, where D, N are co-prime polynomial matrices) and X, L are design parameters (stable rational functions for internal stability). The part of u that is due to the external input r is $u_r = Mr = DXr$ or $D^{-1}u_r = Xr$, and in view of the plant description $Dz = u$, $y = Nz$, $z_r = Xr$. That is X determines the effect of r on the plant's state z. Similarly, from $u_d = Qd_y = DLd_y$, $z_{d_y} = Ld_y$ that is L determines the effect of d_y (or η) on z.

The expressions for u and y can be written as

$$u = D\left[Xr + Ld_y + L\eta + (I + LN)D^{-1}d_u\right] \tag{54}$$

$$y = N\left[Xr + Ld_y + L\eta + (I + LN)D^{-1}d_u\right]. \tag{55}$$

This shows that no matter what r, d_y, η, d_u, C_r, and C_y are, u can always be written as

$$u = D\left[X, L, L, (I + LN)D^{-1}\right]\begin{bmatrix} r \\ d_y \\ \eta \\ d_u \end{bmatrix} = D\xi \tag{56}$$

where ξ is a signal generated by filtered combination of r, d_y, η, and d_u (note that all the filters are stable for internal stability, see also Appendix A.2).

The feedback mechanism always generates a signal u the behavior of which is modified by D, the inverse of the map D^{-1} which is the transfer function between z and u ($z = D^{-1}u$) in the plant. D appears in the numerator of the transfer function between u and r and it has the effect that for such u the behavior of the plant state $z = D^{-1}u = D^{-1}DXr = Xr$ is completely freed from behavior determined by the plant modes.

So feedback does not really generate the inverse of the plant $y = Gu$ (or of the map between y and u)—which may or may not be proper (causal) after all—but it generates the inverse of the map between z and u, namely of $z = D^{-1}u$, which always exists.

More specifically, the expression (54) points out the fact that the potential is there for the feedback to generate the whole D at u. However, depending on the choices for X and L cancellations may take place between D and the denominators of X and L.

To illustrate, consider the case when r is the only external input (d_y, d_u, and η are taken to be zero). If now C_y and C_r are chosen to stabilize the system (see [2], p. 623, Theorem 4.21 and Appendix A.2) then X is stable and can be chosen to cancel all stable poles of D. That is under stability, if $D = D_b D_g$ where D_b contains all the unstable (bad) dynamics and D_g contains all the stable (good) dynamics, $X = D_g^{-1}\widehat{X}$ will produce

$$u = Mr = DXr = D_b D_g D_g^{-1}\widehat{X}r = D_b\widehat{X}r \tag{57}$$

i.e. only D_b (the inverse of D_b^{-1}) will need to appear in u.

On the other hand, if the system $G = ND^{-1}$ is stable, for stability one can select $X = D^{-1}\widehat{X}$ in which case

$$u = Mr = DXr = DD^{-1}\widehat{X}r = \widehat{X}r \tag{58}$$

and no inverse map of D need to be generated at u. So for stability u need contain only all the bad poles of the plant as zeros.

Note that in the case when $G^{-1} = DN^{-1}$ exists and is stable, then if $X = N^{-1}\widehat{X}$

$$u = Mr = DXr = DN^{-1}\widehat{X}r = G^{-1}\widehat{X}r \tag{59}$$

that is the inverse of the plant is generated.

So, the inverse of D^{-1} is generated in u as $u = D\xi$, as in (56), always. In special cases G^{-1} is generated (assuming that the inverse of G exists and is stable). When in addition specific goals are to be satisfied, such as preserving the stable open loop poles in D_g, D_b must be generated; see (57).

In Appendix A.2 the fundamental theorems for internal stability in the 2-degrees of freedom case are given for completeness. It is also shown there that in the most general LTI case $u = Mr = DXr$ and $y = Tr = NXr$ that is control implies the cancellation of the plant dynamics (poles). This is done automatically via feedback when the loop is closed.

In summary, when the loop is closed, an inverse map is generated automatically to cancel all the pole dynamics of the plant. The particular selection of C_y, C_r will determine properties such as stability, and sensitivity, by generating new pole-zero dynamics (via X and L).

Feedback has the truly remarkable property of generating the inverse of the actual plant dynamics (of D^{-1}) exactly. In LTI, this corresponds to generating zeros (in the map from R to U) at the exact pole locations of the plant. The particular values of the feedback gains will determine the new dynamics introduced (recall that $u = DXr$ and X is stable for stability but otherwise (almost) arbitrarily chosen; see Appendix a.2); $DX = M$ must be proper.

8 Two Degree of Freedom Controllers—A Summary of the Analysis

Given a model of the process dynamics $y = Gu$, where G is the transfer function, and $u = Mr$ is the control input, then $y = G(Mr) = Tr$ where

$T = GM$ is the desired input r-output y response map. We typically choose T and M to be stable.

Let now $G = ND^{-1}$ a coprime fractional polynomial matrix representation that corresponds to the internal description $Dz = u$, $y = Nz$ with z the partial state. It is known that to obtain the maps T and M with internal stability, $T = NX$ and $M = DX$ where X is stable (see Appendix A.2). Note that the desired dynamics are introduced via X and the existing dynamics are cancelled via D.

The input $u = Mr = DXr$ can certainly be implemented via open-loop. In fact the two-degrees of freedom controller formulation allows that. The examples in previous sections show the difficulties associated with open-loop control when uncertainties in the process parameters and in the exogenous influences—initial conditions, external disturbances—are present. Note also the amount of dynamics in $M = DX$ that are necessary to be generated, include all the plant dynamics in the denominator of M (D) in addition to all desired dynamics (in X).

The control action $u = Mr$ can be generated via a combination of feedforward and feedback actions. It corresponds to appropriately selecting the design parameters L and X, see Appendix A.2.

Now, the amazing fact is that feedback generates automatically absolutely exact models of the existing dynamics in D. This was shown using the two degree of freedom controller configuration above that contains disturbances and noise signals. Note that as it is well known internal stability in the closed-loop system may be guaranteed by requiring that certain maps between appropriate signals be stable; in view of this we omit the initial conditions in the expressions without loss of generality.

In the expressions for u and y in (54) and (55), first notice that $u = D(.)$, that is $D^{-1}u = z = (.)$ a function of the external inputs and disturbances. In (.), the stable design parameter $X = D^{-1}M$ contains the desired dynamics of the r to y response as discussed above, while $L = D^{-1}Q$ and also $(I + LN)D^{-1}$ must be stable for internal stability (in [2]–p. 625). Additional loop properties such as sensitivity may be addressed by selecting L. Furthermore by selecting L we can reduce the effect of the disturbances from the output y and other signals in the loop.

It can be shown that $Xr = z$ and so $Tr = NXr = Nz = y$ and $Mr = DXr = Dz = u$. A moment's reflection reveals that the control input $u = Dz (= Mr)$, **implements the inverse of the u to z (input to state) map D^{-1}** ($z = D^{-1}u$). Certainly, there are cases where u can insert an (exact or approximate) inverse of the plant that is of the u, y map G. To see this, assume G is invertible, and let $X = N^{-1}X_N$.

Then, $u = DXr = DN^{-1}X_N r = G^{-1}X_N r$. Clearly inverting the plant is a special case that requires conditions on G and a specially chosen X. **What u always implements** is the inverse of $z = D^{-1}u$ which contains all needed information about the plant important dynamics. Recall that the poles produce the dominant characteristics of the response—they appear as exponents of the exponential terms in the modes—while the zeros play an important but secondary role, that of characterizing the coefficients in the modes. So the zeros cannot introduce new exponential terms but they can only reduce the effect or eliminate the effect of existing ones via pole-zero cancellations (see Appendix A.1). So control focuses on the poles, which are included in D. In the figure, $u = Mr$ with $M = DX$ where D is generated

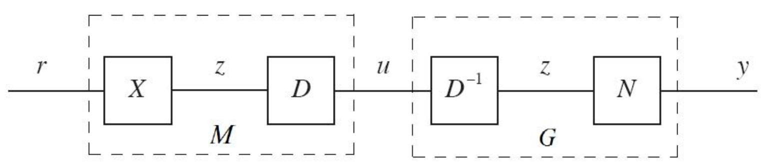

Figure 8:

automatically as soon as the loop closes while X is chosen to assign new desired dynamics.

9 Discussion

FACT: Any arbitrary feedback will result into a closed-loop system with dynamics (poles) different from the open-loop dynamics (poles). This is true under the assumptions of controllability and observability and for any initial conditions. This was demonstrated in the previous sections of this paper using internal polynomial descriptions and transfer function factorizations. It can also be demonstrated as well (although not as easily) via state space representations as it was done in section 1.6 and via other methods as it is briefly shown below.

Given

$$\dot{x} = Ax + Bu, \quad y = Cx + Du.$$

Consider a linear state feedback control law

$$u = -Kx + r$$

or a constant output feedback control

$$u = -HCx + r.$$

Then $A - BK$ or $A - BHC$ define the closed-loop dynamics.

It is known that the eigenvalues of $A - BHC$ will be different from the eigenvalues of A for almost any H; in fact the set of gain H that preserve the eigenvalues of A in $A - BHC$ has measure zero (they are roots of a multivariate polynomial in the gains in H). This is under the assumptions that (A, B) and (A, C) are controllable and observable respectively. If they are not, it is known that the uncontrollable and/or unobservable eigenvalues of the system cannot be altered and they will appear as eigenvalues of $A - BHC$. Corresponding results exist for $A - BK$ where assuming that (A, B) is controllable its eigenvalues are different from the eigenvalues of A for almost any K.

This fact can also be seen from the Root Locus where any feedback gain other than zero assigns the closed-loop poles at locations different from the open-loop pole locations.

In the Diophantine equation under the assumptions of controllability and observability controllers will assign the closed-loop poles to different locations from the open-loop poles almost always. In fact, in the multi-input multi-output case, one can characterize all the controllers that will result to a closed-loop system with poles that contain all the open-loop poles. This can be done for example using the methodology in (see [2]–Appendix) which is based on Polynomial Matrix Interpolation.

So it is a fact that when the loop closes the dynamics are completely reassigned for almost any feedback gains, that is the plant behavior drastically changes.

Remark: It is worth mentioning at this point that when H has a special structure—for example some (block) diagonal structure as is the case in decentralized control even in the case of the system being controllable and observable the characteristic polynomial of $A - BHC$ may contain some fixed zeros which do not change with H.

9.1 Open vs Closed Loop Control

Summarizing, a comparison of open and closed loop control is given and their characteristics are briefly described.

9.1.1 Change of Plant Dynamics

To control a plant, the input u should be so that the undesirable dynamics are somehow cancelled. The mechanism can be seen as a pole/zero cancellation mechanism which takes place automatically in feedback control. In the open-loop case, for the changes to take place the controller needs to know the exact pole locations and the exact initial conditions of the plant so that exact cancellation of dynamics can take place. Although in the closed-loop case all the plant dynamics are changed for almost all controllers, in the open-loop case is the exact opposite and the plant dynamics do NOT change for almost any controller.

So, the complete change of plant dynamics happens

a. almost always in closed-loop feedback control

b. almost never in open-loop feedforward control.

9.1.2 Assigning Desirable Dynamics to the Compensated System

The reason for using control is to the make the plant behave in a desirable manner. So one needs more than just closing the loop, an action that can send the poles in an undesirable region (unstable region for example) with disastrous consequences. If the undesirable plant dynamics have been cancelled, the assignment of new dynamics is much easier in the open-loop case.

The complete change of the dynamics of the compensated system to desirable dynamics is

a. easier in the open-loop feedforward control case

b. relatively harder in the closed-loop feedback case, although typically there is a large range of controller choices that satisfy the requirements for desirable dynamics (control specifications) as one must consider the trade-offs.

In the open-loop after the current plant dynamics have been cancelled out (which is the difficult part) one can simply choose the compensated system dynamics by assigning them to the controller. So the open-loop controller should contain all the desirable dynamics.

In the closed-loop control the choice of the appropriate controller is a nontrivial matter and the field of control theory has been studying this problem intensively for at least the past 50 years. Certainly it is a topic that requires deep understanding. The following table summarizes the above comments.

	Easier	Harder
Open Loop	Assign New Dynamics	Cancel Existing Dynamics
Closed Loop	Cancel Existing Dynamics	Assign New Dynamics

10 On the Role of the Return Difference

The return difference relation forces the cancellation of all plant (and controller) poles. It can be seen as the underlying cause of the fundamental feedback property of canceling automatically the open loop plant dynamics.

Consider the unity feedback configuration as discussed in the section, Open- and Closed-loop Control.

When the initial conditions and the external inputs are zero then, by considering the signals at the output of the plant, the return difference relation is $Y = -GG_cY$ or $(1 + GG_c)Y = 0$.

Considering initial conditions and the external input R, with $G_c = \frac{n_c}{d_c}$, $G = \frac{n}{d}$ then:

$$Y = \frac{n_0}{d} + GU, U = \frac{n_{c0}}{d_c} + G_c(R - Y)$$

from which the return difference relation now becomes

$$Y = \frac{n_0}{d} + G\left[\frac{n_{c0}}{d_c} + G_c(R - Y)\right]$$

or

$$Y = -GG_cY + GG_cR + \left[\frac{n_0}{d} + G\frac{n_{c0}}{d_c}\right]$$

Then

$$(1 + GG_c)Y = \frac{n_0 d_c + nn_{c0}}{dd_c} + GG_cR$$

$$Y = \frac{dd_c}{dd_c + nn_c}\frac{n_0 d_c + nn_{c0}}{dd_c} + \frac{dd_c nn_c}{dd_c(dd_c + nn_c)}R$$

This last relation shows that dd_c cancels throughout to obtain

$$Y = \frac{n_0 d_c + nn_{c0}}{dd_c + nn_c} + \frac{nn_c}{dd_c + nn_c} R$$

That is, in order to satisfy the conditions on Y imposed by the feedback interconnection and expressed in terms of the return difference relations, dd_c must cancel and the new closed loop system has new dynamics imposed by $dd_c + nn_c$ (a Diophantine equation) instead of d (and d_c).

Imposing the conditions of the return difference also causes an automatic change in the gain as discussed in the introduction. Consider Fig. 1.5 with $k = 1$ and $G(s) = A$.

Here the return difference relations for the signals in the loop are

$$U = -AU + R, \quad Y = -AY + AR$$

The term $-AY$ in the Y equation represents the feedback signal. For any input R the signal Y in the output must equal the feedback signal $-AY$ and the signal R through the plant, AR. For large A this can only happen when $Y \approx R$ since the left hand side (Y) is much smaller than AY. In general, from $(1+A)Y = AR$ one can see that $Y < R$ or the gain from R to Y is less than 1 (since $A > 0$, $\frac{A}{1+A} < 1$ always). So the conditions imposed by the return difference relations cause the overall gain to be less than 1. Similarly, they cause the sensitivity $S = \frac{1}{1+A}$ to be less than 1 as well for any $A > 0$.

This automatic reduction in gain is the cause of lower sensitivity in the feedback loop as opposed to the open loop. This is caused by the return difference conditions imposed on the closed loop system by the feedback interconnection.

References

[1] G. Goodwin, S. Graebe, and M. Salgato, *Control System Design*, Prentice Hall, 2001.

[2] P.J. Antsaklis and A.N. Michel, *Linear Systems*, Birkhauser, 2005.

[3] G.F. Franklin, J.D. Powell and A. Emami-Naeini, *Feedback Control of Dynamic Systems*, 5/e, Pearson Prentice Hall, 2006.

[4] B.C. Kuo, *Automatic Control Systems*, 7th edition, Prentice-Hall, 1995.

[5] K. Ogata, *Modern Control Engineering*, 3rd edition, Prentice-Hall, 1997.

[6] R.C. Dorf and R.H. Bishop, *Modern Control Systems*, 10/e, Pearson Prentice-Hall, 2005.

[7] K.J. Astrom, R.M. Murray, *Feedback Systems: An Introduction for Scientists and Engineers*, Princeton University Press, 2008.

[8] J. Truxal, *Control System Synthesis*, McGraw-Hill, 1955.

[9] G.C. Newton, L.A. Gould, J.F. Kaiser, *Analytical Design of Linear Feedback Controls*, Wiley, 1957.

[10] W.A. Wolovich, *Linear Multivariable Systems*, Springer-Verlag, 1974.

[11] T. Kailath, *Linear Systems*, Prentice-Hall, 1980.

[12] O. R. Gonzalez and P. J. Antsaklis, "Internal Models Over Rings," Linear Circuits, Systems and Signal Processing: Theory and Application, C. I. Byrnes, C. F. Martin and R. E. Saeks, Eds., Chapter 1, pp. 41-48, Elsevier Science Pubs. R. V., North Holland, 1988.

[13] O. R. Gonzalez and P. J. Antsaklis, "Internal Models in Regulation, Stabilization and Tracking," Intern. Journal of Control, Vol 53, No 2, pp 411-430, 1991.

Appendix

A The Pole/Zero Cancellation Mechanism: A Review

The simplest way to describe the effect of the pole-zero cancellation mechanism on the system response is to start with a transfer function or polynomial description of the system. This is done first and then the results are also seen directly in the time domain using state variable descriptions. A specific example is used for clarity. The same principles apply to the general case.

Let the plant and the controller be given by

$$G(s) = \frac{1}{s+1}, \quad G_c(s) = \frac{k(as+1)}{s+10} \tag{60}$$

connected in series as in Fig. 9. The transfer function between Y and R is

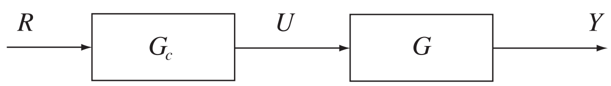

Figure 9:

$$Y/R = GG_c = \frac{k(as+1)}{(s+1)(s+10)} \tag{61}$$

Its inverse Laplace transform, which is the impulse response of the closed-loop system, is given by

$$h(t) = \left[\frac{k}{9}(1-a)e^{-t} - \frac{k}{9}(1-10a)e^{-10t}\right] 1(t) \tag{62}$$

where $1(t)$ denotes the unit step function. Note that the position of the zero (at $-1/a$) affects the behavior, not directly, but via the coefficients of the modes.

Now as a approaches 1 the coefficient of the e^{-t} mode becomes smaller and the effect of the mode e^{-t} on $h(t)$ diminishes. When $a = 1$, a pole/zero (zero at -1) cancellation occurs and the e^{-t} mode disappears from $h(t)$; in that case $h(t) = \left[ke^{-10t}\right] 1(t)$. Similarly when $a = \cdot 1$ a pole/zero cancellation (zero at -10) occurs and $h(t) = \left[\cdot 1 ke^{-t}\right] 1(t)$.

In the above, if R is a constant r_o, that is $r(t) = r_o\delta(t)$, then $U = G_cR = G_cr_o$ can be seen as a signal $u(t) = kr_o\left[a\delta(t) + (1-10a)e^{-10t}\right] 1(t)$ acting on the system $G(s)$ and producing the pole/zero cancellation or the zeroing of the mode coefficient effects.

If $R(s) = 1/s$ (for example), a unit step input $r(t) = 1(t)$ then the plant output will be

$$y(t) = \left[-\frac{k(1-a)}{9} e^{-t} + \frac{k(1-10a)}{90} e^{-10t} + \frac{k}{10} \right] 1(t) \tag{63}$$

and similar effects are observed when pole/zero cancellations occur, in the cases when $a = 1$ and $a = \cdot 1$.

These effects can be seen rather easily using polynomial descriptions for the plant ($q \triangleq d/dt$, the differential operator)

$$(q+1)z(t) = u(t), \quad y(t) = z(t) \tag{64}$$

and the controller

$$(q+10)z_c(t) = r(t), \quad u(t) = k(aq+1)z_c(t). \tag{65}$$

The overall system description is then

$$\begin{gathered} \begin{bmatrix} q+10 & , & 0 \\ -k(aq+1) & , & q+1 \end{bmatrix} \begin{bmatrix} z_c(t) \\ z(t) \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} r(t) \\ y = \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} z_c(t) \\ z(t) \end{bmatrix} \end{gathered} \tag{66}$$

from which when $a = 1$ the -1 eigenvalue is uncontrollable and when $a = \cdot 1$ the -10 eigenvalue is unobservable (see [1], Sect. 3.4). Whether an eigenvalue is uncontrollable or unobservable depends on the polynomial realizations selected. They were chosen here to match the state space development below where controllable realizations are used.

The interpretation of pole/zero cancellations using state variable descriptions is as follows:

The plant $G(s) = \frac{1}{s+1}$ is described by

$$\dot{x} = -x + u, \quad y = x \tag{67}$$

and the controller $G_c(s) = \frac{k(as+1)}{s+10}$ is described by

$$\dot{x}_c = -10x_c + r, \quad u = k(1-10a)x_c + kar. \tag{68}$$

The description of the overall system $\{A, B, C, D\}$ is then

$$\begin{bmatrix} \dot{x}_c \\ \dot{x} \end{bmatrix} = \begin{bmatrix} -10 & 0 \\ k(1-10a) & -1 \end{bmatrix} \begin{bmatrix} x_c \\ x \end{bmatrix} + \begin{bmatrix} 1 \\ ka \end{bmatrix} r$$
$$y = [0 \quad 1] \begin{bmatrix} x_c \\ x \end{bmatrix} \tag{69}$$

From the controllability matrix

$$\mathcal{C} = [B,\ AB] = \begin{bmatrix} 1 & -10 \\ ka & k-11ka \end{bmatrix} \tag{70}$$

$|\mathcal{C}| = k(1-11a)+10ka = k(1-a)$. So for $a = 1$ the system is uncontrollable. In fact, the uncontrollable eigenvalue is at -1 (see [2]–Sect. 3.4). From the observability matrix

$$\mathcal{O} = \begin{bmatrix} C \\ CA \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ k(1-10a) & -1 \end{bmatrix} \tag{71}$$

$|\mathcal{O}| = -k(1-10a)$. So for $a = \cdot 1$ the system is unobservable. In fact, the unobservable eigenvalue is at -10 (see [2]–Sect. 3.4). Again here in the state space setting, pole/zero cancellations can be seen as cancellations between eigenvalues and input or output decoupling zeros see (see [2]–Sect. 3.4) when the cancelled eigenvalues (modes) become uncontrollable or unobservable.

In summary, what we conveniently describe as pole/zero cancellation (in the frequency domain) is a fundamental mechanism of drastically altering the behavior of a system by zeroing the coefficient of the corresponding mode (in the time domain). The internal description interpretation is that a pole/zero cancellation is making the mode (or the corresponding eigenvalue) uncontrollable from an input or unobservable from an output and so invisible from an input/output point of view.

A.0.3 The Effects of Uncertainties in Pole Locations

If the pole of the plant is not exactly at -1 but at $-(1+\varepsilon)$ then the term in (63) that involves this pole of the plant becomes

$$y_1(t) = -\frac{k(1-a+a\varepsilon)}{(9-\varepsilon)(1+\varepsilon)} e^{-(1+\varepsilon)t}. \tag{72}$$

It is then clear that if $a = 1$ the coefficient will not become zero but will be $-k(a\varepsilon)/(9-\varepsilon)(1+\varepsilon)$. For ϵ very small the effect of this mode will still be almost negligible. If, however, the pole of the plant were unstable, say at $+1$ instead of -1, then the mode in this case would be $(\cdot)e^{(1+\varepsilon)t}$ and no matter how small the coefficient is, given enough time the term will grow and so the system is unstable.

In summary, pole/zero cancellation of unstable poles will not work because of the inherent uncertainties in the pole location of the system. Even if the location of the unstable poles were known exactly pole/zero cancellation would not typically produce a stable system because of uncertainties in the initial conditions since the cancelled unstable poles become uncontrollable/unobservable modes (they do not really disappear) and they can be excited by initial conditions.

B Fundamental Theorems for Internal Stability

Let $G = ND^{-1}$ be the proper transfer function of the plant; N and D are right coprime polynomial matrices. Let a desirable stable transfer funcion be T, $y = Tr$, obtained using control $u = Mr$, where M is also stable. Proofs for the following theorems may be found in ([2]–Chapter 7 pp. 627-629).

Theorem 1. *The stable rational function matrices T and M are realizable via a two degrees of freedom control configuration with internal stability if and only if there exists stable X so that*

$$\begin{bmatrix} T \\ M \end{bmatrix} = \begin{bmatrix} N \\ D \end{bmatrix} X \tag{73}$$

Theorem 2. *T, $M \in RH_\infty$ are realizable with internal stability by means of a two degrees of freedom control configuration if and only if there exists $X' \in RH_\infty$ so that*

$$\begin{bmatrix} T \\ M \end{bmatrix} = \begin{bmatrix} N' \\ D' \end{bmatrix} X' \tag{74}$$

Here RH_∞ denotes the set of proper and stable rational function matrices. Let also S denote the desired stable sensitivity matrix (it is S_o in (1.52) in section 1.7).

Theorem 3. *T, M, $S \in RH_\infty$ are realizable with internal stability by a two degrees of freedom control configuration if and only if there exists X', $L' \in RH_\infty$ so that*

$$\begin{bmatrix} T \\ M \\ S \end{bmatrix} = \begin{bmatrix} N' & 0 \\ D' & 0 \\ 0 & N' \end{bmatrix} \begin{bmatrix} X' \\ L' \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ I \end{bmatrix}, \tag{75}$$

where $(I + L'N')D'^{-1} \in RH_\infty$*. Similarly,* T*,* M*,* $Q \in RH_\infty$ *(see section 1.7) are realizable if and only if there exists* X'*,* $L' \in RH_\infty$ *so that*

$$\begin{bmatrix} T \\ M \\ Q \end{bmatrix} = \begin{bmatrix} N' & 0 \\ D' & 0 \\ 0 & D' \end{bmatrix} \begin{bmatrix} X' \\ L' \end{bmatrix}, \tag{76}$$

where $(I + L'N')D'^{-1} \in RH_\infty$*.*

D in $u = DXr$ introduces zeros in the transfer function which cancel out to produce a desired response as in:

$$T = GMR = ND^{-1}DXr = NXr. \tag{77}$$

The following two theorems are the basic internal stability theorems for feedback control (see [2]–pp. 623-625). They give parameterizations of all stabilizing controllers and show that via two design parameters X,L (or M,Q) the feedforward and feedback control actions can be appropriately assigned.

Theorem 4. *Let the plant* $y = Gu$ *have a proper transfer function and let*

$$u = C \begin{bmatrix} y \\ r \end{bmatrix} = [C_y \quad C_r] \begin{bmatrix} \hat{y} \\ \hat{r} \end{bmatrix} \tag{78}$$

be a proper 2-degrees of freedom controller. Let $\det(I - C_yG) \neq 0$*. The closed-loop system in internally stable if and only if*

1. $\hat{u} = C_y\hat{y}$ *internally stabilizes the system* $\hat{y} = G\hat{u}$*,*

2. C_r *is such that the rational matrix* $M = (I - C_yG)^{-1}C_r$ *satisfies* $D^{-1}M = X$*, a stable rational matrix, where* C_y *satisfies (1) and* $G = ND^{-1}$ *is a right coprime polynomial matrix factorization.*

Theorem 5. *Given that the plant* $y = Gu$ *is proper with* $G = ND^{-1} = \widetilde{D}^{-1}\widetilde{N}$ *doubly coprime polynomial MFDs, all internally stabilizing proper controllers* C *in* $u = C\begin{bmatrix} y \\ r \end{bmatrix}$ *are given by:*

$$C = (I + QH)^{-1}[Q, M] = [(I + LN)D^{-1}]^{-1}[L, X], \tag{79}$$

where $Q = KD$ *and* $M = DX$ *are proper with* L*,* X*, and* $D^{-1}(I + QH) = (I + LN)D^{-1}$ *stable, so that* $(I + QH)^{-1}$ *exists and is proper; or by*

$$C = (X_1 - K\widetilde{N})^{-1}[-X_2 + K\widetilde{D}, X], \tag{80}$$

where K and X are stable so that $(X_1 - K\widetilde{N})^{-1}$ exists and C is proper. Also X_1 and X_2 are determined from

$$UU^{-1} = \begin{bmatrix} X_1 & X_2 \\ -\widetilde{N} & \widetilde{D} \end{bmatrix} \begin{bmatrix} D & -X_2 \\ N & X_1 \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}$$

with U unimodular.

If $G = N'D'^{-1} = \widetilde{D}'^{-1}\widetilde{N}'$ are doubly coprime MFDs in RH_∞, then all stabilizing proper C are given by

$$C = (X_1' - K'\widetilde{N}')^{-1}[-X_2' + K'\widetilde{D}_1'), X'], \tag{81}$$

where K', $X' \in RH_\infty$ so that $(X_1' - K'\widetilde{N}')^{-1}$ exists and is proper. Also

$$U'U'^{-1} = \begin{bmatrix} X_1' & X_2' \\ -\widetilde{N}' & \widetilde{D}' \end{bmatrix} \begin{bmatrix} D' & -X_2' \\ N' & X_1' \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}$$

with U', $U'^{-1} \in RH_\infty$; or by

$$C = (I + QH)^{-1}[Q, M] = [(I + L'N')D'^{-1}]^{-1}[L', X'], \tag{82}$$

where $Q = D'L'$, $M = D'X' \in RH_\infty$ with L', X', and $D'^{-1}(I + QH) = (I + L'N')D'^{-1} \in RH_\infty$ and so that $(I + QH)^{-1}$ or $(I + L'N')^{-1}$ exists and is proper.

C Comments on Stability Robustness

Given a plant $G(s)$ there is an infinite number of controllers $G_c(s)$ that stabilize the system in a feedback configuration. They are all given by the Youla parameterization (see [2]–p. 615). To get a sense of how robust the stability of a system is, note that all plants $G(s)$ that are stabilized by a fixed constant gain controller $G_c = k$ in a feedback configuration is given by

$$G = Q(1 - kQ)^{-1} \tag{83}$$

where Q is *any proper and stable matrix, which is a large class of plants.* The closed loop transfer function Y/R in Fig. 5 is then given by

$$Y/R = T = \frac{kG}{1 + kG} = \frac{kQ}{(1 - kQ) + kQ} = kQ \tag{84}$$

and

$$U/R = M = \frac{k}{1 + kG} = \frac{k(1 - kQ)}{(1 - kQ) + kQ} = k(1 - kQ). \tag{85}$$

Note that the sensitivity S to parameter variations is

$$S = \frac{1}{1+kG} = \frac{1-kQ}{(1-kQ)+kQ} = 1 - kQ. \tag{86}$$

The poles of Q are the closed-loop poles. If, for example, Q is selected to be $Q = 1/(as+1+k)$ then (83) becomes $G = \frac{1}{as+1}$ which was used above, in sections 1.3 and 1.4. If Q is chosen to be stable (that is $as+1+k$ Hurwitz or $(1+k)/a > 0$) then the closed-loop is stable as before. (In this case $S = \frac{as+1}{as+1+k}$ which for all $k > 0$ is less than 1 as desired.)

D DRAFT - Internal Models–Stability and Regulation

D.1 Review of Stability in Unity Feedback Configuration

Given a plant P and a controller C, a unity feedback control system is internally stable if and only if $(1+PC)^{-1}$, $(1+PC)^{-1}P$ and $C(1+PC)^{-1}$ are stable (see [2]–p. 584). Having only $(1+PC)^{-1}$ stable does not guarantee that cancellation of unstable P and C poles will not take place, and so we need all conditions. If P and C are stable then internal stability is guaranteed iff $(I+PC)^{-1}$ is stable.

The closed loop characteristic polynomial is $d(1+PC)d_c$ and so a system internally stable if and only if $(d(1+PC)d_c)^{-1}$ is stable. Going from internal polynomial matrices descriptions is the best way to prove these results for the MIMO case as well.

Internal Models: It can be shown that when $(1+PC)^{-1}$ is stable, $(1+PC)^{-1}P$ is stable if and only if no bad poles of P cancel in $1+PC$. This in turn is true if and only if $1+PC$ has an internal model of P.

We will say (not very precisely) that B has an internal model of A if all the bad (undesirable) poles of A are poles of B. For a careful definition (see [13]). So if $A = \frac{s+1}{(s+2)(s-1)}$ then $B = \frac{1}{(s-1)(s+4)}$ has an internal model of A.

The system is internally stable iff the following 3 conditions hold:

1. $(1+PC)^{-1}$ is stable or the zeros of $(1+PC)$ are stable
2. $(1+PC)^{-1}P$ is stable or no cancellations of unstable poles of P take place in (1+PC) , or $(1+PC)$ has an internal model of P
3. $C(1+PC)^{-1}$ is stable or no cancellations of unstable poles of C take place in $(1+PC)$, or $(1+PC)$ has an internal model of C.

Remarks: In 1 above, $(1+PC)^{-1}$ stable is not difficult to satisfy. Use the loop gain $PC(jw)$ and Nyquist to stabilize the system. In 2 above, it is easy to guarantee that no cancellations of bad poles of P occur in the loop gain PC. In 3 above, it is easy to guarantee that no cancellations of bad poles of C occur in the loop gain PC.

Note that $u = Mr = C(1+PC)^{-1}r$. When $(1+PC)^{-1}$ is stable and $(1+PC)$ has an internal model of P then M is has as zeros the unstable poles of P and if $(1+PC)$ has an internal model of C then M is stable and PM is stable.

Regulation: Conditions are best shown using factorizations and internal descriptions. Here Dw is an external disturbance $(1+PC)^{-1}Dw$ must be stable. When $(1+PC)^{-1}$ is stable (which it is when the system is internally stable) then regulation takes place iff $(1+PC)$ has an internal model of Dw.

In [12],[13] it is made very clear that an internal model always exists in the transfer function between the measured output and the plant output. In the unity feedback case this is the return difference $(1+PC)$!

Open Loop: In open loop, we must have PM stable, that is $D^{-1}M$ stable or M^{-1} should have an internal model of P (or PM does not have an internal model of P). So stability iff M stable and M^{-1} has an internal model of P.

Remarks: M stable is easy to satisfy. M^{-1} having an internal model of P is almost impossible to satisfy since the unstable poles of P are rarely known exactly.

The controller M has two roles to play. First it must cancel out all unstable (bad) poles of P in PM-by having these bad poles of P as its zeros, or having M^{-1} have an internal model of P (when M comes for feedback control this part is guaranteed from $(1+PC)$ having an internal model of P). Secondly M must introduce the desirable dynamics (when M comes from feedback M is stable if $(1+PC)$ has an internal model of C-so no unstable poles of C appear in M^{-1} and $(1+PC)^{-1}$ is stable).

Open vs Closed Loop Control

We cannot stabilize via open loop control because of our inability to create an exact internal model of P in the open loop controller. This is very easy to do in closed loop as it corresponds to simply not allowing cancellations of unstable poles of P in PC.

What about initial conditions? From (42) the expression

$$(sI-A)^{-1}\left[(sI-A)(sI-(A-BK))^{-1}\right]x_0$$

has an internal model of the plant, which is generated automatically.

E Connections to Internal Feedback of a Given Plant (FIR and IIR) and to Recursive Relations

In this section, we examine feedback structures internal to given plants. Consider the system

$$\dot{x} = Ax + Bu$$

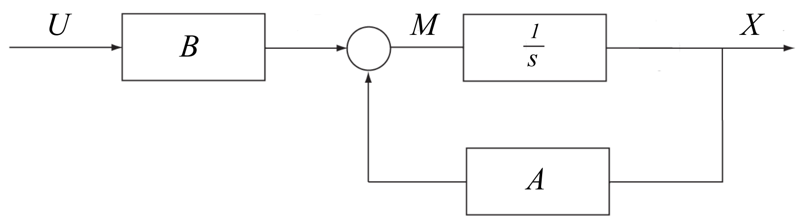

Figure 10:

with initial conditions x_0. From the figure, it is clear that

$$M = BU + AX = BU + A(\frac{1}{s}M + \frac{1}{s}x_0)$$

where U denotes the Laplace transform of u. This is so, because $\dot{x} = m$ from which $sX - x_0 = M$ and $X = \frac{1}{s}M + \frac{1}{s}x_0$. Then

$$(I - A\frac{1}{s})M = BU + A\frac{1}{s}x_0.$$

Note that $(I - A\frac{1}{s})$ is the return difference. Then

$$M = (I - A\frac{1}{s})^{-1}BU + (I - A\frac{1}{s})^{-1}A\frac{1}{s}x_0 = s(sI - A)^{-1}BU + (sI - A)^{-1}Ax_0$$

Substituting in $X = \frac{1}{s}M + \frac{1}{s}x_0$, we obtain

$$\begin{aligned} X &= \frac{1}{s}\left[s(sI - A)^{-1}BU + (sI - A)^{-1}Ax_0\right] + \frac{1}{s}x_0 \\ &= \frac{1}{s}\left[s(sI - A)^{-1}BU\right] + \frac{1}{s}(sI - A)^{-1}\left[A + sI - A\right]x_0 \\ &= (sI - A)^{-1}BU + (sI - A)^{-1}x_0. \end{aligned}$$

So, sI, which is introduced in the last relation by the term that depends on U and the term that depends on x_0, cancels with the pole of $\frac{1}{s}I$ as expected ($X = \frac{1}{s}M$ is the transfer function in the forward path). The new eigenvalues (eigenvalues of A) are (almost always) completely different from the eigenvalues of $\dot{x} = m$ at zero. That is, in this internal feedback case, A plays the role of feedback gain and similar results to the results derived for external feedback are obtained. Specifically, the input M to the open loop plant $\frac{1}{s}I$ is such that the open loop poles (at the origin) cancel automatically.

F Nonlinear Systems

The fundamental feedback property discussed above is applicable to more general systems as well, for example to nonlinear systems: $\dot{x} = f(x, u)$. When feedback $u = h(x, r)$ is applied, the closed loop system is $\dot{x} = f(x, h(x, r))$ which almost always, for almost all h, has different behavior from the open-loop system $\dot{x} = f(x, u)$ assuming the plant is controllable (see [1]–pp. 29-36).

G Sampled Data Systems

The plant poles also cancel in the case of sampled data systems as the following example illustrates

Example 1 Consider the plant $G(s) = \frac{1}{s}$ in a feedback configuration.

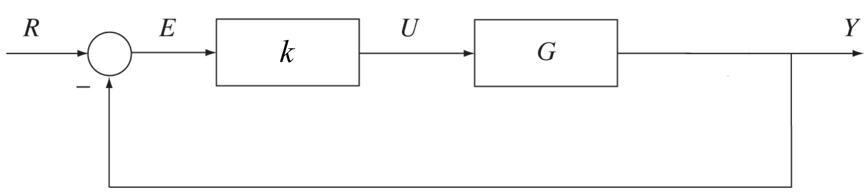

Figure 11:

The overall transfer function is found from

$$Y = GU = GkE = kG(R - Y)$$

from which

$$(1 + kG)Y = kGR$$

and

$$\frac{Y}{R} = \frac{kG}{1 + kG} = \frac{k}{s + k}$$

The closed-loop system has a pole at $-k$ and it is stable for all $k > 0$.

Now if a ZOH and an ideal sampler are used,

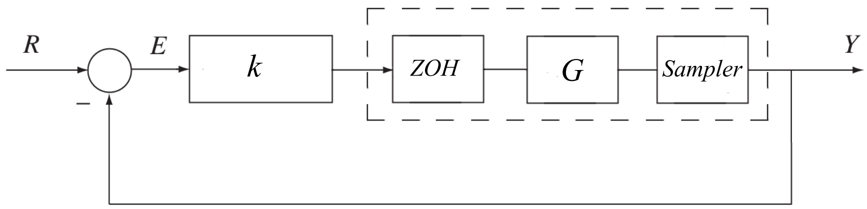

Figure 12:

then

$$\begin{aligned} H(z) &= (1 - z^{-1})Z\{[L^{-1}(\frac{G(s)}{s})]_{t=kT}\} \\ &= (1 - z^{-1})Z\{[L^{-1}(\frac{1}{s^2})]_{t=kT}\} \\ &= (1 - z^{-1})Z\{(kT)\} \\ &= (1 - z^{-1})\frac{Tz}{(z-1)^2} = \frac{T}{z-1} \end{aligned}$$

The closed-loop transfer function is $\frac{kT}{z-1+kT}$, and so the system is stable when

$$\begin{aligned} |1 - kT| &< 1 \\ -1 < 1 - kT &< 1 \\ -2 < -kT &< 0 \\ 0 < k &< \frac{2}{T} \end{aligned}$$

which is much more restrictive than before ($k > 0$). As T becomes larger, the range for acceptable K becomes smaller as expected.

The control input is

$$U(z) = \frac{k}{1 + kH(z)}R(z) = \frac{k(z-1)}{z-1+kT}R(z)$$

and

$$Y(z) = H(z)U(z) = \frac{T}{z-1}\frac{k(z-1)}{z-1+kT}R(z) = \frac{kT}{z-1+kT}R(z)$$

Notice that $U(z)$ as it acts on the plant $H(z)$ cancels the plant dynamics by the pole/zero cancellation of $z-1$. The remarkable fact is that this cancellation happens independently of the sampling period T. Even when T is large, as long as T is finite, the feedback cancels the plant dynamics.

Example 2 consider

$$G(s) = \frac{a}{s+a}$$

in a unity feedback configuration with gain k. For $a > 0$, the closed loop is stable when $k > -1$. When $a < 0$ ($G(s)$ is unstable), the closed loop is stable for $k < -1$. The corresponding sampled data system with T as the sampling period is given by

$$H(z) = \frac{1 - e^{-aT}}{z - e^{-aT}}.$$

The closed loop system is then stable for

$$-1 < k < \frac{1 + e^{-aT}}{1 - e^{-aT}}$$

when $a > 0$ (case when open loop is stable). As T becomes larger the right hand side goes to 1. When $a < 0$ (open loop is unstable), the closed loop system is stable when k satisfies

$$\frac{1 + e^{-aT}}{1 - e^{-aT}} \leq k < -1.$$

As T becomes larger, the left hand side goes to -1 and clearly the range of k for stability is much reduced.

The examples illustrate what happens in the discrete time case with respect to the effect of feeedback; they also show how the stability range for k is reduced in the sampled data case compared to the continuous case. In the discrete time case, every instant of time $k, k+1, K = 2, ...$ the input to the plant is such that the plant dynamics cancel out automatically. This can also be shown directly using the transfer function $H(z)$ and the numerator and denominator polynomials (polynomial matrices) $N(z)$, $D(z)$ or via the state variable description $x(k+1) = Ax(k) + Bu(k)$ in a manner completely analogous to the continuous case.

In the sampled data case, the cancelations happen every T units of time. In between time instants $T, 2T, ...$ the continuous plant is running open loop and no plant pole cancelations are taking place. As T becomes larger, the plant is running open loop for longer time and so it is harder to stabilize it by applying feedback at such infrequent instants in time.

H Systems With Delays

Similar results can be shown for systems with delays. Note that the delay e^{-Ts} can be approximated by (first order approximations) $\frac{1-\frac{T}{2}s}{1+\frac{T}{2}s}$ and if this is done it is clear that similar results regarding open loop pole cancelations, when feedback is used, can be derived. These results can be shown directly.

As an example, consider

$$G(s) = \frac{1}{s+1} \tag{87}$$

in a unity feedback configuration with gain k. Without delay, if k satisfies $-1 < k$, the system is stable. With delay T, k must approximately satisfy $-1 < k < 1 + \frac{2}{T}$ for stability. As T becomes larger, the range of k becomes smaller.

I Discrete Event Systems

Discrete Event Systems (DES) are typically modeled using automata or Petri Nets (also logic, if-then statements, etc.). In the control literature, supervisory control is typically used to control the behavior of DES which restricts possible behaviors but it does not force a specific action. Although, similar results may be shown for automata, below the discussion focuses on supervisory control of DES using Petri Nets.

In the supervisory control of DES described by Petri Nets, the feedback supervisory controller restricts the plant behavior so to satisfy desired constraints. Specifically, appropriate place invariants are generated and a set of initial conditions are specified that restrict the dynamic behavior of the plant in such a way so to satisfy the control specifications which are given here in terms of inequalities on the markings.

Similar ideas apply when feedback is used and the process dynamics are described by automata or Petri nets. In the Petri net case, specifications described by linear inequalities are imposed by introducing new place invariants (new behaviors) and restricting the starting points (suppressing existing behaviors)—place invariants restrict system behaviors, for example if the system starts with its state (markings) in an invariant set it stays there. In standard Supervisory Control feedback information is used to restrict the behavior and not to completely alter it. New dynamics are added (in the PN case via additional control places) and the class in initial conditions is restricted. The mechanism here corresponds to changing some of the modes of the LTI system but not all.

J Sensitivity Considerations

Reduction of uncertainties is the primary reason for using feedback. This is expressed very conveniently using the sensitivity function $S(s) = (1 + G_cG(s))^{-1}$. The percentage change in the total transfer function gain $T(s)$ is determined by multiplying the percentage change of the loop gain $G_cG(s)$ by $S(s)$ all evaluated at the frequencies of interest $s = j\omega$. Recall that

$$\frac{\Delta Y}{Y} \approx S(\frac{\Delta G}{G}).$$

So when the sensitivity of the loop is low, large variations in the plant gain, due to plant parameter variations, translate into small variations in the overall gain of the compensated system.

Sensitivity reduction (and stability) are not automatic when closing the loop but they depend on the choice of the controller. If the controller is not selected appropriately, undesireable effects such as an increase in sensitivity or even destabilization may occur.

It is well known that that the main reason for using closed loop feedback control is the uncertainties in the plant and environment. Open loop control cannot compensate for plant parameter variations or disturbances (unless they can be measured directly).

Closing the loop, however, does not guarantee automatic benefits. In fact, feedback can also be harmful as the following simple examples remind us. Consider

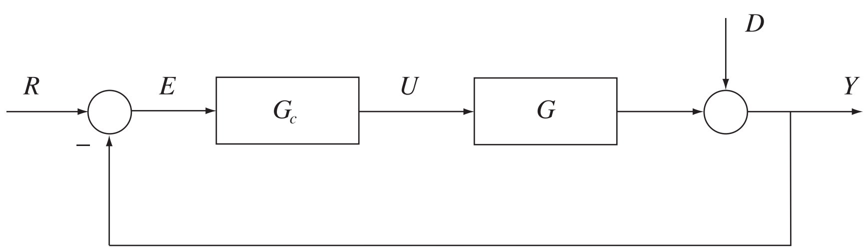

Figure 13:

Note that

$$E = \frac{1}{1+GG_c}R - \frac{1}{1+GG_c}D$$

$$U = \frac{G_c}{1+GG_c}R - \frac{G_c}{1+GG_c}D$$

$$Y = \frac{GG_c}{1+GG_c} R + \frac{1}{1+GG_c} D$$

The sensitivity function (to parameter variations) is

$$S = \frac{1}{1+GG_c}.$$

Example 1. Let $G_c = k$, $G = A$, both static gains, with k to be chosen. The sensitivity to parameter variations should be less than 1, so to derive benefits when closing the loop. That is, $-1 < S < 1$ or $-1 < \frac{1}{1+kA} < 1$. From which kA should satisfy $kA > 0$ or $kA < -2$. For $-2 < kA < 0$ the sensitivity is greater than 1 ($|S| > 1$) and so the closed loop performs worse than the open loop.

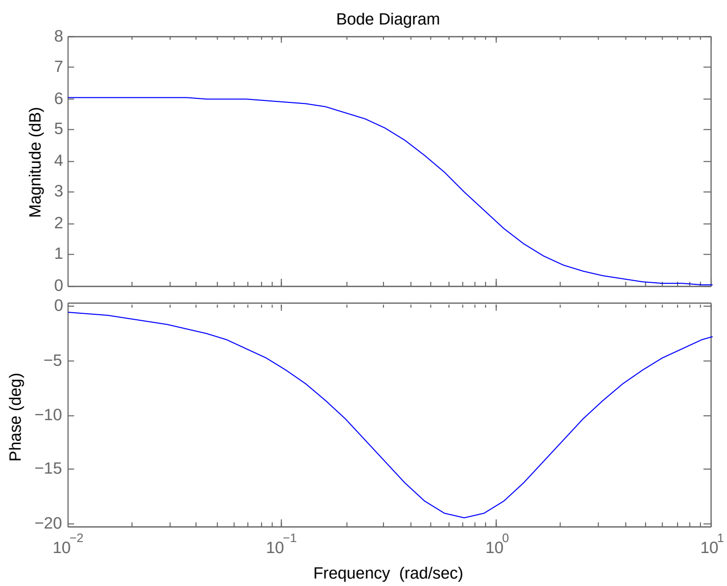

Figure 14:

So sensitivity reduction is not automatic when closing the loop but it depends on appropriately selecting the controller.

Example 2. Let $G_c = k$, $G = \frac{1}{s+1}$. The closed loop characteristic equation is $s + 1 + k = 0$ and so when $1 + k > 0$ or $k > -1$ the system remains stable. When $k < -1$, the system becomes unstable. That is the feedback controller here destabilizes the stable plant.

The sensitivity function is

$$S = \frac{s+1}{s+1+k}.$$

It can be seen that when k satisfies $-1 < k < 0$, the sensitivity magnitude is greater than 1 for all frequencies. For example, for $k = 0.5$ the bode plot is given in Fig. A.6 where it can be seen that S is always greater than 1.

Similarly, if $G_c = k$ and $G = \frac{1}{s-1}$, the closed loop characteristic equation is $s - 1 + k = 0$ and so, when $-1 + k > 0$ or $k > 1$ the system is stabilized. However, if $k < 1$ the system remains unstable. So stability is not automatically guaranteed when closing the loop, but it depends on appropriately selecting the controller.

K Root Locus

Consider $G(s) = \frac{1}{s(s+1)(s+2)}$, $G_c = k$, in the unity feedback configuration. The root locus for $k > 0$ is given below. It can be seen that even for small values of k, the closed loop poles are different from the open loop ones, that is any nonzero feedback gain ompletely changes the open loop poles as expected.

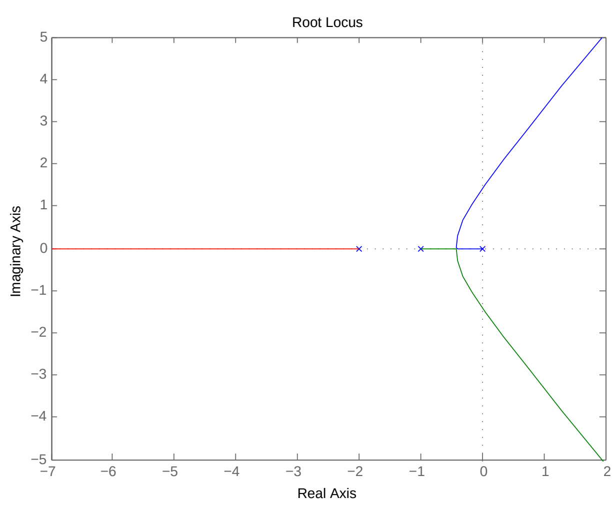

Figure 15:

L On Symmetries

It appears that symmetries are intimately related to the fundamental properties of feedback. In particular, it was seen that feedback automatically generates an inverse (D) of the plant dynamics (D^{-1}) so to cancel them out. This can be seen as creating automatically symmetric mirror images of the undesirable dynamics, implying that feedback is based on creating symmetries. Symmetries are exact when feedback is used since exact copies of the inverse dynamics are generated. Furthermore, note that to reject undesirable dynamics introduced to the loop by external disturbances, copies of the undesirable dynamics should be inserted in appropriate maps (internal models). This, again, is done by creating symmetric mirror images of the undesirable dynamics that need to be cancelled. When the loop gain is high, approximate symmetric mirror images of the zeros are also created, in addition to the symmetric mirror images of the poles, so the whole plant cancels out and the trnsfer function between R and Y is approximately one. In optimal control (LQR) the choice of optimal gains leads to symmetric mirror images as well. Further details and additional connections between feedback and symmetries will be discussed in future versions of this paper.

M Feedbacck Configuration

In the figure below, a standard feedback plant/controller configuration is given for completeness. The cancellation of G_1 and G_2 poles may be easily seen from the expressions describing the relations among the various signals when using polynomial matrix fractional descriptions (see [2]–Chapter 7 pp. 584-586).

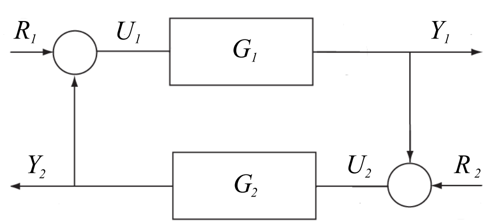

Figure 16: