NATO-ASI on Methods of Model Based Process Control, Antalya, Turkey, Aug. 7-17, 1994.

Frequency Domain Methods for Analysis and Design
II. CONTROLLABILITY ANALYSIS OF SISO SYSTEMS

Sigurd Skogestad *
Chemical Engineering
University of Trondheim, NTH
N-7034 Trondheim, Norway

These notes consists of two parts. Part I gives an overview of modern frequency-domain methods, including H-infinity methods and robust control using the structured singular velue, mu. Part II gives a tutorial introduction to controllability analysis for scalar systems using the frequency domain. Some readers may prefer to read Part II first. This part II is an extended version of some lecture notes used both for a graduate and undergraduate control course. Parts of the notes have previously been presented at the IFAC conferences ADCHEM'94 in Kyoto (mainly Section 3 on the background for the controllability analysis) and IPDC'94 in Baltimore (mainly Section 4 on the pH-application).

Abstract

The objective of this paper is to derive some fundamental results for controllability analysis of single-input single-output (SISO) systems. The effects of disturbances, delays, constraints and RHP-zeros are quantified. These results are applied to a neutralization process where it is shown that the process must be modified to get acceptable controllability.

1 INTRODUCTION

In process control courses the issues of controller design and stability analysis are often emphasized. However, in practice the following three issues are usually more important.

I. How well can the plant be controlled? Before attempting to start any controller design one should have some idea of how easy the plant actually is to control. Is it a difficult control problem? Indeed, does there even exist a controller which meets the required performance objectives?

II. What control strategy should be used? What to measure, what to manipulate, how to pair? In textbooks one finds qualitative rules to address this issue. For example in Seborg et al. (1989) one finds in a chapter called "The art of process control" the rules:

1. Control outputs that are not self-regulating
2. Control outputs that have favorable dynamic and static characteristics, i.e., there should exist an input with a significant, direct and rapid effect.
3. Select inputs that have large effects on the outputs.
4. Select inputs that rapidly effect the controlled variables

These rules are reasonable, but what is "self-regulating", "large", "rapid" and "direct". One objective of this paper is to quantify this.

III. How should the process be changed to improve control ? For example, one may want to design a buffer tank for damping a disturbance, or one may want to know how fast a measurement should be to get acceptable control.

Controllability analysis. All the above three questions are related to the inherent control characteristics of the process itself, that is, to what is denoted the *controllability* of the process. We shall use the following definition:

(Input-output) controllability is the ability to achieve acceptable control performance, that is, to keep the outputs (y) within specified bounds or displacements from their setpoints (r), in spite of unknown variations such as disturbances (d) and plant changes, using available inputs (u) and available measurements (e.g.,y_m or d_m).

In summary, a plant is controllable if there *exists* a controller (connecting measurements and inputs) that yields acceptable performance for all expected plant variations. Thus, controllability is independent of the controller, and is solely a property of the plant (process) only. It can only be affected by changing the plant itself, that is, by *design modifications.* Surprisingly, in spite of the fact that mathematical methods are used extensively for control system design, the methods available when it comes to controllability analysis are usually qualitative. In most cases the "simulation approach" is used. However, this requires a specific controller design and specific values of

*E-mail: skoge@kjemi.unit.no; phone: +47-73-594154; fax: +47-73-594080

disturbances and setpoint changes. In the end one never really knows if the assessment is a fundamental property of the plant or if it depends on the specific choices made.

The objective of this paper is to present quantitative controllability measures which can replace this *ad hoc* procedure. The paper deals with scalar (SISO) systems, but all the tools presented may be generalized to multivariable (MIMO) systems. Disturbances are considered in detail, but model uncertainty, which also necessitates the use of feedback control, is not included in this paper. Linear control theory is used, and most of the tools make use of the frequency response. One reason for this is the very useful idea of "bandwidth" which is a purely frequency-domain concept.

One shortcoming with the controllability analysis presented in this paper is that all the measures are linear. This may seem to be very restrictive, but in most cases it is not. In fact, one of the most important nonlinearities, namely input constraints, can be handled with the linear approach. To deal with slowly varying changes one may perform a controllability analysis at several selected operating points. As a last step of the controllability analysis one should perform some nonlinear simulations to confirm the results of the linear controllability analysis. The experience from a large number of case studies has been that the agreement is generally very good.

Remarks on the definition of controllability. The above definition is in agreement with one's intuitive feeling about the term, and is also how the term was used originally in the control literature. For example, Ziegler and Nichols (1943) define controllability as *"the ability of the process to achieve and maintain the desired equilibrium value"*. Rosenbrock (1970, p. 161) notes that *"in engineering practice, a system is called controllable if it possible to achieve the specified aims of control, whatever these may be"*. Unfortunately, in the 60's the term "controllability" became synonymous with the rather narrow concept of "state controllability" introduced by Kalman, and the term is still used in this restrictive manner by the system theory community. "State controllability" is the ability to bring a system from a given initial state to any final state (but with no regard to the dynamic repsonse between and after these two states). This concept is of interest for realizations and numerical calculations, but as long as we know that all the unstable modes are both controllable and observable, it has little practical significance. For example, Rosenbrock (1970, p. 177) notes that "most industrial plants are controlled quite satisfactorily though they are not [state] controllable". He also remarks that "the chief point to be stressed is that controllability is an engineering term with a wide connotation. To restict its meaning to one particular type of controllability seems wrong, and leads to confusion."To avoid confusion with Kalman's state controllability, Morari (1983) introduced the term "dynamic resilience". However, this term does not capture the fact that "controllability" is related to control, and so instead we propose to use the term "input-output controllability" to make the distinction with "state controllability".

Finally, one should note that the term "controllable" does not quite mean the same as "easy to control". The latter usually means that one can "easily design a simple controller" and get acceptable performance. On the other hand, "controllable" means that there *exists* a controller which yields acceptable performance, although this controller may be very complex and require a detailed model of the plant. It is possible to restict the definition of controllability to make it closer to the term "easy to control". For example, one may require the controller to be linear (as is done throughout this paper), or to be decentralized, or to be of a certain order or form (e.g., PID controller).

One may also consider controllability using feedback control, which is the main topic in this paper, although we do also have some discussion on the use of feedforward control.

The link between process design and control. The terms controllability provides the link between process design and control. This is explained very nicely by Ziegler and Nichols (1943):

"The finest controller made, when applied to a miserably designed process, may not deliver the desired performance. True, on badly designed processes, advanced controllers are able to eke out better results than olde models, but on these processes, there is a definite end point which can be approached by instrumentation and it falls short of perfection. The chronology in process design is evidently wrong. Nowadays an engineer first designs his equipment so that it will be capable of performing its intended function at the normal throughput rate plus a safety factor. The control engineer or instrumentman is then told to put on a controller capable of maintaining the static equilibrium for which the apparatus was designed. .... When the plant is started, however, it may be belatedly discovered that, in spite of the correct equipment design for steady-state condition and the correct instrument selection, control results are not within the desired tolerance. A long expensive process of "cut and try" is then begun in order to make the equipment work. Both engineers realize that some factor in equipment design was neglected but generally can neither identify the missing ingredient nor correct it in future design.

The missing characteristic can be called "controllability", the ability of the process to achieve and maintain the desired equilibrium value. Design for steady-state conditions is not enough if exact maintenance of variables is necessary. "

Ziegler and Nichols then point out that although "a great many factors affecting controllability have been identified" the problem is complex, and "as it now stands the plant designer is almost justified in disregarding the entire matter. .... Sooner or later, however, these factors affecting process controllability will have to be smoked out and reduced to definite "good-practice" rules which will be as much a part of equipement design as safety factors".

It it probably fair to say that progress has been slow, and now, more than 50 years later, such good-practice rules are still not in common use. It is hoped, however, that this tutorial paper will contribute to the "smoking-out" process.

Design modifications. As pointed out above, controllability can only be affected by design modifications. These may include:

1. Change the apparatus itself (type, size, etc.)
2. Relocate sensor and actuators
3. Add new equipment to dampen disturbances, for example, buffer tanks.
4. Add extra sensors for measurement (cascade control)
5. Add extra actuators (parallell control)
6. Change the control objectives
7. Change the control structure of the lower levels

In most cases controllability is improved by bringing the actuator and measurement device closer together in order to improve the speed of response, for example, by reducing the process delay. This applies to the first items above, which usually are quite problem specific and are not treated in this paper.

It it arguable whether or not the last two items are design modifications, but at least they address issues which come before the actual controller design. The last issue is important because control systems are usually designed in a hierarchical manner, and the lower-level loops are assumed closed when designing the control system at a given level. Thus, a change in the lower-level control structure may drastically change the achievable control performance of the levels above, and therefore may be viewed as a design modification as seen from the level above.

Previous work on controllability analysis. The topic has been addressed in many application papers, but mostly on an *ad hoc* basis since the theoretical basis for a controllability analysis has been relatively poor (one reason for this is probably the unfortunate use of the term in the meaning of state controllability, which led to the belief that there was nothing more to).

Except for the initial work by Ziegler and Nichols (1943), there does not seem to have been much progress on input-output controllability analysis until Rosenbrock (1966, 1970) presented a thorough discussion on the various definitions of state controllability and observability, and introduced similar concepts in terms of the *outputs*. This led to the introduction of the term "functional controllability" (which for scalar systems is equivalent to requiring that the transfer function $g(s)$ is not identically equal to zero) and to the important notion of right half plane (RHP) zeros (which for scalar systems is directly related to inverse responses). Another important step towards a quantitative analysis was made by Morari (1983) who made use of the notion of "perfect control" in an attempt to quantify the term controllability. Balchen and Mumme (1988, pp. 16-21, pp.47-48) present some nice controllability guidelines which are more specific than the rules from Seborg er al. (1989) given above, but most of them lack a theoretical justification.

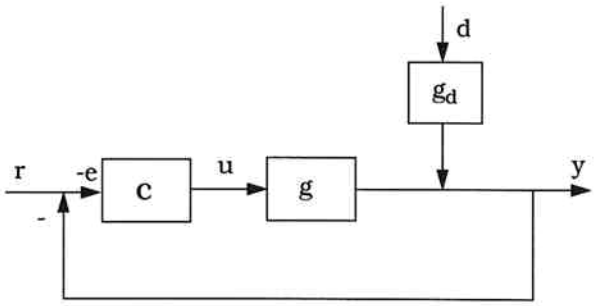

Figure 1: Block diagram of feedback control system.

One important issue which was missing from most of Morari's and Rosenbrock's analyses was an explicit consideration of disturbances. Disturbances have of course been discussed in many application papers, but only recently have their relationship to controllability been treated in a systematic manner (e.g., Skogestad and Wolff, 1992).

The tools for controllability analysis are now reaching a more mature state, but still the fundamental ideas are not well known. The objective of this paper is to present the ideas for scalar systems in a tutorial manner. For decentralized control of multivariable processes the results may be generalized directly by introducing the Closed Loop Disturbance Gain (CLDG) and the Performance Relative Gain Array (PRGA) (Hovd and Skogestad, 1992).

2 LINEAR CONTROL THEORY

Notation. Consider a linear process model in terms of deviation variables

$$y = gu + g_d d \tag{1}$$

Here y denotes the output, u the manipulated input and d the disturbance (including what is of-

ten referred to as "load changes"). $g(s)$ and $g_d(s)$ are transfer function models for the effect on the output of the input and disturbance, and all controllability results in this paper are based on this information. The Laplace variable s is often omitted to simplify notation. The control error e is defined as

$$e = y - r \tag{2}$$

where r denotes the reference value (setpoint) for the output.

Feedback control. Consider a simple feedback scheme

$$u = c(s)(r - y) \tag{3}$$

where $c(s)$ is the controller. Eliminating u from equations (1) and (3) yields the closed-loop response

$$y = Tr + Sg_d d \tag{4}$$

Here the sensitivity is $S = (I + gc)^{-1}$ and the complementary sensitivity is $T = gc(I + gc)^{-1} = 1 - S$. The transfer function around the feedback loop is denoted L. In this case $L = gc$. The corresponding input signal is

$$u = -ce = cSr - cSg_d d \tag{5}$$

The frequency domain. Most of the results in this paper are based on the frequency domain. Unfortunately, few process engineers feel comfortable with this domain, so a simple introduction is given first. Consider the effect of a small change in the input (input signal) u on the output (output signal) y. In the Laplace domain this may be represented as

$$\Delta y(s) = g(s)\Delta u(s)$$

where Δu represents a small change in the input (independent variable), and $\Delta y(s)$ is the resulting change in the output. $g(s)$ is the transfer function of the system. The Δ is included to show explicitly that we are dealing with deviation variables, but since we will only deal with deviation variables in this paper the Δ will be omitted to simplify notation.

Let us now consider the time domain where most engineer feel more comfortable. The problem with the time domain is that we have to consider specific input signals $u(t)$ and have to recompute $y(t)$ for each signal. The favorite input test signal for engineers is a step. However, in general a step response does not provide sufficient information for a controllability analysis. Therefore the frequency domain shoul be used.

The physical interpretation of the frequency domain for a system $y = g(s)u$ is a follows: A persistent sinusoidal input with frequency ω, $u(t) = u_0 sin(\omega t)$, yields a persistent sinusoidal output with the same frequency, $y(t) = y_0 sin(\omega t + \phi)$, but shifted in phase by ϕ. This is shown graphically in Figure 2 for a first-order system with time delay,

$$g(s) = \frac{ke^{-\theta s}}{1 + \tau s}; \quad k = 5, \theta = 2, \tau = 10 \tag{6}$$

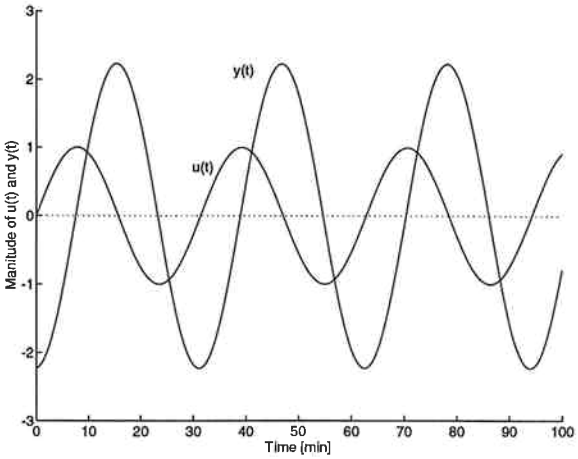

Figure 2: Sinusoidal response for system $g(s) = 5e^{-2s}/(1 + 10s)$ at frequency $\omega = 0.2$ [rad/min]. Period $P = 2\pi/\omega = 31.4$ min. Gain $|g(j\omega)| = 5/\sqrt{1 + (10\omega)^2} = 2.24$. Phase shift $\phi = -\arctan(10\omega) - 2\omega = -1.51$ rad $= -86.3^o$ corresponding to time shift $\Delta t = -\phi/\omega = 7.6$ min.

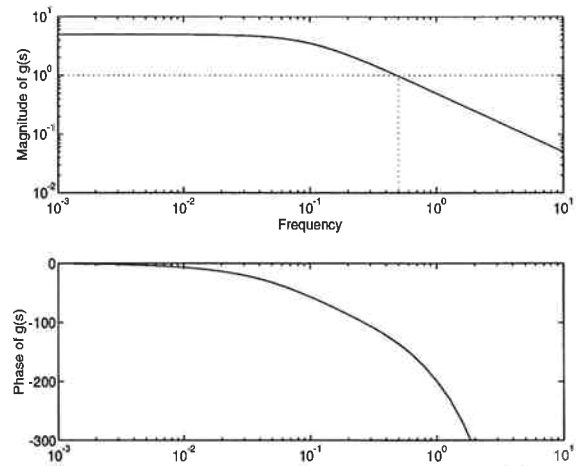

Figure 3: Frequency response of $g(s) = 5e^{-2s}/(1 + 10s)$.

It is useful to have in mind this physical picture of the frequency domain when one interprets the controllability results presented later. The magnitude y_0 and phase shift ϕ is easily computed from the Laplace transform $g(s)$ by inserting the imaginary number $s = j\omega$ and evaluating the magnitude and phase of the resulting complex number:

$$y_0/u_0 = |g(j\omega)|; \quad \phi = \angle g(j\omega) \text{ [rad]} \tag{7}$$

In this paper we use a "frequency-by-frequency" approach and at each frequency consider the response to a sinusoidal input of unit magnitude ($u_0(\omega) = 1$) as illustrated in Fig 2. This results in the "frequency response" of the system where we consider the system gain $y_o(\omega)/u_o(\omega) = |g(j\omega)|$ (and possibly the phase shift $\angle g(j\omega)$) as a function of ω. Graphically, this frequency response is usually represented in a Bode plot with a log-scale for frequency and gain.

In Fig. 3 the frequency response (Bode-plot) is

shown for the example in (6). We note that in this case both the gain (amplitude) and phase falls monotonically with frequency. This is quite common for open-loop (no feedback control) chemical engineering systems. The delay θ will only shift the sinusoid in time, and thus affects the phase but not the gain. The system gain $|g(j\omega)|$ is equal to k at low frequencies (this is the steady-state gain and is obtained by setting $s = 0$). The gain remains relatively constant up to a frequency of about $1/\tau$ where it start falling sharply. Physically, the system responds too slowly to let high-frequency ("fast") inputs have much effect on the outputs, that is, high-frequency sinusoidal inputs are smoothened ("dampened", "attenuated") by the system dynamics.

Assume that $k > 1$ and note for later reference the frequency ω_d where the gain is 1, that is, $|g(j\omega_d)| = 1$ (this frequency is of particular interest when $g(s)$ is the disturbance model, and this is the reason for the subscript d). The exact value is given by $k/\sqrt{1+(\omega_d\tau)^2} = 1$, but we often use the asymptotic approximation $k/(\omega_d\tau) \approx 1$, and obtain

$$\omega_d \approx k/\tau \tag{8}$$

Thus, we see that ω_d is large if the steady-state gain k is large (the input has a large effect on the output) or if the time constant τ is small (the input has a fast effect on the output).

Frequency responses may be obtained for any transfer function. In this paper we consider frequency responses of three transfer functions: $g(j\omega)$ (effect of manipulated inputs u on outputs y), $g_d(j\omega)$ (effect of disturbances d on outputs y), and $L = gc(j\omega)$ (frequency response of loop transfer function). The frequency responses of g and g_d are often similar to the response shown in Fig. 3, whereas the magnitude of $L = gc$ is often infinite at low frequency because the controller $c(s)$ usually contains an integrator.

Bandwidth. Here bandwidth is defined as *the frequency ω_B where the loop gain is one in magnitude*, i.e. $|L(j\omega_B)| = 1$ (or more precicely where the low-frequency asymptote of $|L|$ first crosses 1 from above). This frequency is often called the "crossover frequency".

At frequencies lower than the bandwidth ($\omega < \omega_B$) feedback is effective and will affect the frequency response. However, for sinusoidal signals (for example, a disturbance) with frequencies higher than ω_B the response will not be much affected by the feedback.

Other definitions of bandwidth are also in use, for example, as the frequency where $|S(j\omega)| = 0.7$ or the frequency where $|T(j\omega)| = 0.7$. The above definition in terms of the loop transfer function is preferred because it is simple. It usually yields a value between the two alternative definitions in terms of $|S|$ and $|T|$.

A frequency domain analysis, in particular in the frequency-region corresponding to the bandwidth, is very useful for systems under feedback control. This is the case even when the disturbances and setpoints entering the system are *not* sinusoids. One reason for this is that the feedback control system will usually amplify frequencies corresponding to the closed-loop bandwidth, ω_B. [1]. For example, the effect of disturbances is usually largest around the bandwidth frequency; slower disturbances are attenuated by the feedback control, and faster disturbances are usually attenuated by the process itself. Thus, the magnitude of g_d at the bandwidth frequency, $|g_d(j\omega_B)|$, is usually a very good approximation of the worst-case amplification of a disturbance when using feedback control. This means that if we can somehow estimate the best achievable ω_B, we can say a lot about how sensitive the system is to disturbances under feedback control. The implication for design is to look for plant modifications which makes the plant more "self-regulating" in terms of reducing the magnitude of $|g_d(j\omega_B)|$.

For pure feedforward control the frequency domain may not be quite as relevant. For example, if the disturbances are always steps then a step response analysis may be more relevant. However, in many cases the disturbances are sinusoidal since they are generated from feedback loops in other parts of the system.

3 CONTROLLABILITY ANALYSIS

Scaling. The interpretation of most measures presented in this paper assumes that the transfer functions g and g_d are in terms of scaled variables. The first step in a controllability analysis is therefore to scale (normalize) all variables (input, disturbance, output) to be less than 1 in magnitude (i.e., within the interval -1 to 1).

Thus, in the following *we assume that the signals are persistent sinusoids, and that g and g_d have been scaled, such that at each frequency the allowed input $|u(j\omega)| < 1$, the expected disturbance $|d(j\omega)| < 1$, the allowed control error $|e(j\omega)| < 1$, and the expected reference signal $|r(j\omega)| < R_{max}$*. Note that e and r are measured in the same units so R_{max} is the magnitude of the expected setpoint change relative to the allowed control error. The detailed scaling procedure is outlined in the Appendix.

The ideal controller and plant inversion. The objective of the control system is to manipulate u such that the control error e remains small in spite of disturbances and changes in the setpoint. The ideal controller will accomplish this by inverting the process (Morari, 1983) such that

[1] The bandwidth frequency will often show up as oscillations in the time response and we usually have $\omega_B \approx 2\pi/P$ where P is the period of the oscillations.

the manipulated input becomes (set $y = r$ in (1) and solve for u):

$$u = g^{-1}r - g^{-1}g_d d \tag{9}$$

For example, an ideal feedforward controller operates in this manner. Usually, the disturbance is not measured and feedback control is used instead. As may be expected, the input signal generated under feedback is also given by Eq.(9) at frequencies where feedback is effective. To see this, consider Eq. (5) and use the fact $cS = g^{-1}T$ to derive the following expression for the input signal under feedback control

$$u = g^{-1}Tr - g^{-1}Tg_d d \tag{10}$$

At low frequencies, $\omega < \omega_B$, where $|gc(j\omega)| > 1$ and feedback is effective we have $S \approx 0$ and $T \approx 1$, and we rederive (9). Consequently ideal control (inversion) requires *fast* feedback control (high bandwidth).

On the other hand, inherent limitations of the system may prevent fast control. The limitations may include constraints on the allowed input signal u and non-minimum phase elements in $g(s)$ such as time delay and right half plane zeros. *If these requirements for high and low bandwidth are in conflict then controllability is poor.* The objective of the remaining part of this section is to quantify these statements. The results are derived for feedback control, although some of them also apply to feedforward control.

3.1 Disturbances and bandwidth

The effect of a disturbance on the output at a frequency ω in the absence of control is

$$y(j\omega) = g_d(j\omega)d(j\omega) \tag{11}$$

(we are here assuming that $r = 0$ such that the control error $e = y$). The worst-case disturbance at this frequency has magnitude 1, i.e., $|d(j\omega)| = 1$. Furthermore, at each frequency the output should be less than 1 in magnitude, i.e., we need control if $|y(j\omega)| > 1$. *Consequently, at frequencies where $|g_d(j\omega)| > 1$ we need control (feedforward or feedback) in order to prevent the output exceeding its allowed bound.* Typically, $|g_d(j\omega)|$ is larger than 1 at low frequencies and drops to zero at high frequencies. *In this case the frequency, ω_d, where $|g_d(j\omega_d)| = 1$ is a useful controllability measure:* At frequencies lower than ω_d we need control to reject the disturbance, and thus ω_d provides a minimum bandwidth requirement for control, and we have the approximate requirement

$$\omega_B > \omega_d \tag{12}$$

Example. Consider the disturbance model (recall Fig.3)

$$g_d(s) = k_d e^{-\theta_d s}/(1 + \tau_d s) \tag{13}$$

where $k_d = 5$ and $\tau_d = 10$ [min]. Scaling has been applied to g_d, so this means that with no control, the effect of disturbances on the outputs at low frequencies is $k_d = 5$ times larger than what we allow. Thus control is required, and since g_d crosses 1 at a frequency $\omega_d \approx k_d/\tau_d = 0.5$ rad/min, the minimum bandwidth requirement for disturbance rejection using feedback control is $\omega_B > 0.5$ rad/min.

Remarks.

1. Scaling is critical for any controllability measure involving disturbance rejection.
2. Recall the following rule from the introduction:
   - Control outputs that are not self-regulating

   This rule can be quantified as follows: Control outputs y for which $|g_d(j\omega)| > 1$ at some frequency.
3. In words we have proved that "large disturbances with a fast effect" require fast control. Specifically, if the disturbance is increased, then to get acceptable performance the bandwidth (speed of response) of the control system has to be increased.
4. To be more specific assume that the disturbance is increased by a factor f, and assume that at frequency ω_d the slope of $|g_d(j\omega)|$ on the log-log Bode-plot is $-\beta$, that is, $g_d \sim 1/s^\beta$ at the frequency ω_d (in the example above $\beta = 1$). Then the bandwidth has to be increased by a factor $f^{1/\beta}$ to counteract the increased disturbance.
5. Note that a delay in the disturbance model has no effect on the required bandwidth.
6. On the other hand, with feedforward control where the disturbance is measured, a delay in the disturbance model makes control easier.

3.2 Input constraints

Consider the response to a "worst-case" sinusoidal disturbance of magnitude 1 ($|d(j\omega)| = 1$) and assume $r = 0$. From Eq.(9) the input magnitude needed for perfect control ($e = 0$) is

$$|u| = |g^{-1}g_d d| = |g_d|/|g| \tag{14}$$

(Strictly speaking, perfect control is not required, and the input needed for "acceptable" control ($|e| < 1$) is $|u| = (|g_d| - 1)/|g|$. The difference is small at frequencies where $|g_d|$ is larger than 1, and the input needed for perfect control will be used in the following[2]).

Consider frequencies $\omega < \omega_d$ where control is needed to reject disturbances. The requirement is that $|u(j\omega)| \leq 1$ at each frequency. To fulfill this one must require

$$|g(j\omega)| > |g_d(j\omega)|, \quad \forall \omega < \omega_d \tag{15}$$

Similarly, to perfectly track a setpoint $r(j\omega) = R_{max}$ at each frequency with $|u| < 1$ one must from Eq.(9) require

[2] For multivariable systems the differences between perfect and acceptable control may be large if the plant is ill-conditioned.

$$|g(j\omega)| > R_{max}, \quad \forall \omega < \omega_r \tag{16}$$

where ω_r is the frequency up to which setpoint tracking is desired.

Remarks.

1. Recall the following rule from the introduction:
   - Select inputs that have large effects on the outputs.

   This rule may be quantified as follows: In terms of scaled variables we should have $|g| > |g_d|$ at frequencies where $|g_d| > 1$, and additionally we should have $|g| > R_{max}$ at frequencies where setpoint tracking is desired.
2. The following remark applies also to the previous subsection on disturbances and bandwidth. If there are several disturbances then they should be analyzed individually to identify the most difficult ones. This could be the starting point for proposing design modifications. (The worst-case combined effect of several disturbances may be obtained by simply adding together their individual effects. For example, let the effect of disturbance d_k on y be g_{dk}. Then to consider the worst-case combination one may simply replace $|g_d|$ by $\sum_k |g_{dk}|$ in the above expressions.)
3. For unstable plants we need a minimum bandwidth p to stabilize the system (see below). In this case we need $|g| > |g_d|$ up to the frequency p. Otherwise, the input will saturate, and the plant can not be stabilized.
4. The bounds (15) and (16) are strictly speaking only *necessary* conditions for controllability. This follows since we have used a frequency-by-frequency analysis and have not considered whether there actually exist a causal controller that can achieve the performance required by perfect control. In other words, we must always satisfy the bounds (15) and (16) (or at least the modified bound for "acceptable" control), but this may not be sufficient to avoid input constraints in the presence of delays or RHP-zeros.
5. Since the input needed for perfect control is independent of the control implementation, the bounds (15) and (16) also apply to feedforward control.

3.3 Time delay and right half plane zeros

It is well-known that time delays and right half plane (RHP) zeros limit the achievable speed of response. We shall here quantify this statement in terms of upper bounds on the allowed bandwidth. The derivation makes use of the complementary sensitivity function T which for a controller without a prefilter on r is the transfer function from setpoint to output, i.e., $y = Tr$.

Consider an "ideal" controller which is integral square error (ISE)-optimal for the case with step changes in the setpoint (this controller is "ideal" in the sense that it may not be realizable in practice because the required inputs may be infinite). That is, the objective is to minimize $\int_o^\infty |e(t)|^2 dt$ for the case where $r(t)$ is a step, and with no penalty on the input u. In this case the corresponding "ideal" complementary sensitivity for a plant with RHP-zeros at z_i and a time delay θ is (see Morari and Zafiriou, 1989, p. 58)

$$T = \prod_i \frac{-s + z_i}{s + \bar{z}_i} e^{-\theta s} \tag{17}$$

where $\bar{z}_i$ is the complex conjugate of z_i. Note that T is "all pass" since $|T(j\omega)| = 1$ at all frequencies. Given T we can compute the loop transfer function $L = T/(1-T)$, and then obtain the bandwidth as the frequency where $|L(j\omega)|$ crosses 1.

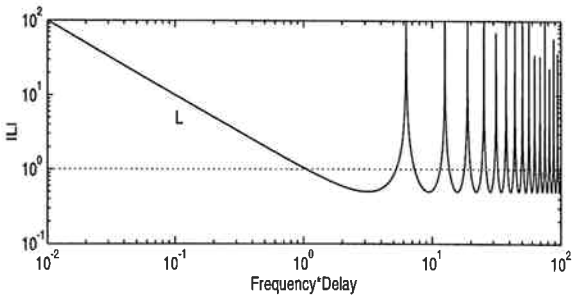

Figure 4: "Ideal" loop transfer function for plant with delay

Time delay. Consider a plant with a time delay, that is, $g(s)$ contains the term $e^{-\theta s}$. The "ideal" controller can "invert away" most of the dynamics in $g(s)$, but it cannot remove the delay. Thus, even the "ideal" complementary sensitivity function will contain the delay,

$$T = e^{-\theta s} \tag{18}$$

The loop transfer function corresponding to this ideal response is $L = T/(1-T) = e^{-\theta s}/(1-e^{-\theta s})$. The magnitude $|L|$ is plotted in Figure 4. At low frequencies, $\omega\theta < 1$, we have $e^{-\theta s} \approx 1 - \theta s$ (by a Taylor series expansion of the exponential) and $L \approx \frac{1}{\theta s}$, and thus the low frequency asymptote of $|L(j\omega)|$ crosses 1 at frequency $1/\theta$ (the exact frequency where $|L(j\omega)|$ crosses 1 in Fig. 4 is $\frac{\pi}{3}\frac{1}{\theta} = 1.05/\theta$). This is the bandwidth frequency. In practice, the "ideal" controller cannot be realized, and so this analysis provides an upper bound on the bandwidth of approximately

$$\omega_B < 1/\theta \tag{19}$$

Real RHP zero. Consider a plant with an inverse response, that is, $g(s)$ contains a term $(-s + z)$ corresponding to a real RHP zero at z. Again, the "ideal" controller cannot remove the effect of this RHP zero. Thus, even the "ideal" complementary sensitivity function will contain the RHP-zero

$$T = \frac{-s + z}{s + z} \tag{20}$$

The loop transfer function corresponding to this ideal response is $L = (-s + z)/2s$. The magnitude $|L|$ is plotted in Figure 5. The low frequency

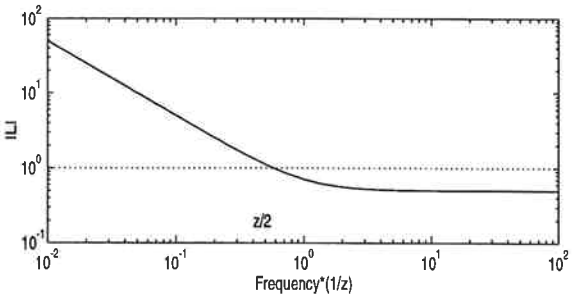

Figure 5: "Ideal" loop transfer function for plant with RHP zero.

asymptote of $|L(j\omega)|$ crosses 1 at frequency $z/2$. In practice, the "ideal" controller cannot be realized, and we obtain an upper bound on the bandwidth of approximately

$$\omega_B < \frac{z}{2} \tag{21}$$

Remarks on bounds (19) and (21).

1. The bounds are independent of scaling.
2. The bounds provide a quantification of the rules
   - Control outputs that have favorable dynamic and static characteristics, i.e., there should exist an input with a significant, direct and rapid effect.
   - Select inputs that rapidly effect the controlled variables
3. To reject a disturbance we obtained the requirement $\omega_B > \omega_d$. Combining this with (19) yields an upper limit on the allowed delay, $\theta < 1/\omega_d$. Similarly, we get $\omega_d < z/2$.
4. It will be possible to have a slightly higher bandwidth than given by these two bounds, but only at the expense of a very oscillatory response (corresponding to a large peak in T and S).
5. The above derivation applies when the delay or RHP zero is in the plant itself (between the input u and the output y). However, with feedback control a delay or RHP zero in the measurement of y yields similar limitations, and the above bounds still apply.
6. The bound (21) for RHP-zeros assumes that we want to use u for "slow control" of y for frequencies lower than $z/2$. However, if this is not the case, then one may instead use u for fast (transient) control of y for frequencies higher than z (with the sign of the controller gain reversed compared to the "normal" case[3]). This is further discussed below. This assumes that we are not concerned with the long-term behavior of the output[4], or that we have a "parallell" control system where another input may be used for long-term control of the output.
7. Zeros in the left half plane, corresponding to "overshoots" in the time response, do not present a *fundamental* limitation on control, but *in practice* a LHP-zero located close to the origin may cause problems. First, one may encounter problems with input constraints at low frequency (because the steady-state gain is often low). Second, a simple controller can probably not be used. Specifically, a simple PID controller contains no poles that can be used to counteract the effect of a LHP zero.
8. Similar restrictions to those given by the bounds above also apply to feedforward control. This follows since the ideal T in (17) corresponds to the input u which minimizes the ISE of the output irrespective of the control implementation.

[3]To see that the controller gain must be reversed one may consider the formulas in Morari and Zafiriou (1989, p. 63) where we see that the sign of $\tilde{q}$ and thus of the feedback controller c is zero if the desired response time τ is such that $\tau = 1/z$.

[4]In process control we are usually concerned with the long-time behavior and often require perfect control at steady-state, but there are cases where the control objective is to reject transient disturbances and the steady-state does not matter. One example is the use of a buffer tank to eliminate high-frequency flowrate disturbances.

Further remarks on the limitation of RHP-zeros and the use of positive feedback.

In remark 6 it was claimed that one may essentially choose whether a RHP-zero should pose control limitations at low or high frequencies. This may need some further discussion, as it was certainly not clear to me when I first looked at it.

Let us start with a simple time domain interpretation. A RHP-zero corresponds to an inverse response, that is, to a gain reversal. Usually, one wants good steady-state control and applies negative feedback. In this case one cannot get good transient response as one has to wait until the effect of the apllied input goes in the right direction (after the gain reversal). On the other hand, if one wants good transient response then one can react immediately, but since the gain is in the opposite direction one must positive feedback, and due to the gain reversal one gets poor steady-state control.

Of course, one can use a controller which itself has a RHP-zero and thus a gain reversal, but as shown below this does not really help. Another, even more tempting approach is to use an unstable controller with a pole at the location of the RHP-zero of the plant. However, it is well known that this does not work as it results in an internally unstable system where something eventually will blow up.

Let us now consider an example in more detail. The problem is to design a feedback controller for the plant

$$g(s) = \frac{-s+z}{s+z} \tag{22}$$

which has a a RHP-zero at z. We shall first consider negative feedback, then positive feedback, and finally the combination of the two.

I. Negative feedback. Let us first consider the conventional case where we want good steady-state control and where the bandwidth is approximately limited to $\omega_B < z/2$. The "ideal" controller in terms of minimizing the integral square error to step setpoints has loop transfer function $L = (-s+z)/2s$ corresponding to a PI-controller

$$c(s) = K\frac{s+z}{s} \tag{23}$$

with gain $K = 0.5$. With the controller (23) the sensitivity function is

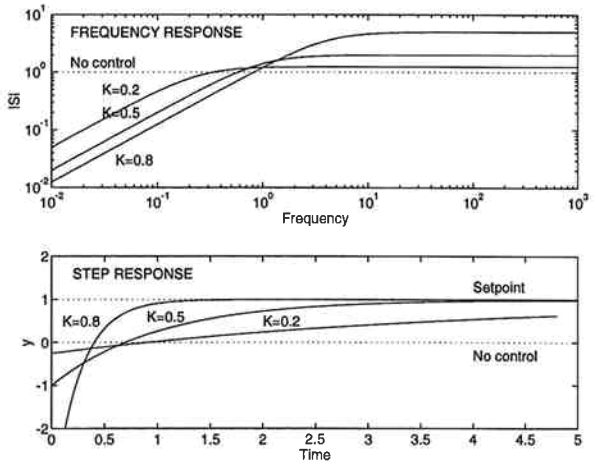

Figure 6: Plant with RHP-zero at $z = 1$ using negative feedback. $c(s) = K\frac{s+1}{s}$

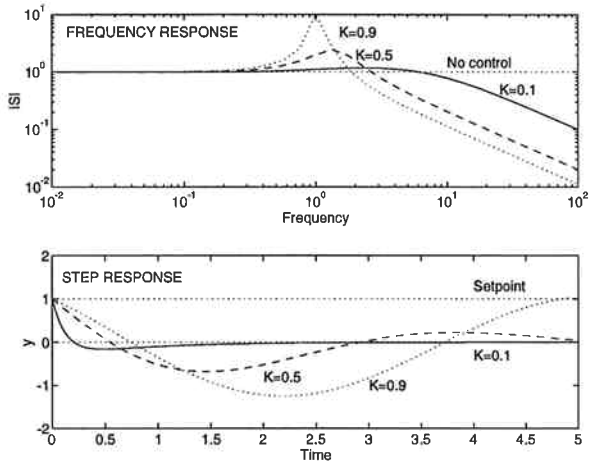

Figure 7: Plant with RHP-zero at $z = 1$ using positive feedback, $c(s) = -Ks$.

$$S = \frac{1}{1+gc} = \frac{s}{(1-K)s + Kz} \tag{24}$$

and we see that the system is stable for $0 < K < 1$. For the numerical calculations let $z = 1$. In Fig. 6 are shown the frequency response of the sensitivity function S and the response to a step setpoint change for various values of K. We note that with higher gains the controller is able to reduce the sensitivity at low frequencies, but only at the expense of a higher peak at some frequency above $z = 1$. Similarly, we see from the time response that an increased gain yields faster settling towards the steady-state, but a poorer initial response (in all cases the output goes in the wrong direction initially). The value $K = 0.5$ is seen to yield a reasonable trade-off between the two situations.

Positive feedback. If we do not care about the steady-state behavior then we may "reverse" the gain of the controller and instead achieve good control at frequencies higher than z. To this effect consider postive feedback using a derivative controller

$$c(s) = -\frac{K}{z}s \tag{25}$$

This yields $S = \frac{s+z}{(K/z)s^2+(1-K)s+z}$. The system is stable for $0 < K < 1$. In Fig. 7 the frequency and step responses for various values of K are shown. We note

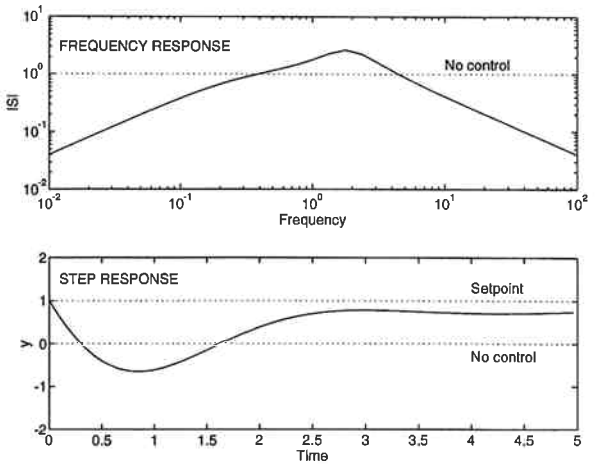

Figure 8: Plant with RHP-zero at $z = 1$ using combined negative and positive feedback, $c(s) = K(-s + \frac{s+1}{s}), K = 0.25$

that with higher gains the controller is able to reduce the sensitivity at high frequencies, but only at the expense of a peak in $|S(j\omega)|$ at intermediate frequencies around $z = 1$. Similarly, we see from the time response that an increased gain yields somewhat better setpoint tracking initially (in all cases the output "jumps" directly up to the desired value of $y = 1$ at $t = 0$), but at the expense of much larger oscillations (in all cases there is no tracking at steady-state). The value $K = 0.5$ yields a reasonable trade-off between the two situations.

III. Combined negative and positive feedback. One may finally consider combining the two control actions at low and high frequency, for example, with the controller

$$c(s) = K(-\frac{s}{z} + \frac{s+z}{s}) = \frac{K}{z}\frac{(-s+1.62z)(s+0.62z)}{s} \tag{26}$$

It is interesting to note that the controller contains a RHP-zero which gives the desired gain reversal. To have stability we must require $0 < K < 0.5$. In Fig. 8 the frequency and step responses for $K = 0.25$ are shown. We are able to reduce the sensitivity below one at all frequencies except around the frequency z corresponding to the RHP-zero. Similarly, we note from the step response that the error is quite large around time $t = 1/z = 1$.

In summary, this combined approach with both negative and positive feedback does not yield much (if any) improvement compared to negative feedback alone. In particular, the settling towards the steady-state is poor. In a practical situation one must, in order to improve the controllability, add another manipulated input to the process to take care of the control at either high or low frequencies. This is commonly done in cases when there is an input with a fast (direct) effect, which has no steady-state effect (i.e., a zero at the origin), and thus can only be used for transient control (Balchen and Mumme, 1988, p.47). In this case a second input must be used for the steady state control.

3.4 Instability

Consider an unstable plant. That is, $g(s)$ contains a term $1/(s-p)$ corresponding to a RHP pole at p. Pure feedforward control can not be used, since even with a feedforward controller with a RHP-zero at p which exactly cancels the RHP-pole, we will have instability because of disturbances entering between the controller and the plant. Thus, the main "limitation" caused by the instability is that *feedback control* is required for stabilization.

To quantify this consider a plant $g(s) = 1/(s-p)$ which is stabilized by a proportional controller, $c(s) = K_c$. The closed-loop pole is at $s = p - K_c$ so we need $K_c > p$ to stabilize the system. Furthermore, for $K_c > p$ the asymptote of the loop transfer function $|L|$ crosses 1 at frequency K_c. Combining these two pieces of information we conclude that the approximate minimum bandwidth needed for stabilization is

$$\omega_B > p \tag{27}$$

Remarks.

1. In words we have found that there is a minimum bandwidth p needed to stabilize the system ("we must respond quicker than the time constant of the instability").
2. For a plant with a time delay we obtained the requirement $\omega_B < 1/\theta$. Combining this with (27) yields the requirement $p < 1/\theta$ or equivalently $\theta < 1/p$.
3. Similarly, for a plant with a RHP-zero we must require $p < z/2$.
4. In theory, any linear rational plant (without time delay) can be stabilized, provided the controller is allowed to be unstable and be nonminimum phase (contain RHP-zeros). Thus, even a plant with a RHP-pole p located to the left of a RHP-zero z (i.e. $p > z$) can in theory be stabilized. This seems to be inconsistent with the above results. It is not, since these results required performance and not just stability. Thus, the requirement $p < z/2$ is indeed needed for obtaining acceptable control performance (at least at low frequencies). Furthermore, if the controller is restricted to being stable, then a plant with a single RHP-pole at $s = p$ can be stabilized (it is "strongly stabilizable") if and only if $p < z$ (Youla et al., 1974).

3.5 Phase lag

Consider a minimum-phase process of the form

$$g(s) = \frac{k}{(1+\tau_1 s)(1+\tau_2 s)\cdots} = \frac{k}{\prod_{i=1}^{n}(1+\tau_i s)} \tag{28}$$

where n is two or larger. At high frequencies the gain drops sharply with frequency ($|g(j\omega)| \approx k/\omega^n \prod \tau_i$) and one may therefore, depending on the value of k, encounter problems with input constraints. Otherwise, the presence of high-order lags does not present any *fundamental* problem.

However, *in practice* the large phase lag at high frequencies ($\angle g(j\omega) \to -n \cdot 90^o$) will usually pose a problem independent of the value of k, because we need the phase of $L = gc$ to be less than -180^o at frequencies lower than the bandwidth ω_B to avoid instability (assuming that $g(s)$ is stable). Thus, zeros in the controller (e.g., derivative action) are needed to counteract the negative phase in the plant. Define the frequency w_{g180} as the frequency where the phase lag in the process itself is -180^o. With a simple PID controller where the derivative action is active over one decade the maximum phase lead is 54.9^o. This is also a reasonable value for the phase margin, and we therefore conclude that with a simple PID controller we must require approximately

$$\text{Practical bound}: \quad \omega_B < w_{g180} \tag{29}$$

Balchen and Mumme (1988, p.17) state that a violation of this bound implies that "feedback control alone will not be satisfactory". This is not strictly correct, as the bound does not pose a *fundamental* limitation if a more complex controller is used. However, in most practical cases the bound in (29) applies since one wants to use simple controllers, and also because the plant model is not known sufficiently well to place zeros in the controller to counteract the poles at high frequency.

3.6 Loop gain requirements for feedback control

The control error under feedback control is given by

$$e = -Sr + Sg_d d \tag{30}$$

We have already used this expression to derive the requirement $\omega_B > \omega_d$, but we may also get more detailed information about the required loop gain $L = gc$ at low frequencies. At a given frequency the worst-case disturbance is $d(j\omega) = 1$ and the requirement is that $|e(j\omega)| < 1$ (assuming that g_d has been appropriately scaled). This requirement is satisfied

$$\text{if and only if} \quad |S(j\omega)| < 1/|g_d(j\omega)|; \quad \omega < \omega_d \tag{31}$$

At low frequencies, $\omega \leq \omega_B$, where feedback is effective and we have $|L(j\omega)| > 1$ and $S(j\omega) \approx 1/L(j\omega)$, (31) becomes

$$|y(j\omega)| < 1 \text{ if and only if } |L(j\omega)| > |g_d(j\omega)|; \quad \omega \leq \omega_d \tag{32}$$

Thus, at frequencies where feedback is needed for disturbance rejection ($|g_d| > 1$), we must require the loop transfer function $|L(j\omega)|$ to be larger than the disturbance transfer function, $|g_d(j\omega)|$ (appropriately scaled).

3.7 Summary of controllability results

Let ω_B denote the closed-loop bandwidth of the system. The following rules apply

1. *Speed of response to reject disturbances. Must require* $\omega_B > \omega_d$. Here ω_d is the frequency at which $|g_d(j\omega_d)|$ first crosses 1 from above. Below this frequency the error will be unacceptable ($|e| > 1$) for a disturbance $d = 1$ unless control is used. More specifically, we must for feedback control require at frequencies lower than ω_d, and $|L| = |gc(j\omega)| > |g_d(j\omega)|$ for $\omega < \omega_d$.

2. *Speed of response to follow setpoints with minimum required response time* $\tau_r = 1/\omega_r$. *Must require* $\omega_B > \omega_r$. The requirement comes in addition to the bandwidth requirement imposed by the disturbances.

3. *Input constraints for disturbances. Must require* $|g(j\omega)| > |g_d(j\omega)|, \quad \forall \omega < \omega_d$. This is needed to avoid input constraints for perfect rejection of disturbance $d(j\omega) = 1$.

4. *Input constraints for setpoints. Must require* $|g(j\omega)| > R_{max}, \forall \omega < \omega_r$. This is needed to avoid input constraints ($|u(j\omega)| < 1$) for perfect tracking of $|r(j\omega)| = R_{max}$. Here ω_r is the frequency up to which setpoint tracking is desired, and R_{max} is the magnitude of the setpoint change relative to the allowed control error. Often $R_{max} = 1$.

   In the frequency range up to the bandwidth ω_B there should not be any time delays, RHP-zeros or high-order plant dynamics that need to be counteracted. We get

5. *Time delay* θ. *Must require* $\omega_B < 1/\theta$.

6. *Real RHP-zero at* $s = z$. *Must require* $\omega_B < z/2$.

7. *Phase lag constraint. In most practical cases:* $\omega_B < \omega_{g180}$.

   Here ω_{g180} is the frequency at which the phase of $g(j\omega)$ is -180^o. This condition is not a fundamental limitation, but more of a practical limitation. In particular it applies if the phase drops rather quickly around the frequency ω_{g180}.

8. *Real open-loop unstable pole at* $s = p$. *We need fast control to stabilize the system and must approximately require* $\omega_B > p$.

4 APPLICATIONS

4.1 Room heating

Consider the problem of maintaining a constant room temperature. A heat balance yields the following differential equation for the temperature T in the room

$$\frac{d}{dt}(C_V T) = Q + k(T_o - T) \qquad (33)$$

Here Q [W] is the heat input, T_o is the outdoor temperature, and the term $k(T_o - T)$ [W] represents the heat loss due to heat conduction through the walls or due to inflow of fresh air [5]. Consider a case where the heat input Q is 2000W and the difference between indoor and outdoor temperature $T - T_o$ is 20K. Then the steady-state energy balance yields $k = 2000/20 = 100$ W/K.

Let the heat capacicity be $C_V = 100$ kJ/K [6]. On introducing deviation variables and taking the Laplace transform we get

$$\Delta T(s) = \frac{1}{\tau s + 1}\left(\frac{1}{k}\Delta Q(s) + \Delta T_o(s)\right); \quad \tau = \frac{C_V}{k} \qquad (34)$$

The time constant for this example is $\tau = 100 \cdot 10^3/100 = 1000$s $= 17$ min which seems reasonable (for a increase in heat input it will take about 17 min for the temperature to reach 63% of its steady-state increase).

Problem statement. Feedback control should be used to maintain approximately constant room temperature. The measurement delay for T is $\theta = 100$s. Assume the acceptable variations in room temperature are ± 1K, i.e., $T_{max} = 1$K. Furthermore, assume that heat input can vary between 0 W and 4000 W, i.e., the heat input is 2000 ± 2000 W so $Q_{max} = 2000$ W. Finally, the expected variations in outdoor temperature are ± 10 K, i.e, $T_{o_{max}} = 10$ K.

- Is the process controllable with repect to disturbances?
- Is the process controllable with respect to setpoint changes[7] of magnitude ± 3 K when the desired response time for setpoint changes is $\tau_r = 1000$ s (17min) ?

Solution. A critical part of the controllability analysis is scaling, and we introduce the following scaled variables

$$y = \Delta T/1\text{K}; \quad u = \Delta Q/2000\text{W}; \quad d = \Delta T_0/10\text{K} \qquad (35)$$

The model in terms of scaled variables then becomes

$$y = g(s)u + g_d(s)d \qquad (36)$$

$$g(s) = \frac{20}{1000s + 1}; \quad g_d(s) = \frac{10}{1000s + 1} \qquad (37)$$

The frequency responses of these transfer functions are shown in Fig. 9.

[5] The heat loss may be represented by $qc_P(T_o - T) + UA(T_o - T)$ where the first term represents the convective heat transfer (difference in energy of inflow and outflow of air) and the second term represents the heat loss through the walls and windows. Thus $k = qc_p + UA$, where q [kg/s] is the flowrate, c_p [J/ kg,K] is the heat capacity, U [W/m^2,K] is the heat transfer coefficient, and A [m^2] is the wall area.

[6] The value C_V=100 kJ/K corresponds approximately to the heat capacity of air in a room of about 100 m^3. Thus we neglect heat accumulation in the walls.

[7] The setpoint change may be due to a desired increase in temperature when we come home from work or get up in the morning.

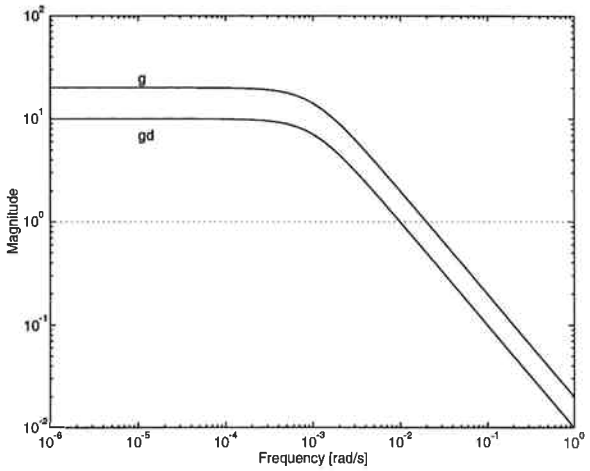

Figure 9: Frequency responses for room heating example

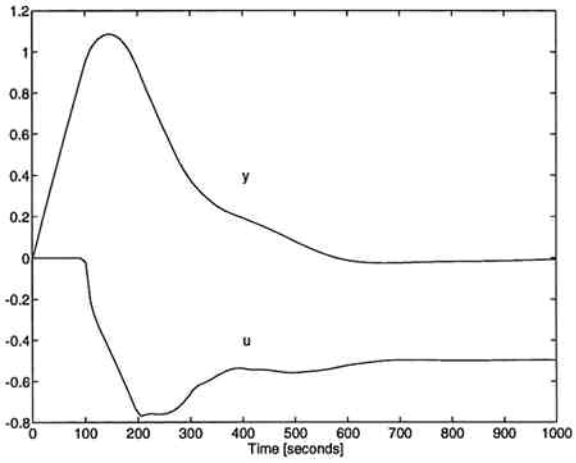

Figure 10: Feedback control for room heating example using PID controller. Step disturbance in outdoor temperature.

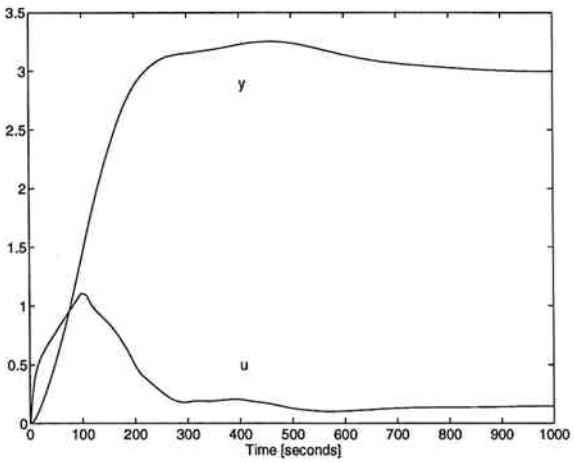

Figure 11: Feedback control for room heating example using PID controller. Setpoint change $3/(150s+1)$.

1. Disturbances. Feedback control is necessary up to the frequency $\omega_d = 10/1000 = 0.01$ rad/s, where $|g_d|$ crosses 1 in magnitude ($\omega_B > \omega_d$). This is exactly the same frequency as the upper bound given by the delay, $1/\theta = 0.01$ rad/s ($\omega_B < 1/\theta$). We therefore conclude that the system is barely controllable for this disturbance. No problems with input constraints are expected since $|g| > g_d|$ at all frequencies. This conclusion is supported by the closed-loop simulation in Fig. 10 for a unit step disturbance (corresponding to a sudden 10 K increase in the outdoor temperature) using a PID-controller with $K_c = 0.4, \tau_I = 200$s, $\tau_D = 60$s (the tunings are in terms of scaled variables for a controller in cascade form with derivative action over one decade). The output error exceeds its allowed value of 1 for a very short time after about 100 s, but then returns quite quickly to zero. The input goes down to about -0.8 and thus remains within its allowed bound of ± 1.

2. Setpoints. The plant is also controllable with respect to the desired setpoint changes. First, the delay is 100s which is much smaller than the desired reponse time of 1000s, and thus poses no problem. Second, $|g(j\omega)| \geq R_{max} = 3$ up to about $\omega = 0.007$ [rad/s] which is significantly higher than the required $w_r = 1/\tau_r = 0.001$ [rad/s]. This means that input constraints pose no problem. In fact, we should be able to achieve response times of about $1/0.007 = 150$s without reaching input constraints. This is confirmed by the simulation in Fig.11 for a desired setpoint change $3/(150s+1)$ using the same PID controller as above.

4.2 First-order with delay process

Consider disturbance rejection for the following process

$$g(s) = k\frac{e^{-\theta s}}{1+\tau s}; \quad g_d(s) = k_d\frac{e^{-\theta_d s}}{1+\tau_d s} \qquad (38)$$

In addition there is a measurement delay θ_m for the output and θ_{md} for the disturbance. All parameters have been appropriately scaled such that at each frequency $|u| < 1, |d| < 1$ and we want $|y| < 1$.

Problem: a) For each of the eight parameters $k, \tau, \theta, k_d, \tau_d, \theta_d, \theta_m and \theta_{md}$ state qualitatively what value you would prefer to have good controllability (large, small, no effect). Give answers both for the case of pure feedforward control and pure feedback control. b) State any quantitative relationships between the parameters that need to be satisfied to have acceptable controllability.

Solution: a) Qualitative results are given in Table 1. Essentially, the effect of the input should be as large and quick as possible, whereas the opposite is true for the disturbance. The main difference between feedback and feedforward control

	Feedback control	Feedforward control
k	Large	Large
τ	Small	Small
θ	Small	Small
k_d	Small	Small
τ_d	Large	Large
θ_d	No effect	Large
θ_m	Small	No effect
θ_{md}	No effect	Small

Table 1: Desired value of parameters to have good controllability.

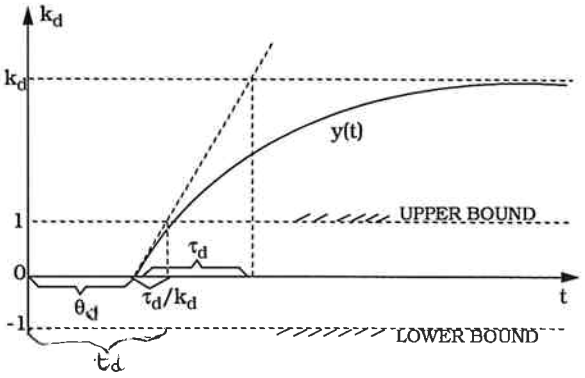

Figure 12: Response for step disturbance, $g_d = \frac{k_d e^{-\theta_d s}}{(1+\tau_d s)}$. Note: θ in plot should be θ_d

is that a delay for the disturbance has no effect for feedback control, while it is an advantage for feedforward control as it leaves more time to take the appropriate control action. We now want to quantify the statements in Table 1.

b) To derive quantitative results we may use of the rules from Section 3.7. Assume $k_d > 1$ such that control is needed to have acceptable performance ($|y| < 1$). From Rule 1 we have that control is needed up to the frequency ω_d where $|g_d(j\omega_d)| = 1$, i.e.,

$$w_d \approx k_d/\tau_d \tag{39}$$

To avoid input saturation (i.e, to have $|u| < 1$) we have from Rule 3 that that $|g(j\omega)| > |g_d(j\omega)|$ for frequencies $\omega < \omega_d$. Specifically, to have $|u| < 1$ we must require for both feedback and feedforward control

$$k > k_d; \quad k/\tau > k_d/\tau_d \tag{40}$$

The required bandwidth for disturbance rejection is ω_d. Thus, we must require for feedback control that $\omega_d < 1/\theta_{tot}$, where θ_{tot} is the total delay around the loop. That is, to have $|y| < 1$ for feedback control we require (combine Rule 1 and Rule 5)

$$\theta + \theta_m < \tau_d/k_d \tag{41}$$

For feedforward control any delay for the disturbance itself yields a smaller "net delay", and to have $|y| < 1$ we require

$$\theta + \theta_{md} < \tau_d/k_d + \theta_d \tag{42}$$

4.3 Step response controllability analysis

The controllability analysis presented in this paper is based on the frequency domain. However, most engineers feel much more comfortable with the time domain. The purpose of this example is to analyze the controllability using step response for a simple first-order process.

Let the model from the disturbance to the output be first-order with delay, i.e.,

$$g_d(s) = k_d e^{-\theta_d s}/(1 + \tau_d s)$$

Consider a unit step disturbance, $d = 1$. Without control the output response is

$$y(t - \theta_d) = k_d(1 - e^{-t/\tau_d}) \tag{43}$$

The response is shown graphically in Fig. 12. Since $k_d > 1$ the output $y(t)$ will exceed 1 after some time. Disregarding for a moment the delay, the time where $y(t) = 1$ is at $t_d = -\tau_d \ln(1-\frac{1}{k_d}) \approx \tau_d/k_d$ (the approximation holds for $k_d >> 1$ and corresponds to the point where the initial tangent of the time response crosses 1, see Fig. 12). Assuming that we measure the disturbance, the "minimum reaction time" to achieve $|y| < 1$ is then (see Fig. 12)

$$t_d = \tau_d/k_d + \theta_d$$

This is then an upper bound on the allowed delay in the process. This is the same value as was obtained in Eq.(42) using the frequency domain in the case of feedforward control.

From this example it seems like a step response controllability analysis yields results similar to the frequency domain, at least for a first-order process and feedforward control. However, for feedback control a step response controllability analysis is generally less suitable, as it explained next.

First, it can *not* be used to estimate the required speed of response for counteracting disturbances.

For example, a possible step response controllability method for disturbances may be based on the following idea: "Generate a unit step disturbance and find the time t_d it takes before the output exceeds its maximum value (which is 1 in terms of the scaled variables used in this paper). Acceptable controllability is possible if the process has a minimum response time (including delay) less than t_d". This statement is *not* correct. We know that the presence of a disturbance delay θ_d should *not* matter with feedback control. A frequency reponse analysis is consistent with this as the frequency ω_d where $|g_d(j\omega_d)| = 1$ is independent of any delay in g_d. On the other hand, from a step response analysis we find that the time t_d *does* depend on the delay in the disturbance (see Fig. 12) and a step response analysis is optimistic.

As another example, consider two disturbances for which the step responses are almost identical, but the magnitude frequency responses are very different

$$g_{d1}(s) = \frac{k_d}{1+\tau_d s} e^{-\theta_d s}; \quad g_{d2}(s) = \frac{k_d}{1+\tau_d s} \frac{1}{(1+\frac{\theta_d}{n}s)^n} \tag{44}$$

Assume $k_d >> 1$. Then the first disturbance is very difficult (in fact, acceptable controllability is impossible), whereas the second disturbance may be counteracted provided the process delay is less than θ_d. A frequency analysis yields this result, whereas the step response analysis again is optimistic for the first disturbance with the delay.

Second, a step response analysis is less suitable for analyzing the effect of constraints at high frequencies. At steady state the two methods yield the same result ($k > k_d$). However, the condition $k/\tau > k_d/\tau_d$ which applies at high frequency is difficult to derive from a step response analysis unless the delays are neglected.

Third, there are disadvantages with an open-loop step response analysis for high-order systems and for systems with oscillations.

However, one advantage with a step response analysis is that nonlinearity may be included.

In conclusion, the frequency domain should generally be used for controllability analysis, and the purpose of this example was *not* to suggest using step responses, but to provide another justification for the usefulness of the frequency domain.

4.4 Design of buffer tanks.

Buffer tanks are frequently used in the process industry to dampen disturbances in temperature, concentration and flow. For "quality" (e.g., temperature and concentration) disturbances the idea is to dampen high-frequency disturbances by use of a well-mixed tank, and level control is not important. For flowrate disturbances the level control is used actively to dampen the disturbance and mixing is not important. Of course, it is possible to use the same tank for both kinds of disturbances - design of the tank must then be based on the most difficult disturbance from a control point of view.

Although buffer tanks are often introduced for control purposes, they are usually sized in a rather *ad hoc* manner without explicitly considering the expected disturbances and desired control objectives. Fortunately, the results on controllability with respect to disturbances presented in this paper, provide the basis for a quantitative approach.

To design the buffer tank consider the controllability of the plant when the disturbance transfer function $g_d(s)$ is replaced by

$$\hat{g}_d(s) = g_d(s)h(s) \tag{45}$$

where $h(s)$ represents the transfer function for the buffer tank(s). Presumably, the controllability is not acceptable without the buffer tank (i.e., with $h(s) = 1$), that is, the effect of the disturbance is too large such that, either the required speed of response is not achievable (typically due to a process delay θ), or the required inputs to reject the disturbance are too large.

The objective of the buffer tank is then to dampen the disturbance such that:

1. The required speed of response *is* achievable, that is, for a process delay θ we must require

$$|\hat{g}_d(j\omega_\theta)| \leq 1; \quad \omega_\theta \stackrel{\text{def}}{=} 1/\theta \tag{46}$$

2. Input constraints cause no problem, that is

$$|g(j\omega)| \leq |\hat{g}_d(j\omega)|, \quad \forall \omega \leq \hat{\omega}_d \tag{47}$$

where $\hat{\omega}_d$ is the frequency where $|\hat{g}_d(j\hat{\omega}_d)| = 1$.

That is, $h(s)$ should be selected such that requirements (46) and (47) are satisfied. Although this is rather straightforward, we consider it in some detail because it yields some rather interesting results.

We shall first consider design of buffer tanks for "quality" disturbances (temperature and concentration) and then consider flow rate disturbances. The main difference between these cases is that for quality disturbances $h(s)$ has to be a series of first-order lags, whereas for flowrate disturbances one may use the level controller to get a desired $h(s)$.

4.4.1 Quality disturbances.

Consider a tank with constant volume V [m^3] and with an inlet and outlet flowrate q [m^3/s]. Let c_{in} denote the inlet concentration or temperature to the tank, and c the corresponding value in the outlet stream. A material or energy balance for a perferctly mixed tank yields

$$V\frac{dc}{dt} = qc_{in} - qc \tag{48}$$

The transfer function for one tank then becomes

$$c(s) = h_1(s)c_{in}(s); \quad h_1(s) = \frac{1}{\tau_h s + 1} \tag{49}$$

where $\tau_h = V/q$ [s] is the residence time in the tank (the subscript h denotes holdup). For n equal tanks in series with total residence time τ_h and total volume V, $h_1(s)$ is replaced by

$$h_n(s) = 1/(\frac{\tau_h}{n}s + 1)^n \tag{50}$$

Typical frequency responses are presented in Figure 13. We have that $|h_n(j\omega)| \approx 1$ at frequencies $w < n/\tau_h$, that is, the buffer tanks are only effective for reducing the effect of disturbances at frequencies higher than the residence time of the individual tanks. The high-frequency asymptote is $|h(j\omega)| \approx (\frac{n}{\tau_h})^n$ so for rejecting high-frequency

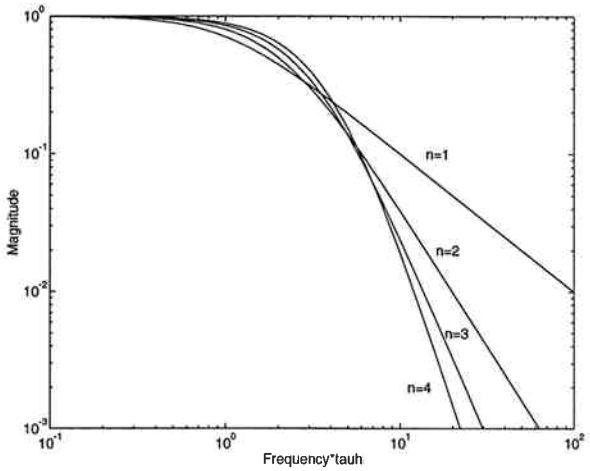

Figure 13: Frequency responses for n tanks in series with total residence time τ_h, $h_n(s) = 1/(\frac{\tau_h}{n}s+1)^n$.

disturbances it is best to use many tanks if the objective is to minimize the total volume. At intermediate frequencies we see from Fig.13 that there is a small frequency range where fewer tanks is better, but the main reason for using fewer tanks is usually to save on equipment cost (including costs for piping, mixing and level control), and not to minimize the total volume.

The design of the buffer tank(s) now depends on which of the two requirements (46) and (47) is most difficult to satisfy.

1. **Delays is the problem.** In this case the objective tanks is to make the curve for $|\hat{g}_d|$ cross 1 in magnitude before the frequency $\omega_\theta = 1/\theta$. Thus, we need to select $h(s)$ such that

$$|h_n(j\omega_\theta)| \cdot |g_d(j\omega_\theta)| \leq 1 \tag{51}$$

Introduce the factor by which the effect of the disturbance must be reduced

$$f = |g_d(j\omega_\theta)| \tag{52}$$

We must at least require $|h_n(j\omega_\theta)| = 1/f$. This may be solved graphically using Fig.13. Alternatively, for n equal tanks in seriesEq.(50) yields the required total residence time

$$\tau_h = \theta n\sqrt{f^{2/n}-1} \tag{53}$$

where θ is the total delay in the feedback loop. The optimal number of tanks can then be found by taking into account cost for equipment, piping, control systems (each tank may require a level controller), etc. As an example, for $f = 10$ we get

No. of tanks, n	1	2	3	4
Total residence time,τ_h	9.94θ	6.00θ	5.72θ	5.88θ

In this case the smallest total volume is obtained with 3 tanks, but with 2 tanks the required volume is only 4% larger and is clearly preferable. In practice one would probably prefer to use only 1 tank which has 66% larger total volume, but which saves additional equipment.

Remark. From (53) we find for large values of f (i.e., $f^{2/n} >> 1$) the following limiting value for the total residence time

$$\tau_h \approx \theta n f^{1/n} \tag{54}$$

Thus, with one tank the residence time should be approximately equal to θf.

2. **Constraints is the problem.** If constraints is the problem then we must in any case require that there are no problems at steady-state, that is, we must have $|g(0)| > |g_d(0)|$. Let ω_e be the frequency where $|g| = |g_d| > 1$. The objective of the buffer tanks in this case is to make $|\hat{g}_d|$ smaller than $|g|$ in the frequency range from ω_e to $\hat{\omega}_d$.

The following procedure may be used to achieve this: Let $n_e > 0$ be the difference in the slope of $|g_d|$ and $|g|$ (on a log-log plot) at frequency ω_e. Assume that the difference in slopes remains constant or decreases in the frequency range from ω_e to the frequency where $|g| = 1$. Select the number of tanks n equal to n_e, and select the holdup of the individual tanks as $1/\omega_e$, that is, select the overall residence time as $\tau_h = n_e/\omega_e$.

Example. Let

$$g(s) = \frac{200}{(55900s+1)(89.4s+1)}$$

$$g_d(s) = \frac{100(5000s+1)}{(55900s+1)(89.4s+1)}$$

We find in this case $\omega_e = \frac{200}{100\cdot 5000} = 0.0004$ [rad/s] (using asymptotic values) and $n_e = 0-(-1) = 1$, and the difference in slopes remains constant at high frequencies. To reduce the effect of the disturbance to an acceptable level such that input constraints are avoided, we then need $n_e = 1$ buffer tank with residence time $1/\omega_e = 2500$s $= 0.7$h.

Problems.

1. The effect of a concentration disturbance must be reduced by a factor of 100 at the frequency 0.5 rad/min. The disturbances should be dampened by use of buffer tanks and the objective is to minimize the total volume. How many tanks in series should one have? What is the total residence time?

2. The feed to a distillation column has large variations in concentration and the use of one buffer tank is suggest to dampen these. The effect of the feed concentration d on the product composition y is given by (scaled variables, time in minutes)
$$g_d(s) = e^{-s}/3s \tag{55}$$
(that is, after a step in d the output y will, after an inital delay of 1 min, increase in a ramplike fashion and reach its maximum allowed value (which is 1) after another 3 minutes). Feedback control should be used and there is a total process delay of 3 min. What should the residence time in the tank be?

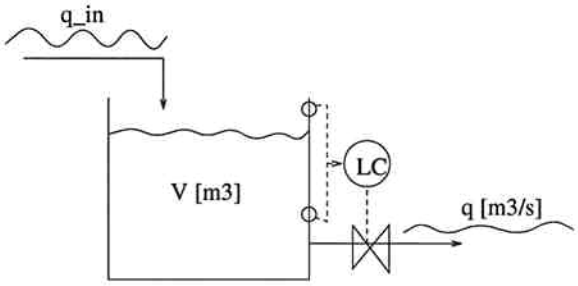

Figure 14: Use of slow level control to dampen flowrate disturbance.

3. In terms of minimizing the total volume it is optimal to have buffer tanks of equal size. The only "exception" is for low-frequency disturbances (frequencies lower than about $1/\theta$, see Figure 13) where it better to use fewer tanks (that is, one tank has zero volume). Consider the case with two buffer tanks with total residence time τ_h. Let the residence time in one of the tanks be x. Then
$$h_2(s) = \frac{1}{(1+xs)(1+(\tau_h - x)s} \quad (56)$$
and the high-frequency asymptote becomes $|h_2(j\omega)| \approx 1/[(\tau_h - x)x\omega^2]$ which is minimized by selecting $x = \tau_h/2$, that is, the tanks should be of equal size to get the best disturbance attentuation with a given total volume. For n tanks in series we get the same result by considering the high-frequency asymptote. Show this.
4. Is there any reason to have buffer tanks in parallell (they must not be of equal size because then one may simply combine them) ?
5. What about parallel pipes in series (pure delay). Is this a good idea?

4.4.2 Flow rate disturbances

Flowrate disturbances may be dampened by use of a slow level controller as illustrated in Fig. 14. Let V [m^3] denote the volume of the buffer tank and let q_{in} and q [m^3/s] be the inlet and outlet flowrates. The dynamic model for the tank and the level control system is
$$V(s) = \frac{1}{s}(q_{in}(s) - q(s); \quad q(s) = c(s)V(s) \quad (57)$$
where $c(s)$ is the transfer function of the level controller (including measurement and actuator devices). We get
$$V(s) = \frac{1}{s + c(s)} q_{in}(s) \quad (58)$$
and the transfer function of interest becomes
$$q(s) = h(s)q_{in}(s); \quad h(s) = \frac{c(s)}{s + c(s)} \quad (59)$$

For flowrate disturbances we have more freedom in selecting $h(s)$ because we can select the algorithm for the level controller, $c(s)$. On the other hand, the level will vary so the size of the tanks must be such that the level does not reach constraints. The design of a buffer tank for flowrate disturbances then consists of two steps

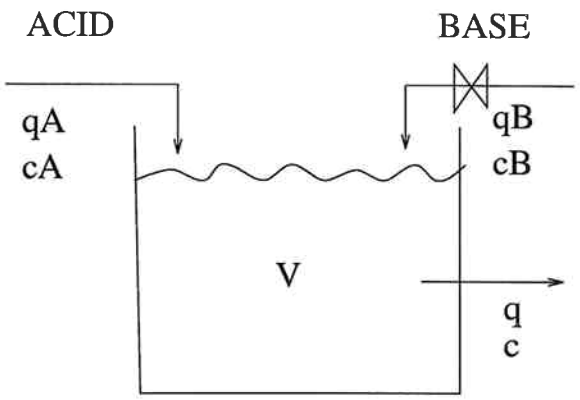

Figure 15: pH-neutralization process.

1. Design the level *controller* such that $h(s)$ has the desired shape, that is, such that (46) and (47) are satisfied.
2. Design the *size of the tank* such that the level remains within the allowed range for the expected disturbances.

First-order filtering. In many cases the desired $h(s)$ has the shape
$$h(s) = 1/(\tau s + 1) \quad (60)$$
and we see from (59) that the required controller is a P-controller with gain $K_c = 1/\tau$. The response for the volume in the tank is given by (58), that is, we get $V(s) = \frac{\tau}{\tau s+1} q_1(s)$. This transfer function has its largest value equal to τ at low frequencies, and if the inlet flowrate varies within its full range $\pm q_{in}$ ($\pm 100\%$), we get that the volume will vary within $\pm\tau q_{in}$. This is in terms of deviation variables, and the total volume of the tank should be $2\tau q_{in}$. We then find, as one probably may expect, that the average residence time in the tank, τ_h, should be equal to the desired filter time constant τ.

Remark. In some cases one may want to add a slow integral action to the controller to reset the volume (level) to its nominal value, but this is not always desired. For example, if q_{in} is at its maximum value, then we may want V to stay at a large value to anticipate a possible large reduction in q_{in}.

Problems.

1. Second-order filtering. Let $h(s) = \frac{1}{(\tau s+1)^2}$. Design the controller and tank in this case.
2. Is there any advantage of using more than one tank in this case?

4.5 Neutralization process

The derived controllability results are next applied to a neutralization process, and we find that more or less heuristic design rules given in the literature follow directly. The key point is to consider disturbances and scale the variables properly.

One mixing tank. Consider the process in Figure 15 where a strong acid (pH= -1) is neutralized by a strong base (pH=15) in a tank with volume V= 10 m^3 to make q=0.01 m^3/s of "salt water". The pH in the product stream is adjusted to be in the range 7 ± 1 ("salt water") by manipulating the amount of base, q_B. The delay for the measurement of pH is $\theta = 10$s. Details about the dynamic model are given in Appendix 2. Introduce the excess of acid c [mol/l] defined as

$$c = c_H - c_{OH} \tag{61}$$

Somewhat surpisingly, we find that in terms of c the dynamic model, which is usually believed to be strongly nonliner, is given by that of a simple mixing process

$$\frac{d}{dt}(Vc) = q_A c_A + q_B c_B - qc \tag{62}$$

Introduce the following scaled variables

$$y = \frac{c}{10^{-6}}; \quad u = \frac{q_B}{q_B^*}; \quad d = \frac{q_A}{0.5 q_A^*} \tag{63}$$

where suprescript * denotes the steady-state value. The appropriately scaled linear model then becomes (see Appendix 2)

$$y = \frac{k_d}{1+\tau s}(-2u + d); \quad k_d = 2.5 \cdot 10^6 \tag{64}$$

where $\tau = V/q = 1000$s. The output is extremely sensitive to both u and d and the large gain is easily explained: A change $d = 1$ corresponds to a 50% increase in the amount of acid which has a concentration of 10 mol/l of H^+ (pH=-1). This increases the amount of H^+ in the product from 0 to 2.5 mol/l, while the largest allowed amount of H^+ in the product is 10^{-6} mol/l (pH=6), thus the gain in terms of scaled variables is $k_d = 2.5/10^{-6}$.

Input constraints do not pose a problem since $|g| = 2|g_d|$ at all frequencies. The main control problem is the high disturbance sensitivity, and from (39) we find the frequency up to which feedback is needed

$$\omega_d \approx k_d/\tau = 5000 \text{ rad/s} \tag{65}$$

This requires a response time of $1/5000 = 0.2$ millisecond. However, there is a delay $\theta = 10s$ so the bandwidth must be less than $\omega_B < 1/\theta = 0.1$ rad/s. From the controllability analysis we therefore conclude that acceptable control using a single tank is impossible.

Design change: Several tanks. The only way to improve the controllability is by design changes. The most useful change in this case is to do the neutralization in several steps. This can be considered as a special case of the buffer tank example considered above: The acid and base is mixed and is then send to one or more buffer tanks, and the measured pH of the final stream is used to adjust the addition of base, as is illustrated with two tanks in Figure 16. The mixing process itself is assumed immediate so in the following

$$g_d(s) = k_d \tag{66}$$

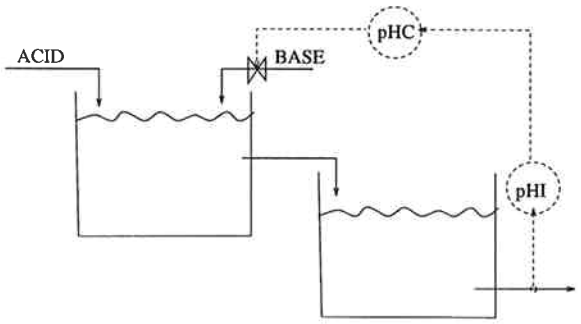

Figure 16: Control of neutralization process using two tanks.

and the objective is to find an appropriate $h(s)$ such that the "new" process $\hat{g}_d(s) = g_d(s)h(s)$ has acceptable controllability in terms of having acceptable self-regulation.

As we already found, the disturbance has a gain of $k_d = 2.5 \cdot 10^6$. Thus, if the main control problem is a delay of θ= 10s, we must design buffer tanks which reduces the effect of the disturbance by a factor $f = 2.5 \cdot 10^6$ at the frequency $\omega_\theta = 1/\theta = 0.1$ [rad/s]. The required total residence time, τ_h, is given by Eq.(53), and the corresponding total volume is

$$V = \tau_h q \tag{67}$$

where q = 0.01 m^3/s. From this we find that the following designs have the same controllability with respect to disturbance rejection:

No. of tanks n	Total volume V [m^3]	Volume each tank [m^3]
1	250000	250000
2	316	158
3	40.7	13.6
4	15.9	3.98
5	9.51	1.90
6	6.96	1.16
7	5.70	0.81
18	3.66	0.20
30	3.89	0.13

With one tank we need a volume corresponding to that of the worlds largest ship to get acceptable controllability. The minimum total volume is obtained with 18 tanks of about 203 l each - giving a total volume of 3.662 m^3. However, taking into the account the additional cost for extra equipment such as piping, mixing and level control, we would probably select a design with 3 or 4 neutralization tanks for this example.

Remarks.

1. Further remarks on some of the practical aspects and comparison with previous work are found in Skogestad (1994).

2. The use of several mixing tanks in series can be compared to playing golf: It is almost impossible to hit the hole with one stroke, but with 5 strokes or more almost anyone can do it.

3. Traditionally, a "feedforward" approach has been taken when considering controllability of such processes, and one key argument has been that control is difficult because on needs to adjust the amount of base extremely accurately to counteract the disturbance in the acid. This is a valid argument for feedforward control, but *not* for feedback control as the feedback control action will be able to adjust the input accurately. As demonstrated above the key problem for feedback control is that the output is extremely sensitive to disturbances (k_d and ω_d are large), which requires an extremely high bandwidth.

4. Of course, feedforward control based on measuring q_A and c_A can be used in addition to feedback to improve performance. According to McMillan (1984) one can typically save one buffer tank using a well designed feedforward controller.

5. The results given above compare well with results by other authors. A simple shortcut method given by McMillan (1984) is to use about one mixing tank for each 2 units change in pH. For example, with a pH change of 8, as in our example (from pH 15 to 7), four tanks is recommended.

5 Conclusions

The paper has presented in a tutorial manner a detailed controllability analysis for SISO systems using the frequency domain. The analysis may be used to answer whether or not a given plant is controllable, and thus extends beyond the traditional use of "controllability indicators". The method has been applied to several examples, including design of buffer tanks for pH-processes. It is found that previously presented heuristic rules follow directly from the analysis. The key point is to consider disturbances and scale the variables properly.

Acknowledgement. Manfred Morari was the first to consider a rigorous approach to controllability analysis. He also pointed out to me to the paper of Ziegler and Nichols (1943) who first introduced the term controllability in the control literature.

References

[1] Balchen, J. G. and K. Mumme, 1988. "Process Control. Structures and Applications", Van Nostrand Reinhold, New York.

[2] Hovd, M. and S. Skogestad, 1992. "Simple Frequency-Dependent Tools for Control Structure Analysis, Structure Selection and Design", *Automatica*, **28**, 989-996.

[3] McMillan, G.K., 1984, *pH Control*, Instrument Society of America.

[4] Morari, M., 1983. "Design of resilient processing plants III, A general framework for the assessment of dynamic resilience", *Chem. Eng. Sci.*, 38, 1881-1891.

[5] Morari, M. and E. Zafirou, 1989. *Robust Process Control*, Prentice Hall.

[6] Rosenbrock, H.H., 1966, "On the design of linear multivariable control systems", *3rd IFAC World Congress*, Paper 1a.

[7] Rosenbrock, H.H., 1970. *State-space and Multivariable Theory*, Nelson, London.

[8] Seborg, D.E., T.F. Edgar and D.A. Mellichamp, 1989, *Process Dynamics and Control*, Wiley, New York.

[9] Skogestad, S., 1994, "A Procedure for SISO controllability analysis - with application to design of pH processes", Preprints IFAC Workshop on Integration between process design and control (IPDC'94), Baltimore, June 1994, 23-28.

[10] Skogestad, S. and E.A. Wolff, 1992, "Controllability measures for disturbance rejection", *Preprints IFAC Workshop on Interactions between process design and control, London, Sept. 1992*, Edited by J.D. Perkins, Pergamon Press, 127-132.

[11] Youla, D.C., J.J. Bongiorno and C.N. Lu, 1974, "Single-loop Feedback Stabilization of Linear Multivariable Dynamical Plants", *Automatica*, **10**, 159-173.

[12] Ziegler, J.G. and N.B. Nichols, 1943, "Process Lags in Automatic Control Circuits", *Trans. ASME*, **65**, 433-444.

APPENDIX 1. Scaling procedure

Let the unscaled variables (in their original units) be identified by a prime ($'$). The model in terms of unscaled variables is

$$y' = g'(s)u' + g'_d(s)d' \tag{68}$$

$$e' = y' - r' \tag{69}$$

where $g'(s)$ and $g'_d(s)$ denote the unscaled ("original") transfer functions.

The normalized or scaled variables are obtained by normalizing each variable by its maximum allowed magnitude

$$d = \frac{d'}{d_{max}}, u = \frac{u'}{u_{max}}, e = \frac{e'}{e_{max}}, y = \frac{y'}{e_{max}}, r = \frac{r'}{e_{max}} \tag{70}$$

Here

- u_{max} - largest allowed magnitude change in u (typically because of saturation constraints)
- d_{max} - largest expected magnitude disturbance
- e_{max} - largest allowed magnitude control error for output
- r_{max} - largest expected magnitude change in setpoint

The maximum control error should typically be chosen by thinking of the largest deviation one can allow as a function of time, and not as the steady-state error. The same applies to the other maximum errors. Note that e, y and r are in the same units and have alll been normalized with respect to e_{max}. Let

$$R_{max} = \frac{r_{max}}{e_{max}} \tag{71}$$

denote the magnitude of the largest setpoint change relative to the allowed control error. In most cases

$R_{max} \geq 1$. With these scalings we have the following requirements at all frequencies

$$|d(j\omega)| \leq 1, |r(j\omega)| \leq R_{max}, |u(j\omega)| \leq 1, |e(j\omega)| \leq 1$$

that is, if we think in terms of sinusoids the variables $d(t)$, $u(t)$ and $e(t)$ should stay within the interval -1 to 1, and $r(t)$ within the interval $\pm R_{max}$.

Introducing (70) into (68) yields

$$e_{max}y = g'(s)u_{max}u + g'_d(s)d_{max}d$$

Define the *scaled* transfer function models as

$$g(s) = g'(s)\frac{u_{max}}{e_{max}}; \quad g_d(s) = g'_d(s)\frac{d_{max}}{e_{max}} \tag{72}$$

We then get the "new" model in terms of scaled variables and scaled transfer functions

$$y = g(s)u + g_d(s)d \tag{73}$$

$$e = y - r \tag{74}$$

In this paper we use the frequency domain, and use the same maximum value at all frequencies, although one may in some cases use frequency-dependent values. [8]. The assumption for the controllability analysis is then: *The (scaled) external signals d and r are persistent sinusoids with magnitude less than 1 and R_{max} at all frequencies, that is, $|d(j\omega)| < 1$ and $|r(j\omega)| < R_{max}$. Similarly, the allowed scaled inputs and allowed scaled control error should be less than 1 at all frequencies, that is, $|u(j\omega)| < 1$, and $|e(j\omega)| < 1$.* For example, we assume that $g_d(s)$ is scaled such that at each frequency the worst (largest) disturbance corresponds to $|d(j\omega)| = 1$, that is, in the time domain we consider a persistent disturbance of magnitude 1, $d(t) = 1 \cdot sin(\omega t)$.

APPENDIX B. Neutralization model

Derivation of model: Consider Fig.15. Let c_H [mol/l] and c_{OH} [mol/l] denote the concentration of H^+ and OH^--ions, respectively. Material balances for these two species yields

$$\frac{d}{dt}(Vc_H) = q_A c_{H,A} + q_B c_{H,B} - qc_H + rV$$

$$\frac{d}{dt}(Vc_{OH}) = q_A c_{OH,A} + q_B c_{OH,B} - qc_{OH} + rV$$

where r [mol/s,m^3] is the rate for the reaction $H_2O = H^+ + OH^-$ which for completely disocciated ("strong") acids and bases this is the only reaction in which H^+ and OH^- participate. We may eliminate r from the equations by taking the difference to get a differential equation in terms of the excess of acid, $c = c_H - c_{OH}$:

$$\frac{d}{dt}(Vc) = q_A c_A + q_B c_B - qc$$

Note: 1. This is the material balance for a mixing tank without reaction. The reason is that the quantity $c = c_H - C_{OH}$ is not affected (invariant) by the reaction. 2. c is the excess of acid and will take on negative values when pH is above 7.

Assume the feed concentrations c_A and c_B are constant. Linearization and Laplace transform yields

$$c(s) = \frac{1}{1+\tau s}\left(\frac{c_A^* - c^*}{q^*}q_A(s) + \frac{c_B^* - c^*}{q^*}q_B(s)\right)$$

where $\tau = V/q^*$ is the residence time and * is used to denote steady-state values. To derive this we have made use of the total material balance $dV/dt = q_A + q_B - q$ (alternatively one may assume V is constant but this is not strictly necessary) and the corresponding steady-state balance $q_A^* + q_B^* = q^*$. We now introduce the following scaled variables

$$y(s) = \frac{c(s)}{c_{max}}; \quad d(s) = \frac{q_A(s)}{q_{Amax}}; \quad u(s) = \frac{q_B(s)}{q_{Bmax}}$$

We use the following numbers: V= 10 m^3, $q_A^* = q_B^* =$ 0.005 m^3/s, $c_{H,A}^* = 10$ mol/l (corresponding to pH= -1 and $c_A = 10 - 10^{-15} \approx 10$ mol/l), $c_{OH,B}^* = 10$ mol/l (corresponding to pH=15 and $c_B^* = 10^{-15} - 10 \approx -10$ mol/l), $c^* = 0$ mol/l (corresponding to pH=7), $q_{Amax} = q_A^*/2 = 0.0025$ m^3/s, $q_{Bmax} = q_B^* =$ 0.005 m^3/s, and $c_{max} = 10^{-6}$ mol/l (i.e., pH= 7 ± 1.. Note from this that the largest disturbance is $\pm 50\%$ of q_A^*, while the largest input is $\pm 100\%$ of q_B^*.

[8] It is usually reasonable to assume that u_{max}, d_{max} and r_{max} are independent of frequency. However, one may select to make $e_{max}(j\omega)$ frequency dependent. Let e_{max}^0 denote the largest allowed steady-state error for the time response, and we usually have $e_{max}^0 < r_{max}$, that is, $R_{max} = r_{max}/e_{max}^0 > 1$. However, at high frequency it is reasonable to assume that $e_{max}(j\omega) \to r_{max}M$ where $M > 1$ (typically, $M = 2$). This follows since we cannot track very fast setpoint changes, and in fact may have to accept an error larger than 100% at high frequencies. A reasonable choice is then

$$e_{max}(s) = e_{max}^0 \frac{\tau_r \frac{r_{max}}{e_{max}^0}s + 1}{\frac{\tau_r}{M}s + 1}$$

where τ_r is the desired response time for the output (we have that $|e_{max j\omega}| \approx r_{max}$ for $\omega = 1/\tau_r$. At low frequencies one may let $e_{max}(j\omega)$ approach 0 to include a requirement of integral action. The approach outlined in this footnote leads directly into the problem of selecting weights for H_∞-control.