IFAC-Symposium ADCHEM'94, Kyoto, Japan, May 25-27, 1994.

CONTROLLABILITY ANALYSIS OF SISO SYSTEMS

Sigurd Skogestad*

Department of Chemical Engineering, University of Trondheim - NTH, N-7034 Trondheim, Norway

Abstract. The objective of this paper is to derive some fundamental results for controllability analysis of scalar 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.

Key Words. Feedback control; Achievable performance; Disturbances; Delay; Zeros; pH process; process design.

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. Is the plant easy to control? 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" these 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. 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 mostly qualitative. In most cases the "simulation approach" is used. However, this requires a specific controller design and specific values of disturbances and setpoint changes. In the end one never really knows if a result is a fundamental property of the plant or if it depends on these specific choices.

The main 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.

Definition of controllability. In this work the term "controllability" (of a plant) has the meaning *"inherent control characteristics of the plant"* or maybe better *"achievable performance" (irrespective of the controller)*. This usage is in agreement with most persons intuitive feeling about the term, and was also how the term was used historically 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"*. Note that controllability is a property of the plant (process) only, and can only be affected by changing the process itself, that is, by design modifications.

Unfortunately, in the 60's Kalman defined the term "controllability" in the very narrow meaning of "state controllability". 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. To distinguish our use of the term from that of Kalman we may use the term "output controllability".

Previous work. Except for the initial work by Ziegler and Nichols (1943), there does not seem to have been much progress on output controllability analysis until Rosenbrock (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 important notion of right half plane (RHP) zeros (which for scalar systems is directly related to inverse responses). The next important step towards a quantitative analysis was made by Morari (1983)

*E-mail: skoge@kjemi.unit.no, Phone: +47-7359-4154, Fax: +47-7359-4080.

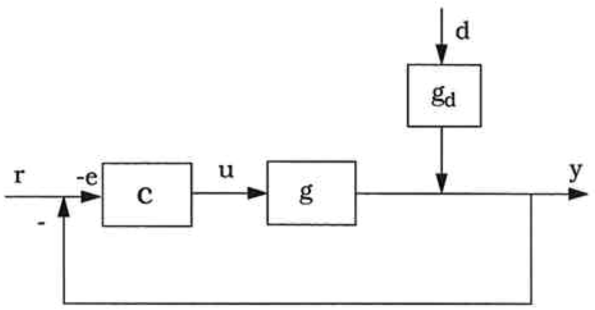

Fig. 1: Block diagram of feedback control system.

who made use of the notion of "perfect control". The main issue which was missing from his analysis was an explicit consideration of disturbances. Balchen and Mumme (1988, pp. 16-21, pp.47-48) present some nice controllability guidelines which are more specific that the rules from Seborg er al. (1989) given above, but most of them lack a theoretical justification. The issue of disturbances has 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 often 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 deleted 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

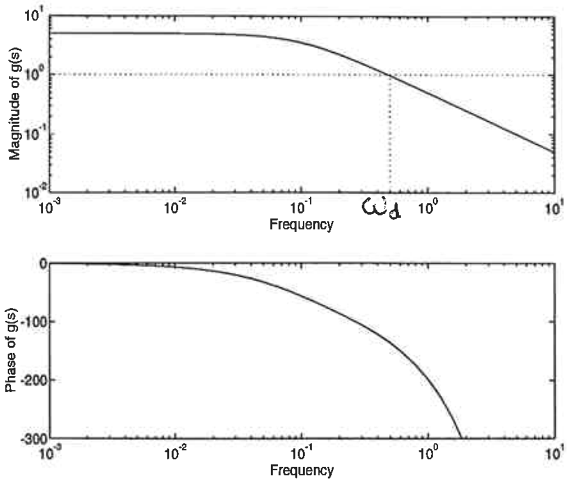

Fig. 2: Frequency response of $g(s) = 5e^{-2s}/(1+10s)$.

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. The physical interpretation 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, $y(t) = y_0 sin(\omega t + \phi)$. 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. We have the system gain $y_0/u_0 = |g(j\omega)|$ and phase shift $\phi = \angle g(j\omega)$ [rad].

In this paper we use a "frequency-by-frequency" approach and at each frequency consider the response to a sinusoidal input of unit magnitude ($u_0(\omega) = 1$). This yields the "frequency response" of the system where we consider the gain $y_o(\omega) = |g(j\omega)|$ (and possibly the phase shift $\angle g(j\omega)$) as a function of ω.

In Fig. 2 the frequency response (Bode-plot) is shown for a first-order system with time delay, $g(s) = ke^{-\theta s}/(1 + \tau s)$. 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

$$\omega_d \approx k/\tau \tag{6}$$

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).

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).

A frequency domain analysis, and in particular of 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. The reason is that the feedback control system usually will amplify frequencies corresponding to the closed-loop bandwidth, ω_B. 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.

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). The detailed scaling procedure is outlined in the Appendix.

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 $|r(j\omega)| < 1$*[1].

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

$$u = g^{-1}r - g^{-1}g_d d \tag{7}$$

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.(7) at frequencies where feedback is effective. Introducing the fact $cS = g^{-1}T$ into (5) yields the following expression for the input signal under feedback control

$$u = g^{-1}Tr - g^{-1}Tg_d d \tag{8}$$

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 (7). Consequently ideal control (inversion) requires *fast* feedback (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.

[1] In this paper we assume that r is less than 1 at at all frequencies, but in general the allowed magnitude of r may be frequency dependent.

3.1 Disturbances and bandwidth

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

$$y(j\omega) = g_d(j\omega)d(j\omega) \tag{9}$$

(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 avoid that the output exceeds 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

$$\omega_B > \omega_d \tag{10}$$

Example. Consider the disturbance model (recall Fig.2)

$$g_d(s) = k_d e^{-\theta_d s}/(1 + \tau_d s) \tag{11}$$

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. The following rule from the introduction is quantified:
    - Control outputs that are not self-regulating
3. The rule can be quantified as follows: Control outputs y for which $|g_d(j\omega)| > 1$ at some frequency.
4. 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.
5. To be more specific assume that the disturbance is increased by a factor α, and assume that at frequency ω_d the slope of $|g_d(j\omega)|$ on the Bode-plot is $-\beta$ [relative change in $|g_d|$/relative change in ω) (in the example above $\beta = 1$). Then the bandwidth has to be increased by a factor $\alpha\beta$ to counteract the increased disturbance.
6. Note that a delay in the disturbance model has no effect on the required bandwidth. On the other hand, for feedforward control a delay will make 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.(7) the input magnitude needed for perfect control ($e = 0$) is

$$|u| = |g_d|/|g| \quad (12)$$

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

$$|g(j\omega)| > |g_d(j\omega)|, \quad \forall \omega < \omega_d \quad (13)$$

Similarly, to be able to track a setpoint of magnitude 1 at each frequency ($r(j\omega) = 1$) one must require

$$|g(j\omega)| > 1, \quad \forall \omega < \omega_r \quad (14)$$

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

Remarks.

1. We have quantified the following rule from the introduction:
   - Select inputs that have large effects on the outputs.
2. The 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 $|g| > 1$ at frequencies where setpoint tracking is desired.
3. This 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 is 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.
4. For unstable plants we need a minimum bandwidth p to stabilize the system (see below). In this case we need $|g| > |g_d|$ and $|g| > 1$ up to the frequency p. Otherwise, the input will saturate, and the plant can not be stabilized.
5. Since the input needed for perfect control is independent of the control implementation, the bounds (13) and (14) apply also 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 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)

$$T = \prod_i \frac{-s + z_i}{s + \bar{z}_i} e^{-\theta s} \quad (15)$$

where $\bar{z}_i$ is the complex conjugate. 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.

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 can not remove this delay. Thus, even the "ideal" complementary sensitivity function will contain the delay,

$$T = e^{-\theta s} \quad (16)$$

The loop transfer function corresponding to this ideal response is $L = e^{-\theta s}/(1 - e^{-\theta s})$. At loe frequencies, $\omega\theta < 1$, we have $e^{-\theta s} \approx 1 - \theta s$ (Taylor series expansion of 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 is at $\frac{\pi}{3}\frac{1}{\theta} = 1.05/\theta$), that is, this is the bandwidth frequency. In practice, the "ideal" controller can not be realized, and this will provide an upper bound on the bandwidth and we have approximately

$$\omega_B < 1/\theta \quad (17)$$

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 can not remove the effect of this RHP zero. Thus, even the "ideal" complementary sensitivity function will contain the RHP-zero

$$T = \frac{-s + z}{s + z} \quad (18)$$

The loop transfer function corresponding to this ideal response is $L = (-s + z)/2s$. The low-frequency asymptote of $|L(j\omega)|$ crosses 1 at frequency $z/2$. In practice, the "ideal" controller can not be realized, and this will provide an upper bound on the bandwidth and we have approximately

$$\omega_B < \frac{z}{2} \quad (19)$$

Remarks on bounds (17) and (19).

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.

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

- Select inputs that rapidly effect the controlled variables

3. To reject a disturbance we obtained the requirement $\omega_B > \omega_d$. Combining this with (17) 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 (19) for RHP-zeros assumes that we want to use u for "slow control" of control 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 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. For example, consider a case where $g(s)$ contains a term s in the numerator (i.e., a zero at the origin). In case the steady-state gain is zero, and this input can be used only for transient 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 bounds apply also to feedforward control. This follows since the the ideal T in (15) corresponds to the input u that minimized the ISE of the output irrespective of the control implementation.

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. For $K_c > p$ the asymptote of the loop transfer function $|L|$ crosses 1 at frequency $1/K_c$. Combining these two pieces of information we conclude that the approximate minimum bandwidth needed for stabilization is

$$\omega_B > p \tag{20}$$

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 (20) 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 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 be stabilized. This seems to be inconsistent with the above result. It is not, since these results required performance and not only stability. Thus, the requirement $p < z/2$ is indeed needed for obtaining acceptable control performance (at least at low frequencies).

3.5 Phase lag

Consider a minimum-phase process of the form

$$g(s) = \frac{k}{(1+\tau_1 s)(1+\tau_2 s)\cdots} = \frac{k}{\prod_{i=1}^{n}(1+\tau_i s)} \tag{21}$$

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 ($\angle g(j\omega) \to -n90^o$) 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

$$\text{Practical bound}: \quad \omega_B < w_{g180} \tag{22}$$

Balchen and Mumme (1989, 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 limitations if a more complex controller is used. However, in

[3] To see that the controller gain must be reversed 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.

most practical cases the bound in (22) applies since one wants to use simple controllers, and since the plant model is not known sufficiently well to place zeros in the controller to counteract the poles at high frequency.

4 NEUTRALIZATION PROCESS

The above controllability results are 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.

Consider a process in which an acid stream (pH= -1) is neutralized by a base (pH=15) to get a final pH of 7 ± 1. Let the output be the excess of acid

$$y = c_H - c_{OH} \tag{23}$$

and

$$u = Flow_{base}, \quad d = Flow_{acid} \tag{24}$$

The appropriately scaled model is

$$y = \frac{k_d}{1+\tau s}(u-d); \quad k_d = 0.25 \cdot 10^7 \tag{25}$$

where $\tau = V/q = 1000$s. The output is extremely sensitive to both u and d, and the frequency up to which feedback is needed is

$$\omega_d \approx k_d/\tau = 2500 \text{ rad/s} \tag{26}$$

This requires a response time of 0.4 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. Note that the fundamental reason is that the process is extremely sensitive to disturbances. For feedback control this is the real control problem, and not that the required precision for the input u is so large (which is the argument usually given in the literature, but which only applies to feedforward control).

The only way to improve the controllability is by design changes. The most obvious change in this case is to do the neutralization in several steps. For n equal tanks in series we have

$$g_d(s) = \frac{k_d}{(1+\tau s)^n} \tag{27}$$

and we find $\omega_d \approx k_d^{1/n}/\tau$ (using the asymptote). Assume that the delay is 10 s and that the total flow is 10 l/s. Then we find that the following designs have the same controllability (with $\omega_d = 1/\theta = 0.1$ rad/s):

3 tanks of about 13500 l each
4 tanks of about 4000 l each
5 tanks of about 1900 l each
6 tanks of about 1160 l each

The minimum total volume is obtained with 16 tanks of about 251 l each - giving 40200 l total volume. However, taking into the account the additional cost for extra equipment, piping, control equipment (each tank must have a pH› controller), etc. , we would probably select a design with 5 neutralization tanks for this example.

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

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

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

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

[7] 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.

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

APPENDIX. Scaling procedure

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

$$y' = g'(s)u' + g_d'(s)d' \tag{28}$$

$$e' = y' - r' \tag{29}$$

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

The normalized or scaled variables (in the interval -1 to 1) are obtained by normalizing each variable by its maximum allowed value.

$$d = \frac{d'}{d^{max}}, \; r = \frac{r'}{r^{max}}, \; u = \frac{u'}{u^{max}}, \; e = \frac{e'}{e^{max}}, \; y = \frac{y'}{e^{max}} \tag{30}$$

Here

- u^{max} - largest allowed change in u (typically because of saturation constraints)
- d^{max} - largest expected disturbance
- r^{max} - largest expected change in setpoint
- e^{max} - largest allowed control error for output

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.

Introducing (30) into (28) yields

$$e^{max} y = g'(s)u^{max} u + g_d'(s)d^{max} d$$

Define the *scaled* transfer function models as

$$g(s) = g'(s)\frac{u^{max}}{e^{max}}; \quad g_d(s) = g_d'(s)\frac{d^{max}}{e^{max}} \tag{31}$$

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

$$y = g(s)u + g_d(s)d \tag{32}$$

$$e = y - \frac{r^{max}}{e^{max}} r \tag{33}$$

In this paper we select $r^{max} = e^{max}$ such that $e = y - r$.

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. 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, $d(t) = 1 \cdot sin(\omega t)$).