CONTROLLABILITY OF PROCESSES WITH LARGE GAINS

Antonio Araújo [*,1] **Sigurd Skogestad** [*]

[*] *Department of Chemical Engineering*
Norwegian University of Science and Technology
N-7491 Trondheim, Norway

Abstract: There is some disagreement in the literature on whether large plant gains are a problem or not when it comes to input-output controllability. In this paper, the effect of two kinds of input errors is studied and controllability requirements are derived. First, input disturbances are studied. These may pose a problem if the plant gain is large at high frequencies. Second, we study the nonlinear effect of limited input resolution which causes limit cycle behavior similar to that found with relay feedback. The magnitude of these limit cycles depends on the high-frequency process gain, but is independent of the controller tuning. They can be reduced by pulse modulating the input signal, but this may cause excessive input movement. In summary, large gains at frequencies corresponding to the closed-loop bandwidth may cause control problems, but large steady-state gains are not by themselves a problem.

Keywords: High gain, input disturbance, valve resolution, quantizer, limit cycle, controllability, PI-controller.

1. INTRODUCTION

The main goals of feedback control systems are to stabilize the process and reduce the effect of unmeasured disturbances on the output to an acceptable level. A fundamental question arises: Is the process input-output controllable? There are many factors that need to be considered, and one of them is the magnitude of the process gain. The gain depends on the frequency and, for multivariable plants, also on the input direction. To quantify this, the singular values $\sigma_i(G(j\omega))$ of the process transfer function $G(s)$ are considered. Of particular interest are the maximum and minimum singular values, denoted $\bar{\sigma}(G)$ and $\underline{\sigma}(G)$, respectively. In this paper, for simplicity, SISO systems where $\bar{\sigma}(G(j\omega)) = \underline{\sigma}(G(j\omega)) = |G(j\omega)|$ are mainly considered.

It is well accepted that small process gains may cause problems, for example, with input saturation. For example, Morari (1983) states that, with unitary scaling of the inputs and desired output changes of magnitude one in terms of the 2-norm, the requirement for avoiding input saturation is $\underline{\sigma}(G) \geq 1$, that is, a minimum gain of one is required to have acceptable control.

It is less clear whether large process gains pose a problem. Skogestad and Postlethwaite (1996) consider the condition number, defined as $\gamma(G) = \bar{\sigma}(G)/\underline{\sigma}(G)$ and make the following conclusion: *A large condition number may be caused by a small value of $\underline{\sigma}(G)$, which is generally undesirable. On the other hand, a large value of $\bar{\sigma}(G)$ is not necessarily a problem.*

On the other hand, intuitively, a large process gain may be troublesome, because the output becomes very sensitive to the input changes. McAvoy and Braatz (2003) argue along these lines and claim that for control purposes the magnitude of steady-state process

[1] Corresponding author: antonio.araujo@chemeng.ntnu.no.

gain ($\bar{\sigma}(G)$) should not exceed about 50. If this is correct then it would have important implications on the design of many processes.

The objective of this work is to study this in more detail. It is clear that the rule of McAvoy and Braatz (2003) is reasonable if feedforward control is considered because there will always be some error when implementing the input and without feedback this cannot be corrected for. However, in terms of feedback control, the rule cannot be generally true because for some classes of processes, it is well known that large process gains are not a problem. Consider, for example, feedback control of liquid level (output) using effluent flow (input). The steady-state gain is infinite due to an integrating transfer function, but it is easily controllable.

Nevertheless, high process gains may cause problems for feedback control at least at high frequencies and the aim of this paper is to study this in terms of input errors. Two main types of input errors are discussed: input (load) disturbance and input inaccuracy caused by limited input (valve) resolution. Most of the results are derived for first-order plus delay processes, otherwise, when appropriated, more general derivations are presented.

2. INPUT LOAD DISTURBANCE

It is well known that "large disturbances" cause control problems. Without control the effect of disturbances on the output is $y(s) = G_d(s)d(s)$, and by "large disturbances" is meant that the product $|G_d|d$ is large. Here, input disturbances are considered, i.e., $G_d = G$. First, a large plant gain $|G|$ may cause problems for feedforward control. This follows because it is necessary to be very precise with the input change (e.g. see eq. (5.70) in Skogestad and Postlethwaite (1996)). Thus, large disturbances motivates the need for feedback control, which is considered in this paper.

With feedback control, "large disturbances" are not necessarily a problem, but they pose limitations on the minimum bandwidth. Consider a single disturbance d and assume that the reference is constant, i.e. $r = 0$. Without control the steady-state sinusoidal response from d to the control error is $e(\omega) = G_d(j\omega)d(\omega)$, where phasor notation is used. Assume that the worst-case disturbance at any frequency is $d(t) = d_0 \sin \omega t$, i.e. $|d(\omega)| = d_0$ (where d_0 is assumed constant at all frequencies), and the control objective is that the controller error is less than $e_{\max}$ at any each frequency, i.e., $|e(\omega)| < e_{\max}$. From this, one can immediately draw the conclusion that *no control is needed if* $|G_d(j\omega)d_0| < e_{\max}$ *at all frequencies (in which case the plant is said to be "self-regulating")*. If $|G_d(j\omega)|d_0 > e_{\max}$ at some frequency, then control is needed (feedforward or feedback). In the following, feedback control is considered, in which case $e(s) = S(s)G_d(s)d(s)$, where $S = (I + GK)^{-1}$ is the sensitivity function. With $|d(\omega)| = d_0$, the requirement $|e(\omega)| < e_{\max}$ then becomes

$$|S(j\omega)| \cdot |G_d(j\omega)|d_0 < e_{\max} \quad \forall\omega \qquad (1)$$

A plant with a small $|G_d|$ is preferable since the need for feedback control is then less, or alternatively, given a feedback controller (and thus given S), the effect of disturbances on the output is small.

$|S|$ is small at low frequencies, so in general it does not matter if $|G_d|$ is large at steady state. However, $|S|$ increases with frequency and crosses 1 at the bandwidth frequency ω_S. At this frequency

$$|G_d(j\omega_S)| < y_{\max}/d_0 \qquad (2)$$

Thus, (2) provides an upper bound on the allowed disturbance gain at the frequency ω_S. In most cases $|G_d|$ becomes smaller at high frequency, so the bound is easier to satisfy if ω_S is increased. However, for stability reasons the value of ω_S is limited, and typically $\omega_S \approx 0.5/\theta$, where θ denotes the "effective delay" around the feedback loop (just consider G as a first-order plus delay model with a PI controller tuned according to Skogestad (2003) and $|S(j\omega_S)| = 1$). The bound (2) then becomes

$$|G_d(j0.5/\theta)| < y_{\max}/d_0 \qquad (3)$$

This bound is independent of the controller, and thus provides a fundamental controllability requirement.

However, the purpose of this paper is not to consider plants for which $|G_d|$ is large, but rather plants for which $|G|$ is large. In practice, these are related because all plants have disturbances at the input to the plant. To this effect, consider input (load) disturbances with $G_d(s) = G(s)\alpha_d$ where α_d is a constant gain. (3) then gives the following limit on the allowed plant gain at frequency ω_S

$$|G_d(j0.5/\theta)| < 1/\alpha_d \cdot y_{\max}/d_0 \qquad (4)$$

Input disturbances are very common and have many sources. For example, in many cases the input is a valve which receives its power from a hydraulic system (e.g. the brakes of a car) or from pressured air (many process control applications). A change (disturbance) in the power system will then cause an input disturbance. The value of α_d will vary depending on the application. If it is assumed that the system has been scaled such that the largest expected input u is of magnitude 1, then it seems reasonable that α_d is at least 0.01, and that a typical value is 0.1 or larger.

As an example consider the following plant $G(s) = ke^{-\theta s}/(\tau s + 1); G_d(s) = \alpha_d G(s)$, where $k = |G(0)|$ is the steady-state gain of the plant. The high-frequency asymptote is $|G(j\omega)| \approx k/\tau\omega = k'/\omega$, where $k' = k/\tau$ is the initial slope of the step response. (4) gives the controllability requirement

$$k/\tau = k' < 1/\alpha_d \cdot 0.5/\theta \cdot y_{\max}/d_0 \qquad (5)$$

Thus, there exist an upper bound on the allowed value of k'.

Remark 1. (5) seems to indicate that a plant with a large steady-state gain k is fundamentally difficult to control. However, this is usually not true, because a large value of k is usually accompanied by a large time constant τ. For example, for an integrating process $G(s) = k'e^{-\theta s}/s$. Thus, there is an infinite steady-state gain k and also an infinite time constant τ.

3. LIMITED INPUT RESOLUTION

The implications of limited input resolution is studied here. The main reason for this is that McAvoy and Braatz (2003), based on a case study, claim that this imposes limitations on the allowed steady-state process gain.

3.1 Controllability requirement assuming sinusoids

Consider a simple SISO example where the plant is given by

$$G(s) = 100/[(10s + 1)(s + 1)^2] \tag{6}$$

and the controller is

$$K(s) = K_c(\tau_I s + 1)/\tau_I s, \tag{7}$$

which contains a dominant time constant $\tau_I = 10$, that cancels the pole in $G(s)$, and $K_c = 0.04$ is selected.

The block diagram of the feedback system is depicted in Figure 1.

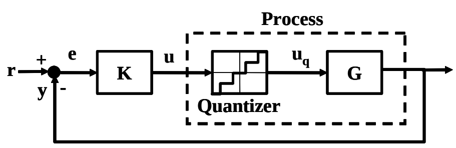

Fig. 1. Feedback control configuration for the valve inaccuracy problem.

In this Figure, r is the set point, y is the plant output, u is the controller (K) output, and G is the plant. The element called quantizer has been used to simulate valve inaccuracy.

The effect is thus to quantize a smooth signal u into a stair-step output u_q.

$$u_q = q \cdot \text{round}\,(u/q)\,, \tag{8}$$

Here q is the quantization step and the $round$ function takes its argument to the nearest integer. The limited valve resolution results in a stepwise input "disturbance" of magnitude equal to the quantization step, q.

For the example given by (6) and (7), $q = 0.03$ is taking as the quantizer step. Figure 2 shows the closed-loop response for a step change in the reference of magnitude 1 ($r_0 = 1$). From this figure, the magnitude and the period of oscillations in y are measured to be $a = 0.189$ and $T = 6.72s$, respectively.

Limit cycles are inevitable if there is a quantizer and integral action in the controller. This follows because on average the input must equal the steady-state value $u_{ss} = y_{ss}/G(0) = r/G(0)$, and if this does not happen to exactly correspond to one of the quantizer level, the quantized input u_q will cycle between the two neighboring quantizer levels, q_1 and q_2. Let f and $1 - f$ denote the fraction of time spend at each level. Then, at steady state $u_{ss} = fq_1 + (1 - f)q_2$ and from this f can be found. Note that the closer u_{ss} is to one of these values, the longer the time u_q must remain on it. In the example, $u_{ss} = y_{ss}/G(0) = 1/100 = 0.01$, which is closer to $q_1 = 0$ than $q_2 = 0.03$. The fraction of time u_q remains on $q_1 = 0$ is $f = 1 - 0.01/0.3 = 0.67$. As expected, this agrees with the simulations.

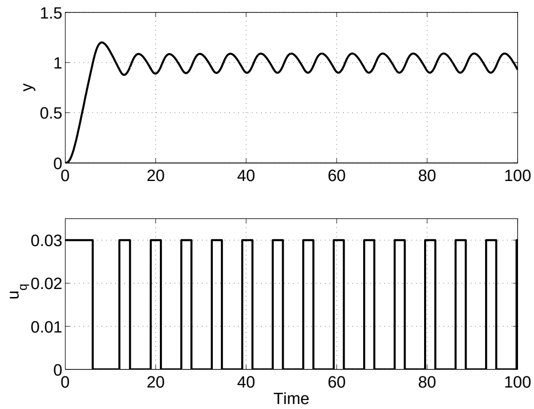

Fig. 2. Simulation results for the system given by (6) and (7) for $r_0 = 1$.

Moreover, when the limit cycle is established the quantizer can be regarded as a relay without hysteresis and thus can be treated as such. The amplitude of the oscillations can then be found analytically by considering the harmonic linearization or describing function of the nonlinearity in the loop showed in Figure 1.

For a relay without hysteresis, the describing function is given by (see Slotine and Li (1991)):

$$N(a) = 4q/\pi a, \tag{9}$$

where a is the amplitude of the oscillations and q is the relay amplitude (like the quantization step).

For the system depicted in Figure 1, the condition for oscillation is simply given by

$$N(a)L(j\omega) = -1, \tag{10}$$

where $L(j\omega) = G(j\omega)K(j\omega)$ is the open-loop transfer function.

Since according to (9) $N(a)$ is a real number, it follows from (10) that ω is the ultimate frequency $\omega_{L,180}$ and $K_u = N(a) = 4q/\pi a$ is the ultimate gain (Aström and Hägglund, 1988). As long as τ_I in

(7) is sufficiently large, that is, $\frac{1}{\tau_I}$ is much smaller than $\omega_{L,180}$, $\angle K = -\frac{\pi}{2} + \arctan(\omega_{L,180} \cdot \tau_I) \approx 0$. Then, $\angle L = \angle G + \angle K \approx \angle G$, $\omega_{L,180} \approx \omega_{G,180}$ ($\omega_{L,180}$ is independent of both K_c and τ_I), and $K_u = 1/|G(j\omega_{L,180})|$ which leads to

$$|G(j\omega_{L,180})| \approx \pi a/4q \tag{11}$$

Let a_{max} denote the maximum allowed amplitude of the oscillations is y. Then, from (11) the following controllability requirement applies

$$|G(j\omega_{L,180})| < \pi a_{max}/4q, \tag{12}$$

Typically, $a_{\max}$ will be considerably smaller than $y_{\max}$, e.g. $a_{\max} = 0.1y_{\max}$. (12) gives an upper limit on plant gain at frequency where $\angle L = -\pi$ (-180^o). Usually, $\omega_{L,180} \approx 1.5/\theta$ (just consider G as a first-order plus delay model with a PI controller tuned according to Skogestad (2003) and $\angle L = -\pi$).

For the system given by (6) and (7), $\angle L(j\omega_{L,180}) = -\arctan(10\omega_{L,180}) - 2\arctan(1\omega_{L,180}) = -\pi$ which gives $\omega_{L,180} = 1.09$ and the period of oscillation is found to be $T = \frac{2\pi}{\omega_{L,180}} = 5.8$. Moreover, $|G(j\omega_{L,180})| = 4.13$ and from (11), $a = \frac{4}{\pi}q|G(j\omega_{L,180})| = 0.158$. This agrees quite well with the simulation results ($T = 6.72, a = 0.189$).

It has been assumed here that the resulting oscillations are sinusoidal, but this is not quite true. Then, two questions arise: What happens if the response in y is non-sinusoidal? Does (12) still hold? The answer for the last question is *yes*, as discussed in section 3.3.

3.2 Non sinusoid responses

By taking the system described by (6) and (7) and using the configuration of Figure 1 with $q = 1$ (representing the worst case, an on/off valve), the simulation results for the output y are depicted in Figure 3.

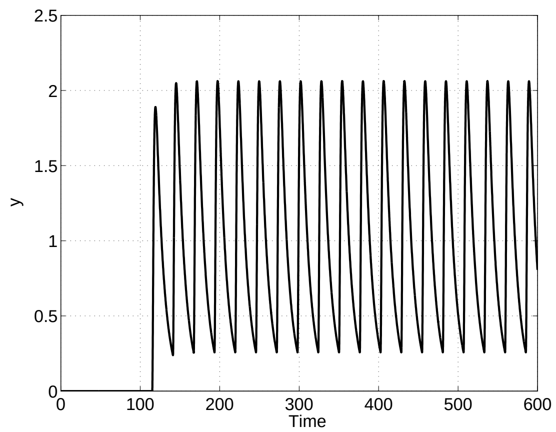

Fig. 3. Simulation results for the system given by (6) and (7) for $q = 1$.

From the figure, it is clear that the oscillations are not sinusoid-type. A deeper analysis by computing the power spectrum of the limit cycle confirms this hypothesis. In Figure 4, there is a second peak of about 50 at 4 rad/s which shows the data inconsistency, i. e. the limit cycles cannot be properly fitted to a sinusoid-type curve. From (11) and $T = 2\pi/\omega_{L,180}$

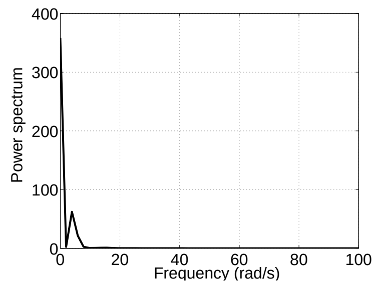

Fig. 4. Power spectrum of the limit cycle of the system described by (6) and (7) for $q = 1$.

the amplitude and period of the limit cycle are found to be $a = 6.23$ and $T = 6.28s$ which are very different from the measured results, $a = 1.82$ and $T = 26.48s$. Consequently, (11) cannot predict the amplitude for a non sinusoid response and (12) should not be used to assess the controllability of such systems.

We would like to perform an exact analysis. This is difficult, but we have derived exact results for a first-order plus delay process (see next section).

3.3 Controllability requirement for first-order plus time delay processes in the time domain

In this section, non sinusoid-type quantitized responses for a first-order with delay plant controlled by a PI controller is discussed. The following example is considered

$$G(s) = ke^{-\theta s}/(\tau s + 1) \tag{13}$$

$$K(s) = K_c(\tau_I s + 1)/\tau_I s, \tag{14}$$

with $k = 100$, $\theta = 1$, $\tau = 10$, $K_c = 0.04$, and $\tau_I = 10$.

The loop is set up according to Figure 1. The simulation results for $q = 0.03$ and a step change of 0.2 in the reference ($r_0 = 0.2$) are given in Figure 5.

The amplitude and period of the limit cycle of y can be predicted for first-order plus delay processes as it is shown later. For this particular example they are measured to be $a = 0.3$ and $T = 16.07s$, respectively. It can be seen that the output of the quantizer, u_q, oscillates between 0 and 0.03. The steady-state value is $u_{ss} = 0.2/100 = 0.002$, which means it stays $f = 0.93$ (93%) of the time (15s) at 0 and $f = 0.07$ (7%) of the time (1.07s) at 0.03.

Again, it can be argued that the response depicted in Figure 5 is non sinusoid and then again (12) cannot be

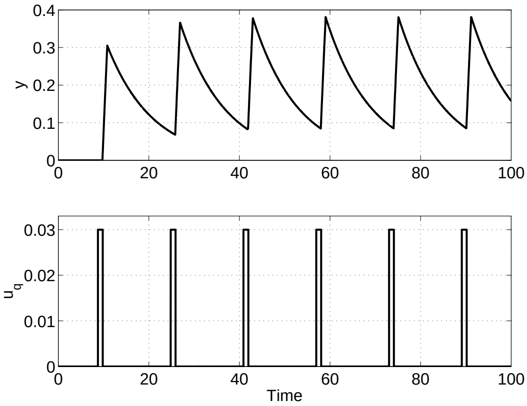

Fig. 5. Simulation results for the system given by (13) and (14).

applied to assess the controllability of the system. This suggests a different approach from the one derived in section 3.2 must be investigated. We have derived an exact analytical expressions for the amplitude and period of oscillation of the limit cycle for a first-order plus delay process. The main result is presented below without proof.

Theorem: *For the system given by (13) and (14) set up according to the configuration of Figure (1) with quantizer level* q*, the amplitude and period of the limit cycle oscillations are given by*

$$a = kq\frac{1 - e^{-t_1/\tau} + e^{-T/\tau} - e^{-(T-t_1)/\tau}}{1 - e^{-T/\tau}} \quad (15)$$

$$T = \theta(\frac{1}{1-f} + \frac{1}{f}), \quad (16)$$

where $t_1 = \theta/(1-f)$ *and* f *is calculated from* $u_{ss} = fq_1 + (1-f)q_2$. $\square$

For the example presented at the beginning of this section, the amplitude and period of oscillation calculated using (15) and (16) are $a = 0.3$ and $T = 16.07s$, respectively which match exactly the observed results. For this case, (11) gives $a = 0.24$ and $T = 4\theta = 4s$, that is, (12) can be considered a more conservative bound.

In general, the minimum value for T and the maximum amplitude a occurs when the set point change, r_0, is such that $f = 0.5$. In this case, $T = 4\theta$ and $a = kq[(1 - e^{-2\theta/\tau})^2/(1 - e^{-4\theta/\tau})]$ and the results also compare well with the describing function analysis based on sinusoids in (11).

Moreover, for the plant given by (13) and (14), assuming $\tau_I \approx \tau \Rightarrow \angle L = -\omega_{L,180}\theta - \pi/2 = -\pi \Rightarrow \omega_{L,180} = \pi/2\theta$, the corresponding period and amplitude of oscillation are $T = 2\pi/\omega_{L,180} = 4\theta$ and $a = kq\sqrt{\frac{16}{\pi^2}\frac{4(\theta/\tau)^2}{\pi^2+4(\theta/\tau)^2}}$, which for small values of θ/τ agrees quite well with the previous expression for a in (15); see Figure 6.

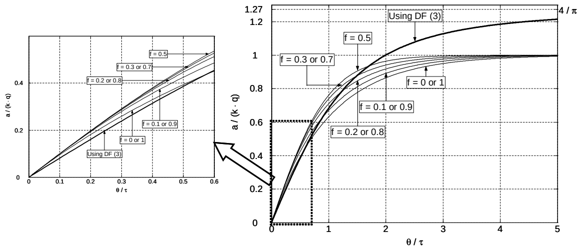

Fig. 6. The amplitude in (15) agrees surprisingly well with the describing function analysis (11) provided θ/τ is small.

Remark 2. Since (15) is derived taking into account the approximation $\tau = \tau_I$ *which applies for well-tuned controllers (see Skogestad (2003)), the amplitude and period of the limit cycle are independent of the controller parameters.*

Again, it is required that $a < a_{max}$ and the controllability requirement for first-order plus time delay processes is

$$|G(0)| < \frac{a_{max}}{q}\frac{1 - e^{-T/\tau}}{1 - e^{-t_1/\tau} + e^{-T/\tau}(1 - e^{t_1/\tau})} \quad (17)$$

3.4 How to avoid oscillations

The oscillations in the output of the system showed in Figure 1 can be avoided by the following ways:

a. Change the valve so that the resolution is enhanced (small quantization step);
b. Redesign the process in order to change the values of k, τ, and θ (smaller effective delay);
c. Take away the integral action leaving solely a P-controller (may give a poor performance);
d. Introduce fast, forced cycles at the input with a higher frequency than those generated “naturally”. For example, one may use high-frequency pulse modulation or add high-frequency sinusoids (some $d_u = \sin\omega t$ which may wear out the valve).

The use of a P-controller (item c) can eliminate oscillations as long as steady-state offset can be afforded. In order to make the offset as small as possible, bounds on the controller gain, K_c, are found to be (again, for the sake of compactness, the derivation of those bounds are not to be shown here):

$$\frac{n_{max}q}{r_0 - n_{max}qG(0)} \leq K_c < \frac{(n_{max}+0.5)q}{r_0 - n_{max}qG(0)}, \quad (18)$$

where $n_{max} = \lfloor r_0/qG(0) \rfloor$.

An attractive alternative, at least from a theoretical point of view, is to introduce high-frequency cycling at the input (item d). The problem is that the fast cycling may be difficult to handle in practice, for example, because the valve cannot be moved so fast or because of excessive wear. One approach is to introduce a pulse modulator in the controller before the quantizer. By applying this method, the response of the system given by (13) and (14) is depicted in Figure 7. As it can be seen, the amplitude is drastically reduced.

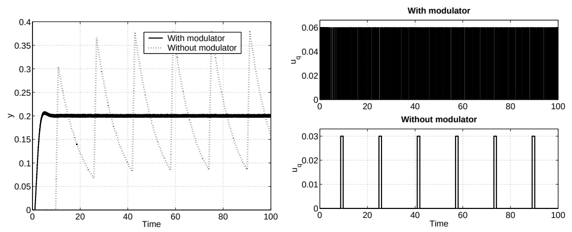

Fig. 7. Simulation results for the system given by (13) and (14) using a modulator ($r_0 = 0.2$).

4. DISCUSSION

McAvoy and Braatz (2003) claim that an upper limit for $\bar{\sigma}(G)$ should be imposed. It is suggested that a reasonable limit is 50 because essentially all control systems are eventually implemented with analogue devices which typically have an accuracy on the order of 0.5%. Actually this is true only at bandwidth frequency whereas no such limit exists at steady-state. Furthermore, it is also claimed that it is impossible in practice to get the fine manipulation of the control valves that is required for control because these valves would be limited to move in a very small region. Actually this is only true for feed forward systems without pulsing. There will be no problem with feedback, but some cycling must be accepted.

McAvoy and Braatz (2003) also claim that the cycling can be avoided by detuning the controller, but this is not generally true, unless one is willing to remove the integral action and accept an offset. The simulation used by McAvoy and Braatz (2003) to illustrate this claim is misleading, because oscillations do start if the simulation time is increased.

An important distinction between input load disturbance and valve inaccuracy is that, in general, in the latter a high bandwidth has no effect on the controllability of the system since controller parameters do not generally affect the limit cycle as showed for first-order plus delay process. Moreover, the requirement in (4) is more restrictive than the requirement in (12) if $|G_d|d_0 > q/2.42$ (to see this, consider the ratio $|G(j\omega_S)|/|G(j\omega_{L,180})|$ and (4) and (12)).

Two basic approaches to assess controllability are discussed in this paper. But, in general, to make use of one or the other, the resulting limit cycle has to be characterized. If the process is a first-order plus time delay the controllability requirement is directly given by (17). Otherwise, simulations must be performed in order to determine if the limit cycle is sinusoid-type, for example, by performing a spectral power analysis. If the limit cycle is proved to be sinusoid-type, (12) is used as the controllability requirement.

5. CONCLUSION

Processes with large gains are a major problem when input load disturbance and valve inaccuracy problems arise. For input load disturbance, high gain implies the need of a high bandwidth which cannot always be achieved in practice.

When dealing with valve inaccuracy problems a different approach has to be used since, in general, the controller parameters do not affect the bandwidth. Besides, high gains give a large amplitude of the resulting limit cycles. For sinusoid-type limit cycles, the simple approach using harmonic linearization approximation are derived to assess controllability. As for first-order plus time delay processes, on the one hand, more complicated expressions are needed to assess controllability but on the other hand, the results are exact. A general approach to deal with valve inaccuracy is proposed.

In order to avoid oscillations due to valve inaccuracy one may use a P-controller and performance may then degrade due to offset. Alternatively, the pulse modulation approach yields much better results since the remaining oscillations are of very low amplitude, but the problem is that the valve may wear out severely.

REFERENCES

Aström, K. J. and T. Hägglund (1988). *Automatic Tuning of PID Controllers*. Instrument Society of America. USA.

McAvoy, T. J. and R. D. Braatz (2003). Controllability of process with large singular values. *Ind. Eng. Chem. Res.* **42**, 6155–6165.

Morari, M. (1983). Design of resilient processing plants III - a general framework for the assessment of dynamic resilience. *Chemical Engineering Science* **38**, 1881–1891.

Skogestad, S. (2003). Simple analytic rules for model reduction and PID controller tuning. *Journal of Process Control* **13**, 291–309.

Skogestad, S. and I. Postlethwaite (1996). *Multivariable Feedback Control: Analysis and design*. John Wiley & Sons. Chichester, UK.

Slotine, J. E. and W. Li (1991). *Applied Nonlinear Control*. Prentice-Hall International Editions. New Jersey, USA.