Reglermöte '94
Västerås, 25-26 Oct '94
(Published in proceedings)

INPUT-OUTPUT CONTROLLABILITY ANALYSIS

SIGURD SKOGESTAD *

Chemical Engineering Dept., University of Trondheim, NTH
N-7034 Trondheim, Norway

Abstract

The objective of the talk is to derive some fundamental results for controllability analysis (achievable control performance). The effects of disturbances, constraints, delays and RHP-zeros are quantified. The results are applied to a neutralization process where it is shown that the process must be modified to get acceptable controllability.

1 Introduction

In 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?

III. How should the process be changed to improve control ? For example, one may want to find the required size of 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 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.

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".* 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.

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

$$y = gu + g_d d \quad (1)$$

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 closed-loop response is

$$y = Tr + Sg_d d \quad (2)$$

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

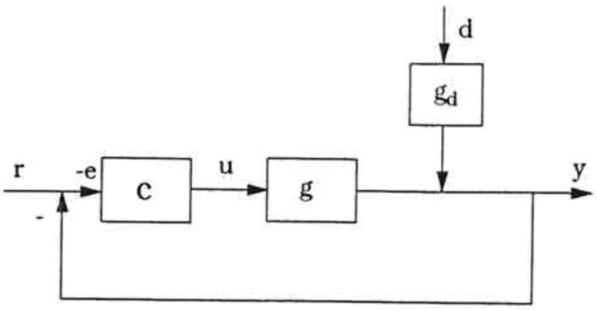

Figure 1: Block diagram of feedback control system.

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 $L = gc$.

Bandwidth. In this paper bandwidth is defined as *the frequency* ω_B *where the loop gain is one in magnitude*, i.e. $|L(j\omega_B)| = 1$.

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.

2 Summary of rules

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

Rule 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.

Proof:. Without control $y = g_d d$. Scaling has been applied such that the largest disturbance at a given frequency has $|d(j\omega)| = 1$. Thus, at frequencies $\omega < \omega_d$ the output y will be unacceptable ($|y| > 1$) for a disturbance $d = 1$, so control is needed at these frequencies, and we must require $\omega_B \geq \omega_d$.

More specifically, we must with feedback control require $|L| = |gc(j\omega)| > |g_d(j\omega)|$ for $\omega < \omega_d$. *Proof:* With feedback control $y = Sg_d d \approx (g_d/L)d$, and to have $y < 1$ for $d = 1$ we must require $|L| > |g_d|$.

Rule 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. *"Proof"*: This is really the definition of what is meant by the minimum response time and requires no proof.

Rule 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 a disturbance $d(j\omega) = 1$.

Proof. From $y = gu + g_d d = 0$ we get $u = -(g_d/g)d$ and with $d = 1$ we need $|u| = |g_d|/|g| < 1$ to avoid input constraints.

Strictly speaking, perfect control is not required, and the input needed for "acceptable" control ($|y| < 1$) is $|u| = (|g_d| - 1)/|g|$. The difference is small at low frequencies where $|g_d|$ is larger than 1. (However, for multivariable systems the difference may be large for ill-conditioned plants).

Rule 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.

Proof. From $y = gu$ and $y = r$ (perfect control) we get $u = r/g$, and with $r = R_{max}$ we need $|u| = R_{max}/|g| < 1$ to avoid input constraints.

Rule 5 *Time delay* θ. *Must require* $\omega_B < 1/\theta$ *to have acceptable control performance.*

Proof. It is impossible to remove the effect of the delay and $L(s)$ must contain a term $e^{-\theta s}$. The ideal controller which minimizes $J = \int_0^\infty |e(t) - r(t)|^2 dt$ when $r(t)$ is a step and there is no penalty on the inputs has complementary sensitivity $T = e^{-\theta s}$. The corresponding loop gain $L = T/(1 - T)$ crosses 1 in magnitude at about frequency $1/\theta$. In practice, the ideal controller cannot be realized so this value provides an upper bound on the bandwidth.

Rule 6 *Real RHP-zero at* $s = z$. *Must require* $\omega_B < z/2$ *to have acceptable control performance at low frequencies.*

Proof. Again, it is impossible to remove the effect of a RHP-zero and we must have $L(s = z) = 0$. This may be used a starting point to derive the above bound. An alternative proof is as follows: The ideal controller which minimizes $J = \int_0^\infty |e(t) - r(t)|^2 dt$ when $r(t)$ is a step and there is no penalty on the inputs has complementary sensitivity $T = \frac{-s+z}{s+z}$. The corresponding loop gain $L = T/(1-T)$ crosses 1 in magnitude at about frequency $z/2$. In practice, the ideal controller cannot be realized so this value provides an upper bound on the bandwidth.

Remark. Strictly speaking, a RHP-zero only makes it impossible to have tight control in the frequency range close to the location of the RHP-zero. If we do not need tight control at low frequencies, then we may reverse the sign of the controller gain, and instead achieve tight control at frequncies higher than z. One special example is for plants with a zero at the origin ($g(s)$ contains a term s) where one can achieve good transient control, but where there is no effect of the control at steady-state.

Rule 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, since for minimum phase plants (no delays or RHP-zeros), any phase lag may be in theory counteracted by the controller. However, in practice this is not possible, in particular if the phase drops quickly around the frequency ω_{g180}.

Proof for PID-controller. With a PID-controller the maximum phase lead is 54.9^o for a controller with derivative action over one decade. Thus, if we require at phase margin larger than 54.9^o we must require $|L| \leq 1$ at frequency w_{180} and the rule follows.

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

Proof. One possible starting point to prove this is to use the fact that L must contain the term $1/(s-p)$ so we have $L(s=p) = \infty$.

The above rules are necessary conditions ("minimum requirements") in order to achieve acceptable control performance. The reason they are not sufficient is that they are based on considering only "one effect at the time".

In summary, Rules 1, 2 and 8 tell us that we need high feedback gain ("fast control") in order to reject disturbances, to track setpoints and to stabilize the plant. On the orher hand, Rules 5, 6 and 7 tell us that we must use low feedback gains in the frequency range where there is RHP-zeros or delays or where the plant has a lot of phase lag. We have formulated these requirements for high and low gain as bandwidth requirements. If they somehow are in conflict then the plant is not controllable and the only remidy is to introduce a design modification which 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 (feedforward, cascade control)
5. Add extra actuators (parallell control)
6. Change the control objectives
7. Change the control structure of the lower levels

3 Applications

Several applications will be discussed in the talk. For further details the reader is referred to Skogestad (1994a, 1994b, 1994c).

References

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

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

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

[4] S. Skogestad, 1994a, "Controllability analysis of SISO systems", *Preprints IFAC Symposium ADCHEM'94*, Kyoto, Japan, May 1994, 471-476.

[5] Skogestad, S., 1994b, "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.

[6] S. Skogestad, 1994c, "Frequency-domain methods for analysis and design. II. Controllability analysis of SISO Systems", *Proc. NATO-ASI on Methods of Model Based Process Control*, Antalya, Turkey, Aug. 1994.

[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.