Dynamics and Control of Quasirational Systems

Systems having transfer functions of the form

\[ G_p(s) = \frac{P_1(s) - P_2(s)e^{-us}}{Q(s)}, \]

where \( P_1(s), P_2(s) \) and \( Q(s) \) are polynomials, are called quasirational distributed systems (QRDS). They are encountered in processes modeled by hyperbolic partial differential equations. QRDS can have an infinity of right half-plane zeros which causes large phase lags and can result in poor performance of the closed-loop system with PID controllers. Theory on the asymptotic location of zeros of quasipolynomials is used to predict the nonminimum phase characteristics of QRDS and formulas are presented for factoring QRDS models into minimum and nonminimum phase elements.

A generalized Smith predictor controller design procedure for QRDS, based on this factorization, is derived. It uses pole placement to obtain a controller parameterization that introduces free poles which are selected to satisfy robustness specifications. The use of pole placement allows for the design of robust control systems in a transparent manner. Controller selection is generally better, simpler and more direct with this procedure than searching for optimal PID controller settings.

Introduction

This paper deals with the dynamics and control of a class of distributed processes that can be modeled by hyperbolic partial differential equations. Important industrial examples include heat exchangers, tubular and packed bed reactors, and particulate systems with nucleation and growth. These systems, when linearized, yield transfer functions having the following form

\[ G_p(s) = \frac{P_1(s) - P_2(s)e^{-us}}{Q(s)} \]

where \( P_1(s), P_2(s) \) and \( Q(s) \) are polynomials in \( s \), with the order of \( P_1(s) \) and \( P_2(s) \) being less than or equal to the order of \( Q(s) \).

The numerator of Eq. 1 is a function called an exponential polynomial or quasipolynomial (Bellman and Cooke, 1963). Systems described by Eq. 1 are here given the name quasirational distributed systems (QRDS). The QRDS transfer function represents a large class of processes with a wide range of dynamic behavior. Models for the dynamics of particle size distribution in fluidized bed calciners and crystallizers are presented here as examples. Further, quasipolynomials are also encountered in hereditary systems (Bellman and Cooke, 1963) and in multivariable systems where one or more elements of the transfer function matrix have time delays (Ray, 1980; Jerome and Ray, 1986).

QRDS do not exhibit the simple delayed response characteristic of lumped parameter systems with time delays unless \( P_1(s) = 0 \). Some QRDS exhibit nonminimum phase behavior and including them in a closed loop can yield oscillatory responses with poor performance and poor stability characteristics with PID controllers. Other QRDS exhibit minimum phase behavior and can be controlled satisfactorily using PI or PID controllers. The presence of large phase lags, similar to systems with dead time, motivates the use of predictive controllers similar to the Smith predictor (SP) (Smith, 1957) for nonminimum phase QRDS.

Quasipolynomials are known to have an infinity of zeros.
which could lie either in the left half-plane (LHP) or in the right half-plane (RHP) (Bellman and Cooke, 1963). This information, however, has received little attention in the chemical engineering literature even though transfer functions for many distributed processes contain quasipolynomial functions. A predictive controller design procedure, called here the Generalized Smith Predictor (GSP) procedure, which uses information on the zeros, was first presented by the authors (1987). A similar procedure was recently reported by Jerome and Ray (1987) for quasirational functions arising from multivariable delay systems. The present paper describes our general design procedure.

The SP control scheme with the structure shown in Figure 1a is the oldest predictive controller (Smith, 1957). The SP scheme requires, for pure delay systems, a compensator with a transfer function that is the difference of the model transfer functions without and with the delay. This requires that the process model be factored into delay (nonminimum phase) and delay free (minimum phase) parts. The compensator removes the delay from the feedback signal, yielding a signal that contains a prediction of the response of the nondelayed plant. The main controller \( G_c(s) \) was usually chosen to be of the PI or PID form.

A perfect model, the SP scheme is equivalent to controlling the nondelayed plant with the delay being left in open loop. This is illustrated by the equivalent block diagram in Figure 1b.

Past applications of the SP approach have suffered from poor robustness properties (see Laughlin and Morari, 1987). This was due to the following factors:  

i) Lack of a framework for robustness analysis in terms of model uncertainty before the 1980s.

ii) In the past it was attempted to design \( G_c(s) \) independent of (ignoring) the error level in the dead time.

In addition, proposed extensions of the SP concept to general nonminimum phase systems were not based on an appropriate factorization of the transfer function (see Krishnan and Friedly, 1980).

The correct interpretation of the Smith predictor idea is that the parameterization of \( G_c(s) \) should be based on the delay free or equivalent minimum phase part of the model, but the final design should be based on robustness and performance considerations for the overall system. These ideas are discussed in detail in subsequent sections.

In this paper, a new Generalized Smith Predictor (GSP) design procedure is presented. It provides a correct extension of the SP concept to general nonminimum phase systems, with QRDS being used as examples. Further, the use of pole-placement to synthesize the main controller allows the GSP procedure to deal with robustness in a transparent manner. With these extensions, the GSP procedure can be used to design robust control systems and yields results equivalent to internal model control (IMC) (Garcia and Morari, 1982).

Early attempts to extend the SP compensator to QRDS were unsuccessful because the compensator was constructed to be the difference of the model transfer functions without and with the delay. Hence, the "delay free" or "minimum phase factor" was obtained by neglecting the delay from the model transfer function. Such an approach does not correctly account for the nonminimum phase properties of quasipolynomials and can lead to the closed-loop system being unstable or, even when stable, to poor performance. Krishnan and Friedly (1980) demonstrated problems with this approach for an evaporator subject to flow oscillations. They suggested a particular form for the compensator that worked for their system.

The GSP control scheme, with the structure shown in Figure 2a, extends the SP scheme to general nonminimum phase systems. It requires a factorization of the process model into minimum and nonminimum phase factors (Smith, 1958). The GSP compensator transfer function should be the difference of the model transfer function without and with the nonminimum phase factor, where the latter is in the form of an all-pass. The compensator removes the excess phase lag due to the nonminimum phase factor from the feedback signal, yielding a signal that contains a prediction of the response of the corresponding minimum phase system.

With a perfect model, the GSP scheme is equivalent to controlling the corresponding minimum phase system, which has the same magnitude response as the plant. This is illustrated by the block diagram in Figure 2b. The nonminimum phase component, which is responsible for the excess phase lag, is left in open-loop.

The GSP scheme is applicable to any nonminimum phase system provided that the minimum and nonminimum phase factors can be determined. Formulas are presented in this paper for performing the factorization for QRDS. For distributed systems that cannot be factored as easily, the GSP scheme may still be used by substituting an approximation for the factors. Vit (1979) describes a procedure for approximating the behavior of the minimum and nonminimum phase factors in systems modeled by a parabolic partial differential equation using lag and delay terms.

In the next section, QRDS models for particle size distribution in terms of cumulative mass fraction are presented for a fluidized bed calciner and a mixed suspension, mixed product removal (MSMPR) crystallizer. The dynamics that QRDS can exhibit are demonstrated through these examples. This is followed by a summary of results from the literature, on the location of zeros of quasipolynomials, relevant to the developments in this paper.

A controller design procedure for QRDS is then presented. It uses the GSP structure and pole placement to obtain a controller parameterization that includes adjustable parameters that are selected to satisfy robustness specifications. Robustness condi-

Figure 1a. Structure of the Smith predictor control scheme.

\[
\begin{align*}
G_p(s) & \rightarrow y' \\
(1 - e^{-1s}) G_{po}(s) & \rightarrow y \\
\end{align*}
\]

where \( G_p(s) = G_m(s)e^{-aw} \)

Figure 1b. Block diagram equivalent to the Smith predictor control scheme.

\[
\begin{align*}
\gamma_p & \rightarrow G_c(s) \\
G_{po}(s) & \rightarrow y' \\
(1 - e^{-1s}) & \rightarrow y \\
\end{align*}
\]
Figure 2a. Structure of the generalized Smith predictor control scheme.

\[
G_p(s) = G_s(s) + G_e(s) G_p(s)\]

where \(G_s(s) = G_r(s) G_e(s)\)

\[G_e(s) = \text{minimum phase}\]

\[|G_e(\omega)| = 1\]

The GSP design procedure is illustrated using examples dealing with the control of cumulative mass fraction in a fluidized bed calciner and a MSMPR crystallizer. The examples provide a comparison of the performance of GSP controllers with that of optimal PID controllers. A discussion on techniques for on-line implementation of the GSP controller is also included.

Dynamics of QRDS

Systems modeled by hyperbolic partial differential equations yield complicated transcendental transfer functions that can frequently be simplified or approximated to the form of Eq. 1 (Friedly, 1972). Hyperbolic partial differential equations are used commonly to describe numerous engineering processes where convective transport dominates dispersive phenomena. Some examples include heat exchangers, tubular and packed bed reactors, fixed- and countercurrent moving-bed adsorption systems and particulate systems with nucleation and growth. The objective of this section is to introduce examples of QRDS: fluidized bed calciners and crystallizers.

Fluidized bed calciners

An example of a fluidized bed calciner is the Dorr-Oliver Fluo-Solids lime-mud reburning calciner (Moran and Wall, 1965). It consists of a bed of heated particles, kept fluidized by air. The bed is kept at 1,500°F (816°C) by in-bed combustion of fuel. Solid feed material, as a very fine powder, is sprayed into the bed where it is calcined and sticks to the bed particles leading to their growth. Product particles are withdrawn from the bed at a controlled rate to maintain a constant bed mass. Seed particles, obtained by crushing part of the product, are intermittently added to the bed to maintain the cumulative mass fraction of +20 mesh (0.8 mm and larger) particles between 80% and 90%.

Control of particle size distribution can be achieved through control of a moment of the particle size distribution, a mean diameter, or the cumulative mass fraction above a cut-point size. However, the lack of on-line instrumentation for measuring particulate size distribution or low order moments suggests choosing the cumulative mass fraction, which can be measured fairly easily. Unlike the moments, the cumulative mass fraction does not require the counting of particles within different size fractions. Measurement of cumulative mass fraction requires only separation of a sample into two fractions with respect to a cut-point size and weighing them. This operation can be automated with existing solid handling equipment. In this paper, a control scheme based on measuring the cumulative mass fraction above a cut-point size, and manipulating the seeding rate is presented.

The dynamics of particle size distribution can be modeled using a population balance. The transfer function for the cumulative mass fraction above a cut-point size \(Z_c\) for seed rate manipulation is presented below; the modeling assumptions and a detailed derivation is available in Ramanathan (1988, p. 22).

\[
G_p(s) = \frac{[P_{Z_c}(s)/P_{Z_c}(0)]e^{-\tau Z_c} - 1}{s(s + 1)(s^2 + 3s + 3)}
\]

where

\[
P_{Z_c}(s) = (s + 1)^2Z_c^2 + 3(s + 1)^2Z_c^2 + 6(s + 1)Z_c + 6
\]

\[
P_{Z_c}(0) = Z_c^2 + 3Z_c^2 + 6Z_c + 6
\]

The cumulative mass fraction has a transfer function of the QRDS form given in Eq. 1. A \(Z_c\) of 2 was used for the example below (and in later sections) as it corresponds to a steady-state cumulative mass fraction of 85% above 20 mesh, a typical industrial choice (Moran and Wall, 1965). With this choice of cut-point size, the polynomials \(P_{Z_c}\) become

\[
P_{Z_c}(s) = 8s^3 + 36s^3 + 60s + 38
\]

\[
P_{Z_c}(0) = 38
\]

Figure 3 shows the response of the cumulative mass fraction to a step increase in seed rate. This system has negative gain with the final value tending to -0.1403. The response shown in Figure 3 has been normalized with respect to this value. It changes rapidly after an initial period of sluggish behavior and

Figure 2b. Block diagram equivalent to the generalized Smith predictor control scheme.

Figure 3. Open-loop step response for the cumulative mass fraction \(Z_c = 2\) in a fluidized bed calciner.
exhibits an overshoot with a discontinuous slope at dimensionless time equal to 2 before settling to the new steady state. The delayed term in Eq. 2 becomes active at $Z_2 - 2$, which corresponds to the time of crossing of the growth front of new seed particles into the size range of the cumulative mass fraction. The complicated dynamics of this system can be explained in terms of the interactions between particle size distribution, the growth rate and the distribution at the seed size (boundary condition) through an inherent feedback loop (Ramanathan, 1988).

A Bode plot for the cumulative mass fraction is shown in Figure 4. The amplitude has a high frequency asymptote with a slope of $-1.0$ and an intermediate region ($\omega = 3$ to 5) with a noninteger slope. The phase lag increases proportionally at high frequency, which is characteristic of systems with dead-time. However, it is evident from the open-loop time response and the transfer function in Eq. 2 that this system is more complicated than a pure first-order lag plus dead-time system.

**MSMPR crystallizer**

Developments in modeling, steady-state design and simulation techniques for crystallizers are fairly advanced (Randolph, 1984; Moyers and Randolph, 1973; Randolph and Larson, 1971). However, developments on the dynamics and control of particle size distribution have been limited.

Almost all the dynamic and stability studies on crystallizers have been in terms of the number of the number size distribution function. There have been very few studies on the dynamics of number or mass size distribution functions (Beckman and Randolph, 1977; Lei et al., 1971; Sherwin et al., 1969) and no published studies on dynamics of cumulative mass fraction.

A model for the cumulative mass fraction (above a cut-point size $Z_2$) in an isothermal, constant volume, MSMPR crystallizer is presented below for inlet concentration and residence time forcing; again the modeling details are available in Ramanathan (1988, p. 50).

$$M_z(s) = \left[ c_1 \right] + \left[ \frac{1}{(1 - C_2)s} \right] \frac{P_{Z2}(s)e^{-Ze_2}}{P_{Z2}(0)(s + 1)^4} + \frac{1}{(1 - C_2)s} \frac{P_{Z2}(s)e^{-Ze_2}}{P_{Z2}(0)(s + 1)^4} M_z(s)$$

$$= \left[ \frac{P_{Z2}(s)/P_{Z2}(0)e^{-Ze_2} - (s + 1)^4}{s(s + 1)^4} \right] T(s)$$

The second and third moments of the size distribution ($M_2(s)$ and $M_3(s)$, respectively) are given by

$$\frac{M_2(s)}{C_f(s)} = \frac{c_1 f_1(s)}{(1 - C_2)(s + c_2)(s + 1)^4 - c_4} + (c_1 - 1) f_i(s)$$

$$\frac{M_3(s)}{C_f(s)} = \frac{c_1 [(s + 1)^3 + (s + 1)(s + 2) + l]}{(1 - C_2)(s + c_2)(s + 1)^4 - c_4} + (c_1 - 1) f_i(s)$$

$$\frac{M_3(s)}{T(s)} = \frac{1}{(1 - C_2)(s + c_2)(c_4 + (s + 1)^2(s + 2) + (s + 1)) - (c_1 - 1) f_i(s)}$$

and the $P_{Z2}$'s are the same as those defined in Eq. 3.

As expected, the crystallizer cumulative mass fraction transfer functions are more complicated than those of the fluidized bed calciner. However, their QRDS nature is apparent. Similar models can be developed for other crystallizer configurations though the algebra becomes very lengthy.

The normalized open-loop response of cumulative mass fraction to a step increase in the residence time $T$ is shown in Figure 5 for three values of the nucleation to growth sensitivity parameter, $i = 1, 5$ and 15, with the secondary nucleation parameter, $j = 1$. The remaining parameters were taken from the crystallization literature (Zumstein and Rousseau, 1987). An abrupt change in the slope of the response can be seen in some cases, when the delayed term becomes active. With increasing $i$, the

![Figure 4. Bode plot for cumulative mass fraction ($Z_2 - 2$) in a fluidized bed calciner.](image)

![Figure 5. Open-loop step response for the cumulative mass fraction ($Z_2 - 2$) in a MSMPR crystallizer.](image)
response becomes more oscillatory and is unstable for \( i > 19 \). Similar results for the moments have been reported in the crystallization literature. For \( i > 19 \), the cumulative mass fraction transfer function has a pole in the RHP causing the system to be open-loop unstable.

A Bode plot for the cumulative mass fraction transfer function with respect to residence time is shown in Figure 6 for \( i = 1, 5 \) and 15, with \( j = 1 \). All the amplitude plots show resonance, with the magnitude of the resonance peaks being dependent on the parameter \( i \). At high frequency, the phase plot for \( i = 1 \) oscillates about a constant phase lag, while for \( i = 5 \) and 15 the phase lag tends to increase at a constant rate.

**Characteristics of QRDS**

The examples in the previous section illustrate the range of dynamic behavior that QRDS can exhibit. The following points can be noted:

i) Response to a step input is not delayed. However, in practice it is difficult to differentiate between the initial sluggish response of a high-order nondelayed system and the response of a delayed system.

ii) There may be abrupt change in slope when the delayed term becomes active. Sometimes the slope can switch signs. The abrupt change in slope is not always visible and, in practice, it would be difficult to detect, like the change from zero to nonzero output for systems with dead time.

iii) Some QRDS exhibit large phase lags at high frequency, indicating nonminimum phase behavior. Others exhibit minimum phase behavior.

iv) Sometimes they exhibit continuous resonance at high frequency.

v) QRDS may occur in series with time delays, in which case the time delay introduces a nonminimum phase term.

Unfortunately, the above list does not provide a unique characteristic that would make identification of a system as a QRDS easy. Theoretical models of the process are required for this purpose. Further, it is possible for a QRDS to switch from minimum to nonminimum phase behavior and vice versa over a certain range of parameter values making controller design very difficult.

The observed range in phase behavior of QRDS can be explained in terms of the location of the zeros of quasipolynomials. Quasipolynomials have been noted to have an infinity of zeros. A general theory for the asymptotic locations of zeros of quasipolynomials exists (Bellman and Cooke, 1967). Only the results relevant to our work are outlined below.

The zeros of quasipolynomial equations of the form

\[
N(s) = P_1(s) - P_2(s)e^{-\tau s} = 0
\]  

(11)

are of interest. Equation 11, which is the numerator of the QRDS transfer function, can be written in the expanded form

\[
N(s) = (s^a + a_1s^{a-1} + \cdots + a_m)
\]

\[- K(s^b + b_1s^{b-1} + \cdots + b_n)e^{-\tau s} = 0
\]  

(12)

The coefficients of the dominant powers of the two polynomials in Eq. 12 have been reduced to unity, and \( K \) represents the ratio of the coefficients of the dominant powers of \( P_2(s) \) and \( P_1(s) \). For \( |s| \) large, Eq. 12 can be approximated by the comparison function

\[
N_c(s) = s^a - Ks^b e^{-\tau s} = 0
\]  

(13)

which has an infinity of zeros lying along a curve of the form

\[
|s^{a-b}e^{-\tau s}| = |K|
\]  

(14)

The curve described by Eq. 14 has two symmetric branches with respect to the real axis. The roots of large modulus in this curve have the form

\[
s_j = (1/\tau_d)
\]

\[
\times [\ln|K| + (n-m) \ln \tau_d - (n-m) \ln |2k\pi + (n-m)\pi/2|]
\]

\[+ i|2k\pi + (n-m)\pi/2| + \epsilon(k)
\]  

(15)

where \( k = 0, \pm 1, \pm 2, \ldots \) and \( \epsilon(k) \rightarrow 0 \) as \( |k| \rightarrow \infty \). The upper sign is used for \( K > 0 \) while the lower sign corresponds to \( K < 0 \). For \( |s| \) large, all but a finite number of zeros of Eq. 11 will be asymptotic to the curve given in Eq. 14 and will be spaced at an asymptotic distance of \( 2\pi/\tau_d \) apart. The following cases can be distinguished:

In case \( n = m \), the formula in Eq. 15 simplifies to

\[
s_j = (1/\tau_d)[\ln|K| + i(2k\pi)] + \epsilon(k), \quad k = 0, \pm 1, \ldots
\]  

(16)

The curve is in the form of a vertical line and the zeros of Eq. 11,
excluding zeros at the origin, will be asymptotic to this vertical line. For \(|K| > 1\) the line will be in the RHP, while for \(|K| < 1\) it will be in the LHP.

In case \(n \neq m\), Eq. 11 will have an infinity of roots asymptotically given by Eq. 15 and the curve in Eq. 14. This curve is similar to an exponential curve for large \(|s|\). Further, if \(n > m\) the curve lies entirely in the LHP, while for \(n < m\) it lies entirely in the RHP.

The existing theory on the zeros of quasipolynomials only provides information on the asymptotic location of zeros. It does not yield information on the location of low modulus zeros, which have to be determined using a numerical root solving program.

Summarizing these results, if \(n < m\) or \(n = m, K > 1\), then the QRDS will have a chain of infinite RHP zeros. If \(n > m\) or \(n = m, K < 1\), then the QRDS will have a chain of infinite LHP zeros and at most a finite number of low modulus zeros in the RHP.

A program to calculate a finite number of roots of Eq. 11 was developed as a part of this investigation. It uses high-order Padé approximation for the delay term to calculate accurate values for the low modulus zeros and Eq. 15 to obtain initial guesses for the large modulus roots and Newton’s method for refinement of the approximation.

Figure 7 shows the location of zeros for the cumulative mass fraction transfer function of the fluidized bed calciner (Eq. 2) for a cut-point, \(Z_c = 2\). Figure 8 shows similar results for a MSMPR crystallizer for residence time changes (see Eq. 5), for three values of the nucleation to growth sensitivity parameter, \(i = 1, 5\) and 15, with the secondary nucleation parameter, \(j = 1\). The remaining parameters were taken from the crystallization literature (Zumstein and Rousseau, 1987). Only the upper half of the complex plane is shown. A symmetric branch of zeros located in the lower half plane has been omitted.

The cumulative mass fraction for the fluidized bed calciner has an infinity of RHP zeros and three low modulus LHP zeros plus a zero at the origin. The asymptotic theory predicts an infinity of RHP zeros for this system, as it corresponds to the case \(n < m\) (\(n = 0\) and \(m = 3\)). The zeros predicted by Eq. 15 are also shown in Figure 7 by the solid line. The asymptotic theory predicts the location of zeros very accurately except for the low modulus zeros.

The chain of zeros of the cumulative mass fraction transfer function for the MSMPR crystallizer, shown in Figure 8, goes from right to left (here \(n > m\)) with increasing modulus. For \(i = 1\), there are no RHP zeros; however, as \(i\) increases, the chain moves into the RHP. Hence for \(i = 5\) and \(i = 15\) there are a finite but large number of RHP zeros. Further, this example illustrates that certain design or operating variables can drastically shift the location of zeros and thus have a significant influence on controller design.

From the above results, it is evident that QRDS have a chain of infinite zeros lying either in the LHP or RHP. From a control perspective, three cases can be distinguished corresponding to:

i) No RHP zeros
ii) A finite number of RHP zeros (low modulus zeros)
iii) An infinity of RHP zeros

The location of the chain determines the frequency response behavior. Open-loop stable QRDS with no RHP zeros will be minimum phase and have phase plots that tend to a constant phase lag at high frequency. QRDS with an infinity or finite number of RHP zeros will be nonminimum phase. The closeness of the chain to the imaginary axis determines the resonance behavior.

### A Generalized Smith Predictor Design Procedure

Smith (1957) proposed the dead-time compensator or Smith predictor (SP) for improved control of SISO systems with a time delay. The SP, with the structure shown in Figure 1a, uses a model of the undelayed plant \(G_{mu}(s)\) and delay \(e^{-\theta d}\) to cancel the plant delay, leaving the undelayed signal \(y^*\) in the feedback path. Thus, controller parametrization can be based on the undelayed system.

There have been many modifications and extensions of the original SP. Reviews of these are in Jerome and Ray (1986), and Wong and Seborg (1986). Other model-based predictive controllers with closely similar structures, such as the analytical predictor of Moore et al. (1970), the inferential controller of Brossow (1979), internal model control (IMC) of Garcia and Morari (1982), and the generalized analytical predictor of Wong and Seborg (1986) and Wellons and Edgar (1985) have been introduced recently. Further, the work of Doyle and Stein (1981) and Morari and coworkers (Mandler et al., 1986; Zafiriou and Morari, 1986) have introduced a powerful framework...
for designing robust controllers in the face of model uncertainties.

The following extension of the SP, called the Generalized Smith Predictor (GSP) (see Figure 2), is proposed: given a factorization of the process transfer function

$$G_p(s) = G_{p+}(s)G_{p-}(s)$$

(17)

where

$$G_{p-}(s) = \text{minimum phase, causal and proper}$$

$$G_{p+}(s) = \text{nonminimum phase, causal, proper and satisfies } G_{p+}(0) = 1$$

one can obtain a feedback variable $y^*$ as if $G_{p-}(s)$ were the process transfer function is

$$y^*(s) = y(s) + G_{p-}(s)(1 - G_{p+}(s))u(s)$$

(18)

Once $y^*$ is computed on-line, the error $y_p - y^*$ can be fed to a controller $G_c(s)$, parameterized as if $G_{p-}(s)$ were the process transfer function.

For the predictor to be stable it is necessary and sufficient that the process be open-loop stable (assumed in this work). The GSP is equivalent to classical feedback (see Figure 9) with

$$C(s) = \frac{G_c(s)}{1 + G_c(s)G_{p-}(s)(1 - G_{p+}(s))}$$

(19)

However, the GSP structure will make the control system’s design more transparent. The closed-loop response of the GSP is

$$y(s) = \frac{G_c(s)G_{p+}(s)}{1 + G_c(s)G_{p-}(s)G_{p+}(s)}y_p(s) + \left[1 - \frac{G_c(s)G_{p-}(s)}{1 + G_c(s)G_{p-}(s)G_{p+}(s)}\right]d(s)$$

(20)

For ISE-optimal pole placement with respect to step changes

in $y_p$ or $d$, the closed-loop transfer function must have poles at the LHP zeros of the open-loop transfer function and at the mirror images of its RHP zeros with respect to the imaginary axis; the remaining poles must be placed “far left” in the complex plane (Letov, 1960; Chang, 1961). This result is rigorously valid for systems with rational transfer functions only. However, it is known that for any arbitrary transfer function there exists a rational function arbitrarily close to it. Therefore, one intuitively expects that such a placement will be optimal (or at least “nearly” optimal) for distributed systems as well. This suggests the following design procedure:

i) Select the factorization of $G_p(s)$ so that in addition to the previous requirements, $G_{p+}(s)$ is of the form

$$G_{p+}(s) = \frac{R(s)}{R^*(s)}$$

(21)

where $R(s)$ is an entire function whose zeros are exactly the RHP zeros of $G_p(s)$, $R^*(s)$ is an entire function whose zeros are exactly at the mirror images of the RHP zeros of $G_p(s)$. The factor $R^*(0)/R(0)$ is included to make $G_{p+}(0) = 1$.

ii) Parameterize $G_c(s)$ according to

$$G_c(s) = \frac{1}{W(s) - G_{p-}(s)}$$

(22)

where $W(s) = \Pi_{r=1}^r(p_r,s + 1)$ is the polynomial of the “free” poles, whose order $r$ is such that $G_c(s)$ is proper. This is the standard zero-pole placement formula for minimum phase lumped parameter systems, where $[W(s)]^{-1}$ is the desired closed-loop transfer function; it has been referred to as the synthesis formula (Smith and Corripio, 1985). When there is more than one free pole, to facilitate the search for the $p_i$’s, a reasonable requirement is that the closed-loop response be critically damped. This is equivalent to selecting all the $p_i$’s to be equal.

iii) The free poles are understood to be placed “significantly far left” in the complex plane, but how far left will depend on robustness. This is addressed in the next section. With the above choice of factorization and parameterization of $G_c(s)$, the equivalent unity feedback controller is given by

$$C(s) = \frac{1}{W(s) - G_{p-}(s)G_{p+}(s)}$$

(23)

and the closed-loop response (for a perfect model) becomes

$$y(s) = \frac{1}{W(s)}G_{p+}(s)y_p(s) + \left(1 - \frac{1}{W(s)}G_{p-}(s)\right)d(s)$$

(24)

Equation 24 gives the optimal response when the free poles are placed at $-\infty (p_i = 0)$, that is, when $W(s) = 1$. Equations 22 and 23 indicate that this would require infinite control action and is impossible to achieve. However, the response with $p_i = 0$ provides an ideal performance that can be used to evaluate the performance of different designs. Further, Eq. 24 clearly shows that the achievable performance is limited by the presence of the nonminimum phase factor $G_{p+}(s)$. This limit cannot be exceeded by controller design. Improvement beyond this limit would require modification of the plant to eliminate the nonminimum phase terms.
Selection of $G_p(s)$

Three cases were identified for QRDS that correspond to: 1) no RHP zeros; 2) finite number of RHP zeros; and 3) an infinite chain of RHP zeros.

Case 1. $G_p(s) = 1$ and the GSP reduces to a unity feedback controller: $C(s) = G(s)$. The controller can be parameterized according to Eq. 22.

Case 2. The process has a finite number of RHP zeros. Given their location, say $z_1, \ldots, z_n$, then

$$R(s) = \prod_i (z_i - s)$$

and

$$R^*(s) = \prod_i (z_i + s)$$

Case 3. The process has an infinity of RHP zeros and possibly a finite number of LHP zeros ($-z_i$). Then

$$R(s) = \frac{P_1(s) - P_2(s)e^{-\omega s}}{\Pi_i(z_i - s)}$$

$$R^*(s) = \frac{P_1(s) - P_2(s)e^{\omega s}}{\Pi_i(z_i - s)}$$

and Eq. 21 defines $G_p(s)$. Equation 28 was obtained from Eq. 27 by substituting $-s$ for $s$ and by multiplying $e^{-\omega s}$, which is required for realizability.

The expressions for $R(s)$ and $R^*(s)$ given by Eqs. 27 and 28 represent the quotient of the perfect division of a quasipolynomial by a polynomial with roots at its own zeros. This quotient cannot be written in closed form; one can write it as an infinite product but not with a finite formula in terms of elementary functions. It is more compact and easier to write it as a fraction with the perfect division being implied.

Remarks

The key issue in the control of QRDS is how to cope with the transcendental nature of the transfer function, which in general will produce an infinity of zeros; it is not the choice of predictive control framework. Alternative predictive control frameworks like the IMC would in principle yield equivalent results. The proposed framework seems to be the most appealing for the following reasons:

i) The Smith predictor idea provides a nice way to “decouple” the minimum phase part of the process, which will be controlled, from the nonminimum phase part, which will stay in open loop. This “decoupling,” is particularly convenient in delay and general distributed systems.

ii) The pole placement philosophy leads to a natural split of $G_p(s)$ into a minimum phase and a nonminimum phase part. At the same time, it provides a natural parameterization of the controller $C(s)$ that guarantees closed-loop stability. The adjustable parameters $p_i$’s are interpreted as free poles, to be placed significantly “far left,” rather than time constants of an ad-hoc filter.

Robust Design

The results of the previous section are under the assumption of a perfect model. This assumption is of course unrealistic and one would wish to guarantee stability and “high performance” for any process model within a specified uncertainty band.

The most popular way to specify model uncertainty bounds has been in the frequency domain in terms of a multiplicative or “percent” uncertainty. According to Doyle and Stein (1981), one can represent the “true” process transfer function $G_p^T(s)$ as

$$G_p^T(s) = G_p(s)[1 + L_m(s)]$$

where $G_p(s)$ is the process model and $L_m(s)$ the multiplicative uncertainty. However, $L_m(s)$ is seldom known exactly. Usually only an upper bound $I(\omega)$ defined according to

$$|L_m(\omega)| < I(\omega)$$

can be estimated. $G_p^T(s)$ therefore represents a family of plants that includes the process model.

The ISE-optimal pole placement will be applicable to the specific $G_p^T(s) = G_p(s)$ and not to the entire family of $G_p^T$’s. Although in principle it is possible to reformulate the optimization problem to optimize the “worst” performance over the family of $G_p^T$’s (Doyle, 1984), the following two step procedure is more practical:

Step 1. Design a controller assuming perfect model, using the framework described in the previous section.

Step 2. Select the free poles ($p_i$’s) so as to meet robust stability and performance specifications.

The same two-step idea is central in the robust design of IMC controllers (Mandler et al., 1986; Zafiriou and Morari, 1986; Morari and Doyle, 1986). The difference is of course the origin of the controller parameterization (Q-parameterization vs. closed-loop pole parameterization). It should be stressed that the robust design should be based on an overall uncertainty bound between the plant and model. A design based on a partial uncertainty bound or considering only uncertainties in $G_p(s)$, for example, could lead to closed-loop systems with poor robustness.

The robust stability condition for the GSP, following the approach of Doyle and Stein (1981), simplifies to

$$|W(\omega)| > I(\omega)$$

The robust performance condition, usually imposed over a certain low frequency band, is

$$\frac{P_2(\omega) + 1}{1 - I(\omega)} < \frac{1}{|W(\omega)|}$$

for $0 < \omega < \omega^*$

where the performance function $P_2(\omega)$ is a priori determined by the designer and is associated to the ability of the closed-loop system to attenuate disturbances within the specified frequency band.

One of the major difficulties hindering application of the above theory is the determination of an overall uncertainty bound $I(\omega)$ in an actual process. When it is too difficult or too costly to estimate $I(\omega)$, the free parameters ($p_i$’s) will have to be tuned on-line.
Examples: Control of Cumulative Mass Fraction

**Fluidized bed calciner**

The transfer function for the cumulative mass fraction in a fluidized bed calciner is reproduced below from Eq. 2 for a cut-point size $Z_c = 2$:

$$G_p(s) = \frac{(8s^3 + 36s^2 + 60s + 38)e^{-2s} - 1}{s(s + 1)(s^2 + 3s + 3)}$$

(33)

The numerator quasipolynomial can be written as

$$N(s) = \frac{(8s^3 + 36s^2 + 60s + 38)}{38}e^{-2s} - 1$$

(34)

while the corresponding image quasi-polynomial is

$$N_i(s) = \frac{(-8s^3 + 36s^2 - 60s + 38)}{38}e^{-2s}$$

(35)

The zeros of Eq. 34 are plotted in Figure 7. This system has three LHP zeros, a zero at the origin and an infinity of RHP zeros. The LHP factors are $(s + 1.545)$ and $(s^2 + 2.319s + 1.73)$.

**Selection of $G_p^+(s)$.** This system corresponds to case 3. The nonminimum phase factor, $G_p^+(s)$ defined by eqs. 21, 27 and 28, is

$$G_p^+(s) = \frac{N(s)(-s + 1.545)(s^2 - 2.319s + 1.73)}{(s + 1.545)(s^2 + 2.319s + 1.73)N_i(s)}$$

(36)

while the minimum phase factor is given by

$$G_p^-(s) = \frac{-N_i(s)(s + 1.545)(s^2 + 2.319s + 1.73)}{s(s + 1)(s^2 + 3s + 3)(-s + 1.545)(s^2 - 2.319s + 1.73)}$$

(37)

**Controller Synthesis.** The desired closed-loop transfer function, $[W(s)]^{-1}$, has to be selected in order to use Eq. 22 for the synthesis of $G_c(s)$. In this case, a first order polynomial

$$W(s) = p_1s + 1$$

(38)

makes the controller proper. As pointed out earlier, the righthand side of Eqs. 36 and 37 involve pole-zero cancellations. For this reason, special care must be taken when simulating $G_c(s)$ and the predictor $(1 - G_p^+(s))G_p^-(s)$ on-line.

Using a 10% uncertainty in the cut-point size, for illustrative purposes, the uncertainty bound is

$$l(\omega) = \left| \frac{G_p(Z_c = 2.2) - G_p(Z_c = 2.0)}{G_p(Z_c = 2.0)} \right|$$

(39)

This bound is shown in Figure 10 together with $|W(\omega)|$ for $p_1 = 0.2$ and 0.3. The robust stability condition in Eq. 31 is satisfied for all $p_1 \geq 0.2$. A value of 0.3 was chosen to limit the overshoot.

The closed-loop cumulative mass fraction response of a plant at the nominal cut-point ($Z_c = 2$) and plants at the extreme cut-points ($Z_c = 1.8$ and 2.2) are given in Figure 11 for a unit step change in the set-point. For comparison, the closed-loop responses of the cumulative mass fraction at the nominal cut-point are shown in Figure 12 for PID controllers at the Ziegler-Nichols ($K_c = -5.3$, $\tau_i = 1.48$ and $\tau_d = 0.37$) and optimal settings ($K_c = -4.5$, $\tau_i = 0.8$ and $\tau_d = 0.37$). The optimal PID settings were determined from simulations using the Ziegler-Nichols settings as a starting point for the parameter search. Many trials were required to arrive at the optimal settings.

The performance and robustness of the PID controller at the optimal settings are comparable to that of the GSP controller for this system. As expected, the PID controller at the Ziegler-Nichols settings is considerably poorer. The time histories of the

**Figure 10. Uncertainty bound for cumulative mass fraction in a fluidized bed calciner for a 10% uncertainty in the cut-point ($Z_c$).**

**Figure 11. Fluidized bed calciner, response of the cumulative mass fraction for the nominal ($Z_c = 2$) and extreme ($Z_c = 1.8$ or 2.2) plants to a step change in set-point using a GSP controller ($p_1 = 0.3$).**
Figure 12. Fluidized bed calciner, response of the cumulative mass fraction for the nominal ($Z_c = 2$) plant to a step change in set-point using a PID controller at the optimal and Ziegler-Nichols settings.

The manipulated variable corresponding to the simulations for the nominal cut-point are shown in Figure 13 for the GSP and optimal PID controllers. The GSP controller requires larger initial control action. This will limit the magnitude of the set-point or disturbance inputs that the controller can tolerate without saturation.

**MSMPR crystallizer**

This example presents results for the control of a MSMPR crystallizer, with $i = 5$ and $j = 1$, using residence time as the manipulated variable. Figure 8 shows that this system has a finite number of (nine) RHP zeros. Hence, this system corresponds to case 2. The nonminimum phase factor, $G_{p+}(s)$, can be calculated using Eqs. 21, 25 and 26, while the controller is defined by Eq. 23. Here also a first order polynomial, $W(s) = p_1s + 1$, would make the controller proper. Assuming an uncertainty bound of 10% in the cut-point size and using a procedure similar to that for the fluidized bed calciner, a $p_1 \approx 0.3$ was found to satisfy the robust stability condition.

The closed-loop responses of the cumulative mass fraction for the nominal and extreme cut-points are given in Figure 14 for a unit step change in set-point with $p_1 = 0.3$. For comparison, the responses with an optimally tuned ($K_p = 0.9$, $\tau_i = 0.85$, and $\tau_d = 0.37$) PID control system are shown in Figure 15. Here the performance of the GSP controller is much better than that of the optimal PID controller.

The difference between the performance of GSP and optimal PID controllers for the fluidized bed calciner and MSMPR crystallizer examples can be attributed to the number and location of the RHP zeros with respect to the imaginary axis.

**Controller Implementation Aspects**

As noted earlier, the expressions for $R(s)$, $R^*(s)$ and therefore those for $G_{p+}(s)$ and $G_{p-}(s)$ (see e.g., Eqs. 36 and 37) can involve the quotient of the perfect division of a quasipolynomial by a polynomial with roots at some of the RHP zeros of the quasipolynomial. The immediate question that arises is how can it be simulated on-line when those zeros are only known approximately (since they can only be calculated numerically.) There are two major techniques for the on-line simulation of transcendental transfer functions:

i) Off-line calculation of the controller's impulse response by numerical inversion of the Laplace transform and implementation as a time-discretized convolution integral.

ii) Use of a sufficiently high-order Padé approximation of the exponential term to obtain a rational approximation of the controller transfer function.

Our experience with numerical inversion techniques of the Laplace transform has indicated that they are completely insensitive to small zero-pole mismatch. The reason is that any such small mismatch will only affect the very-high-$|s|$ values of the transfer function on the Bromwich contour. Consequently, this effect will not be sensed after numerical truncation.

In the case where Padé approximations are used, it has been shown by Takahashi et al. (1987) that the low modulus zeros of a quasipolynomial can be calculated to any desired degree of
accuracy by using a sufficiently high-order Padé approximation. As only low modulus roots participate in zero-pole cancellations, the use of Padé approximation for the delay terms yields a stable rational approximation for the quotient of the transcendental transfer function after the zero-pole cancellation is performed. This allows approximation of a transcendental controller transfer function as a stable rational transfer function.

The use of high-order Padé approximation is further supported by the fact that the mismatch between the transcendental and rational (Padé approximation) transfer functions occurs at large |s|. Therefore the relative contribution of the mismatch, introduced by Padé approximation, to the overall dynamics is insignificant. The disadvantages with using Padé approximation are:

i) The Padé coefficients tend to increase very rapidly with Padé order, leading to numerical problems.

ii) The minimum Padé order required to yield satisfactory accuracy will depend on the system transfer function. Hence, simulations are required.

Conclusions

The literature on the design of model-based predictive controllers for time delay and finite dimensional nonminimum phase systems is quite extensive. However, extensions to distributed systems have been limited. In this paper, a class of distributed systems modeled by hyperbolic PDE’s is identified. This class, called QRDS, has an infinity of zeros lying in the LHP or RHP. Formulas for factoring QRDS into minimum and nonminimum phase factors and a generalized Smith predictor (GSP) controller design procedure were presented.

The GSP design procedure uses pole-placement to obtain a controller parameterization that guarantees closed-loop stability. The free parameters are selected to satisfy robustness and performance specifications. The examples on control of the cumulative mass fraction demonstrates the advantages of the GSP design procedure. A robust design is obtained in a rather straightforward manner without having to perform many simulations to search for the best settings.

Further, the cumulative mass fraction examples are significant in that they provide an implementable scheme for controlling particle size distribution in particulate systems. The advantage of controlling cumulative mass fraction is that it can be measured fairly easily on-line compared to the size distribution or its moments. In the past, the lack of instrumentation and suitable controller design techniques coupled with complicated dynamics have prevented the automatic control of particle size distribution.

Notation

c = solute concentration in crystallizer, based on clear liquor
c_f = solute concentration in crystallizer feed
c_s = solute saturation concentration
c_r = crystallizer parameter, \( c_r/c \)
c_m = crystallizer parameter, \( (1 - k_m Z_m / (1 - C)) \)
c_c = crystallizer parameter, \( (1 - c_c / C) \)
c_n = crystallizer parameter, \( (1 - c_n / C) \)
c_c = dimensionless solute concentration, \( c / c \)
c_c = dimensionless saturation concentration, \( c / c \)
C(s) = unity feedback controller transfer function
d(s) = process disturbance
g = linear particle growth rate
G = dimensionless particle growth rate
\( G_p(s) \) = controller transfer function
\( G_p(s) \) = process transfer function
\( G_{ps}(s) \) = nonminimum phase, allpass component of plant model
\( G_{pc}(s) \) = minimum phase component of plant model
i = nucleation to growth rate kinetic sensitivity parameter, or \( \sqrt{-1} \)
j = solid concentration dependent nucleation kinetics exponent
\( k_n \) = area shape factor
\( k_g \) = growth rate kinetic constant
\( k_n \) = nucleation rate kinetic constant
\( k_v \) = volume shape factor
K = process gain, or ratio of coefficient of dominant power of \( P_1(s) \) by coefficient of dominant of power of \( P_1(s) \)
K_r = controller gain
\( l(\omega) \) = upper bound for multiplicative uncertainty
\( L_m(s) \) = multiplicative uncertainty
m = order of polynomial \( P_1(s) \)
m_2, m_3 = second, third moment of the number distribution function
\( M_2, M_3 \) = cumulative mass fraction normalized with respect to the initial steady-state value
\( M_2, M_3 \) = dimensionless second, third moment
n = order of polynomial \( P_1(s) \)
\( N(s) \) = QRDS numerator quasi polynomial
\( N(s) \) = QRDS image quasi polynomial
\( P_1(s) \) = performance function
\( R(s) \) = an entire function containing all the RHP zeros of \( G_p(s) \)
\( R^*(s) \) = an entire function whose zeros are exactly at the mirror image of the RHP zeros of \( G_p(s) \)
s = Laplace variable
t = time
\( t_d \) = time delay
T = dimensionless residence time
\( W(s) \) = polynomial of “free” poles
\( y(s) \) = controlled variable
\( y_p(s) \) = set-point
Z = dimensionless particle size
Z_c = dimensionless cut-point size

Greek letters

\( \rho \) = density
\( \epsilon \) = liquid fraction in crystallizer
\( \tau \) = time constant or residence time
\( r_i \) = Integral time of PID controller
\( r_d \) = Derivative time of PID controller
\( \omega \) = frequency
Literature Cited


Doyle, J. C., Lecture notes, ONR/Honeywell Workshop on Advances in Multivariable Control, Minneapolis (1984).


Manuscript received June 14, 1988, and revision received Feb. 16, 1989.