Wiener-Hopf Methods for Unstable, Nonminimum Phase Processes

H. C. LIM and S. G. BANKOFF
Northwestern University, Evanston, Illinois

Part I. Optimal Regulators

It is shown that existing methods for design of optimal regulators by the Wiener-Hopf procedure must be modified in order to be applicable to unstable and/or nonminimum phase plant or disturbance transfer functions, such as are frequently encountered in the chemical industry. Solutions are developed for three cases: I. stable, but possibly nonminimum phase, plant and disturbance transfer functions; II. minimum phase, but possibly unstable, plant with no restrictions on the disturbance transfer functions; and III. prestabilized, with a proper modification to retain the original control effort inequality constraints, but possibly nonminimum phase plant and disturbance transfer functions. Case III gives the general solution for regulation of linear, time-invariant, lumped-parameter systems. When prestabilization is not necessary, it reduces to case I. Where applicable, solutions by the method of case II frequently involve less algebra than in case III.

The Wiener-Hopf theory of spectral factorization in the complex Fourier plane was developed more than two decades ago. One of the first applications was in the design of optimal filters, for which auxiliary inequality constraints were necessary. Newton (*1*) modified the method to synthesize optimum servo controllers, subject to mean-square constraints on controller effort, by means of a Lagrangian multiplier technique. The system to be controlled must be described by constant-coefficient, linear, ordinary differential equations, possibly involving a time delay. The Wiener-Hopf equation, which corresponds to the Euler-Lagrange equation in the calculus of variations, is an integral equation solvable in either the time or frequency domain (*1* to *9*).

Newton (*2*) pointed out that a filter in open-loop cascade with the plant was, in principle, equivalent to a feedback servo-controller loop. For unstable plants, he advocated that an internal stabilizing loop be placed around the plant before applying the method. However, in the presence of one or more mean-square constraints on the control efforts, this approach will lead to a different problem, since the constraints are then placed on signals related to the original control efforts.

Chang (*3*) considered a fairly general control problem where the disturbance transfer function could be other than unity. However, it will be shown that his solution formulation must be modified if the disturbance transfer function is unstable and/or nonminimum phase (NMP). In particular, one obtains nonrealizable (predictive) controllers if the disturbance and plant transfer functions exhibit dead time, as found by Luecke and McGuire (*10*).

In the chemical industry, plant transfer functions frequently involve dead time (a form of NMP) and sometimes are unstable. Hence, there is a need for a general solution to handle all possible cases, wherein either or both the plant and disturbance transfer functions may be unstable and/or NMP. In this paper, the synthesis of optimum feedback regulators by the Wiener-Hopf method is extended to any physical system which can be described by constant-coefficient, linear, ordinary differential equations with or without a delayed time argument.

H. C. Lim is at Purdue University, West Lafayette, Indiana.

FORMULATION OF THE CONTROL PROBLEM

The control system represented by Figure 1 is subjected to a stationary random disturbance whose autocorrelation function is known. The problem is to determine the linear regulator $\hat{G}_c$ which minimizes the mean-square error, subject to mean-square constraints on the control efforts.

Figure 1 represents a typical control system for systems described by a time-invariant linear differential equation of the form

$$A(p)c(t) = L(p)m(t-\tau_p) + U(p)d(t-\tau_d) \quad (1)$$

where

$$A(p) = \sum_{i=0}^{n} a_i p^i,\ L(p) = \sum_{j=0}^{m} l_j p^j,\ U(p) = \sum_{k=0}^{l} u_k p^k$$

and

$$p^i = \frac{d^i}{dt^i},\ n > m, l$$

τ_d and τ_p are associated with the disturbance variable $d(t)$ and manipulated variable $m(t)$, respectively. Therefore, in the Laplace transform domain

$$G_d(s) \triangleq \frac{C(s)}{D(s)} = \frac{U(s)e^{-\tau_d s}}{A(s)} = \frac{a\, e^{-\tau_d s} \prod_k (s - z_k)}{\prod_i (s - p_i)} \quad (2)$$

and

$$G_p(s) \triangleq \frac{C(s)}{M(s)} = \frac{L(s)e^{-\tau_p s}}{A(s)} = \frac{b\, e^{-\tau_p s} \prod_j (s - z_j)}{\prod_i (s - p_i)} \quad (3)$$

If the real part of a pole or zero is positive, it lies in the right-hand plane (RHP). RHP zeros correspond to nonminimum phase and RHP poles to instability. Let $B_j(s)$, $j = 1,2 \ldots, q$ be physically realizable linear operators on the control variable $M(s)$, such that $M_j(s) \triangleq B_j(s)M(s)$ represent the control effort upon which a mean-square constraint is to be placed. For example, $M_1 = M$ and $M_2 = \frac{M}{s}$ represent a situation in which it is desired to place upper bounds both on $\overline{m^2(t)}$ and $\overline{[\int m dt]^2}$, where the overbar denotes a time average.

Mathematically, the problem is to minimize $\overline{e^2(t)}$, subject to the constraints $\overline{m_j^2(t)} \leqq W_j^2$, $j = 1,2, \ldots q$, which is equivalent to finding the stationary point of the Lagrangian function (*11*):

$$I = \overline{e^2(t)} + \sum_{j=1}^{q} \lambda_j^2 \, \overline{m_j^2(t)} \tag{4}$$

Equation (4) can be written in terms of spectral densities (*9, 12*) as

$$I = \frac{1}{2\pi i} \int_{-i\infty}^{+i\infty} \left[\Phi_{ee}(s) + \sum_{j=1}^{q} \lambda_j^2 \, \Phi_{m_j m_j}(s) \right] ds \tag{5}$$

where

$$i = \sqrt{-1}, \; s = i\omega$$

$$\Phi_{ee}(s) = \int_{-\infty}^{+\infty} \phi_{ee}(\tau) \, e^{-s\tau} d\tau$$

$$\phi_{ee}(\tau) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} e(t) \, e(t+\tau) \, dt$$

and

$$\Phi_{m_j m_j}(s) = \int_{-\infty}^{\infty} \phi_{m_j m_j}(\tau) e^{-s\tau} \, d\tau = \Phi_{mm}(s) \; |B_j(s)|^2 \tag{6}$$

The autocorrelation function $\phi(\tau)$ and the spectral density function $\Phi(s)$ thus constitute a Fourier transform pair. From Figure 1

$$-E = [G_m G_d/(1 + G_c G_m G_p)]D \tag{7}$$

$$M = [-G_c G_d G_m/(1 + G_c G_m G_p)]D \tag{8}$$

$$= (C/D - G_d) \; D/G_p$$

Hence, Equation (5) can be written as

$$I(G_c) = \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} \left\{ |G_m G_d/(1 + G_c G_m G_p)|^2 + \sum_{j=1}^{q} \lambda_j^2 \, |G_c G_d G_m/(1 + G_c G_m G_p)|^2 \right\} \Phi_{dd} \; ds \tag{9}$$

where Φ_{dd} is the spectral density of the stationary random disturbance $d(t)$.

In order to satisfy the necessary conditions for optimality, the objective is to find the function G_c so that the functional $I(G_c)$ is stationary. However, G_c appears nonlinearly in Equation (9), which complicates the solution. A new function, $H(s)$, containing the unknown $G_c(s)$ is defined such that $E(s)$ and $M(s)$ are linear functions of $H(s)$. From (6), (7), and (8), $I[H(s)]$ will then be quadratic in $H(s)$. The Wiener-Hopf method is then used to solve for $\hat{H}$ and hence $(\widehat{C/D})$, where the brevet denotes an optimal function. From (6), the corresponding optimum controller is then given by

$$\hat{G}_c = \frac{G_d - (\widehat{C/D})}{(\widehat{C/D}) \; G_m G_p} \tag{10}$$

However, it is also necessary to insure that: the overall system transfer function (C/D) is physically realizable; the transfer functions (M_j/D) are stable and nonanticipatory (respond in the present to future signals), so that mean-square constraints can be imposed; G_c is stable and nonanticipatory; and (C/D) must not be destabilized owing to infinitesimal changes in the plant parameters. The last requirement implies that exact cancellation of RHP zeros and poles in the fixed member functions by factors in G_c must be avoided. For physical realizability, (C/D) must be stable, nonanticipatory (no $e^{\tau s}$ factor, corresponding to negative dead time), and must have the order of the numerator polynomial less than or equal to the order of the denominator polynomial. Some of these functional constraints are frequently not obvious, and the difficulties are not readily recognized. In the next section, therefore, a solution form given previously (3), for a restricted class of problems is examined, and some illustrations are given to demonstrate the difficulties imposed by the functional constraints. Let G_d and G_p be stable transfer functions. Inspection of Equation (7) suggests that one can choose, as done by Chang (3) and Luecke and McGuire (*10*)

$$(C/D - G_d) = N_p Q \tag{11}$$

where $N_p \triangleq e^{-\tau_p s} \prod_k (s - z_k)$, $z_k > 0$ which are the NMP factors in G_p, and Q is a stable, nonanticipatory, but otherwise arbitrary function. It is readily seen from Equations (7) and (8) that E and M are linear functions of the unknown function Q, so that the first step of the solution has been achieved. Before we proceed, however, the previous list of functional constraints must be checked. Since G_d is stable and nonanticipatory, so is $C/D = N_p Q + G_d$. The second functional constraint is also satisfied, since

$$M/D = (C/D - G_d)/G_p = N_p Q/G_p = Q/G_p' \tag{12}$$

where $G_p' \triangleq G_p/N_p$

The third functional constraint is placed on

$$G_c = (G_d - C/D)/[G_m G_p(C/D)] = -N_p Q/[G_m G_p(G_d + N_p Q)] \tag{13}$$

Hence, if G_d and G_p both have dead time (τ_d and τ_p nonzero), the controller obtained from Equation (13) is anticipatory and hence physically unrealizable. To demonstrate this, we let $G_d = e^{-\tau_d s} G_d'$, $G_m = 1$, and $G_p = e^{-\tau_p s} G_p'$. Then

$$G_c = Q/[G_p'(e^{-\tau_d s} G_d' + e^{-\tau_p s} Q]$$

$$= e^{\tau_d s} Q/[G_p'(G_d' + e^{(\tau_d - \tau_p)s} Q)] \quad \text{for} \quad \tau_d \leqq \tau_p$$

$$= e^{\tau_p s} Q/[G_p'(e^{(\tau_p - \tau_d)s} G_d' + Q)] \quad \text{for} \quad \tau_d > \tau_p$$

Therefore, Equation (11) fails, since the corresponding optimum regulator will be physically unrealizable. Note that this does not imply the nonexistence of the optimum controller but simply the inapplicability of Equation (11) without further modification. Furthermore, it is clear from Equation (12) that if G_d and G_p have a common RHP zero, this will appear as an RHP pole in G_c, so that the form of Equation (11) requires modification for this case also.

The fourth functional constraint is now investigated by writing Equation (11) in the form

$$C/D = G_d + N_pQ = G_d/(1 + G_cG_mG_p) \quad (14)$$

This implies that the RHP zeros and dead time of G_d must appear in C/D, unless exactly cancelled by $(1 + G_cG_mG_p)$. However, exact cancellation of dead time is not possible, and in practice exact cancellation of the RHP zeros also fails, owing to small variations in their location. To illustrate this, let $G_d = \frac{s-1}{(s+1)(s+2)}$, $G_m = 1$, and $G_p = \frac{1}{(s+1)(s+2)}$. The optimal controller is obtained from Equation (13):

$$\hat{G}_c = - \hat{Q} \Big/ \left\{ \frac{1}{(s+1)(s+2)} \left[\frac{s-1}{(s+1)(s+2)} + \hat{Q} \right] \right\} \quad (15)$$

$$= \frac{-\hat{Q}(s+1)^2(s+2)^2}{(s-1) + \hat{Q}(s+1)(s+2)}$$

Now, let $\tilde{G}_d = \frac{s-1-\epsilon}{(s+1)(s+2)}$. The overall system transfer function is then

$$C/D = \tilde{G}_d/(1 + \hat{G}_cG_mG_p) \quad (16)$$

$$= \frac{(s-1-\epsilon)[(s-1) + \hat{Q}(s+1)(s+2)]}{(s-1)(s+1)(s+2)}$$

which is unstable unless $\hat{Q}$ contains $(s-1)$ explicitly.

Similarly, from Equation (13), the RHP poles of G_p appear as RHP zeros in G_c, which implies, from Equation (14), that stabilization of (C/D) also then depends upon exact cancellation and is hence inadmissible. Thus, Equation (11) can be used only when G_d is stable and minimum phase (MP) and G_p is stable. If any of these conditions is violated, a different approach is required.

Case I: Stable G_d and G_p

It is easy to show that for stable open-loop systems [G_d and G_p admit poles in the left-half plane (LHP) only] all functional constraints are met with

$$C/D = H = G_d + N_dN_pQ \quad (17)$$

where $N_d \triangleq e^{-\tau_d s} Z_d$ and $N_p \triangleq e^{-\tau_p s} Z_p$ are then the NMP factors in G_d and G_p, respectively, and Q is a stable, but otherwise, arbitrary function to be determined. Here $Z_d \triangleq \prod_i (s - z_i)$, $z_i > 0$, for example, represents the RHP zeros of G_d.

In terms of Equation (17), E and M can be written as

$$E = R - G_mC = -(G_d + N_dN_pQ)G_mD = -N_d(G_d' + N_pQ)G_mD \quad (18)$$

where $G_d' \triangleq G_d/N_d$. Similarly

$$M = (C - G_dD)/G_p = (N_dQ/G_p')D \quad (19)$$

$$M_j = B_jM, \quad j = 1, 2, \ldots, q \quad (20)$$

where $G_p' \triangleq G_p/N_p$. The corresponding spectral densities are

$$\Phi_{ee} = |N_d(G_d' + N_pQ)G_m|^2 \, \Phi_{dd} \quad (21)$$

$$\Phi_{mm} = |N_dQ/G_p'|^2 \, \Phi_{dd} \quad (22)$$

together with Equation (6). In terms of these spectral densities, Equation (5) can be written as

$$I = \frac{1}{2\pi i} \int_{-i\infty}^{+i\infty} \left[|(G_d' + N_pQ)G_m|^2 + \sum_{j=1}^{q} |\lambda_jB_j|^2 \, |N_pQ/G_p|^2 \right] |N_d|^2 \, \Phi_{dd} \, ds \quad (23)$$

Let $\hat{Q}$ minimize the functional $I(Q)$, and let

$$Q = \hat{Q} + \epsilon K \quad (24)$$

so that

$$Q^* = \hat{Q}^* + K^* \quad (25)$$

where ϵ is a small real constant, K is an arbitrary stable function, and $Q^*(s) = Q(-s)$. The expansion of Equation (23) in terms of Equations (24) and (25) yields

$$I = I_1 + \epsilon(I_2 + I_2^*) + \epsilon^2 I_3 \quad (26)$$

where

$$I_1 = \hat{I} = \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} \left[|G_m(G_d' + N_p\hat{Q})|^2 + \sum_{j=1}^{q} |\lambda_jB_j|^2 \, |N_p\hat{Q}/G_p|^2 \right] |N_d|^2 \, \Phi_{dd} \, ds \quad (27)$$

$$I_2 = \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} K^* \left[G_d'N_p^* \, |G_m|^2 + |N_p|^2 \, \hat{Q} \left\{ |G_m|^2 + \sum_{j=1}^{q} |\lambda_jB_j/G_p|^2 \right\} \right] |N_d|^2 \, \Phi_{dd} \, ds \quad (28)$$

$$I_3 = \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} \left\{ |G_m|^2 + \sum_{j=1}^{q} |\lambda_jB_j/G_p|^2 \right\} |N_dN_pK|^2 \, \Phi_{dd} \, ds \geq 0 \quad (29)$$

Therefore, the necessary and sufficient condition is that $I_2 = 0$. By following the well-known spectral factorization (3) procedure $\hat{Q}$ is found to be

$$\hat{Q} = -[G_d' \, |G_m|^2 \, Z_d^*N_p^* \, \{\Phi_{dd}\}^+/(Z_p \, \{V\}^-]_+/(Z_d^*Z_p^* \, \{\Phi_{dd}\}^+ \, \{V\}^+) \quad (30)$$

$$= -[G_d \, |G_m|^2 \, N_d^*N_p^* \, \{\Phi_{dd}\}^+/(Z_dZ_p \, \{V\}^-]_+/(Z_d^*Z_p^* \, \{\Phi_{dd}\}^+ \, \{V\}^+)$$

where

$$\{V\}^+ \, \{V\}^- \triangleq |G_m|^2 + \sum_{j=1}^{q} |\lambda_jB_j/G_p|^2 \quad (31)$$

and

$$\{\Phi_{dd}\}^+ \, \{\Phi_{dd}\}^- \triangleq \Phi_{dd} \quad (32)$$

For example, this implies that Φ_{dd} has been split into two factors, $\{\Phi_{dd}\}^+$ and $\{\Phi_{dd}\}^-$, which have all their poles and zeros in the LHP and the RHP, respectively. Also, a decomposition into positive and negative time functions is implied by

$$[F(s)] \triangleq [F(s)]_- + [F(s)]_+ = \int_{-\infty}^{0} f(t)e^{-st}dt + \int_{0}^{\infty} f(t)e^{-st}dt, \quad s = i\omega \quad (33)$$

When $F(s)$ is rational in s, the operation $[F(s)]_+$ is equivalent to partial decomposition into fractions and then keeping those terms with poles in the LHP interior. Then

$$\hat{H} = (\hat{C/D}) = N_d N_p \hat{Q} + G_d \quad (34)$$

and the corresponding optimum feedback regulator is

$$\hat{G}_c = (G_d - \hat{H})/G_m G_p \hat{H} = -\hat{Q}/[G_m G_p' (G_d' + N_p \hat{Q})] \quad (35)$$

The minimum value of the Lagrangian function is given by Equation (27). However, $\hat{H}$ as given by Equations (30) and (34) is a function of the Lagrangian multipliers λ_j, which must be determined from the control constraints in order to complete the solution. This evaluation will be discussed later. The optimum feedback regulator for a single-disturbance input which can be obtained by the approach used by Chang (*3*) and Luecke and McGuire (*10*) is a special case of the solution given here when $G_m = 1$ and G_d has no NMP factors.

Having completed the general solution for open-loop stable systems, we now proceed to cases where the open-loop systems are unstable. Once again, a somewhat restricted case will be considered first, since the computational difficulties due to the involved algebra are frequently reduced for this case.

Case II: Minimum Phase G_p

The only restriction we shall now impose is that G_p has no RHP zero and no dead time. No restriction on G_d is required. The form of H which meets all the functional constraints is

$$H = C/D = N_d P_{pd} F \quad (36)$$

where N_d comprises NMP in G_d, P_{pd} represents RHP poles in G_p not also present in G_d, and F is a stable, but otherwise arbitrary function. As an example, if $G_d = 1/[(s-1)(s+2)]$ and $G_p = 1/[(s-1)(s-2)(s+2)]$, then $P_{pd} = (s-2)$ is explicitly included in Equation (36) to avoid exact cancellation of the RHP poles in G_p not also present in G_d by the RHP zeros in the controller $\hat{G}_c$, since P_{pd} will appear as RHP zeros in $\hat{G}_c$. Note that if the disturbance transfer function is minimum phase and contains all the unstable poles of G_p, this expression reduces to $H = F$, which is of the form recommended by Chang (*3*) to reduce the amount of computation over the form of Equation (17). The Lagrangian to be minimized is now

$$I = \overline{e^2(t)} + \sum_{j=1}^{q} \lambda_j^2 \overline{m_j^2(t)} = \frac{1}{2\pi i}\int_{-i\infty}^{+i\infty} \left[|G_m N_d P_{pd} F|^2 + \sum_{j=1}^{q} |\lambda_j^2 B_j (N_d P_{pd} F - G_d)/G_p|^2 \right] \Phi_{dd} ds \quad (37)$$

By following the same procedure as above, the optimum F is given by

$$F = \left[\sum_{j=1}^{q} |\lambda_j B_j/G_p|^2 N_d^* P_{pd}^* G_d \{\Phi_{dd}\}^+ / (Z_d P_{pd} \{V\}^-) \right]_+ / (Z_d^* P_{pd}^* \{\Phi_{dd}\}^+ \{V\}^+) \quad (38)$$

where the symbols have been defined in Equations (31) and (33). The optimum overall and feedback regulator transfer functions are given by

$$\hat{H} = (\hat{C/D}) = N_d P_{pd} \hat{F} \quad (39)$$

and

$$\hat{G}_c = (G_d' - P_{pd}\hat{F})/(G_m G_p P_{pd} \hat{F}) \quad (40)$$

Thus, $\hat{G}_c$ will be a stable controller so long as G_p is minimum phase. The minimum value of I is

$$\hat{I} = \frac{1}{2\pi i}\int_{-i\infty}^{+i\infty} \left[|N_d P_{pd} \hat{F}\, G_m|^2 + \sum_{j=1}^{q} \lambda_j^2 |B_j (N_d P_{pd} \hat{F} - G_d)/G_p|^2 \right] \Phi_{dd} ds \quad (41)$$

We next proceed to solve the most general case. We note that the method used in case II is inapplicable if G_p is NMP, and no acceptable internal feedback loop (not depending upon exact cancellation) can remove the NMP factors in G_p. However, the formulation of case I, with appropriate modifications, can be used if G_d and G_p are both stabilized, if necessary, by an internal feedback loop which does not depend upon exact cancellation of the RHP poles.

Case III: General Case

No restriction whatsoever is required on G_d and G_p. The approach used here is to convert this to the equivalent form in case I. This principle is best described with Figure 2. Figure 2(a) is a typical control system, where G_d and G_p are both unstable and nonminimum phase. From this is constructed Figure 2(b), in which a fictitious internal feedback loop is placed around G_p by using G_s which is a transfer function required to stabilize G_p and G_d; that is, the zeros of $1 + G_s G_p$ are in the LHP interior. Figures 2(a) and (b) are forced to be equivalent through the use of $\overline{G}_c$ in Figure 2(b) in place of G_c in Figure 2(a). Here, by equivalence we mean

$$(C/D)_a = (C/D)_b \text{ and } (M_j/D)_a = (M_j/D)_b \quad (42)$$

where the subscripts a and b refer to Figures 2(a) and (b) since only these terms are involved in the Lagrangian function I. It is obvious that

$$\overline{G}_c = G_c - G_s/G_m \quad (43)$$

It is also possible to construct Figure 2(c) which is equivalent to Figure 2(b); that is

$$(C/D)_b = (C/D)_c \text{ and } (M_j/D)_b = (M_j/D)_c \quad (44)$$

This is accomplished by associating the fictitious minor stabilizing loop with G_d and G_p and by introducing a necessary modification factor G for control effort. Hence

$$\overline{G}_p = G_p/(1 + G_p G_s), \quad (45)$$

$$\overline{G}_d = G_d/(1 + G_p G_s) \quad (46)$$

and

$$G = 1 + G_s/(G_m \overline{G}_c) \quad (47)$$

Note that the omission of G corresponds to solving a different problem, which fact has not been previously pointed out. In Figure 2(c), both $\overline{G}_d$ and $\overline{G}_p$ are stable; hence, the approach adopted in case I is directly applicable. Note that one can regard $\overline{G}_d$ and $\overline{G}_p$ in Figure 2(c) as equivalent to G_d and G_p in Figure 2(a). However, in Figure 2(c) the input to $\overline{G}_p$ is multiplied by B_jG to obtain M_j, while the input to G_p in Figure 2(a) is multiplied by B_j in order to obtain the same signal. We will develop the solutions for optimum overall transfer function for the system given by Figure 2(c) and obtain the corresponding $\hat{\overline{G}}_c$. Then, the optimum feedback regulator $\hat{G}_c$ for the original system given by Figure 2(a) is obtained by

$$\hat{G}_c = \hat{\overline{G}}_c + G_s/G_m \tag{48}$$

We now adopt the approach used in case I to Figure 2(c). The form of H which meets all the functional constraints is

$$H = C/D = \overline{G}_d + N_dN_pT \tag{49}$$

where N_d and N_p are NMP factors in G_d and G_p, respectively, as defined in Equation (17). $\overline{G}_d$ is a stabilized G_d, that is, $G_d/(1 + G_sG_p)$, and T is some arbitrary, but stable, function.

In terms of Equation (49), (E/D) and (M_j/D) can be written as

$$(E/D) = -(\overline{G}_d + N_dN_pT)G_m \tag{50}$$

$$M_j/D = (C/D) - \overline{G}_d)GB_j/\overline{G}_p$$
$$= B_j\,[N_dN_p(1/\overline{G}_p - G_s)T - G_s\overline{G}_d] \tag{51}$$

after some manipulations based upon Equations (43), (47), and (49).

The performance criterion I is given by

$$I = \overline{e^2(t)} + \sum_{j=1}^{q} \lambda_j{}^2\overline{m_j{}^2(t)}$$
$$= \frac{1}{2\pi i}\int_{-i\infty}^{+i\infty} \Big\{ |G_m(\overline{G}_d + N_dN_pT)|^2 + \sum_{j=1}^{q} |\lambda_jB_j\,[N_dN_p(1/\overline{G}_p - G_s)T - G_s\overline{G}_d]|^2 \Big\} \Phi_{dd}\,ds \tag{52}$$

By following the same procedure used in case I, the optimum T is given by

$$\hat{T} = -\left[N_d{}^*N_p{}^*\overline{G}_d \left\{ |G_m|^2 - G_s(1/\overline{G}_p{}^* - G_s{}^*) \sum_{j=1}^{q} |\lambda_jB_j|^2 \right\} \{\Phi_{dd}\}^+/(Z_dZ_p\,\{V\}^-) \right]_+ \Big/ (Z_d{}^*Z_p{}^*\,\{V\}^+\,\{\Phi_{dd}\}^+) \tag{53}$$

where

$$\overline{G}_d = G_d/(1 + G_pG_s),\quad \overline{G}_p = G_p/(1 + G_pG_s) \tag{54}$$

$$\Phi_{dd} = \{\Phi_{dd}\}^+\,\{\Phi_{dd}\}^-,\quad \{V\}^+\,\{V\}^-$$
$$= |G_m|^2 + \sum_{j=1}^{q} |\lambda_jB_j(1/\overline{G}_p - G_s)|^2$$

and Z_d and Z_p are the RHP zeros of G_d and G_p, respectively. Therefore, the optimum overall transfer function is given by

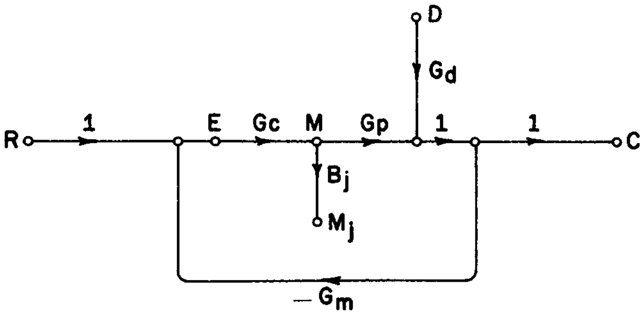

Fig. 1. A signal flow graph for a typical control system.

$$\hat{H} = (\hat{C/D}) = \overline{G}_d + N_dN_p\hat{T} \tag{55}$$

and the optimum $\overline{G}_c$ is given by

$$\hat{\overline{G}}_c = (\overline{G}_d - \hat{H})/(\hat{H}G_m\overline{G}_p)$$
$$= -(1 + G_pG_s)^2\;N_dN_p\hat{T}/(G_mG_p\,\{G_d + (1 + G_sG_p)N_dN_p\hat{T}\}) \tag{56}$$

and the optimum regulator to the original problem of Figure 2(a) is

$$\hat{G}_c = \hat{\overline{G}}_c + G_s/G_m$$
$$= [G_dG_pG_s + (1 + G_pG_s)\;N_dN_p\hat{T}]/[G_mG_p\,\{G_d + (1 + G_pG_s)\;N_dN_p\hat{T}\}] \tag{57}$$

It is clear that $\hat{T}$ given by Equation (53) should reduce to $\hat{Q}$ given by Equation (30), and Equation (54) to Equation (31), when G_d and G_p are stable, that is, $G_s = 0$, $\overline{G}_d = G_d$ and $\overline{G}_p = G_p$. This can be readily verified.

Thus, the most general solutions have been obtained in case III, which reduces to the solutions for stable case as treated in case I. When G_p is minimum phase, then the solutions given in case II must be same as those given in case III. One should note, however, the superior simplicity of the solutions given in case II, in so far as computations are concerned, to those given in case III. The operations involved in Equations (53) and (54) are considerably more than those indicated in Equation (38), since in Equations (53) and (54) we are dealing, in general, with higher-order terms. In addition, one must predetermine G_s in case III, which may also be time consuming. Since the choice of G_s is arbitrary, so long as the roots of the characteristic equation $1 + G_pG_s = 0$ all lie in the LHP interior, it is recommended that G_s be chosen judiciously so that computational effort be reduced in Equations (53) and (54).

EVALUATION OF THE LAGRANGIAN MULTIPLIERS AND THE RELATIONSHIP BETWEEN THE MEAN-SQUARE CONSTRAINTS AND INSTANTANEOUS CONSTRAINTS

We now proceed to evaluate the Lagrangian multipliers in order to complete the solutions developed previously. The constraints to be satisfied are

$$\overline{m^2{}_j(t)} = \Phi_{m_jm_j}(0)$$
$$= \frac{1}{2\pi i}\int_{-i\infty}^{+i\infty} \Phi_{m_jm_j}(s;\lambda_1,\ldots\ldots\lambda_q) \leq W_j{}^2,\quad j = 1,\ldots q \tag{58}$$

TABLE 1. PROBABILITY OF MEETING INSTANTANEOUS CONSTRAINT

$k_j = a_j/W_j$*	Gaussian distribution $Pr\ (\lvert m_j(t)\rvert \leq a_j) =$	Any zero-mean distribution $Pr\ (\lvert m_j(t)\rvert < a_j) >$
2	0.9545	0.750
2.5	0.9876	0.837
3.0	0.9973	0.889
3.5	0.9995	0.919
4.0	0.9999	0.937
5.0	0.9999	0.960

* Ratios of instantaneous constraint to root-mean-square constraint.

where

$$\Phi_{m_j m_j} = |B_j N_d \hat{Q}/G_p'|^2\, \Phi_{dd} \qquad \text{for case I}$$

$$\Phi_{m_j m_j} = |B_j (N_d P_{pd} \hat{F} - G_d)/G_p|^2\, \Phi_{dd} \qquad \text{for case II}$$

and

$$\Phi_{m_j m_j} = |B_j\, [N_d N_p (1/\overline{G}_p - G_s)\hat{T} - G_s \overline{G}_d]|^2\, \Phi_{dd} \qquad \text{for case III}$$

and $\hat{Q}$, $\hat{F}$, and $\hat{T}$ are given by Equations (30), (38), and (53), respectively. Equation (58), choosing the appropriate equality relationships, determines the values of λ_j, $j = 1, \ldots q$, which completes the solution. However, a general comment on this type of constraint is in order. The constraints given by Equation (58) are mean-square constraints, whereas physical problems usually call for instantaneous constraints of the form $|m_j(t)|^2 \leq a_j^2$. Since the disturbance is a random input, it is impossible to satisfy an instantaneous constraint for all time, for the input is not known precisely at all time. One, therefore, requires the control system to satisfy the instantaneous constraint with some probability less than 1. The relationship between the mean-square values W_j^2 and the instantaneous values a_j^2 for a Gaussian distribution with zero mean is given by

$$Pr\ (|m_j(t)| \leq a_j) = \frac{1}{\sqrt{2}\ \sigma_{m_j}} \int_{-a_j}^{+a_j} \exp\ (-x^2/2\sigma^2_{m_j})\, dx = \text{erf}\ [a_j/(\sqrt{2}\ \sigma_{m_j})] \qquad (59)$$

For any zero mean distribution, the relationship is given by the Chebyshev inequality (*13*), that is

$$Pr\ (|m_j(t)| < a_j) = 1 - Pr\ (|m_j(t)| \geq a_j) \geq 1 - (W_j/a_j)^2 \qquad (60)$$

According to Table 1, if the root mean-squared value is taken to be, for instance, one-quarter of the instantaneous constraint value, it is guaranteed that at least 93.7% of the time the instantaneous constraint would be satisfied regardless of the type of distribution, while for Gaussian inputs at least 99.99% of the time the physical (instantaneous) constraint would be satisfied. Therefore, we now have a means to satisfy the instantaneous physical constraint for any desired fraction of time, and Equation (58) can be written as

$$\frac{1}{2\pi i} \int_{-i\infty}^{+i\infty} \Phi_{m_j m_j}\ (s;\ \lambda_1, \ldots \lambda_q)\, ds = (a_j/k_j)^2 \qquad (61)$$

where $k_j = a_j/W_j$ is selected by referring to Table 1. There are other ways of handling this saturation type of constraint (*14*).

CONCLUDING REMARKS

Three classes of problems have been discussed. Case I, which refers to systems with stable G_d and G_p, is of the form given in the literature, except that a modification is necessary for physical realizability if G_d is NMP. Case II deals with minimum-phase G_p and leads to an extension of a procedure recommended by Chang to reduce computational effort, which takes into account NMP G_d and the possibility that G_p contains unstable poles not in G_d. An alternative procedure would be to prestabilize G_p and G_d by inner-loop feedback, which is the usual recommendation. However, a direct application of inner-loop stabilization, followed by the formulation, due to Chang, of Equation (17), leads to the solution of a different problem than the original, due to the form of the inequality constraint. This point is not explicitly stated in the literature and hence is further expanded in case III. This shows how to modify the prestabilized system so as to retain all features of the original problem and hence represents a general method applicable to all systems. If prestabilization is unnecessary, the method reduces to that of case I. For minimum-phase plants, the method of case II is still preferable, since the algebra is less involved.

Some remarks seem in order on the form of optimum controller resulting from the general solutions, Equations (35), (40), and (57). It is clear from Equation (40) that the optimum controller $\hat{G}_c$ may turn out to be unstable if $\hat{F}$ admits any RHP zero, since any zero of $\hat{F}$ is a pole of $\hat{G}$. This situation may arise owing to the particular process involved in Equation (38). When all functions involved are rational in s, the process $[\ ]_+$ in Equation (38) is equivalent to decomposing into partial fractions and retaining only those terms containing poles in the LHP interior. Hence, if G_d is stable and higher order in s than G_p, the above process gives a linear combination of stable first-order terms. By combining terms, the numerator is a linear combination of products of first-order terms. For example

$$[\ ]_+ = \sum_{j=1}^{3} \frac{c_j}{s + b_j} = \frac{c_1(s + b_2)(s + b_3) + c_2(s + b_1)(s + b_3) + c_3(s + b_1)(s + b_2)}{\prod_{j=1}^{3} (s + b_j)} \qquad (62)$$

so that the numerator polynomial may contain one or more RHP zero(s). Hence, the probability of $\hat{F}$ having a RHP zero may be high for open-loop stable systems when $G_d(s)$ is higher order than $G_p(s)$. Physically, such a situation arises in systems in which the disturbance enters far ahead of the control actions, as in a series of reactors in which the disturbance enters into the first reactor while the control action is in the last reactor.

Unstable controllers are theoretically possible, since all of the functional constraints are met, and the overall system is stable. Frequently, however, its implementation

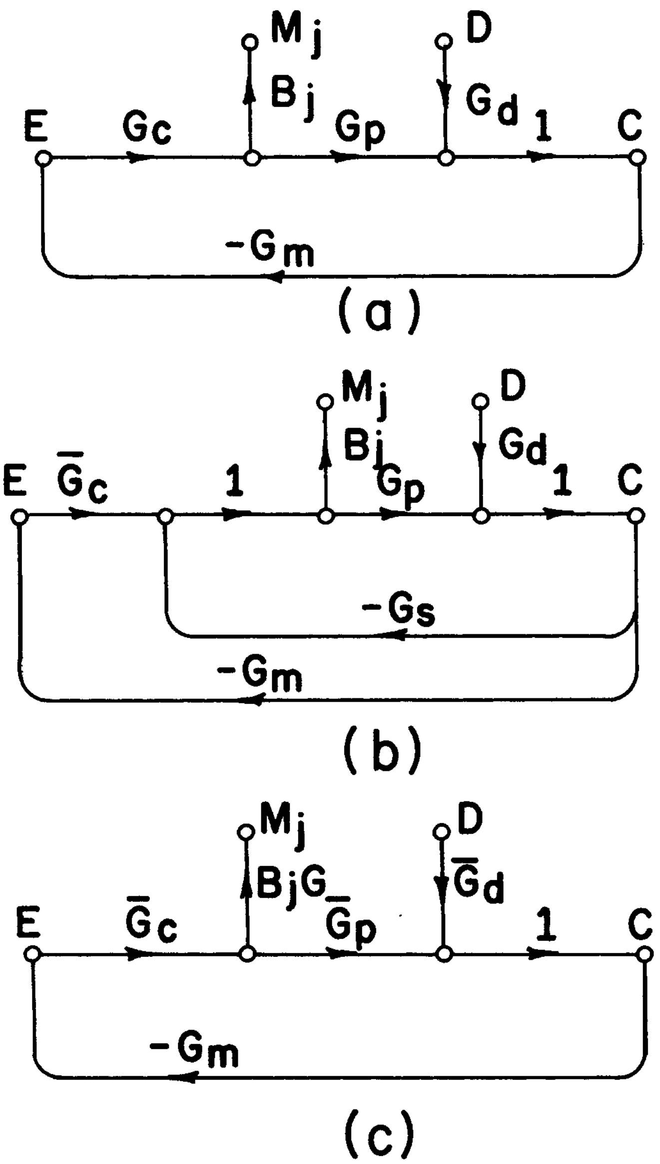

Fig. 2. Unstable and nonminimum phase G_d and G_p.

may be impractical, and the relative stability and safety of the resulting system will be marginal. Note, however, that, theoretically speaking, regulation of a stable plant by an unstable controller in a closed-loop system is equivalent to regulating an unstable plant by a stable controller, a procedure often employed. This follows from the fact that G_p and G_c are symmetrical in the closed-loop equation $G_cG_p/(1 + G_cG_p)$. It is also clear that since $\hat{G}_c$ admits at most the RHP zero(s) of $\hat{F}$ as RHP poles, C/R is always stable. The same situation may arise, due to the RHP zero(s) of $(G_d' + \hat{Q})$ in Equation (35) and of $\{G_d' + (1 + G_sG_p)N_p\hat{P}\}$ in Equation (56) for cases I and III, respectively.

Thus, for the case where the order of G_d is higher than that of G_p, it is possible for the optimum feedback controller to be unstable. Some approximation techniques over a limited bandwidth, perhaps in the form of the conventional proportional-integral derivative (PID) controller, may be desirable for practical implementation. One approach would be to approximate a high-order $G_d(s)$ by a lower-order transfer function before the synthesis procedure is employed. Note, however, that one cannot thereby eliminate an RHP zero of the original G_d, since this would be equivalent to cancellation and would lead to difficulties in stability and realizability. Another may be a direct approximation of the unstable controller over a small frequency range of importance by a known stable controller.

In part II, the application of these methods to unstable and/or nonminimum phase stirred-tank chemical reactors will be demonstrated and the practical implications assessed.

It should also be pointed out that the synthesis procedure presented here can be readily extended to design problems associated with deterministic inputs. It is necessary to replace mean-square error by integral-square error, mean-square constraints by integral-square constraints, and the disturbance autocorrelation function by the autotranslation function (9), which is defined as

$$\phi_{dd}(\tau) = \int_0^\infty d(t)\ d(t + \tau)\ dt$$

ACKNOWLEDGMENT

Financial support by the Walter P. Murphy Foundation, the Union Carbide Company, the National Science Foundation (GK-1127) and the Ford Foundation is greatly appreciated.

NOTATION

$A(p)$ = differential operator in Equation (1)
a_i = constants describing system in Equation (1)
B_j, $B_j(s)$ = Laplace transform of linear operator
C, $C(s)$ = Laplace transform of $c(t)$
$c(t)$ = output
D, $D(s)$ = Laplace transform of $d(t)$
$d(t)$ = random disturbance or load
E, $E(s)$ = Laplace transform of $e(t)$
$e(t)$ = error
$\overline{e^2(t)}$ = mean-square error
erf = error function
F, $F(s)$ = portion of overall function given in Equation (36)
$\hat{F}$, $\hat{F}(s)$ = optimum F
G, $G(s)$ = Laplace transform operator defined in Equation (47)
G_c, $G_c(s)$ = controller transfer function
$\hat{G}_c$ = optimum controller transfer function
$\overline{G}_c$ = controller transfer function for Figures 2(b) and (c)
$\hat{\overline{G}}_c$ = optimum $\overline{G}_c$
G_d, $G_d(s)$ = disturbance transfer function $C(s)/D(s)$
$\overline{G}_d$ = stabilized G_d as defined in Equation (54)
G_d' = G_d without NMP factor as defined in Equation (12)
G_m = transfer function of measuring element
G_p, $G_p(s)$ = plant transfer function $C(s)/M(s)$
$\overline{G}_p$ = stabilized G_p as defined in Equation (54)
G_p' = G_p without NMP factor
G_s = transfer function of stabilizing loop
H = overall transfer function C/D
$\hat{H}$ = optimum H
I = Lagrangian function as defined in Equation (4)
i = imaginary number $\sqrt{-1}$
j = summing index for constraints
K, $K(s)$ = arbitrary function of the same class as F
k_j = constants, ratios of instantaneous to root-mean-square constraints

$L(p)$ = differential operator as defined in Equation (1)
l_k = constants in system describing equation
$M, M(s)$ = Laplace transform of $m(t)$
$m(t)$ = manipulated, or control variable
$M_j, M_j(s)$ = Laplace transform of $m_j(t)$
$m_j(t)$ = signal obtained from $m(t)$
$\overline{m_j^2(t)}$ = mean square value of $m_j(t)$
N_d = NMP factors in G_d as defined in Equation (17)
N_p = NMP factors in G_p as defined in Equation (17)
P_{pd} = right-half plane poles of G_p not present as poles in G_d
p = poles
$Pr(\ )$ = probability of ( )
q = number of constraints
Q = portion of overall transfer function as defined in Equation (11)
$\hat{Q}$ = optimum Q
$R, R(s)$ = Laplace transform of set point
T = portion of overall transfer function as defined in Equation (49)
$\hat{T}$ = optimum T
t = time
$U(p)$ = differential operator as defined in Equation (1)
u_k = constants in Equation (1)
$\{V\}^+$ = portion of V having poles and zeros in the LHP
$\{V\}^-$ = portion of V having poles and zeros in the RHP
W_j^2 = mean-square constraint values of $m_j(t)$
z = zeros
Z_d = RHP zeros of G_d, $\pi_r(s + z_r)$, $z_r < 0$
Z_p = RHP zeros of G_p, $\pi_k(s + z_k)$, $z_k < 0$
$[\ ]_+$ = operation defined in Equation (33)
$[\ ]_-$ = operation defined in Equation (33)
$(\ )^*$ = complex conjugate

Greek Letters

ϵ = real and small constant
λ_j = Lagrangian multipliers
$\sigma^2_{m_j}$ = variance of $m_j(t)$
τ_d = dead time associated with G_d
τ_p = dead time associated with G_p
$\phi_{dd}, \phi_{dd}(\tau)$ = autocorrelation function of $d(t)$
$\phi_{ee}, \phi_{ee}(\tau)$ = autocorrelation function of $e(t)$
$\phi_{mm}, \phi_{mm}(\tau)$ = autocorrelation function of $m(t)$
$\phi_{m_jm_j}, \phi_{m_jm_j}(\tau)$ = autocorrelation function of $m_j(t)$
$\Phi_{dd}, \Phi_{dd}(s)$ = spectral density of $d(t)$
$\Phi_{ee}, \Phi_{ee}(s)$ = spectral density of $e(t)$
$\Phi_{mm}, \Phi_{mm}(s)$ = spectral density of $m(t)$
$\Phi_{m_jm_j}, \Phi_{m_jm_j}(s)$ = spectral density of $m_j(t)$
$\{\Phi_{dd}\}^-$ = portion of Φ_{dd} having poles and zeros in the RHP
$\{\Phi_{dd}\}^+$ = portion of Φ_d having poles and zeros in the LHP
ω = frequency

LITERATURE CITED

1. Newton, G. C., Jr., *J. Franklin Inst.*, **254**, 391 (1952).
2. Newton, G. C., Jr., L. A. Gould, and J. F. Kaiser, "Analytical Design of Linear Feedback Controls," pp. 34-38, 160-214, Wiley, New York (1957).
3. Chang, S. S. L., "Synthesis of Optimum Control Systems," pp. 63-99, McGraw-Hill, New York (1961).
4. Wiener, N., "Extrapolation, Interpolation and Smoothing of Stationary Time Series," pp. 84-86, Wiley, New York (1949).
5. Bendat, J. S., "Principles and Applications of Random Noise Theory," pp. 144-180, Wiley, New York (1958).
6. Horowitz, I. M., "Synthesis of Feedback Systems," pp. 495-524, Academic Press, New York (1963).
7. Lee, Y. W., "Statistical Theory of Communication," pp. 355-395, Wiley, New York (1960).
8. Seifert, W. W., and C. W. Steeg, "Control Systems Engineering," pp. 625-640, McGraw-Hill, New York (1960).
9. Merriam, C. W., III, "Optimization Theory and the Design of Feedback Control Systems," pp. 31-62, McGraw-Hill, New York (1964).
10. Luecke, R. H., and M. L. McGuire, *AIChE J.*, **14**, 173 (1968).
11. Hildebrand, F. H., "Methods of Applied Mathematics," pp. 121-168, Prentice-Hall, Englewood Cliffs, N. J. (1952).
12. Papoulis, A., "The Fourier Integral and its Applications," pp. 240-264, McGraw-Hill, New York (1962).
13. Brunk, H. C., "An Introduction to Mathematical Statistics," pp. 122-125, Ginn and Company, New York (1960).
14. Pugachev, V. S., "Proceedings Second IFAC Congress, Basle, Switzerland," Vol. 2, Pt. 2,1, Butterworths, England (1963).

Part II. Application to a Stirred Tank Reactor

The Wiener-Hopf procedure for synthesis of optimum constrained linear feedback regulators has been extended in part I to all linear time-invariant lumped parameter systems. The solution is here applied to the control of the output concentration of an exothermic, stirred tank reactor operating close to an unstable steady state, by constrained manipulation of a cooling water flow rate, in the presence of a randomly fluctuating inlet concentration.

When the spectral density of disturbance is given (for example, white noise through a first-order time delay, or a series of randomly alternating steps), the optimum controller has three modes: proportional, derivative, and integral with minor feedback. The responses of the nonlinear reactor and the linearized reactor control systems to a series of alternating deterministic step inputs and Gaussian distributed inputs are simulated, and a sensitivity study of the linearized system with respect to variations in process, controller, and disturbance parameters is made in order to demonstrate the feasibility of the method.

The identification of chemical systems from responses to random inputs has been examined by several authors (*1* to *5*). More recently, the synthesis of optimum controllers by Wiener-Hopf methods for chemical systems subject to random disturbances has been treated (*6, 7*).

In part I (*8*) of this two-part series, it was shown how to extend the Wiener-Hopf synthesis of optimum constrained linear feedback regulators to the general single-input, single-output linear, time-invariant system, with or without dead time. In this paper, the solution is applied to the control of the output concentration of an exothermic, stirred tank reactor in the presence of a randomly fluctuating inlet concentration. The reactor is to be operated close to its unstable steady state by manipulation of a constrained cooling water flow rate. Within a small operating region, the system dynamics may be represented by an unstable plant transfer function and an unstable, nonminimum phase disturbance transfer function. Therefore, the solution de-