Automatica, Vol. 12, pp. 457–465. Pergamon Press, 1976. Printed in Great Britain

The Internal Model Principle of Control Theory*

B. A. FRANCIS† and W. M. WONHAM‡

In multivariable servomechanisms designed to accommodate parameter uncertainty, the controller must have special qualitative structural features which may be derived for linear and weakly nonlinear systems.

Summary—The classical regulator problem is posed in the context of linear, time-invariant, finite-dimensional systems with deterministic disturbance and reference signals. Control action is generated by a compensator which is required to provide closed loop stability and output regulation in the face of small variations in certain system parameters. It is shown, using the geometric approach, that such a structurally stable synthesis must utilize feedback of the regulated variable, and incorporate in the feedback path a suitably reduplicated model of the dynamic structure of the disturbance and reference signals. The necessity of this control structure constitutes the Internal Model Principle. It is shown that, in the frequency domain, the purpose of the internal model is to supply closed loop transmission zeros which cancel the unstable poles of the disturbance and reference signals. Finally, the Internal Model Principle is extended to weakly nonlinear systems subjected to step disturbances and reference signals.

1. INTRODUCTION

A CLASSICAL problem of control theory is that of synthesizing controllers which regulate systems despite uncertainty in plant and controller parameters. In general terms, we are concerned in this paper with the regulator problem illustrated schematically in Fig. 1. A plant subjected to external disturbances is controlled by a compensator processing certain plant measurements, a reference command signal r, and possibly the feedforward disturbance signal.

The purpose of the compensator is twofold. First, it is to provide closed loop stability. Second, it is to regulate a variable z which is a given function of the plant output c and the reference signal r; typically z may be the tracking error $r-c$. A plant-compensator combination with these two properties is termed a *synthesis*, and a synthesis is called *structurally stable* if these two properties are preserved when certain system parameters are perturbed.

In the context of linear multivariable systems, synthesis procedures for designing controllers providing structural stability have been given by Davison[2], [3], Staats and Pearson[4], and Wonham[1]. In [5] we considered the converse problem by asking: What controller structure is necessary for structural stability? The results of [5] were interpreted in [6] in terms of closed loop transmission zeros. The purpose of this paper is to summarize and amplify the results of [5] and [6], and then to extend some of them to weakly nonlinear systems with constant reference and disturbance signals.

Notation

R (resp. **C**) denotes the field of real (resp. complex) numbers. $\mathbf{C}^-$ (resp. $\mathbf{C}^+$) is the open left-half (resp. closed right-half) complex plane. $\|\cdot\|$ is the Euclidean norm on $\mathbf{R}^n$.

If $A: \mathscr{X} \to \mathscr{X}$ is a linear map and $\mathscr{V} \subset \mathscr{X}$ a subspace, Im A (resp. Ker A) is the image (resp. kernel) of A, $\sigma(A)$ is its complex spectrum, $d(\mathscr{V})$ is the dimension of $\mathscr{V}$, and $\langle A|\mathscr{V}\rangle$ is the subspace $\sum_{i\geq 0} A^i\mathscr{V}$. If $C: \mathscr{X} \to \mathscr{Y}$ is a map, $C|\mathscr{V}$ is the restriction of C to $\mathscr{V}$.

Let L be the space of linear maps $\mathbf{R}^n \to \mathbf{R}^m$ endowed with its usual norm topology. If $f: \mathbf{R}^n \to \mathbf{R}^m$ is a differentiable function, let f': $\mathbf{R}^n \to L$ denote the differential of f: $f'(x)$ is represented, in the standard basis, by the Jacobian matrix of f at $x \in \mathbf{R}^n$. We say f is of class C^1 if f' is continuous.

*Received 1 October 1975; revised 5 March 1976. The original version of this paper was presented at the *6th IFAC Congress* which was held in Boston/Cambridge, MA U.S.A. during August 1975. The published Proceedings of this IFAC Meeting may be ordered from: ISA—Instrument Society of America, 400 Stanwix Street, Pittsburgh, PA 15222, U.S.A. This paper was recommended for publication in revised form by associate editor H. Kwakernaak.

†Department of Electrical Engineering and Computer Science, University of California, Berkeley, CA 94720, U.S.A.

‡Department of Electrical Engineering, University of Toronto, Canada M5S 1A4.

This research was partially supported by the National Research Council of Canada, Grant No. A-7399.

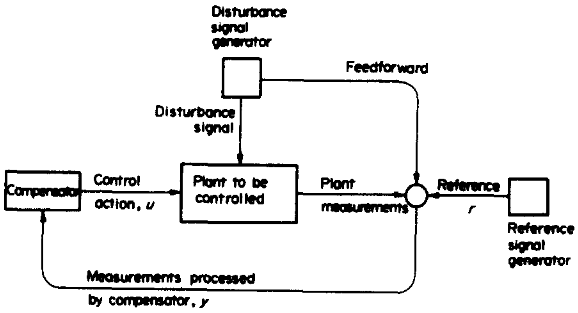

FIG. 1. The system considered in the regulator problem.

If $n \geq 1$ is an integer, $\mathbf{n}$ is the set $\{1, \ldots, n\}$. Equality by definition is denoted by $:=$.

2. THE INTERNAL MODEL PRINCIPLE FOR LINEAR MULTIVARIABLE SYSTEMS

We begin by posing the above classical regulator problem in the context of linear, time-invariant, finite-dimensional, multivariable systems with deterministic exogenous signals. Consider the system described by the equations

$$\dot{x}_1 = A_1x_1 + A_3x_2 + B_1u \qquad \text{(1a)}$$

$$\dot{x}_2 = A_2x_2 \qquad \text{(1b)}$$

$$z = D_1x_1 + D_2x_2 \qquad \text{(1c)}$$

$$y = C_1x_1 + C_2x_2 \qquad \text{(1d)}$$

$$\dot{x}_c = A_cx_c + B_cy \qquad \text{(2a)}$$

$$u = F_cx_c + Fy. \qquad \text{(2b)}$$

Here x_1 is the state vector of the plant and u is the vector of control inputs, x_2 represents both disturbance and reference signals, z is the output to be regulated, y the output available for measurement, and x_c is the state vector of the compensator. Equation (1b) defines the class of exogenous signals to be processed (e.g. steps, ramps), and equation (2) describes the operation of the compensator.

The vectors u, x_1, x_2, x_c, y, z belong to fixed real finite-dimensional linear spaces

$$\mathscr{U}, \mathscr{X}_1, \mathscr{X}_2, \mathscr{X}_c, \mathscr{Y}, \mathscr{Z} \qquad (3)$$

respectively. The time-invariant linear maps A_1, $A_3, \ldots$ in (1) and (2) are defined on the appropriate spaces. The vector spaces in (3) are assumed to have fixed bases. The maps A_1, $A_3, \ldots$ then have matrix representations referred to these bases; these matrices will be denoted by the same letter as the corresponding maps. The signal flow graph of the system is shown in Fig. 2.

The state space of the closed loop is

$$\mathscr{X}_L := \mathscr{X}_1 \oplus \mathscr{X}_c,$$

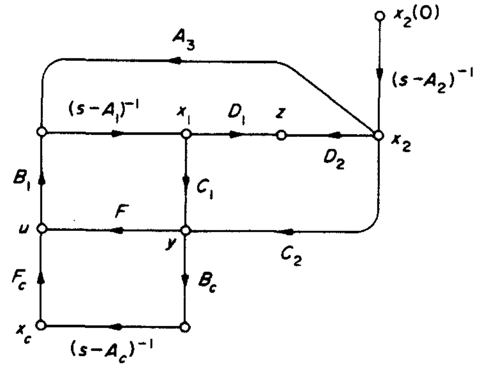

FIG. 2. Overall system signal flow graph.

and the closed loop state vector is

$$x_L := \begin{bmatrix} x_1 \\ x_c \end{bmatrix} \in \mathscr{X}_L.$$

By combining (1a), (1d), and (2), and by defining

$$A_L = \begin{bmatrix} A_1 + B_1FC_1 & B_1F_c \\ B_cC_1 & A_c \end{bmatrix}, \quad B = \begin{bmatrix} A_3 + B_1FC_2 \\ B_cC_2 \end{bmatrix},$$

we see that the closed loop is described by

$$\dot{x}_L = A_Lx_L + B_Lx_2. \qquad (4)$$

In addition define

$$D_L = [D_1 \quad 0] : \mathscr{X}_1 \oplus \mathscr{X}_c \to Z$$

so that the output to be regulated is

$$z = D_Lx_L + D_2x_2. \qquad (5)$$

The composite system is now described by (1b), (4), and (5).

As stated earlier, the purpose of the compensator is to stabilize the closed loop and regulate the output. *Closed loop stability* means that A_L is stable; that is

$$\sigma(A_L) \subset \mathbf{C}^-.$$

Output regulation means that $z(t) \to 0$ as $t \to \infty$ for all $x_L(0)$ and $x_2(0)$. Without loss of generality we make the following standing assumptions:

$$\sigma(A_2) \subset \mathbf{C}^+$$

$$\operatorname{Im} C_1 + \operatorname{Im} C_2 = \mathscr{Y},$$

$$\operatorname{Im} D_1 = \mathscr{Z}.$$

See [5] for justification of these assumptions.

Now let $\mathscr{P}$ be a *data point* in $\mathbf{R}^n$; that is, $\mathscr{P}$ is a list of n real numbers selected from among the elements of the matrices in (1) and (2). For example $\mathcal{P} := (A_1, B_1)$ is a data point in $\mathbf{R}^n$ where $n := d(\mathscr{X}_1)^2 + d(\mathscr{X}_1) \times d(\mathscr{U})$. Let Π be a property of points in $\mathbf{R}^n$, and assign to $\mathbf{R}^n$ the usual topology. Then we say that Π is *stable* at $\mathcal{P}$ if Π holds throughout some open neighborhood (nbhd) of $\mathcal{P}$, we say that *the synthesis is structurally stable at* $\mathcal{P}$. If A_L is stable and its elements are perturbed a little, then of course the resulting matrix is still stable: closed loop stability is a stable property. Hence structural stability at $\mathcal{P}$ is equivalent to the two conditions:

(i) closed loop stability holds at $\mathcal{P}$

(ii) output regulation is a property which is stable at $\mathcal{P}$.

As stated in the Introduction, our object is to describe the controller structure which is necessary for structural stability. To this end, we introduce the concepts readability and internal model.

We say that *z is readable from y* if there exists a map $Q: \mathscr{Y} \to \mathscr{Z}$ such that $z = Qy$, which is to say

$$D_1 = QC_1, \quad D_2 = QC_2.$$

If this is the case, then $\mathscr{Z}$ can be embedded in $\mathscr{Y}$: write

$$\mathscr{Y} = \mathscr{W} \oplus \mathscr{Z} \tag{6}$$

for a suitable linear space $\mathscr{W}$. Then

$$C_1 = \begin{bmatrix} E_1 \\ D_1 \end{bmatrix}, \quad C_2 = \begin{bmatrix} E_2 \\ D_2 \end{bmatrix} \tag{7}$$

for suitable maps $E_i : \mathscr{X}_i \to \mathscr{W}$ $(i = 1, 2)$. Defining $w = E_1x_1 + E_2x_2$, we have

$$y = \begin{bmatrix} w \\ z \end{bmatrix} \in \mathscr{W} \oplus \mathscr{Z}.$$

Finally, corresponding to (6) the maps F and B_c have the representations

$$F = [F_w \quad F_z], \quad B_c = [B_{cw} \quad B_{cz}]. \tag{8}$$

We say that *a map* $A : \mathscr{X} \to \mathscr{X}$ *incorporates an internal model of* A_2 if the minimal polynomial (m.p.) of A_2 divides at least $d(Z)$ invariant factors (i.f.) of A. Thus an internal model is an ℓ-fold reduplication in A of the maximal cyclic component of A_2, where $\ell \geq d(Z)$ = the number of independent outputs to be regulated.

To say that *the compensator incorporates an internal model of* A_2 means the following:

(i) There is an A_c-invariant subspace $\mathscr{R}_c \subset \mathscr{X}_c$ such that the map $\bar{A}_c$ induced by A_c in $\bar{\mathscr{X}}_c := \mathscr{X}_c/\mathscr{R}_c$ incorporates an internal model of A_2.

Now assume that, in addition to (i),

$$\text{(ii) Ker } F_c \cap \text{Ker}\,(A_c - \lambda) = 0, \quad \lambda \in \sigma(A_2).$$

Then we say that *the internal model is observable by u*. In view of Lemma 8.1 of [1], (ii) says that the A_2-modes of A_c are observable by F_c.

Finally, suppose z is readable from y and (i) holds. Adopt the representations in (8), and let $P_c : \mathscr{X}_c \to \bar{\mathscr{X}}_c$ be the canonical projection. We say that *the internal model is controllable by z* if

(iii) $\text{Im } B_{cw} \subset \mathscr{R}_c$

and

(iv) $\bar{\mathscr{X}}_c = \text{Im}\,(\bar{A}_c - \lambda) + \text{Im}\,\bar{B}_{cz}, \quad \lambda \in \sigma(A_2)$

where $\bar{B}_{cz} := P_cB_{cz}$. Condition (iv) says that the A_2—modes of $\bar{A}_c$ are controllable by $\bar{B}_{cz}$.

To see what these concepts mean in terms of signal flow, assume (i) holds, and write

$$\mathscr{X}_c = \mathscr{R}_c \oplus \mathscr{X}_{c2}$$

for any complement $\mathscr{X}_{c2}$. Corresponding to this decomposition write

$$A_c = \begin{bmatrix} A_{c1} & A_{c3} \\ 0 & A_{c2} \end{bmatrix}, \quad B_{cw} = \begin{bmatrix} B_{cw1} \\ B_{cw2} \end{bmatrix} \tag{9a}$$

$$B_{cz} = \begin{bmatrix} B_{cz1} \\ B_{cz2} \end{bmatrix}, \quad F_c = [F_{c1} \quad F_{c2}]. \tag{9b}$$

Then (i) implies that A_{c2} incorporates an internal model of A_2, (iii) is equivalent to $B_{cw2} = 0$, and (iv) is equivalent to

$$\mathscr{X}_{c2} = \text{Im}\,(A_{c2} - \lambda) + \text{Im}\,B_{cz2}, \quad \lambda \in \sigma(A_2).$$

Applying (7)–(9) to Fig. 2 yields Fig. 3.

Briefly, the main result of [5] is that a synthesis is structurally stable only if z is readable from y, and the compensator incorporates an internal model of A_2 which is controllable by z and observable by u. This result is established by the following two propositions and two theorems.

Define $z_1 := D_1x_1$ and $y_1 := C_1x_1$.

Proposition 1

The synthesis is structurally stable at A_3 *only if* z_1 *is readable from* y_1. (Proof: See [5], Proposition 2). Considering structural stability at A_3, we may assume, by Proposition 1, that there exists $Q: \mathscr{Y} \to \mathscr{Z}$ such that $D_1 = QC_1$. Hence $d(\mathscr{Y}) \geq d(\mathscr{Z})$, and so $\mathscr{Z}$ can be embedded in $\mathscr{Y}$: write

$$\mathscr{Y} = \mathscr{Y}_1 \oplus \mathscr{Z}$$

for any complement $\mathscr{Y}_1$.

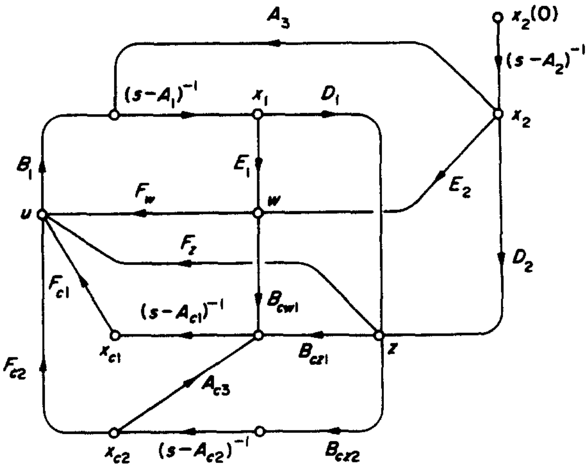

FIG. 3. The canonical synthesis: z is readable from y, the compensator incorporates an internal model of A_2 which is controllable by z and observable by u.

Theorem 1 (*Necessity of readability*)

The synthesis is structurally stable at $(A_3, B_c|\mathcal{X})$ *only if* z *is readable from* y. (Proof: See [5], Theorem 1). We suppose now that z is readable from y, and hence adopt the representations (7) and (8). For the main result on the internal model we shall assume that

$$\operatorname{Im} B_{cw} \subset \langle A_c | B_{cw} E_1 \operatorname{Ker} D_1 \rangle. \qquad (10)$$

Intuitively, this technical condition says that information sent from w and processed by the compensator pertains only to the plant and is unavailable from z. That (10) is necessary for structural stability is established by:

Proposition 2

Assume that z *is readable from* y. *If* (10) *does not hold then the synthesis is not structurally stable at* (A_3, B_{cw}). (Proof: See [5], Proposition 3).

Theorem 2 (*Necessity of the internal model*)

Suppose z *is readable from* y *and* (10) *holds. The synthesis is structurally stable at* A_3 *only if the compensator incorporates an internal model of* A_2 *which is controllable by* z *and observable by* u. (Proof: See [5], Theorem 2). In drawing conclusions from Theorem 2, it is important to note that if some of the elements of A_3 are not subject to variation then it may be that a full internal model is not required: a smaller class of parameter variations in general requires less reduplication in the internal model.

Intuitively, on the basis of information about z, the internal model injects into the closed loop signals which asymptotically counterbalance the exogenous signals. We remark that plausibility arguments in support of an 'internal model' idea have been presented by Kelley [7] and Conant and Ashby [8]. In addition internal models have played a rôle in theories of visual perception (Gregory [9]) and brain functioning (Oatley [10]).

We can see the function of the internal model more concretely in the frequency domain. For this we assume for simplicity that $y = z$; that is

$$\mathcal{Y} = \mathcal{Z}, \quad C_1 = D_1, \quad C_2 = D_2.$$

Let $G_L(s)$ be the closed loop transfer matrix,

$$G_L(s) := D_L(s - A_L)^{-1} B_L + D_2,$$

and let its McMillan form be

$$\left[\begin{array}{cccc|c} \frac{\phi_1}{\psi_1} & & & & \\ & \cdot & & & \\ & & \cdot & & 0 \\ & & & \cdot & \\ & & & \frac{\phi_r}{\psi_r} & \\ \hline & & 0 & & 0 \end{array}\right]$$

where $r := \operatorname{rank} G_L$, $\phi_i(s)$ and $\psi_i(s)$ are comprime polynomials ($i \in \mathbf{r}$), ϕ_i divides ϕ_{i+1}, and ψ_{i+1} divides ψ_i.

The complex roots, counting multiplicities, of the polynomials $\phi_i(s)$ ($i \in \mathbf{r}$) are the closed loop transmission zeros. The roots of the m.p. of A_2 are the poles of the exogenous signal. The purpose of the internal model is to supply right half-plane closed loop transmission zeros to cancel the unstable poles of the exogenous signal. This fact is contained in:

Theorem 3

Assume that $y = z$ *and the closed loop is stable, and let* $\mathcal{P} = (A_1, A_3, B_1)$. *The following are equivalent.* (a) *Output regulation is a property which is stable at* $\mathcal{P}$; (b) A_c *incorporates an internal model of* A_2; (c) *The m.p. of* A_2 *divides* ϕ_i (*all* $i \in \mathbf{r}$), *and this property is stable at* $\mathcal{P}$. (Proof: See [6], Theorem 2.)

Our major conclusion of this section can be summarized as: *The Internal Model Principle*: A regulator synthesis is structurally stable only if the controller utilizes feedback of the regulated variable, and incorporates in the feedback path a suitably reduplicated model of the dynamic structure of the exogenous signals which the regulator is required to process.

3. ON THE SUFFICIENCY OF READABILITY AND THE INTERNAL MODEL

Theorem 2 considered parameter variations only in A_3, but in fact a synthesis in which the compensator incorporates an internal model of A_2 can accommodate a much larger class of parameter variations. We have:

Theorem 4

Suppose z is readable from y, the closed loop is stable, and the compensator incorporates an internal model of A_2 which is controllable by z and observable by u. Then the synthesis is structurally stable at the data point.

$$\mathcal{P} = (A_1, A_3, B_1, F_{c1}, F_{c2}, F_w, F_z, A_{c1}, A_{c3}, B_{cw1}, B_{cz1}, B_{cz2}).$$

(Proof: See Appendix).

The only part of the compensator not included in the data point $\mathcal{P}$ is A_{c2}, that part containing the internal model of A_2. We shall now discuss what happens when A_{c2} is allowed to vary.

Suppose first that the exogenous signals are step functions of time; that is, assume $A_2 = 0$. From (1b), (4), and (5) the system equations are

$$\dot{x}_L = A_L x_L + B_L x_2 \quad (11a)$$

$$\dot{x}_2 = 0 \quad (11b)$$

$$z = D_L x_L + + D_2 x_2. \quad (11c)$$

Since $A_2 = 0$, A_{c2} incorporates an internal model of A_2 if and only if

$$d(\text{Ker } A_{c2}) \geq d(Z). \quad (12)$$

Under the hypotheses of Theorem 4, A_L is stable, and the steady state solution of (11) is

$$z_\infty := \lim_{t\to\infty} z(t) = (-D_L A_L^{-1} B_L + D_2)x_2(0). \quad (13)$$

For the nominal data and any $x_2(0)$, $z_\infty = 0$; i.e. output regulation holds.

Now fix $x_2(0) \neq 0$ and a nbhd of A_{c2} throughout which A_L remains stable. From (13) we see that, for A_{c2} in this nbhd, z_∞ is a rational function of the elements of A_{c2}. Thus, defined on this nbhd, z_∞ is a continuous function of the elements of A_{c2}. Hence if A_{c2} is perturbed slightly so that (12) no longer holds, then (generally) z_∞ is perturbed slightly from 0; that is, there results a small steady state offset error.

We next consider the case, illustrated by the following example, where the exogenous signals are not bounded time functions.

Example

In Fig. 4 is shown a synthesis in which the output of a first order plant is required to track a ramp reference signal. The system equations are

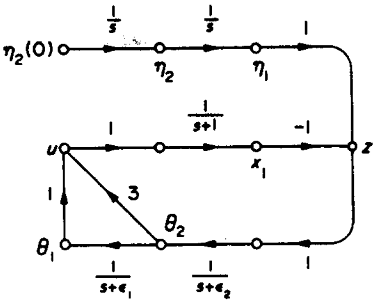

FIG. 4. Illustrative example.

$$\dot{x}_1 = -x_1 + u$$
$$\dot{\eta}_1 = \eta_2$$
$$\dot{\eta}_2 = 0$$
$$y = z = \eta_1 - x_1$$

$$\dot{\theta}_1 = -\epsilon_1\theta_1 + \theta_2$$
$$\dot{\theta}_2 = -\epsilon_2\theta_2 + z$$
$$u = \theta_1 + 3\theta_2.$$

Here

$$x_2 = \begin{bmatrix}\eta_1\\ \eta_2\end{bmatrix}, \quad x_c = \begin{bmatrix}\theta_1\\ \theta_2\end{bmatrix}, \quad A_c = \begin{bmatrix}-\epsilon_1 & 1\\ 0 & -\epsilon_2\end{bmatrix}.$$

For $\epsilon_1 = \epsilon_2 = 0$ A_L is stable and A_c incorporates an internal model of

$$A_2 = \begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix};$$

hence output regulation holds. If $\epsilon_1 = 0$ and ϵ_2 is small, then

$$z(t) \to \epsilon_2\eta_2(0) \quad \text{as} \quad t \to \infty.$$

If $\epsilon_2 = 0$ and ϵ_1 is small, then

$$z(t) \to \frac{\epsilon_1}{3\epsilon_1 + 1}\eta_2(0) \quad \text{as} \quad t \to \infty.$$

But if ϵ_1, ϵ_2, and $\eta_2(0)$ are all nonzero, then

$$|z(t)| \to \infty \quad \text{as} \quad t \to \infty.$$

It is apparent from these two examples that the accuracy of regulation depends on the fidelity of the internal model.

4. EXTENSION TO WEAKLY NONLINEAR SYSTEMS

In this section we treat the case where the plant equation is weakly nonlinear, the exogenous signals are constant, and $y = z$. We extend Theorems 2 and 4 to this case.

The plant is assumed to be described by the equation

$$\dot{x}_1 = A_1x_1 + f_1(x_1) + A_3x_2 + B_1u, \qquad (14)$$

where $x_1 \in \mathbf{R}^{n_1}$, $x_2 \in \mathbf{R}^{n_2}$, and $u \in \mathbf{R}^m$ have the same meaning as in Section 2. The function $f_1 : \mathbf{R}^{n_1} \to \mathbf{R}^{n_1}$ is assumed of class C^1 with $f_1(x_1) = o(\|x_1\|)$ as $x_1 \to 0$. The real matrices A_1, A_3, and B_1 are of appropriate dimensions. In writing (14) we have in mind the following situation. The plant is initially described by the equation

$$\dot{x}_1 = f(x_1) + A_3x_2 + B_1u; \qquad (14a)$$

then (14) is obtained from (14a) by expanding f in a Taylor's series about $x_1 = 0$.

We consider only constant exogenous signals

$$\dot{x}_2 = 0. \qquad (15)$$

The output to be regulated, $z \in \mathbf{R}^q$, is given by

$$z = D_1x_1 + D_2x_2. \qquad (16)$$

For simplicity we assume a compensator processing z directly

$$\dot{x}_c = A_cx_c + B_cz \qquad (17a)$$

$$u = F_cx_c + Fz. \qquad (17b)$$

Here $x_c \in \mathbf{R}^{n_c}$.

Define

$$x_L = \begin{bmatrix} x_1 \\ x_c \end{bmatrix}, \quad A_L = \begin{bmatrix} A_1 + B_1FD_1 & B_1F_c \\ B_cD_1 & A_c \end{bmatrix}$$

$$B_L = \begin{bmatrix} A_3 + B_1FD_2 \\ B_cD_2 \end{bmatrix}, \quad D_L = [D_1 \quad 0]$$

$$f_L(x_L) = \begin{bmatrix} f_1(x_1) \\ 0 \end{bmatrix}.$$

Then (14), (16), and (17) become

$$\dot{x}_L = A_Lx_L + f_L(x_L) + B_Lx_2 \qquad (18a)$$

$$z = D_Lx_L + D_2x_2. \qquad (18b)$$

Closed loop stability here means, as in Section 2, that A_L is stable; that is, the linear part of the closed loop is stable. We revise the definition of *output regulation* to be: the solution $z = 0$ of (18) is asymptotically stable. Recall (see [12]) that this means that the following two conditions hold.

(i) For any $\epsilon > 0$ there exists $\delta > 0$ such that if $\|x_L(0)\| < \delta$ and $\|x_2(0)\| < \delta$ then $\|z(t)\| < \epsilon$ $(t \geq 0)$.

(ii) There exists $b > 0$ such that $z(t) \to 0$ as $t \to \infty$ whenever $\|x_L(0)\| < b$ and $\|x_2(0)\| < b$.

Thus output regulation is a local condition. We adopt this definition, of course, so that during the time evolution of the system, the linear part of the system is dominant. With closed loop stability and output regulation as just defined, structural stability has the same meaning as in Section 2.

Theorem 5 (Necessity of the internal model). *Suppose f_1 is a C^1-function with*

$$\lim_{x_1 \to 0} \frac{\|f_1(x_1)\|}{\|x_1\|} = 0. \qquad (19)$$

Then the synthesis is structurally stable at A_3 only if A_c incorporates an internal model of $A_2 = 0$ which is controllable by z and observable by u. Condition (19) says that near the origin of $\mathbf{R}^{n_1}$, $\|f(x_1)\|$ is dominated by $k\|x_1\|$ for some $k > 0$. This condition holds for example if f_1 is derived, as described earlier, from Taylor's series expansion of a sufficiently smooth function f.

The proof requires two lemmas. Let S_ϵ be the open ball in $\mathbf{R}^n$:

$$S_\epsilon := \{x : x \in \mathbf{R}^n, \|x\| < \epsilon\}.$$

Lemma 1

Suppose f: $\mathbf{R}^n \to \mathbf{R}^n$ is a function of class C^1 with

$$\lim_{x \to 0} \frac{\|f(x)\|}{\|x\|} = 0, \qquad (20)$$

and A is a stable $n \times n$ matrix. Then there exist $\epsilon > 0$ and $\delta > 0$ such that for each $a \in S_\epsilon$ the equation

$$\dot{x} = Ax + f(x) + a \qquad (21)$$

has a constant solution $x(t; a) = \bar{x}(a)$ which is the unique constant solution in S_δ. Moreover this solution is asymptotically stable and $\bar{x}(0) = 0$.

Proof:

Define $g(x) = Ax + f(x)$. Then from (20) $g(0) = 0$. Now by the continuity of f' (the differential of f) there exists $\delta_1 > 0$ such that

$$g'(x) = A + f'(x) \text{ is stable} \quad (x \in S_{\delta_1}). \qquad (22)$$

Thus $\det\, g'(0) \neq 0$, and so by the Inverse

Function Theorem there exist $\epsilon > 0$ and $\delta > 0$ such that the function

$$g: S_\delta \to S_\epsilon$$

has a differentiable inverse

$$g^{-1}: S_\epsilon \to S_\delta.$$

Clearly we may assume that $\delta \leq \delta_1$. Then for each $a \in S_\epsilon$ there is a unique $\bar{x} \in S_\delta$ such that

$$g(\bar{x}) + a = 0. \tag{23}$$

Moreover, since $g(0) = 0$, $\bar{x} = 0$ if $a = 0$.

Fix $a \in S_\epsilon$. Then $\bar{x}$, defined by (23), is a solution to (21) and is the unique constant solution in S_δ. It remains to show it is asymptotically stable.

Let $\phi(t) = \bar{x} + \psi(t)$ be a solution of (21). Then

$$\dot{\phi} = A\phi + f(\phi) + a = g(\phi) + a;$$

so

$$\dot{\psi} = g(\bar{x} + \psi) + a. \tag{24}$$

By the Mean Value Theorem

$$g(\bar{x} + \psi) = g(\bar{x}) + g'(\bar{x})\psi + h(\psi)$$

for some continuous function h: $\mathbf{R}^n \to \mathbf{R}^n$ with

$$\lim_{x\to 0} \frac{\|h(x)\|}{\|x\|} = 0.$$

Thus from (23) and (24)

$$\dot{\psi} = g'(\bar{x})\psi + h(\psi). \tag{25}$$

Now in view of (22) the zero solution of (25) is asymptotically stable ([12], Theorem 1.1, p. 314). Thus $\bar{x}$ is asymptotically stable. □

Lemma 2

Suppose f: $\mathbf{R}^n \to \mathbf{R}^p$ is a function of class C^1 and the interior of Im *f is nonempty. Then $n \geq p$.* For the proof of this lemma we shall need the following terminology from [13]. A point $x \in \mathbf{R}^n$ is a *critical point* of f if $f'(x)$: $\mathbf{R}^n \to \mathbf{R}^p$ is not surjective. A *critical value* of f is a point $y \in \mathbf{R}^p$ such that $y = f(x)$ for some critical point x.

Proof of Lemma 2

Suppose $n < p$. Then every point $x \in \mathbf{R}^n$ is a critical point of f. Thus Im f equals the set of critical values of f. By Sard's Theorem [13] the set of critical values, and hence Im f, has Lebesgue measure zero. This contradicts the fact that Im f contains a nonempty open set. □

Proof of Theorem 5

Notice that, since $A_2 = 0$, A_c incorporates an internal model of A_2 if and only if

$$d(\text{Ker } A_c) \geq q.$$

Suppose A_L is stable and output regulation holds throughout a nbhd of A_3. Applying Lemma 1 with $A := A_L$, $f := f_L$, and $a := B_L x_2$, we find that there exists $\epsilon > 0$ such that throughout a nbhd of A_3, if $\|x_L(0)\| < \epsilon$ and $\|x_2(0)\| < \epsilon$ then in (18) $x_L(t)$ tends to a constant and $z(t)$ tends to zero as $t \to \infty$.

Fix $x_2(0) \neq 0$ with $\|x_2(0)\| < \epsilon$, and set $x_L(0) = 0$. By taking the limit as $t \to \infty$ in (18) we conclude that throughout a nbhd of A_3 there exists a vector x_L such that

$$0 = A_L x_L + f_L(x_L) + B_L x_2(0) \tag{26a}$$

$$0 = D_L x_L + D_2 x_2(0). \tag{26b}$$

This is equivalent to

$$\begin{aligned} 0 &= (A_1 + B_1 F D_1)x_1 + B_1 F_c x_c + f_1(x_1) \\ &\quad + (A_3 + B_1 F D_2)x_2(0) \\ 0 &= B_c D_1 x_1 + A_c x_c + B_c D_2 x_2(0) \\ 0 &= D_1 x_1 + D_2 x_2(0), \end{aligned}$$

which implies

$$0 = A_1 x_1 + B_1 F_c x_c + f_1(x_1) + A_3 x_2(0) \tag{27a}$$

$$0 = A_c x_c \tag{27b}$$

$$0 = D_1 x_1 + D_2 x_2(0). \tag{27c}$$

Let $\mathscr{X}_{11}$ be an arbitrary complement of Ker D_1 in $\mathbf{R}^{n_1}$. There is a unique $x_{11} \in \mathscr{X}_{11}$ such that $D_1 x_{11} + D_2 x_2(0) = 0$. Then (27c) is equivalent to the condition that $x_1 = x_{11} + \hat{x}_1$ for some $\hat{x}_1 \in \text{Ker } D_1$. Hence (27a) becomes

$$-A_3 x_2(0) - A_1 x_{11} = A_1 \hat{x}_1 + B_1 F_c x_c + f_1(x_{11} + \hat{x}_1), \tag{28}$$

where, from (27b), $x_c \in \text{Ker } A_c$. Define

$$g_1: \text{Ker } D_1 \oplus \text{Ker } A_c \to \mathbf{R}^{n_1}$$

$$g_1(\hat{x}_1, x_c) = A_1 \hat{x}_1 + B_1 F_c x_c + f_1(x_{11} + \hat{x}_1).$$

From (28), Im g_1 contains a nonempty open set, namely

$$\{-\tilde{A}_3 x_2(0) - A_1 x_{11}: \tilde{A}_3 \text{ is in a nbhd of } A_3\}.$$

Hence by Lemma 2

$$d(\text{Ker } D_1 \oplus \text{Ker } A_c) \geq n_1,$$

which implies that

$$d(\text{Ker } A_c) \geq q.$$

Closed loop stability implies that A_L is invertible. Hence in particular

$$\text{Ker}\,(B_1F_c) \cap \text{Ker } A_c = 0$$

and

$$\text{Im } A_c + \text{Im}\,(B_cD_1) = \mathbf{R}^{n_c}.$$

These conditions imply respectively that

$$\text{Ker } F_c \cap \text{Ker } A_c = 0$$

and

$$\text{Im } A_c + \text{Im } B_c = \mathbf{R}^{n_c}; \tag{29}$$

that is, the internal model is observable by u and controllable by z. □

Now assume that A_c incorporates an internal model of $A_2 = 0$. Decompose $\mathbf{R}^{n_c}$ as

$$\mathbf{R}^{n_c} = \mathscr{X}_{c1} \oplus \mathscr{X}_{c2},$$

where $\mathscr{X}_{c1} = \text{Im } A_c$ and $\mathscr{X}_{c2}$ is an arbitrary complement. Corresponding to this decomposition write

$$A_c = \begin{bmatrix} A_{c1} & A_{c3} \\ 0 & 0 \end{bmatrix}.$$

Let $\mathscr{P}$ be the data point

$$(A_1, A_3, B_1, F_c, F, A_{c1}, A_{c3}, B_c).$$

Theorem 6

Suppose f_1 is a C^1 function satisfying (19). *Suppose also that the closed loop is stable and that A_c incorporates an internal model of $A_2 = 0$. Then the synthesis is structurally stable at $\mathscr{P}$.*

Proof

Choose a nbhd of $\mathscr{P}$ throughout which closed loop stability holds, and fix an arbitrary point, again denoted by $\mathscr{P}$, in this nbhd. From Lemma 1 there exists $\epsilon > 0$ such that if $\|x_L(0)\| < \epsilon$ and $\|x_2(0)\| < \epsilon$ then in (18) $x_L(t)$ tends to a constant as $t \to \infty$. Fix such $x_L(0)$, $x_2(0)$. Then taking the limit as $t \to \infty$ in (18a), we find that

$$0 = A_Lx_L + f_L(x_L) + B_Lx_2(0),$$

where $x_L := \lim_{t\to\infty} x_L(t)$. Thus in particular

$$0 = B_cD_1x_1 + A_cx_c + B_cD_2x_2(0),$$

and hence

$$0 = B_cz + A_cx_c, \tag{30}$$

where $z := \lim_{t\to\infty} z(t)$.

Now closed loop stability implies that, as in (29),

$$\text{Im } A_c + \text{Im } B_c = \mathbf{R}^{n_c},$$

which implies, since B_c: $\mathbf{R}^q \to \mathbf{R}^{n_c}$ and $d(\text{Ker } A_c) \geq q$,

$$\text{Im } A_c \cap \text{Im } B_c = 0 \text{ and } B_c \text{ is monic.}$$

Hence from (30) $z = 0$. □

The significance of Theorem 6 is the following: If the plant is described by

$$\dot{x}_1 = f(x_1) + A_3x_2 + B_1u,$$

and if the compensator is designed on the basis of a first order linear approximation of f and it incorporates an internal model, then, subject to mild smoothness assumptions on f, the synthesis is locally structurally stable for the nonlinear plant.

5. CONCLUSION

The regulator problem which we have considered is somewhat idealized: for example, we have demanded perfect asymptotic disturbance rejection. In practice the regulator problem is posed in fuzzy terms; thus, one may require attenuation of disturbances only to a certain degree. Nonetheless, our idealization has allowed a precise formulation of the problem which in turn has permitted a rigorous treatment. The result is a rational foundation for, and qualitative insight into, the practical design of multivariable regulators.

REFERENCES

[1] W. M. WONHAM: *Linear Multivariable Control: A Geometric Approach.* Lecture Notes in Economics and Math. Systems, Vol. 101, Springer-Verlag, New York (1974).

[2] E. J. DAVISON: The feedforward and feedback control of a general servomechanism problem, Parts I and II, *Proc. Eleventh Allerton Conf. on Circuit and Systems Theory,* pp. 343–362. Univ. of Illinois (1973).

[3] E. J. DAVISON and A. GOLDENBERG: The robust control of a general servomechanism problem: the servo compensator. *Automatica* 11, 461–471 (1975).

[4] P. W. STAATS JR. and J. B. PEARSON, Robust solution of the linear servomechanism problem, *Proc. IFAC Sixth Triennial World Congress,* Boston/Cambridge, MA (1975).

[5] B. A. FRANCIS and W. M. WONHAM: The internal model principle for linear multivariable regulators, *J. Appl. Maths. Optimization* 2(2), 170–194 (1975).
[6] B. A. FRANCIS and W. M. WONHAM, The rôle of transmission zeros in linear multivariable regulators. *Int. J. Control* 22(5), 657–681 (1975).
[7] C. R. KELLEY: *Manual and Automatic Control*. Wiley, New York (1968).
[8] R. C. CONANT and W. R. ASHBY, Every good regulator of a system must be a model of that system. *Int. J. Systems Sci.* 1(2) 89–97 (1970).
[9] R. L. GREGORY: *The Intelligent Eye*. Weidenfeld & Nicolson, London (1970).
[10] K. OATLEY: *Brain Mechanisms and Mind*. Thames & Hudson, London (1972).
[11] B. A. FRANCIS: Ph.D. Thesis, Dept. of Electrical Eng., Univ. of Toronto (1975).
[12] E. A. CODDINGTON and N. LEVINSON: *Theory of Ordinary Differential Equations*. McGraw-Hill, New York (1955).
[13] S. STERNBERG: *Lectures on Differential Geometry*, Prentice-hall, Englewood Cliffs, N.J. (1964).

APPENDIX: OUTLINE OF PROOF OF THEOREM 4

A complete proof of Theorem 4 is given in [11, Theorem 6.1]. As this reference is not widely accessible we here present a sketch of the proof.

We suppose the hypotheses of Theorem 4 hold. Since z is readable from y we may adopt the representations (7), (8) and (9). Corresponding to the partition (9) write $x_c = \begin{bmatrix} x_{c1} \\ x_{c2} \end{bmatrix}$. Now by redefining x_1 to be $\begin{bmatrix} x_1 \\ x_{c1} \end{bmatrix}$ and x_c to be x_{c2} it may readily be seen that we may assume at the outset that $y = z$. With this assumption the appropriate data point $\mathcal{P}$ is $(A_1, A_3, B_1, F_c, F, B_c)$.

Choose a nbhd of $\mathcal{P}$ small enough so that closed loop stability is maintained throughout this nbhd, and choose any point (again denoted by $\mathcal{P}$) in this nbhd. We shall show that output regulation holds at this point.

From (1b), (4) and (5) the system equations are

$$\dot{x}_L = A_L x_L + B_L x_2 \tag{31a}$$

$$\dot{x}_2 = A_2 x_2 \tag{31b}$$

$$z = D_L x_L + D_2 x_2. \tag{31c}$$

If A_L is stable the equation

$$A_L X - X A_2 = B_L \tag{32}$$

has a unique solution X. Define $\bar{x}_L = x_L + X x_2$. Then (31) is transformed into

$$\dot{\bar{x}}_L = A_L \bar{x}_L$$

$$\dot{x} = A_2 x_2$$

$$z = D_L \bar{x}_L + (D_2 - D_L X) x_2.$$

Thus output regulation will be guaranteed once we show that $D_2 = D_L X$, or equivalently

$$D_2 = D_1 V \tag{33}$$

where we have written

$$X = \begin{bmatrix} V \\ W \end{bmatrix} : \mathscr{X}_2 \to \mathscr{X}_1 \oplus \mathscr{X}_c.$$

It follows from (32) that

$$B_c D_1 V + A_c W - W A_2 = B_c D_2$$

or

$$B_c Z + A_c W - W A_2 = 0 \tag{34}$$

where $Z = D_1 V - D_2$. According to (33) we must show that $Z = 0$.

By considering a Jordan decomposition of $\mathscr{X}_2$ relative to A_2 and by restricting the maps in (34) to any cyclic subspace in this decomposition, we may as well assume that A_2 is a $k \times k$ Jordan matrix

$$\begin{bmatrix} \lambda & 1 & & & 0 \\ & \lambda & \cdot & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & 1 \\ 0 & & & & \lambda \end{bmatrix}.$$

Taking matrix representations for Z and W we write

$$Z = [z_1, \ldots, z_k], \quad W = [w_1, \ldots, w_k]$$

with $z_i \in \mathscr{Z}$ and $w_i \in \mathscr{X}_c (i \in \mathbf{k})$. Then (34) implies

$$\begin{aligned} &B_c z_1 + (A_c - \lambda) w_1 = 0 \\ &B_c z_2 + (A_c - \lambda) w_2 - w_1 = 0 \\ &\vdots \\ &B_c z_k + (A_c - \lambda) w_k - w_{k-1} = 0. \end{aligned} \tag{35}$$

From the fact that the closed loop is stable and A_c incorporates an internal model of A_2 it can be deduced that

$$\text{Ker}\, B_c = 0 \tag{36a}$$

$$\text{Im}\, B_c \cap \text{Im}\,(A_c - \lambda) = 0 \tag{36b}$$

and

$$\text{Ker}\,(A_c - \lambda)^{k-1} \subset \text{Im}\,(A_c - \lambda). \tag{37}$$

Now (36) and the first equation in (35) show that $z_1 = 0$ and $w_1 \in \text{Ker}\,(A_c - \lambda)$; hence from (37) $w_1 \in \text{Im}\,(A_c - \lambda)$. Then (36) and the second equation in (35) show that $z_2 = 0$ and $w_2 \in \text{Ker}\,(A_c - \lambda)^2$. Continuing in this fashion we find that $Z = 0$. □