Feedback and Optimal Sensitivity: Model Reference Transformations, Multiplicative Seminorms, and Approximate Inverses

G. ZAMES, FELLOW, IEEE

Abstract—In this paper, the problem of sensitivity reduction by feedback is formulated as an optimization problem and separated from the problem of stabilization. Stable feedback schemes obtainable from a given plant are parameterized. Salient properties of sensitivity reducing schemes are derived, and it is shown that plant uncertainty reduces the ability of feedback to reduce sensitivity.

The theory is developed for input–output systems in a general setting of Banach algebras, and then specialized to a class of multivariable, time-invariant systems characterized by $n \times n$ matrices of H^{∞} frequency response functions, either with or without zeros in the right half-plane.

The approach is based on the use of a *weighted seminorm* on the algebra of operators to measure sensitivity, and on the concept of an *approximate inverse*. Approximate invertibility of the plant is shown to be a necessary and sufficient condition for sensitivity reduction. An indicator of approximate invertibility, called a *measure of singularity*, is introduced.

The measure of singularity of a linear time-invariant plant is shown to be determined by the location of its right half-plane zeros. In the absence of plant uncertainty, the sensitivity to output disturbances can be reduced to an optimal value approaching the singularity measure. In particular, if there are no right half-plane zeros, sensitivity can be made arbitrarily small.

The feedback schemes used in the optimization of sensitivity resemble the lead–lag networks of classical control design. Some of their properties, and methods of constructing them in special cases are presented.

I. INTRODUCTION

IN THIS paper we shall be concerned with the effects of feedback on uncertainty, where uncertainty occurs either in the form of an additive disturbance d at the output of a linear plant P (Fig. 1), or an additive perturbation in P representing "plant uncertainty." We shall approach this subject from the point of view of classical sensitivity theory, with the difference that feedbacks will not only reduce but actually optimize sensitivity in an appropriate sense.

The theory will be developed at two levels of generality. At the higher level, a framework will be sought in which the essence of the classical ideas can be captured. To this end, systems will be represented by mappings belonging to a normed algebra. The object here is to obtain general answers to such questions as: how does the usefulness of feedback depend on plant invertibility? are there measures

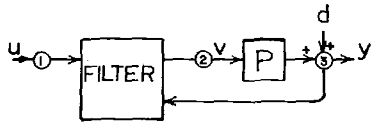


Fig. 1.

of sensitivity or plant uncertainty that are natural for optimization? how does plant uncertainty affect the possibility of designing a feedback scheme to reduce plant uncertainty?

At a more practical level, the theory will be illustrated by simple examples involving single variable and multivariable frequency responses. The questions here are: can the classical "lead–lag" controllers be derived from an optimization problem? How do RHP (right half-plane) zeros restrict sensitivity? in multivariable systems without RHP zeros, can sensitivity be made arbitrarily small, and if so how?

A. Motivation

A few observations might serve to motivate this reexamination of feedback theory.

One way of attenuating disturbances is to introduce a filter of the WHK (Wiener–Hopf–Kalman) type in the feedback path. Despite the unquestioned success of the WHK and state-space approaches, the classical methods, which rely on lead–lag "compensators" to reduce sensitivity, have continued to dominate many areas of design. On and off, there have been attempts to develop analogous methods for multivariable systems. However, the classical techniques have been difficult to pin down in a mathematical theory, partly because the purpose of compensation has not been clearly stated. One of our objectives is to formulate the compensation problem as the solution to a well defined optimization problem.

Another motivating factor is the gradual realization that classical theory is not just an old-fashioned way of doing WHK, but is concerned with a different category of mathematical problems. In a typical WHK problem, the quadratic norm of the response to a disturbance d is minimized by a projection method (see Sections III′-A′ and IV-C); in a deterministic version, the power spectrum

Manuscript received October 8, 1979; revised December 4, 1980 and December 17, 1980. Paper recommended by A. Z. Manitius, Past Chairman of the Optimal Systems Committee. An earlier version of this paper [23] was presented the 17th Allerton Conference, October 1979.

The author is with the Department of Electrical Engineering, McGill University, Montreal, P.Q., Canada.

0018-9286/81/0400-0301$00.75 ©1981 IEEE

$|\hat{d}(j\omega)|$ is a *single, known* vector in, e.g., the space $L_2(-\infty,\infty)$; in stochastic versions, d belongs to a *single* random process of *known* covariance properties. However, there are many practical problems in which $|\hat{d}(j\omega)|$ is unknown but belongs to a prescribed set, or d belongs to a class of random processes whose covariances are uncertain but belong to a prescribed set. For example, in audio design, d is often one of a set of narrow-band signals in the 20–20K Hz interval, as opposed to a single, wide-band signal in the same interval. Problems involving such more general disturbance sets are not tractable by WHK or projection techniques. In a feedback context, they are now usually handled by empirical methods resembling those of classical sensitivity. One objective here is to find a systematic approach to problems involving such sets of disturbances.

Another observation is that many problems of plant uncertainty can be stated easily in the classical theory, e.g., in terms of a tolerance-band on a frequency response as in [2], but are difficult to express in a linear-quadratic-state-space framework. One reason for this is that frequency-response descriptions and, more generally, input–output descriptions preserve the operations of system addition and multiplication, whereas state-space descriptions do not. Another reason is that the quadratic norm is hard to estimate for system products (see Sections III′-A′ and IV-B1), whereas the induced norm (or "gain") that is implicit in the classical theory is easier to estimate. We would like to exploit these advantages in the study of plant uncertainty.

Finally, sensitivity theory is one of the few tools available for the study of organization structure: feedback versus open-loop, aggregated versus disaggregated, etc. For example, feedback reduces complexity of identification roughly for the same reason that it reduces sensitivity [12], [13]. However, it is hard to draw definitive conclusions about the effects of organization without some notion of optimality, and such a notion is missing in the old theory.

B. Weighted Seminorms and Approximate Inverses

One way of defining the optimal sensitivity of a feedback system, and of addressing some of the issues mentioned in Section I-A, is in terms of an induced norm of the sensitivity operator. However, it will be shown in Section III′-B′ that the primary norm of an operator in a normed algebra is useless for this purpose. Perhaps that is why operator norm optimization has not been pursued extensively in the past.

Instead, we shall introduce an auxiliary "weighted" seminorm, which retains some of the multiplicative properties of the induced norm, but is amenable to optimization. Plant uncertainty will be described in terms of belonging to a sphere in the weighted seminorm.

Approximate invertibility of the plant is one of the features which distinguishes control from, say, communication problems. We shall define the concept of an approximate inverse under a weighted seminorm, and show that sensitivity reduction is possible if there is such an inverse.

C. Background

Many of the ideas in this paper are foreshadowed in the classical theory [1], [2] of single-input single-output convolution systems, especially as presented by Horowitz [2], who derived various limits on sensitivity imposed by the plant, and stressed the need to consider plant uncertainty in design. The author posed the feedback problem in a normed algebra of operators on a Banach space, and introduced [4], [5] perturbation formulas of the type

$$(I-P)^{-1}-(I-P_0)^{-1}=(I-P)^{-1}(P-P_0)(I-P_0)^{-1} \tag{1.1}$$

which were used to show that high-gain feedback reduces the sensitivity of linear amplifiers to large nonlinear perturbations [3]–[5]. Desoer studied a related problem in [6], and recently [7] has obtained results for the case of P and P_0 both nonlinear (also see footnote 8). Perkins and Cruz [8] used perturbation formulas similar to (1.1) to calculate the sensitivity of linear multivariable systems. Porter [9] posed various sensitivity problems in Hilbert space, and in a paper with Desantis [10] obtained circle type conditions for sensitivity reduction. Willems [11] has stressed the Banach algebraic aspects of feedback theory.

In [1]–[10], the disturbance is either a fixed vector, or lies in some band of frequencies, and sensitivity is measured in terms of an output norm, as opposed to an induced operator norm. The approach of using weighted operator norms, and relating optimal sensitivity to weighted invertibility via a fractional transformation was used in [12], but has since been reworked and expanded.

D. Two Problems

We shall be concerned with the system of Fig. 1. Here, P is a given *plant* with a single (possibly multivariable) input v accessible to control, and an output y to which a disturbance d, not accessible to control, has been added. The plant input v is generated by a filter whose only inputs consist of observations on the plant output y and a reference input u. Two types of problems will be considered.

Problem 1—Disturbance Attenuation: This problem will be the subject of Sections V–VII. Suppose that $u=0$. The input–output behavior of the system between the nodes (2, 3) can be modeled by the flowgraph of Fig. 2, which consists of the plant P and a single additional operator F in the feedback path. The disturbance d is uncertain in the sense that it can be any one of a set of disturbances. Initially (through Section VI) P is assumed to be known exactly, but later (Section VII) to be uncertain. We would like to *characterize the feedback operators F which attenuate the response y to d in some appropriately optimal sense, and examine the effects of uncertainty about P on disturbance attenuation.*

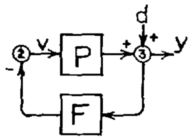
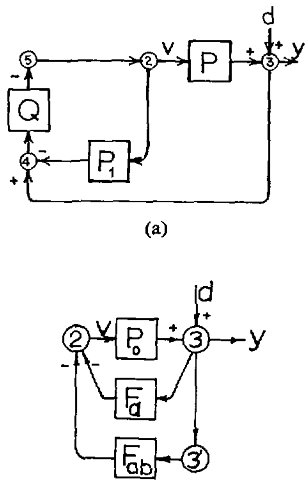


Fig. 2.



(a)

(b)

Fig. 3.

Problem 2—Plant Uncertainty Attenuation: (Problem 2 is the subject of Section VIII.) Suppose that $d=0$, and the plant P is uncertain to the extent that it can be any one of a "ball" of possible plants centered around some nominal value P_1. If the filter is linear, the behavior of the system between nodes $(1,2,3)$ can be modeled by the flowgraph of Fig. 4. The filter can be characterized by a pair of operators (U, F). We would like to *find operators* (U, F) *which shrink the ball of uncertainty but leave the nominal plant invariant*; *to find bounds on the optimal shrinkage and to look at its dependence on plant uncertainty.*

E. *Outline of the Paper*

See Synopsis following Appendixes.

II. Spaces and Algebras of Systems

The purpose of this section is to specify the meaning which will be attached to the terms "frequency-response" and "linear system," and to summarize their properties for later use.

A feature of the input–output approach is that systems can be added, multiplied by other systems or by scalars, and the sums or products obtained are still systems, i.e., they form an *algebra*. Frequently, it will be assumed that the largest amplification produced by a system can be measured by a norm, typically the maximum frequency response amplitude over some region of analyticity; under this assumption the algebra of systems becomes a *normed algebra*. Normed algebras provide the natural setting for the study of system interconnections such as feedback. Their elementary properties will be used freely here, and

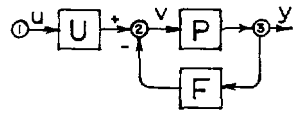


Fig. 4.

may be found in such texts as Naimark [19]. Occasionally, it will be assumed that a normed algebra is a *Banach algebra*, i.e., has the property that every convergent sequence of elements of the algebra has a limit in the algebra.

It will be assumed that all linear spaces and algebras are over the real field.

A. *Algebras of Frequency Response Systems*

The frequency response of a stable, causal, linear time-invariant system is a function analytic in the right-half of the complex plane. An accepted setting for such functions involves the H^P Hardy spaces [15], which we shall employ with some modifications to accommodate unstable systems.

The algebra H_e^∞ consists of functions $\hat{p}(\cdot)$ of a complex variable $s=\sigma+j\omega$, each of which is analytic in some open half-plane $\mathrm{Re}(s)>\sigma_p$ possibly depending on $\hat{p}(\cdot)$, and is bounded there, i.e., $\hat{p}(s)\leqslant$const. for $\mathrm{Re}(s)>\sigma_p$. The functions in H_e^∞ will be referred to as *causal* frequency responses. If $\hat{p}$ is in H_e^∞, then the domain of definition of $\hat{p}$ can be extended by analytic continuation to a unique, maximal, open half-plane of analyticity $\mathrm{Re}(s)>\sigma_{pm}$, where $\sigma_{pm}\leqslant\sigma_p$. In general, $\hat{p}$ need not be bounded on this maximal open RHP, but if it is, then it can further be extended to the boundary by the limit $\hat{p}(\sigma_{pm}+j\omega \triangleq \lim_{\sigma\to\sigma_{pm}}\hat{p}(\sigma+j\omega)$, which exists for almost all ω provided $\sigma+j\omega$ approaches the boundary nontangentially from the right. Assume that all functions in H_e^∞ have been so extended.

The algebra H^∞ (of *stable causal* frequency responses) consists of functions $\hat{p}$ of H_e^∞ for which $\sigma_p\leqslant 0$, i.e., the region of bounded analyticity includes the RHP. The norm $\|\hat{p}\|=\sup\{|\hat{p}(s)|\colon \mathrm{Re}(s)>0\}$ is defined on H^∞, making H^∞ a normed and, indeed, Banach algebra.

A *strictly proper* function in H_e^∞ satisfies the condition $\hat{p}(s)\to 0$ as $|s|\to\infty$ in $\mathrm{Re}(s)\geqslant\sigma_p$. The symbols H_{e0}^∞ and H_0^∞ will denote the algebras of strictly proper frequency responses in H_e^∞ and H^∞, respectively. By a straightforward application of the maximum modulus principle, the normed algebra H_0^∞ of strictly proper stable frequency responses has the property that $\|\hat{p}\|=\operatorname{ess\,sup}\{|\hat{p}(j\omega)|\colon \omega$ real$\}$ for any $\hat{p}$ in H_0^∞, i.e., the norm can be computed from $j\omega$-axis measurements.

Spaces of Inputs and Outputs: For any integer $1\leqslant q<\infty$, the linear space H_e^q consists of functions $\hat{u}(\cdot)$ of a complex variable, each $\hat{u}$ being analytic in some open half-plane $\mathrm{Re}(s)>\sigma_u$ in which the restriction of $\hat{u}$ to any vertical line is in L^q and $\int_{-\infty}^{\infty}|\hat{u}(\sigma+j\omega)|^q\,d\omega\leqslant$const. for all $\sigma>\sigma_u$. Again, the domain of definition of each $\hat{u}$ in H_e^q is extended by analytic continuation to a maximal open RHP of analyticity and then, if $\hat{u}$ is L^q-bounded in this RHP, to its boundary by a nontangential limit. The space H^q consists of functions $\hat{u}$ in H_e^q for which $\sigma_u \leqslant 0$, and is a Banach space under the norm $\|\hat{u}\| = \sup_{\sigma>0}\{\int_{-\infty}^{\infty}|\hat{u}(\sigma+j\omega)|^q d\omega\}^{1/q}$.

Inputs and outputs will belong either to H_e^∞ or one of the H_e^q spaces, $1 \leqslant q < \infty$. In the special case $q=2$ it follows from the Paley–Wiener theory that every function of H^2 is a Laplace transform of a time function in $L^2(0,\infty)$ and vice versa. In the general case of H^q, some functions can be viewed as transforms of time functions, and the others as frequency functions that do not appear in physical applications and can be disregarded.

Frequency Response Operators: Let q be any integer, $1 \leqslant q \leqslant \infty$, which will be held fixed. For any causal frequency response $\hat{p}(\cdot)$ in H_e^∞ an operator P: $H_e^q \to H_e^q$ is defined by the multiplication $\widehat{Pu}(s)=\hat{p}(s)\hat{u}(s)$. P will be called a causal frequency response operator, and the algebra of all such operators will be denoted by the bordered capital $\mathbb{H}_e^\infty$. Similarly, for each of the algebras of frequency responses H^∞, H_{e0}^∞, and H_0^∞, an algebra of operators mapping H_e^q into itself is defined and denoted by the corresponding bordered capital, i.e., $\mathbb{H}^\infty$, $\mathbb{H}_{e0}^\infty$, or $\mathbb{H}_0^\infty$. In the case of normed frequency response algebras, the corresponding operator algebras are similarly normed, e.g., $\|P\|_{\mathbb{H}^\infty} = \|\hat{p}\|_{H^\infty}$.

The stable operators in $\mathbb{H}^\infty$ map H^q, which is a proper subspace of H_e^q, into H^q, and are in fact completely determined by their behavior on H^q. They can therefore be represented by their restrictions of the form P: $H^q \to H^q$. In sections devoted entirely to stable systems, we shall concentrate on operators of the form P: $H^q \to H^q$ without distinguishing them as restrictions of operators on H_e^q.

B. More General Algebras of Systems

We would like to take an axiomatic approach to the problem of sensitivity reduction by feedback, i.e., to single out the relevant properties of linear systems and postulate them as axioms. For example, the related properties of causality, realizability, and strong or strict causality have definitions [see, e.g., [14] and [11]) reflecting the fact that the response to a sudden input to a physical system can not anticipate the input, and cannot occur instantaneously. These properties of physical systems preclude the pathological phenomena associated with instantaneous response around a feedback loop, and ensure that the feedback operator $(I+P)^{-1}$ is well defined. However, these details are not relevant here. The only items of interest are that causal systems form an algebra of mappings, and that strictly causal systems form a subalgebra whose salient feature is the existence of the inverse $(I+P)^{-1}$ for all of its members, i.e., a "radical." Accordingly we postulate the following.

$\mathfrak{X}$ is a linear space whose elements will be called *inputs* or *outputs*. $\mathbb{A}$ is a linear algebra of linear mappings P: $\mathfrak{X} \to \mathfrak{X}$ with identity I, whose elements will be called *causal operators*. $\mathbb{A}_s$ is a radical of $\mathbb{A}$, i.e.,[1] a proper nontrivial subalgebra of $\mathbb{A}$ with the property that for any P in $\mathbb{A}_s$, the inverse $(I+P)^{-1}$ exists in $\mathbb{A}$, and for any F in $\mathbb{A}$ the products PF and FP are in $\mathbb{A}_s$. The elements of $\mathbb{A}_s$ will be called *strictly causal operators.* (The concept of an algebra of "realizable" systems was introduced by the author in [14]; the related notions of strong causality of Willems [11], and later strict causality of Porter, Saeks, and Desantis are compared in a paper of Feintuch [20].)

An example of the space $\mathfrak{X}$ of inputs and the algebra $\mathbb{A}$ of causal operators is provided by the space H_e^q of (transforms of) inputs and the algebra $\mathbb{H}_e^\infty$ of causal frequency response operators. In this context the algebra $\mathbb{H}_{e0}^\infty$ of strictly proper frequency response operators is an example of an algebra $\mathbb{A}_s$ of strictly causal operators. Henceforth, we shall refer to the strictly proper operators as strictly causal.

In the case of stability, we shall need the fact that a stable input–output system produces a finite amplification of inputs that can be measured by a suitable norm, and that stable systems form an algebra (see, e.g. [5]). Accordingly, we postulate the following.

$\mathfrak{B}$ is a Banach subspace of $\mathfrak{X}$ whose elements will be called *bounded* inputs or outputs [for example, H^2 or $L_2(0,\infty)$]. $\mathbb{B}$ is a normed subalgebra of $\mathbb{A}$ containing the identity I, whose elements will be called *stable* causal operators, under the following assumption: the norm of any P in $\mathbb{B}$ is the $\mathfrak{B}$-induced norm, that is, $\|P\| \triangleq \sup\{\|Pu\|/\|u\|$: u in $\mathfrak{B}$, $u \neq 0\}$, the sup being finite.[2]

If $\mathbb{H}_e^\infty$ is taken as an example of $\mathbb{A}$, then $\mathbb{H}^\infty$ is an example of $\mathbb{B}$.

$\mathbb{B}_s$ is the subalgebra of $\mathbb{B}$ obtained by intersecting $\mathbb{A}_s$ and $\mathbb{B}$, consisting of *strictly-causal stable* operators. It should be noted that $\mathbb{B}_s$ is not a radical of $\mathbb{B}$, as P stable does not imply that $(I+P)^{-1}$ is stable.

For the purpose of estimating the effects of small perturbations, it will be assumed that $\mathbb{B}_s$ has the *small-gain property*, i.e., for any P in $\mathbb{B}_s$, if $\|P\|<1$ then $(I+P)^{-1}$ is in $\mathbb{B}$. If $\mathbb{B}$ is complete, i.e., a Banach space, this assumption is redundant for then the series $I-P+P^2-\cdots$ converges to the inverse in $\mathbb{B}$ of $(I+P)$. However, we have applications in mind in which completeness of $\mathbb{B}$ is replaced by other assumptions.

A frequency response $\hat{p} \in H^\infty$ which is not strictly proper can not be realized exactly, but can be approximated by a sequence of strictly proper responses of the form $n(s+n)^{-1}\hat{p}(s)$, $n=1,2,\cdots$. The sequence $n(s+n)^{-1}$ is an example of an "identity sequence." More generally, identity sequences will be used to construct strictly proper approximations to improper responses, and are defined as follows: an *identity sequence* $\{I_n\}_{n=1}^\infty$ for $\mathbb{B}$ is a sequence of operators in $\mathbb{B}$ with the property that for any F in $\mathbb{B}$ the sequences $\|I_nF-F\|$ and $\|FI_n-F\|$ approach 0 as $n\to\infty$. It is assumed that $\mathbb{B}_s$ contains an identity sequence for $\mathbb{B}$.

The following well-known (cf. Naimark [19, p. 162]) properties of a normed algebra will be crucial in many

[1] The properties of radicals are discussed in Naimark [19, p. 162].

[2] Whenever the norm of x is not identified by a subscript, it should be taken to be the principal norm of the space to which x belongs.

parts of the paper. For convenience they are proved in Appendix I.

Let P and Q be in B.

Proposition 2.1: a) If $(I+PQ)^{-1}$ is in B, then $(I+QP)^{-1}$ is in B, and the formula $P(I+QP)^{-1}=(I+PQ)^{-1}P$ is valid. b) If R is a radical in B and P is in R, then P has no inverse in B. (Strictly causal operators have no inverses in $\mathbb{B}$.) c) If P and $(I+P)^{-1}$ are in B and $\|P\|<1$, then $\|(I+P)^{-1}\|\leqslant(1-\|P\|)^{-1}$.

III. Feedback Decomposition: Stabilizing and Stabilized Stages

We proceed to derive a decomposition principle to be employed in disturbance attenuation. Suppose that there is no plant uncertainty, and that the plant and feedback are constrained not to be simultaneously unstable. Under these hypotheses, any closed-loop stable feedback design can be decomposed into two stages: a first stage involving plant stabilization (which can be omitted for stable plants); and a second stage, involving a model reference scheme in which only stable elements are used, and which is automatically closed-loop stable. The choice of a stabilizing stage is independent of, and does not prejudice the choice of the second stage. Having established this fact, we shall be free to concentrate on the second stage of the disturbance attenuation problem under the condition that the plant is stable (or has been stabilized), without loss of generality.

Consider the system of Fig. 2. The plant input v, output y, and disturbance d are all in $\mathfrak{X}$ and satisfy the equations

$$y=Pv+d \tag{3.1a}$$

$$v=-Fy \tag{3.1b}$$

in which P and F are operators in $\mathbb{A}_s$ [see Remark 3.1c)]. We shall refer to (3.1) as a *feedback scheme* with plant P and feedback F. Since P is strictly causal, the inverse $(I+PF)^{-1}$ exists in $\mathbb{A}$. Therefore, for each d in $\mathfrak{X}$, (3.1) have unique solutions for v and y in $\mathfrak{X}$, given by the formulas

$$y=(I+PF)^{-1}d \tag{3.2a}$$

$$v=-F(I+PF)^{-1}d. \tag{3.2b}$$

Let K_{32}: $\mathfrak{X}\to\mathfrak{X}$ denote the "closed-loop" operator mapping d to v. K_{32} is an operator in $\mathbb{A}$ given by $K_{32}=-F(I+PF)^{-1}$.

The flowgraphs in this paper are simple, and will be approached informally in order to avoid lengthy definitions. Expressions for some of the subsidiary c.l. (closed-loop) operators, which can be found by inspection, will be listed without derivation as needed.

For a system to be physically realizable on an infinite time interval, it is usual to postulate [14] that all c.l. input–output operators must be stable, though "open-loop" operators such as the plant P and feedback F may be unstable. The set of c.l. operators for (3.1) consists of: $K_{22}=(I+FP)^{-1}$, $K_{23}=PK_{22}$, $K_{33}=(I+PF)^{-1}$, and K_{32} specified above. Accordingly, the feedback scheme (3.1) will be called *c.l. stable* if K_{ij} is in $\mathbb{B}$ for $i, j=2$ or 3.

We shall be interested in situations in which P is at or near some nominal value P_1, and the feedback F appears as an operator variable in an optimization problem whose object is to minimize response to d. Unstable operator variables are difficult to handle, and so our first step will be to show that F has an equivalent realization in terms of a stable operator.

A. The Model Reference Transformation

The flowgraph of Fig. 3 is described by the equations

$$y=Pv+d \tag{3.3a}$$

$$v=-Q(y-P_1v) \tag{3.3b}$$

in which y, v, and d are in $\mathfrak{X}$, and P, P_1, Q are in $\mathbb{A}_s$. Equation (3.3) will be called a *model reference scheme* with *comparator* Q, as the output of the plant P with disturbance d added is compared to the output of a model P_1 of the plant without disturbance, and the difference actuates Q.

The two sets of equations (3.1) and (3.3) are called *equivalent* iff every input–output triple (d, v, y) in $\mathfrak{X}^3$ satisfying (3.1) satisfies (3.3), and vice versa. Their equivalence will be established under the assumption that the equations

$$Q=F(I+P_1F)^{-1} \tag{3.4a}$$

$$F=Q(I-P_1Q)^{-1} \tag{3.4b}$$

hold. If either equation in (3.4) is valid, then so is the other, and

$$(I-P_1Q)=(I+P_1F)^{-1}. \tag{3.5}$$

To derive (3.5), suppose (3.4a) is valid. Therefore,

$$\begin{aligned}I-P_1Q&=I-P_1F(I+P_1F)^{-1}\\&=(I+P_1F)(I+P_1F)^{-1}-P_1F(I+P_1F)^{-1}\\&=(I+P_1F)^{-1}\end{aligned}$$

and (3.5) is true; here the expression $I=(I+P_1F)(I+P_1F)^{-1}$ and the distributive law for multiplication on the right was used. Equations (3.4a) and (3.5) can now be used to give the identities

$$Q(I-P_1Q)^{-1}=F(I+P_1F)^{-1}(I+P_1F)=F,$$

so (3.4b) is true as claimed. The converse proposition is proved similarly.

Assumption: For the present, and until the end of Section VI, assume that $P=P_1$, i.e., there is no plant uncertainty.

Theorem 1:

a) Any closed-loop stable feedback scheme (3.1) with stable plant $P\in\mathbb{B}_s$ and (not necessarily stable) feedback $F\in\mathbb{A}_s$ is equivalent to a model reference scheme whose branches are all stable, i.e., $Q\in\mathbb{B}_s$ and $P_1=P$, where F and Q are related by (3.4). Conversely, any model reference scheme with stable branches is closed-loop stable, and equivalent to a closed-loop stable feedback scheme subject to (3.4).

b) If (3.4) holds, and d and y satisfy either the feedback or model reference equations they satisfy the equation

$$y=(I-P_1Q)d. \tag{3.6}$$

Proof:

a) For any feedback scheme (3.1), if $(d,v,y)\in\mathfrak{X}^3$ satisfies (3.1) and F is given by (3.4b), then (d,v,y) satisfies (3.3), and conversely. Therefore, (3.1) is equivalent to (3.3). If the feedback scheme is c.l. stable, then Q must be stable as it equals $-K_{32}$. If, in addition, P is assumed stable, then all branches in (3.3) are stable, as claimed.

Conversely, by a similar argument, any model reference scheme (3.1) is equivalent to a feedback scheme (3.3). Suppose that the branches of (3.3), namely, P_1, Q, and $P=P_1$ are all stable. Then, all the c.l. operators of (3.3), namely, $\{K_{ij}\}_{i,j=2,3,4,5}$, must be stable because they can be expressed in terms of sums and products of the stable operators P, P_1, Q, and I. The last assertion follows from the expressions for the diagonal c.l. operators K_{ii} of (3.3), namely,

$$K_{22}=I-QP_1,\ K_{33}=I-P_1Q,\ K_{44}=K_{55}=I$$

and the fact, easily checked by inspection, that the remaining c.l. operators K_{ij}, $i\neq j$, are products of the K_{ii} by P, P_1, Q, or I. It follows that (3.3) is c.l. stable.

b) If $P=P_1$ and (d,y) satisfies (3.1) or (3.3), and (3.4) holds, then (3.6) is obtained by substitution of (3.5) into (3.2a). Q.E.D.

The operator $(I-P_1Q)$ appearing in (3.6) will reappear as a factor in most expressions for sensitivity. It will be called the *sensitivity operator* and denoted by E. For equivalent schemes $E=(I+P_1F)^{-1}$.

Remarks 3.1:

a) The model-reference scheme has some remarkable features. Unlike most feedback arrangements, it is a realization which cannot be made unstable by any choice of Q, at least for stable plants in the absence of plant uncertainty. Under these assumptions,[3] any allowable feedback law can be realized in the form of an equivalent model reference scheme, with the guarantee that all branches will be stable, and the closed-loop system automatically stable. *The design of Q, whether for small sensitivity or other purposes, can be accomplished without concern for closed-loop stability.*

In engineering applications, model reference schemes are realizable in principle, but may have undesirable features. For example, they may have high sensitivity to errors in the realization of Q. Unstable inner loops, obtained whenever $(I-P_1Q)^{-1}$ is unstable, may present reliability problems. Even then, the fractional transformation remains advantageous from the viewpoint of theory, as potentially unstable feedbacks F are replaced by stable operators Q. In later sections on plant uncertainty, the flowgraph interpretation of the model reference scheme will provide a convenient guide to perturbation analysis. It will also appear that model-reference schemes have a useful plant-invariance property.

b) Implicit in our notion of an allowable feedback is the view that each feedback realization involves a graph, and that although most of the internal details of the realization may be unimportant, closed-loop stability at *all* internal nodes is essential.

c) Theorem 1 holds even if F and Q are in $\mathbb{A}$ but not strictly causal. However, strict causality is a prerequisite for physical realizability, and will therefore have to be assumed in subsequent theorems.

B. Unstable Plants

The assumption in Theorem 1 that the plant P is stable will now be relaxed. Consider a plant $P_0\in\mathbb{A}_s$ with disturbance d at the output, which is unstable but for which there exists a *stabilizing feedback*, i.e., an operator $F_0\in\mathbb{A}_s$ which gives a c.l. stable feedback scheme on being fed back around P_0. The stabilized system can be incorporated in a model reference scheme, by letting P be the stabilized c.l. operator $P_0(I+F_0P_0)^{-1}$ and d be the stabilized disturbance $(I+P_0F_0)^{-1}d_0$, and Q (or F) can be selected as for a stable plant. At this point the question arises: "can F_0 be selected independently of Q (or F), or could the prior choice of F_0 prejudice the class of achievable systems?"

In general, the choices are not independent, even for stable plants, because the application of two unstable feedbacks in succession may give a result different from the application of a single feedback equal to their sum. Consider the following frequency response example in $\mathbb{H}_e^\infty$. Let $\hat{p}(s)=1$ and $\hat{f}_1(s)=\hat{f}_2(s)=s^{-1}$. The application of a single feedback $\hat{f}(s)$ equal to $\hat{f}_1(s)+\hat{f}_2(s)$ gives a c.l. stable feedback scheme, with c.l. responses 1, $(s+2)^{-1}$, and $s(s+2)^{-1}$. However, if the feedback is split into two branches, the c.l. response across either one of these branches is $(s+1)/s(s+2)$, i.e., the system is not c.l. stable. Popular belief notwithstanding, *c.l. stable systems do not form an additive group under feedback if the complete set of c.l. operators is considered.*

However, if feedbacks are constrained to be stable then choices are independent, as the following construction shows.

Let $P_0\in\mathbb{A}_s$ be an unstable plant which can be stabilized by either one of two feedbacks, F_a and F_b in $\mathbb{A}_s$, and label the resulting feedback schemes (a) and (b), respectively. We would like to find an operator $F_{ab}\in\mathbb{A}_s$ which on being fed back around scheme (a), as shown in Fig. 3(a), produces a *two-stage feedback scheme* equivalent to scheme (b). (Observe that the two-stage feedback scheme has extra nodes in the feedback branches to allow for the possibility of noise sources there.)

Proposition 3.2: If F_a and F_b are stable, then the stable feedback $F_{ab}\triangleq F_b-F_a$ makes the two-stage feedback scheme c.l. stable and equivalent to scheme (b).

Proof: The two-stage scheme is obviously equivalent to scheme (b), and is c.l. stable because its c.l. operators consist of: i) the c.l. operators K_{22}, K_{23}, K_{32}, and K_{33} of scheme (a) or scheme (b), which are stable by hypothesis, or ii) sums and products of the operators listed in i), and

[3] If P is unstable, the parameterization of c.l. stable schemes by a single operator Q is obviously still possible. However, some of our other conclusions, concerning existence of a feedback realization with stable elements, structural stability, or decomposition properties, may no longer be valid.

the operators F_a or F_b which are stable by hypotheses. Q.E.D.

The following *decomposition principle* can be obtained immediately from Theorem 1 and Proposition 3.2. Let $P_0 \in \mathbb{A}_s$ be any unstable plant stabilizable by a set of stable feedbacks in $\mathbb{B}_s$. *Any closed-loop stable feedback scheme employing a stable feedback around the plant P_0 is equivalent to a closed-loop stable scheme consisting of: i) a stabilizing feedback $F_0 \in \mathbb{B}_s$ which can be selected arbitrarily, followed by ii) a model reference scheme with stable operators Q and P_1.*

It follows that under our hypotheses,[3] and in particular under the assumption that plant and feedback are not simultaneously unstable, the problem of sensitivity reduction can be decomposed into two independent problems: stabilization followed by desensitization of a stable system. Henceforth, we shall confine ourselves to the second problem.

III′. Approaches to Feedback-Sensitivity Minimization

A′. Quadratic versus Induced Norms

The main properties of feedback cannot be deduced without some notion of uncertainty. Suppose that the disturbance d is uncertain but belongs to some subset $\mathfrak{D}$ of possible disturbances in $\mathfrak{X}$. From (3.6) it is clear that for disturbances to be attenuated, $(I-P_1Q)$ must be small on $\mathfrak{D}$, i.e., Q must act as an approximate inverse of P_1 on $\mathfrak{D}$. The various approaches to the disturbance attenuation problem are differentiated by the way in which uncertainty is described, and this approximate inversion is metricized and calculated.

A typical WHK approach in a deterministic version could be viewed as follows: $\mathfrak{D}$ consists of the set of disturbances d in $L_2(0,\infty)$ possessing a single, fixed, known power spectrum $|\hat{d}(j\omega)|$ in $L_2(-\infty,\infty)$, and the object of design is to find a filter Q that minimizes the quadratic distance $\|d-P_1Qd\|_{L_2}$, where in general Q depends on $|\hat{d}(j\omega)|$. In the stochastic analog of this problem $\mathfrak{D}$ is a random process characterized by probability-covariance functions and metricized by a quadratic norm. This description of uncertainty has certain limitations that we would like to circumvent, namely, the following.

1) The covariance properties of the random process must be known. In practice, they are often unknown elements of prescribed sets.

This is merely a limitation on the class of random processes for which WHK is valid. More serious from the point of view of feedback theory is the following observation.

2) The quadratic norm on plants employed in the WHK method lacks the multiplicative property $\|PQ\|_{L_2} \leqslant \|P\|_{L_2}\|Q\|_{L_2}$, and in general it may be difficult or impossible to estimate the norm of a product PQ from the norms of P and Q. The product norm $\|PQ\|_{L_2}$ may be large even though $\|P\|_{L_2}$ and $\|Q\|_{L_2}$ are small.

Consequently, if plant uncertainty is metricized by the quadratic norm, its propagation through products and inverses is hard to study. This is a serious limitation in feedback problems in which expressions such as $(I+PF)^{-1}$ play a major role.

By contrast, the "M_m" spec which is widely used in classical design measures the maximum frequency response magnitude, and is essentially the induced operator norm $\|P\| \triangleq \sup\{\|Px\|_{L_2}/\|x\|_{L_2}: x \in L_2\}$, which has the multiplicative property, and is therefore convenient to estimate in cascaded systems. By describing plant uncertainty in terms of a sphere of specified radius in a norm having such a multiplicative property, it is possible to obtain a general approach to problems involving disturbances/random processes which are unknown but belong to prescribed sets, as we shall see in Section IV.

Minimization of an induced norm in effect amounts to a minimax solution. Minimax methods do not necessarily represent uncertainty with greater fidelity than quadratic methods. However, the concern here is less with fidelity than with the ability to handle product systems.

B′. Constraints on the Norm of a Sensitivity Operator

It is natural then to try to pose sensitivity reduction problems in terms of the minimization of norm of the sensitivity operator, and to employ a norm having multiplicative properties. The primary norm of a Banach algebra has such properties, but the following propositions show it to be useless for this purpose.

Proposition 3.3: If P and Q are in a Banach algebra $\mathbb{B}$ and $\|I-PQ\|<1$, then PQ has an inverse in $\mathbb{B}$.

Proof: Denote $(I-PQ)$ by E. As $\|E\|<1$, the power series $I-E+E^2-\cdots$ converges to an operator which inverts $(I-E)$. As $(I-E)=PQ$, $(PQ)^{-1}$ exists. Q.E.D.

Proposition 3.3 has occasionally been interpreted as showing that invertibility is necessary for sensitivity reduction. This interpretation is empty. In fact, since strongly causal operators never have inverses in $\mathbb{B}$ [see Proposition 2.1b)], we have the following.

Corollary 3.4: If PQ is in $\mathbb{B}_s$, then $\|I-PQ\| \geqslant 1$.

It is impossible to make the sensitivity operator less than 1 in the original $\mathbb{B}$ norm. In H_0^∞ this simply means that the frequency response of PQ approaches 0 at infinite frequencies and $(I-PQ)$ approaches 1.

An obvious idea at this point is to make $(I-P_1Q)$ small in norm over some finite frequency band, i.e., over an invariant subspace. The next proposition shows that norms over invariant subspaces usually are not useful measures of sensitivity for optimization purposes. Let $\mathfrak{B}_1$ be a subspace of $\mathfrak{B}$, Π a projection operator onto $\mathfrak{B}_1$, and suppose that $\mathfrak{B}_1$ is invariant[4] under $\mathbb{B}$, i.e., $R\Pi=\Pi R\Pi$ for each R in $\mathbb{B}$. Let α denote the norm of $(I-P_1Q)$ restricted to the subspace $\mathfrak{B}_1$ and optimized over all Q, i.e., $\alpha \triangleq \inf_{Q\in\mathbb{B}_s}\sup\{\|(I-P_1Q)\Pi d\|: d\in\mathfrak{B} \text{ and } \|d\|=1\}$.

Proposition 3.5: For any P in $\mathbb{B}_s$, $\alpha=1$ or $\alpha=0$.

The proof is in Appendix I. If $\alpha=0$, sensitivity can be made arbitrarily small over the subspace $\mathfrak{B}_1$. In practice there are special cases involving "minimum phase" systems

[4] More generally, Proposition 3.5 holds if $\mathbb{B}$ is replaced by any of its norm preserving extensions. Note that Π is not necessarily in the algebra $\mathbb{B}$ or causal.

in which solutions that approach $\alpha=0$ may be useful. More typically, this result is achieved at the expense of increasing the sensitivity without bound on complements of $\mathcal{B}_1$; in such cases, the norm of a restriction of $(I-P_1Q)$ is not a candidate for minimization.

Corollary 3.4 and Proposition 3.5 delineate some of the peculiarities of the sensitivity optimization problem. In one form or another these peculiarities were recognized in the classical theory, and are probably the reason why it stopped short of optimization. We shall try to circumvent them by introducing an auxiliary (semi) norm to which they do not apply.

IV. Multiplicative Seminorms and Approximate Inverses

Uncertainty in a disturbance (or plant) in a linear space can be specified in terms of belonging to a ball of disturbances (or plants) centered at some nominal value, and of radius specified in some norm. Such a description of uncertainty may be cruder than a probabilistic description, but is usually more tractable in feedback problems.

One of the axioms of a norm asserts that only the zero element has zero norm. This axiom is often not needed, and with its elimination a norm is replaced by the slightly more general concept of a seminorm.

A ball in any seminorm can be shown to be a convex set.[4] Conversely, any convex set[5] in a linear space generates a seminorm (see Rudin [22, p. 24]) known as the Minkowski functional of that set. In linear spaces convex sets[4] of uncertainty can therefore always be described in terms of seminorms. We shall employ seminorms to obtain a systematic approach to such sets of uncertainty (cf. the objectives outlined in Section I-A).

In the next section, we shall define classes of left and right seminorms. To motivate the definitions, let us find seminorm descriptions for two disturbance sets which can be generated by the interaction of filters and certain "flat" disturbance sets.

Henceforth, W will denote a stable causal operator of unit norm, which will play the role of a weighting filter. For concreteness, W can be thought of as an operator in $\mathbb{H}_0^\infty$, W: $H^\infty \to H^\infty$, with response $\hat{w}(s)=k(s+k)^{-1}$. $\mathfrak{D}$ will denote a flat disturbance set (analogous to white noise) consisting of the unit ball in the space of inputs, in this case in H^∞:

$$\mathfrak{D}_1=\{d_1\in H^\infty: |\hat{d}_1(j\omega)|\leqslant 1\}$$

whose elements are frequency functions of unknown but bounded magnitude. Consider two situations.

1) Let $\mathfrak{D}$ be the set mapped by W into flat disturbances, i.e., $W\mathfrak{D}=\mathfrak{D}_1$. $\mathfrak{D}$ can be described as the unit ball in the seminorm $\|\cdot\|_\ell$, defined on H^∞ by the equation $\|d\|_\ell \triangleq \|Wd\|$.

2) Let $\mathfrak{D}$ be the set into which W maps flat disturbances, i.e., $W\mathfrak{D}_1=\mathfrak{D}$. $\mathfrak{D}$ is also the set

$$\mathfrak{D}=\{d\in \text{range}(W): |\hat{d}(j\omega)|\leqslant|\hat{w}(j\omega)|\}.$$

$\mathfrak{D}$ can be described as the unit ball in the seminorm $\|\cdot\|_r$ defined on the range of W (which is a proper subspace of H_0^∞) by the equation $\|d\|_r \triangleq \|W^{-1}d\|$.

There are many engineering problems in which the apriori information about disturbances is in the form of an upper bound to the magnitudes of their possible frequency responses. The seminorm description 2) is natural for such problems, and 1) occurs in inverse problems.

The seminorm $\|\cdot\|_r$ employs an up-weighting, and $\|\cdot\|_\ell$ employs a down-weighting. These two examples generalize into the notion of left and right seminorms, defined as follows.

A. *Seminorms for Inputs and Outputs*[6]

Let $\mathcal{Y}$ be a $\|\cdot\|$-normed linear space. Let $\|\cdot\|_\ell$ be any seminorm defined on all of $\mathcal{Y}$, and $\|\cdot\|_r$ a seminorm defined on some nontrivial subspace $\mathcal{Y}_r$ of $\mathcal{Y}$. The seminorm $\|\cdot\|_r$ is said to *dominate* $\|\cdot\|_\ell$ iff $\|y\|_\ell\leqslant\|y\|_r$ for all y in $\mathcal{Y}_r$; this dominance is denoted by $\|\cdot\|_\ell\leqslant\|\cdot\|_r$.

Definition: A *left seminorm* is any seminorm defined on all of $\mathcal{Y}$ with the property that $\|\cdot\|_\ell\leqslant\|\cdot\|$. A *right seminorm* is any seminorm defined on a nontrivial subspace $\mathcal{Y}_r$ of $\mathcal{Y}$ with the property that $\|\cdot\|\leqslant\|\cdot\|_r$.

For example, if W is any $\mathbb{H}^\infty$ filter of unit norm, the expressions $\|y\|_\ell\triangleq\|Wy\|$ and $\|y\|_r\triangleq\|W^{-1}y\|$ define left and right seminorms, on H^∞ and the range of W, respectively. The range of W is a subspace of H^∞, proper whenever $\hat{w}(s)$ has zeros in the right half-plane or at ∞.

B. *Weighted Seminorms for Plants*

Definition: A weighted seminorm is any seminorm $\|\cdot\|_w$ on the $\mathbb{B}$ with the property that $\|\cdot\|_W\leqslant\|\cdot\|$.

The terms "weighted" and "left" are synonymous. We shall use the term "weighted" to distinguish the left seminorms on $\mathbb{B}$ used as measures of plant sensitivity from the others.

A weighted seminorm on $\mathbb{B}$ is *induced* by a pair $(\|\cdot\|_\ell,\|\cdot\|_r)$, where $\|\cdot\|_\ell$ is a left seminorm on the space $\mathcal{B}$ (of outputs), and $\|\cdot\|_r$ is a right seminorm on a subspace $\mathcal{B}_r\subset\mathcal{B}$ (of inputs), iff $\|\cdot\|_W$ is defined for $A\in\mathbb{B}$ by the equation

$$\|A\|_W=\sup\{\|Au\|_\ell/\|u\|_r: u\in\mathcal{B}_r \text{ and } \|u\|_r\neq 0\}.$$

It follows that $\|I\|_W\leqslant 1$.

In control problems, weightings are often introduced by filters, which act on disturbances either before entering a plant or after leaving it. For example, let W_ℓ and W_r be linear mappings in $\mathfrak{X}\times\mathfrak{X}$, each of unit $\mathcal{B}$-induced norm. A

[5]Satisfying the following additional assumptions: 1) if x is in the linear space, then αx is in the set for some real α; 2) if y is in the set then so is $-y$.

[6]*Convention*: Whenever x belongs to a space on which several norms are defined, the unsubscripted norm $\|x\|$ denotes the principal norm. Weighted norms will be designated by subscripts.

left seminorm is defined on $\mathfrak{B}$ by the equation $\|y\|_\ell \triangleq \|W_\ell y\|$. Let $\mathfrak{B}_r$ be the range of W_r; if W is 1:1, a right seminorm is defined on $\mathfrak{B}_r$ by the equation $\|u\|_r \triangleq \|W_r^{-1}u\|$. The pair $(\|\cdot\|_r, \|\cdot\|_\ell)$ induces the weighted seminorm $\|A\|_W = \|W_\ell A W_r\|$ on the space $\mathbb{B}$.

Although weightings produced by filters will be emphasized in this paper, they can be produced by other means. For example, a weighted seminorm on $\mathbb{H}_0^\infty$ is given by the supremum over a shifted half-plane,

$$\|P\|_W = \sup\{|\hat{p}(s)|: \operatorname{Re}(s) \geqslant \alpha\}, \alpha > 0.$$

1) Multiplicative Seminorms—Symmetric Case: In general, weighted seminorms lack the multiplicative property $\|CD\|_W \leqslant \|C\|_W \|D\|_W$ of algebra norms. In problems involving the attenuation of a single disturbance (or single random process) this need not matter, as multiplications can be avoided. However, in problems involving plant uncertainty, closed-loop perturbations have the product form $(I-PQ)\Delta P$. We shall employ seminorms with weaker multiplicative properties suitable for such products.

Definition: A *symmetric (weighted) seminorm* on the algebra $\mathbb{B}$ is a weighted seminorm $\|\cdot\|_W$ on the space $\mathbb{B}$ which satisfies the multiplicative inequalities

$$\|CD\|_W \leqslant \|C\|_W \|D\|, \quad \|CD\|_W \leqslant \|C\| \cdot \|D\|_W. \quad (4.1)$$

Any operator $W: \mathfrak{X} \to \mathfrak{X}$ of unit $\mathfrak{B}$-induced norm which commutes with all operators of $\mathbb{B}$ defines a symmetric (weighted) seminorm by the equation $\|A\|_W \triangleq \|WA\| = \|AW\|$.

Symmetric seminorms have the property that $\|I\|_W = 1$.

2) Multiplicative Seminorms—General Case: In multivariable systems the plant perturbation ΔP always appears on the right of the product $(I-PQ)\Delta P$, and often does not commute with $(I-PQ)$. In such cases a more general class of multiplicative seminorms will be used. If ΔP is strictly causal, then the product $(I-PQ)\Delta P$ lies in a proper subspace of $\mathbb{B}$ (which is a left ideal, although we have no immediate use for this fact). With such products in mind, we make the following definition.

Definition: Let $\mathbb{B}_r$ be any subspace of $\mathbb{B}$, and $\mathbb{B} \cdot \mathbb{B}_r$ denote the space of products $\{CD: C \text{ in } \mathbb{B} \text{ and } D \text{ in } \mathbb{B}_r\}$. A *multiplicative seminorm* on the space of products $\mathbb{B} \cdot \mathbb{B}_r$ is any weighted seminorm on the space $\mathbb{B} \cdot \mathbb{B}_r$ with the following additional property: there is a left seminorm $\|\cdot\|_{W\ell}$ defined on $\mathbb{B}$, a seminorm $\|\cdot\|_{W1}$ defined on $\mathbb{B}_r$, and the inequality $\|CD\|_W \leqslant \|C\|_{W\ell} \|D\|_{W1}$ holds[7] for all C in $\mathbb{B}$ and D in $\mathbb{B}_r$.

Note that $\|\cdot\|_{W1}$ is not necessarily a right seminorm.

An example of a multiplicative seminorm is obtained as follows. Let $\|\cdot\|_W$ be the seminorm produced by operators W_ℓ and W_r in the example of Section IV-B [preceding Section IV-B1)]; $V: \mathfrak{X} \to \mathfrak{X}$ be any 1:1 map of unit norm; and $\mathbb{B}_r \triangleq \{VD_1: D_1 \in \mathbb{B}\}$. Define the seminorms $\|C\|_{W\ell} \triangleq \|W_\ell C V\|$ for $C \in \mathbb{B}$, and $\|D\|_{W1} \triangleq \|V^{-1} D W_r\|$ for $D \in \mathbb{B}_r$. Then, $\|\cdot\|_W$ has the multiplicative property claimed, as

$$\|CD\|_W = \|W_\ell C D W_r\| \leqslant \|W_\ell C V\| \cdot \|V^{-1} D W_r\|$$
$$= \|C\|_{W\ell} \|D\|_{W1}.$$

A *symmetric* seminorm can be viewed as a special case of a multiplicative seminorm on the space of products $\mathbb{B} \cdot \mathbb{B}_r$, in which $\mathbb{B}_r = \mathbb{B}$, and the multiplicative inequality holds for each of two pairs of seminorms, namely, $(\|\cdot\|_W, \|\cdot\|)$ and $(\|\cdot\|, \|\cdot\|_W)$.

C. Approximate Inverses and Singularity Measures

Many problems of feedback theory, both classical and modern, can be reduced to the construction of an approximate inverse. Let $\|\cdot\|_W$ be a fixed, weighted seminorm on the space $\mathbb{B}$.

Definition: For any operator P in $\mathbb{B}$, an *approximate right inverse* (in $\mathbb{B}_s$) of P is any operator Q in $\mathbb{B}_s$ for which[8] $\|I-PQ\|_W < \|I\|_W$; the *right singularity measure* (in $\mathbb{B}_s$) of P (under $\|\cdot\|$), denoted by $\mu(P)$, is

$$\mu(P) = \inf\{\|I-PQ\|_W: Q \text{ in } \mathbb{B}_s\}.$$

In general, $\mu(P)$ is a number in the interval $0 \leqslant \mu(P) \leqslant 1$. The last inequality follows from the observation that $Q=0$ gives $\|I-PQ\|_W = \|I\|_W$.

In all of the following $\mathbb{H}^\infty$ examples, $\|\cdot\|_W$ will be the symmetric weighted norm defined by $\|A\|_W \triangleq \|WA\|$ for $A \in \mathbb{H}^\infty$, where $W \in \mathbb{H}_0^\infty$ is a fixed (strictly causal) operator of unit norm. For example, W can be the "low-pass" frequency response $\hat{w}(s) = k(s+k)^{-1}$, $k>0$. As $\|\cdot\|_W$ is symmetric, it has the multiplicative properties defined in Section IV-B1).

Example 4.1: P_a is a plant in $\mathbb{H}_0^\infty$ with frequency response $\hat{p}_a(s) = \alpha(s+\alpha)^{-1}$, $\alpha > 0$. The sequence of operators Q_n in $\mathbb{H}_0^\infty$ with frequency responses $\hat{q}_n(s) = \alpha^{-1}(s+\alpha) \cdot n^2(s+n)^{-2}$, $n = 1, 2, \cdots$, satisfies the equation

$$\|\hat{w}(1-\hat{p}\hat{q}_n)\|_{H^\infty} = \sup_{\operatorname{Re}(s)\geqslant 0} \left|\hat{w}(s)\left(1 - \frac{n^2}{(s+n)^2}\right)\right|. \quad (4.1)$$

The right-hand side (RHS) of (4.1) approaches 0 as $n \to \infty$. Therefore, the singularity measure in $\mathbb{H}_0^\infty$ of P_a under $\|\cdot\|_W$ is $\mu(P_a) = 0$. The operators Q_n are approximate inverses, and the sequence Q_n is an example of what will be called an inverting sequence.

Example 4.2: P_a is the operator of Example 4.1; P_b is the "nonminimum phase" operator in $\mathbb{H}^\infty$ with frequency response $\hat{p}_b(s) = (\beta - s)(\beta + s)^{-1}$, $\beta > 0$; P_1 is the product, $P_1 \triangleq P_b P_a$. For any Q in $\mathbb{H}_0^\infty$, $\hat{w}(s)[1 - \hat{p}_1(s)\hat{q}(s)]$ has the value $\hat{w}(\beta)$ at the zero of $\hat{p}_b(s)$. Therefore, $\|I - P_1 Q\|_W \geqslant |\hat{w}(\beta)|$, and we get the lower bound $\mu(P_1) \geqslant |\hat{w}(\beta)|$. In

[7]The definition can obviously be generalized for the case of perturbations appearing on the left.

[8]Recall that $\|I\|_W \leqslant 1$, and $\|I\|_W = 1$ if $\|\cdot\|_W$ is symmetric.

Section V-A and Corollary 6.1 it will be established that in fact $\mu(P_1)=|\hat{w}(\beta)|$.

In these examples we have emphasized approximate inverses under a multiplicative seminorm. In passing, it may be worth mentioning that WHK problems can be viewed as approximate inversion problems in which the weighted seminorm of $(I-PQ)$ is obtained by weighting by a fixed vector $d\in L^2(0,\infty)$, to obtain $\|I-PQ\|_W=\|(I-PQ)d\|_{L^2}$. Here $\mu(P)$ is the irreducible error. However, $\|\cdot\|_W$ lacks the multiplicative properties.

No matter which seminorm is used $\mu(\cdot)$ has the following property.

Proposition 4.3: For any P_a and P_b in $\mathbb{B}$, $\mu(P_bP_a)\geqslant\mu(P_b)$.

Proof: If the contrary is assumed to be true, there is a Q in $\mathbb{B}_s$ for which $\|I-P_bP_aQ\|_W<\mu(P_b)$; P_aQ now acts as an approximate inverse for P_b, and there is a contradiction.

V. Sensitivity to Disturbances and Approximate Invertibility

We shall show that approximate invertibility of the plant is a necessary and sufficient condition for the existence of a feedback to attenuate disturbances, and the optimal sensitivity depends on the measure of singularity of the plant.

Consider the feedback scheme, (3.1), and suppose the plant P in $\mathbb{B}$ equals the nominal P_1. Let $\|\cdot\|_r$ be a fixed right seminorm defined on some subspace $\mathfrak{B}_r$ of (inputs), $\|\cdot\|_\ell$ a fixed left seminorm on (outputs) $\mathfrak{B}$, and $\|\cdot\|_W$ the resulting weighted seminorm induced on $\mathbb{B}$. The *sensitivity to disturbances* η_1 is defined by the equation $\eta_1\triangleq\sup\{\|y\|_\ell/\|d\|_r: d \text{ in } \mathbb{B}_r,\ \|d\|_r\neq 0\}$. From (3.2a) it follows that $\eta_1=\|(I+P_1F)^{-1}\|_W$. Whenever the equivalence equations (3.4) hold, η_1 also equals $\|I-P_1Q\|_W$.

The sensitivity η_0 obtained when $F=0$ will be called the *open-loop sensitivity*. As $\eta_0=\|I\|_W$, $\eta_0\leqslant 1$, and $\eta_0=1$ whenever $\|\cdot\|_W$ is symmetric.

Theorem 2:

a) A necessary and sufficient condition for the existence of a feedback F in $\mathbb{A}_s$ for which (3.1) is closed-loop stable and the sensitivity η_1 is less than the open-loop value η_0, is that P_1 have an approximate (stable) right inverse Q in the algebra[9] $\mathbb{B}_s$.

b) For any $\epsilon>0$, there is a feedback F in $\mathbb{A}_s$ for which (3.1) is closed-loop stable and for which $\eta_1<\mu(P_1)+\epsilon$, but no F for which (3.1) is closed-loop stable and $\eta_1<\mu(P_1)$, where $\mu(P_1)$ is the measure of right singularity in $\mathbb{B}_s$ of P_1 under $\|\cdot\|_W$.

Proof: By Theorem 1, any closed-loop stable feedback scheme (3.1) with F in $\mathbb{A}_s$ is equivalent to a model reference scheme (3.3) with Q in $\mathbb{B}_s$. Let (3.3) be equivalent to (3.1). η_1 can be expressed by $\eta_1=\|I-P_1Q\|_W$. Therefore, a) $\eta_1<\eta_0$ iff Q is an approximate right inverse of P_1. Also, b) by definition of $\mu(P_1)$, there is a Q in $\mathbb{B}_s$ for which $\|I-P_1Q\|_W<\mu(P_1)+\epsilon$, but none for which $\|I-P_1Q\|_W<\mu(P_1)$. The conclusion concerning F follows from the equivalence of (3.1) and (3.3). Q.E.D.

In general, it may be impossible to attain the sensitivity $\mu(P_1)$.

Definition: A sequence $Q_n\in\mathbb{A}$, $n=1,2,\cdots$, will be called *optimal* for P_1 iff the sequence of sensitivities $\|I-P_1Q_n\|_W$ approaches $\mu(P_1)$ as $n\to\infty$. $\mu(P_1)$ will be called the *optimal sensitivity* for P_1.

A sequence of feedbacks $F_n\in\mathbb{A}_s$ will be called optimal iff the equivalent comparators Q_n are optimal.

Remark: If the disturbance d lies in a balanced set $\mathfrak{D}$ (i.e., d in $\mathfrak{D}$ implies $-d$ in $\mathfrak{D}$), it is simple to show that no open-loop control, obtained by letting $Q=0$ and applying an input at node 2, can make $\|y\|_\ell$ less than $\|d\|_\ell$ for all d in $\mathfrak{D}$. It follows from Theorem 2 that, for right invertible plants, optimal sensitivity achievable with feedback is smaller than without; in other words, ability of control schemes to cope with unknown disturbances depends on their configuration. This can be viewed as a continuation of the internal model principle [16] to seminormed disturbances.

A. An Example of Sensitivity Optimization

We would like to show that feedback optimization is feasible under a seminorm that has the multiplicative properties (4.1), unlike the quadratic norm of WHK methods. If weighting is obtained from a filter, the optimal feedback in $\mathbb{H}^\infty$ resembles a classical lead–lag network. We shall try to demonstrate these points by an example. A more comprehensive theory of H^∞ optimization would take too long to present here.

Let P_1 in $\mathbb{H}_0^\infty$ be any plant with a single $\mathrm{Re}(s)>0$ zero at $s=\beta$, subject to the high frequency restriction $|\omega\hat{p}_1(j\omega)|^{-1}\leqslant$const. for $|\omega|>1$. Let $\|\cdot\|_W$ be the weighted seminorm, $\|P\|_W\triangleq\|WP\|$, where $\hat{w}$ in H_0^∞ satisfies the condition $|\omega^\ell\hat{w}(j\omega)|^{-1}\leqslant$const. for some integer ℓ. An optimal sequence of comparators Q_n in $\mathbb{H}_0^\infty$ is sought.

It will be shown that *a sequence with frequency responses*

$$\hat{q}_m(s)=c_mP_b^{-1}(s)\left[1-\hat{w}(\beta)\hat{w}^{-1}(s)\right](s+m)^{-\ell}(s+n_m)^{-2},\qquad m,n_m=1,2,\cdots \tag{5.1}$$

in which: P_b *is the operator of Example 4.2*; $c_m=m^ln_m^2$; *and for each* m, n_m *is a sufficiently large integer, is such an optimal sequence. In fact*, $\mu(P_1)=|\hat{w}(\beta)|$, *and the sequence of sensitivities* $\|I-P_1Q_m\|_W$ *approaches* $\mu(P_1)$ *as* $m\to\infty$.

Q_m will be constructed here, but some of the proof details will be postponed until Section VI. P_1 can be factored into the product[10] $P_1=P_aP_b$, in which $\hat{p}_b(s)=(\beta-s)(\beta+s)^{-1}$, and $|\hat{p}_b(j\omega)|=1$; and $\hat{p}_a$ is in H_0^∞, $\hat{p}_a(j\omega)=$

[9] The fact that given $P\in\mathbb{B}$, the closed-loop system is stable iff $F(I+P_1F)^{-1}$ is stable is pointed out by Desoer–Chan [21, Theorem 3], who scrutinize the relationship between open- and closed-loop stability in the context of convolution algebras.

[10] P_a/P_b are often referred to as all-pass/minimum phase or inner/outer factors.

$\hat{p}_1(j\omega)/\hat{p}_b(j\omega)$ for all real ω, and $\hat{p}_a$ has no zeros in $\mathrm{Re}(s)\geqslant 0$. Optimal sequences will be constructed for P_a and P_b separately.

P_b coincides with the operator of Example 4.2. By the reasoning of that example we get the lower bound $\mu(P_1)\geqslant |\hat{w}(\beta)|$; in fact, for the $\hat{p}_b$ factor alone, $\|\hat{w}(1-\hat{p}_b\hat{q})\|_{H^\infty}\geqslant |\hat{w}(\beta)|$ whenever $\hat{q}$ preserves the analyticity of $\hat{w}\hat{p}_b\hat{q}$ in $\mathrm{Re}(s)\geqslant 0$.

We observe next that the function $\hat{q}_b$: $\mathbb{C}\to\mathbb{C}$, $\hat{q}_b \triangleq \hat{p}_b^{-1}(s)[1-\hat{w}(\beta)\hat{w}^{-1}(s)]$ exactly minimizes $\|\hat{w}(1-\hat{p}_b\hat{q}_b)\|$, since $\hat{w}(s)[1-\hat{p}_b(s)\hat{q}_b(s)]$ equals the lower bound $\hat{w}(\beta)$ for all s in $\mathbb{C}$. Also, $\hat{q}_b$ is analytic in $\mathrm{Re}(s)\geqslant 0$, as $[1-\hat{w}(\beta)\hat{w}^{-1}(s)]$ has a zero at β to cancel the pole $\hat{p}_b^{-1}$. We now combine $\hat{q}_b$ with enough high frequency attenuation to obtain an H^∞ sequence:

$$\hat{q}_{bm}(s)=\hat{p}_b^{-1}(s)\left[1-\hat{w}(\beta)\hat{w}^{-1}(s)\right]m^\ell(s+m)^{-\ell}\qquad m=1,2,\cdots.$$

(It will become apparent that the attenuation $m\ell(s+m)^{-\ell}$ will be at a high enough frequency to have arbitrarily little effect on weighted sensitivity.) The factor $P_a(s)$ can be inverted by the sequence in H_0^∞,

$$\hat{q}_{an}(s)=\hat{p}_a^{-1}(s)n^2(s+n)^{-2},\qquad n=1,2,\cdots.$$

The sequence $\hat{q}_m$ of (5.1) is constructed from the product of $\hat{q}_{an}$ and $\hat{q}_{bm}$. The validity of this construction is established in Corollary 6.1.

Remarks: It may be worth looking at a special case of (5.1) to get a better feel for the kind of feedbacks our approach generates. Let $\hat{p}_a(s)=\alpha(s+\alpha)^{-1}$ and $\hat{w}(s)=k(s+k)^{-1}$, $\alpha>0$, $k>0$. Equation (5.1) gives

$$\hat{q}_m(s)=c_m(s+\beta)(s+m)^{-1}(s+n_m)^{-2}. \tag{5.2}$$

$c_m \triangleq m^2n_m(k+\beta)^{-1}$. This optimal $\hat{q}_m$ consists of the "lead" factor $c(s+\beta)$, and high frequency poles whose purpose is to make $\hat{q}_m(s)$ strictly proper. The feedback law $\hat{f}(s)$ produced by $\hat{q}_m(s)$ [via (3.4b)] is a lead–lag network typical of classical control.

If the high frequency poles are neglected, the sensitivity operator $E\triangleq I-P_1Q_m$ has the frequency response $\hat{e}(j\omega)=\hat{w}(\beta)\hat{w}^{-1}(s)$; $\hat{e}$ is small at highly weighted frequencies and vice versa. If $\hat{e}(j\omega)$ is compared to the value that would be obtained by letting $\hat{q}(s)=1$, it appears that $\hat{q}_m(s)$ has the effect of trading undesirable low-frequency phase lags $\arg\hat{p}(j\omega)$, for undesirably large magnitudes $|\hat{p}(j\omega)\hat{q}(j\omega)|$ at high frequencies, where they matter less, at least if the specified weighting is correct. This, too, is a typical strategy of classical design.

The growth of $|\hat{e}(j\omega)|$ as $\omega\to\infty$ depends on the decay of $|\hat{w}(j\omega)|$; if the former is too high, the choice of the latter was inappropriate, if $\|\hat{w}(j\omega)|$ is bounded from below, then $|\hat{e}(j\omega)|$ is bounded from above. The dependence of the optimal filter on the weighting W is not surprising, as W describes the convex set of disturbances to be attenuated.

Filters optimal in this sense are known to be very sensitive to plant uncertainty, and are practical only where accurate plant models are available.

B. Unstable Plants

If P_0 is an unstable plant in $\mathbb{A}_s$ with output disturbance d_0 in $\mathfrak{B}$, and there is a *stabilizing feedback* F_0 with the property that $P_0(I+F_0P_0)^{-1}$ and $(I+P_0F_0)^{-1}$ are in $\mathbb{B}$, then we can let P_1 in Theorem 2 be the stabilized system $P_0(I+F_0P_0)^{-1}$, and d be the stabilized disturbance $(I+P_0F_0)^{-1}d_0$. In that case, (3.6) takes the form

$$y=(I-P_1Q)(I+P_0F_0)^{-1}d_0. \tag{5.3}$$

The term $(I+P_0F_0)^{-1}$ contributed by stabilization appears in (5.3) as an extra right weighting on $(I-P_1Q)$. Provided $\|\cdot\|_W$ is modified to inlude the extra weighting, Theorem 2 remains valid.

Remark: We have preferred to separate feedback synthesis into two consecutive stages: 1) stabilization and 2) desensitization of (input–output) stable systems. From the point of view of input–output sensitivity theory, the separation is in a certain sense unavoidable, as the perturbations allowed for robust stabilization are radically different from those for desensitization. This point is elaborated in Appendix II.

C. Optimal Sequences: Symmetric Case

Usually, optimal filters for sensitivity reduction are not strictly proper and can not be attained, although they can be approached by sequences of filters of increasing bandwidth. The behavior of such sequences is conveniently described in terms of the concept of an "identity sequence," or "approximate identity" drawn from Banach algebras. Here we shall summarize those properties of identity sequences which are employed in our filter construction.

Let $\|\cdot\|_W$ be a fixed, symmetric[11] weighted seminorm on the algebra $\mathbb{B}$. A *weighted identity sequence* (widseq) is any sequence I_m, $m=1,2,\cdots$ in $\mathbb{B}$ with the property that $\|I_m\|=1$ and, for any A in $\mathbb{B}$, $\|A-I_nA\|_W$ and $\|A-AI_n\|_W$ approach 0 as $n\to\infty$. For any P in $\mathbb{A}$, a sequence Q_n, $n=1,2,\cdots$, is a *right weighted inverting sequence* (winvseq) iff PQ_n is a widseq.

Whenever $\|I-PQ_n\|_W\to 0$ as $n\to\infty$, Q_n must be a right winvseq for P; for then, the inequalities

$$\|A-PQ_nA\|_W\leqslant\|I-PQ_n\|_W\|A\|,$$
$$\|A-APQ_n\|_W\leqslant\|A\|\cdot\|I-PQ_n\|_W,$$

obtained using the multiplicative property of symmetric seminorms, have left-hand sides which converge to 0 as $n\to\infty$.

For example, in $\mathbb{H}^\infty$ the sequence of operators with frequency responses $n^r(s+n)^{-r}$, $n=1,2,\cdots$ and $r>0$ any constant integer, is a widseq for the weighted norm of

[11]Without symmetry a distinction between left and right identity sequences, etc., must be made.

Example 4.1. The sequence Q_n of that example is a winvseq for P_a.

It will be shown that in certain cases an optimal sequence of approximate inverses for a product P_bP_a can be obtained from separate sequences for the factors P_a and P_b, and consequently some optimization problems can be decomposed into simpler ones. Let $\|\cdot\|_a \geq \|\cdot\|_b$ be a pair of symmetric, weighted seminorms on the algebra $\mathbb{B}$. Let P_a in $\mathbb{B}_s$ and P_b in $\mathbb{B}$ have singularity measures $\mu_a(P_a)$ and $\mu_b(P_b)$ under $\|\cdot\|_a$ and $\|\cdot\|_b$, respectively, and optimal sequences $\{Q_{am}\}_{m=1}^{\infty}$ in $\mathbb{B}_s$, $\{Q_{bn}\}_{n=1}^{\infty}$ in $\mathbb{B}$, i.e., $\|I-P_iQ_{in}\|_i \to \mu_i(P_i)$ as $n\to\infty$ for $i=a$ or b.

Lemma 5.1: If $\mu_a(P_a)=0$, then $\mu_b(P_bP_a)=\mu_b(P_b)$, and P_bP_a has an optimal sequence of $\|\cdot\|_b$-weighted right approximate inverses in $\mathbb{B}_s$ of the form $Q_{an_m}Q_{bm}$, where for any $m=1,2,\cdots,n_m$ is a sufficiently large integer.

In fact, for any $\epsilon>0$ and $0<\alpha<1$, if we pick integers m and n_m such that $\|I-P_bQ_{bm}\|_b \leq \mu_b(P_b)+\alpha\epsilon$, and $\|I-P_aQ_{an_m}\|_a \leq (1-\alpha)\epsilon\|P_b\|^{-1}\|Q_{bm}\|^{-1}$, then

$$\|I-P_bP_aQ_{an_m}Q_{bm}\|_b \leq \mu_b(P_b)+\epsilon. \tag{5.4}$$

Proof:

a) $\mu_b(P_bP_a)\geq\mu_b(P_b)$ by Proposition 4.3. Let us establish the opposite inequality. For any $m>0$, $n>0$, we have

$$\begin{aligned}\|I-P_bP_aQ_{an}Q_{bm}\|_b &= \|I-P_bQ_{bm}-P_n(P_aQ_{an}-I)Q_{bm}\|_b\\ &\leq \|I-P_bQ_{bm}\|_b+\|P_b\|\cdot\|P_aQ_{an}-I\|_a\|Q_{bm}\|\end{aligned} \tag{5.5}$$

where the triangle inequality, the multiplicative property of symmetric seminorms, and the dominance $\|\cdot\|_b\leq\|\cdot\|_a$ have been used. The integers m and n specified in the hypothesis exist by definition of $\mu_b(P_b)$ and by the hypothesis that $\mu_a(P_a)=0$. (5.4) now follows from (5.5). As ϵ was arbitrary, (5.4) implies that $\mu_b(P_bP_a)\leq\mu_b(P_b)$. Therefore, $\mu_b(P_bP_a)=\mu_b(P_b)$, and $Q_{an_m}Q_m$ is an optimal sequence. Q.E.D.

VI. Multivariable Systems in $\mathbb{H}^{\infty N}$

At present there is much interest in the design of multivariable systems along classical lines, but there is little formal theory. Our next objective is to show that the theory outlined in Sections III–V provides a suitable framework for multivariable design.

Multivariable frequency responses will be viewed as elements of an n-dimensional version of the Hardy space H^{∞}. Our approach will be to decompose the frequency response matrix into a product of nearly invertible and noninvertible parts, and to design a feedback for each part separately. The validity of approaches based on decompositions will be established in Theorem 3. The optimal sensitivity scheme of Section V-A will be obtained as a corollary. Another corollary will validate the common hypothesis that sensitivity can be made arbitrarily small if the plant is nearly invertible. Finally, a lower bound to sensitivity in terms of the location of RHP zeros will be derived.

A. *Terminology*

N is a fixed integer. H^{2N} is the Banach space consisting of the N-fold product of H^2, on which the norm of any vector $\hat{u}=(\hat{u}_1,\hat{u}_2,\cdots,\hat{u}_N)$ is $\|\hat{u}\| \triangleq (\Sigma_{i=1}^N\|\hat{u}_i\|^2)^{1/2}$. The elements of H^{2N} can be represented by columns of input or output Laplace transforms.

$H^{\infty N}$ is the algebra of $N\times N$ matrices, $\hat{P}=[\hat{p}_{ij}]$, whose elements $\hat{p}_{ij}$ are causal, stable, frequency responses in H^{∞}. For any $\hat{P}$ in $H^{\infty N}$ and s in $\mathbb{C}$, $\sigma[\hat{P}(s)]$ denotes the square root of the magnitude of the largest eigenvalue of $[\hat{P}(s)]^*\hat{P}(s)$, where $*$ denotes the conjugate transpose of a matrix. $H^{\infty N}$ is a Banach algebra under the norm $\|\hat{P}\|=\sup_{\omega}\sigma[\hat{P}(j\omega)]$.

Each frequency response matrix $\hat{P}$ in $H^{\infty N}$ determines an operator P: $H^{2N}\to H^{2N}$, $\widehat{Pu}=\hat{P}\hat{u}$. The algebra of such operators is denoted by $\mathbb{H}^{\infty N}$, and is a Banach algebra under the norm $\|P\|_{\mathbb{H}^{\infty N}}=\|\hat{P}\|_{H^{\infty N}}$ The norm defined in this way coincides with the norm induced on $\mathbb{H}^{\infty N}$ by H^{2N}, i.e., $\sup\{\|Pu\|: u \text{ in } H^{2N} \text{ and } \|u\|=1\}$.

A frequency response $\hat{P}$ in $H^{\infty N}$ is called *strictly proper* iff each of its components $\hat{p}_{ij}$ is strictly proper; the corresponding operator P in $H^{\infty N}$ is called *strictly causal.* The subalgebras of strictly proper or strictly causal elements in $H^{\infty N}$ or $\mathbb{H}^{\infty N}$ are denoted by $H_0^{\infty N}$ or $\mathbb{H}_0^{\infty N}$, respectively.

Let $\hat{w}$ be a (scalar) weighting function in H_0^{∞}. The weighted seminorm $\|P\|_W \triangleq \|\hat{w}\hat{P}\|_{H^{\infty N}}$ is defined on $\mathbb{H}^{\infty}$. Observe that $\hat{w}$ commutes with all operators in $\mathbb{H}^{\infty N}$, so that $\|\cdot\|_W$ is symmetric, and the multiplicative inequalities (4.1) are satisfied.

B. *A Decomposition Theorem*

We shall be interested in frequency response matrices whose rate of approach to zero as $s\to\infty$ is comparable to that of a power of $(1+s)$, and employ the following.

Notation: For any integer n, $(1+s)^nH^{\infty N}$ denotes the set of functions $\hat{P}(s)$ with the property that $(1+s)^{-n}\hat{P}(s)$ is in $H^{\infty N}$.

In the rest of Section VI, plants will be decomposed into products P_aP_b of strictly causal and approximately invertible parts under the following *assumption*:

a) P_a is a strictly causal operator in $\mathbb{H}_0^{\infty N}$; and

b) P_b is in $\mathbb{H}^{\infty N}$. The frequency response $\hat{P}_b$ has a *minimal* approximate right inverse $\hat{Q}_b$ in $(1+s)^fH^{\infty N}$; i.e., $\hat{w}\hat{P}_b\hat{Q}_b$ is in $H^{\infty N}$ and $\|\hat{w}(1-\hat{P}_b\hat{Q}_b)\|_{H^{\infty N}}=\mu(P_b)$.

Theorem 3: If i) $\det\hat{P}_a(s)\neq 0$ for $\mathrm{Re}(s)\geq 0$ and ii) there are constants $c>0$, $\rho>0$, and an integer $k>0$ for which the inequality $|s|^k\sigma[\hat{P}_a(s)]\geq c$ is valid in the region $|s|>\rho$, $\mathrm{Re}(s)\geq 0$, then

a) $\mu(P_bP_a)=\mu(P_b)$, and

b) P_bP_a has an optimal sequence of approximate right inverses Q_m in $\mathbb{H}_0^{\infty N}$, with frequency responses

$$\hat{Q}_m(s) \triangleq \hat{P}_a^{-1}(s)\hat{Q}_b(s)\left[m(s+m)^{-1}\right]^f\left[n_m(s+n_m)^{-1}\right]^{k+1}$$

where for any $m=1,2,\cdots,n_m$ is a sufficiently large integer (given explicitly in Lemma 5.1).

Theorem 3 is an application of Lemma 5.1. Q_m will be constructed out of separate optimal sequences for P_a and P_b, both under the symmetric $\|\cdot\|_W$ norm. First, however, a lemma will be proved.

For any integer $r>0$, let J_n^r denote the sequence in $\mathbb{H}^{\infty N}$ with "low-pass" frequency responses

$$\hat{J}_n^r(s)=n^r(s+n)^{-r}\hat{I}, \qquad n=1,2,\cdots.$$

Lemma 6.1: $\|J_n^r\|=1$; J_n^r is a weighted identity sequence,[12] i.e., for any $r>0$, $\|I-J_n^r\|_W\to 0$ as $n\to\infty$.

Proof: First of all, $\|J_n^r\|=\|I\|\sup_\omega|n(j\omega+n)^{-1}|^r=1$. Second, we have

$$\|I-J_n^r\|_W=\sup_\omega|\hat{w}(j\omega)|\cdot\left|1-\left(\frac{n}{j\omega+n}\right)^r\right| \tag{6.1}$$

$$\leqslant 2\sup_{|\omega|\geqslant\delta}|\hat{w}(j\omega)|+\|\hat{w}\|\sup_{|\omega|<\delta}\left|1-\left(\frac{n}{j\omega+n}\right)^r\right| \tag{6.2}$$

$\delta>0$ being any number. Let $\epsilon>0$ be given. As $\hat{w}$ is in H_0^∞, there is a $\delta>0$ for which the first term in (6.2) is less than $\epsilon/2$; for fixed δ there is an integer n for which the second term is less than $\epsilon/2$; i.e., $\lim_{n\to\infty}\|I-J_n^r\|_W=0$. Q.E.D.

Proof of Theorem 3: First, let us show that the sequence Q_{an} with frequency responses

$$\hat{Q}_{an}\triangleq\hat{P}_a^{-1}\hat{J}_n^{k+1}, \qquad n=1,2,\cdots \tag{6.3}$$

is a right weighted inverting sequence in $\mathbb{H}_0^{\infty N}$ for P_a. As $\det(\hat{P}_a(s))\neq 0$, $\hat{P}_a^{-1}(s)$ is analytic in $\mathrm{Re}(s)\geqslant 0$. The inequality $|s|^k\sigma[\hat{P}_a(s)]\geqslant c$ ensures that the functions defined by (6.3) are in $H_0^{\infty N}$. Now, $\|I-P_aQ_{an}\|_W=\|I-J_n^{k+1}\|_W\to 0$ as $n\to\infty$, by Lemma 6.1. Therefore, $\mu(P_a)=0$ and Q_{an} is a right winvseq in $\mathbb{H}_0^{\infty N}$ for P_a.

Next, consider P_b. By hypothesis, $\hat{P}_b$ has a minimal approximate right inverse $\hat{Q}_b$ in $(1+s)^\ell H^{\infty N}$. The frequency responses

$$\hat{Q}_{bm}\triangleq\hat{Q}_b\hat{J}_m^\ell, \qquad m=1,2,\cdots \tag{6.4}$$

lie in $H^{\infty N}$, and determine operators Q_{bm} in $\mathbb{H}^{\infty N}$. Let us show that Q_{bm} is optimal. Certainly, $\|I-P_bQ_{bm}\|_W\geqslant\mu(P_b)$. Now,

$$\|I-P_bQ_{bm}\|_W=\|(I-P_bQ_b)J_m^\ell+I-J_m^\ell\|_W$$

$$\leqslant\|I-P_bQ_b\|_W\|J_m^{\ell+1}\|+\|I-J_m^\ell\|_W$$

by the triangle inequality and the multiplicative property of symmetric seminorms. As $\|I-P_bQ_b\|_W=\mu(P_b)$ by hypothesis, and by Lemma 6.1 $\|J_m^{\ell+1}\|=1$ and $\lim_{m\to\infty}\|I-J_m^\ell\|_W\to 0$, we obtain $\lim_{m\to\infty}\|I-P_bQ_{bm}\|_W\leqslant\mu(P_b)$. As this upper bound coincides with the lower bound obtained above, $\lim_{m\to\infty}\|I-P_bQ_{bm}\|_W=\mu(P_b)$, i.e., Q_{bm} is a right winvseq in $\mathbb{H}^{\infty N}$ for P_b.

[12] For a symmetric weighting, $\|A-J_n^rA\|_W\leqslant\|I-J_n^r\|_W\|A\|$, so $\|A-J_n^rA\|_W\to 0$ whenever $\|I-J_n^r\|_W$ does; similarly for $\|A-AJ_n^r\|_W$. Therefore, J_n^r is a widseq iff $\lim_{n\to\infty}\|I-J_n^r\|_W=0$.

The conclusions of Theorem 3 are now true by Lemma 5.1. Here $\|\cdot\|_a=\|\cdot\|_b=\|\cdot\|_W$, and Q_m is the product $P_a^{-1}J_{n_m}^{k+1}Q_bJ_m^\ell$. Q.E.D.

Corollary 6.1: The conclusion of the example of Section V-A is true.[13] That example is simply a special case of Theorem 3 for $N=k=1$, and $\hat{q}_b=\hat{p}_b^{-1}(1-\hat{w}(\beta)\hat{w}^{-1})$ in $(1+s)^\ell H^\infty$.

C. Sensitivity Reduction in Nearly Invertible Multivariable Systems

Consider the feedback scheme of equations (3.1)–(3.3). Suppose that the plant $P_1(s)$ is in $\mathbb{H}_0^{\infty N}$ and satisfies the restrictions: $\det P_1(s)\neq 0$ for $\mathrm{Re}(s)\geqslant 0$, and $|s|^k\sigma[\hat{P}_1(s)]>c$ in some region $|s|>p$, $\mathrm{Re}(s)\geqslant 0$, where c, p are constants and k is an integer. The sensitivity $\eta\triangleq\|I-P_1Q\|_W$ is defined[14] as in Section V.

Corollary 6.2:[15] The sensitivity $\|I-P_1Q_n\|_W$ can be made smaller than any $\epsilon>0$ by a comparator Q_n in $\mathbb{H}_0^{\infty N}$ with frequency response

$$\hat{Q}_n(s)=\hat{p}_1^{-1}(s)\left[n(s+n)^{-1}\right]^{k+1}$$

n being a (sufficiently large) integer.

Proof: The hypotheses of Theorem 3 are satisfied, with $P_a=P_1$ and $P_b=I$. Therefore, $\mu(P_1)=0$ and $\|I-P_1Q_n\|_W\to 0$ as $n\to\infty$. Q.E.D.

D. Lower Bounds to Sensitivity

It might be expected that the singularity measure of an $H^{\infty N}$ frequency response matrix would be limited by the location of its RHP zeros, and that the optimal sensitivity would be similarly limited. The following theorem shows this to be true.

Let $\|\cdot\|_W$ be any weighted seminorm on $\mathbb{H}^{\infty N}$ of the form $\|P\|_W=\|\hat{w}\hat{P}\|$, $\hat{w}$ being (a scalar function) in H_0^∞. For any plant P in $\mathbb{H}^{\infty N}$, let the *RHP zeros* of P be the points s_i in $\mathrm{Re}(s)\geqslant 0$ $i=1,2,\cdots$, at which $\det[\hat{P}(s_i)]=0$.

Theorem 4:

$$\mu(P)\geqslant\sup\{|\hat{w}(s_i)|: i=1,2,\cdots\}.$$

Proof: For any Q in $\mathbb{H}^{\infty N}$ let $E\triangleq(I-PQ)$. For any RHP zero s_i, let ξ be a unit vector in the nullspace of the matrix $\hat{P}(s_i)\hat{Q}(s_i)$. Let f: $C\to C$ be the function $f(s)=\xi^*\hat{w}(s)\hat{E}(s)\xi$. Since $\hat{P}(s_i)\hat{Q}(s_i)=0$, $f(s_i)=\xi^*\xi\hat{w}(s_i)=\hat{w}(s_i)$. Now $f(s)$ is an H_0^∞ function, and by the maximum modulus principle attains its maximum on the $j\omega$ axis, i.e., $|f(j\omega)|\geqslant|\hat{w}(s_i)|$ for some ω. But for any transformation A in Euclidean N-space, and any unit N-vector ξ, $\sigma(A)\geqslant\xi^*A\xi$. The last assertion is established by the inequalities

[13] For simplicity we have considered inputs in H^2 here, but the conclusion is easily extended to H^p, $1\leqslant p\leqslant\infty$.

[14] Recall that sensitivity of a feedback scheme equals $\|I-P_1Q\|_W$, by Theorem 1.

[15] This corollary is based on results obtained with D. Bensoussan [17].

$$|\xi^* A\xi|^2 \leqslant |\xi^*|^2 |A\xi|^2 = |A\xi|^2 = (A^*A\xi)^*\xi \leqslant |A^*A\xi|\cdot|\xi| \leqslant \sigma^2(A)|\xi|^2 = \sigma^2(A).$$

We employ this fact to obtain

$$\mu(P) \geqslant \|E\| = \sup_\omega \sigma\left[\hat{w}(j\omega)\hat{E}(j\omega)\right] \geqslant \sup_\omega |f(j\omega)| \geqslant |\hat{w}(s_i)|.$$

The theorem follows. Q.E.D.

Remark: Theorem 4 implies that no feedback can produce small sensitivity if a plant zero is present in any heavily weighted part of the right half-plane. For "low-pass" weightings such as $k(s+k)^{-1}$, $k>0$, this means that if sensitivity is to be reduced, the only RHP zeros possibly allowed are those at very high complex frequencies, $|s_i| \gg k$.

VII. Effects of Plant Uncertainty on Disturbance Attenuation

Two opposing tendencies can be found in most feedback systems. On the one hand, to the extent that feedback reduces sensitivity it reduces the need for plant identification. On the other hand, the less information is available about the plant, the less possible it is to select a feedback to reduce sensitivity. The balance between these tendencies establishes a maximum to the amount of tolerable plant uncertainty and, equivalently, a minimum to the amount of identification needed.

It can be argued that the search for such a minimum should be basic to the theory of adaptive systems. Actually, even the existence of such a minimum appears not to have been stated, perhaps because plant uncertainty is so difficult to study in the WHK framework in the absence of the multiplicative properties (4.1), and because there is no notion of optimality in the classical setup.

Here, we would like to take a step in the direction of articulating these issues, by defining the tradeoff between minimal sensitivity and plant uncertainty and deducing its simpler properties. Sensitivity to disturbances will be considered in this section, and to plant uncertainty in the next.

Let $\mathbb{B}_r$ be a subspace of the causal operators $\mathbb{B}_s$. Consider the feedback (3.1) and model reference (3.3) schemes. Suppose that some nominal plant P_1 in $\mathbb{B}_r$ is specified and (3.4) hold, so the two schemes are equivalent. The true plant P in $\mathbb{B}_r$ will differ from P_1 in general. Let K_{33}: $\mathfrak{X} \to \mathfrak{X}$, $K_{33}(d)=y$ be the operator in $\mathbb{A}$ mapping disturbances into outputs. K_{33} can be expressed in terms of F [using (3.2a)],

$$K_{33} = (I+PF)^{-1}$$

or in terms of P_1 and Q [using 3.4b)],

$$\begin{aligned} K_{33} &= \left[I + PQ(I-P_1Q)^{-1}\right]^{-1} \\ &= \left[(I-P_1Q+PQ)(I-P_1Q)^{-1}\right]^{-1} \\ &= (I-P_1Q)\left[I+(P-P_1)Q\right]^{-1}. \end{aligned} \tag{7.1}$$

We shall be interested in the way in which closed-loop operators such as K_{33} behave as functions of the open-loop operators P, P_1, Q, and F. In particular, let us define two functions mapping open-loop into closed-loop operators. Let $\mathbb{E}$: $\mathbb{A}_s^3 \to \mathbb{A}$, $\mathbb{E}(P, P_1, Q) = K_{33}$ be the function relating K_{33} to the model reference variables, and $\mathbb{E}_f: \mathbb{A}_s^2 \to \mathbb{A}$, $\mathbb{E}_f(P, F) = K_{33}$ the function for the feedback variables. Any pair (d, y) in $\mathfrak{X}^2$ satisfying (3.1) or (3.3) also satisfies the equations

$$y = \mathbb{E}_f(P, F)d = \mathbb{E}(P, P_1, Q)d. \tag{7.2}$$

The c.l. (closed-loop) operators K_{ij}, $i, j = 2, 3$, of the feedback scheme (3.1) were introduced in Section III, and are well defined for the model reference scheme (3.3) under the assumed equivalence. The model reference scheme shown in Fig. 3 has two extra nodes labeled 4 and 5. We shall avoid the lengthly but straightforward calculation of the remaining c.l. operators, and instead assume the following elementary properties of the model reference scheme: define the operators $K_{44} \triangleq [I+(P-P_1)Q]^{-1}$ and $K_{55} \triangleq [I+Q(P-P_1)]^{-1}$ and let K be the sextuplet of operators $\{K_{44}, K_{55}, P, P_1, Q, I\}$; the set $\{K_{ij}\}_{i,j=2}^5$ of all c.l. operators consists of algebraic combinations (i.e., involving sums and products only of) the operators in K; furthermore $K_{32} = -QK_{44}$.

A. Stabilizing Feedbacks

It is well known that any stable plant which is stabilized by feedback is surrounded by a ball of "admissible" perturbations which preserve closed-loop stability. Here, the radius of such a ball will be calculated, and sensitivity will be defined with respect to plant uncertainty within the ball.

Suppose a right seminorm $\|\cdot\|_r$ to be defined on $\mathbb{B}_r$, and recall that $\|\cdot\|_r \geqslant \|\cdot\|$. The true plant P will be supposed to lie in a *ball of uncertainty of radius* $\delta \geqslant 0$ in $\mathbb{B}_r$ *around the nominal* P_1, $b(P_1, \delta) \triangleq \{P \text{ in } \mathbb{B}_r : \|P-P_1\|_r \leqslant \delta\}$. Since $\|P-P_1\| \leqslant \|P-P_1\|_r$, $b(P_1, \delta)$ is a subset of the ball of radius δ in $\mathbb{B}$.

An operator K in $\mathbb{A}$ which depends on $P \in \mathbb{B}_r$ will be called *bounded* over a ball $b(P_1, \delta)$ iff K is in $\mathbb{B}$ and there is a constant $c \geqslant 0$ with the property that $\|K\| \leqslant c$ for all P in $b(P_1, \delta)$. A feedback or model reference scheme will be *c.l. (closed-loop) bounded* on $b(P_1, \delta)$ iff all its c.l. operators are bounded on $b(P_1, \delta)$.

Proposition 7.1: The following statements are equivalent on any ball $b(P_1, \delta)$. i) The feedback scheme is closed-loop bounded; ii) Q is in $\mathbb{B}$ and $K_{44} \triangleq [I+(P-P_1)Q]^{-1}$ is bounded; iii) the model reference scheme is closed-loop bounded.

Proof: We shall prove the sequence of implications, i)⇒ii)⇒iii)⇒i). If i) is true, then K_{44} is bounded by definition. Also, K_{32} is bounded by definition, and therefore in $\mathbb{B}$ for all P in $b(P_1, \delta)$. But P_1 is in $b(P_1, \delta)$, and $K_{32} = -Q$ when $P = P_1$; therefore, Q is in $\mathbb{B}$, and ii) is true. If ii) is true then K_{55} is also bounded, as the equations $K_{55} = [I+Q(P-P_1)]^{-1} = I - QK_{44}(P-P_1)$ imply that

$\|K_{55}\| \leq 1+\|Q\| \cdot \|K_{44}\| \cdot \delta$, for any P in $b(P_1, \delta)$. Therefore, and by the assumed property of the K_{ij}, each c.l. operator K_{ij}, $i, j=2,\cdots,5$, is an algebraic combination of the operators of K, which are bounded on $b(P_1, \delta)$. It follows that each K_{ij} is bounded there, and that (3.3) is c.l. bounded, i.e., iii) is true. If iii) is true then i) is true, as (3.1) and (3.3) are equivalent. Q.E.D.

Again, only stable Q need be considered. The operators F in $\mathbb{A}_s$ and Q in $\mathbb{B}_s$ will be called *stabilizing* for $b(P_1, \delta)$ iff (3.1) and (3.3), respectively, are c.l. bounded on $b(P_1, \delta)$. The set of all Q in $\mathbb{B}_s$ stabilizing for $b(P_1, \delta)$ will be denoted by $\mathbb{B}_s(P_1, \delta)$. For any P_1 and Q in $\mathbb{B}_s$, the set of real points δ for which Q is stabilizing for $b(P_1, \delta)$ will be denoted by $\Delta(P_1, Q)$, and abbreviated to Δ when dependence on (P_1, Q) is not of interest.

The next lemma shows that any stable Q gives a c.l. bounded scheme (3.1), and is therefore stabilizing, for δ small enough. Let P_1 and Q be in $\mathbb{B}_s$.

Proposition 7.2: If δ satisfies $0 \leq \delta < \|Q\|^{-1}$, then δ is in Δ and for any P in $b(P_1, \delta)$ the inequality

$$\|\{I+(P-P_1)Q\}^{-1}\| < (1-\delta\|Q\|)^{-1} \tag{7.3}$$

holds. Δ is a half-open interval, $[0, \delta_1)$.

Proof: Under the hypothesis, $\|(P-P_1)Q\| < 1$. Therefore, the small gain property ensures that $[I+(P-P_1)Q]^{-1}$ is in $\mathbb{B}$, and (7.3) is true by Proposition 2.1c. Δ is an interval beginning at 0, because $\delta \in \Delta$ and $0 \leq \delta' \leq \delta$ implies that $b(P_1, \delta)$ contains $b(P_1, \delta')$, which implies that $\delta' \in \Delta$. Δ is half open by a standard perturbation argument for the openness of resolvent sets which will be omitted, as this property is not important here. Δ contains $[0, \|Q\|^{-1}]$ because, for any δ in $[0, \|Q\|^{-1}]$, (7.3) implies that K_{44} is bounded on $b(P_1, \delta)$, so Q is stabilizing by Proposition 7.1.

B. Sensitivity

Suppose that right and left seminorms are defined on $\mathfrak{B}$ as in Section V, and induce a weighted seminorm $\|\cdot\|_W$ on $\mathbb{B}$. For the plant P in any ball of uncertainty $b(P_1, \delta)$, and any feedback F in $\mathbb{A}_s$ or comparator Q in $\mathbb{B}_s$, the *sensitivity to disturbances under plant uncertainty* of the equivalent schemes (3.1), (3.3) is defined to be

$$\eta(P_1, Q; \delta) = \sup_{P \in b(P_1, \delta)} \sup\{\|y\|_\ell / \|d\|_r : d \text{ in } \mathfrak{B}_r, \|d\|_r \neq 0\}$$

if Q is (stabilizing) in $\mathbb{B}_s(P_1, \delta)$, and $\eta(P_1, Q; \delta) = \infty$ otherwise. For any stabilizing Q,

$$\eta(P_1, Q; \delta) = \sup_{P \in b(P_1, \delta)} \|\mathbb{E}(P, P_1, Q)\|_W$$

by (7.2). (Recall, also, that $\mathbb{E}(P, P_1, Q) = \mathbb{E}_f(P, F)$.) $\eta(P_1, Q, \delta)$ is the smallest assured sensitivity for P in the ball. We are interested in finding Q to minimize this sensitivity. Accordingly, we define the *minimal sensitivity to plant uncertainty* to be

$$\eta(P_1, \delta) = \inf_{Q \in \mathbb{B}_s} \sup_{P \in b(P_1, \delta)} \|\mathbb{E}(P, P_1, Q)\|_W$$

with the proviso that the sup is replaced by ∞ if Q is not stabilizing. For any fixed nominal plant[16] P_1 in $\mathbb{B}_s$, $\eta(P_1, \delta)$ is a positive-real valued function of $\delta > 0$.

Theorem 5: For any nominal plant P_1 in $\mathbb{B}_s$, the minimal sensitivity to plant uncertainty $\eta(P_1, \delta)$ is a monotone nondecreasing function of δ for $\delta \geq 0$; $\eta(P_1, \delta)$ approaches the singularity measure $\mu(P_1)$ as $\delta \to 0$; and $\eta(P_1, \delta) = \|I\|_W$ for $\delta = \|P_1\|_r$.

Lemma 7.3: For any P_1 and Q in $\mathbb{B}_s$, $\eta(P_1, Q; \delta)$ is a monotone nondecreasing function of $\delta \geq 0$, finite for δ in Δ. Δ contains the interval $[0, \|Q\|^{-1}]$. For any δ in $[0, \|Q\|^{-1}]$, the inequalities

$$\|I-P_1Q\|_W \leq \eta(P_1, Q; \delta) \leq \|I-P_1Q\|_W + \|I-P_1Q\|_W \delta\|Q\|(1-\delta\|Q\|)^{-1} \tag{7.4}$$

and, if $\|\cdot\|_W$ is symmetric, also the inequality

$$\eta(P_1, Q; \delta) \leq \|I-P_1Q\|_W (1-\delta\|Q\|)^{-1} \tag{7.5}$$

are satisfied.

Proof of Lemma 7.3: The symbol $\uparrow$ will denote a monotone nondecreasing function of $\delta \geq 0$. By definition of Δ, Q is stabilizing for any $b(P_1, \delta)$ with δ in Δ, so $\eta(P_1, Q; \delta)$ is finite on Δ. By Proposition 7.2, Δ is an interval containing $[0, \|Q\|^{-1}]$. δ is a $\uparrow$ function on Δ since it is a sup over sets $b(P_1, \delta)$ which are nested and nondecreasing as δ increases.

From (7.1) we get

$$\mathbb{E}(P, P_1, Q) = (I-P_1Q)\{I-(P_1-P)Q[I+(P-P_1)Q]^{-1}\}. \tag{7.6}$$

The triangle inequality and dominance condition $\|\cdot\|_W \leq \|\cdot\|$ give

$$\|\mathbb{E}(P, P_1, Q)\|_W \leq \|I-P_1Q\|_W + \|(I-P_1Q)(P_1-P)Q[I+(P-P_1)Q]^{-1}. \tag{7.7}$$

For δ in $[0, \|Q\|^{-1}]$, the upper bound in (7.4) is obtained from (7.7) by (7.3) and the multiplicative property of norms. The lower bound is valid as P_1 is in $b(P_1\delta)$, and $E(P_1, P_1, Q) = I - P_1Q$. If $\|\cdot\|_W$ is symmetric, the multiplicative property of *symmetric* seminorms applied to (7.1) gives

$$\|\mathbb{E}(P, P_1, Q)\|_W \leq \|I-P_1Q\|_W \|\{I+(P-P_1)Q\}^{-1}\|$$

and (7.4) follows. Q.E.D.

Proof of Theorem 5: From Lemma 7.3 and the inequality $\|I-P_1Q\|_W \geq \mu(P_1)$, it can be concluded that for any Q in $\mathbb{B}_s$, $\eta(P_1, Q; \delta)$ is a $\uparrow$ function with values in $[\mu(P_1), \infty]$. As $\eta(P_1, \delta) = \inf_{Q \in \mathbb{B}} \eta(P_1, Q; \delta)$ by definition, $\eta(P_1, \delta)$ is the inf of a set of $\uparrow$ functions with values in $[\mu(P_1, \infty]$. Therefore, $\eta(P_1, \delta)$ is $\uparrow$ and $\eta(P_1, \delta) \geq \mu(P_1)$. Since 0 is in

[16]Again, the assumption that plant and feedback are both strictly causal can be relaxed in Theorem 5, Lemma 5.3, etc.

$\mathbb{B}_s$, $\eta(P_1,0;\delta)$ is in the set, and $\eta(P_1,\delta)\leqslant\eta(P_1,0;\delta)=\|I\|_W$, where the last identity is true by (7.1).

Let us show that $\lim_{\delta\to 0}\eta(P_1,\delta)=\mu(P_1)$. By definition of $\mu(P_1)$, for any $\epsilon>0$ there is a Q in $\mathbb{B}_s$ satisfying $\|I-P_1Q\|_W\leqslant\mu(P_1)+\epsilon/2$; as the RHS of (7.4) approaches $\|I-P_1Q\|_W$ as $\delta\to 0$, there is a $\delta>0$ for which $\eta(P,Q;\delta)\leqslant\mu(P_1)+\epsilon$. Since ϵ was arbitrary, and it has been shown that $\eta(P_1,\delta)\geqslant\mu(P_1)$, the conclusion follows.

Finally, let us show that $\eta(P_1,\|P_1\|_r)=\|I\|_W$. Suppose the contrary to be true. It has been shown that $\eta(P_1,\|P_1\|_r)\leqslant\|I\|_W$, so $\eta(P_1,\|P_1\|_r)\leqslant\|I\|_W$. Therefore, there is a Q in $\mathbb{B}_s$ for which $\|\mathbb{E}(P,P_1,Q)\|_W<\|I\|_W$. Now, 0 is in $b(P_1,\|P_1\|_r)$ and, by (7.1), $\eta(P_1,\|P_1\|_r)\geqslant\|\mathbb{E}(0,P_1,Q)\|_W=\|I\|_W$ which is a contradiction. Q.E.D.

Remark: Theorem 5 implies that whenever feedback reduces sensitivity for the nominal plant P_1, there are points in the interval $[0,\|P_1\|]$ at which $\eta(P_1,\delta)$ increases, i.e., where "the less we know about P, the less able we are to construct a feedback to attenuate disturbances."

VIII. Filtering of Plant Uncertainty: Invariant Schemes

We now turn to the second problem outlined in the introduction, Section I-D. The plant P lies in a ball of uncertainty[17] around some nominal value P_1, and one object of using feedback is to shrink the size of the uncertainty. Of course, uncertainty can be reduced to zero by disconnecting the input (u in Fig. 4) from the system, but then P_1 is also transformed into zero. Clearly, the problem is trivial unless there is a normalization or constraint on the control law that transforms P_1 into a closed-loop system.

We would prefer as far as possible to separate the reduction of uncertainty from the transformation of P_1, and therefore seek a definition of uncertainty which is independent of the eventual closed-loop system.

If P_1 were a real number, uncertainty could be normalized by specifying it as a percentage of the nominal value. This possibility is not open for noninvertible plants. Instead, we shall achieve a normalized definition of uncertainty by employing the device of a plant-invariant scheme, which leaves the nominal plant invariant while shrinking the ball of uncertainty. Such a scheme will be shown always to be realizable in the form of a model reference scheme.

This device will also enable us to separate the design process into two consecutive stages: 1) reduction of uncertainty and 2) transformation of the nominal plant into a nominal closed-loop system (cf. the separation into estimation and control stages in Kalman filtering).

A possible disadvantage to this approach is that the two stages may be dependent, and yield a suboptimal sensitivity. We shall try to get the best of both worlds, and simultaneously formulate normalized and unnormalized

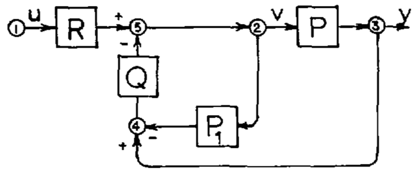


Fig. 5.

versions of the problem, by allowing the plant invariant scheme to be multiplied by a constant R_1, representing a desired nominal control law.

A. Plant Invariant Schemes

Any feedback arrangement of the type shown in Fig. 1, incorporating a plant P with a single accessible input, has an equivalent description in terms of the flowgraph of Fig. 4. This flowgraph is completely specified by the two additional operators U and F, which may be unstable even though the original arrangement is closed-loop stable. Note that Fig. 4 is an enlargement of Fig. 2 by a new branch, U, connected to a new node labeled 1. For simplicity, it will be assumed that $d=0$. The new *feedback scheme* equations are

$$y=Pv \tag{8.1a}$$

$$v=Uu-Fy \tag{8.1b}$$

in which $P\in\mathbb{A}_s$, $F\in\mathbb{A}_s$, and[18] $U\in\mathbb{A}$ are operators, and u, v, and y are in $\mathfrak{X}$. (Equations (8.1) represent an enlargement of (3.1) by the term Uu, under the constraint that $d=0$.) As P is strictly causal, $(I+FP)^{-1}$ exists in $\mathbb{A}$, and for each u in $\mathfrak{X}$, (8.1) have unique solutions for v and y in $\mathfrak{X}$, given by the equations

$$v=(I+FP)^{-1}Uu \tag{8.2a}$$

$$y=P(I+FP)^{-1}Uu. \tag{8.2b}$$

Let K_{12} and K_{13} be the operators mapping $\mathfrak{X}$ into $\mathfrak{X}$ which satisfy, $K_{12}(u)=v$, $K_{13}(u)=y$. By (8.2), $K_{12}=(I+FP)^{-1}U$ and $K_{13}=P(I+FP)^{-1}U$.

The full set of closed-loop operators of (8.1) consists of K_{12}, K_{13}, $K_{11}\triangleq I$, $K_{i1}\triangleq 0$, and the operators K_{ij} defined in Section III, where $i,j=2,3$.

A variant of the model-reference transformation of Section III-A will be used. The flowgraph of Fig. 5 is described by the new *model reference scheme* equations

$$y=Pv \tag{8.3a}$$

$$v=Ru-Q(y-P_1v) \tag{8.3b}$$

in which $P,P_1\in\mathbb{B}_s$, $R\in\mathbb{A}$, and $u,v,y\in\mathfrak{X}$. For any $u\in\mathfrak{X}$, (8.3) have unique solutions for v and y in $\mathfrak{X}$, as P, P_1, and Q are strictly causal. (Equations (8.3) can be viewed as an enlargement of (3.3) by the term Ru, and subject to the constraint that $d=0$.)

The schemes (8.1) and (8.3) will be called *equivalent* iff

[17]Departures of the true plant from the nominal can be interpreted as uncertain in some applications; in others, they may simply represent known perturbations to be attenuated.

[18]In order to be physically realized, U would have to be approximated by a strictly causal operator. We prefer not to assume that $U\in\mathbb{A}_s$ as this simplifies the presentation. As U is followed by a strictly causal element, and has no feedback around it, there is no loss of generality.

every triple (u, v, y) in $\mathfrak{X}^3$ satisfying (8.1) satisfies (8.3) and vice versa. Equivalence will be established subject to the equations

$$Q=F(I+P_1F)^{-1} \tag{8.4a}$$

$$F=Q(I-P_1Q)^{-1} \tag{8.4b}$$

$$R=(I+FP_1)^{-1}U \tag{8.5a}$$

$$U=(I+FP_1)R. \tag{8.5b}$$

Equation (8.4) is a repetition of (3.4). If either of (8.5) holds then, clearly, both are true.

For equivalent schemes, let $\mathbb{E}(P, P_1, Q)$ be defined as in Section VII; Let $\mathbb{K}$ be the function $\mathbb{K}$: $\mathbb{A}_s^3 x \mathbb{A} \to \mathbb{A}$, $\mathbb{K}(P, P_1, Q, R)=K_{13}$, which maps operators appearing in (8.3) into the c.l. (closed-loop) operator K_{13}. Whenever all operators except P are regarded as fixed, $\mathbb{K}(P, P_1, Q, R)$ will be denoted by $\mathbb{K}(P)$. $\mathbb{K}(P)$ can be expressed in terms of the model-reference scheme operators, by the formula

$$\mathbb{K}(P)=P\left[I+Q(P-P_1)\right]^{-1}R \tag{8.6a}$$

which is obtained from the following sequence of equations. For any (u, v, y) in $\mathfrak{X}^3$ satisfying (8.3), we have

$$v=Ru-Q(P-P_1)v$$

$$\left[I+Q(P-P_1)\right]v=Ru$$

$$y=Pv=P\left[I+Q(P-P_1)\right]^{-1}Ru \tag{8.6b}$$

in which the inverse exists in $\mathbb{A}$ as $Q(P-P_1)$ is strictly causal. Since K_{13} maps u into y, K_{13} must coincide with the last operator of (8.6b), and (8.6a) is true.

The set of c.l. operators of the model reference scheme (8.3) is defined as in the preceding Section VII, except that K is augmented by the operator R.

For any P_1 in $\mathbb{A}_s$, $\mathbb{K}$ will be called (nominal) *plant invariant* iff $\mathbb{K}(P_1)=P_1$; and, for any R_1 in $\mathbb{A}$, *plant invariant* $(\times R_1)$ iff $\mathbb{K}(P_1)=P_1R_1$. From (8.4) it is clear that $\mathbb{K}$ has these properties whenever $R=I$ or $R=R_1$, respectively.

An operator $0_r \in \mathbb{B}$ is a *right zero*[19] of P_1 iff $P_1 0_r=0$.

Theorem 7:

a) Any feedback scheme (8.1) with (P, F, U) in $\mathbb{A}_s^2 \times \mathbb{A}$ is equivalent to a model reference scheme (8.3) with (P, P_1, Q, R) in $\mathbb{A}_s^3 \times \mathbb{A}$, in which (F, U) and (P_1, Q, R) are related by (8.4)–(8.5); and vice versa. If (8.4)–(5) hold, then we have the following.

b) If P assumes the (nominal) value $P_1 \in \mathbb{A}_s$, then $\mathbb{K}(P_1)=P_1R$, and if the feedback scheme is closed-loop stable, then Q and R of the model reference scheme are stable (although F and U may be unstable).

c) The differences $\mathbb{K}(P)-\mathbb{K}(P_1)$ and $P-P_1$ are related by the formulas

$$\mathbb{K}(P)-\mathbb{K}(P_1)=(I-P_1Q)(P-P_1)\{I+Q(P-P_1)\}^{-1}R \tag{8.7a}$$

$$=\mathbb{E}(P, P_1, Q)(P-P_1)R. \tag{8.7b}$$

d) For any $R_1 \in \mathbb{B}$, $\mathbb{K}$ is plant invariant $(\times R_1)$ iff $R=R_1+0_r$, where $0_r \in \mathbb{B}$ is any right zero of P_1.

Proof:

a) If (u, v, y) satisfies (8.1), then we have

$$\begin{aligned}(I-QP_1)v&=(I-QP_1)(Uu-Fy) && [\text{by } (8.1\text{b})]\\ &=(I-QP_1)\left[(I+FP_1)Ru-Q(I-P_1Q)^{-1}Y\right] && [\text{by } (8.5\text{b}) \text{ and } (8.4\text{b})]\\ &=(I-QP_1)\left[(I-QP_1)^{-1}Ru-(I-QP_1)^{-1}Qy\right]\\ &=Ru-Qu\end{aligned}$$

where the second last equation was obtained using the identity $(I+FP_1)=(I-QP_1)^{-1}$ [see (3.5)] and Property 2.1a. Hence, (u, v, y) satisfies (8.3). The reverse assertion is proved similarly. It follows that (8.1) and (8.3) are equivalent.

b) From (8.4), $\mathbb{K}(P_1)=P_1R$; if (8.1) is closed-loop stable, then the c.l. operators $K_{32}=F(I+P_1F)^{-1}$ and $K_{12}=(I+FP)^{-1}U$ are stable, and equal Q and R.

c) From (8.4) the following sequence of equations is obtained:

$$\mathbb{K}(P)-\mathbb{K}(P_1)=P\left[I+Q(P-P_1)\right]^{-1}R-P_1R$$

$$=\left\{P-P_1\left[I+Q(P-P_1)\right]\right\}\left[I+Q(P-P_1)\right]^{-1}R$$

from which (8.7a) follows; (8.7b) is obtained by Proposition 2.1a.

d) $\mathbb{K}$ is plant invariant $(\times R_1)$ iff $\mathbb{K}(P_1)-P_1R_1=P_1(R-R_1)=0$, i.e., $R-R_1=0_r$. Q.E.D.

Remark: Plant invariant schemes allow us to divide any control-law synthesis into two stages: 1) filtering of plant uncertainty $P-P_1$ and 2) design of a control law for a nominal plant P_1, with the assurance at least that the filtering stage will improve the design. In general, there is no "separation principle" to guarantee optimality of the division.

Either stage may come first. Therefore, in our theorems, P can be interpreted either as a plant without controller, for which a controller will be designed eventually; or, as a plant with controller attached, which requires additional filtering only to the extent that P differs from P_1.

B. *Stabilizing Feedbacks*

In the rest of Section VIII it will be assumed that: P and P_1 belong to a subspace $\mathbb{B}_r$ of the strongly causal operators $\mathbb{B}_s$, on which a right seminorm $\|\cdot\|_r$ is defined; P lies in a ball $b(P_1, \delta)$, defined as in Section VII-A, of what can be interpreted either as uncertainty or perturbations; and the equivalence conditions (8.4), (8.5) hold, so that the feedback and model reference schemes (8.1), (8.3) are equivalent.

Q in $\mathbb{A}_s$ and R in $\mathbb{A}$ are sought which give low sensitivity and maintain c.l. boundedness. In view of Theorem 7b), the assumption that Q is in $\mathbb{B}_s$ and R in $\mathbb{B}$, i.e., that both are stable, can be made without loss of generality.

[19]More simply, 0_r is an operator whose range is in the nullspace of P_1. The term "zero" is appropriate for a normed algebra.

For any $b(P_1,\delta)$, the definitions of a bounded operator, c.l. bounded scheme, and stabilizing Q were given in Section VII-A. The set $\mathbb{B}_s(P_1,\delta)$ of Q in $\mathbb{B}_s$ stabilizing for $b(P_1,\delta)$ was introduced.

Proposition 8.1: Under the present hypotheses, Propositions 7.1 and 7.2, are valid for the feedback scheme (8.1) and model reference scheme (8.3), and the set of Q in $\mathbb{B}_s$ stabilizing for $b(P_1,\delta)$ coincides with $\mathbb{B}_s(P_1,\delta)$.

Proof: As R is in $\mathbb{B}$, the set K augmented by R consists of operators bounded on $b(P_1,\delta)$ iff the unaugmented set K consists of operators bounded on $b(P_1,\delta)$. Therefore, $\{K_{ij}\}_{i,j=1}^{6}$ are bounded iff $\{K_{ij}\}_{i,j=1}^{5}$ are bounded. The conclusion follows. Q.E.D.

C. Sensitivity to Plant Perturbations or Uncertainty

Suppose that a pair of seminorms $(\|\cdot\|_W,\|\cdot\|_1)$ is defined on the spaces $(\mathbb{B},\mathbb{B}_r)$, respectively. We assume $\|\cdot\|_W$ to be a left seminorm, but leave open the possibility that $\|\cdot\|_1$ is not a right seminorm, and assume instead that $\|\cdot\|_1\leqslant\|\cdot\|_r$, i.e., $\|\cdot\|_1$ is dominated by the right seminorm $\|\cdot\|_r$. A weighted seminorm $\|\cdot\|_v$ is assumed to be defined on the set of products $\mathbb{B}\cdot\mathbb{B}_r$, and to have the multiplicative property, $\|CD\|_v\leqslant\|C\|_W\|D\|_1$ for all C in $\mathbb{B}$ and D in $\mathbb{B}_r$. The pair $(\|\cdot\|_W,\|\cdot\|_1)$ will be called *aligned* iff there is a $D\in\mathbb{B}_r$ for which the preceding inequality is an equality for all C in $\mathbb{B}$.

For the plant P in any ball of uncertainty $b(P_1,\delta)$, and any $(F\in\mathbb{A}_s,\ U\in\mathbb{A})$, or $(Q\in\mathbb{B}_s,\ R\in\mathbb{B})$, the *sensitivity to plant perturbations* (or uncertainty) of the equivalent schemes (8.1)–(8.3) is defined to be

$$\nu(P_1,Q,R;\delta)=\sup_{P\in b(P_1,\delta)}\left\{\|\mathbb{K}(P)-\mathbb{K}(P_1)\|_v\delta^{-1}\right\} \tag{8.8}$$

for Q stabilizing (i.e., in $\mathbb{B}_s(P_1,\delta)$), and $\nu(P_1,Q,R;\delta)=\infty$ otherwise. For any stabilizing Q, (8.7) gives the equation

$$\nu(P_1,Q,R;\delta)=\sup_{P\in b(P_1,\delta)}\left\{\|\mathbb{E}(P,P_1,Q)(P-P_1)R\|_v\delta^{-1}\right\}. \tag{8.9}$$

The following lemma relates the disturbance and plant perturbation sensitivities, η and ν, to each other, and to the $\|\cdot\|_W$-singularity measure $\mu(P_1)$, when $R=I$.

Lemma 8.2: $\nu(P_1,Q,I;\delta)$ is a monotone nondecreasing function of $\delta\geqslant 0$ satisfying the inequality

$$\nu(P_1,Q,I;\delta)\leqslant\eta(P_1,Q;\delta); \tag{8.10}$$

and if $\|\cdot\|_1=\|\cdot\|_r$ and the pair $(\|\cdot\|_W,\|\cdot\|_r)$ is aligned then

$$\nu(P_1,Q,I;\delta)\geqslant\mu(P_1). \tag{8.11}$$

Remark: $\|\cdot\|_v$ can coincide with the principal norm $\|\cdot\|$. For example, if W is in $\mathbb{B}_r$, $\|W\|=1$, $\|C\|_W\triangleq\|CW\|$, and $\|D\|_1\triangleq\|W^{-1}D\|$, then the principal norm has the multiplicative property $\|CD\|_v\leqslant\|C\|_W\|D\|_1$ and may be used as the $\|\cdot\|_v$ norm. Even though $\|\cdot\|$ provides no weighting, (8.10) shows that sensitivities ν smaller than 1 can be achieved whenever $\nu(P_1)<1$. In effect, the plant perturbations supply their own weighting.

This example has the alignment property, as $\|CW\|_v=\|C\|_W\|W\|_1$.

Proof of Lemma 8.2: $\nu(P_1,Q,I;\delta)$ is a $\uparrow$ function of δ, as it is a sup over sets $b(P_1,\delta)$ which are nested and nondecreasing with $\delta\in\Delta$, and is ∞ for $\delta\notin\Delta$. For any P in $b(P_1,\delta)$ and stabilizing Q, we have the inequalities

$$\begin{aligned}\|\mathbb{E}(P,P_1,Q)(P-P_1)\|_v\delta^{-1}&\leqslant\|\mathbb{E}(P,P_1,Q)\|_W\|P-P_1\|_1\delta^{-1}\\&\leqslant\|\mathbb{E}(P,P_1,Q)\|_W\end{aligned} \tag{8.12}$$

by the multiplicative property of $\|\cdot\|_v$, and as

$$\|P-P_1\|_1\leqslant\|P-P_1\|_r\leqslant\delta$$

is obtained by taking the sup of both sides of (8.12) over all P in $b(P_1,\delta)$. If $\|\cdot\|_1=\|\cdot\|_r$ and $(\|\cdot\|_W,\|\cdot\|_r)$ is aligned there is an operator $D\in\mathbb{B}_r$ for which $\|\mathbb{E}(P,P_1,Q)D\|_v=\|\mathbb{E}(P,P_1,Q)\|_W\|D\|_r$. The operator $P_0\triangleq P_1+\delta D\|D\|_r^{-1}$ is in $b(P_1,\delta)$ and $\|\mathbb{E}(P_0,P_1,\delta)(P-P_1)\|_v\delta^{-1}=\|\mathbb{E}(P_1,P_1,Q)\|_W\geqslant\mu(P_1)$. The sup in (8.9) must have the lower bound $\mu(P_1)$, and (8.11) follows. Q.E.D.

D. Assumptions on R

As $R=0$ gives $\nu=0$, the attainment of a small sensitivity is trivial unless R is constrained. In many problems a target value of the c.l. operator K_{13} is specified for some nominal value of the plant. By Theorem 7b), the target value can be attained iff it has the form P_1R. We therefore make the following assumption.

Assumption 1: An $R_1\in\mathbb{B}$ is given for which the equation $\mathbb{K}(P_1)=P_1R_1$ must be satisfied, i.e., $\mathbb{K}$ is plant invariant $(\times R_1)$. By Theorem 7d), $R=R_1+O_r$, where $O_r\in\mathbb{B}$ is any right zero of P_1. There are now two variables to be optimized, Q and O_r. Their simultaneous optimization can be difficult, and we make the following simplifying assumption.

Assumption 2: $O_r=0$. Assumption 2 may constrain the class of allowable feedbacks and thereby give suboptimal sensitivities. However, it is clear from (8.7b) that there is no such constraint if the condition

$$(P-P_1)O_r=0 \tag{8.13}$$

is fulfilled. Equation (8.13) is fulfilled whenever $\mathsf{N}(P_1)$ (nullspace of P_1) is contained in $\mathsf{N}(P)$ for all P in $b(P_1,\delta)$; for then, for any $u\in\mathcal{X}$, either $O_ru=0$, or $O_ru\in\mathsf{N}(P_1)$ and so $O_ru\in\mathsf{N}(P)$. In either case, $(P-P_1)O_ru=0$, i.e., (8.13) is true. If $\mathsf{N}(P_1)$ is trivial, then, of course, (8.13) is fulfilled. For example, $\mathsf{N}(P_1)$ is trivial for any nonzero P_1 in $\mathbb{H}_0^\infty$, because functions analytic and not identically zero in $\mathrm{Re}(s)\geqslant 0$ have at most a countable number of $\mathrm{Re}(s)\geqslant 0$ zeros.

As R in (8.6) is now fixed, cR can be absorbed into the weighting for some constant λ in $(0,1]$ by replacing $\|\cdot\|_v$ with the new seminorm $\|C\|_v\triangleq\lambda\|CR\|_v$, and similarly for $\|\cdot\|_1$. Without loss of generality, we can therefore make the following assumption.

Assumption 3: $R=I$. We prefer to absorb R rather than to show it explicitly, as this approach reduces the number of variables and allows sensitivity to be related to weighted invertibility.

E. Minimal Sensitivity

For any P_1 in $\mathbb{B}_s$ and $\delta>0$, the *minimal sensitivity to plant perturbations* is defined to be

$$\begin{aligned}\nu(P_1,\delta) &= \inf_{Q\in\mathbb{B}} \{\nu(P_1,Q,I;\delta)\}\\ &= \inf_{Q\in\mathbb{B}} \sup_{P\in b(P_1,\delta)} \left\{\|\mathbb{E}(P,P_1,Q)(P-P_1)\|_v \delta^{-1}\right\}\end{aligned}$$

provided the last sup is replaced by ∞ for nonstabilizing Q. For fixed P_1, $\nu(P_1,\delta)$ is a function of $\delta\geqslant 0$ which represents the dependence of perturbation sensitivity on plant uncertainty.

Theorem 8: For any nominal plant P_1 in $\mathbb{B}_s$, the minimal sensitivity to plant perturbations $\nu(P_1,\delta)$ is a monotone nondecreasing function of $\delta\geqslant 0$.

$\nu(P_1,\delta)$ satisfies the upper bound conditions $\nu(P_1,\delta)\leqslant \eta(P_1,\delta)$ and $\lim_{\delta\to 0}\nu(P_1,\delta)\leqslant\mu(P_1)$. If $\|\cdot\|_1=\|\cdot\|_r$, and $(\|\cdot\|_W,\|\cdot\|_1)$ are aligned, then $\lim_{\delta\to 0}\nu(P_1,\delta)=\mu(P_1)$ and, for $\delta=\|P_1\|_r$, $\|P_1\|_v\|P_1\|_r^{-1}\leqslant\nu(P_1,\delta)\leqslant\|I\|_W$.

Proof: $\nu(P_1,\delta)$ is a $\uparrow$ function of $\delta\geqslant 0$, as by definition it is a sup over Q in $\mathbb{B}_s$ of functions which are $\uparrow$ by Lemma 8.2. The inequality $\nu(P_1,\delta)\leqslant\eta(P_1,\delta)$ is obtained by taking $\sup_{Q\in\mathbb{B}}$ of both sides of (8.10). It follows by Theorem 5 that $\lim_{\delta\to 0}u(P_1,\delta)\leqslant\mu(P_1)$, and that $\nu(P_1,\delta)\leqslant\|I\|_W$ for $\delta=\|P_1\|_r$. If alignment and $\|\cdot\|_1=\|\cdot\|_r$ are assumed, then taking the $\inf_{Q\in\mathbb{B}}$ of both sides of (8.11) gives the inequality $\nu(P_1,\delta)\geqslant\mu(P_1)$, and the conclusion that $\lim_{\delta\to 0}\nu(P_1,\delta)=\mu(P_1)$ follows.

As the operator $P=0$ lies in $b(P_1,\|P_1\|_r)$ and, for each Q in $\mathbb{B}_s$, $\|E(0,P_1,Q)(-P_1)\|_v\|P_1\|_r^{-1}=\|P_1\|_v\|P_1\|_r^{-1}$; therefore $\|P_1\|_v\|P_1\|_r^{-1}\leqslant\nu(P_1,\|P_1\|_r)$. Q.E.D.

Remark: The ratio $\|P_1\|_v\|P_1\|_r^{-1}$ is a measure of the amount by which the $\|\cdot\|_v$ norm weighs down the nominal plant. Provided norms are used for which the weighting is not excessive, in fact whenever the ratio exceeds $\mu(P_1)$, there must be points in the interval $[0,\|P_1\|_r]$ at which any increase in plant uncertainty causes a worsening of minimal sensitivity.

Appendix I

Proof of Proposition 2.1:

a) If $(I+PQ)^{-1}$ exists in B, let $R\triangleq I-Q(I+PQ)^{-1}P$. Now, $(I+QP)R=I$, because

$$\begin{aligned}(I+QP)R &= I-Q(I+PQ)^{-1}P+QP-QPQ(I+PQ)^{-1}P\\ &= I-Q(I+PQ)^{-1}P+Q\left[I-PQ(I+PQ)^{-1}\right]P\\ &= I-Q(I+PQ)^{-1}P+Q(I+PQ)^{-1}P=I.\end{aligned}$$

Similarly, $R(I+QP)=I$, so R is the inverse of $(I+QP)$ in B. The required formula is obtained by multiplying both sides of the equation $(I+PQ)P=P(I+QP)$ by $(I+PQ)^{-1}$ on the left, and by $(I+QP)^{-1}$ on the right.

b) If P is in R and P^{-1} in B, there is a contradiction: any F in B belongs to R as $F=PP^{-1}F$, so R is not a proper subspace and not a radical.

c) Let $R\triangleq(I+P)^{-1}$ and observe that $R=I-PR$. By the triangle inequality, $\|R\|\leqslant 1+\|P\|\|R\|$ and, as $\|P\|<1$, $\|R\|\leqslant(1-\|P\|)^{-1}$. Q.E.D.

Proof of Proposition 3.5: Since $\|(I-PQ)\Pi\|=1$ when $Q=0$, the infimum α satisfies $\alpha\leqslant 1$. Suppose $\alpha\neq 1$; then there is a Q in $\mathbb{B}_s$ for which $\|(I-PQ)\Pi\|\triangleq\alpha_1<1$. It will now be shown that given any $\epsilon>0$, there is a Q_1 in $\mathbb{B}_s$ for which $\|(I-PQ_1)\Pi\|<\epsilon$, and so the proposition is true.

Let n be any integer for which $\alpha^n<\epsilon$. Now, $(I-PQ)^n=I+a_1PQ+\cdots+a_n(PQ)^n$ is a polynomial in PQ, in which every term except the first has P as a left factor, and Q in $\mathbb{B}_s$ as a factor. Therefore, there is a Q_1 in $\mathbb{B}_s$ for which the equation $(I-PQ)^n=(I-PQ_1)$ holds. Now as $\Pi\mathbb{B}$ is invariant under all operators in $\mathbb{B}$, we get $\|(I-PQ_1)\Pi\|=\|\{(I-PQ)\Pi\}^n\|\leqslant\|(I-PQ)\Pi\|^n\leqslant\alpha_1^n<\epsilon$, as claimed. Q.E.D.

Appendix II
Unstable Plants: Remarks

Consider the problem of input–output stabilizing an unstable plant P_0 in $\mathbb{H}_{e0}^{\infty}$ by identity feedback, given that the *a priori* information about P consists of $\hat{p}(j\omega)$ and the number N_+ of $\mathrm{Re}(s)\geqslant 0$ poles of P. By Nyquist's criterion, uncertainty as to N_+ translates into uncertainty[20] as to the number of unstable poles of $(I+P_0)^{-1}$. N_+ must not be underestimated if stability is to be assured. However, it is impossible to be sure that N_+ has not been underestimated from purely input–output measurements. For example, let s_i, $i=1,2,\cdots n$, be any finite set of frequencies, and $\hat{p}_m(s_i)$ a set of n measurements of $\hat{p}_0(s)$ of tolerance ϵ, i.e., $|\hat{p}_m(s_i)-\hat{p}_0(s)|<\epsilon$. It is impossible to deduce N_+ from the measurements, as it is always possible to find a dipole, $(s+\alpha)(s+\alpha+\delta)^{-1}$, with $\delta>0$ so small that multiplication of $\hat{p}_0(s)$ by the dipole changes $\hat{p}_0(s_i)$ by less than ϵ, but increases N_+ by 1. In our particular setting, it appears that input–output stabilization cannot be accomplished using solely input–output data. The implication is that some information concerning internal structure is essential, and that stabilization is not a legitimate problem for a purely input–output theory.

On the other hand, desensitization can be achieved using input–output data. If P is in H_0^{∞}, and $p(t)$ satisfies appropriate restrictions on smoothness and convergence to 0 as $t\to\infty$, then $\hat{p}(j\omega)$ has a finite modulus of continuity, and $\hat{p}$ can be located in a ϵ-ball of H^{∞} by a finite set of measurements [14]. This is sufficient for the purpose of desensitization, as sensitivity depends continuously on $\hat{p}$ in H^{∞} (see Theorems 5 and 7).

[20] A more complete discussion is in Zames and El-Sakkary [18].

Synopsis

The main results of this paper are the Theorems 1–7.

Section II: Spaces of frequency responses and algebras of input–output mappings are defined, and their relevant properties summarized.

Section III: A decomposition principle is derived. The disturbance attenuation problem is separated into two independent stages: stabilization, followed by desensitization of a stable system. The second stage is a model reference scheme.

Section III′: Some constraints on sensitivity norms are displayed. It is shown that the induced operator norm is useless as a measure of sensitivity.

Section IV: The concepts of a weighted seminorm, approximate inverse, and measure of singularity are introduced and illustrated by examples.

Section V: The plant is assumed to be stable and known precisely. Sensitivity is defined. Approximate invertibility is shown to be a necessary and sufficient condition for sensitivity reduction in Theorem 2, and optimal sensitivity is shown to be equal to the measure of singularity. An example of sensitivity minimization is solved in Section V-A. Unstable plants are discussed in Section V-B. Identity and inverting sequences are introduced in Section V-C, and a lemma on products of inverting sequences is proved.

Section VI: Devoted to multivariable systems. In Theorem 3, an optimal sequence of compensators is derived for a plant factorable into a product of nearly invertible and noninvertible factors. Corollary 6.2 specializes this result to a single-input single-output plant with a RHP zero. Corollary 6.2 gives conditions under which the sensitivity of multivariable systems without RHP zeros can be made arbitrarily small. Theorem 4 shows that sensitivity can never be made small if there are zeros in any heavily weighted part of the RHP.

Section VII: The plant is assumed to lie in a ball of uncertainty around a nominal value. Feedbacks stabilizing over a ball are defined. Optimal sensitivity is shown to be a monotone nondecreasing function of uncertainty in Theorem 5, and various bounds are obtained.

Section VIII: Problem 2 is formulated, again in terms of an equivalence between feedback and model reference schemes, this time subject to a plant invariance property, in Theorem 6. Sensitivity to plant perturbations is defined, bounded, and optimal sensitivity is shown to be a monotone nondecreasing function of plant uncertainty in Theorem 7.

Acknowledgment

The author thanks C. A. Desoer for many useful suggestions concerning the draft of this paper.

References

[1] H. W. Bode, *Network Analysis and Feedback Amplifier Design.* Princeton, NJ: Van Nostrand, 1945.
[2] M. Horowitz, *Synthesis of Feedback Systems.* New York: Academic, 1963.
[3] G. Zames, "Nonlinear operators—Cascading inversion, and feedback" M.I.T., *R.L.E., Quart. P.R.*, vol. 53, pp. 93–107, Apr. 1959.
[4] ——, "Nonlinear operators for system analysis," Res. Lab. Electron., M.I.T., Tech. Rep. 370, Sept. 1960.
[5] ——, "functional analysis applied to nonlinear feedback systems," *IEEE Trans. Circuit Theory*, vol. CT-10, pp. 392–404, Sept. 1963.
[6] C. A. Desoer, "Nonlinear distortion in feedback amplifiers," *IRE Trans. Circuit Theory*, vol. CT-9, pp. 2–6, Mar. 1962.
[7] ——, "Perturbation in the 1/0 map of a nonlinear feedback system caused by large plant perturbation," *J. Franklin Inst.*, vol. 306, pp. 225–235, Sept. 1978.
[8] J. B. Cruz, Jr. and W. R. Perkins, "A new approach to the sensitivity problem in multivariable feedback system design," *IEEE Trans. Automat. Contr.*, vol. AC-9, pp. 216–223, 1964.
[9] W. A. Porter and C. L. Zahm, "Basic concepts in system theory," System Eng. Lab., Univ. of Michigan, Tech. Rep. 33, 1969.
[10] R. M. Desantis and W. A. Porter, "A generalized Nyquist plot and its use in sensitivity analysis," *Int. J. Syst. Sci.*, vol. 5–12, pp. 1143–1153, 1974.
[11] J. C. Willems, *The Analysis of Feedback Systems.* Cambridge, MA: M.I.T., 1971.
[12] G. Zames, "Feedback hierarchies and complexity," in *Proc. IEEE Conf. Decision Contr.*, Dec. 1976, addenda.
[13] ——, "On the metric complexity of causal linear systems, ϵ-entropy and ϵ-dimension for continuous time," *IEEE Trans. Automat. Contr.*, vol. AC-12, pp. 222–230, Apr. 1979.
[14] ——, "Realizability conditions for nonlinear feedback systems," *IEEE Trans. Circuit Theory*, vol. CT-11, pp. 186–194, June 1964.
[15] R. G. Douglas, *Banach Algebra Techniques in Operator Theory.* New York: Academic, 1972.
[16] B. A. Francis and W. M. Wonham, "The internal model principle for linear multivariable regulators," *J. Appl. Math. Optimiz.*, vol. 2, pp. 170–194, 1975.
[17] D. M. Bensoussan, "High-gain feedback and sensitivity," Ph.D. dissertation, Dep. Elec. Eng., McGill Univ., to appear.
[18] A. K. El-Sakkary, "The gap metric for unstable systems," Ph.D. dissertation, Dep. Elec. Eng., McGill Univ.; also in G. Zames and A. K. El-Sakkary, *Proc. 18th Allerton Conf.*, Oct. 1980.
[19] M. A. Naimark, *Normed Rings.* Groningen: Noordhooff, 1964.
[20] A. Feintuch, "Strong Causality Conditions and Causal Invertibility," *SIAM J. Contr. Optimiz.*, vol. 18, pp. 317–324, May 1980.
[21] C. A. Desoer and W. S. Chan, "Feedback interconnection of lumped linear time invariant systems," *J. Franklin Inst.*, vol. 300, pp. 335–351, 1975.
[22] W. Rudin, *Functional Analysis.* New York: McGraw Hill, 1973.
[23] G. Zames, "Feedback and optimal sensitivity: Model reference transformations, weighted seminorms, and approximate inverses," in *Proc. 17th Allerton Conf.*, Oct. 1979, pp. 744–752.

G. Zames (S'57–M'61–SM'78–F'79) received the Sc.D. degree in electrical engineering from the Massachusetts Institute of Technology, Cambridge, in 1960.

He is currently Professor of Electrical Engineering at McGill University, Montreal, P.Q., Canada. He was a Senior Scientist at the NASA Electronics Research Center, Cambridge (1965–1970), and has served on the Faculties of M.I.T. (1960–1965), Harvard University (1962–1963), and the Technion (1972–1974). Among the awards he has received are the Guggenheim Fellowship, National Academy of Sciences R.R. Associateship, Athlone Fellowship (Imperial College), British Association Medal (McGill University), American Automatic Control Council Outstanding Paper Award, and two IEEE TRANSACTIONS ON AUTOMATIC CONTROL Outstanding Paper Awards (Honorable Mention). He is a member of the Editorial Boards of the *SIAM Journal on Control* and *Systems Control Letters*, and author of many papers on control and nonlinear systems.