NASA TN D-3479

OPTIMAL CONTROL OF SATURATING SYSTEMS WITH STOCHASTIC INPUTS

By Elwood C. Stewart and William P. Kavanaugh

Ames Research Center
Moffett Field, Calif.

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

For sale by the Clearinghouse for Federal Scientific and Technical Information
Springfield, Virginia 22151 – Price $1.00

OPTIMAL CONTROL OF SATURATING SYSTEMS WITH STOCHASTIC INPUTS[1]

By Elwood C. Stewart and William P. Kavanaugh
Ames Research Center

29216

SUMMARY

This study is concerned with the optimal control of nonlinear systems which operate in the presence of random unwanted disturbances. Although the results are intended to be applicable to a certain class of nonlinear systems, saturating systems are emphasized because of their practical importance. Such systems are important when operation occurs in critical regions, for example, in the control of vehicles in situations where the available thrust is marginal.

This report presents an approximate method of synthesizing the optimal controller with noisy measurements which utilizes the limited control function in the best possible manner. The principal results show that: (a) although there is no minimum, the performance can approach arbitrarily closely to a greatest lower bound, and (b) striking improvement in performance can be obtained in comparison with linear theory when operating in critical regions. Computer results are used for verification. In addition, dimensionless curves of the optimal performance are given as a function of a dimensionless parameter involving the input statistics and saturation value of the forcing function.

Author

INTRODUCTION

The optimal control of linear systems which operate in the presence of random unwanted disturbances has been the subject of intensive research for many years and appears to be fairly well understood. However much less has been done on the more important and more difficult nonlinear problem.

In one approach, when there is a linear region of operation of the equations, the nonlinear regions are intentionally avoided (ref. 1). The problem is treated as a linear variational problem with constraints by minimizing some function of the error with a constraint on another function of the control. Generally the constraint is placed on only the expected value or time average of the squared control function. Thus to insure that the equations remain linear, the constraint must be chosen sufficient so that the probability of exceeding the linear range is small. This method works quite well as long as

---

[1]Presented at the Third Congress of International Federation of Automatic Control, London, June 1966.

the linear range of operation is large. However, in critical situations where the linear range of operation is not large, this approach gives excessively large errors.

Several other studies in recent years have been concerned with a rigorous account of the nonlinear aspect (as well as other aspects) of the stochastic control problem. These studies take various forms depending on the properties of the inputs, the sources of noise, the form of the nonlinearity, and the form of the plant. Recent extensive bibliographies and summaries of these results are given in references 2 and 3. The setting for most of these studies has been the Chapman-Kolmogoroff equations or the related Fokker-Planck equations which govern the behavior of the first probability density function. Such approaches hold much promise. At the present stage of development, however, solutions to only the simplest examples can be obtained.

What is needed is an intermediate approach. In this paper we will take a middle ground more closely allied with the former approach. The problem to be considered is roughly the Wiener filtering and control problem with the addition of a zero-memory nonlinearity preceding the plant. The scalar problem is illustrated in figure 1. Here as in the Wiener problem the plant we wish to control is given by its transfer function. Preceding the plant is a zero-memory nonlinearity f. By far the most important nonlinearity and the one to be emphasized here is that due to saturation. However, a good deal can be learned by specifying only general properties of the nonlinearity. The input $s(t)$ is the true signal entering the system. The measurement of the state of the system, or, alternatively, the incoming signal, is contaminated with unwanted noise $n(t)$. It is assumed that the inputs are stationary ergodic processes with known spectral densities, as in the Wiener theory. Thus we will not consider the first probability density function directly in the optimization. However, in the example we will take the noise to have a gaussian distribution as is usually the case, and will show that results are not very sensitive to large variations in the signal distribution. And last, the controller will be free to choose so as to minimize $E[\epsilon^2(t)]$ or $\lim_{T\to\infty}(1/2T)\int_{-T}^{T} \epsilon^2(t)dt$ while allowing operation of the system to extend into the nonlinear regions. That is, we want to have an optimization procedure for finding the best controller in which the system is allowed to operate in the nonlinear region.

APPROXIMATE SOLUTION FOR THE OPTIMUM CONTROLLER

The approach we take here utilizes the linearization of the nonlinearity as developed by Booton (ref. 4). This linearization idea has been widely used in analysis problems with much success, but not in synthesis problems. According to this linearization method the nonlinearity f is replaced by a scalar k which is dependent on $E[u^2(t)]$, the expected value, or $\overline{u^2(t)}$, the time average of the input to the nonlinearity (if the input is stationary and ergodic). An inherent assumption in the choice of the scalar k is that the input to the nonlinearity is gaussian. This assumption may not be accurate due

to the nonlinearity in the system and due possibly to a nongaussian input signal. Nevertheless, we will see that the controllers derived here will provide sufficient filtering that the approximation will be fairly good. Thus the approach is consistent with not considering the exact propagation of the first probability density function through the system. As a result of this approach it is clear that both the controller and the equivalent gain k of the nonlinearity are free and need to be determined so as to result in minimum error.

As is generally recognized nowadays, a very convenient setting for stochastic problems and one promising generalization is a Hilbert space. Although there is more than one way the space can be defined, here the stochastic process $x(t,i)$ will be taken to be an element of a Hilbert space in which the inner product is either

$$(v(t,i),\ w(\tau,j)) = E[v(t,i)w(\tau,j)]$$

or

$$(v(t,i),\ w(t+\tau,j)) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} v(t,i)w(t+\tau,j)dt$$

depending on whether the events or the times are fixed. Although these inner products will be identical due to the assumed stationarity and ergodicity, time averages will be more convenient for experimental purposes. Usually the event i is suppressed in the notation. The norm generated from the above inner product is

$$\|v(t)\|^2 = (v(t),\ v(t)) < \infty$$

Thus we take

$$\mathcal{H} = \{x(t,i) : \|x(t,i)\| < \infty,\ -\infty < t < \infty\}$$

Now redrawing figure 1 in the usual equivalent open-loop form because of linearity, we have figure 2. The box labelled H is assumed to be a linear operator $H: x(t) \to u(t)$ of the form

$$u(t) = Hx(t) = \int_{-\infty}^{\infty} x(\tau)h(t-\tau)d\tau$$

where $x(t) = s(t) + n(t)$, and is in $\mathcal{H}$. The operator P is similarly defined.

The problem to be solved, stated abstractly, is to find

$$\min_{k,H} \|s(t) - PkHx(t)\|^2 \tag{1}$$

subject to

$$g(\|Hx(t)\|^2) - k = 0 \tag{2}$$

Note that the describing function g, which is the exact relation between k and $\|Hx(t)\|^2$, is not specified. One could attempt to solve equations (1) and (2) by looking for a stationary value by means of classical variational methods. However, this approach will not lead to a solution as will be seen later. Instead we will follow a different route. Separating the minimization into two steps and interchanging the order of the operations on $x(t)$, we have the equivalent problem

$$\underset{k}{\text{Min}}\{\underset{H}{\text{Min}}\|s(t) - kPHx(t)\|^2\} \tag{3}$$

subject to

$$g(\|Hx(t)\|^2) - k = 0 \tag{4}$$

Assuming g has an inverse, letting $Z = kP$ and $T = kPH$, we see that equations (3) and (4) become

$$\underset{k}{\text{Min}}\{\underset{H}{\text{Min}}\|s(t) - Tx(t)\|^2\} \tag{5}$$

subject to

$$\|Z^{-1}Tx(t)\|^2 = g^{-1}(k) \tag{6}$$

Now consider the inner minimization of equation (5) subject to equation (6). It is known that a sufficient condition for an absolute minimum of this inner minimum is that

$$\|s(t) - Tx(t)\|^2 + \lambda\|Z^{-1}Tx(t)\|^2 \tag{7}$$

attain an absolute minimum, where λ is an unknown multiplier. The absolute minimum of equation (7) is expressed by the following.

THEOREM: Let Y be a closed convex subset of a Hilbert space $\mathcal{H}$. Given any $s(t) \in \mathcal{H}$, then a necessary and sufficient condition for the existence of a unique $y_o \in Y$ satisfying

$$\|s(t) - y_o(t)\|^2 + \lambda\|Z^{-1}y_o(t)\|^2 \leq \|s(t) - y(t)\|^2 + \lambda\|Z^{-1}y(t)\|^2 \quad \text{for all } y \tag{8}$$

is that

$$(y(t),\ s(t) - Uy_o(t)) = 0 \quad \text{for all } y(t) \tag{9}$$

where

$$U = I + \lambda Z^{-1*}Z^{-1}$$

and $*$ denotes the adjoint operator.

The proof is sketched in appendix A. Equation (9) is another form of the Wiener-Hopf equation which accounts for the restriction on the control function first developed by Newton (ref. 1) by variational principles. It should be noted that this theorem insures the existence and uniqueness of an absolute minimum rather than only a stationary value. This equation is appealing because it maintains close contact with the geometrical notion of orthogonal projection. That is, it says that the optimum y_o vector is one which makes the optimum error vector $s(t) - Uy_o(t)$ orthogonal to all arbitrary $y(t) \epsilon Y$.

The applicability of the preceding theorem is clear from an examination of the set Y, consisting of output vectors, which is generated by the mapping $T: x(t) \rightarrow y(t)$. This mapping produces the output set Y such that

$$Y = \{y(t) : y(t) = \int_{-\infty}^{\infty} x(\tau)q(t - \tau)d\tau \quad \text{for all} \quad t, x \epsilon \mathcal{H}\}$$

It is well known that as long as T satisfies the condition

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |q(t - \tau)|^2 \, dt \, d\tau < \infty$$

T will be linear, bounded, and completely continuous. Thus from the boundedness, $\|y(t)\| \leq M\|x(t)\| < \infty$ for all $x(t) \epsilon \mathcal{H}$ for some $M > 0$ so that $y(t) \epsilon \mathcal{H}$ or $Y \subset \mathcal{H}$. Also from the complete continuity of T, Y is compact and therefore closed; Y is also convex.

Now consider the minimization over k as given in equation (5). We can show what this step involves by examining the properties of the operator T which are contained in the orthogonal projection theorem of equation (8). Let k be arbitrary. Then take λ_1 and λ_2 to be two values of λ in equation (8) such that $0 < \lambda_1 < \lambda_2$, and let y_1 and y_2 be the corresponding optimum output vectors. Then for these two values of λ we have, from equation (8),

$$\|s(t) - y_2(t)\|^2 + \lambda_2\|Z^{-1}y_2(t)\|^2 \leq \|s(t) - y_1(t)\|^2 + \lambda_2\|Z^{-1}y_1(t)\|^2 \tag{10}$$

$$\|s(t) - y_1(t)\|^2 + \lambda_1\|Z^{-1}y_1(t)\|^2 \leq \|s(t) - y_2(t)\|^2 + \lambda_1\|Z^{-1}y_2(t)\|^2 \tag{11}$$

Solving equation (11) for $\|s(t) - y_1(t)\|^2$ and substituting in equation (10) we see that

$$\|s(t) - y_2(t)\|^2 + \lambda_2\|Z^{-1}y_2(t)\|^2 \leq \|s(t) - y_2(t)\|^2 + \lambda_1[\|Z^{-1}y_2(t)\|^2 - \|Z^{-1}y_1(t)\|^2] + \lambda_2\|Z^{-1}y_1(t)\|^2$$

or

$$\lambda_2[\|Z^{-1}y_2(t)\|^2 - \|Z^{-1}y_1(t)\|^2] \leq \lambda_1[\|Z^{-1}y_2(t)\|^2 - \|Z^{-1}y_1(t)\|^2]$$

Thus we conclude

$$\|Z^{-1}y_2(t)\|^2 \leq \|Z^{-1}y_1(t)\|^2 \tag{12}$$

Considering equations (12) and (11) we see

$$\|s(t) - y_1(t)\|^2 \leq \|s(t) - y_2(t)\|^2 \tag{13}$$

Thus the operator T which makes equation (7) attain an absolute minimum has the monotonic property

$$\lambda_1 < \lambda_2 \Rightarrow \|Z^{-1}y_2(t)\|^2 \leq \|Z^{-1}y_1(t)\|^2$$

$$\Rightarrow \|P^{-1}y_2(t)\|^2 \leq \|P^{-1}y_1(t)\|^2$$

$$\Rightarrow \|s(t) - y_1(t)\|^2 \leq \|s(t) - y_2(t)\|^2 \tag{14}$$

Obviously, this property holds for arbitrary linear plants.

To continue the minimization over k we must assume some properties for the describing function g. A particularly important class of describing functions is that described by the property that $k^2g^{-1}(k)$ is a monotonically decreasing function. For example, the describing function for a saturation nonlinearity or a relay satisfies this property. Now consider two values of k, $k_1 < k_2$. Then from the above property and equation (6), we will have

$$k_1 < k_2 \Rightarrow \|P^{-1}y_2(t)\|^2 \leq \|P^{-1}y_1(t)\|^2 \tag{15}$$

Then from the property of the optimum operator given in equation (14),

$$k_1 < k_2 \Rightarrow \|s(t) - y_1(t)\|^2 \leq \|s(t) - y_2(t)\|^2 \tag{16}$$

Now if the describing function is assumed to be restricted further by the commonly occurring property

$$0 < k < 1$$

it will be clear from the monotone property in equation (16) that there will be no minimum error, that is, no minimum of equation (5) subject to equation (6). Note that this conclusion has not required the exact form of the describing function g to be specified other than by its general properties

that it have an inverse, be greater than zero, and k^2g^{-1} be monotonically decreasing. This conclusion also shows that the stationary value which is sought in the variational approach does not exist.

Even though there is no minimum, from a practical viewpoint, one would be interested in determining the greatest lower bound, glb, of performance, that is,

$$\underset{k}{\text{glb}}\{\underset{H}{\text{Min}}\|s(t) - Tx(t)\|^2\} \tag{17}$$

subject to

$$\|Z^{-1}Tx(t)\|^2 = g^{-1}(k) \tag{18}$$

However, to do so, more must be assumed about the nonlinearity. Because of its obvious practical importance it will be assumed that the nonlinearity is due to saturation. In this case we may find the exact form of g^{-1} in equation (18) as $k \to 0$. For, from an analysis of the saturation describing function, it is known (ref. 4) that

$$k \to \frac{\sqrt{2/\pi}}{\|Z^{-1}y(t)\|/L} \quad \text{as} \quad \frac{\|Z^{-1}y(t)\|}{L} \to \infty$$

where L is the saturation value for the nonlinearity. In other words,

$$\|Z^{-1}y(t)\|^2 = \|Z^{-1}Tx(t)\|^2 \to \frac{2}{\pi}\frac{L^2}{k^2} \quad \text{as} \quad k \to 0 \tag{19}$$

Thus as the error becomes smaller as a result of decreasing k, this expression may be interpreted as the behavior of the restriction equation (18) by defining g^{-1}. But since, by definition $Z = kP$, equation (19) becomes

$$\|P^{-1}Tx(t)\|^2 \to \frac{2}{\pi}L^2 \quad \text{as} \quad k \to 0$$

Then since $\text{glb}\{k\} = 0$, the above relationship together with the property in equation (15) implies $\underset{k}{\text{lub}}\{\|P^{-1}Tx(t)\|^2\} = 2L^2/\pi$. Using this fact and the property in equation (14), equations (17) and (18) are equivalent to

$$\underset{H}{\text{Min}}\|s(t) - Tx(t)\|^2 \tag{20}$$

subject to

$$\|P^{-1}Tx(t)\|^2 = \|f[u(t)]\|^2 = \frac{2}{\pi}L^2 \tag{21}$$

The solution to these equations by conventional methods then gives the limiting value of performance. Note that no assumptions other than linearity are made concerning the plant. Although equation (19) is valid for a saturation nonlinearity, one could arrive at similar results for other nonlinearities by examining the limiting behavior of equation (18) for that particular nonlinearity.

It is apparent now that the optimum performance for the saturating case is related directly and simply to the performance for the completely linear case. In the linear case (denoted by subscript L), in which $f = 1$, equations (20) and (21) become

$$\underset{H}{\text{Min}} \|s(t) - Tx(t)\|_L^2 \tag{22}$$

subject to

$$\|u(t)\|_L^2 = \text{constant} \tag{23}$$

In comparing the performance for the saturation case (denoted by subscript s) with the linear case, we see that

$$\|u(t)\|_L = \|f[u(t)]\|_s = \sqrt{\frac{2}{\pi}}\, L$$

will imply

$$\|s(t) - Tx(t)\|_s = \|s(t) - Tx(t)\|_L$$

Thus the glb of the nonlinear error performance, as a function of the saturation level L, is the same as the optimum linear error performance as a function of $\|u(t)\|$ where $\|u(t)\|$ is replaced by $\sqrt{2/\pi}\, L$.

In order to illustrate the comparative performance between the linear and nonlinear cases it will be expedient to take an example. Since the details of the example are unimportant, we mention only that the example represents a hypothetical interception problem taken from a later section. For this example the optimum linear performance, where $f = 1$, is given in figure 3. In order to achieve results predicted by linear theory, it is customary to design such that $\|u(t)\|_L = L/2$ so that the system remains essentially linear. However in the nonlinear case we have seen that the glb of the error performance occurs when $\|f[u(t)]\| = \sqrt{2/\pi}\, L$. Thus there is a 60 percent increase in effective control in comparison to the linear case. This increase may be significant depending on the value of L. For example, in figure 3 it is clear that an increase in $\|u(t)\|$ by a factor of 1.6 will not be of much value in reducing minimum error at point A where $\|u(t)\|$ is large, but it will offer great reduction in error at point B where $\|u(t)\|$ is small. This comparison is shown even better perhaps in figure 4 (obtained in an obvious way by the discussion of the preceding paragraph) where the performances of these two systems

are given as a function of the saturation level L. (Note that for metric units in figures 3 and 4, 1 ft = 0.3048 m and 1 g = 980.7 cm/sec^2.)

Up to now only the glb of the error performance for the saturating system has been discussed. Even though this performance is not attainable from a practical viewpoint we are interested in obtaining performance which is reasonably close to the glb. Thus it will be desirable to consider how the system should be designed so that its performance will be close to this glb. This can be done by considering the describing function characteristic given in figure 5 where the dimensionless $m = \|f[u(t)]\|/L$ is given as a function of the equivalent gain k. For a given L and any choice of m,

$$\|f[u(t)]\| = mL , \qquad 0 < m < \sqrt{\frac{2}{\pi}}$$

and the corresponding error performance is obtainable from figure 3. Hence one may choose the value of m so that the error will be satisfactorily close to the glb performance. The decision regarding this choice of m can best be shown in figure 4 where the performance is shown for several values of m.

DIMENSIONLESS PERFORMANCE CURVES

It will be the purpose of this section to present some dimensionless performance curves for the saturating system. The motivation here is that one would like to be able to draw some general conclusions regarding optimal performance. A similar approach for linear systems was taken at about the same time by Coales (ref. 5) and Stewart (ref. 6), although the details of the two studies are slightly different. Since it is not possible to choose a system which is optimal for all inputs and plants, some narrowing of the task was required to carry out this approach. In both the above studies, the class of inputs and plants was restricted to a special but yet very important class practically. From these assumptions a set of dimensionless curves were obtained which are useful in determining optimal linear performance and the design of the optimal linear system.

Similar dimensionless performance curves for the saturating system may be obtained because of the relationships developed earlier. Since we utilize the results of reference 6 for linear systems, it will be desirable to summarize the assumptions in this reference along with a few remarks:

1. The class of inputs was restricted to a very common and important form which occurs in many physical problems; the power spectra of the signal and noise were defined, respectively, by $\Phi_S(\omega) = \sigma/\omega^4(\omega^2 + \xi^2)$ and $\Phi_N(\omega) = N$. This form for the signal is valid for many stationary and nonstationary processes. Furthermore, this form may often approximate a variety of experimentally determined input data. The other function for the noise is approximated by a constant, and this is a good approximation when the noise bandwidth is larger than the resultant system bandwidth.

2. The plant was assumed to be of the form c/s^2. Although the plant will generally be much more complicated, it was found that the more complex form is not only often unnecessary but undesirable. That is, the more complex form can often be approximated by the simpler form c/s^2. Moreover, it was shown (ref. 6) that the effect of additional dynamic terms is detrimental to optimum performance so that the simpler form should be striven for when there is control over the plant dynamics.

With these assumptions, the dimensionless error can be determined as a function of the dimensionless control function when the system is linear. This relationship is given in figure 6, where the dimensionless parameters β and ν are defined as: $\beta = \sqrt[6]{\sigma/N}$ and $\nu = \xi/\beta$. (Note that in fig. 6, 32.2 is a scalar without dimensions. Thus $\|u(t)\|/\sqrt{N\beta^5}$ is dimensionless.)

The corresponding dimensionless curves for the saturating system parallel figures 4 and 5 for the specific example discussed earlier. First, figure 7 gives the dimensionless glb of the error as a function of a dimensionless parameter l defined by $l = L/32.2\ N^{1/2}\beta^{5/2}$, where L is the saturation level of the forcing function. (Again note that both 32.2 and $L/N^{1/2}\beta^{5/2}$ are dimensionless.) These curves, valid for any set of conditions, might be useful in several ways. They might be used to determine the best performance which could be achieved in a specific case for purposes of comparing with other systems to indicate possibilities for improvement. They might also be used in preliminary design to evaluate the relative importance of those factors affecting the optimum performance. Second, in figure 8, the nonlinearity characteristic is given in dimensionless form. This figure, in conjunction with figure 6, is useful in deciding how close to the glb curve one wants to be. The detailed design of the optimum controller can be accomplished by combining the results in this paper with those of reference 6.

EXPERIMENTAL RESULTS

Since the results presented earlier depend on the describing function approximation, it is important to investigate the validity of this approximation. One expedient way of investigating this approximation is by means of an analog computer simulation, the results of which will be described briefly.

The example chosen for the simulation was a hypothetical interception problem in which the target executes a constant amplitude switching of acceleration with a Poisson distribution for the length of time between switches. The spectrum for such a function is known to be

$$\Phi_s(\omega) = Ka^2/\pi\omega^4(K^2 + \omega^2)$$

so that the dimensionless curves just presented are applicable. The other functions, the noise and plant, are assumed to be of the form described earlier, that is, the noise spectrum is a constant and the plant is of the form c/s^2. The following typical numerical values for the parameters were

taken: a = 0.968 g, K = 0.431, and N = 15 $ft^2/rad/sec$. Two values of the saturation level of the forcing function were chosen as L = 2.87 g and 2.3 g. The systems were designed so that the theoretical errors were reasonably close to the glb curve as indicated in figure 4.

The experimental results which were obtained (using norms based on time averages for convenience) are shown in figure 4, and the errors may be seen to be reasonably close to the theoretical values with which they should be compared. Also see table I for the comparison. It is difficult to estimate the accuracy of the experimental results because although the measurements extend over a few minutes the results varied somewhat with different noise and signal sample time histories.

Several other interesting experimental results were obtained which are difficult or impossible to examine analytically. We itemize these results in the following. First, measurements were made directly from experimental data of the distribution at the input to the nonlinearity and the equivalent gain. Although data cannot be given here, the distribution was exceptionally normal and the equivalent gain calculated from the well-known relation

$$k = \frac{\int_{-\infty}^{\infty} uf(u)p(u)du}{\int_{-\infty}^{\infty} u^2p(u)du}$$

was in excellent agreement with theory. Second, the saturating element was replaced by a constant gain equal to the theoretical value for which the system was designed. The experimental value for the error performance is given in table I; it agrees well with the theoretical value as it should. Third, the effect of the first probability density function of the signal on the results was examined. Interest centers on this effect because the theory presented depends on only the second probability density function. However, it is clear that a rigorous account of nonlinear system behavior must depend on the first probability density function. Thus if the method is to be useful, the results must not be very sensitive to the first probability density function. To investigate this effect two signals were chosen with the same spectral densities but with widely different first probability density functions. The distribution of the random square-wave of acceleration used earlier was, of course, two impulses, at +a and -a. In addition, a signal with the same spectrum but with a gaussian distribution was used. From the results indicated in table I it appears that the error for the gaussian case is fairly close to but slightly greater than for the random square-wave signal. Since the difference between these two distributions is extreme, it appears that the results are fairly insensitive to distribution. Fourth, a series of experimental runs were made to investigate the local behavior of the error performance for the saturating system with the random square-wave signal. That is, since the theory presented is only approximate, it is of interest to measure the gradient of the error with respect to the parameters, α_i, of the optimum controller, or more properly $\Delta\epsilon/\Delta\alpha_i$. The controller, which is given by the ratio of a second-order polynomial to a third-order polynomial, contains six parameters. By varying each of these parameters individually and sufficiently to cause a

small change in error, $\Delta\epsilon/\Delta\alpha_i$ was found to be positive for each parameter α_i. Thus the design was experimentally verified to be optimum, at least locally.

CONCLUDING REMARKS

The method given here shows that when operating in critical regions one can obtain significant improvement over linear performance by nonlinear operation. The method is appealing because it agrees with ones intuitive notions concerning optimum performance for saturating systems. However, the validity of the method depends on the accuracy of the describing function concept which generally cannot be predicted a priori. The simulation results indicate the approximation is valid to about the same accuracy as is usually associated with the describing function.

Ames Research Center
National Aeronautics and Space Administration
Moffett Field, Calif., Feb. 24, 1966

APPENDIX A

PROOF OF THEOREM

THEOREM: Let Y be a closed convex subset of a Hilbert space $\mathcal{H}$. Given any $s(t) \in \mathcal{H}$, then a necessary and sufficient condition for the existence of a unique $y_o \in Y$ satisfying

$$\|s(t) - y_o(t)\|^2 + \lambda\|Z^{-1}y_o(t)\|^2 \leq \|s(t) - y(t)\|^2 + \lambda\|Z^{-1}y(t)\|^2 \quad \text{for all } y \tag{A1}$$

is that

$$(y(t),\ s(t) - Uy_o(t)) = 0 \quad \text{for all } y \tag{A2}$$

where

$$U = I + \lambda Z^{-1*}Z^{-1} \tag{A3}$$

and $*$ denotes the adjoint operator.

To prove the existence of a minimizing vector, let

$$m = \text{glb}\{\|s(t) - y(t)\|^2 + \lambda\|Z^{-1}y(t)\|^2 : y \in Y\}$$

Now choose a sequence $\{y_q\}$ such that

$$\|s(t) - y_q(t)\|^2 + \lambda\|Z^{-1}y_q(t)\|^2 \to m$$

It can be shown that $\{y_q\}$ is a Cauchy sequence. For by the parallelogram law,

$$\begin{aligned}\|y_p - y_q\|^2 + \lambda\|Z^{-1}(y_p - y_q)\|^2 = {} & 2(\|y_p - s\|^2 + \lambda\|Z^{-1}y_p\|^2) \\ & + 2(\|s - y_q\|^2 + \lambda\|Z^{-1}y_q\|^2) \\ & - 4\|\tfrac{1}{2}(y_p + y_q) - s\|^2 \\ & - 4\|Z^{-1}\tfrac{\lambda^2}{2}(y_p + y_q)\|^2 \end{aligned} \tag{A4}$$

Due to the assumed convexity property of Y and the definition of m, the last two terms are less than $-4m$. Hence

$$\|y_p - y_q\|^2 + \lambda\|Z^{-1}(y_p - y_q)\|^2 \leq 2(\|y_p - s\|^2 + \lambda\|Z^{-1}y_p\|^2)$$

$$+ 2(\|s - y_q\|^2 + \lambda\|Z^{-1}y_q\|^2) - 4m$$

$$\to 0$$

because of the way in which the sequence was chosen as described following equation (A3). Thus $\{y_q\}$ is a Cauchy sequence. Since Y is a closed subset of $\mathcal{H}$, it is complete, so that $y_q \to y_o \in \mathcal{H}$. It will then follow that

$$\|s - y_q\|^2 + \lambda\|Z^{-1}y_q\|^2 \to \|s - y_o\|^2 + \lambda\|Z^{-1}y_o\|^2$$

Thus

$$\|s - y_o\|^2 + \lambda\|Z^{-1}y_o\|^2 = m$$

and the minimum is actually attained.

To prove uniqueness, assume there are two elements y_o and z_o in Y that are optimum. Let the minimum value be m. Now consider

$$\|y_o - z_o\|^2 + \lambda\|Z^{-1}(y_o - z_o)\|^2$$

By means of the parallelogram law we can show, as in the preceding paragraph, that

$$\|y_o - z_o\|^2 + \lambda\|Z^{-1}(y_o - z_o)\|^2 \leq 0$$

Hence it follows $y_o = z_o$ and the optimum is unique.

To find the necessary and sufficient condition which y_o must satisfy, we consider the following expression which can be expanded as follows

$$\|s - y\|^2 + \lambda\|Z^{-1}y\|^2 = \|s - y_o\|^2 + \lambda\|Z^{-1}y_o\|^2 + \|y - y_o\|^2$$

$$+ \lambda\|Z^{-1}(y - y_o)\|^2 - (y - y_o, s - y_o)$$

$$+ \lambda(Z^{-1}(y - y_o), Z^{-1}y_o) - \overline{(y - y_o, s - y_o)}$$

$$+ \overline{\lambda(Z^{-1}y_o, Z^{-1}(y - y_o))} \qquad (A5)$$

Using the adjoint operator denoted by $*$

$$(y - y_o, s - y_o) + \lambda(Z^{-1}(y - y_o), Z^{-1}y_o) = (y - y_o, s - Uy_o) = (y, s - Uy_o) \tag{A6}$$

where

$$U = I + \lambda Z^{-1*}Z^{-1}$$

and $y - y_o$ is an arbitrary element of Y and is relabelled as y. Now equation (A5) becomes

$$\|s - y\|^2 + \lambda\|Z^{-1}y\|^2 = \|s - y_o\|^2 + \lambda\|Z^{-1}y_o\|^2 + \|y - y_o\|^2$$

$$+ \lambda\|Z^{-1}(y - y_o)\|^2 - 2R_e(y, s - Uy_o) \tag{A7}$$

It is clear that a sufficient condition for y_o to satisfy equation (A1) is that

$$(y, s - Uy_o) = 0 \quad \text{for all } y \tag{A8}$$

For necessity, we assume equation (A1) holds. Then in view of equation (A7), it will be necessary that

$$\|y - y_o\|^2 + \lambda\|Z^{-1}(y - y_o)\|^2 - 2R_e(y, s - Uy_o) \geq 0 \quad \text{for all } y \tag{A9}$$

If ρ is any scalar, ρy belongs to Y. Hence for (A9) to hold for all y, it will be necessary that

$$\|y - y_o\|^2 + \lambda\|Z^{-1}(y - y_o)\|^2 - 2\rho R_e(y, s - Uy_o) \geq 0$$

for any scalar ρ. This will require

$$(y, s - Uy_o) = 0 \quad \text{for all } y \tag{A10}$$

REFERENCES

1. Newton, G. C., Jr.: Compensation of Feedback Control Systems Subject to Saturation. J. Franklin Inst., vol. 254, no. 4, Oct. 1952, pp. 281-296; and no. 5, Nov. 1952, pp. 391-413.

2. Wonham, W. M.: Stochastic Problems in Optimal Control. IEEE Internatl. Convention Record, vol. 11, part 2, March 1963, pp. 114-124.

3. Kushner, H. J.: On the Differential Equations Satisfied by Conditional Probability Densities of Markov Processes, With Applications. Tech. Rep. 64-6, RIAS, March 1964.

4. Booton, R. C., Jr.: Nonlinear Servomechanisms With Random Inputs. Rep. 70, Massachusetts Inst. Tech., Aug. 1953.

5. Coales, J. F.; and Lawrence, P. J.: The Preparation of Charts and Tables for the Optimization of Automatic Control Systems With Random Inputs. Proceedings of the First International Congress of the International Federation of Automatic Control, vol. II, Butterworths, London, 1961, pp. 753-760.

6. Stewart, E. C.: An Explicit Linear Filtering Solution for the Optimization of Guidance Systems With Statistical Inputs. NASA TN D-685, 1961.

TABLE 1.- COMPARISON OF THEORETICAL AND EXPERIMENTAL RESULTS

	Signal distribution	Saturation value L, g	m	glb$\|\|\epsilon(t)\|\|$, ft	Theoretical performance $\|\|\epsilon(t)\|\|$, ft	Experimental performance $\|\|\epsilon(t)\|\|$, ft
Saturating system	Random square-wave signal	2.87	0.753	31.55	33.4	35-38
	Random square-wave signal	2.3	.738	41.0	45.4	44-47
Saturating element replaced by gain	Random square-wave signal	2.87	.753	31.55	33.4	34-35
Effect of signal distribution	Gaussian	2.87	.753	31.55	33.4	41

Note that for metric units, 1 ft = 0.3048 m and 1 g = 980.7 cm/sec^2.

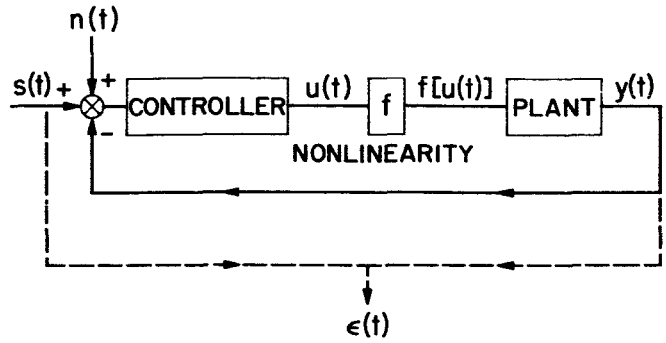

Figure 1.- Nonlinear filtering and control problem.

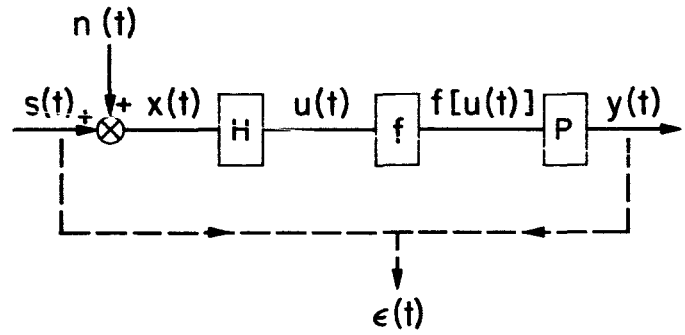

Figure 2.- Equivalent open-loop problem.

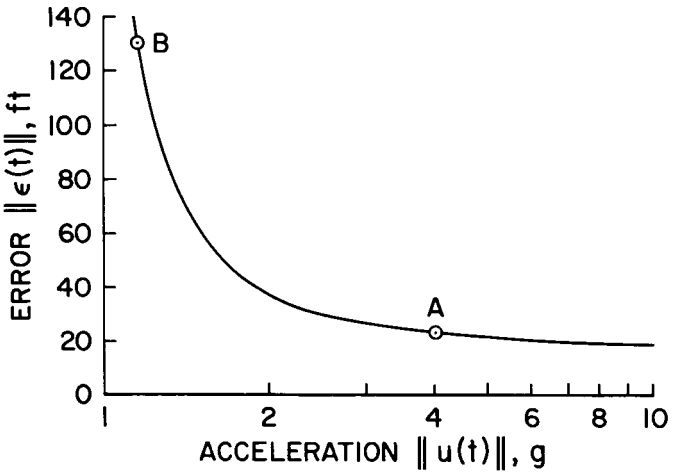

Figure 3.- Example error performance for linear system.

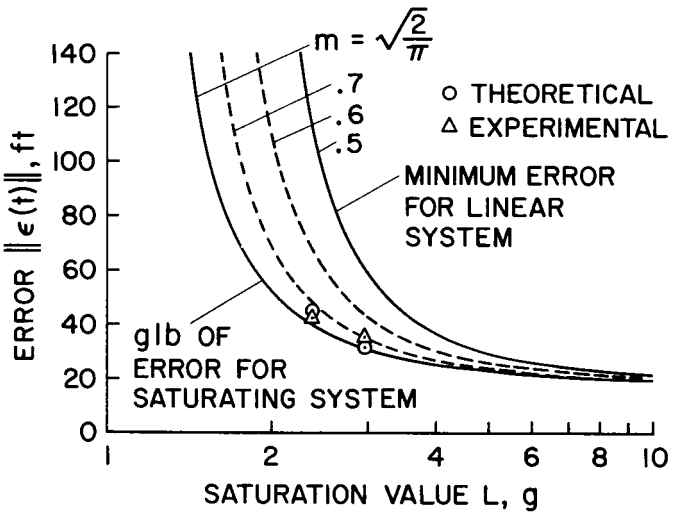

Figure 4.- Example error performance as a function of saturation level.

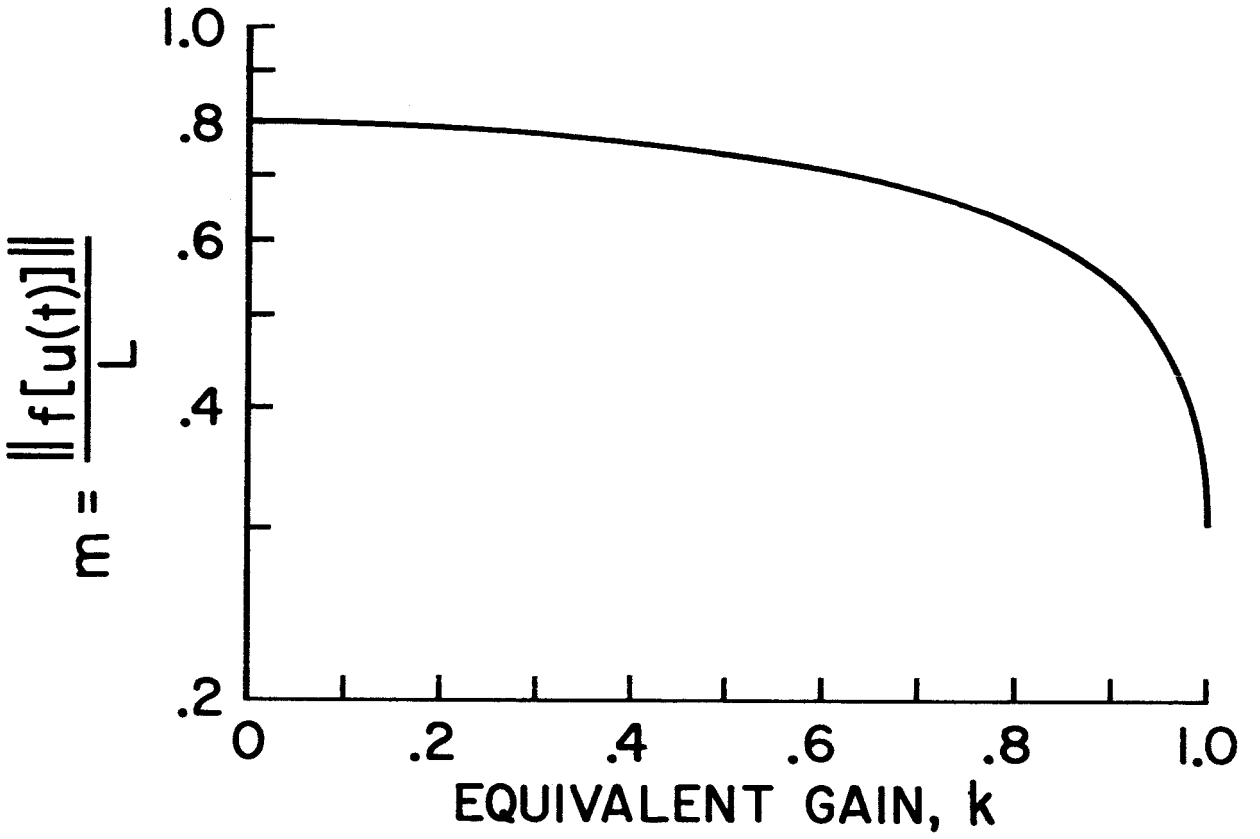

Figure 5.- Describing function characteristic.

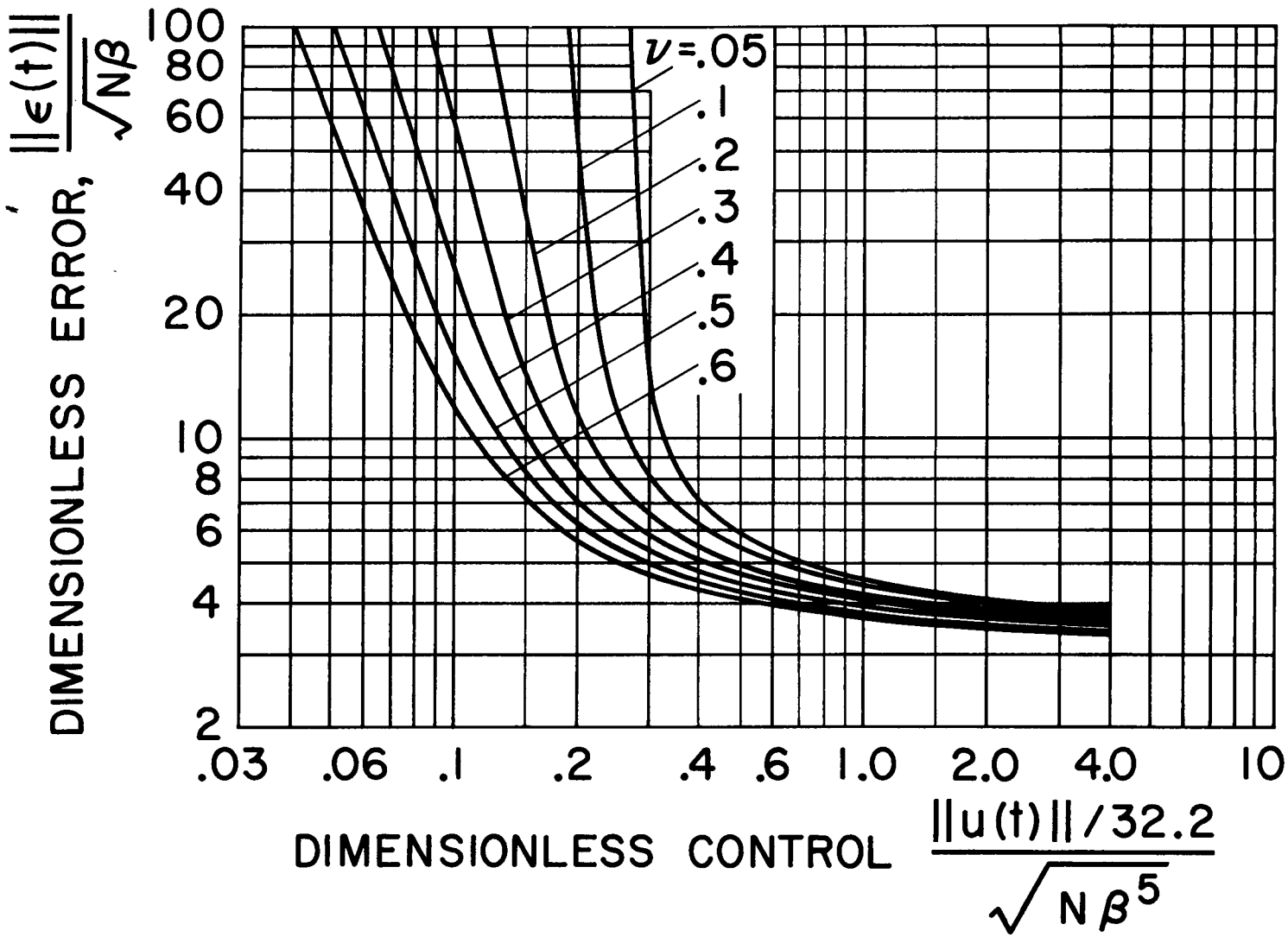

Figure 6.- Dimensionless performance for linear system.

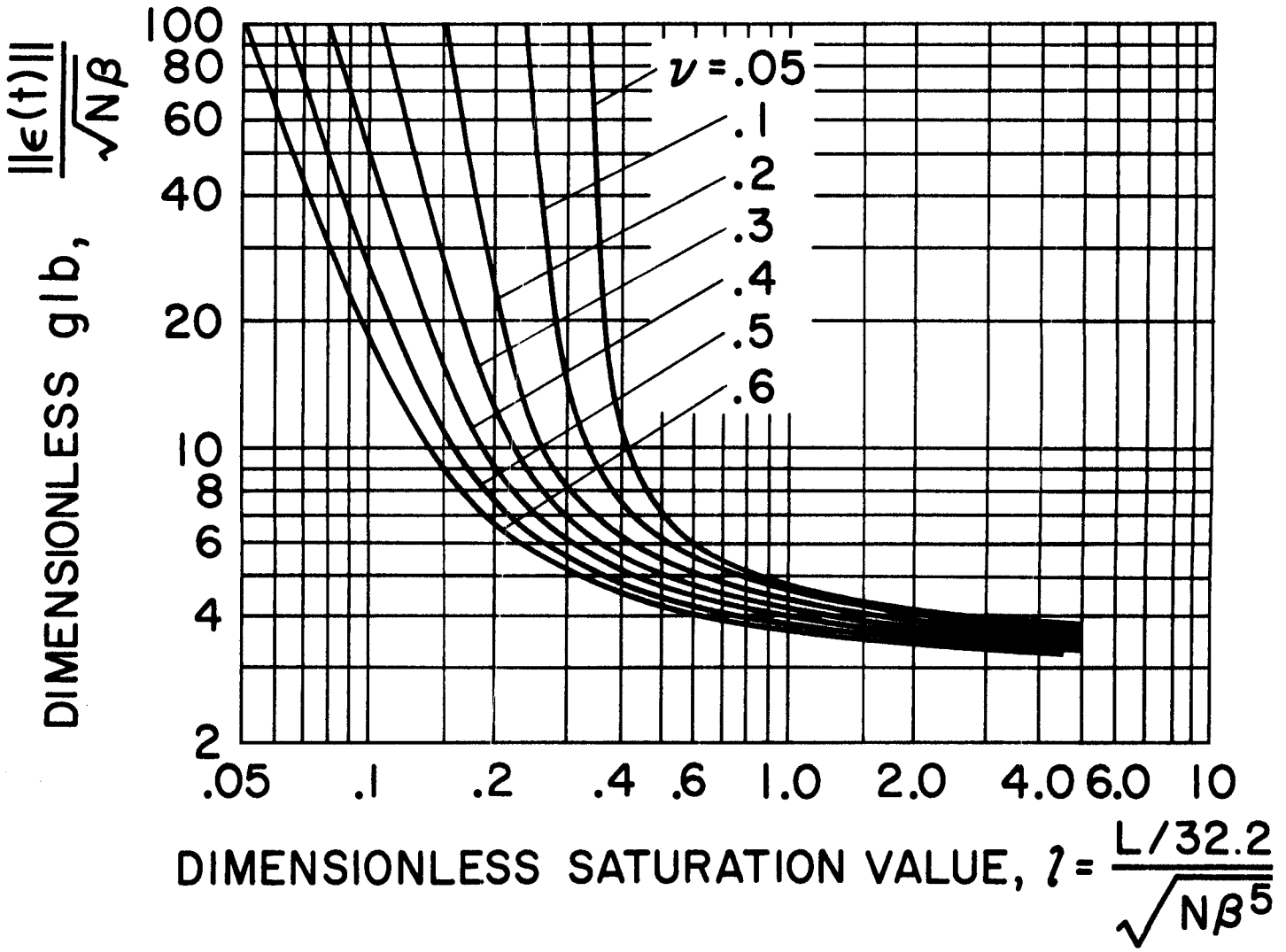

Figure 7.- Dimensionless performance for saturating system.

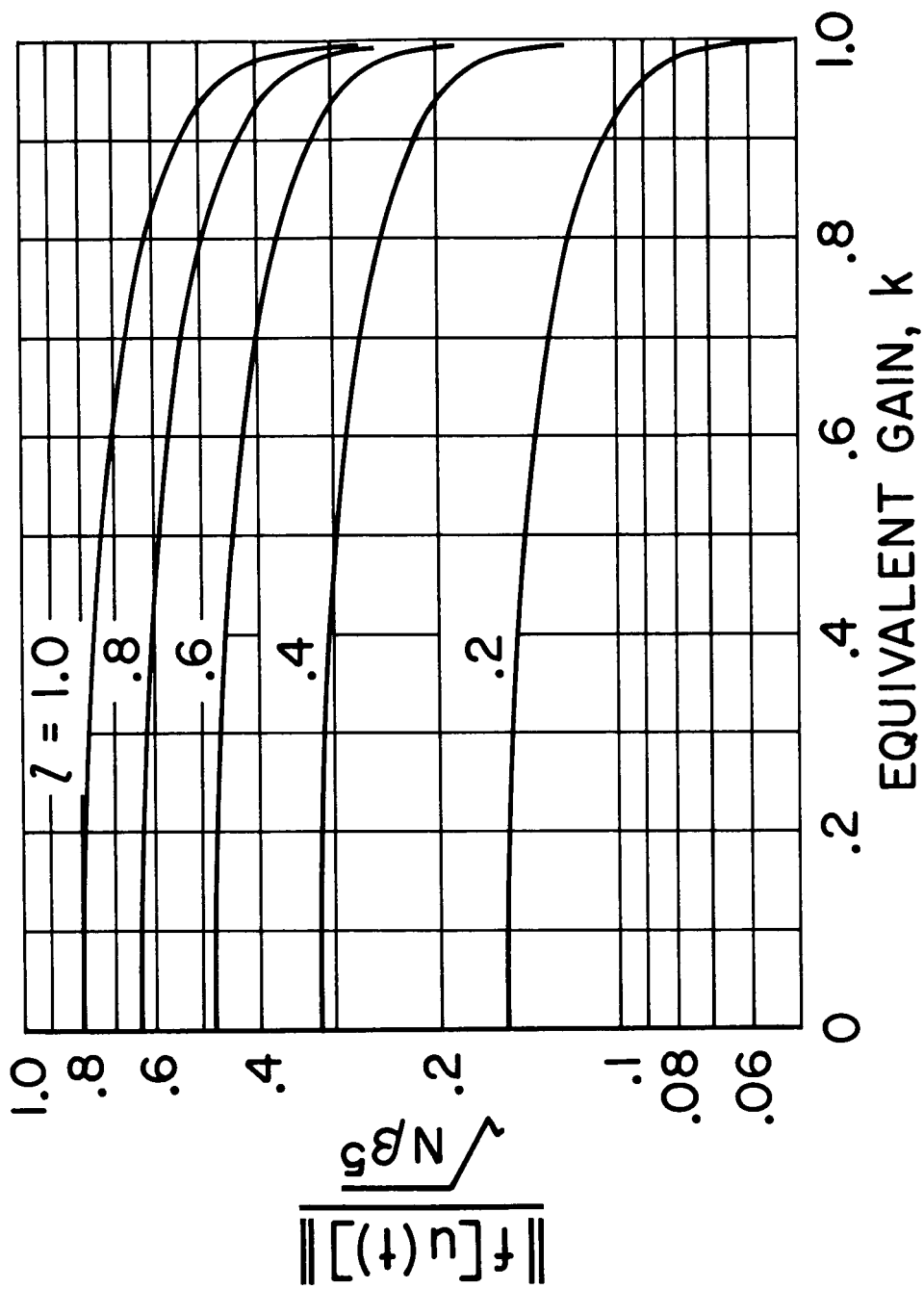

Figure 8.- Dimensionless describing function characteristic.

NASA-Langley, 1966 A-1098