This report presents a general study of the stability of nonlinear as compared to linear control systems. The analysis is general and, therefore, applies to other types of nonlinear biological control systems as well as the cardiovascular control system models. Both inherent and numerical stability are discussed for corresponding analytical and graphic methods and numerical methods.
STUDY REPORT

ON

GUIDELINES AND TEST PROCEDURES FOR

INVESTIGATING STABILITY OF NONLINEAR

CARDIOVASCULAR CONTROL SYSTEM MODELS

11 November 1974

Prepared By

D. G. Fitzjerrell
<table>
<thead>
<tr>
<th>No.</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>INTRODUCTION AND CONCLUSIONS</td>
<td>1</td>
</tr>
<tr>
<td>2.0</td>
<td>STABILITY OF LINEAR VERSUS NONLINEAR CONTROL SYSTEMS</td>
<td>2</td>
</tr>
<tr>
<td>3.0</td>
<td>ANALYTICAL AND GRAPHIC METHODS</td>
<td>5</td>
</tr>
<tr>
<td>4.0</td>
<td>NUMERICAL METHODS</td>
<td>9</td>
</tr>
</tbody>
</table>
1.0 INTRODUCTION AND CONCLUSIONS

In reviewing the literature on stability of control systems, one finds a great deal of very powerful and satisfying analytical machinery which is available for the study of linear systems. Most of these analytical tools are developed to a very high degree for application to non-biological control systems because most of these systems are linear. Even the several good references for the analysis of biological control systems, Riggs (1970), Grodins (1963), Milhorn (1966), etc., have excellent developments of general theory for linear control systems. There is, by comparison, very little treatment of the nonlinear control system. This seems particularly discouraging since even relatively simple biological systems are often nonlinear. The fact is that no such general theory exists for the analysis of nonlinear systems. Furthermore, the nature of such systems is such that such a theory may never be found. Whereas a single transfer function characterizes the entire behavior of a linear system (i.e., specifies its response to any arbitrary input), a nonlinear system presents a whole new problem whenever the form or even the amplitude of forcing is changed, or whenever initial conditions change.

This outlook is not quite so bleak since the large digital computer has become available. Most, even large sets of, complex nonlinear differential equations can be solved with numerical methods very satisfactorily with high speed digital computers. The numerical analysis of nonlinear differential equations still become important when considering the optimization of solutions for accuracy versus computation time. The question of stability is also not so critical to the bioengineer or physiologist (e.g., the design engineer is vitally concerned with stability because he does not want to build an unstable system; the physiologist is confronted by an existing system whose stability is usually obvious). Stability considerations for cardiovascular control systems (which is certainly a nonlinear system) are even less critical because of the inherent stability of the real system. Instabilities encountered in models of the cardiovascular system are usually, then, injected into the system model through inaccurate formulation or they exist as a result of numerical instability. In either case, such instabilities are usually moved out of
the operating range of the model with relative ease by reformulation, by adding various filters or lags, or by decreasing the numerical integration step size.

This study report will review the literature to substantiate these remarks and conclusions and to set forth some general guidelines and techniques which various modelers have found useful in dealing with the stability of nonlinear systems. Most of the remarks will be generally applicable to all biological control systems except where specific note is made to the cardiovascular system. This report will also review some of the more general or useful analytical and graphic methods which can be applied to the analysis of certain nonlinear systems. Care has been taken to make note of the applicability of each method and its limitations.

2.0 STABILITY OF LINEAR VERSUS NONLINEAR CONTROL SYSTEMS

The following remarks are typical of those found in the literature regarding the stability of linear versus nonlinear control systems:

"There is a certain feeling of disappointment in not finding a sufficiently general theory of nonlinear systems available for our use". - Grodins, (1963)

"The question of applicability of the conclusions of the simple stability analysis to more general nonlinear equations remains unanswered." - Carnahan, Luther, and Wilkes (1969).

"Testing the stability of a linear system is a simple matter. If the characteristic equation for the system has any roots with non-negative real parts, it is unstable, and that's that. The output of a stable linear system is bounded for any bounded input. Furthermore, for any given constant input, the output will approach a single constant value as the transient response dies away with the passage of time. In nonlinear systems, considerations of stability are much more complicated, . . . . There are no truly general tests for stability in nonlinear systems." - Riggs (1970).

Stability is a somewhat ambiguous term and appears in the literature with a variety of qualifying adjectives (inherent, partial, relative, weak, strong,
absolute, etc.). In general, a control system (or its representative differential equations) is unstable if perturbation of the input signals or error signals (or numerical errors from erroneous initial conditions or local truncation or round-off errors) are propagated without bound throughout the system (or subsequent calculations). Note that the question of stability of the real system cannot be answered from studies of the stability of the differential equations representing the real system unless they are exact representations. Approximations of the real system or numerical approximations of the solution of such equations, either one or both, can introduce instability. Numerical instability, then, is not a different phenomenon, but rather is related to the source of the instability. Certain equations with specified initial conditions cannot be solved by any step-by-step (numerical) integration procedure without exhibiting instability, and are said to be inherently unstable. Inherent instability is associated with the equation being solved and the initial conditions specified, and does not depend on the particular algorithm being used to approximate its solution. Since the stability of the real system is usually not in question for biological systems, the discussion of stability in this report will center upon inherent instability of the equations simulating the real system and numerical instability of their solution approximations.

For a better understanding of linear versus nonlinear systems, consider a nonlinear system where the output, $y(t)$, is governed by one or more known differential equations. Suppose the system has come into equilibrium with some constant forcing function (which may, in fact, be zero) so that $y(t)$ is no longer changing with time. But, if $y(t)$ is no longer changing with time, its equilibrium value, $y_{eq}$, must satisfy the differential equations with all of their time derivatives set to zero, and with time approaching infinity. For a given constant input, a linear system can have but one such equilibrium point, but a nonlinear system may have none, one, or several. If there are several, the stability of the system in the neighborhood of each equilibrium point must be investigated individually. Assume an autonomous system, i.e., systems with constant coefficients which, when displaced from some equilibrium state at time zero, undergo their natural
response. Systems with a constant input imposed as a step function at time zero are still considered to be autonomous because a step input at time zero is equivalent to imposing initial conditions with respect to the new equilibrium state required by the constant input. It is, in fact, often convenient to define $y(t)$ as its deviation from whatever $y_{eq}$ is being investigated. When this is done, $y_{eq} = 0$, and there is little point in distinguishing between the instantaneous removal of a previous input at time zero (which is really what initial conditions amount to) and the instantaneous imposition of a constant input at time zero. To study the stability of a given equilibrium state, we displace $y(t)$ from $y_{eq}$ by a small amount at time zero and see whether or not $y(t)$ returns asymptotically to $y_{eq}$ as $t \to \infty$. If it does, that particular equilibrium state is at least locally asymptotically stable. If it does not, that equilibrium state is unstable. (Note that in a nonlinear system, instability at a given equilibrium point does not imply that the outputs will necessarily increase without bound as $t \to \infty$. The nonlinear system may approach some other equilibrium point which is stable, or it may approach a stable cycle of activity as a limit which will be discussed below under practical stability.) If $y(t)$ eventually returns to $y_{eq}$ regardless of how far it is displaced, then $y_{eq}$ is said to be globally asymptotically stable.

This discussion brings up the point of practical stability as compared to theoretical stability (La Salle and Lefschetz (1961)). The question of practical stability shows why some stability investigations should not be taken too seriously. Investigations that take into account only the linear approximation fall into this category. The point is that theoretical stability and even asymptotic stability by themselves may not assure practical stability. One needs to know the size of the region of asymptotic stability, and then based on estimates of the conditions under which the system will actually operate, requirements on its performance, etc., one can judge whether or not the system is sufficiently stable to function properly.
and one may be able to see how to improve stability. Having decided that asymptotic stability is not, by itself, sufficient for practical stability, one might be inclined to conclude that it is, however, always a necessary condition. This, too, is incorrect. The desired state of a system may be mathematically unstable and yet the real system may oscillate sufficiently near this state that its performance is acceptable. Many aircraft and missiles behave in this manner. (One of the most aerodynamically "stable" aircraft, the DC-3, flew with a built-in "dutch roll" oscillation.)

3.0 ANALYTICAL AND GRAPHIC METHODS

With all the proper cautions on inferring stability of real systems from their representative equations, it is still necessary to understand the basic question of stability inherent in some equations which one might use to represent real systems. Consider the simple ordinary differential equation,

$$\frac{dy}{dx} = x + y,$$

(1)

for which the analytical solution subject to the initial condition $y(x_0) = y_0$ can be found to be

$$y(x) = -x - 1 + \left[1 + x_0 + y(x_0)\right] e^{-x_0} e^x.$$

(2)

With the initial condition $y(x_0) = -1$, the analytical solution is

$$y(x) = -x - 1.$$

(3)

Thus, the exponential term in the general solution vanishes because of the particular choice of the initial condition. Even a very tiny change in the initial condition (for example, $y(x_0) = -0.99999$) will eventually cause a drastic change in the magnitude (even the sign in this case) of the solution for large values of $x$. 
Therefore, even though the multiple of the exponential term is quite small, the contribution of the exponential term will eventually swamp the contribution of the linear terms in the solution. It can readily be seen that any attempt at solving such an equation with numerical methods with an integration step size however small will rapidly encounter an unstable condition. For example, even if the initial condition is error-free for the first step, the initial conditions for subsequent steps will inevitably contain errors introduced by truncation and round-off in preceding steps; the calculated solution for large \( x \) will bear no resemblance to the true solution. It is, therefore, apparent that, however frustrating analytical and graphic techniques may be in studying stability of nonlinear systems, information regarding inherent stability can oftentimes only be attained through the use of such methods. Of course, there is always the temptation to search for new information regarding the real system operation and reformulate the model to attempt to circumvent the problem rather than encounter this frustration.

There are, however, a few methods which can be applied with relative ease to certain nonlinear systems (Truxal (1955), Mishkin and Braun (1961), Thaler and Pastel (1962), Grodins (1963), Riggs (1970)). These several techniques will be discussed in general terms with particular attention to their applicability and limitations. Detailed development of the methods with examples can be found in the several good references listed above.

The *describing function* method is one of the most used (and abused) methods for studying systems which exhibit a limit cycle behavior. Its usefulness is restricted to predicting the approximate amplitude and frequency of the limit cycle in one general class of nonlinear feedback systems. The output of a linear feedback system can show sustained oscillations only in response to a periodic input. But certain kinds of nonlinear feedback systems may exhibit sustained oscillations, called limit cycles, even when the input is constant. If the limit cycle is stable, the cyclic behavior characteristic of the system is approached as a limit with increasing time even if the starting point in state space (determined by the initial conditions) is some distance away from the path of the limit cycle in state space. The variation of bath temperature or the operation of a thermostat is a
limit cycle, but limit cycles may also occur in systems where the nonlinearity is of the "no-memory" type. Many cyclic phenomena in biological systems, e.g., variations in population density of certain species, cyclic alteration of wakefulness and sleep even in a constant environment, the menstrual cycle, "biological clocks", pacemaker activity, are either probably or certainly best explained as limit cycles.

The describing function technique assumes that there is only a single nonlinear block included in an otherwise linear system and that the input to this block is a pure sinusoid. This amounts to assuming that the linear portion of the system filters out all higher harmonics from the output of the nonlinear block, so that the input to the latter can indeed be a pure sinusoid. Under these conditions, the fundamental term in the Fourier expansion of the nonlinear block output need be the only one considered, and it is used to define an approximate frequency transfer function (or describing function) for this block. Nyquist type stability tests may be applied and closed-loop frequency response curves obtained. However, the method yields no information about transient response, and cannot be used if more than one nonlinearity is present. This technique often correctly predicts the existence of a limit cycle, and, what is more, forecasts the frequency and amplitude of the cycle quite accurately. But there its talents end. It is not a method for determining the stability of the limit cycle. The basic assumption that a stable cycle exists for one particular amplitude and frequency and the steady-state condition are counter to stability testing.

Phase plane analysis is a graphic technique which gives information about transient behavior. It does not resemble any of the linear techniques, but involves plotting (for a second-order system) a family of curves relating $\dot{y}$ and $y$ for each of a large number of initial conditions. The $(y, \dot{y})$ coordinate system is called the phase space, each curve is called a phase trajectory, and the family of curves is the phase portrait. From the latter, a variety of information about the transient response can be derived. Equilibrium points can be determined by setting $\dot{y}(t)=0$ (the $y$ intercepts on the phase plane plot), and by examining the behavior of the trajectory in the neighborhood of the equilibrium points, some ideas about stability
can be derived. Although theoretically applicable to systems of any order, phase plane analysis is practical only for first and second-order systems.

Since most biological systems contain several nonlinearities and/or are greater than second-order, neither of the first two methods are satisfactory for this application.

Another method for examining transient behavior (i.e., stability) in the neighborhood of equilibrium points which has somewhat more general application is the perturbation method, where \( y(t) \) is "perturbed" by moving it a small distance away from \( y \) in state space. The differential equations describing the system are assumed to be linear for this small perturbation, and the familiar tests for stability of linear systems are applied. The set of linear differential equations for small perturbations from equilibrium may equally well be obtained by applying Taylor's series for functions of several variables and retaining only the linear terms of the series expansion. To write a Taylor's series for a function of several variables, one deals with each variable and its corresponding small increment separately. The total change in the value of the function, then, is simply the sum of all of the partial changes thus obtained. The resulting linear approximations of the nonlinear equations can be analyzed provided that the attention is confined to a sufficiently small neighborhood of an equilibrium point. This method is a valid test of stability only within the immediate neighborhood of an equilibrium point.

The second method of Liapunov is a much more general way of proving that a system, starting from some specified initial state, will approach an equilibrium point. Note the wording! Failure of the second method of Liapunov to prove stability does not necessarily imply that the system is unstable. Consider a system with \( n \)-dimensional state space. Defining time as an additional coordinate, the corresponding solution space is \( (n+1) \)-dimensional. Then asymptotic stability about the equilibrium point demands that, as \( t \) moves from zero toward infinity, the solution must approach closer and closer to the time axis where all the other coordinates are zero. One then tries to find a guiding function (called a Liapunov function) whose slope, with respect to the time axis, will surely be negative (or, at
worse, zero for only an instant). The Liapunov function must depend upon the state variables in such a way that at time, $t$, any point $t, y(t)$ in solution space lies upon the "surface" defined by the equilibrium point. It is not necessary to find a specific Liapunov function, it is only necessary to demonstrate that one exists.

### 4.0 NUMERICAL METHODS

Since, as has been pointed out earlier, there is no sufficiently general method for the analysis of nonlinear systems, the biological modeler usually turns to the computer and numerical methods for their solution. This section will, therefore, discuss stability with respect to various one-step and multistep algorithms for step-by-step numerical integration.

Inherent instability, as mentioned above, is associated with the equation being solved and the initial conditions specified, and does not depend on the particular algorithm being used. Depending on the equation being solved, its initial conditions, and the particular one-step method being used, another form of instability, numerical instability, may be observed, even when the equation is not inherently unstable. This phenomenon is related to the step size chosen, and is perhaps seen most easily by examining the Euler method where the total error at $x_{i+1}$ is related to the total error at $x_i$ by

$$
epsilon_{i+1} = \epsilon_i + h \left[ f(x_i, y_i) - f(x_i, y(x_i)) \right] - \frac{h^2}{2} f' \left( \xi, y(\xi) \right)$$

where $x_i < \xi < x_{i+1}$. From the differential mean-value theorem we may write

$$f(x_i, y_i) - f(x_i, y(x_i)) = (y_i - y(x_i)) \frac{\partial f}{\partial y} \bigg|_{x_i, \alpha}$$

with $\alpha$ in $(y_i, y(x_i))$. Since $\left[ y_i - y(x_i) \right]$ is just $\epsilon_i$, equation (4) may be written

$$\epsilon_{i+1} = \epsilon_i \left( 1 + h \frac{\partial f}{\partial y} \bigg|_{x_i, \alpha} \right) - \frac{h^2}{2} f' \left( \xi, y(\xi) \right),$$

$\xi$ in $(x_i, x_{i+1})$, $\alpha$ in $(y_i, y(x_i))$. (6)
The first term on the right-hand side of (6) is the contribution of the propagated error to the error at \( x_{i+1} \) while the second term is the local truncation error. Clearly, if \( \frac{\partial f}{\partial y} \) is negative, then a value of \( h \) can be found which will make \( \left[ 1 + h \left( \frac{\partial f}{\partial y} \right) \right] < 1 \), and the error will tend to diminish or die out: the solution will be stable. If \( \left[ 1 + h \left( \frac{\partial f}{\partial y} \right) \right] > 1 \), that is, for \( \frac{\partial f}{\partial y} \) positive, the error at \( x_i \) will be amplified in traversing the \( i \)th step, and the solution will tend toward instability. Even in these cases, however, it may be possible to keep the propagation error under control, especially during the early course of the integration, by choosing a sufficiently small \( h \), that is, by keeping the propagation factor \( \left[ 1 + h \left( \frac{\partial f}{\partial y} \right) \right] \) close to 1.

Suppose that \( \frac{\partial f}{\partial y} \) is positive and constant, so that the propagation factor is greater than one for all \( h \) and the error does increase without bound for increasing \( x \) as shown by equation (6). For example, consider the following equation

\[
\frac{dy}{dx} = 2y,
\]

for which the solution is

\[
y(x) = y(x_0) e^{2(x-x_0)}.
\]

Will the unbounded growth of the error invalidate the computed solution? Not necessarily, since the solution itself is unbounded for increasing \( x \). The most important criterion is not that the absolute error \( \epsilon_i \) be bounded, but that the relative error \( \frac{\epsilon_i}{y_i} \) not grow appreciably.

Similarly, though more complicated, propagation factors can be developed for higher-order one-step methods (Hildebrand (1956)). The quantity \( h(\frac{\partial f}{\partial y}) \), sometimes called the step factor, contributes to these propagation factors in a manner comparable to that for Euler's method. The solution of a set of \( n \) simultaneous first-order ordinary differential equations is, at least in principle, no more difficult than the solution of a single first-order equation. The algorithm selected
is applied to each of the \( n \) equations in parallel at each step.

Error analyses are virtually impossible to implement for the higher-order Runge-Kutta schemes for systems of differential equations. The step-size control mechanisms and stability considerations outlined above carry over to the multiple-equation case without appreciable change. In practice, one often solves the equations using different step sizes and observes the behavior of the solutions with regard to apparent convergence and stability.

Stability analysis of the multistep methods is somewhat more complicated than for the one-step methods discussed above. Local truncation error is an important consideration in stability analysis, but will not be presented here. Carnahan, Luther, and Wilkes (1969), page 386, presents an excellent step-by-step procedure for analyzing truncation error.

Step size control is another important consideration for analyzing multistep methods since their principal advantage, namely that they require fewer derivative evaluations per step than do the one-step methods of comparable accuracy, will be lost if the step size is chosen to be smaller than necessary. The step size must be small enough to satisfy the convergence criterion for the corrector equation, preferably small enough to insure convergence in just one or two iterations, and it must be small enough to satisfy restrictions on the magnitude of the local truncation error. On the other hand, the step size should preferably be large enough so that round-off errors and the number of derivative evaluations will be minimized. The latter consideration is especially important when the derivative function is complicated, and each evaluation requires substantial computing time. Fortunately, the truncation error analysis yields enough information to determine when the step size should be increased or decreased. Unfortunately, the mechanism for implementing such changes is not very straightforward. The difficulty is that when the step size is changed, the necessary starting values for the predictor and corrector will not usually be available. One approach is to use a one-step method to compute these starting values. In practice, the step size is usually changed by doubling or halving it. Clearly, too-frequent changes in the step size will vitiate the principal advantage of the multistep methods — their computational speed.
In the early 1950's, when computers were first widely used for solving ordinary differential equations, investigators discovered that for some equations certain of the multistep methods led to computational errors far larger than expected from local truncation errors. It was also discovered that a decrease in step size often resulted in an increase in the observed error, even when round-off errors were known to be insignificant. In some cases, the numerical solution showed little, if any, relationship to the true solution of the equation being solved.

Subsequent analysis has shown that, under certain conditions, some of the multistep methods exhibit catastrophic instabilities which render the numerical solution meaningless. Such instabilities develop even though the equations are inherently stable and cannot in general be removed by step size adjustment (hence the instability cannot be a numerical instability of the type discussed for the one-step methods).

Virtually all stability analyses reported in the literature have been for the simple linear ordinary differential equation

$$\frac{dy}{dx} = f(x, y) = ay,$$  \hspace{1cm} (8)

where $a$ is a constant. Often the analyses begin with an arbitrary equation, but assumptions to retain only linear terms and to require that $\frac{\partial f}{\partial y}$ be constant, soon follow and the equation actually being studied is equation (8).

The basic procedure followed for any of the multistep methods is about the same. In each case, the appropriate homogenous linear difference equation with constant coefficients is found for equation (8). The characteristic equation is then solved for the roots $\gamma_1, \gamma_2, \ldots, \gamma_n$, where $n$, the order of the equation, is equal to the difference of the largest and smallest subscript which appears in the multistep equation. For example, for the Milne corrector, the largest subscript is $i+1$ and the smallest is $i-1$; the order of the difference equation and thus the number of solutions is $(i+1) - (i-1)$ or 2.
Let $\gamma_1$ be the root of the characteristic equation which is associated with the true solution of equation (8); then $\gamma_1$ is called the principal zero or root. If

$$|\gamma_j| < |\gamma_1|, \quad j = 2, 3, \ldots, n,$$

the method is called relatively stable. In this case, the approximation to the true solution will dominate the parasitic solutions (see Carnahan, Luther, and Wilkes (1969) page 389) associated with the roots $\gamma_j, \quad j = 2, 3, \ldots, n$. If the true solution of equation (8) decays with increasing $i$, then the parasitic solutions must decay even more rapidly. If the true solution of equation (8) grows with increasing $i$, then the parasitic solutions must either decay or grow less rapidly than the true solution.

A method is called strongly stable if

$$|\gamma_j| < 1, \quad j = 2, 3, \ldots, n.$$

Thus, for a method to be strongly stable, all parasitic solutions must decay with increasing $i$.

Dahlquist (1956) has shown that if the truncation error for a multistep method is of $O(h^{p+1})$, and $n$ is the order of the difference equation, then to achieve either strong or marginal stability, $p$ must satisfy the relationship

$$p \leq n + 2.$$

Moreover, $p = n + 2$ is possibly only when $n$ is even and the method is marginally stable (for example, Milne's fourth-order corrector for which $n = 2$ and $p = 4$). Thus, one can increase the accuracy of a method (increase the order of the truncation error) only by increasing the value of $n$, and consequently, the number of
parasitic solutions associated with the difference equation. Thus, diminished stability is the price one must pay for increased accuracy in a multistep method.

Simultaneous ordinary differential equations may be solved by multistep methods as well as by one-step methods. The appropriate algorithm is implemented for each equation in parallel at each step. The criteria for solution of simultaneous corrector equations is by the usual method of successive substitutions.

All the preceding stability analyses have dealt with the solution of the very simple linear equation (8). As has been stated before, the question of applicability of these conclusions to more general nonlinear equations remains unanswered. Henrici (1962) believes that the variability of $\partial f/\partial x$ may make a significant difference in the stability of a method, for example, changing a strongly stable method into a marginally stable one for certain equations. Computational experiments have shown, however, that conclusions about stability, following from a stability analysis for equation (8), correlate rather well with observed behavior of the multistep methods when used for other equations as well.
REFERENCES


