EQUIVALENT DIFFERENTIAL EQUATIONS FOR NONLINEAR DYNAMICAL SYSTEMS

Thesis by

Edward John Patula

In Partial Fulfillment of the Requirements

For the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1970

(Submitted May 8, 1970)

ACKNOWLEDGEMENTS

The author wishes to express his sincere appreciation to Professor W.D. Iwan, whose guidance and encouragement made this investigation possible. The assistance of the faculty in the Applied Mechanics Department, who have been so helpful in so many ways, is gratefully acknowledged.

The author wishes to thank Mrs. Odessa Walker, Miss Julie Wright, and Mrs. Phyllis Henderson for the very competent typing of the manuscript, and also Miss Cecilia Lin for her assistance.

The author would also like to express his appreciation to the California Institute of Technology for the financial assistance received during the course of this study.

The author wishes to thank his wife, Paulette, for her endless patience and encouragement throughout the years of graduate study.

ABSTRACT

A technique for obtaining approximate periodic solutions to nonlinear ordinary differential equations is investigated. The approach is based on defining an equivalent differential equation whose exact periodic solution is known. Emphasis is placed on the mathematical justification of the approach. The relationship between the differential equation error and the solution error is investigated, and, under certain conditions, bounds are obtained on the latter. The technique employed is to consider the equation governing the exact solution error as a two point boundary value problem. Among other things, the analysis indicates that if an exact periodic solution to the original system exists, it is always possible to bound the error by selecting an appropriate equivalent system.

Three equivalence criteria for minimizing the differential equation error are compared, namely, minimum mean square error, minimum mean absolute value error, and minimum maximum absolute value error. The problem is analyzed by way of example, and it is concluded that, on the average, the minimum mean square error is the most appropriate criterion to use.

A comparison is made between the use of linear and cubic auxiliary systems for obtaining approximate solutions. In the examples considered, the cubic system provides noticeable improvement over the linear system in describing periodic response.

A comparison of the present approach to some of the more classical techniques is included. It is shown that certain of the standard approaches where a solution form is assumed can yield erroneous qualitative results.

TABLE OF CONTENTS

I. INTRODUCTION

The area of nonlinear ordinary differential equations has been investigated by mathematicians for centuries. However, it has only been in recent times that the engineer and the applied theoretician have developed an interest in this area. One major reason for this interest is that it is not always possible to neglect nonlinearities in many of today's complex problems. In situations where a more detailed understanding of the qualitative and quantitative behavior of systems is desired, it is often necessary to include nonlinear effects. Although nonlinear equations have occupied the mathematician for quite some time, the techniques available for obtaining exact closed form solutions are rather limited. Under suitable conditions, existence and uniqueness of solutions can be proved, but for only a relatively small number of nonlinear equations are the exact solutions known. A good treatment of the existence and uniqueness issue is given in reference (1).

The inability to obtain exact solutions has necessitated the development of approximate analysis for studying nonlinear problems. This analysis may loosely be divided into two categories: topological methods and approximate solution methods. The former usually involve phase plane or functional analysis methods. The Poincaré theory[2] for the singular points of two-dimensional autonomous systems and the Second Method of Liapunov [3] for the stability of nonlinear systems are examples of topological methods. Approximate solution

methods, on the other hand, usually involve obtaining closed form approximate solutions for the nonlinear system of interest. Typical examples of this category are the Poincaré-Linstead perturbation technique[4] and the asymptotic methods of Krylov, Bogoliubov, and Mitropolsky[5].

With the increased speed and flexibility of today's digital computers, one might be tempted to say that the usefulness and importance of approximate analysis have virtually been eliminated. However, a strong case can still be made for approximate analysis. It is true that, if accurate quantitative results are desired for specific cases, the computer is the tool to be utilized. However, if the general behavior of the solution is of interest, the computer can become cumbersome and quite expensive. It is difficult to determine trends and the dependence of the solution on differential equation parameters using a digital computer. It is usually necessary to calculate a large number of cases, and, even then, it may be difficult to determine whether or not some phenomenon or characteristic is being overlooked or concealed. Therefore, if it is possible to perform a meaningful approximate analysis, the behavior of the solution is usually more easily determined. Once the general nature of the solution is known, the computer may then be used to obtain accurate quantitative results if desired. Therefore, the importance of approximate analysis has not been diminished by the digital computer. On the contrary, it has been increased since a good understanding of the basic phenomenon is always desirable prior to utilizing the computer.

The present investigation involves an approximate technique which may be classified as an approximate solution method. The approach, called the equivalent equation approach, was presented in several recent papers by Iwan [6, 7] and is designed to provide approximate periodic solutions to nonlinear systems. The technique is a generalization of the method of equivalent linearization and is based on defining an alternative or auxiliary differential system whose exact periodic solution is known.

Most standard approximating techniques involve assuming a certain solution form containing unspecified parameters. These parameters may be prescribed by minimizing, in some sense, the error residual obtained by substituting the assumed solution into the original differential system. Since periodic motions are of interest, the usual solution form involves linear combinations of trigonometric functions. Typical methods which fall into this class are: The Poincare'-Linstead perturbation technique [4], Krylov-Bogoliubov-Mitropolsky asymptotic methods [5], Galerkin's technique [9], and the methods of collocation, subdomain, and least squares [10]. One major limitation on most of the above techniques is that the usual first order approximation they provide is accurate only for equations which are nearly linear. Increased accuracy is possible by including more terms in the approximation, but the additional computational effort required rapidly becomes excessive. Therefore, there exists a need to develop techniques which yield more accurate results for equations containing

moderately large nonlinearities and which, at the same time, involve levels of computational effort comparable to the standard first order techniques.

Various other authors have utilized non-trigonometric solution forms in order to achieve more accuracy. Eringen postulated a generalized Galerkin's procedure utilizing non-trigonometric solutions[11]. Klotter and Cobb used a polynomial approximation to represent the quarter period wave form[12]. The parameters were determined utilizing Galerkin's procedure. Recently, Barkham and Soudack used solution forms which involve Jacobian elliptic functions[13, 14]. Their technique utilized the method of slowly varying parameters and enabled transient behavior to be analyzed. However, they make several simplifying assumptions which detract from the rigor of the approach. Furthermore, their results apply only to second order equations which are "Duffing-like". The equivalent equation approach is like the above in that it represents an attempt to provide an unambiguous technique for systematically treating equations containing moderately large nonlinearities.

The idea of using one differential system to model another differential system is not new. The method of equivalent linearization has been an accepted approximate technique for quite some time. Various authors have suggested modifications but these mainly concern the manner in which the linear system is made equivalent to the original system. The standard method[8] minimizes the mean square differential equation error. Denman and Liu have suggested using an ultraspherical

polynomial approximation where the nonlinearity is expanded in a series of ultraspherical polynomials[15, 16]. Only the linear term is utilized, and, therefore, an equivalent linear system is generated.

An example of using a nonlinear auxiliary system was given by Helfenstein[17]. He utilized Duffing's equation with Jacobian cosine excitation to model Duffing's equation with trigonometric cosine excitation. However, the equivalent system is obtained by merely equating the coefficients of all like terms. This idea of using one nonlinear system to model another nonlinear system was the motivating factor in the development of the equivalent equation approach. By using nonlinear auxiliary systems, it is felt that better approximations could be obtained since some of the features peculiar to nonlinear problems would be incorporated into the analysis in a very natural manner.

A complete description of the equivalent equation approach is given in Chapter II. As stated earlier, the approach is based on defining equivalent differential equations whose exact periodic solutions are known. By developing an alternative differential equation which is, in some sense, equivalent to the original system of interest, it is hoped that the corresponding periodic solution will also be equivalent to the exact solution of the original system. To illustrate the technique, Section 2.2 presents an example where the equivalent equation approach is used to obtain an approximate periodic solution to the undamped Duffing's equation with trigonometric excitation. The auxiliary system utilized is Duffing's equation with Jacobian cosine excitation.

In Chapter III, the relationship between the differential equation error (the difference between the original system and the equivalent system) and the solution error (the difference between the exact periodic solution and the solution of the equivalent system) is investigated. Under certain conditions, bounds are obtained on the solution error in terms of the differential equation error. The technique employed is to consider the equation governing the exact solution error as a two point boundary value problem. Rewriting the problem as an integral equation and using the Green's function, the method of successive approximations is applied to obtain a bound on the exact error. Among other things, the above analysis indicates that if an exact periodic solution to the original system exists, it is always possible to bound the error by selecting an appropriate equivalent system. Other authors have obtained results similar to those presented in Chapter III. Some are Cesari[18, 19], Urabe[20, 21, 22], McLaughlin[23, 24], Holtzman[25], and Lazer[26]. Unfortunately, most of the above consider either weakly nonlinear equations or specific auxiliary systems. The present analysis attempts to be more general by considering two arbitrary differential systems. That research which seems to be most closely related to the present work is discussed in Section 3.1.

Section 3.1 presents an error bound analysis for first order n-dimensional vector systems. A discussion of autonomous systems is included which shows that the successive approximations techniques never apply to non-trivial periodic solutions of autonomous systems.

In Section 3.2, the results are specialized for the case of second order scalar equations. Furthermore, the Green's function used in Section 3.1 is associated with the unique linear part of the error differential equation. In general, this equation contains periodic coefficients which makes the determination of the Green's function very tedious. To avoid this difficulty the problem is reformulated utilizing the Green's function for a system with constant coefficients. These coefficients are then selected by minimizing the resulting error bound.

Section 3.3 presents an example where the theory of Section 3.2 is applied to the example considered in Section 2.2, i.e., the trigonometrically excited undamped Duffing's equation. Bounds are obtained for both the linear and cubic approximations.

An example of a conservative autonomous system is considered in Section 3.4. As mentioned above, the theory of Section 3.2 yields very little information concerning autonomous systems. Consequently, an alternative comparison technique is developed for second order scalar conservative autonomous systems. The autonomous example considered is the undamped Duffing's equation.

The manner in which an alternative system is made equivalent to the original system is examined in Chapter IV. Various equivalence criteria are compared; namely, minimization of the mean square differential equation error, minimization of the mean absolute value differential equation error, and minimization of the maximum absolute value differential equation error. The errors are minimized with

respect to parameters appearing in the auxiliary system. The above errors were selected because of their physical relevance and their relation to the error bound analysis performed in Chapter III.

Since it was of interest to determine which equivalence criterion yielded the smallest actual solution error on the average, it was impossible to use analytical techniques to investigate the problem. Therefore, the problem is analyzed by way of example, and it is concluded that, on the average, the minimum mean square error is the most appropriate equivalence criterion to use.

In Chapter V, a comparison is made between the linear and the cubic approximations for second order scalar systems. The general approximations are developed, and the determining equations are presented. The two approximations are compared by way of example. The specific examples considered are Duffing's equation with trigonometric excitation, and a system of the form

$$\ddot{x} + \frac{\gamma x}{1+\alpha|x|} = F\cos(\omega t) \ . \qquad (1.1)$$

In both cases, the cubic approximation provides considerable improvement in solution accuracy over the linear system. In addition, in the second example there exists some indication that the cubic approximation, which is, primarily, a harmonic approximation, is trying to yield some information about the ultraharmonic behavior of (1.1).

In Chapter VI, a brief comparison is presented between the equivalent equation approach and various other classical approximate

techniques. The techniques considered are collocation, subdomain, least squares, and Galerkin's procedure. The relation of these techniques to the general method of weighted residuals[27] is included.

In Section 6.2, some peculiarities associated with the method of least squares are presented. Examples are considered which illustrate that the method of least squares may generate extraneous approximate solutions. These solutions correspond to maximums of the mean square differential equation error instead of minimums. Similar difficulties may arise with other techniques, such as the equivalent equation approach, which are based on an averaging principle. However, if the differential equation parameters appear linearly in the differential equation error, which is often the case, the equivalent equation approach always generates solutions which minimize the differential equation error.

II. EQUIVALENT EQUATION APPROACH

In this chapter, the Equivalent Equation Approach is presented. The description closely parallels that given by Iwan [6]. An example of the use of the technique to obtain an approximate periodic solution for the undamped Duffing's equation with trigonometric excitation is included.

2.1. Description of the Technique.

Consider the problem of obtaining an approximate solution for the periodic motions of a system of ordinary differential equations. The system of interest, called the original system, will be represented as

$$\underline{D}(\underline{x}(t), t) = \underline{0} \quad , \tag{2.1}$$

where $\underline{D}$ is a vector system which may contain differential operators operating on the dependent vector $\underline{x}$ and functions of the independent variable t. Furthermore, (2.1) is assumed to possess periodic solutions with least period T_s. For nonautonomous systems, T_s may be prescribed by the excitation, but in the case of autonomous systems, T_s may be an unknown of the problem.

In order to obtain an approximate solution of (2.1), consider another system of equations, the auxiliary system, represented as

$$\underline{D}^*(\underline{x}(t), \alpha_1, \alpha_2, \ldots, \alpha_r, t) = \underline{0} \quad , \tag{2.2}$$

where $\alpha_i (i=1,\ldots, r)$ are parameters of the equations. Let (2.2) have

known periodic solution forms that are members of some class of functions C having the form

$$\underline{x}(t) = \underline{y}(\beta_1, \ldots, \beta_s, t) \quad , \tag{2.3}$$

where $\beta_j (j=1, \ldots, s)$ are parameters which define the members of C. If the solution of (2.2) is unique, there will exist s relations between α_i and β_j which come directly from (2.2) plus any periodicity and/or initial conditions that may apply. Knowledge of the $\alpha_i (i=1, \ldots, r)$ implies a unique determination of the $\beta_j (j=1, \ldots, s)$, but the converse is not necessarily true.

It is possible to obtain an approximate solution of (2.1) by using the solution of an auxiliary system (2.2) where the auxiliary system is chosen to be as "close" to the original system (2.1) as possible. By close, it is meant that the equations comprising the original system and the auxiliary system are very similar in form. It is then hoped that, by making the difference in the governing equations small, the difference in their respective solutions will also be small. In Chapter III, the nature of the relation between the difference in the governing equations and the difference in their solutions is investigated, and the following statement is proved.

"Under certain conditions, given any bound on the difference between the solutions of the two differential systems (2.1) and (2.2), it is always possible to select an auxiliary system (2.2) such that the magnitude of the actual difference between the solutions is less than the prescribed bound."

Motivated by the above argument, one may select certain of the parameters $\alpha_i (i=1,\ldots, r)$ so as to make some of the terms in $\underline{D}^*$ identical in form to terms in $\underline{D}$. Let

$$\underline{D}^*(\underline{x}(t), \alpha_1, \ldots, \alpha_r, t) = \underline{D}_1^*(\underline{x}(t), \alpha_1, \ldots, \alpha_p, t) + \underline{D}_2^*(\underline{x}(t), \alpha_{p+1}, \ldots, \alpha_r, t) \quad , \quad (2.4)$$

and

$$\underline{D}(\underline{x}(t), t) = \underline{D}_1(\underline{x}(t), t) + \underline{D}_2(\underline{x}(t), t) \quad , \quad (2.5)$$

where $\alpha_i (i=1,\ldots, p)$ are selected so that

$$\underline{D}_1^*(\underline{x}(t), \alpha_1, \ldots, \alpha_p, t) \equiv \underline{D}_1(\underline{x}(t), t) \quad . \quad (2.6)$$

The additional $\alpha_i (i=p+1,\ldots, r)$ parameters are determined in some manner so as to minimize the remaining difference between $\underline{D}^*$ and $\underline{D}$ for all members of the class C; i.e. for all $\underline{x}(t)$ having the form $\underline{x}(t) = \underline{y}(t)$ where $\underline{y}(t)$ is notation for $\underline{y}(\beta_1, \ldots, \beta_s, t)$.

Define the differential equation error $\underline{\varepsilon}(t)$ as the difference obtained between $\underline{D}(\underline{x}(t), t)$ and $\underline{D}^*(\underline{x}(t), t)$ when both are evaluated at the solution form $\underline{y}(t)$. Then,

$$\underline{\varepsilon}(t) \equiv \underline{D}(\underline{y}(t), t) - \underline{D}^*(\underline{y}(t), t) \quad . \quad (2.7)$$

In (2.7), $\underline{D}^*(\underline{y}(t), t)$ does not vanish since the s relations between the $\alpha_i (i=1,\ldots, r)$ and the $\beta_j (j=1,\ldots, s)$ have not been utilized. Using (2.4), (2.5), and (2.6), equation (2.7) can be written as

$$\underline{\varepsilon}(\alpha_{p+1}, \ldots, \alpha_r, t) = \underline{D}_2(\underline{y}(t), t) - \underline{D}_2^*(\underline{y}(t), \alpha_{p+1}, \ldots, \alpha_r, t) \quad , \quad (2.8)$$

where the explicit dependence of $\underline{\varepsilon}$ on the $\alpha_i (i=p+1,\ldots, r)$ is indicated. The parameters $\beta_j (j=1,\ldots, s)$ are to be considered arbitrary in (2.8).

The $\alpha_i (i=p+1,\ldots, r)$ are now selected so that (2.8) is minimized. However, there are many ways that (2.8) could be minimized depending on the specific type of minimization desired. For example, the maximum norm of $\underline{\varepsilon}$ could be minimized, or the mean value of the norm of $\underline{\varepsilon}$ over one cycle of the motion could be minimized, or $\underline{\varepsilon}$ could be made to vanish at certain preselected points. In Chapter IV, approximations obtained by using these various minimization conditions are compared. As expected, no one minimization condition gives the "best" approximation in all cases. However, it may be concluded, in a very broad sense, that the optimum minimization criterion is

$$\frac{1}{T_s} \int_{t_1}^{t_1+T_s} \underline{\varepsilon}^T \underline{\varepsilon} dt = \text{minimum} \quad , \tag{2.9}$$

where T denotes the transpose, and t_1 is an arbitrary time. Since the minimization of the integral is with respect to $\alpha_i (i=p+1,\ldots, r)$, (2.9) is equivalent to

$$\frac{\partial}{\partial \alpha_i}\left[\frac{1}{T_s} \int_{t_1}^{t_1+T_s} \underline{\varepsilon}^T \underline{\varepsilon} dt \right] = 0 \quad , \quad i=p+1,\ldots, r \ . \tag{2.10}$$

Since T_s does not depend on $\alpha_i (i=p+1,\ldots, r)$ (remembering that the $\beta_j (j=1,\ldots, s)$ are still arbitrary), (2.10) becomes

$$\int_{t_1}^{t_1+T_s} \left[\frac{\partial \underline{\varepsilon}^T}{\partial \alpha_i} \underline{\varepsilon} + \underline{\varepsilon}^T \frac{\partial \underline{\varepsilon}}{\partial \alpha_i} \right] dt = 0 \quad , \quad i=p+1,\ldots, r \ . \tag{2.11}$$

Because the s relations between the $\alpha_i (i=1,\ldots, r)$ and the $\beta_j (j=1,\ldots, s)$ have not yet been utilized, equations (2.6) and (2.11) provide relations

which determine the auxiliary system parameters $\alpha_i (i=1,\ldots,r)$ which are valid for all members of C. If these s relations are now introduced, equation (2.11) can be further reduced. Since the differentiation in equation (2.11) is with respect to explicit $\alpha_i (i=p+1,\ldots,r)$, the derivatives of $\underline{\varepsilon}^T$ may be expressed, using equation (2.8), in terms of $\underline{D}_2^{*T}(\underline{y}(t), \alpha_{p+1},\ldots,\alpha_r, t)$ only. Furthermore, $\underline{\varepsilon}$ may be expressed in terms of $\underline{D}(\underline{y}(t))$ only from equation (2.7). $\underline{D}^*(\underline{y}(t),t)$ vanishes once the s relations are utilized. From the above considerations, equation (2.11) becomes

$$\int_{t_1}^{t_1+T_s} \frac{\partial}{\partial \alpha_i}\left[\underline{D}_2^{*T}(\underline{y}(t), \alpha_{p+1},\ldots,\alpha_r, t)\right]\underline{D}(\underline{y}(t), t)dt = 0\ ,\quad i=p+1,\ldots,r. \tag{2.12}$$

It is assumed that the relations in (2.12) will be of such a form that it is possible to determine the auxiliary system parameters $\alpha_i (i=p+1,\ldots,r)$ and $\beta_j (j=1,\ldots,s)$ so that meaningful approximate solutions and equivalent system equations are obtained. The values of the parameters generated by (2.12) correspond to extremums of the mean square error (2.9). These extremums may be either maximums or minimums. Care must be taken to select only those values of $\alpha_i (i=p+1,\ldots,r)$ and $\beta_j (j=1,\ldots,s)$ which minimize (2.9). If the weight functions

$$W_i(t) = \frac{\partial}{\partial \alpha_i}\left[\underline{D}_2^{*T}(\underline{y}(t), \alpha_{p+1},\ldots,\alpha_r)\right],\quad i=p+1,\ldots,r\ , \tag{2.13}$$

are independent of $\alpha_i (i=p+1,\ldots,r)$, it can be shown that all solutions generated by (2.12) correspond to minimums of (2.9). This situation occurs if $\underline{\varepsilon}$ is linear in the $\alpha_i (i=p+1,\ldots,r)$, which is often the case. However, if the $W_i(t)$ depend on some of the $\alpha_i (i=p+1,\ldots,r)$, various anomalies may arise. For example, 1) no approximate solution may

be generated, even though an exact solution exists, or 2) some extraneous approximate solutions could be introduced, or 3) a combination of 1) and 2) above could occur. This particular point is investigated in more detail in Chapter VI, where the equivalent equation approach and the method of least squares are compared. It suffices at this point to assume that $\frac{\partial}{\partial \alpha_i}\left[\underline{D}_2^{*T}(\underline{y}(t), \alpha_{p+1}, \ldots, \alpha_r, t)\right]$ $(i=p+1, \ldots, r)$ are of a form that provide meaningful approximate solutions.

Let q be the number of independent equations generated by the minimization condition (2.12). Then, if q=r-p, the equations from (2.12) plus the s relations from the auxiliary system (2.2) combined with the p preselected parameters $\alpha_i (i=1, \ldots, p)$ satisfying (2.6) will determine all of the parameters $\alpha_i (i=p+1, \ldots, r)$ and $\beta_j (j=1, \ldots, s)$. One obtains not only an approximate solution, but also an "equivalent" auxiliary system. This additional information may be quite useful. For example, it might be of interest to know the equivalent mass or the equivalent excitation level or the equivalent stiffness of some original system, and the equivalent system approach would provide an auxiliary system whose behavior, presumably would be better understood. If $q < r-p$, it means that there are not enough independent relations to determine all of the parameters, and that an additional r-p-q relations have to be supplied. There are several ways in which these additional relations may be obtained. One approach might be to simply prescribe an additional r-p-q parameters in the auxiliary system. However, depending on the specific parameters being prescribed, fewer relations might be obtained from the minimization condition (2.12), and, consequently, more auxiliary system parameters

would have to be specified until enough independent equations were obtained to determine all of the $\alpha_i (i=1,\ldots, r)$ and $\beta_j (j=1,\ldots, s)$. An alternative approach to prescribing additional auxiliary system parameters is to generate an additional r-p-q equations from the q equations resulting from (2.12). Consider an alternative form of equations (2.12),

$$\int_{t_1}^{t_1+T_s} \frac{\partial}{\partial \alpha_i}\left[\underline{D}_2^{*T}(\underline{y}(t), \alpha_{p+1},\ldots, \alpha_r)\right]\Big(\underline{D}_2(\underline{y}(t), t) - \underline{D}_2^{*}(\underline{y}(t), \alpha_{p+1},\ldots, \alpha_r)\Big)dt = 0 \quad , \quad i=p+1,\ldots, \quad , \qquad (2.14)$$

which is obtained by using equations (2.4), (2.5), (2.6) and the s relations for the auxiliary system. In general, any differential system $\underline{D}$ contains terms which can be put into the following categories: 1) terms containing only the highest order derivative of the vector function $\underline{x}, (\underline{D})_A$; 2) terms containing only lower order derivatives of $\underline{x}, (\underline{D})_B$; and 3) terms which depend only on the independent variable $t, (\underline{D})_C$. If $\underline{D}_2$ and $\underline{D}_2^*$ are both separated into the above terms, equations (2.14) become

$$\int_{t_1}^{t_1+T_s} \frac{\partial}{\partial \alpha_i}\left[\underline{D}_2^{*T}(\underline{y}, \alpha_{p+1},\ldots, \alpha_r)\right]\Big[(\underline{D}_2)_A - (\underline{D}_2^*)_A + (\underline{D}_2)_B - (\underline{D}_2^*)_B + (\underline{D}_2)_C - (\underline{D}_2^*)_C\Big]dt = 0 \quad , \quad i=p+1,\ldots, r, \qquad (2.15)$$

where the functional dependence of $\underline{D}_2$ and $\underline{D}_2^*$ has been dropped for brevity. It is now possible to generate more equations by requiring that the individual terms in some of the equations vanish instead of the

entire combination. Utilizing this approach, (2.15) may be decomposed to give

$$\left.\begin{aligned}
&\int_{t_1}^{t_1+T_s} \frac{\partial}{\partial \alpha_i}\left[\underline{D}_2^{*T}(\underline{y}, \alpha_{p+1}, \ldots, \alpha_r)\right]\left[(\underline{D}_2)_A - (\underline{D}_2^*)_A\right] dt = 0 \,, \\
&\int_{t_1}^{t_1+T_s} \frac{\partial}{\partial \alpha_i}\left[\underline{D}_2^{*T}(\underline{y}, \alpha_{p+1}, \ldots, \alpha_r)\right]\left[(\underline{D}_2)_B - (\underline{D}_2^*)_B\right] dt = 0 \,, \\
&\int_{t_1}^{t_1+T_s} \frac{\partial}{\partial \alpha_i}\left[\underline{D}_2^{*T}(\underline{y}, \alpha_{p+1}, \ldots, \alpha_r)\right]\left[(\underline{D}_2)_C - (\underline{D}_2^*)_C\right] dt = 0 \,,
\end{aligned}\right\} \qquad (2.16)$$

where i can take on as many values in the set $p+1, \ldots, r$ as are needed to generate enough equations so that all of the auxiliary system parameters $\alpha_i (i=1, \ldots, r)$ and $\beta_j (j=1, \ldots, s)$ can be determined. Depending on the specific $\underline{D}$ and $\underline{D}^*$ chosen, certain of the equations in (2.14) will lend themselves more naturally to the type of separation given in (2.16). Equations (2.16) have a physical interpretation as well. By separating terms, one is attempting to represent certain types of terms in the original system by the same type of terms in the auxiliary system; that is, one is asking that the highest derivative terms in $\underline{D}_2^*$ model the highest derivative terms in $\underline{D}_2$, and that the terms depending only on t in $\underline{D}_2^*$ model the terms depending only on t in $\underline{D}_2$, etc. If the separated equations (2.16) still do not provide sufficient equations to determine the auxiliary system parameters, it is possible to further divide the terms in $\underline{D}$ into more categories. In this way, it is always possible to generate a sufficient number of equations to determine all of the α_i and β_j.

The problem of determining the solution parameters $\beta_j (j=1,\ldots, s)$ is simplified considerably if $\underline{D}_2^*$ is a linear function of $\alpha_i (i=p+1,\ldots, r)$ and $q=s$. In this case, $\beta_j (j=1,\ldots, s)$ may be determined directly from equations (2.12) without using the differential equations (2.2) or first determining the parameters $\alpha_i (i=p+1,\ldots, r)$. This would certainly be the case when $\underline{D}^*$ is a set of linear differential equations with constant coefficients $\alpha_i (i=1,\ldots, r)$ and the class C contains the least number of functions necessary to describe the periodic solution. The above formulation then becomes a generalization of the method of equivalent linearization[8]. However, the equivalent equation approach is not restricted to using only linear auxiliary systems. Indeed, one of the more important aspects of the equivalent equation approach is that it allows for the possibility of using one nonlinear system to model another nonlinear system.

The similarity in the form of equations (2.12) to those obtained by application of Galerkin's method is apparent[9]. In fact, the two approaches can give identical approximations depending on the set of trial functions used in Galerkin's method. This point is considered in more detail in Chapter VI, where the two techniques are compared. In general, however, the results of the equivalent equation approach will differ from those of Galerkin's method.

As noted earlier, the present approach is essentially that of defining an equivalent system for the set $\underline{D}$. As such, the $\beta_j (j=1,\ldots, s)$ remain arbitrary in equation (2.8), and $\underline{\varepsilon}$ has the significance of a difference term. The functional relationship between $\alpha_i (i=1,\ldots, r)$ and

$\beta_j (j=1,\ldots, s)$ is not introduced until (2.12). However, it is clear from (2.8) that $\underline{\varepsilon}$ could also have been thought of as the error residual of the set of equations $\underline{D}(\underline{y}(t), t)$ if the relations between $\alpha_i (i=1,\ldots, r)$ and $\beta_j (j=1,\ldots, s)$ had been used at that earlier stage of development. In this way, $\underline{\varepsilon}$ would no longer have been an explicit function of $\alpha_i (i=p+1,\ldots, r)$, and, consequently, the minimization specified by (2.9) would have been made with respect to the solution parameters $\beta (j=1,\ldots, s)$. This is the so called method of least squares [10]. Although the two approaches appear to be very similar, they can lead to quite different results even for the same class of approximating functions $\underline{y}$. In fact, the present approach will usually result in a cleaner mathematical formulation since the $\alpha_i (i=1,\ldots, r)$ normally appear quite simply in the well behaved auxiliary equations whereas the $\beta_j (j=1,\ldots, s)$ frequently appear in a complicated manner in a nonlinear $\underline{D}(\underline{y}(t), t)$. This complicated nature leads to some fundamental difficulties with the method of least squares related to generating meaningless approximate solutions as described previously. This difficulty is investigated in more detail in Chapter VI where the method of least squares and the equivalent equation approach are compared.

2.2. Example.

In the previous section, the equivalent equation approach is developed and discussed. In this section, the use of the technique is illustrated by way of an example.

Consider the problem of finding an approximate periodic solution to the undamped Duffing's equation with trigonometric excitation.

In this case, the original system may be written in the form

$$\underline{D}(\underline{x}(t), t) \equiv \ddot{x} + ax + bx^3 - B\cos(\omega t) = 0 \quad , \tag{2.17}$$

where dots denote differentiation with respect to t, and a, b, B, and ω are constants. As an auxiliary system, choose

$$\underline{D}^*(\underline{x}, \alpha_1, \alpha_2, \alpha_3, t) = \ddot{x} + \alpha_1 x + \alpha_2 x^3 - \alpha_3 \, cn(\eta t, k) = 0 \quad , \tag{2.18}$$

where α_1, α_2, α_3, η, and k are constants and cn (u, k) is the Jacobian elliptic cosine function with modulus k. Since the forced response of system (2.17) is of interest, the response will possess the same period as the excitation. Consequently, η is selected so that the periods of the excitations in (2.17) and (2.18) are the same. Therefore,

$$\eta = \frac{2K(k)\omega}{\pi} \quad , \tag{2.19}$$

where K(k) is the complete elliptic integral of the first kind. In an attempt to make the original system and the auxiliary system similar in form, prescribe α_1 and α_2 such that

$$\alpha_1 = a \quad \text{and} \quad \alpha_2 = b \quad . \tag{2.20}$$

Then, the auxiliary system becomes

$$\underline{D}^*(\underline{x}, \alpha, t) = \ddot{x} + ax + bx^3 - \alpha \, cn(\eta t, k) = 0 \quad , \tag{2.21}$$

where the subscript on α_3 has been dropped for convenience.

The exact steady-state solution of (2.21) is of the form

$$y = \beta \, cn(\eta t, k) \quad , \tag{2.22}$$

where the frequency is the same as that of the excitation. Satisfaction of the differential equation (2.21) requires

$$b\beta^3 + (1-\eta^2)\beta - \alpha = 0$$
$$k^2 = \frac{b\beta^2}{2\eta^2} \,. \tag{2.23}$$

Referring to equation (2.8), the difference term ϵ is

$$\epsilon(t, \alpha) = B\cos(\omega t) - \alpha\, cn(\eta t, k) \,. \tag{2.24}$$

Hence, minimization of ϵ is with respect to α. Applying condition (2.12) gives

$$\int_0^{T_s/4} cn(\eta t, k)\Big[\beta\left(a-\eta^2(1-2k^2)\right) cn(\eta t, k) + (b\beta^3 - 2\eta^2 k^2\beta)\, cn^3(\eta t, k) - B\cos(\omega t)\Big]dt = 0 \,, \tag{2.25}$$

where t_1 was set to zero, and the symmetry of the integrand was used to replace T_s by $T_s/4$. The integrals involving the cn functions are available[28]. The integral involving the cn and the cos functions may be evaluated by first expanding the cn function in a Fourier series[36] and then using the orthogonality of the trigonometric functions to show that only one term in the expansion makes any contribution. When these results are substituted into (2.25), the relation becomes, for $b > 0$,

$$\frac{\beta}{\eta k^2}(a-\eta^2+b\beta^2)\left(E(k) - (1-k^2)K(k)\right) - \frac{B\pi^2}{4\omega k K(k)}\,\mathrm{sech}\left(\frac{\pi K(k')}{2K(k)}\right) = 0 \,, \tag{2.26}$$

where E(k) is the complete elliptic integral of the second kind and $k' = (1-k^2)^{1/2}$ is the complimentary modulus. Equations (2.19) and (2.23) may be used to eliminate the dependence of η and ω giving

$$\left(\frac{\beta^3 b}{k}\left(1 - \frac{1}{2k^2}\right) + \frac{\beta a}{k}\right)\left(E(k) - k'^2 K(k)\right) - \frac{B\pi}{2}\operatorname{sech}\left(\frac{\pi K(k')}{2K(k)}\right) = 0 \quad . \tag{2.27}$$

When b is negative, k is pure imaginary, and the reduction of (2.25) gives

$$\left(\beta^3 b\left(1 + \frac{1}{2r^2}\right) + \beta a\right)\left(K(k_1) - E(k_1)\right) + \frac{B\pi k_1}{2}\operatorname{csch}\left(\frac{\pi K(k_1')}{2K(k_1)}\right) = 0 \quad , \tag{2.28}$$

where $k_1 = r(1+r^2)^{1/2}$ and $k = ir$. The most efficient procedure for obtaining a frequency-response curve is to first assume a value of k (or k_1, if b is negative), then obtain β from equation (2.27) (or (2.28) if b is negative), use equation (2.23) to determine η, and then finally use equation (2.19) to calculate ω. If b is negative, equation (2.19) and (2.23) become

$$\eta = \frac{2k_1 K(k_1)\omega}{\pi} \tag{2.19}'$$

and

$$b\beta^3 + (1-\eta^2)\beta - \alpha = 0$$
$$r^2 = -\frac{b\beta^2}{2\eta^2} \quad , \tag{2.23}'$$

where k_1 and r are defined in (2.28).

An indication of the accuracy of the cubic approximation may be obtained by considering some specific examples. Figure 1 shows the steady-state response amplitude β as a function of excitation frequency ω for $B = 0.1$, $b = 0.1$, and $a = 1.0$. Also shown is the approximation obtained using equivalent linearization. Several exact solution points were obtained using direct numerical integration of (2.17), and these are also included in the figure. It will be noted that the cubic approximation shows considerable improvement over the usual first order approximation particularly for frequencies significantly different from one. This is not surprising since most of the standard solution techniques require b to be a small number, B to be of order b, and $1-\omega^2$ also to be of order b. On the other hand, the accuracy of the present approach is primarily a function of the magnitude of B and only indirectly a function of b and ω. When B approaches zero, the cubic approximation gives an exact solution of the original system while the equivalent linearization solution obviously does not. The cubic approximation gives similar improvements when b is negative[6]. Recently, Iwan[7] studied Duffing's equation with linear viscous damping and showed that an equivalent cubic approximation describes the steady-state behavior much more accurately than the usual approximation obtained using equivalent linearization.

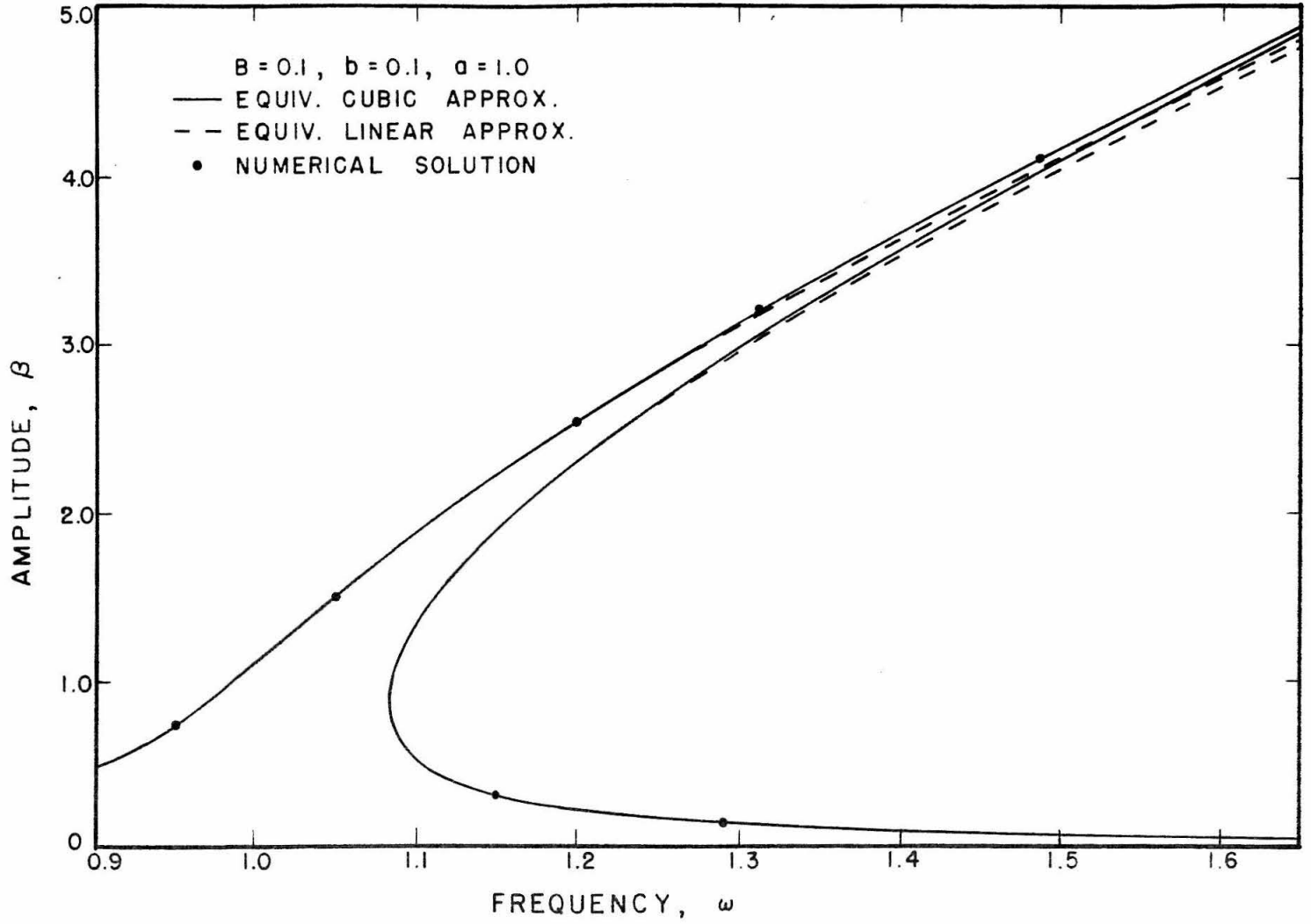

Figure 1: Amplitude versus Frequency for Duffing's Equation

III. ERROR BOUND ANALYSIS

In this chapter, the relationship between the difference of two systems of differential equations and the corresponding difference in their respective solutions is examined. The first section deals with two first order n-dimensional vector systems, and a bound is obtained on the norm of the difference between the two solutions. The second section treats a special subclass of vector systems, namely second order scalar equations, where sharper estimates can be made and better results can be obtained. Sections three and four are devoted to examples which illustrate the use of the theory to obtain bounds for nonautonomous systems (Section 3.3) and conservative autonomous system (Section 3.4).

3.1. Error Bounds for General Vector Systems.

Before entering into the details of formulating the problem and deriving bounds, it is convenient to introduce some notation which will prove useful throughout the analysis.

Notation

The norm of a vector $\underline{x}$, denoted by $\|\underline{x}\|$, is a scalar function that provides a measure of the magnitude of $\underline{x}$. A valid norm is any scalar function of $\underline{x}$ satisfying the following conditions:

$$\left.\begin{aligned} &\text{i)}\ \|\underline{x}\| \geq 0 \quad , \quad \text{for all } \underline{x} \in E^n \ , \\ &\text{ii)}\ \|\underline{x}\| = 0 \quad , \quad \text{iff } \underline{x} \equiv 0 \quad , \end{aligned}\right\} \qquad (3.1)$$

iii) $\|\underline{x}+\underline{y}\| \leq \|\underline{x}\| + \|\underline{y}\|$, for all $\underline{x}, \underline{y} \in E^n$,

iv) $\|c\underline{x}\| = |c| \, \|\underline{x}\|$, for any real scalar c and all $\underline{x} \in E^n$, (3.1) cont.

where E^n denotes n-dimensional Euclidean space. Corresponding to each valid vector norm is an associated matrix norm. The norm of a matrix A, denoted by $\|A\|$, is a scalar function that indicates the magnitude of A. A valid vector along with an associated matrix norm satisfy the following relations [29]:

i) $\|A+B\| \leq \|A\| + \|B\|$, for all nxn matricies A and B,

ii) $\|AB\| \leq \|A\| \, \|B\|$, for all nxn matricies A and B,

iii) $\|cA\| = |c| \, \|A\|$, for any real scalar c and any matrix A,

iv) $\|A\underline{x}\| \leq \|A\| \, \|\underline{x}\|$, for any nxn matrix A and any vector $\underline{x} \in E^n$, (3.2)

It should be noted that the appropriate matrix norm associated with any specific vector norm is not necessarily unique. For any valid vector norm $\|\underline{x}\|$, an associated matrix norm may always be defined by [29]

$$\|A\| = \max_{\|\underline{x}\| = 1} \|A\underline{x}\|, \tag{3.3}$$

for all vectors $\underline{x} \in E^n$ satisfying $\|\underline{x}\| = 1$. Examples of valid vector norms and associated matrix norms are:

i) $\|\underline{x}\| = \sum_{i=1}^{n} |x_i|$ and $\|A\| = \sum_{i,j=1}^{n} |a_{ij}|$ (Taxicab norm);

ii) $\|\underline{x}\| = \left(\sum_{i=1}^{n} x_i^2 \right)^{1/2}$ and $\|A\| = \left(\sum_{ij=1}^{n} |a_{ij}|^2 \right)^{1/2}$ (Euclidean norm);

iii) $\|\underline{x}\| = \max_i |x_i|$ and $\|A\| = \max_i \left(\sum_{j=1}^{n} |a_{ij}| \right)$ (Maximum modulus norm).

The following analysis is done using general vector and matrix norms, the only requirement being that they satisfy (3.1) and (3.2).

Formulation

Let R be a domain in E^n and L be the real line. Consider the problem of finding an approximate periodic solution for the system

$$\frac{d\underline{x}}{d\tau} = \underline{F}(\underline{x}, \tau) \quad , \tag{3.4}$$

where $\underline{x}$ and $\underline{F}$ are n-dimensional vectors, $\underline{F}$ is periodic with period T_0 for fixed $\underline{x}$, and $\underline{F}(\underline{x}, \tau)$ is C^1 for $\underline{x} \in R$ and C^0 for $\tau \in L$. (Actually, this condition can be relaxed to $\underline{F}$ being C^0 for $\underline{x} \in R$ and $\tau \in L$ and $\underline{F}$ satisfying a modified Lipschitz condition in $\underline{x}$. This point is discussed in the section on generalizations following the basic analysis.) The period T_0 need not be the least period of $\underline{F}(\underline{x}, \tau)$. In general, (3.4) possesses a periodic solution with period T_0, but it may also possess other periodic solutions having periods different from T_0. Assume that the periodic solution of (3.4) with period T_s is of interest. For $\underline{x}(\tau)$ to actually exist, T_s cannot be completely independent of T_0. Since $\underline{x}(\tau)$ has period T_s, $\frac{d\underline{x}(\tau)}{d\tau}$ also has period T_s, and, consequently, $\underline{F}(\underline{x}, \tau)$ is periodic with period T_s. Therefore, $\underline{F}\left(\underline{x}(\tau), \tau\right) = \underline{F}\left(\underline{x}(\tau + T_s), \tau + T_s\right) = \underline{F}\left(\underline{x}(\tau), \tau + T_s\right)$; i.e. $\underline{F}\left(\underline{x}(\tau), \tau\right)$ is periodic in τ for

fixed $\underline{x}$ with period T_s. From above, it must happen that $T_s = n T_0$ for $n=1, 2, \ldots$. If $n=1$, the solution is called the harmonic solution; the solution having the same period as the least period of the excitation. In $n \geq 2$, the solution is called a subharmonic solution of order n. From the periodicity requirement on $\underline{F}(\underline{x}, \tau)$, it is clear that there exist no solutions with periods satisfying $T_0 = j T_s$ for $j=2, 3, \ldots$. Therefore, no periodic solution exists having a period less than the least period of the differential equation. If $\underline{F}(\underline{x}, \tau)$ is an autonomous system, it possesses all periods in τ, and consequently it may possess solutions with all periods. For example, if there exists a constant vector $\underline{c}$ such that $\underline{F}(\underline{c}) = \underline{0}$, then $\underline{x}(\tau) = \underline{c}$ is a solution possessing all periods.

It is possible to obtain an approximate periodic solution by considering an auxiliary system which is represented as

$$\frac{d\underline{y}}{d\tau} = \underline{G}(\underline{y}, \alpha_1, \ldots, \alpha_r, \tau) \quad , \tag{3.5}$$

where $\alpha_i (i=1, \ldots, r)$ are parameters of the system, $\underline{y}$ and $\underline{G}$ are n-dimensional vectors, and $\underline{G}$ is periodic in τ and is C^1 for $\underline{y} \in R$ and C^0 for $\tau \in L$. It is assumed that (3.5) has known exact periodic solutions $\underline{y}(\beta_1, \ldots, \beta_s, \tau)$, where $\beta_j (j=1, \ldots, s)$ are solution parameters. Since $\underline{y}$ will ultimately be the approximate solution, the period of $\underline{y}$ is required to be the same as the period of the desired solution $\underline{x}$. Therefore, $\underline{y}(\tau + T_s) = \underline{y}(\tau)$. Depending on the system (3.5) chosen and on the particular exact solution $\underline{y}$ that is known, the period of $\underline{G}$ will be determined once the period of the solution is specified. Therefore, $\underline{G}$ will be

periodic in τ with period T_A, such that $T_s = mT_A$ for m = 1, or 2, ... Again T_A need not be the least period of $\underline{G}$. The equivalent equation approach may now be used to determine all of the $\alpha_i (i=1, \ldots, r)$ and $\beta_j (j=1, \ldots, s)$.

Having found an approximate solution, it is desirable to obtain some indication of its accuracy. For practical purposes, any comparison between the exact solution and the approximate solution must involve only those quantities which are accessible; specifically the differential equations, periodicity conditions, and the approximate solution. Thus, one is motivated to obtain differences in solutions by considering differences in the corresponding differential equations. To this end, normalize the independent variable τ in the following manner. Since $\underline{F}(\underline{x}, \tau), \underline{G}(\underline{y}, \tau), \underline{x}(\tau)$ and $\underline{y}(\tau)$ are all periodic with period T_s, let

$$\tau = t T_s \quad . \tag{3.6}$$

Then all derivatives with respect to τ can be written in terms of t as

$$\frac{d}{d\tau}(\cdot) = \frac{1}{T_s} \frac{d}{dt}(\cdot) \quad . \tag{3.7}$$

Using (3.7), equation (3.4) becomes

$$\frac{d\underline{x}}{dt} = \underline{f}(\underline{x}; t) \quad , \tag{3.8}$$

where $\underline{f}(\underline{x}, t) = T_s \underline{F}(\underline{x}, T_s t)$, and $\underline{f}(\underline{x}, t)$ is periodic in t with period 1. Equation (3.5) becomes

$$\frac{d\underline{y}}{dt} = \underline{g}(\underline{y}, \alpha_1, \ldots, \alpha_r, t) \quad , \tag{3.9}$$

where $\underline{g} = T_s \underline{G}$, and $\underline{g}$ is periodic in t with period 1. It is clear that $\underline{f}$ and $\underline{g}$ possess the same continuity and differentiability properties as do $\underline{F}$ and $\underline{G}$ respectively. The approximate solution $\underline{y}$ is now periodic in t with period 1.

Denoting the difference between $\underline{x}(t)$ and $\underline{y}(t)$ by $\underline{z}(t)$, then

$$\underline{z}(t) = \underline{x}(t) - \underline{y}(t) \quad . \tag{3.10}$$

Differentiating (3.10) and using equations (3.8), (3.9) and (3.10), the equation for the exact solution error $\underline{z}(t)$ is

$$\frac{d\underline{z}}{dt} = \underline{f}(\underline{z}+\underline{y}, t) - \underline{g}(\underline{y}, t) \quad . \tag{3.11}$$

In general, it will not be possible to solve (3.11) exactly. Therefore, it seems reasonable to try to obtain a bound on $\| \underline{z} \|$ in terms of known quantities, specifically the differential equation error $\underline{e}(t)$ as defined by (2.9). Since $\underline{x}(t)$ and $\underline{y}(t)$ are both periodic with period 1, $\underline{z}(t)$ is also periodic with period 1. Therefore, the problem for $\underline{z}$ can be recast in terms of a two point boundary value problem over the interval $0 \leq t \leq 1$ with mixed boundary conditions

$$\underline{z}(0) = \underline{z}(1) \quad . \tag{3.12}$$

The problem for $\underline{z}(t)$ then consists of equation (3.11) subject to the boundary conditions (3.12).

In order to proceed further, it is convenient to reformulate the problem for $\underline{z}$ in terms of an integral representation. Consider the following homogenous problem

$$\frac{d\underline{\varphi}}{dt} = A(t)\underline{\varphi}$$
$$\underline{\varphi}(0) = \underline{\varphi}(1) \quad , \tag{3.13}$$

where $\underline{\varphi}$ is a n-dimensional vector and A(t) is a nxn coefficient matrix which is C^0 in t and is chosen such that the only solution of (3.13) is the trivial solution $\underline{\varphi} \equiv \underline{0}$. A(t) could be the Jacobian matrix $\frac{\partial \underline{f}(\underline{y}, t)}{\partial \underline{x}}$ evaluated at the approximate solution $\underline{y}$ if it possessed the above property of having only the trivial solution. Related to (3.13) is an Associated Matrix Equation

$$\frac{dZ}{dt} = A(t)Z \quad , \tag{3.14}$$

where Z is a nxn matrix. Let Z(t) be the principal matrix solution of (3.14) satisfying Z(0) = I, the identity matrix. System (3.13) will have only the trivial solution if and only if the matrix Q,

$$Q = Z(0) - Z(1) = I - Z(1) \quad , \tag{3.15}$$

is non-singular. Since (3.13) has only the trivial solution, it posseses a Green's function G(t, s) defined as

$$G(t, s) = \begin{cases} Z(t)Q^{-1}Z(1)Z^{-1}(s) & , \text{ for } t < s \\ Z(t)Q^{-1}Z^{-1}(s) & , \text{ for } t \geq s, \end{cases} \tag{3.16}$$

where Q^{-1} denotes the inverse of Q, and $Z^{-1}(s)$ is the inverse of Z(s), which is non-singular since it is a principal matrix solution. G(t, s) is a matrix which is continuously differentiable except at the point t=s.

Consider now the following inhomogenous problem

$$\frac{d\underline{\varphi}}{dt} = A(t)\underline{\varphi} + \underline{\theta}(t)$$
$$\underline{\varphi}(0) = \underline{\varphi}(1) \ , \tag{3.17}$$

where A(t) is the same matrix as in (3.13) and $\underline{\theta}(t)$ is a continuous vector function. Using the Green's function defined in (3.16), the unique solution of (3.17) can be written as

$$\underline{\varphi}(t) = \int_0^1 G(t, s)\underline{\theta}(s)\, ds \quad . \tag{3.18}$$

The reader interested in proofs of the above statements concerning the existence of the Green's function and the validity of the representation (3.18) is referred to reference (30).

The problem for $\underline{z}(t)$ can now be written as an integral equation. Equation (3.11) may be written as

$$\frac{d\underline{z}}{dt} = A(t)\underline{z} + \underline{\varepsilon}(t) + \underline{f}^*(\underline{z}, t) \quad , \tag{3.19}$$

where A(t) is the same matrix as in (3.13), $\underline{\varepsilon}(t)$ is the differential equation error

$$\underline{\varepsilon}(t) = \underline{f}(\underline{y}, t) - \underline{g}(\underline{y}, t) \ , \tag{3.20}$$

and

$$\underline{f}^*(\underline{z}, t) = \underline{f}(\underline{z}+\underline{y}, t) - \underline{f}(\underline{y}, t) - A(t)\underline{z} \ . \tag{3.21}$$

Equation (3.19) is still subject to the boundary conditions (3.12). As mentioned earlier, A(t) could be the Jacobian matrix $\frac{\partial \underline{f}(\underline{y}, t)}{\partial \underline{x}}$ if it

possessed a Green's function . It would be desirable to use this matrix for A(t) since all linear terms in $\underline{z}$ would then be eliminated from $\underline{f}^*(\underline{z}, t)$, and, consequently, $\|\underline{f}^*(\underline{z}, t)\|$ would be $o(\|\underline{z}\|)$ as $\|\underline{z}\| \to 0$. Using the Green's function (3.16) and applying (3.18), where $\underline{\epsilon}(t)$ and $\underline{f}^*(\underline{z}, t)$ are considered inhomogenous terms, the integral equation determining $\underline{z}$ is

$$\underline{z}(t) = \int_0^1 G(t, s)\Big(\underline{\epsilon}(s) + \underline{f}^*(\underline{z}(s), s)\Big) ds \quad . \tag{3.22}$$

It is possible to prove the existence of a solution to (3.22) using the method of successive approximations and, as a consequence, a bound on $\|\underline{z}\|$ is obtained. These results are presented in the form of a Lemma.

Lemma 1

Consider the integral equation (3.22). If the following conditions are satisfied:

i) $\underline{\epsilon}(s)$ is a continuous vector function for $s \in [0, 1]$.

ii) $\underline{f}^*(\underline{z}(s), s)$ is a continuous function of s for $s \in [0, 1]$ and of $\underline{z}$ for $\underline{z}$ such that $\|\underline{z}\| \leq \delta$ and satisfies a modified Lipschitz condition of the following form,

$$\|\underline{f}^*(\underline{z}_1(s), s) - \underline{f}^*(\underline{z}_2(s), s)\| \leq k(s) \|\underline{z}_1(s) - \underline{z}_2(s)\| \tag{3.23}$$

for all $\underline{z}_1$ and $\underline{z}_2$ such that $\|\underline{z}_1\| \leq \delta$ and $\|\underline{z}_2\| \leq \delta$ where k(s) is a positive continuous function of s for $s \in [0, 1]$.

iii) The kernel G(t, s) in (3.22) can be bounded by

$$\|G(t, s)\| \leq p(t)\, q(s) \tag{3.24}$$

for all $t \in [0, 1]$ and $s \in [0, 1]$, where $p(t)$ is a bounded non-negative integrable function for $t \in [0, 1]$, and $q(s)$ is a positive continuous function for $s \in [0, 1]$.

iv) $$K \equiv \int_0^1 k(s)\, p(s)\, q(s)\, ds < 1 \quad . \tag{3.25}$$

v) $$p(t) \Big|_{\max} \frac{E}{1 - K} \leq \delta \quad , \tag{3.26}$$

where δ is defined in ii) and

$$E = \int_0^1 q(s) \| \underline{\varepsilon}(s) \| \, ds \quad .$$

Then (3.22) possesses an exact unique solution $\underline{z}(t)$ and

$$\| \underline{z}(t) \| \leq p(t)(1-K)^{-1} \int_0^1 q(s) \| \underline{\varepsilon}(s) \| \, ds \quad . \tag{3.27}$$

Proof: Existence of a solution is shown using the method of successive approximations. Consider an iteration scheme

$$\underline{z}_0 = \int_0^1 G(t, s)\, \underline{\varepsilon}(s)\, ds$$

$$\underline{z}_n = \underline{z}_0 + \int_0^1 G(t, s) \underline{f}^*(\underline{z}_{n-1}(s), s)\, ds \quad . \tag{3.28}$$

It is first necessary to show that every iterate satisfies

$$\| \underline{z}_n(t) \| \leq \delta \quad , \qquad n = 0, 1, \ldots \quad . \tag{3.29}$$

Taking norms of the first of (3.28) and using (3.24), the initial guess

satisfies

$$\|\underline{z}_0(t)\| \leq p(t)\, E \quad , \tag{3.30}$$

where E is defined in (3.26). But since $K < 1$,

$$\|\underline{z}_0(t)\| \leq p(t) \max (1-K)^{-1} E \quad .$$

(3.26) implies that $\|\underline{z}_0(t)\| \leq \delta$. To prove the general case, it is only necessary to show that

$$\|\underline{z}_n\| \leq p(t) E \sum_{i=0}^{n} K^i \quad , \tag{3.31}$$

where K is given in (3.25). Then since $K < 1$, the properties of geometric series may be utilized to show that $\sum_{i=0}^{n} K^i < (1-K)^{-1}$. Consequently, $\|\underline{z}_n\| \leq \delta$. To prove (3.31) use induction. Taking norms of the second relation in (3.28), $\|\underline{z}_n\|$ satisfies

$$\|\underline{z}_n\| \leq \|\underline{z}_0\| + p(t) \int_0^1 q(s) \|\underline{f}^*(\underline{z}_{n-1}, s)\| ds \quad .$$

Since $\underline{z}_{n-1}$ is assumed to satisfy (3.31), it is permissible to use the Lipschitz condition (3.23). This, combined with (3.30) and (3.31), gives

$$\|\underline{z}_n\| \leq p(t) \int_0^1 q(s) \|\underline{\varepsilon}(s)\| ds + p(t) \int_0^1 q(s) k(s) p(s) *$$

$$\int_0^1 q(r) \|\underline{\varepsilon}(r)\| dr \sum_{i=0}^{n-1} K^i \, ds \quad .$$

Combining terms and noting the definition of K and E, $\|\underline{z}_n\|$ is bounded

as

$$\|\underline{z}_n\| \leq p(t)\, E \sum_{i=0}^{n} K^i \quad , \tag{3.32}$$

which is the desired result. Therefore, (3.32) shows that the use of the Lipschitz condition (3.23) is valid for any pair of iterates. Before proving that the sequence of functions $\{\underline{z}_n(t)\}$ is uniformly convergent for $t \in [0, 1]$, it is necessary to determine bounds on the difference between successive iterates. Consider the difference

$$\underline{z}_1 - \underline{z}_0 = \int_0^1 G(t, s)\, \underline{f}^*(\underline{z}_0, s)\, ds \quad .$$

Taking norms and using (3.23), (3.24), and (3.32) for $n = 0$, the above relation becomes

$$\|\underline{z}_1 - \underline{z}_0\| \leq p(t) \int_0^1 q(s)k(s)p(s)ds \int_0^1 q(s)\|\underline{\varepsilon}(s)\|ds$$

or

$$\|\underline{z}_1 - \underline{z}_0\| \leq p(t)\, KE \quad . \tag{3.33}$$

Use induction to show that

$$\|\underline{z}_n - \underline{z}_{n-1}\| \leq p(t)\, EK^n \quad . \tag{3.34}$$

Assume that (3.34) is valid for n-1 and consider the norm of $\underline{z}_n - \underline{z}_{n-1}$. From (3.28), the difference may be written as

$$\|\underline{z}_n - \underline{z}_{n-1}\| = \| \int_0^1 G(t, s) \left(\underline{f}^*(\underline{z}_{n-1}, s) - \underline{f}^*(\underline{z}_{n-2}, s) \right) ds \| \quad .$$

Taking norms under the integral and using the Lipschitz condition and the bound on G(t, s), the above equation becomes

$$\|\underline{z}_n - \underline{z}_{n-1}\| \leq p(t) \int_0^1 q(s)k(s)\|\underline{z}_{n-1} - \underline{z}_{n-2}\| ds \quad .$$

However, by the inductive hypothesis,

$$\|\underline{z}_{n-1} - \underline{z}_{n-2}\| \leq p(s) K^{n-1} E \quad .$$

Therefore, using the definition of K, the bound on successive iterates is

$$\|\underline{z}_n - \underline{z}_{n-1}\| \leq p(t) K^n E \quad .$$

Returning to the task of showing that $\{z_n\}$ is uniformly convergent, consider $\underline{z}_m - \underline{z}_n$ for two integers m and n such that $m > n$. Writing this difference as a collapsing sum,

$$\underline{z}_m - \underline{z}_n = \sum_{j=n+1}^{m} (\underline{z}_j - \underline{z}_{j-1}) \quad .$$

Taking norms and using the triangle inequality,

$$\|\underline{z}_m - \underline{z}_n\| \leq \sum_{j=n+1}^{m} \|\underline{z}_j - \underline{z}_{j-1}\| \quad .$$

Using (3.34), the above relation becomes

$$\|\underline{z}_m - \underline{z}_n\| \leq p(t) E \sum_{j=n+1}^{m} K^j \quad .$$

But by assumption p(t) is bounded for $t \in [0, 1]$, and $K < 1$ which implies that the sequence $\{\underline{z}_n(t)\}$ of continuous functions converges uniformly to a continuous function $\underline{z}(t)$ for $t \in [0, 1]$ by Cauchy's criterion. $\underline{z}(t)$ satisfies (3.22), since the limit as $n \to \infty$ can be taken in (3.28). A bound is obtained on $\|\underline{z}(t)\|$ by taking the limit as $n \to \infty$ in (3.32). Consequently,

$$\|\underline{z}(t)\| \leq p(t)\,(1-K)^{-1}\,E \quad .$$

Showing that the solution $\underline{z}(t)$ is unique is relatively straightforward. Assume there exists two solutions $\underline{z}_1(t)$ and $\underline{z}_2(t)$ satisfying (3.22), and consider their difference

$$\|\underline{z}_2 - \underline{z}_1\| \leq \int_0^1 \|G(t, s)\|\,\|\underline{f}^*(\underline{z}_2(s), s) - \underline{f}^*(\underline{z}_1(s), s)\|\,ds \quad .$$

Since the limit function $\underline{z}$ must also satisfy $\|\underline{z}\| < \delta$, the Lipschitz condition (3.23) and (3.24) may be used to obtain

$$\|\underline{z}_2 - \underline{z}_1\| \leq p(t) \int_0^1 q(s)k(s)\|\underline{z}_2 - \underline{z}_1\|\,ds \quad .$$

Multiplying by q(t), k(t), and integrating, this relation becomes

$$\int_0^1 q(t)k(t)\|\underline{z}_2 - \underline{z}_1\|dt \leq K \int_0^1 q(s)k(s)\|\underline{z}_2 - \underline{z}_1\|\,ds \quad .$$

Since $K < 1$, and the integral is non-negative, the only possibility is

$$\int_0^1 q(t)k(t)\|\underline{z}_2 - \underline{z}_1\|\,dt \equiv 0 \quad .$$

Since the integrand is continuous and non-negative, it must vanish everywhere. But by assumption q(t) and k(t) are positive for $t \in [0, 1]$, therefore

$$\| \underline{z}_2 - \underline{z}_1 \| = 0 \quad ,$$

which implies $\underline{z}_2 \equiv \underline{z}_1$ for $t \in [0, 1]$. Q.E.D.

In Lemma 1, the integral equation (3.22) is considered as a separate entity. However, in the present analysis, (3.22) is related to the differential system (3.19). Therefore, the following theorem applies to system (3.19).

Theorem 1

Let R be a domain in E^n, and let L be the real line. Consider the original system

$$\begin{aligned} \frac{d\underline{x}}{dt} &= \underline{f}(\underline{x}, t) \\ \underline{x}(0) &= \underline{x}(1) \quad , \end{aligned} \tag{3.35}$$

where $\underline{f}(\underline{x}, t)$ is C^1 for $\underline{x} \in R$ and C^0 for $t \in L$, and $\underline{f}(\underline{x}, t) = \underline{f}(\underline{x}, t+1)$ for fixed $\underline{x}$. Consider also the auxiliary system

$$\begin{aligned} \frac{d\underline{y}}{dt} &= \underline{g}(\underline{y}, t) \\ \underline{y}(0) &= \underline{y}(1) \quad , \end{aligned} \tag{3.36}$$

where $\underline{g}(\underline{y}, t)$ is C^1 for $\underline{y} \in R$ and C^0 for $t \in L$, and $\underline{g}$ is periodic in t with period 1 for fixed $\underline{y}$. If the difference $\underline{z}$ is formed

$$\underline{z} = \underline{x} - \underline{y} \quad , \tag{3.37}$$

then the differential system governing $\underline{z}$ is

$$\frac{d\underline{z}}{dt} = A(t)\underline{z} + \underline{\epsilon}(t) + \underline{f}^*(\underline{z}, t)$$
$$\underline{z}(0) = \underline{z}(1) \quad , \tag{3.38}$$

where $\underline{f}^*(\underline{z}, t) = \underline{f}(\underline{z}+\underline{y}, t) - \underline{f}(\underline{y}, t) - A(t)\underline{z}$, $\underline{\epsilon}(t) = \underline{f}(\underline{y}, t) - \underline{g}(\underline{y}, t)$, and $A(t)$ is a continuous matrix function of t. If the homogenous problem

$$\frac{d\underline{\varphi}}{dt} = A(t)\underline{\varphi}$$
$$\underline{\varphi}(0) = \underline{\varphi}(1) \tag{3.39}$$

has only the trivial solution, it possesses a Green's function defined by (3.16). Furthermore, if

$$K = \int_0^1 k(s)q(s)p(s)ds \quad < 1 \quad , \tag{3.40}$$

where $q(s) = \max\left(\|Q^{-1}Z(1)Z^{-1}(s)\|, \|Q^{-1}Z^{-1}(s)\|\right)$, $p(s) = \|Z(s)\|$ (Q and Z are defined in (3.14) and (3.15)), and $k(s)$ is a Lipschitz constant for $\underline{f}^*(\underline{z}, t)$, and if

$$p(t)\Big|_{\max} \quad (1-K)^{-1} E \leq \delta \quad , \tag{3.41}$$

where $E = \int_0^1 q(s)\|\underline{\epsilon}(s)\| ds$ and δ defines a region $\|\underline{z}\| \leq \delta$ for which $\underline{f}^*(\underline{z}, t)$ is Lipschitzian, then the following conclusions may be reached. The original system possesses an exact unique periodic solution for $\underline{x} \in R$. Furthermore, the norm of the error $\underline{z}(t)$ can be bounded as

$$\|z(t)\| \leq p(t)(1-K)^{-1} E \quad . \tag{3.42}$$

If, in addition, $\underline{f}^*(\underline{z}, t)$ satisfies

$$\max_{\|\underline{z}\| \le \delta} \left\| \frac{\partial \underline{f}^*}{\partial \underline{z}}(\underline{z}, t) \right\| \le k(t) \tag{3.43}$$

the exact unique solution $\underline{x}(t)$ is an isolated solution. (An isolated periodic solution $\underline{x}$ is one such that the equation of first variation associated with it possesses no non-trivial solution with the same period as $\underline{x}$.)

Proof: Differentiating (3.37) and using (3.35), (3.36), and (3.37), the governing equation for $\underline{z}(t)$ is found to be

$$\frac{d\underline{z}}{dt} = \underline{f}(\underline{z}+\underline{y}, t) - \underline{g}(\underline{y}, t)$$

$$\underline{z}(0) = \underline{z}(1) \quad .$$

Adding and subtracting $A(t)\underline{z}+\underline{f}(\underline{y}, t)$ to the right hand side, (3.38) is obtained. However, by assumption, the homogenous problem (3.39) possesses a Green's function. Therefore, by (3.18), the error $\underline{z}$ satisfies

$$\underline{z}(t) = \int_0^1 G(t, s)\Big(\underline{\epsilon}(s) + \underline{f}^*(\underline{z}(s), s)\Big) ds \quad , \tag{3.44}$$

where $G(t, s)$ is

$$G(t, s) = \begin{cases} Z(t) Q^{-1} Z(1) Z^{-1}(s) & , \text{ for } t < s \\ Z(t) Q^{-1} Z^{-1}(s) & , \text{ for } t \ge s \quad . \end{cases}$$

$Z(t)$ is the principal matrix solution of (3.39), and Q is a non-singular matrix given by $Q \equiv I - Z(1)$.

Define

$$p(t) = \| Z(t) \|, \; q(s) = \max\Big(\| Q^{-1} Z(1) Z^{-1}(s) \|, \; \| Q^{-1} Z(s) \| \Big) .$$

Since $\underline{f}(\underline{x}, t), \underline{g}(\underline{y}, t)$, and $A(t)$ are continuous, $\underline{\epsilon}(t)$ and $\underline{f}^*(\underline{z}, t)$ in (3.38)

will also be continuous. Since $\underline{f}(\underline{x}, t)$ is continuously differentiable with respect to $\underline{x}$, $\underline{f}^*(\underline{z}, t)$ will also be continuously differentiable in $\underline{z}$. Therefore, for all $\|\underline{z}\| \leq \delta$, $\underline{f}^*(\underline{z}, t)$ will satisfy a modified Lipschitz condition of the form

$$\|\underline{f}^*(\underline{z}_2, t) - \underline{f}^*(\underline{z}_1, t)\| \leq k(t)\|\underline{z}_2 - \underline{z}_1\| \quad ,$$

where k(t) is continuous and positive for $t \in [0, 1]$. Using assumptions (3.40) and (3.41), it is clear that all of the hypotheses of Lemma 1 are satisfied, and consequently, (3.44) has an exact unique solution $\underline{z}(t)$ with

$$\|\underline{z}(t)\| \leq p(t)(1-K)^{-1} E \quad .$$

Since (3.44) has an exact unique solution, (3.35) must also possess an exact unique solution. Since the solution of the original system is $\underline{x} = \underline{y} + \underline{z}$ where $\underline{y}$ is a known prescribed function, $\underline{z}(t)$ existing and being unique implies that $\underline{x}(t)$ exists and is unique.

To show that $\underline{x}$ is an isolated solution, if (3.43) is satisfied, is straightforward. Consider the equation of first variation of (3.35),

$$\frac{d\underline{\xi}}{dt} = \frac{\partial \underline{f}(\underline{x}, t)}{\partial \underline{x}} \underline{\xi} \quad . \tag{3.45}$$

Adding and subtracting $A(t)\underline{\xi}$ to the right hand side, the representation (3.18) can be used to obtain

$$\underline{\xi}(t) = \int_0^1 G(t, s)\left[\frac{\partial \underline{f}}{\partial \underline{x}}(\underline{x}(s), s) - A(s)\right]\underline{\xi}(s)ds \quad , \tag{3.46}$$

where G(t, s) is the Green's function for the homogenous problem (3.39).

From (3.37) and (3.38), $\underline{f}(\underline{x}, t)$ may be written as

$$\underline{f}(\underline{x}, t) = \underline{f}^{*}(\underline{x}-\underline{y}, t) + \underline{f}(\underline{y}, t) + A(t)(\underline{x}+\underline{y}) \quad .$$

Noting that $\underline{f}^{*}(\underline{z}, t)$ is continuously differentiable, the Jacobian matrix may be formed yielding

$$\frac{\partial \underline{f}(\underline{x}, t)}{\partial \underline{x}} = \frac{\partial \underline{f}^{*}}{\partial \underline{z}}(\underline{x}-\underline{y}, t) + A(t) \quad .$$

Taking norms of (3.46) and using the above relation, $\underline{\xi}(t)$ satisfies

$$\| \underline{\xi}(t) \| \leq p(t) \int_0^1 q(s)k(s) \| \underline{\xi}(s) \| ds \quad ,$$

where it is assumed that $\|\underline{x}-\underline{y}\| \leq \delta$ so that the use of (3.43) is justified. Multiplying by q(s) and k(s) and integrating, the above relation becomes,

$$(1-K) \int_0^1 q(s)k(s) \| \underline{\xi}(s) \| ds \leq 0 \quad ,$$

where the definition of K has been used. Since $K < 1$ and the integral is non-negative, the only possibility is that the integral must vanish. Since the integrand is assumed continuous for $t \in [0, 1]$, and q(s) and k(s) are positive, the above relation implies that

$$\| \underline{\xi}(s) \| \equiv 0 \quad .$$

Therefore, the exact periodic solution $\underline{x}$ of the original system (3.35) is isolated. Q.E.D.

It is also possible to prove the following.

Theorem 2

Assume that the original system possesses an exact isolated

periodic solution $\underline{x}$ with period 1. Then for any small number δ, it is always possible to choose an auxiliary system (3.36) with an exact periodic solution $\underline{y}$ such that the differential equation error is sufficiently small so that the norm of the difference between $\underline{x}$ and $\underline{y}$ satisfies $\|\underline{x}-\underline{y}\| \leq \delta$.

Proof: Consider the integral equation

$$\underline{z}(t) = \int_0^1 G(t, s)\Big(\underline{\varepsilon}(s) + \underline{f}^*(\underline{z}(s), s)\Big)\,ds \quad ,$$

where $\underline{\varepsilon}(s) = \mu\, \underline{g}(\underline{x}(s), s)$ and $\underline{g}$ is continuously differentiable in $\underline{x}$ for $s \in [0, 1]$ and

$$\underline{f}^*(\underline{z}(s), s) = \underline{f}(\underline{x}, s) - \underline{f}(\underline{x}-\underline{z}, s) - \frac{\partial \underline{f}}{\partial \underline{x}}(\underline{x}, s)\, z - \mu\Big(\underline{g}(\underline{x}, s) - \underline{g}(\underline{x}-\underline{z}, s)\Big) \quad .$$

$G(t, s)$ is the Green's function for the system (3.39) with $A(t) = \frac{\partial \underline{f}}{\partial \underline{x}}(\underline{x}, t)$ which exists since the equation of first variation possesses no nontrivial solution with period 1, i.e. $\underline{x}$ is isolated. For $\underline{z}_1$, $\underline{z}_2$, and μ satisfying $\|\underline{z}_1\| \leq \eta \leq \delta$, $\|\underline{z}_2\| \leq \eta \leq \delta$, and $|\mu| \leq \mu_1$ where η and μ_1 are sufficiently small positive numbers, $\underline{f}^*(\underline{z}, s)$ satisfies a modified Lipschitz condition,

$$\|\underline{f}^*(\underline{z}_1, s) - \underline{f}^*(\underline{z}_2, s)\| \leq k(t)\|\underline{z}_1 - \underline{z}_2\| \quad ,$$

such that $\max_t k(t)$ can be taken to be as small as desired. Select η and μ_1 sufficiently small such that $k(t)$ may be chosen small enough so that condition (3.40) is satisfied (i.e. $K < 1$), where $p(t)$ and $q(s)$ are

bounds on G(t, s). Furthermore, let μ_2 be the small positive number such that for $|\mu| \leq \mu_2$,

$$|\mu| \max_{p(t)} (1-K)^{-1} \int_0^1 q(s) \| \underline{g}(\underline{x}, s) \| \, ds \leq \eta \quad .$$

Then, for $\| \underline{z} \| \leq \eta$ and $|\mu| \leq \min(\mu_1, \mu_2)$, all the conditions of Lemma 1 are fulfilled. Therefore, the above integral equation possesses an exact unique solution such that $\| \underline{z} \| \leq \eta \leq \delta$.

However, it is shown previously that the integral equation is equivalent to the differential system

$$\frac{d\underline{z}}{dt} = \frac{\partial \underline{f}(\underline{x}, t)}{\partial \underline{x}} \underline{z} + \underline{f}^*(\underline{z}, t) + \underline{\epsilon}(t)$$

$$\underline{z}(0) = \underline{z}(1) \quad .$$

Using the definitions of $\underline{\epsilon}(t)$ and $\underline{f}^*(\underline{z}, t)$, the system becomes,

$$\frac{d\underline{z}}{dt} = \underline{f}(\underline{x}, t) - \underline{f}(\underline{x}-\underline{z}, t) + \mu \underline{g}(\underline{x}-\underline{z}, t)$$

$$\underline{z}(0) = \underline{z}(1) \quad .$$

Subtracting this system from the original system (3.35) and defining $\underline{y} \equiv \underline{x} - \underline{z}$, one has

$$\frac{d\underline{y}}{dt} = \underline{f}(\underline{y}, t) - \mu \underline{g}(\underline{y}, t)$$

$$\underline{y}(0) = \underline{y}(1) \quad .$$

For $|\mu| \leq \min(\mu_1, \mu_2)$, the above system possesses an exact unique solution of period 1 such that $\| \underline{x} - \underline{y} \| \leq \delta$. Therefore, the above system may be chosen as the auxiliary system. Q.E.D.

δ represents a bound on the error which is uniform in t. Usually it is desirable to obtain the smallest bound possible, therefore the equality sign is used in equation (3.41) for determining δ.

Generalizations

Some of the hypotheses in Lemma 1 and Theorem 1 can be weakened to include more general systems. The condition that $\underline{f}(\underline{x}, t)$ be C^1 in $\underline{x}$ can be replaced by assuming that $\underline{f}(\underline{x}, t)$ is C^0 in $\underline{x}$ and satisfies a modified Lipschitz condition in $\underline{x}$. The proofs of Lemma 1 and Theorem 1 are only slightly modified with the exception that in Theorem 1 it is no longer possible to conclude that the exact solution is isolated since the equation of first variation is not defined.

The assumption in Lemma 1 that q(s) and k(s) be positive continuous functions for $s \in [0, 1]$ can be weakened to q(s) and k(s) being non-negative, integrable functions for $s \in [0, 1]$ with the loss of "strict" uniqueness. It is possible to conclude only that if there exists two solutions to (3.22), $\underline{z}_1$ and $\underline{z}_2$, then

$$\int_0^1 k(s)\, q(s) \| \underline{z}_1 - \underline{z}_2 \| \, ds = 0 \quad .$$

Therefore, the integrand vanishes everywhere except at a set of points with zero measure. For all values of t where $k(t)q(t) > 0$, $\| \underline{z}_1 - \underline{z}_2 \|$ is zero, which implies $\underline{z}_1 = \underline{z}_2$. Consequently, uniqueness is obtained only over a subset of $t \in [0, 1]$.

Autonomous Systems

Theorem 1 may also be applied to autonomous systems but in a negative manner. In Theorem 1, it is shown that, if certain

conditions are satisfied, the equation of first variation of the original system associated with the exact periodic solution $\underline{x}(t)$ possesses no non-trivial solution with the same period as $\underline{x}(t)$. However, for autonomous systems it is well known that, if the original system possesses a non-trivial periodic solution $\underline{x}(t)$, the equation of first variation associated with $\underline{x}(t)$ has a non-trivial periodic solution $\frac{d\underline{x}(t)}{dt}$ with the same period as $\underline{x}(t)$. In this situation, hypothesis (3.40) in Theorem 1 can never be satisfied. Consider the following autonomous system

$$\frac{d\underline{x}}{dt} = \underline{f}(\underline{x}) \quad . \tag{3.47}$$

Assume $\underline{f}(\underline{x})$ is continuously differentiable with respect to $\underline{x}$ and that the Lipschitz constant $k(s)$ for $\underline{f}^*(\underline{z}, t)$ satisfies (3.43). Differentiating with respect to t, (3.47) becomes

$$\frac{d}{dt}\left(\frac{d\underline{x}}{dt}\right) = \frac{\partial \underline{f}(\underline{x})}{\partial \underline{x}} \frac{d\underline{x}}{dt} \quad . \tag{3.48}$$

Assume that (3.47) possesses a non-trivial periodic solution with period 1. Then, $\frac{d\underline{x}}{dt}$ will also be periodic with period 1 and will satisfy the equation of first variation (3.48). Adding and subtracting $A(t)\ \frac{d\underline{x}}{dt}$ to (3.48), where $A(t)$ is the same matrix that appears in (3.39), and writing this modified equation in integral form, $\frac{d\underline{x}}{dt}$ satisfies

$$\frac{d\underline{x}}{dt} = \int_0^1 G(t, s)\left[\frac{\partial \underline{f}(\underline{x})}{\partial \underline{x}} - A(t)\right]\frac{d\underline{x}}{dt}\, ds \quad , \tag{3.49}$$

where $G(t, s)$ is the Green's function for (3.39). From (3.38), $\underline{f}^*(\underline{x}-\underline{y}, t)$

satisfies

$$\underline{f}^*(\underline{x}-\underline{y}, t) = \underline{f}(\underline{x}) - \underline{f}(\underline{y}) - A(t)(\underline{x}+\underline{y}) \quad , \tag{3.50}$$

where $\underline{y}$ is some periodic function such that for $\underline{z} \equiv \underline{x}-\underline{y}$, $\underline{f}^*(\underline{z}, t)$ satisfies a modified Lipschitz condition given in (3.23) for $\|\underline{z}\| \leq \delta$. Forming the Jacobian matrix for $\underline{f}^*(\underline{x}-\underline{y}, t)$

$$\frac{\partial \underline{f}^*(x-y, t)}{\partial \underline{x}} = \frac{\partial \underline{f}(x)}{\partial x} - A(t) \quad . \tag{3.51}$$

Taking norms of (3.49) and using (3.51) and the bound on $\|G(t, s)\|$ given in (3.24), $\|\frac{d\underline{x}}{dt}\|$ satisfies

$$\|\frac{d\underline{x}}{dt}\| \leq p(t) \int_0^1 q(s)\, k(s) \|\frac{d\underline{x}}{dt}\| ds \quad , \tag{3.52}$$

where it has been assumed that $\|\underline{x}-\underline{y}\| \leq \delta$ so that use of (3.43) is justified. Multiplying by $q(t)\, k(t)$ and integrating, the above relation becomes

$$(1-K) \int_0^1 q(t)\, k(t) \|\frac{d\underline{x}}{dt}\| dt \leq 0 \quad , \tag{3.53}$$

where the definition of K has been used. But if $\|\frac{d\underline{x}}{dt}\| \neq 0$ and $q(t)\, k(t) \neq 0$, the integral is non-zero, and, consequently, (3.53) implies $K \geq 1$. Therefore, Theorem 1 will never apply to an autonomous system whenever there exists a non-trivial periodic solution lying in $\|\underline{x}-\underline{y}\| \leq \delta$.

However, there are situations when Theorem 1 is applicable to autonomous systems. Suppose that all of the hypotheses of Theorem 1 are satisfied. Then the only solution of the equation of first variation (3.48) is the trivial solution. Therefore,

$$\underline{x}(t) = \underline{c} \quad , \tag{3.54}$$

where $\underline{c}$ is a constant vector and satisfies $\underline{f}(\underline{c}) = \underline{0}$. Consequently, the only exact solution lying in $\|\underline{x}-\underline{y}\| \leq \delta$ is the degenerate solution (3.54). Theorem 1 may be construed as a "non-existence" theorem when applied to autonomous systems. A region $\|\underline{x}-\underline{y}\| \leq \delta$ is obtained where the only solution to the original system (3.47) is the degenerate solution (3.54).

Discussion

For nonautonomous systems, Theorem 1 provides a means for obtaining bounds on the norm of the exact solution error $\underline{z}(t)$ in terms of the magnitude of the differential equation error $\underline{\varepsilon}(t)$. If the philosophy behind the equivalent equation approach is used, the mean differential equation error E is minimized with respect to the differential equation parameters $\alpha_i (i=1, \ldots, r)$. Although the form of E is not exactly the same as the form minimized in Chapter II, it still represents a valid equivalence criterion. Furthermore, E can be related to the mean square differential equation error in such a manner that an equation similar to (3.41) may be obtained expressly in terms of the mean square error. Selecting α_i such that

$$\{E(\alpha_1, \ldots, \alpha_r)\} = \text{minimum} \quad , \tag{3.55}$$

equation (3.41) implies that the error bound is also minimized with respect to α_i. Define δ_1^* as the minimum value of δ satisfying (3.41) whenever the equal sign is utilized. Since p(t) and $K(\delta_1^*)$ depend only on

the approximate solution form (i. e., β_j) and the Green's function, they are independent of explicit α_i. Therefore, the dependence of δ_1^* on α_i may be determined from the dependence of δ_1^* on E. Since $\delta_1^*(1-K(\delta_1^*))$ is a monotone increasing function of δ_1^* and vice versa (see Figure 2), δ_1^* is a monotone increasing function of E. Consequently, the minimum value of δ_1^* occurs whenever E is minimized. Therefore, the equivalent equation approach implies that the error bound δ_1^* is minimized in the space of the differential equation parameters $\alpha_i (i=1, \ldots, r)$.

Although the above analysis is performed within the framework of the equivalent equation approach, it is by no means restricted to that approach. The primary reason for using this approach is that it provides a convenient vehicle for carrying out the details and gives a definite approximate solution $\underline{y}$. However, the analysis still applies for approximate solutions obtained using other techniques, i. e., Galerkin's method, method of least squares, etc. In the approaches where a solution form is assumed, the differential equation error $\underline{\varepsilon}(t)$ can be interpreted as an error residual obtained by substituting the approximate solution $\underline{y}$ into the original system (3. 8).

Another aspect of Theorem 1 which deserves some discussion is the fact that the homogenous system (3. 39) generating the Green's function is, in a sense, arbitrary. As pointed out earlier, if A(t) is chosen to be the Jacobian matrix of the original system evaluated at the approximate solution $\underline{y}$, then $\|\underline{f}^*(\underline{z}, t)\|$ is of an order higher than $\|\underline{z}\|$. This is advantageous because the Lipschitz constant k(t) for $\underline{f}^*(\underline{z}, t)$ will have no term independent of δ. However, if A(t) is

any other matrix, in general $\underline{f}^*(\underline{z}, t)$ will have terms linear in $\underline{z}$, and consequently k(t) will have terms independent of δ. This increases k(t) and thereby increases the bound δ which in turn reduces the region in the parameter space of the original system where Theorem 1 will apply. But there is one great advantage in choosing A(t) to be a matrix other than the Jacobian matrix. In general, the Jacobian matrix will be a function of t, and determining a fundamental matrix solution for (3.39) could be a most difficult, if not impossible, task. By selecting A(t) such that a fundamental matrix solution is known, this difficulty is avoided. If A(t) depends on some parameters, they can be considered arbitrary in the analysis and then may be specified, using (3.41), by minimizing the bound δ with respect to them. In this manner, the homogenous system (3.39) is in some sense optimized.

It is also worth noting that Theorem 1 gives no information concerning the stability of the exact periodic solution $\underline{x}(t)$. Since the problem for $\underline{z}(t)$ is recast as a two point boundary value problem over a finite range in t, there is not sufficient time for an asymptotically stable or unstable perturbation to decay or grow. In general, Theorem 1 will apply equally well to stable or unstable periodic solutions. The fact that the equation of first variation possesses no nontrivial solution of period 1 is sufficient to insure that the exact periodic solution $\underline{x}(t)$ varies continuously for small changes in the parameters of the governing original differential equation (3.8)[31]. This is essentially the result given in Theorem 2.

Several other authors have presented results similar to those

given in Lemma 1 and Theorem 1. Those works which seem most closely related to the present analysis are discussed here.

One of the most straightforward approaches is the one utilized by McLaughlin [23, 24]. This work is mainly concerned with second order scalar equations of the type $\ddot{x}+\omega_0 x = \epsilon X(\omega t, x, \dot{x}, \epsilon)$. An approximate periodic solution is obtained using the Poincaré-Linstead perturbation technique. A differential equation governing the difference between the exact periodic solution and a truncated expansion in the small parameter ϵ is formed. The equation is transformed to an integral equation, and bounds on the norms of the error and its first derivative are obtained using a consistency argument. The error $x(t)$ and the first derivative $\dot{x}(t)$ are assumed to satisfy $\|x(t)\| \leq u$ and $\|\dot{x}(t)\| \leq v$. Using standard bounds and inequalities on the integral representation, functions $F_1(u, v, \epsilon)$ and $F_2(u, v, \epsilon)$ are obtained which are bounds on $\|x\|$ and $\|\dot{x}\|$ respectively. Then requiring that F_1 and F_2 satisfy $F_1 \leq u$, $F_2 \leq v$ makes the entire argument consistent. The equal sign is used to obtain the smallest bound. The implicit function theorem guarantees the existence of a unique solution of the above equations so long as the Jacobian matrix,

$$\begin{vmatrix} \frac{\partial F_1}{\partial u} & \frac{\partial F_1}{\partial v} \\ \frac{\partial F_2}{\partial u} & \frac{\partial F_2}{\partial v} \end{vmatrix} ,$$

is non-singular. The limit of applicability is obtained by setting the determinant equal to zero. McLaughlin applies the above method only

to the approximations obtained by the perturbation technique, but it is also valid for approximations obtained by other means. Although the above procedure is straightforward, it suffers from the fact that the estimates used to get F_1 and F_2 are rather poor. Consequently, the bound obtained is poorer and the region of applicability is smaller than the region obtained using Theorem 1. This fact is illustrated in Section 3.3 where actual comparisons are made in a specific example.

Another approach which uses the contraction mapping theorem is presented by Holtzman [25]. He is primarily interested in obtaining bounds on the error between the exact solution and the approximate solution obtained from the method of equivalent linearization. The results obtained are similar to those of Theorem 1 with the major difference arising again in the accuracy of the bound and the region of applicability of the approach. Holtzman uses the unique linear part with constant coefficients of the error differential equation as the homogenous system to generate the Green's function. Whereas in Theorem 1, the homogenous system is left arbitrary in the analysis and can later be selected in such a manner so as to optimize the error bound. In addition, the estimates involved in applying Holtzman's modification of the contraction mapping principle are somewhat poorer than those used in Theorem 1. Also, Holtzman's approach does not allow for the generalizations of Lemma 1 and Theorem 1 discussed earlier. Comparisons between Holtzman's results and Theorem 1 are also given in Section 3.3.

Another approach is that presented by Urabe [20, 21, 22].

Actually, Theorems 1 and 2 are quite similar to Urabe's work in that they both use the method of successive approximations. However, Urabe is primarily interested in Galerkin's procedure and its relation to the exact periodic solution. He shows that if there exists a Galerkin's approximation of sufficiently high order and if there exists an exact periodic solution, it is always possible to obtain a bound on the magnitude of the difference between the exact solution and the Galerkin's approximation. Urabe also proves the converse. However, in practice, one is usually interested in very low order approximations, i.e. one or two term approximations, and, therefore, it is necessary to make as sharp estimates as possible in any bound analysis. Since Urabe is not confined to low order approximations, he can afford to use poorer estimates in obtaining his results because he can simply increase the order of the approximation to where his estimates are sufficient. Because of the similarity between Urabe's approach and Theorem 1, no comparison between the two is included in Section 3.3, although the sharper estimates in Theorem 1 would seem to indicate that the corresponding results would show some improvement.

3.2 Error Bounds for Second Order Scalar Systems

In this section some of the results presented in Section 3.1 are specialized for the case of second order scalar equations. The norm to be used is the absolute value. The initial result is a theorem which is the second order scalar equivalent of Theorem 1.

Formulation

Let R be a region in E^2 and L be the real line. For convenience,

let $\underline{x}$ denote the point $(x, \dot{x})$. Dots over functions mean differentiation with respect to t. Consider the following original system

$$\ddot{x} + f(x, \dot{x}, t) = F(t) \quad , \tag{3.56}$$

where f is C^1 for $\underline{x} \in R$ and f and F are C^0 for $t \in L$. Furthermore, assume f and F are periodic in explicit t with period 1. It is of interest to obtain an approximate solution of (3.56) with period 1 and a bound on the error associated with this approximation. Use the equivalent equation approach by considering the auxiliary system

$$\ddot{y} + g(y, \dot{y}, \alpha_1, \ldots, \alpha_j, t) = G(\alpha_{j+1}, \ldots, \alpha_r, t) \ , \tag{3.57}$$

where g is C^1 for $y \in R$, g and G are C^0 for $t \in L$, and $\alpha_i (i=1, \ldots, r)$ are differential equation parameters. g and G are also assumed to be periodic in explicit t with period 1. Assume further that (3.57) possesses known periodic solutions $\underline{y}(\beta_1, \ldots, \beta_s, t)$ of period 1 where $\beta_j (j=1, \ldots, s)$ are solution parameters.

Define the error as

$$z(t) = x(t) - y(t) \quad . \tag{3.58}$$

Differentiating (3.58) and using (3.56), (3.57), and (3.58), the equation governing z(t) is

$$\ddot{z} + f(y+z, \dot{y}+\dot{z}, t) - f(y, \dot{y}, t) = \epsilon(t) \quad , \tag{3.59}$$

where $\epsilon(t)$ is the differential equation error given by

$$\epsilon(t) = F(t) - f(y, \dot{y}, t) + g(y, \dot{y}, \alpha_1, \ldots, \alpha_j, t) - G(\alpha_{j+1}, \ldots, \alpha_r, t). \tag{3.60}$$

Since x and y are periodic with period 1, z will satisfy

$$z(0) = z(1) \quad , \quad \dot{z}(0) = \dot{z}(1) \ . \tag{3.61}$$

Assume that the following homogenous problem possesses only the trivial solution

$$\ddot{\varphi} + a(t)\dot{\varphi} + b(t)\varphi = 0$$
$$\varphi(0) = \varphi(1) \quad , \quad \dot{\varphi}(0) = \dot{\varphi}(1) \quad , \tag{3.62}$$

where a(t) and b(t) are C^0 for $t \in [0, 1]$. (3.62) will then possess a Green's function G(t, s) which enables the two point boundary value problem (3.59) and (3.61) to be represented as

$$z(t) = \int_0^1 G(t, s)\Big[\epsilon(s) + f^*(z(s), \dot{z}(s), s)\Big]\, ds$$
$$\dot{z}(t) = \int_0^1 \frac{\partial G(t, s)}{\partial t}\Big[\epsilon(s) + f^*(z(s), \dot{z}(s), s)\Big]\, ds \quad , \tag{3.63}$$

where

$$f^*(z, \dot{z}, t) \equiv f(y, \dot{y}, t) - f(y+z, \dot{y}+\dot{z}, t) + a(t)\dot{z} + b(t)z \quad . \tag{3.64}$$

Making use of the above formulation, the following result is possible.

Theorem 3

If the following conditions hold:

i) Systems (3.56) and (3.57) possess sufficient smoothness so that the formulation in (3.63) is justified.

ii) $f^*(z, \dot{z}, t)$ satisfies a modified Lipschitz condition

$$|f^*(z_2, \dot{z}_2, s) - f^*(z_1, \dot{z}_1, s)| \le k(s)|z_2 - z_1| + l(s)|\dot{z}_2 - \dot{z}_1| \quad , \tag{3.65}$$

for $|z_1| \le \delta$, $|z_2| \le \delta$, $|\dot{z}_1| \le \dot{\delta}$, and $|\dot{z}_2| \le \dot{\delta}$ (δ and $\dot{\delta}$ are constants) where k(s) and l(s) are positive continuous functions for $s \in [0, 1]$.

iii) The Green's function G(t, s) can be bounded as

$$|G(t,s)| \le p(t)q(s) \quad , \quad \left|\frac{\partial G(t,s)}{\partial t}\right| \le m(t)\,q(s) \quad , \tag{3.66}$$

for all $t \in [0,1]$ and $s \in [0,1]$ where $p(t)$ and $m(t)$ are non-negative integrable bounded functions for $t \in [0,1]$ and $q(s)$ is a positive continuous function for $t \in [0,1]$.

iv) $$K \equiv \int_0^1 q(s)\Big(p(s)k(s) + m(s)\,l(s)\Big)\,ds < 1 \quad . \tag{3.67}$$

v) $$\begin{aligned} p(t)\Big|_{\max}\,(1-K)^{-1}E &\le \delta \\ m(t)\Big|_{\max}\,(1-K)^{-1}E &\le \dot{\delta} \quad , \end{aligned} \tag{3.68}$$

where

$$E = \int_0^1 q(s)\,|\epsilon(s)|\,ds \quad . \tag{3.69}$$

Then (3.56) possesses an exact unique solution $x(t)$ with period 1. Also the error and its derivative, z and $\dot{z}$, are bounded by

$$\begin{aligned} |z(t)| &\le p(t)\,(1-K)^{-1}E \quad , \\ |\dot{z}(t)| &\le m(t)\,(1-K)^{-1}E \quad . \end{aligned} \tag{3.70}$$

Furthermore, if $\left|\frac{\partial f^*}{\partial z}\right| \le k(s)$ and $\left|\frac{\partial f^*}{\partial \dot{z}}\right| \le l(s)$ for all z and $\dot{z}$ such that $|z| \le \delta$ and $|\dot{z}| \le \dot{\delta}$, then the exact unique solution is an isolated solution.

The proof of the above theorem is essentially the same as the proof of Theorem 1 except for some additional details concerning the convergence of two iteration schemes, one for z and another for $\dot{z}$. Because of the similarity, the proof of Theorem 3 is omitted.

A result analogous to Theorem 2 may also be proven for the second order scalar case. However, no additional information or

insight is gained from this specialization since Theorem 2 applies to second order scalar systems as well. Consequently, it is not included.

Discussion

The remarks made in Section 3.1 concerning generalizations of Theorem 1 also apply to Theorem 3. The restrictions on $f(x, \dot{x}, t)$, $k(t)$, and $q(s)$ can be weakened somewhat with a corresponding weakening of the results. Furthermore, much of the discussion appearing in Section 3.1 concerning Theorem 1 is pertinent to Theorem 3 also. In particular, the homogenous system (3.62) generating the Green's function $G(t, s)$ is essentially arbitrary, the only restriction being that $a(t)$ and $b(t)$ are such that $G(t, s)$ indeed exists. (3.62) may be chosen to be either the Jacobian of $f(x, \dot{x}, t)$ evaluated at the approximate solution $\underline{y}$ $\left(\text{i.e. } a(t) = \frac{\partial f(\overline{y})}{\partial \dot{x}},\ b(t) = \frac{\partial f}{\partial x}(\overline{y})\right)$ or any other system whose Green's function is known. The former is desirable since the Lipschitz constants $k(s)$ and $l(s)$, and correspondingly the bounds δ and $\dot{\delta}$, would be made small, the latter is desirable since determining the Green's function for the system using the Jacobian of $f(x, \dot{x})$ may be quite difficult. The discussion concerning autonomous systems in Section 3.1 applies to Theorem 3 as well. Theorem 3 may be interpreted as a non-existence theorem when applied to autonomous systems.

A Particular Green's Function

As mentioned above, the homogenous system (3.62) may be chosen to be a system possessing a known Green's function. One possible system is

$$\ddot{\varphi} + 2\zeta\gamma + \gamma^2\varphi = 0$$
$$\varphi(0) = \varphi(1) \quad , \quad \dot{\varphi}(0) = \dot{\varphi}(1) \quad , \tag{3.71}$$

where ζ and γ are non-negative real constants. A necessary and sufficient condition for (3.71) to possess a Green's function is that the only solution of (3.71) be the trivial solution. Therefore, in the following development, precautions must be taken to insure this. The two linearly independent solutions of (3.71) are well know, and their behavior is different depending on the particular value of ζ. Consequently, the Green's function and bounds are obtained for $\zeta = 0$, $0 < \zeta < 1$, $\zeta = 1$, and $\zeta > 1$. When $\zeta = 0$, it is clear that (3.71) possesses only the trivial solution if γ is restricted such that $\gamma \neq 2n\pi$ for $n=0,1,2,\ldots$ For $\zeta > 0$, (3.71) possesses only the trivial solution for all $\gamma > 0$. Imposing the above restrictions, the Green's function satisfies the following problem,

$$\frac{\partial^2 G(t,s)}{\partial t^2} + 2\zeta\gamma\frac{\partial G(t,s)}{\partial t} + \gamma^2 G(t,s) = 0, \quad t \neq s \quad , \tag{3.72}$$

$$\left.\begin{aligned} G(0,s) = G(1,s) \quad , \quad \frac{\partial G(0,s)}{\partial t} = \frac{\partial G(1,s)}{\partial t} \\ \frac{\partial G(s^+,s)}{\partial t} - \frac{\partial G(s^-,s)}{\partial t} = 1 \\ G(s^+,s) - G(s^-,s) = 0 \quad . \end{aligned}\right\} \tag{3.72'}$$

For $\zeta=0$, two linearly independent solutions of (3.72) are $\sin(\gamma t)$ and $\cos(\gamma t)$. The Green's function can be written as

$$G(t,s) = \begin{cases} C_1 \sin(\gamma t) + C_2 \cos(\gamma t), & 0 \le t \le s \le 1 \\ C_3 \sin(\gamma t) + C_4 \cos(\gamma t), & 0 \le s \le t \le 1 \quad . \end{cases}$$

The constants C_1, C_2, C_3, and C_4 (which depend on s) are determined using the four conditions in (3.72′). Performing the algebra, G(t, s) is, for $\zeta = 0$ and $\gamma \neq 2n\pi$

$$G(t,s) = \begin{cases} \dfrac{\sin\big(\gamma(t-s+1)\big) - \sin\big(\gamma(t-s)\big)}{2\gamma(1-\cos\gamma)}, & 0 \le t \le s \le 1 \\ \dfrac{\sin\big(\gamma(t-s)\big) - \sin\big(\gamma(t-s-1)\big)}{2\gamma(1-\cos\gamma)}, & 0 \le s \le t \le 1 \ . \end{cases} \tag{3.73}$$

For $0<\zeta<1$, two linearly independent solutions of (3.72) are $e^{-\zeta\gamma t}\sin(\gamma(1-\zeta^2)^{1/2}t)$ and $e^{-\zeta\gamma t}\cos(\gamma(1-\zeta^2)^{1/2}t)$. Writing the Green's function for the two regions $0 \le t \le s \le 1$ and $0 \le s \le t \le 1$, as above, and using (3.72′) to determine the coefficients, G(t, s) is found to be, for $0<\zeta<1$,

$$G(t,s) = \begin{cases} \dfrac{e^{-\zeta\gamma(t-s+1)}}{D_1}\Big[\sin\big(\gamma(1-\zeta^2)^{1/2}(t-s+1)\big) \\ \qquad\qquad -e^{-\zeta\gamma}\sin\big(\gamma(1-\zeta^2)^{1/2}(t-s)\big)\Big], \quad 0 \le t \le s \le 1 \\ \dfrac{e^{-\zeta\gamma(t-s)}}{D_1}\Big[\sin\big(\gamma(1-\zeta^2)^{1/2}(t-s)\big) \\ \qquad\qquad -e^{-\zeta\gamma}\sin\big(\gamma(1-\zeta^2)^{1/2}(t-s-1)\big)\Big], \quad 0 \le s \le t \le 1 \ , \end{cases} \tag{3.74}$$

where $D_1 = \gamma(1-\zeta^2)^{1/2}\big(1+e^{-2\zeta\gamma}-2e^{-\zeta\gamma}\cos(\gamma(1-\zeta^2)^{1/2})\big)$.

For $\zeta = 1$, two linearly independent solutions of (3.72) are $e^{-\gamma t}$ and $te^{-\gamma t}$. Performing the same operations as above, G(t, s) is, for $\zeta = 1$,

$$G(t,s)=\begin{cases}\dfrac{e^{-\gamma(t-s+1)}}{(1-e^{-\gamma})^2}\left[1+(t-s)(1-e^{-\gamma})\right], & 0\le t\le s\le 1\\[2ex] \dfrac{e^{-\gamma(t-s)}}{(1-e^{-\gamma})^2}\left[1+(t-s-1)(1-e^{-\gamma})\right], & 0\le s\le t\le 1\ .\end{cases} \tag{3.75}$$

For $\zeta>1$, two linearly independent solutions of (3.72) are $e^{-\gamma(\zeta+(\zeta^2-1)^{1/2})t}$ and $e^{-\gamma(\zeta-(\zeta^2-1)^{1/2})t}$. Again performing the same calculations as described above, G(t, s) is, for $\zeta>1$,

$$G(t,s)=\begin{cases}\dfrac{e^{-\zeta\gamma(t-s+1)}}{D_2}\Big[\sinh\Big(\gamma(\zeta^2-1)^{1/2}(t-s+1)\Big) \\ \qquad -e^{-\zeta\gamma}\sinh\Big(\gamma(\zeta^2-1)(t-s)\Big)\Big], & 0\le t\le s\le 1\ ,\\[2ex] \dfrac{e^{-\zeta\gamma(t-s)}}{D_2}\Big[\sinh\Big(\gamma(\zeta^2-1)^{1/2}(t-s)\Big) \\ \qquad -e^{-\zeta\gamma}\sinh\Big(\gamma(\zeta^2-1)^{1/2}(t-s-1)\Big)\Big], & 0\le s\le t\le 1\ ,\end{cases} \tag{3.76}$$

where $D_2=\gamma(\zeta^2-1)^{1/2}\left[1+e^{-2\zeta\gamma}-2e^{-\zeta\gamma}\cosh(\gamma(\zeta^2-1)^{1/2})\right]$.

For Theorem 3, it is necessary to obtain bounds on $|G(t,s)|$ and $\left|\frac{\partial G(t,s)}{\partial t}\right|$ satisfying (3.66). For all values of $\zeta\ge 0$, G(t, s) and $\frac{\partial G(t,s)}{\partial t}$ is a function of the variable (t-s). For convenience, we choose as bounds for $|G(t,s)|$ and $\left|\frac{\partial G(t,s)}{\partial t}\right|$ the maximum value attained as (t-s) varies over the allowable range. Consequently, the bounds will be independent of t and s. Obtaining this maximum is relatively straightforward, although it tends to be somewhat lengthy. It is necessary to consider endpoints and all relative extrema. Since the

procedure is basically algebra, the details are bypassed, and only the results are given. For $\zeta = 0$ and $\gamma \neq 2n\pi$,

$$|G(t, s)| \leq \frac{1}{\gamma\left(2(1-\cos\gamma)\right)^{1/2}} \quad ,$$

$$\left|\frac{\partial G(t, s)}{\partial t}\right| \leq \frac{1}{\left(2(1-\cos\gamma)\right)^{1/2}} \quad . \tag{3.77}$$

For $0 < \zeta < 1$,

$$|G(t, s)| \leq \frac{1}{\gamma(1-\zeta^2)^{1/2}\left[1+e^{-2\zeta\gamma}-2e^{-\zeta\gamma}\cos\left(\gamma(1-\zeta^2)^{1/2}\right)\right]} \quad ,$$

$$\left|\frac{\partial G(t, s)}{\partial t}\right| \leq \frac{1}{(1-\zeta^2)^{1/2}\left[1+e^{-2\zeta\gamma}-2e^{-\zeta\gamma}\cos\left(\gamma(1-\zeta^2)^{1/2}\right)\right]} \quad . \tag{3.78}$$

For $\zeta = 1$,

$$|G(t, s)| \leq \frac{e^{\frac{\gamma}{1-e^{-\gamma}}-\gamma-1}}{\gamma(1-e^{-\gamma})} \quad ,$$

$$\left|\frac{\partial G(t, s)}{\partial t}\right| \leq \frac{1-(1+\gamma)e^{-\gamma}}{(1-e^{-\gamma})^2} \quad . \tag{3.79}$$

For $\zeta > 1$,

$$|G(t, s)| \leq \frac{e^{-\zeta\gamma}\left[\left(\zeta-(\zeta^2-1)^{\frac{1}{2}}\right)\left(-e^{-\zeta\gamma}+e^{\gamma(\zeta^2+1)^{\frac{1}{2}}}\right)\right]\frac{\zeta}{(\zeta^2-1)^{\frac{1}{2}}}}{\gamma\left[1+e^{-2\zeta\gamma}-2e^{-\zeta\gamma}\cosh\left(\gamma(\zeta^2-1)^{\frac{1}{2}}\right)\right]\frac{\zeta}{2(\zeta^2-1)^{\frac{1}{2}+\frac{1}{2}}}}$$

$$\left|\frac{\partial G(t,s)}{\partial t}\right| \leq \frac{\zeta e^{-\zeta\gamma}\sinh\left(\gamma(\zeta^2-1)^{\frac{1}{2}}\right)-(\zeta^2-1)^{\frac{1}{2}}\left(1-e^{-\zeta\gamma}\cosh\left(\gamma(\zeta^2-1)^{\frac{1}{2}}\right)\right)}{(\zeta^2-1)^{1/2}\left[1+e^{-2\zeta\gamma}-2e^{-\zeta\gamma}\cosh\left(\gamma(\zeta^2-1)^{1/2}\right)\right]} \quad . \tag{3.80}$$

In the development of these bounds, ζ and γ are left unspecified. They may now be determined by minimizing the error bound δ with respect to them.

Particularization of Theorem 3 using (3.71).

The relations for determining the bounds δ and $\dot{\delta}$ using Theorem 3 and the homogenous system (3.71) are now developed. Consider a specific original system and let the auxiliary system and approximate solution be obtained in any manner whatever. A corresponding differential equation error (or residual) $\epsilon(t)$, given by (3.60), will also be generated. Satisfying the postulates of Theorem 3 will enable a bound on the error z and $\dot{z}$ to be obtained.

Assume that $f(x, \dot{x}, t)$ satisfies the continuity conditions in (3.56). The next requirement is that the homogenous system (3.71) possesses only the trivial solution. As shown previously, this is satisfied for all values of γ(except zero) whenever ζ is non-zero and for all values of γ, except $\gamma = 2n\pi$, for n=0, 1,..., wherever ζ vanishes. Imposing these restrictions on ζ and γ guarantees that the only solution of (3.71) is the trivial solution.

It is now necessary to show that $f^*(\underline{z}) = f^*(z, \dot{z}, t)$, satisfies a Lipschitz condition (3.65). From (3.65), $f^*(z)$ is

$$f^*(\underline{z}) = f(\underline{y}) - f(\underline{y}+\underline{z}) + 2\zeta\gamma\dot{z} + \gamma^2 z \quad .$$

Considering the four variables z_1, z_2, $\dot{z}_1$, and $\dot{z}_2$, the difference $f^*(\underline{z}_2) - f^*(\underline{z}_1)$ is

$$f^*(\underline{z}_2) - f^*(\underline{z}_1) = f(\underline{y}+\underline{z}_1) - \underline{f}(\underline{y}+\underline{z}_2) + 2\gamma\zeta(\dot{z}_2 - \dot{z}_1) + \gamma^2(z_2 - z_1) \quad . \qquad (3.81)$$

Since $f(\underline{x})$ is C^1 for $\underline{x} \in R$, the mean value theorem may be used to give

$$f(\underline{y}+\underline{z}_1) - f(\underline{y}+\underline{z}_2) = \frac{\partial f}{\partial x}\Big(\underline{y}+\underline{z}_1+\lambda(\underline{z}_2-\underline{z}_1)\Big)(\underline{z}_1-\underline{z}_2) + \frac{\partial f}{\partial \dot{x}}\Big(\underline{y}+\underline{z}_1+\lambda(\underline{z}_2-\underline{z}_1)\Big)(\dot{z}_1-\dot{z}_2) \quad , \qquad (3.82)$$

where λ satisfies $0<\lambda<1$. Using this, (3.81) becomes

$$f^*(\underline{z}_2) - f^*(\underline{z}_1) = \left[\gamma^2 - \frac{\partial f}{\partial x}\Big(\underline{y}+\underline{z}_1+\lambda(\underline{z}_2-\underline{z}_1)\Big)\right](z_2-z_1) + \left[2\zeta\gamma - \frac{\partial f}{\partial \dot{x}}\Big(\underline{y}+\underline{z}_1+\lambda(\underline{z}_2-z_1)\Big)\right](\dot{z}_2-\dot{z}_1) \quad . \qquad (3.83)$$

Taking absolute values and using the triangle inequality, (3, 83) becomes, for $|z_1| \le \delta$, $|z_2| \le \delta$, $|\dot{z}_1| \le \dot{\delta}$, $|\dot{z}_2| \le \dot{\delta}$,

$$|f^*(\underline{z}_2) - f^*(\underline{z}_1)| \le k(t)\,|z_2-z_1| + l(t)|\dot{z}_2-\dot{z}_1| \quad , \qquad (3.84)$$

where

$$\left.\begin{aligned} k(t) &= \max_{\substack{0<\lambda<1 \\ |z_1|\,\&\,|z_2|\le\delta \\ |\dot{z}_1|\,\&\,|\dot{z}_2|\le\dot{\delta}}} \left|\gamma^2 - \frac{\partial f}{\partial x}\Big(\underline{y}+\underline{z}_1+\lambda(\underline{z}_2-\underline{z}_1)\Big)\right| \quad , \\ l(t) &= \max_{\substack{0<\lambda<1 \\ |z_1|\,\&\,|z_2|<\delta \\ |\dot{z}_1|\,\&\,|\dot{z}_2|<\dot{\delta}}} \left|2\zeta\gamma - \frac{\partial f}{\partial \dot{x}}\Big(\underline{y}+\underline{z}_1+\lambda(\underline{z}_2-\underline{z}_1)\Big)\right| \quad . \end{aligned}\right\} \qquad (3.85)$$

Therefore, $f^*(\underline{z})$ satisfies a Lipschitz condition with the Lipschitz constants being given by (3.85). Actually, k(t) and l(t) can be any functions greater than or equal to the values given in (3.85), and the Lipschitz condition will still hold. This fact is used for the examples discussed in Sections 3.3 and 3.4.

Bounds on G(t, s) and $\frac{\partial G(t, s)}{\partial t}$ have already been developed in (3.77) - (3.80). Because of the particular bounds chosen, p(t), m(t), and q(s) appearing in (3.66) are all constants. Without loss in generality, choose q(s) equal to one. p(t) and m(t) are then equal to the expressions in (3.77) - (3.80).

The remaining hypotheses to be satisfied are (3.67) and (3.68), i.e.

$$K \equiv p\int_0^1 k(s)dt + m\int_0^1 l(s)ds < 1 \quad , \tag{3.86}$$

and

$$p(1-K)^{-1}E \leq \delta \quad , \quad m(1-K)^{-1}E \leq \dot{\delta} \quad , \tag{3.87}$$

where $E \equiv \int_0^1 |\epsilon(s)|ds$ and p and m are the constant bounds on G(t, s) and $\frac{\partial G(t,s)}{\partial t}$. If (3.86) is true, (3.87) may be written as

$$pE \leq \delta\,(1-K) \quad , \tag{3.88}$$

and

$$mE \leq \dot{\delta}\,(1-K) \quad . \tag{3.89}$$

(3.88) and (3.89) are two relations for determining the bounds δ and $\dot{\delta}$. Since they are bounds and since p, m, and E are positive, δ and $\dot{\delta}$ must be positive. Therefore, if $\delta>0$ and $\dot{\delta}>0$ can be found such that (3.88) and (3.89) are satisfied, Theorem 3 applies, and δ and $\dot{\delta}$ are bounds on the error $|z|$ and $|\dot{z}|$.

In general, there may be more than one pair (δ, $\dot{\delta}$) satisfying (3.88) and (3.89). m, p, and E are independent of δ and $\dot{\delta}$. Since k(t) and l(t) will, in general, depend on δ and $\dot{\delta}$, K will also depend on

δ and $\dot{\delta}$. Define

$$H_1(\delta,\dot{\delta}) = \delta\left(1-K(\delta,\dot{\delta})\right) \quad , \quad H_2(\delta,\dot{\delta}) = \dot{\delta}\left(1-K(\delta,\dot{\delta})\right) \quad . \tag{3.90}$$

(3.88) and (3.89) then become

$$H_1(\delta,\dot{\delta}) \geq pE \quad , \quad H_2(\delta,\dot{\delta}) \geq mE \quad . \tag{3.91}$$

Consider H_1 and H_2 as functions of the two variables δ and $\dot{\delta}$ for $\delta \geq 0$ and $\dot{\delta} \geq 0$. Let S_1 denote the set of points $(\delta, \dot{\delta})$ such that $S_1 = \left\{(\delta, \dot{\delta}) \mid H_1(\delta, \dot{\delta}) \geq pE\right\}$, and S_2 denote the set of points $S_2 = \left\{(\delta, \dot{\delta}) \mid H_2(\delta, \dot{\delta}) \geq mE\right\}$. In order to apply Theorem 3, both relations in (3.91) must hold, therefore the set S_3 which is the intersection of S_1 and S_2 are points $(\delta, \dot{\delta})$ for which Theorem 3 applies. Since S_3 is a closed set, there will exist a point $(\delta_1^*, \dot{\delta}_1^*)$ in S_3 where the bound on $|z|$ is the smallest possible. Similarly, there will exist a point $(\delta_2^*, \dot{\delta}_2^*)$ where the bound on $|z|$ is the largest possible. If it is of interest to obtain a bound on the error $|z|$ between an approximate solution and the exact solution, the point $(\delta_1^*, \dot{\delta}_1^*)$ would be used when applying Theorem 3. However, if it is of interest to prove the existence and uniqueness (or "non-existence" in case of an autonomous system) of a solution the point $(\delta_2^*, \dot{\delta}_2^*)$ would be chosen.

In addition to having some freedom in selecting δ and $\dot{\delta}$, it is possible to minimize (or maximize) the bound δ with respect to the Green's function parameters γ and ζ. By varying γ and ζ, the set S_3 will change, and γ and ζ may be selected such that the particular S_3 is obtained which contains the point $(\delta_i^*, \dot{\delta}_i^*)$ for which the bound on $|z|$ attains its minimum (or maximum) value. Because the relations

yielding S_3 are in the form of inequalities, it is very difficult in most cases to perform the minimization indicated above. Considerable simplification (along with some restriction) may be obtained by requiring that strict equality holds in (3.91) (and, therefore, (3.88 and (3.89)). The equations determining δ and $\dot{\delta}$ then become

$$pE = \delta\left(1-K(\delta, \dot{\delta})\right) \quad , \tag{3.92}$$

and

$$mE = \delta\left(1-K(\delta, \dot{\delta})\right) \quad . \tag{3.93}$$

It is now relatively straightforward to perform the above minimization. The minimum (or maximum) obtained using (3.92) and (3.93) will, in general, be larger (or smaller) than the true minimum (or maximum) of S_3. However, it still will be valid for applying Theorem 3.

Dividing (3.92) by p and (3.93) by m and noting that $K<1$, δ and $\dot{\delta}$ satisfy

$$\dot{\delta} = \frac{m}{p}\,\delta \quad . \tag{3.94}$$

This shows that δ and $\dot{\delta}$ are no longer independent, and, therefore, it is not possible to minimize both δ and $\dot{\delta}$ with respect to γ and ζ. Since the accuracy of the solution and not its time derivative is usually the quantity of most interest, δ is minimized with respect to γ and δ. These two relations, combined with (3.92) and (3.94), are sufficient to determine γ, ζ, $\dot{\delta}$, and δ. Substituting (3.94) into (3.92) eliminates $\dot{\delta}$ from the formulation yielding

$$E = \delta\left(p^{-1} - p^{-1}K(\delta, \frac{m}{p}\,\delta)\right) \quad . \tag{3.95}$$

From (3.90), it is seen that $H_1(\delta, \frac{m}{p}\delta)$ is independent of $\dot{\delta}$. Consider $H_1(\delta, \frac{m}{p}\delta)$ as a function of δ, and let $\hat{\delta}$ be the value of δ such

$K(\hat{\delta}, \frac{m}{p}\hat{\delta}) = 1$. From the definition of H_1, it is clear that

$$H_1(\delta, \frac{m}{p}\delta)\Big|_{\delta=0} = 0 \quad \text{and} \quad H_1(\delta, \frac{m}{p}\delta)\Big|_{\delta=\hat{\delta}} = 0 \ .$$

Since $K(\delta, \frac{m}{p}\delta)$ is a single valued function of δ, $H_1(\delta, \frac{m}{p}\delta)$ will also be single valued in δ. H_1 is non-negative because $\delta > 0$ and $K < 1$. Therefore $H_1(\delta, \frac{m}{p}\delta)$ possesses the general character indicated in Figure 2. (3.95) is satisfied for δ_1^* and δ_2^* which, from Figure 2, are the minimum and maximum values for which Theorem 3 applies. Any δ satisfying $\delta_1^* \leq \delta \leq \delta_2^*$ will be a valid δ for applying Theorem 3.

From Figure 2 it is possible to obtain the boundary of applicability of Theorem 3. When the parameters in the original system, the auxiliary system, and the Green's function are such that δ_1^* and δ_2^* coalesce, the valid region in δ degenerates to a point δ_m. Any change in the parameters so as to increase pE beyond $(pE)_m$ makes it impossible to satisfy (3.95) for any real positive δ. Consequently, the boundary is given by

$$H_1(\delta_m, \frac{m}{p}\delta_m) = (pE)_m \ . \tag{3.96}$$

Returning to (3.95), the minimization of δ with respect to γ and ζ may now be performed. Assuming δ to be a continuous function of γ and ζ, a necessary condition for a minimum is $\frac{\partial \delta}{\partial \gamma} = 0$ and $\frac{\partial \delta}{\partial \zeta} = 0$. δ is first minimized with respect to γ by implicitly differentiating (3.95) and noting that E is independent of δ and γ, p is independent of δ, and K depends on δ and γ. Recalling the definition of K, (3.86), the

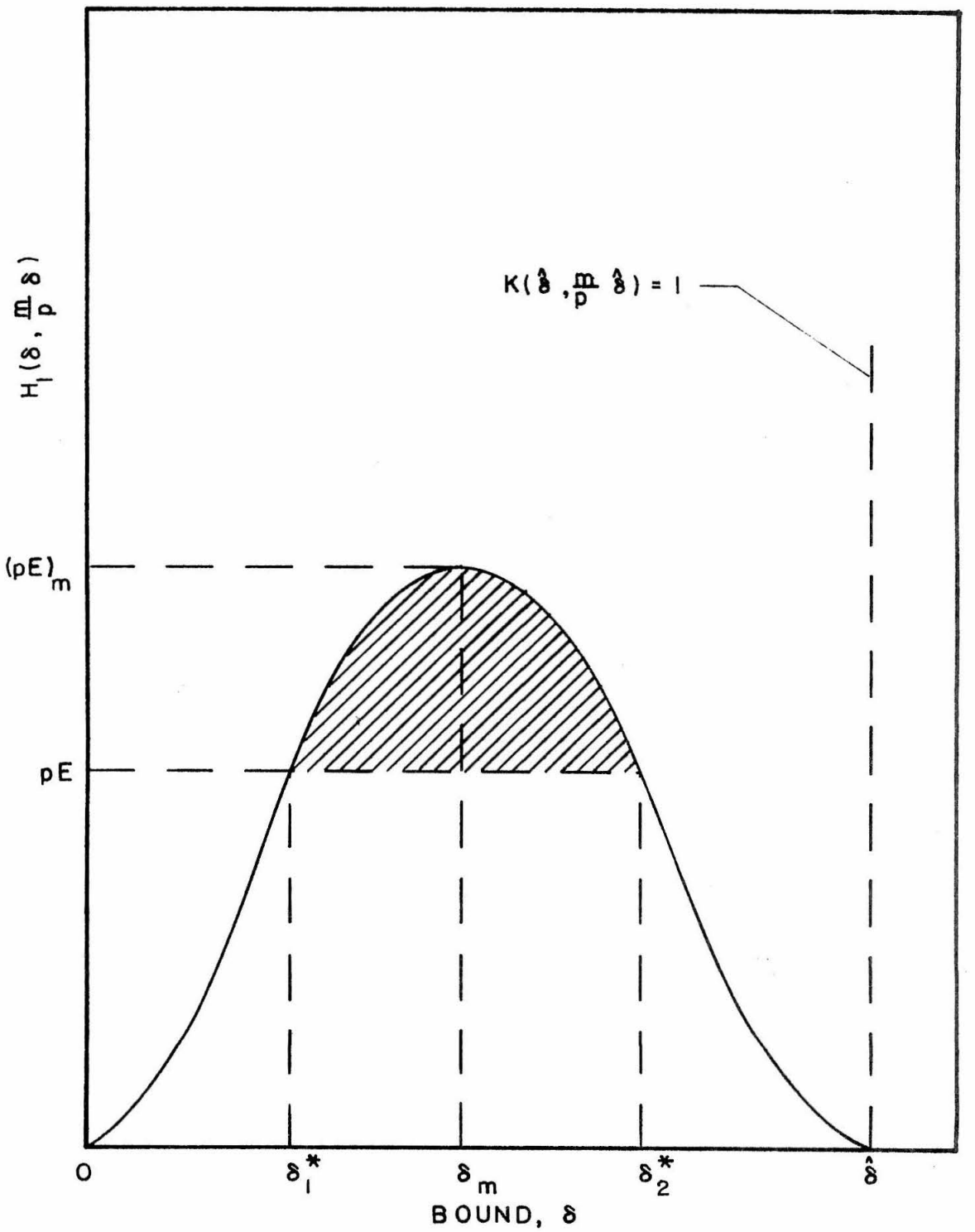

Figure 2: $H_1(\delta, \frac{m}{p}\delta)$ versus δ

condition $\frac{\partial \delta}{\partial \gamma} = 0$ implies

$$\frac{\partial}{\partial \gamma}\left(p^{-1}\right) = \int_0^1 \frac{\partial k(s)}{\partial \gamma} ds + \frac{\partial}{\partial \gamma}\left(\frac{m}{p}\right) \int_0^1 l(s) ds + \frac{m}{p} \int_0^1 \frac{\partial l(s)}{\partial \gamma} ds \quad . \tag{3.97}$$

Since all of the functions in (3.95) which depend on γ also depend on ζ and vice versa, the relation generated by $\frac{\partial \delta}{\partial \zeta} = 0$ can be obtained by replacing explicit γ by ζ in (3.97). Therefore, $\frac{\partial \delta}{\partial \zeta} = 0$ implies

$$\frac{\partial}{\partial \zeta}\left(p^{-1}\right) = \int_0^1 \frac{\partial k(s)}{\partial \zeta} ds + \frac{\partial}{\partial \zeta}\left(\frac{m}{p}\right) \int_0^1 l(s) ds + \frac{m}{p} \int_0^1 \frac{\partial l(s)}{\partial \zeta} ds \quad . \tag{3.98}$$

Equation (3.95), (3.97), and (3.98) will, in general, have to be solved simultaneously to determine ζ, γ, and δ. (3.94) is then used to determine $\dot{\delta}$. Expressions (3.97) and (3.98) may be rewritten using the values of m and p for the particular Green's function discussed earlier*. This operation is straightforward, and it would serve no useful purpose to include these calculations for all cases. However, in Section 3.3 some examples are presented, and it is convenient to simplify (3.97) and (3.98) for the specific cases considered there.

The examples come from a general class of problems where the original system contains no dissipative terms i.e. $f(x, \dot{x}, t)$ is independent of $\dot{x}$. This being the case, the auxiliary system will usually have the same property, and therefore it seems appropriate

---

*Equations (3.95), (3.97), and (3.98) remain valid for any Green's function depending on two parameters γ and ζ such that p and m are constants and q is unity.

to consider a Green's function where $\zeta \equiv 0$. Furthermore, the Lipschitz condition is simplified because $l(t) \equiv 0$. This fact essentially uncouples the equations (3.88) and (3.89) for determining δ and $\dot{\delta}$ since (3.88) can now be solved separately for δ. Therefore, the restriction imposed by requiring (3.92) and (3.93) is actually no restriction at all. (3.88) and (3.89) being separable implies that the true minimum (or maximum) of S_3 occurs when (3.92) and (3.93) are satisfied.

(3.77) gives the values of m and p for the case where $\zeta \equiv 0$. It is worthwhile noting again that γ is restricted such that $\gamma \neq 2n\pi$ to insure the existence of a Green's function. From (3.77),

$$p = \frac{1}{\gamma\left(2(1-\cos\gamma)\right)^{1/2}} \quad , \tag{3.99}$$

and

$$m = \frac{1}{\left(2(1-\cos\gamma)\right)^{1/2}} \quad . \tag{3.100}$$

Substituting these values into (3.95) and rearranging, the equation becomes

$$E = \delta\left[\gamma\left(2(1-\cos\gamma)\right)^{1/2} - \int_0^1 k(s)ds\right] \quad . \tag{3.101}$$

Using (3.99) and (3.100) in (3.97) and performing the indicated differentiation, the following relation is obtained for determining the optimum γ,

$$\frac{2(1-\cos\gamma)+\gamma\sin\gamma}{2\left(2(1-\cos\gamma)\right)^{1/2}} = \frac{1}{\gamma}\int_0^1 \frac{\partial k(s)}{\partial \gamma}\,ds \quad . \tag{3.102}$$

$\dot{\delta}$ is determined using (3.94), (3.99), and (3.100) and is

$$\dot{\delta} = \gamma\delta \quad .$$

The problem of applying Theorem 3 to non-dissipative systems is reduced to solving (3.101) and (3.102) for γ and δ and then using the above relation to calculate $\dot{\delta}$. Given an original system and a fully determined auxiliary system, it is necessary to determine k(s) greater than or equal to the one given in (3.85). Using k(s), the integrals in (3.101) and (3.102) may be evaluated, and consequently δ and $\dot{\delta}$ may be determined. The following section uses these results for a specific nonautonomous non-dissipative system.

3.3. Error Bounds for a Specific Nonautonomous System

The system considered is

$$\frac{d^2x}{d\tau^2} + \hat{a}x + \hat{\mu}x^3 = \hat{F}\cos(\omega\tau) \quad . \tag{3.103}$$

It is of interest to determine the accuracy of the approximate periodic solution of (3.103) obtained using the equivalent equation approach. Two approximations are considered. The first is obtained using equivalent linearization. The second is the approximation discussed in Section 2.2, where an equivalent cubic equation is used.

The method of equivalent linearization uses the auxiliary system

$$\frac{d^2y}{d\tau^2} + \hat{K}y = \hat{F}\cos(\omega\tau) \quad . \tag{3.104}$$

(3.104) has known periodic solutions of the form

$$y = A\cos(\omega\tau) \quad , \tag{3.105}$$

where A satisfies

$$A = \frac{\hat{F}}{\hat{K} - \omega^2} \quad . \tag{3.106}$$

$\hat{F}$ and ω in (3.104) are selected to have the same values as $\hat{F}$ and ω in (3.103), and $\hat{K}$ may be determined by the method of Chapter II which is identical to the standard method of equivalent linearization. $\hat{K}$ satisfies

$$\hat{K} = \hat{a} + \tfrac{3}{4}\hat{\mu} A^2 \quad . \tag{3.107}$$

Given $\hat{a}, \hat{\mu}, \hat{F}$, and ω, the approximate solution (3.105) is determined using (3.106) and (3.107). If the approximation has more than one solution, it is necessary to select a particular solution of interest.

Having obtained the approximate solution (3.105), it is of interest to determine its accuracy. The independent variable τ is normalized so that (3.103), (3.104), and (3.105) are periodic with period 1. Let

$$\tau = \frac{2\pi}{\omega} t \quad . \tag{3.108}$$

Equation (3.103) then becomes

$$\ddot{x} + ax + \mu x^3 = F \cos(2\pi t) \quad , \tag{3.109}$$

where $a = \hat{a}(\frac{2\pi}{\omega})^2$, $\mu = \hat{\mu}(\frac{2\pi}{\omega})^2$, and $F = \hat{F}(\frac{2\pi}{\omega})^2$. Equation (3.104) reduces to

$$\ddot{y} + Ky = F \cos(2\pi t) \quad , \tag{3.110}$$

where $K = \hat{K}(\frac{2\pi}{\omega})^2$ and F is the same as in (3.109). The approximate solution is

$$y = A \cos(2\pi t) \quad . \tag{3.111}$$

To use the general analysis developed in Section 3.2, it is

necessary to determine the Lipschitz constant defined in (3.85). Using the notation of the general system (3.56),

$$f(x, \dot{x}) = ax + \mu x^3 \quad , \tag{3.112}$$

and

$$F(t) = F \cos(2\pi t) \quad . \tag{3.113}$$

$f(x, \dot{x})$ is independent of $\dot{x}$; consequently, the Lipschitz constant $l(t)$ for the derivative can be taken to be zero, and the two relations (3.92) and (3.93) determining δ and $\dot{\delta}$ are uncoupled. δ can be found first. Furthermore, since there is no dissipation, the Green's function(3.73) for $\zeta=0$ is used. The bounds on $|G(t, s)|$ and $\left|\frac{\partial G(t, s)}{\partial t}\right|$ are given by (3.99) and (3.100) respectively.

The Lipschitz constant $k(t)$ has to be greater than or equal to the expression given in (3.85). Therefore, for $|z_1|<\delta$ and $|z_2|<\delta$, $k(t)$ satisfies

$$k(t) = \max_{\substack{|z_1|\&|z_2|<\delta \\ 0<\lambda<1}} \left|\gamma^2 - a - 3\mu\left(y + z_1 + \lambda(z_2 - z_1)\right)^2\right| \leq \left|\gamma^2 - a - 3\mu y^2\right|$$

$$+ \max_{\substack{0<\lambda<1 \\ |z_1|\&|z_2|<\delta}} \left[\left|6\mu y\left(z_1 + \lambda(z_2 - z_1)\right)\right| + \left|3\mu\left(z_1 + \lambda(z_2 - z_1)\right)^2\right|\right] .$$

But if $|z_1| \leq \delta$, $|z_2| \leq \delta$, and $0 < \lambda < 1$, $|z_1 + \lambda(z_2 - z_1)| \leq \delta$. Therefore

$$k(t) \leq \left|\gamma^2 - a - 3\mu y^2\right| + 6|\mu||y|\delta + 3|\mu|\delta^2 \equiv \hat{k}(t) \quad . \tag{3.114}$$

The Lipschitz constant may be taken to be $\hat{k}(t)$ defined in (3.114). For convenience, the $\hat{\ }$ is dropped.

To determine the optimum γ using (3.102), the integral $\frac{1}{\gamma}\int_0^1 \frac{\partial k}{\partial \gamma}\,dt$ must be evaluated. Dividing the integral at the zeros of the first term in k(t) eliminates the absolute value sign in the first term. This term must contain zeros since γ is restricted to the range $a<\gamma^2<a+3\mu A^2$, $(\mu>0)$ or $a+3\mu A^2<\gamma^2<a$, $(\mu<0)$, so that (3.102) has a solution. Differentiating with respect to γ eliminates the remaining terms in k(t). Performing the resulting integration yields

$$\frac{1}{\gamma}\int_0^1 \frac{\partial k}{\partial \gamma}\,dt = \frac{\mu}{|\mu|}\,2\left(1-\frac{4}{\pi}\cos^{-1}\left(\frac{\gamma^2-a}{3\mu A^2}\right)^{1/2}\right) \qquad (3.115)$$

In order to determine δ, it is necessary to compute $\int_0^1 k\,dt$ and E in (3.101). $\int_0^1 k\,dt$ is determined by integrating (3.114) term by term. Each resulting integral is divided at the zeros of the integrand eliminating the absolute value signs. The result is

$$\int_0^1 k(t)\,dt = \frac{\mu}{|\mu|}(\gamma^2-a)(1-8\hat{t}) + 3|\mu|A^2(4\hat{t}-1/2+\frac{1}{\pi}\sin 4\pi\hat{t}) + \frac{12}{\pi}|\mu|A\delta + 3|\mu|\delta^2 \qquad (3.116)$$

where

$$\hat{t} = \frac{1}{2\pi}\cos^{-1}\left(\frac{\gamma^2-a}{3\mu A^2}\right)^{1/2}.$$

To compute E, it is convenient, in this example, to modify the definition given in (3.87). Although the form given in (3.87) is readily calculable for the present approximation, it becomes awkward for the approximation generated using the equivalent cubic equation. For comparison purposes, it is desirable to use the same form for E in both approximations. From Figure 2, it is clear that if E is

increased, the region for valid δ is reduced. If E is increased but is still required to satisfy equations (3.68), all of the arguments in Theorem 3 remain valid. Consequently, Theorem 3 may still be used. However, as mentioned above, increasing E increases δ_1^* and decreases δ_2^*. Therefore, the error bound δ_1^* is larger (i.e. poorer) for the increased value of E. In the present example, the added convenience of using an alternative form of E more than compensates for what can be shown to be a slight (10%) increase in δ_1^*.

A convenient E to use is the following

$$E = \left(\int_0^1 \epsilon^2(t)t \right)^{1/2} . \tag{3.117}$$

By Schwartz's inequality,

$$\int_0^1 |\epsilon(t)|dt \leq \left(\int_0^1 \epsilon^2(t)dt \right)^{1/2} ,$$

so that E given by (3.117) is a valid definition to use. From (3.60), $\epsilon(t)$ is given by

$$\epsilon(t) = ay + \mu y^3 - Ky = \frac{\mu A^3}{4} \cos(6\pi t) , \tag{3.118}$$

where (3.107) and (3.111) have been used. Determining E in (3.117) then gives

$$E = \frac{|\mu| A^3}{4\sqrt{2}} . \tag{3.119}$$

Substituting (3.115) into (3.102), the relation for determining the optimum γ is

$$\frac{2(1-\cos\gamma)+\gamma\sin\gamma}{2\gamma(1-\cos\gamma)} = \frac{\mu}{|\mu|}\,2\left(1-\frac{4}{\pi}\cos^{-1}\left(\frac{\gamma^2-a}{3\mu A^2}\right)^{1/2}\right) \quad . \tag{3.120}$$

Once γ is determined, δ can be found from (3.101) using (3.116) and (3.119). The relation for δ is

$$\delta\left[\gamma\left(2(1-\cos\gamma)\right)^{1/2} - \int_0^1 k(t)dt\right] = E \quad . \tag{3.121}$$

$\dot{\delta}$ is determined from (3.94) and is given by

$$\dot{\delta} = \frac{1}{\gamma}\,\delta \quad . \tag{3.122}$$

Equation (3.120), (3.121), and (3.122) are the equations of interest for the approximation obtained using equivalent linearization. These were solved for several numerical values of $\hat{a}$, $\hat{\mu}$, $\hat{F}$, and ω. The results for $\hat{a}=1$, $\hat{\mu}=0.1$, and $\hat{F}=0.1$ are given in Figures 3 and 4 in the form of plots of δ versus ω. The results for $\hat{a}=1$, $\hat{\mu}=-0.2$, and $\hat{F}=0.2$ are given in Figures 5 and 6.

Prior to discussing the results, it is convenient to develop the bound for the approximation obtained using the equivalent cubic equation. The auxiliary system is

$$\frac{d^2x}{d\tau^2} + \hat{a}y + \hat{\mu}y^3 = \hat{F}'\,\mathrm{cn}(\hat{\eta}\tau, \bar{k}) \quad , \tag{3.123}$$

where $\hat{a}$ and $\hat{\mu}$ have the same values as in (3.103), and $\hat{\eta}$ and $\bar{k}$ are related so that the period of the excitation in (3.123) is the same as the period in (3.103), i.e. $2\pi/\omega$. (3.123) has exact periodic solutions in terms of Jacobian elliptic functions of the form

$$y = A\,\mathrm{cn}(\hat{\eta}\tau, \bar{k}) \quad . \tag{3.124}$$

From Section 2.2, the approximate periodic solution for (3.103) for $\hat{\mu}>0$ is determined using (2.27), (2.23), and (2.19) once a value $\bar{k}$ (modulus of the elliptic function) has been assumed. If $\hat{\mu}<0$, (2.28), (2.23)′, and (2.19)′ are used once a value of $\bar{k}_1$ is assumed. ($\bar{k}_1$ is defined in (2.28).) It is necessary to exercise caution when using the above equations from Section 2.2 so that the difference in notation is correctly taken into account.

Once the approximate solution is completely determined, τ is again normalized using (3.108) so that equations (3.103), (3.123), and (3.124) have period 1. Equation (3.123) becomes

$$\ddot{y} + ay + \mu y^3 = F' cn(\pi t, \bar{k}) \quad , \tag{3.125}$$

where a and μ are given in (3.109), $F'=\hat{F}'\left(\frac{2\pi}{\omega}\right)^2$ and $\pi = \hat{\pi}\,\frac{2\pi}{\omega}$. Equation (3.124) becomes

$$y = A\, cn(\pi t, \bar{k}) \quad . \tag{3.126}$$

In order to apply the theory of Section 3.2, it is necessary to determine a Lipschitz constant k(t). For the same reasons as in the first approximation, l(t) is set to zero, and the Green's function for $\zeta=0$ is used. Again, δ and the optimum γ are given by (3.101) and (3.102). Since the arguments in developing the Lipschitz constant in (3.114) are independent of the particular approximation used, (3.114) is also valid for the present approximation. Consequently, performing the same operations as indicated above, k(t) is found to be, for $|z|<\delta$,

$$k(t) = |\gamma^2 - a - 3\mu y^2| + 6|\mu|\,|y|\,\delta + 3|\mu|\,\delta^2 \quad , \tag{3.127}$$

where $y = A\, cn(\pi t, \bar{k})$.

It is now possible to determine the integral $\frac{1}{\gamma}\int_0^1 \frac{\partial k(s)}{\partial \gamma} dt$ needed for calculating the optimum γ. Again, separating the integral at the zeros of the first term in k(t), taking the derivative, and evaluating the resulting integrals yields

$$\frac{1}{\gamma}\int_0^1 \frac{\partial k}{\partial \gamma}\, d\hat{t} = 2\,\frac{\mu}{|\mu|}(1-8\hat{t}) \quad , \tag{3.128}$$

where

$$\hat{t} = \frac{1}{\eta}\, cn^{-1}\left(\frac{\gamma^2-a}{3\mu A^2}\right)^{1/2} \quad , \tag{3.129}$$

such that $0<\hat{t}<1/4$.

To determine δ, it is necessary to calculate $\int_0^1 k dt$ and E. The integral is evaluated using the same techniques as in the earlier approximation. The final results involve integrals of products of elliptic functions which can be evaluated using standard reference tables [28]. Performing the algebra, the following result is obtained.

$$\int_0^1 k(t)dt = \frac{\mu}{|\mu|}(\gamma^2-a)(1-8\hat{t}) + \frac{3|\mu|A^2}{4\bar{K}(\bar{k})}\left[\frac{4}{\bar{k}^2}\left(\bar{E}(\bar{k}) - (1-\bar{k}^2)\,\bar{K}(\bar{k})\right)\right.$$

$$\left. + \frac{8}{\bar{k}^2}\left(\bar{E}(am(\eta\hat{t},\bar{k}),\bar{k}\right) - \bar{E}(\bar{k}) + (1-\bar{k}^2)(\bar{K}(\bar{k}) - \eta\hat{t})\right]$$

$$+6|\mu|A\,\frac{\sin^{-1}(\bar{k})}{\bar{K}(\bar{k})}\delta + 3|\mu|\delta^2 \quad , \tag{3.130}$$

where $\bar{K}(\bar{k})$ and $\bar{E}(\bar{k})$ are the complete elliptic integrals of the first and second kind with modulus $\bar{k}$, $\bar{E}(\varphi,\bar{k})$ is the incomplete elliptic integral of the second kind, and $am(u,\bar{k})$ is the Jacobian amplitude function.

Finally, it is necessary to compute E. From (3.60), $\epsilon(t)$ is

$$\epsilon(t) = F \cos(2\pi t) - F' \operatorname{cn}(\eta t, \bar{k}) \quad . \tag{3.131}$$

For convenience, (3.117) is used to calculate E, which gives

$$E = \left[\frac{F^2}{2} - \frac{FF'\pi}{2\bar{k}\bar{K}(\bar{k})} \operatorname{sech}\left(\frac{\pi \bar{K}(\bar{k}')}{2\bar{K}(\bar{k})} \right) \right]^{1/2} . \tag{3.132}$$

where $\bar{k}'^2 = 1 - \bar{k}^2$. In the numerical results to be presented, an accuracy problem developed in computing E for small values of $\bar{k}$. Consequently, a power series in terms of $\bar{k}$ was determined, and the first few terms were used. For $\bar{k}$ small, E was determined using

$$E = \frac{F}{\sqrt{2}} \frac{\bar{k}^2}{4} + O(\bar{k}^4) \quad \text{as} \quad \bar{k} \to 0 \quad . \tag{3.133}$$

Equations (3.120), (3.129), (3.130), and (3.132) are convenient for numerical work only if $\hat{\mu} > 0$. If $\hat{\mu} < 0$, $\bar{k}$ is pure imaginary, which necessitates some modification of these equations prior to performing any numerical computations.

All of the quantities needed for determining the bound have now been obtained. Substituting (3.128) in (3.102) yields an expression for the optimum γ. Once γ is determined, equations (3.130) and (3.132) are substituted into (3.101), which gives a relation for determining the bound δ. δ^* is found using (3.122). Numerical results were obtained for the same values of $\hat{a}$, $\hat{\mu}$, and $\hat{F}$ as those values used in the linear approximation. The results for $\hat{a}=1$, $\hat{\mu}=0.1$, and $\hat{F}=0.1$ are given in Figures 3 and 4 in the form of plots of δ versus ω. The results for $\hat{a}=1$, $\hat{\mu}=-0.2$, and $\hat{F}=0.2$ are given in Figures 5 and 6.

The exact solution errors for the linear and the cubic approximations were obtained by numerically integrating the appropriate differential equations. For the linear approximation, the equation describing the difference $z_L(t)$ between the approximate solution (3.111) and the exact solution is

$$\frac{d^2 z_L}{dt^2} = (K-a)y - az_L - \mu(z_L+y)^3 \quad ,$$

where $y=A\cos(2\pi t)$, K, μ, and a are given in (3.109). $z_L(t)$ is periodic with period 1. The measure for the exact linear error used on Figures 3 and 4 is $\max|z_L(t)|$ for $t\in[0, 1]$. The exact error $z_c(t)$ for the cubic approximation is determined from

$$z_c(t) = z_L(t) + y_L(t) - y_c(t) \quad ,$$

where y_L and y_c are the linear and cubic approximations respectively. As above, the measure used for the exact cubic error on Figures 3 and 4 is $\max|z_c(t)|$ for $t\in[0, 1]$. The curves for the exact error have some portions which are dashed. These indicate extrapolation of the curves. There exists some scatter in the exact solution points for errors smaller than 10^{-5}. It was felt that the accuracy of the computation determining the exact error was only of this order, and to indicate this, the curves are dashed for this portion also.

Discussion

Figures 3 and 4 give the results of the above analysis for the values $\hat{a}=0.1$, $\hat{\mu}=0.1$, and $\hat{F}=0.1$. Since $\hat{\mu}>0$, the restoring force is termed "hardening". Figures 5 and 6 give the results for the case

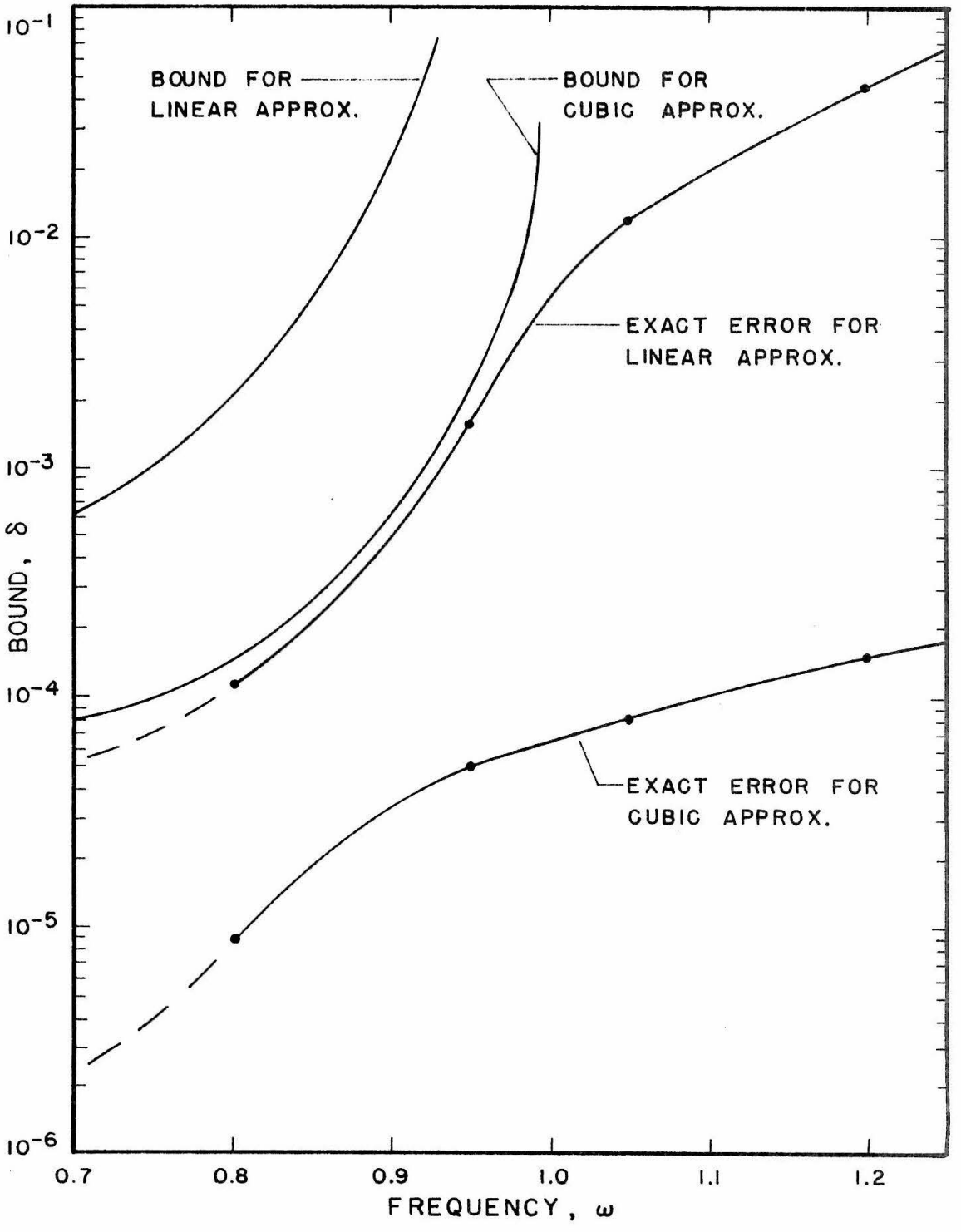

Figure 3: Error Bounds for $\ddot{x}+x+0.1x^3=0.1\cos(\omega\tau)$

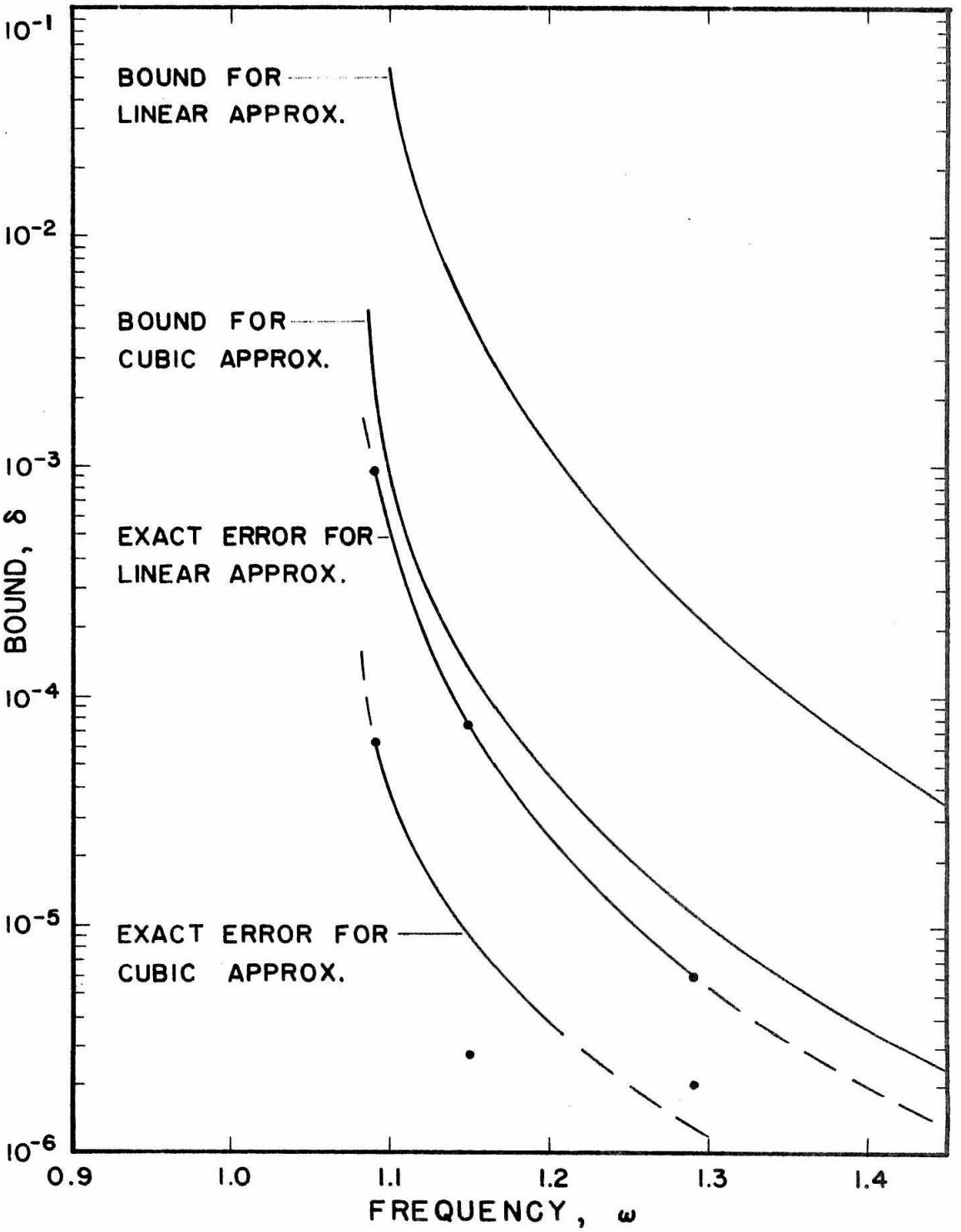

Figure 4: Error Bounds for $\ddot{x} + x + 0.1x^3 = 0.1 \cos(\omega\tau)$

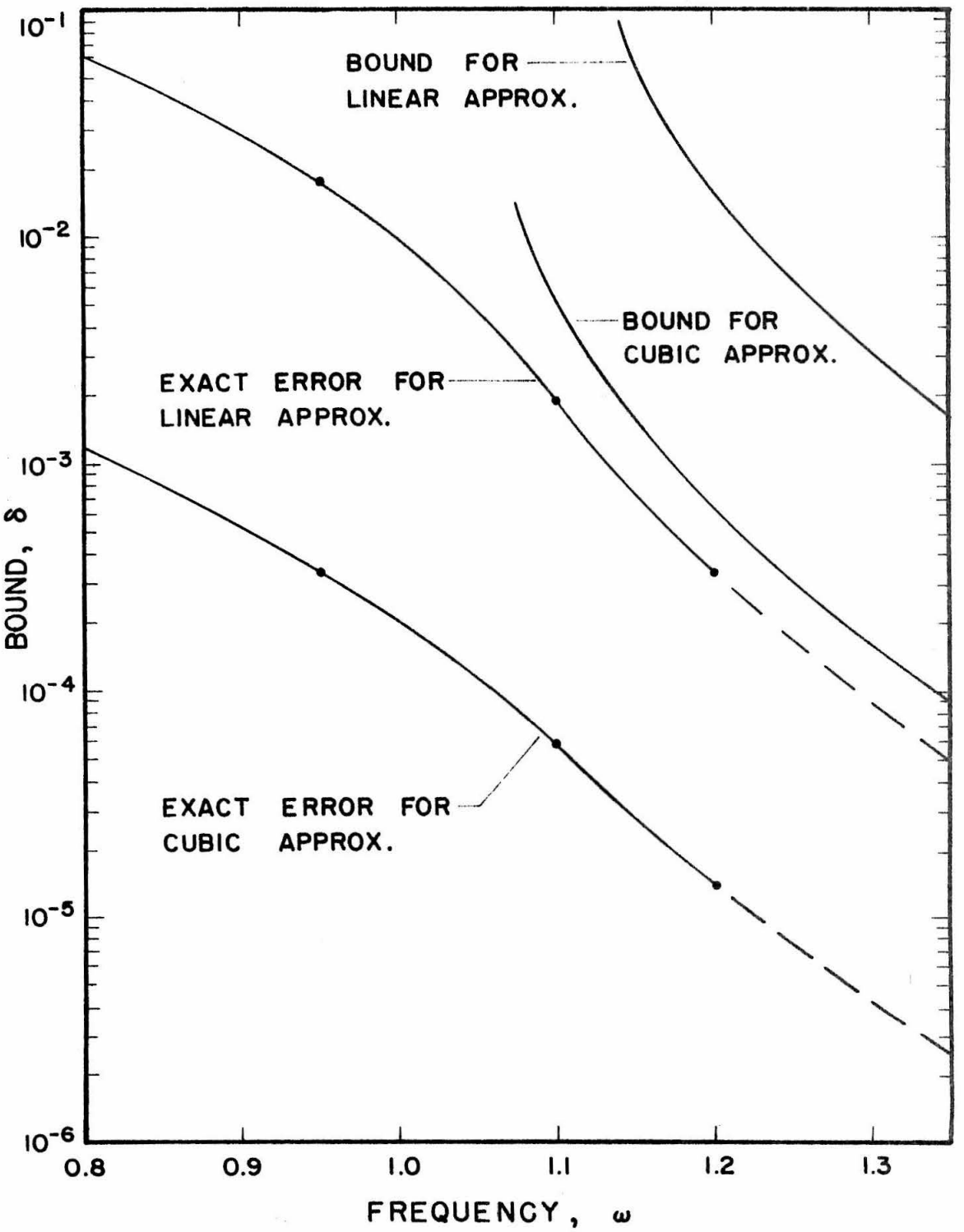

Figure 5: Error Bounds for $\ddot{x} + x - 0.2x^3 = 0.2\cos(\omega\tau)$

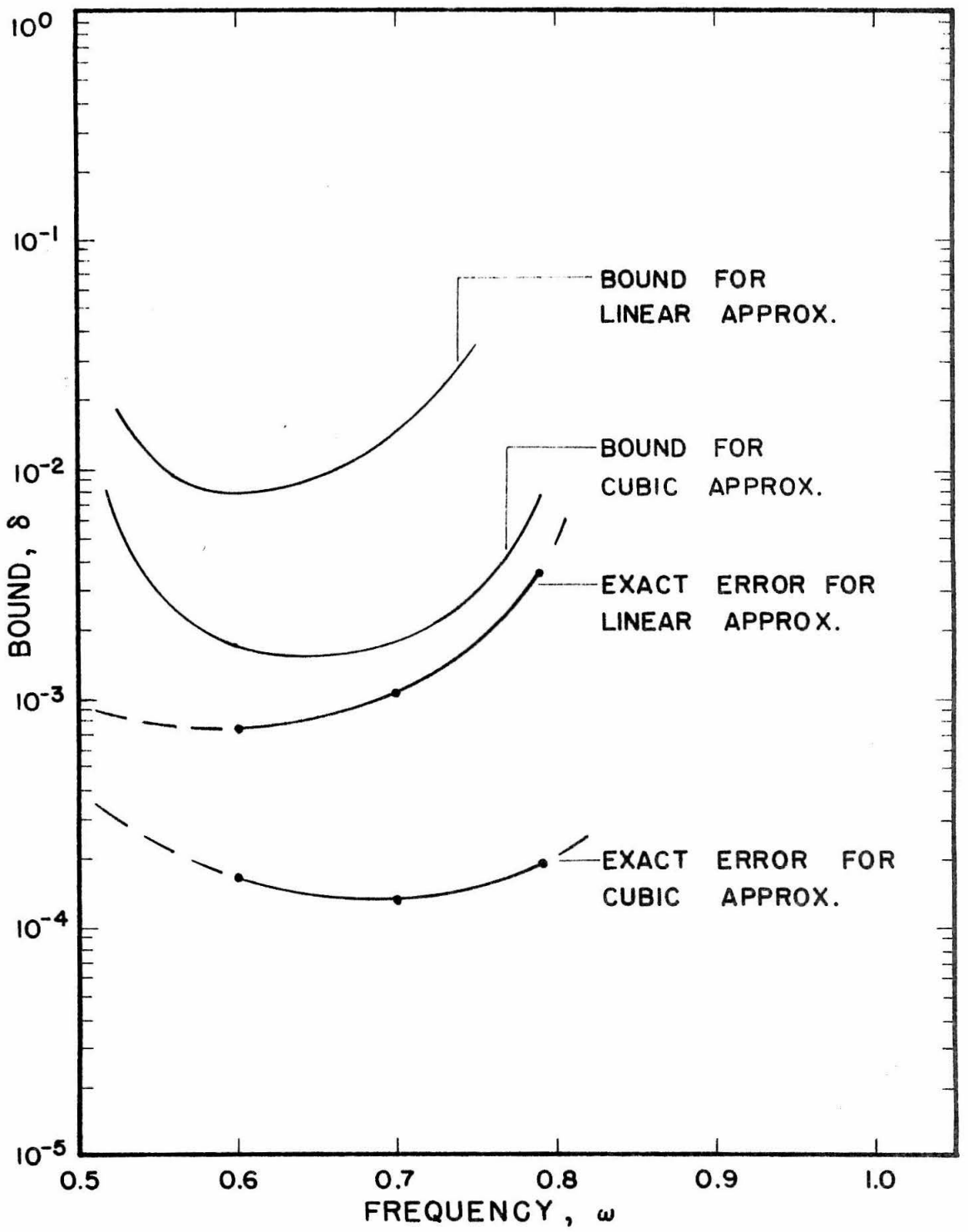

Figure 6: Error Bounds for $\ddot{x} + x - 0.2x^3 = 0.2 \cos(\omega\tau)$

$\hat{a} = 1.0$, $\hat{\mu} = -0.2$, and $\hat{F} = 0.2$. Since $\hat{\mu} < 0$, the restoring force is called "softening".

Figure 3 gives the bound on the magnitude of the difference between the exact solution and the linear approximation, the bound on the magnitude of the difference between the exact solution and the cubic approximation, and the exact error for both approximations for the upper branch of the response curve. (See Figure 1.) Figure 4 presents the same quantities for the stable portion of the lower branch of the response curve. Figures 5 and 6 for the softening case include the bounds for both approximations and the exact errors for the upper branch of the response curve and for the stable portion of the lower branch respectively.

Figures 3 through 6 indicate that it is not possible, using the present analysis, to obtain a bound for all ω. In fact, the expression relating the parameters $\hat{a}, \hat{\mu}, \hat{F}$, and ω specifying the boundary of applicability of Theorem 3 could be determined using (3.96). However, it adds nothing to the discussion to include it. From Figures 3 through 6, it is also clear that the bound for the cubic approximation is roughly an order of magnitude smaller than the bound for the linear approximation. This is in agreement with the actual difference between the exact errors for the two approximations over the range of ω where bounds are obtainable. This does not necessarily mean that, if one approximation leads to a smaller bound than another approximation, the first approximation is better, i.e. its actual error

is smaller. However, without knowledge of the exact error, one is usually more confident in using the approximation providing the smaller error bound.

Figures 3 and 5 also show that a bound is obtainable over a larger range in ω for the cubic approximation than for the linear approximation. The primary reason is that the mean differential equation error E for the cubic approximation is considerably smaller than E for the linear approximation. In a sense, therefore, equation (3.123) better models the original equation (3.103) than does equation (3.104).

Another point worth noting in Figures 3 through 6 is that the bound δ possesses the same general dependence on ω as does the exact error. This seems to indicate that, in the present example, the qualitative behavior of the exact error is described fairly accurately by the qualitative behavior of the bound. Although the actual bounds obtained are an order of magnitude larger than the exact error, smaller bounds could be obtained by using the Green's function for the Jacobian matrix of f(x). In addition, the original definition of E could be used, and/or smaller bounds for the Green's function could be obtained. These improvements would lead to closer bounds.

Comparison with Other Published Work

As mentioned earlier, other authors have obtained error bounds using slightly different techniques. It is therefore of interest to compare the bounds obtained by others to the bounds given by the present approach. Such a comparison has been made for the case

$\hat{a}=1$, $\hat{\mu}=0.1$, and $\hat{F}=0.1$ for the solution corresponding to the upper branch of the response curve and to the stable portion of the lower branch. The same approximate solution, namely the linear approximation (3.104) and (3.105) was used for computing all the bounds in an attempt to compare the various approaches and not the particular approximate solution used.

McLaughlin uses a consistency argument described in Section 3.1 [23, 24]. Although he applies the technique exclusively to approximate solutions obtained using the Poincare'-Linstead perturbation technique, the procedure is valid for approximations obtained by other means as well. In reference (23), McLaughlin considers equation (3.103). He obtains an approximate solution of the form (3.105) where the amplitude A is determined using the perturbation techniques. If this aspect is modified so that A (actually K) is determined using equivalent linearization, the equations developed by McLaughlin apply directly, since his arguments are still valid.

Assuming the error $|z(t)|$ to satisfy $|z| \leq u$, the equation for u (which is (E1.11) in reference (23)) is

$$u - \mu\rho_1\left(u^3 + 3|A|u^2 + 3A^2u + |A^3|\right) = 0 \quad , \tag{3.134}$$

where

$$\rho_1 = \max_{n \geq 0} \left| \frac{1}{1-n^2\omega^2} \right| \quad , \quad n=0,1,2,\ldots \quad .$$

u is the smallest positive real root of (3.134). Equation (3.134) was solved for the values of $\hat{a}$, $\hat{\mu}$. and $\hat{F}$ given above and for the frequency range appearing in Figure 3.

The bound obtained is given in Figures 7 and 8. Figure 7 is the bound for the solution corresponding to the upper branch of the response curve, and Figure 8 gives the bound for the stable portion of the lower branch. Before discussing the results, it is convenient to present the bounds obtained by yet another investigator.

As mentioned in Section 3.1, Holtzman utilizes a modification of the contraction mapping principle to obtain bounds. In reference (25), he also considers equation (3.106) as an example. He uses the approximation obtained by equivalent linearization, consequently the relations he obtains are directly applicable to the present example.

The bound δ is determined by first determining a contraction constant α. α is a root of (equation (59) in reference (25)),

$$3|\mu|T\left(|A|+\frac{cT|\mu A^3|}{4(1-\alpha)}\right)\leq\alpha \quad , \tag{3.135}$$

where $T=\frac{2\pi}{\omega}$, $c=\frac{1}{2|\sin T/2|}$, and A is the approximate solution amplitude. Once α is found, the bound δ satisfies

$$\delta=\frac{k}{1-\alpha} \quad ,$$

where $k=cT\left|\frac{\mu A^3}{4}\right|$. α is the smallest real root of (3.135) satisfying $0\leq\alpha<1$. Equation (3.135) and the above equation were also solved for the hardening case of the present example, and the results are given in Figures 7 and 8. In addition, Figures 7 and 8 include the bounds obtained using the analysis of Section 3.2. The original definition of E (3.87) was used in computing δ. The exact error is also presented.

From the figures, it is clear that the present analysis provides a somewhat sharper bound and is applicable over a larger range in ω than either of the alternative bounds discussed above. It is interesting to note that, although Holtzman's bound is sharper than McLaughlin's, its range of applicability in ω is smaller. This results mainly from simplifications which Holtzman introduces. Although Holtzman's use of the contraction mapping principle is similar to the present analysis, it differs in the manner in which the bound is obtained. The contraction mapping principle enables him to conclude the existence of a unique solution to the error equation, but for bound purposes, he must, in addition, determine the region in which the mapping is a contraction. This region constitutes the bound. The manner in which Holtzman chooses to do this accounts for the somewhat poorer bound and the smaller range of applicability. However, Holtzman's primary interest is not to obtain a bound but rather to determine under what conditions does the existence of a linear approximation imply the existence of an exact solution. No doubt, as Holtzman points out, other methods for determining the contracting region exist which could improve the bound and increase the range of applicability. The figures also show that all three bounds possess the same qualitative behavior as does the exact error.

From the standpoint of convenience, McLaughlin's and Holtzman's bounds are easier to obtain. They both involve a cubic

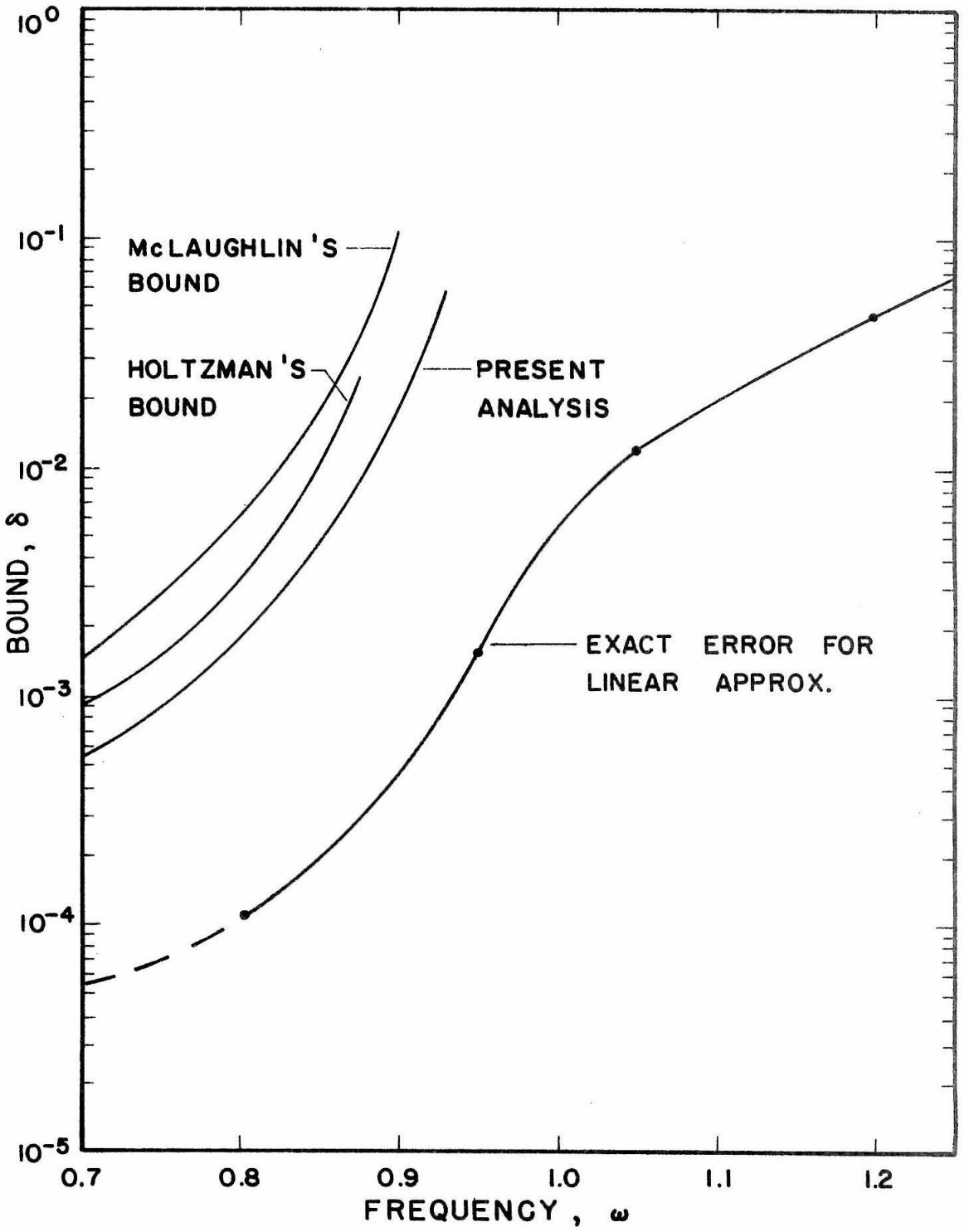

Figure 7: Comparison of Error Bounds

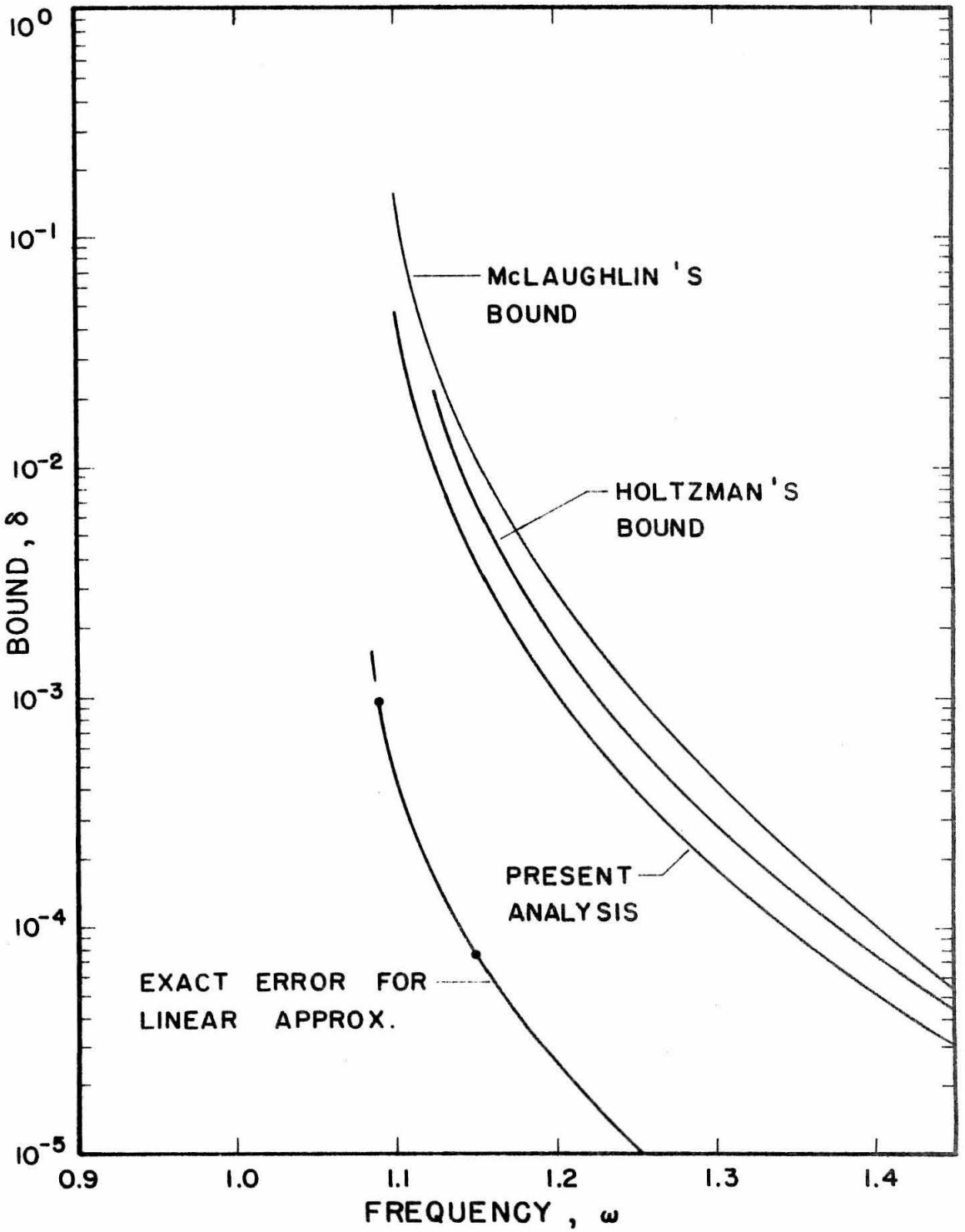

Figure 8: Comparison of Error Bounds

equation. In addition to a cubic equation, the present analysis has a transcendental equation for determining the optimum γ. However, this relation (3.102) is not so involved since it can be put into a very convenient form. The left hand side of (3.102) is a function of γ only and can be plotted once for all γ of practical interest. Then, in a particular situation, it is necessary only to compute the right hand side $\frac{1}{\gamma}\int_0^1 \frac{\partial k}{\partial \gamma}\,dt$ and to locate the point of intersection on the above plot. In the present example, the right hand side is given in (3.115), and the range for γ is $a < \gamma^2 < a + 3\mu A^2$, $(\mu > 0)$, or $a + 3\mu A^2 < \gamma^2 < a$, $(\mu < 0)$. Having determined γ, the cubic equation is then solved for δ. The additional effort needed to use the present analysis is relatively small, consequently the present approach remains manageable and easy to apply. Furthermore, the improvement in the accuracy and the increase in the range of applicability of the bound obtained appear to be ample compensation for the slight increase in effort.

3.4. Error Bounds for a Specific Autonomous System

In this section, Theorem 3 is used to determine a region on the response diagram of an autonomous system where there exists only the trivial solution. Since this approach yields a negative result in that it provides a region where a non-trivial solution cannot exist, a second approach, for conservative autonomous systems, is discussed which yields a region where the non-trivial solution must lie.

Bounds Using Theorem 3.

The system of interest is the following

$$\frac{d^2x}{d\tau^2} + ax + bx^3 = 0 \quad , \tag{3.136}$$

where a and b are constants and $a > 0$. The periodic solution of (3.136) which is symmetric about the origin is of interest. Assuming that (3.136) possesses a solution of period $\frac{2\pi}{\omega}$, τ is normalized so that the solution has period 1. Consequently, using $\tau = \frac{2\pi}{\omega} t$, equation (3.136) becomes

$$\ddot{x} + a\left(\frac{2\pi}{\omega}\right)^2 x + b\left(\frac{2\pi}{\omega}\right)^2 x^3 = 0 \quad . \tag{3.137}$$

The following comparison system is used,

$$\ddot{y} + (2\pi)^2 y = 0 \quad , \tag{3.138}$$

which possesses known periodic solutions of the form

$$y = \hat{A} \cos(2\pi t) \quad , \tag{3.139}$$

Usually, $\hat{A}$ is assumed given and ω is determined approximately using equivalent linearization, however, it is convenient to let ω and $\hat{A}$ be unspecified for the present.

Using Theorem 3, a bound δ on the error $z = x-y$ can be obtained. As mentioned previously, there exists a range of valid δ satisfying Theorem 3. From Figure 2, δ_1^* represents the smallest permissible bound, and δ_2^* represents the largest. When determining a bound on an error, it is clear that δ_1^* is chosen when applying Theorem 3. However, in the present example, an exact solution is known (i.e. $x \equiv 0$). The primary interest is to determine the <u>largest</u> region where there exists <u>only</u> the trivial solution. Consequently, δ_2^*

is chosen when applying Theorem 3. $|z| \leq \delta_2^*$ then represents a region where there exists only the trivial solution. δ_2^* still represents an error bound, but it is not the smallest possible. Since equation (3.137) and (3.138) are identical to equations (3.109) and (3.110) studied in the previous section, except that the excitations are zero, equations (3.121) and (3.120) can be employed directly to determine δ and γ with only one slight modification. In equation (3.121), E is calculated using the fact that $\omega^2 = a + \frac{3}{4} b\hat{A}^2$ obtained from equivalent linearization. However, in the present example, ω is, as yet, unspecified. Therefore, E must be evaluated accordingly. Using equations (3.60) to determine $\epsilon(t)$ and equation (3.117), E satisfies

$$E = \left(\frac{2\pi}{\omega}\right)^2 \hat{A}^2 \left[\frac{(a-\omega^2)^2}{2} + \frac{3}{4}(a-\omega^2) b\hat{A}^2 + \frac{5}{16} b^3 \hat{A}^4\right] . \qquad (3.140)$$

Using the above relation for E in equation (3.121) gives a valid expression for determining δ in the present example.

ω and $\hat{A}$ are now specified in the following manner. ω may be considered known; it is the frequency of the desired solution. $\hat{A}$ may be determined by applying Theorem 3. In addition to providing an error bound δ, Theorem 3 also establishes the existence and uniqueness of the exact periodic solution $x(t)$. Furthermore, Theorem 3 shows that the exact solution is isolated. However, in Section 3.1, it is shown that whenever Theorem 1 (or its second order scalar counterpart Theorem 3) applies to an autonomous system, there can exist only the degenerate solution (3.54) satisfying $|z| \leq \delta$. In the present example, the degenerate solution is the trivial solution.

Consequently, since it is of interest to determine the largest possible region where there can exist only the trivial solution, the quantity $\hat{A} + \delta$ may be maximized with respect to $\hat{A}$ and δ. Since $\hat{A}$ is considered an independent variable, the maximization of $\hat{A}+\delta$ with respect to γ yields $\frac{\partial \delta}{\partial \gamma} = 0$, which is the same relation obtained previously. Therefore, equation (3.120) still represents the relation for maximizing $\hat{A}+\delta$ with respect to γ. Maximizing $\hat{A}+\delta$ with respect to $\hat{A}$ yields

$$1 + \frac{\partial \delta}{\partial \hat{A}} = 0 \quad . \tag{3.141}$$

Implicitly differentiating equation (3.121) to obtain $\frac{\partial \delta}{\partial A}$, (3.141) becomes

$$-\frac{E}{\delta} + \delta \int_0^1 \left(\frac{\partial k}{\partial \delta} - \frac{\partial k}{\partial \hat{A}} \right) dt = \frac{\partial E}{\partial \hat{A}} , \tag{3.142}$$

where the differentiation is with respect to explicit δ and $\hat{A}$. Given ω, equations (3.120), (3.121), and (3.142) are sufficient for determining $\hat{A}$, γ, and δ.

For convenience, in the present example equation (3.142) is not used to determine the optimum $\hat{A}$, but $\hat{A}$ is chosen arbitrarily to be zero. Therefore, the <u>exact</u> solution is chosen as the approximation. Taking the limit as $A \to 0$ in equation (3.120) yields the value of γ to be

$$\gamma = \left(\frac{2\pi}{\omega} \right)^2 a \quad . \tag{3.143}$$

Setting A equal to zero in equation (3.140) gives E=0. Consequently, the equation (3.121) for determining δ reduces to

$$\delta^2 \leq \frac{a\left(2(1 - \cos \gamma)\right)^{1/2}}{3\gamma |b|} . \tag{3.144}$$

Equation (3.144) represents the region in the δ, ω plane (actually A, ω plane, where A is the amplitude of x(t)) where there exists only the trivial solution $x(t) \equiv 0$. Theorem 3 provides a region, $|z| \leq \delta$, where there exists only one solution for (3.136). Since $z = x-y$ and since y is taken to be zero, the region reduces to $|x| \leq \delta$. However, an exact solution is known to be $x \equiv 0$. By invoking Theorem 3 again, $x = 0$ is the only solution satisfying $|x| \leq \delta$. Consequently, if there exist non-trivial solutions to (3.136), they must be exterior to the region $|x| \leq \delta$. It is convenient to define the following dimensionless parameters:

$$\Omega \equiv \frac{\omega^2 - a}{a} \quad \text{and} \quad A^2 = \xi \frac{a}{b} \quad . \tag{3.145}$$

Using the above definitions, equation (3.144) becomes,

for $\xi \geq 0$,

$$\xi \leq \frac{\left(2(1-\cos\gamma)\right)^{1/2}}{3\gamma} \quad ,$$

and, for $\xi \leq 0$, (3.146)

$$\xi \geq -\frac{\left(2(1-\cos\gamma)\right)^{1/2}}{3\gamma} \quad ,$$

where $\gamma = 2\pi(1+\Omega)^{-1/2}$.

Equation (3.146) is plotted in Figure 9. Also included is the exact non-trivial solution of (3.137) which is obtainable using quadratures and involves elliptic integrals. The exact solution in terms of the variables defined in (3.145), is easily found to be

$$\Omega = \frac{\pi^2}{4I^2}\left(1 + \frac{\xi}{2}\right) - 1 \quad , \tag{3.147}$$

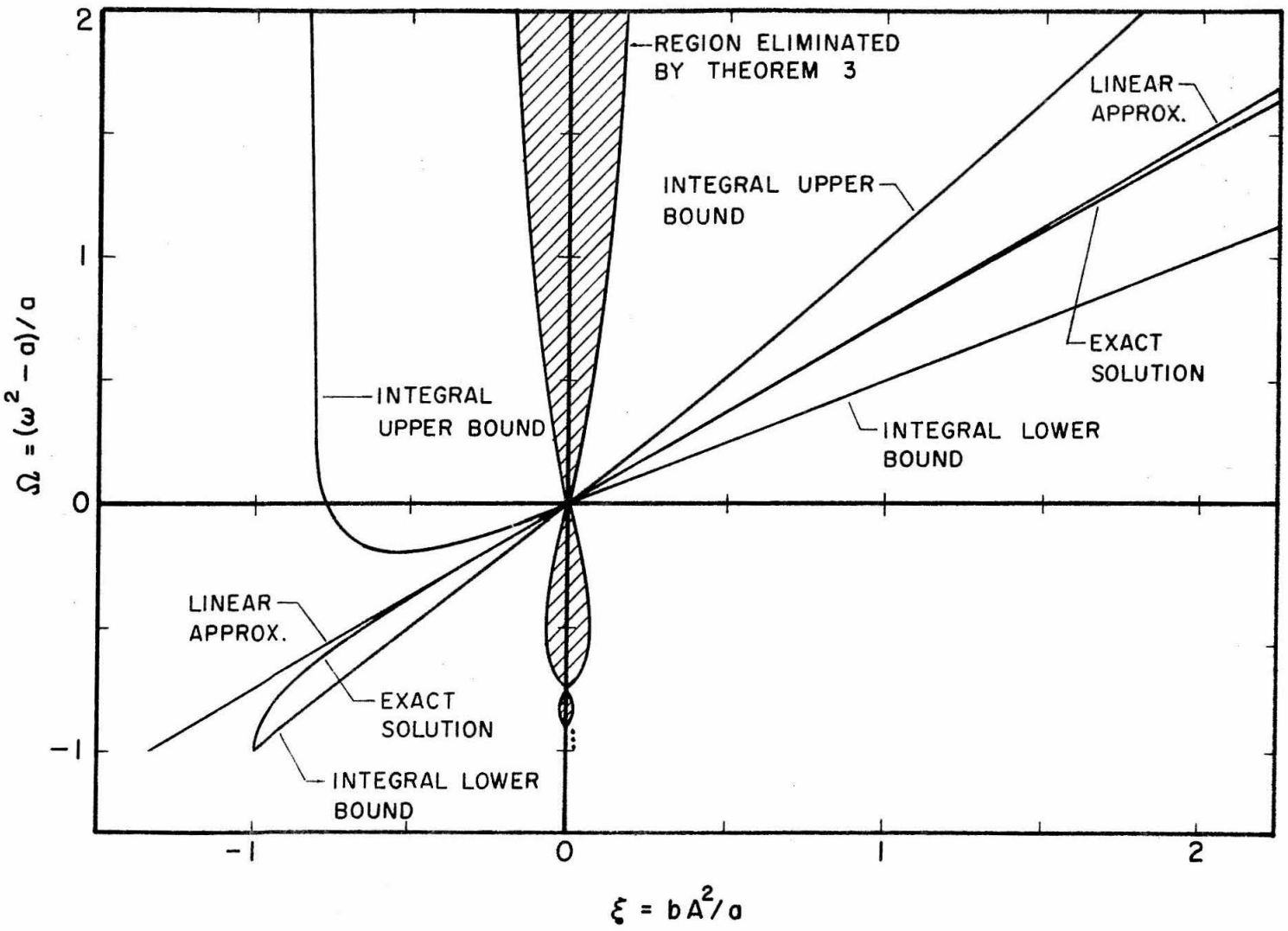

Figure 9: Bounds for Autonomous Duffing's Equation: a > 0

where

$$I = \begin{cases} \frac{1}{\sqrt{2}} \left(\frac{2+\xi}{1+\xi}\right)^{1/2} \bar{K}\left(\frac{1}{\sqrt{2}}\left(\frac{\xi}{1+\xi}\right)^{1/2}\right), & \text{for } \xi > 0 \\ \frac{1}{\sqrt{2}} \left(\frac{2+\xi}{1+\xi}\right)^{1/2} \frac{\bar{K}\left(\frac{\bar{k}}{(1+\bar{k}^2)^{1/2}}\right)}{(1+\bar{k}^2)^{1/2}}, & \text{for } -1 < \xi < 0 \end{cases},$$

and $\bar{K}$ is the complete elliptic integral of the first kind, and $\bar{k} = \frac{1}{\sqrt{2}} \left(\frac{-\xi}{1+\xi}\right)^{1/2}$. No exact periodic solution, symmetric about the origin, exists for $\xi < -1$. $\xi < -1$ represents an initial amplitude so large that the potential energy is outside the potential well for oscillatory motion. Figure 9 also contains bounds obtained using a second approach to be described shortly.

Figure 9 shows that the region defined by (3.146) is quite small. The asymptotic value of the boundary as $\Omega \to \infty$ is $\xi = \pm \frac{1}{3}$. In addition, the region degenerates to a point for values of Ω such that $\Omega = \frac{1}{n^2} - 1$, $n = 1, 2, \ldots$ Practically speaking, the region in the Ω, ξ plane excluded using Theorem 3 is too small to provide meaningful information for the exact non-trivial solution. However, the region does indicate that the bifurcation points for (3.137) for $|b|$ small are associated with the eigenvalues of the linear problem for $b=0$. Therefore, non-trivial solutions of (3.137) for $|b|$ small can emerge only from the points $\Omega = \frac{1}{n^2} - 1$, $n = 1, \ldots$ The above result is well known.

Since Theorem 3 does not provide practical information concerning bounds on the non-trivial solution of (3.137), one is motivated to consider another approach.

Bounds for Conservative Autonomous Systems

In the study of nonautonomous systems, the frequency of the exact periodic solution is specified by some external time varying mechanism. However, for autonomous systems the frequency of the steady-state response is not known beforehand, and most often this is the quantity of interest. For nonautonomous systems, a reasonable criterion for comparing the exact and approximate solution is the maximum of the absolute value of the difference between the solutions. However, this criterion is no longer meaningful for autonomous systems. In the autonomous case, the initial amplitude is usually prescribed, and knowledge of the frequency of the response is desired. The accuracy of an approximate frequency cannot be estimated by considering differences between the exact and the approximate solutions. If both solutions have the same initial conditions (i. e., $x(0) = A$; $\dot{x}(0) = 0$), then since the two frequencies presumably would not be identical, the two solutions would be completely out of phase after a sufficient length of time. At this point, the magnitude of the difference between the two solutions would be 2A regardless of the manner in which the approximate solution is obtained. Therefore, for autonomous systems, it seems more meaningful to seek differences between the exact frequency (or period) and the approximate frequency (or period), given that both solutions started with the same initial conditions.

Consider the conservative autonomous system of the form

$$\ddot{x} + f(x) = 0 \quad , \tag{3.148}$$

where $f(x)$ is C^0 for all x and is an odd function of x. The non-trivial periodic solution of (3.148) that is symmetric about $x=0$ is of interest. In principle, (3.148) can be solved by quadratures using the initial conditions

$$x(0) = A \quad , \quad \dot{x}(0) = 0 \quad . \tag{3.149}$$

Performing the algebra, the exact period T_e is

$$T_e = 2\sqrt{2} \int_0^A \left(\int_s^A f(\zeta)\, d\zeta \right)^{-1/2} ds \quad . \tag{3.150}$$

In many situations, (3.150) can only be solved numerically. It is of interest, however, to be able to determine bounds on T_e without having to resort to numerical computation. For this reason, one is motivated to consider an auxiliary system whose period T_a is determinable in closed form. Hopefully, bounds on the difference between T_a and T_e will be obtained in terms of the difference between the corresponding differential equations.

Assume the auxiliary system to be of the form

$$\ddot{y} + g(y) = 0 \quad , \tag{3.151}$$

where $g(y)$ is C^0 for all y and is an odd function of y. (3.151) is selected so that its period T_a, given by

$$T_a = 2\sqrt{2} \int_0^A \left(\int_s^A g(v)\, dv \right)^{-1/2} ds \quad , \tag{3.152}$$

is known. The initial conditions (3.149) have been utilized in obtaining (3.154). The difference between T_a and T_e is

$$T_e - T_a = 2\sqrt{2} \int_0^A \left[\left(\frac{\int_s^A g(v)dv}{\int_s^A f(v)dv} \right)^{1/2} - 1 \right] \left(\int_s^A g(v)dv \right)^{-1/2} ds \quad . \qquad (3.153)$$

Define

$$\Lambda(u) = \frac{\int_{\breve{u}A}^A g(v)dv}{\int_{\breve{u}A}^A f(v)dv} \quad . \qquad (3.154)$$

Taking absolute values of (3.153), the difference satisfies

$$|T_e - T_a| \leq \max_{0 \leq u \leq 1} |\Lambda^{1/2}(u) - 1| \, T_a \quad . \qquad (3.155)$$

Equation (3.155) represents a bound on the magnitude of $T_e - T_a$ in terms of T_a and the ratio of the potential energies associated with g(y) and f(x). From the form of the bound, it is clear that as g(v) and f(v) tend to the same function, Λ tends to 1, and the bound tends to zero. This implies that the closer in form g(v) and f(v) are, the more accurate the bound is. Furthermore, as the difference between g(v) and f(v) tends to zero, so does the difference between their corresponding periods.

In the present example, $f(x) = a \left(\frac{\omega}{2\pi}\right)^2 x + b \left(\frac{\omega}{2\pi}\right)^2 x^3$. It is again convenient to use (3.138) as the auxiliary system. In addition, let ω be determined using equivalent linearization. ω is given by.

$$\omega^2 = a + \frac{3}{4} b A^2 \quad , \qquad (3.156)$$

where A is in initial amplitude. In order that a periodic approximate solution exists, ω^2 must be positive. Using the variables defined in (3.145), (3.156) becomes

$$\Omega = \frac{3}{4}\xi \quad , \tag{3.157}$$

where $\Omega > -1$ so that periodic solutions exist. Equation (3.157) is also included on Figure 9.

To determine the bound, it is necessary to compute Λ. Using (3.154), Λ is easily found to be

$$\Lambda = \frac{2\omega^2}{2a + bA^2(1+u^2)} \quad . \tag{3.158}$$

It is convenient to rewrite the bound given in (3.155) in terms of the variables defined in (3.145). First, note that, if $M = \max_{0 \le u \le 1} |\Lambda^{1/2}(u) - 1| \le 1$, equation (3.155) implies that

$$\omega_e \le \frac{\omega_a}{1-M} \quad ,$$

or, in terms of the dimensionless variables Ω and ξ,

$$\Omega_e \le \frac{1+\Omega_a}{(1-M)^2} - 1 \quad . \tag{3.159}$$

Similarly, equation (3.155) implies that

$$\Omega_e \ge \frac{1+\Omega_a}{(1+M)^2} - 1 \quad . \tag{3.160}$$

M can be written in terms of Ω and ξ to give

$$M = \max_{0 \le u \le 1} \left| \left(\frac{1+\Omega_a}{1 + \frac{\xi}{2}(1+u^2)} \right)^{1/2} - 1 \right| \quad . \tag{3.161}$$

It is easily shown, using (3.157), that

$$M = \begin{cases} \left(\dfrac{1+\Omega_a}{1+\frac{\xi}{2}}\right)^{1/2} - 1 \ , & \text{for } \xi \geq 0 \\ \left(\dfrac{1+\Omega_a}{1+\xi}\right)^{1/2} - 1 \ , & \text{for } \xi \leq 0 \ . \end{cases} \tag{3.162}$$

The bounds on Ω_e given in equations (3.159) and (3.160) are included in Figure 9.

It is clear from Figure 9 that the bounds obtained using the comparison approach are much more meaningful than those obtained using Theorem 3. The bounds from the comparison approach can be determined over the entire range in ξ and Ω where there exists an exact periodic solution. As ξ approaches -1, the upper bound goes to infinity. This occurs since there exists no exact periodic solution for $\xi = -1$. Consequently, in the present example, the bound going to infinity indicates the non-existence of an exact periodic solution. This fact is important since the approximate analysis implies the existence of a periodic solution for all $\xi > -\frac{4}{3}$. However, since no bound is obtainable for $\xi \leq -1$, the bound analysis indicates that the region for existence of an exact periodic solution is actually $\xi > -1$. Presumably, there exist other techniques which give closer bounds than the approach described above. However, the above approach is conceptually simple and easy to apply.

IV. COMPARISON OF DIFFERENT EQUIVALENCE CRITERIA

In the description of the equivalent equation approach given in Chapter II, two differential systems are said to be equivalent when the mean square of the difference between them is minimized. The minimization is performed with respect to certain parameters contained in the auxiliary system. However, there is no a priori assurance that minimizing the mean square differential equation error will lead to the smallest solution error. There are many other possibilities for minimizing the differential equation error. The purpose of the present chapter is to study the relationship between the solution error and the manner in which the differential equation error is minimized. The investigation is concerned exclusively with second order scalar equations. Three minimization schemes are considered, namely, mean square error minimization, mean absolute value error minimization, and maximum absolute value error minimization. Only the case of periodic motions is considered, therefore, the interval used for the above schemes is one period of the solution. The problem does not appear to be amenable to analytical approaches, therefore, examples will be used to indicate the major results.

The first section presents some preliminary considerations and a formulation of the problem. Section 4.2 gives a description of the

three minimization schemes to be used. Sections 4.3 through 4.6 present specific examples. Section 4.7 contains the results and conclusions of the analysis.

4.1 Preliminaries.

In Chapter II, it is shown that the equivalent equation approach can be used to obtain approximate periodic solutions for equations of the following form

$$\frac{d^2x}{dt^2} + f(x, \dot{x}, t) = F(t) \quad , \tag{4.1}$$

where $f(x, \dot{x}, t)$ and $F(t)$ are periodic in explicit t with period one. This represents no loss in generality since the independent variable may always be normalized so that (4.1) has period one. The procedure is based on considering an auxiliary system having known periodic solutions. This system can be represented as

$$\frac{d^2y}{dt^2} + g(y, \dot{y}, t, \alpha_1, \ldots, \alpha_j) = G(t, \alpha_{j+1}, \ldots, \alpha_r) \quad , \tag{4.2}$$

where g and G are periodic in explicit t with period one, and $\alpha_i (i=1,\ldots,r)$ are parameters which are selected so that equations (4.1) and (4.2) are, in some sense, equivalent.

The manner in which equation (4.2) is made equivalent to (4.1) is of primary interest in the present chapter. Equivalence is based on making the difference between (4.1) and (4.2) small. Specifically, the differential equation error $\epsilon(t)$, given by

$$\epsilon(t) = F(t) - f(y, \dot{y}, t) + g(y, \dot{y}, t, \alpha_1, \ldots \alpha_j) - G(t, \alpha_{j+1}, \ldots \alpha_r) \quad , \tag{4.3}$$

is minimized, in some manner, with respect to the parameters $\alpha_i (i=1,\ldots r)$. By making equations (4.1) and (4.2) similar, it is assumed that the corresponding periodic solutions will also be similar. The relationship between the differential equation error and the solution error is investigated in Chapter III. It is shown that, under certain conditions, the above assumption is justified.

When obtaining an approximate solution for any system, the primary objective usually is to make the error z(t) between the approximate solution y(t) and the exact solution x(t) as small as possible. The ideal situation would be to minimize z(t) with respect to $\alpha_i (i=1,\ldots,r)$. However, z(t) is not known exactly. In general, the only information available concerning the error is that it satisfies $|z(t)| \leq \delta$, where δ is a bound. The next alternative is to minimize δ with respect to $\alpha_i (i=1,\ldots,r)$. In Section (3.2), it is shown that, under certain conditions, δ satisfies

$$\delta = \max_{\forall t} p(t)(1-K)^{-1} E \quad , \tag{4.4}$$

where p(t) and $K(\delta)$ are defined in (3.66) and (3.67). E is an average differential equation error given by

$$E = \int_0^1 |\epsilon(t)| dt \quad . \tag{4.5}$$

From the Cauchy-Schwartz inequality, it is clear that

$$E \leq \left(\int_0^1 \epsilon^2(t) dt \right)^{1/2} \quad . \tag{4.6}$$

Furthermore, from (4.5) E also satisfies

$$E \leq \max_{\forall t} |\epsilon(t)| \quad . \tag{4.7}$$

From the arguments presented in Section 3.3 concerning various definitions of E, it is evident that the definitions given in (4.5), (4.6) and (4.7) are all valid. Consequently, a bound δ can be obtained using (4.4) and any of the above definitions.

For a particular definition of E, it is possible to minimize δ using (4.4). If the differential equation parameters $\alpha_i (i=1,\ldots,r)$ are considered independent of the solution parameters $\beta_j (j=1,\ldots,s)$, δ can be minimized with respect to explicit α_i. After the minimization has been performed, the s relations can then be used to completely determine the α_i and β_j. Since p(t) and $K(\delta)$ depend only on the approximate solution y and the Green's function G(t, s), the minimization of δ with respect to explicit α_i implies

$$\{E(\alpha_1, \ldots, \alpha_r)\} = \text{minimum} \quad , \tag{4.8}$$

where use has been made of (4.4). Hence, minimizing δ leads one, very naturally, to a condition of the form (4.8). Furthermore, the differential equation error is independent of δ which enables the approximation to be obtained independent of the bound. This is advantageous since the conditions necessary for the existence of a bound are not satisfied in general. Therefore, an approximate solution may be obtained even though the bound analysis of Chapter III does not apply. In addition, the minimization procedure (4.8) is unambiguous and,

usually, easy to implement. Since α_i very often appear linearly in $\epsilon(t)$, the relations resulting from (4.8) are quite simple mathematically.

Although the above arguments provide motivation for taking (4.8) as the appropriate condition for determining $\alpha_i (i=1,\ldots,r)$, they do not indicate which definition of E yields the smallest actual solution error. Selecting E so that the smallest bound is obtained does not necessarily mean that the smallest actual error is obtained. Since it is of interest to determine the particular form of E which provides the smallest actual error, it is necessary to consider the exact error and and not a bound. Since the exact error is, in general, unobtainable using analytical techniques, the only recourse is to consider specific examples where the exact error can be determined numerically.

4.2 Description of the Minimization Procedure.

In the examples to follow, three specific definitions of E are considered, namely (4.5), (4.6), and (4.7). The corresponding minimization conditions are

$$\left(\int_0^1 \epsilon^2(t)dt \right)^{1/2} = \text{minimum} \quad , \tag{4.9}$$

$$\int_0^1 |\epsilon(t)|dt = \text{minimum} \quad , \tag{4.10}$$

and

$$\max_{\forall t} |\epsilon(t)| = \text{minimum} \quad . \tag{4.11}$$

The above equations are minimized with respect to $\alpha_i (i=1,\ldots r)$ to generate relations for determining α_i. As discussed in Chapter II, all of the relations resulting from the minimization procedure may not be independent. If this occurs, certain of the α_i must be specified , or the independent relations must be specified, or the independent relations have to be separated, so that a sufficient number of independent relations are generated. A necessary condition for a relative minimum is equation (4. 8). Throughout this chapter, the three conditions (4. 9), (4. 10), and (4. 11) will occur frequently. It is convenient to define the following shorthand notation. ASE, symbolizing Average Square Error minimization, is used to represent (4. 9). AAVE, symbolizing Average Absolute Value Error minimization, is used to represent (4. 10). MAVE, symbolizing Maximum Absolute Value Error minimization, is used for condition (4. 11).

ASE

ASE is one of the most common techniques used. It is easy to apply, and the resulting relations are usually quite simple in form. Using (4. 8) and the condition (4. 9), an alternative form of ASE is

$$\frac{\partial}{\partial \alpha_i}\left(\int_0^1 \epsilon^2(t)dt\right) = 0 \quad , \quad i=1,\ldots,r \quad , \tag{4. 12}$$

or

$$\int_0^1 \frac{\partial \epsilon(t)}{\partial \alpha_i}\epsilon(t)dt = 0 \quad , \quad i=1,\ldots,r \quad . \tag{4. 13}$$

Equations (4. 13) determine $\alpha_i (i=1,\ldots,r)$.

AAVE

AAVE is a somewhat more complicated technique than ASE. The condition in (4. 10) may be written as

$$\int_0^1 \left(\epsilon^2(t)\right)^{1/2} dt = \text{minimum} \quad . \tag{4.14}$$

Minimizing (4. 14) with respect to α_i using (4. 8) yields

$$\frac{\partial}{\partial \alpha_i}\left(\int_0^1 \left(\epsilon^2(t)\right)^{1/2} dt\right) = 0 \quad , \quad i=1,\ldots,r \quad . \tag{4.15}$$

Assuming that $\epsilon(t)$ vanishes for only a finite number of $t \in [0, 1]$, (4. 15) can be written as

$$\int_0^1 \operatorname{sgn}(\epsilon(t)) \frac{\partial \epsilon(t)}{\partial \alpha_i} dt = 0 \quad , \quad i=1,\ldots,r \quad , \tag{4.16}$$

where

$$\operatorname{sgn}(z) = \begin{cases} 1 \ , & \text{for } z>0 \\ 0 \ , & \text{for } z=0 \\ -1 \ , & \text{for } z<0 \end{cases} \quad .$$

Let t_j $(j=1,\ldots,N)$ denote the zeros of $\epsilon(t)$ where $0<t_1<t_2\ldots<t_N\leq 1$. Then, (4. 16) becomes

$$\int_0^{t_1} \frac{\partial \epsilon}{\partial \alpha_i} dt - \int_{t_1}^{t_2} \frac{\partial \epsilon}{\partial \alpha_i} dt \ldots + (-1)^N \int_{t_N}^1 \frac{\partial \epsilon}{\partial \alpha_i} dt \quad , \quad i=1,\ldots,r \quad . \tag{4.17}$$

Since the analysis is primarily concerned with periodic motions, $\epsilon(t)$ is periodic with period one. Therefore, there can exist only an even number of zeros of $\epsilon(t)$. Adding and subtracting the second, the fourth, the sixth, etc., integrals to (4. 17), one obtains

$$\frac{1}{2}\int_0^1 \frac{\partial \epsilon}{\partial \alpha_i}\,dt = \int_{t_1}^{t_2} \frac{\partial \epsilon}{\partial \alpha_i}\,dt + \int_{t_3}^{t_4} \frac{\partial \epsilon}{\partial \alpha_i}\,dt + \ldots + \int_{t_{N-1}}^{t_N} \frac{\partial \epsilon}{\partial \alpha_i}\,dt \quad , \quad i=1,\ldots,r \quad . \quad (4.18)$$

Equations (4.18) determine $\alpha_i (i=1,\ldots,r)$. It is interesting to note the increased complexity of (4.18) compared with equations (4.13).

MAVE

MAVE is an extremely simple minimization scheme conceptually. However, practically speaking, it is the most tedious of the three schemes considered. MAVE minimizes the maximum error for all time. Since $\epsilon(t)$ is periodic, its absolute maximum can occur only at a relative extremum. Define the set of points Φ as

$$\Phi = \left\{ t_j \,\middle|\, 0 \le t_j \le 1 \text{ and } \frac{\partial \epsilon(t_j)}{\partial t} = 0 \right\} .$$

Φ contains all the possible points where $\epsilon(t)$ could attain an absolute maximum. $\alpha_i (i=1,\ldots,r)$ are selected such that

$$\max_{t \in \Phi} |\epsilon(t)| = \text{minimum} \quad . \quad (4.19)$$

(4.19) can become very involved especially if r is larger than 2 or 3 or if Φ contains more than 2 or 3 points where $|\epsilon(t)|$ possesses different values. The difficulties in applying (4.19) are better illustrated in the examples.

4.3 Example 1.

In this example, the following autonomous system is of interest.

$$\frac{d^2x}{d\tau^2} + \tan^{-1}(x) = 0 \quad . \quad (4.20)$$

An approximate amplitude-frequency relation for periodic oscillations symmetric about x=0 is desired. The initial conditions are

$$x(0) = A \quad , \quad \frac{dx(0)}{d\tau} = 0 \quad . \tag{4.21}$$

The equivalent equation approach is used with the auxiliary system

$$\frac{d^2y}{d\tau^2} + Ky = 0 \quad . \tag{4.22}$$

(4.22) is made equivalent to (4.20) by determining K such that the differential equation error is minimized. The periodic solution of (4.22) is

$$y = A\cos(\omega\tau) \quad , \tag{4.23}$$

which satisfies the initial conditions (4.21). Normalize τ, using $t = \frac{\omega}{2\pi}\tau$, so that the solution has period one. Equations (4.20), (4.21), and (4.22) become respectively,

$$\frac{d^2x}{dt^2} + \left(\frac{2\pi}{\omega}\right)^2 \tan^{-1}(x) = 0 \quad , \tag{4.24}$$

$$\frac{d^2y}{dt^2} + \left(\frac{2\pi}{\omega}\right)^2 Ky = 0 \quad , \tag{4.25}$$

and

$$y = A\cos(2\pi t) \quad . \tag{4.26}$$

Using (4.3), the differential equation error $\epsilon(t)$ is

$$\epsilon(t) = \left(\frac{2\pi}{\omega}\right)^2 \left[\tan^{-1}\left(A\cos(2\pi t)\right) - KA\cos(2\pi t)\right] \quad . \tag{4.27}$$

The three equivalence criteria are now used to determine K generating amplitude-frequency relations. These relations are then compared to the exact relation obtained by numerically integrating (4.20).

ASE

The minimization condition using ASE is (4.13). In the present example, there is only one parameter, K. Using (4.27), (4.13) reduce to

$$K = \frac{2}{A}\int_0^1 \tan^{-1}\left(A\cos(2\pi t)\right)\cos(2\pi t)dt \quad . \tag{4.28}$$

The integration may be evaluated by parts. For (4.23) to be a solution of (4.22), K must equal ω^2. Therefore, the amplitude-frequency relation generated by ASE is

$$\omega^2 = \frac{2}{A^2}\left((1+A^2)^{1/2}-1\right) \quad . \tag{4.29}$$

AAVE

The minimization condition generated by AAVE is (4.18). It is necessary to determine the zeros of $\epsilon(t)$. From (4.27), $\epsilon(t)$ is zero whenever

$$\tan^{-1}\left(A\cos(2\pi t)\right) = KA\cos(2\pi t) \quad . \tag{4.30}$$

Letting $\xi = A\cos(2\pi t)$, (4.30) becomes

$$\tan^{-1}(\xi) = K\xi \quad , \tag{4.31}$$

where ξ is restricted to $-A \le \xi \le A$. $\xi=0$ is one root of (4.31). Under further inspection, it is seen that (4.31) possesses two non-trivial roots of equal magnitude and opposite sign if $K<1$. K is expected to be less than one for $A \neq 0$ because (4.20) is a softening system, and one typical feature of softening systems is that the response frequency decreases with increased initial amplitude. Let ξ^* denote the positive

root of (4.31). Because of the symmetry of the cosine function, the zeros of $\epsilon(t)$ may be written as

$$\left.\begin{aligned} t_1 &= t^* \quad , \quad t_2 = 1/4 \quad , \quad t_3 = 1/2 - t^* \quad , \\ t_4 &= 1/2 + t^* \quad , \quad t_5 = 3/4 \quad , \quad t_6 = 1 - t^* \quad , \end{aligned}\right\} \tag{4.32}$$

where

$$t^* = \frac{1}{2\pi}\cos^{-1}\left(\frac{\xi^*}{A}\right) \quad , \quad \text{for } 0<t^*<1/4 \quad .$$

Returning to (4.18), there is only one minimizing parameter, i.e., K. Using (4.32), (4.18) reduces to

$$\frac{1}{2}\int_0^1 \frac{\partial \epsilon}{\partial K}dt = \int_{t_1}^{t_2} \frac{\partial \epsilon}{\partial K}dt + \int_{t_3}^{t_4} \frac{\partial \epsilon}{\partial K}dt + \int_{t_5}^{t_6} \frac{\partial \epsilon}{\partial K}dt \quad . \tag{4.33}$$

For $\epsilon(t)$ given in (4.27), it is easily shown that the first integral in (4.33) vanishes. Using (4.32) and performing the remaining integrations, (4.33) reduces to

$$\sin(2\pi t^*) = 1/2 \quad .$$

This implies that $t^* = 1/12$. ξ^* is found to be $\xi^* = \sqrt{3}/2\, A$. (4.31) then determines K. For (4.23) to be a solution of (4.22), K must equal ω^2. Therefore, the amplitude-frequency relation is

$$\omega^2 = \frac{2}{\sqrt{3}\,A}\tan^{-1}\left(\frac{A\sqrt{3}}{2}\right) \quad . \tag{4.34}$$

MAVE

The remaining minimization condition MAVE is given in (4.19). It is first necessary to determine the members of the set Φ. $\epsilon(t)$ has relative extrema at t satisfying

$$2\pi A \sin(2\pi t)\left[K - \frac{1}{1+A^2\cos^2(2\pi t)}\right] = 0 \quad . \tag{4.35}$$

Roots of (4.35) occur at

$$\left.\begin{aligned} &t_1 = 0 \quad , \quad t_2 = t^* \quad , \quad t_3 = 1/2 - t^* \quad , \\ &t_4 = 1/2 \quad , \quad t_5 = 1/2 + t^* \quad , \quad t_6 = 1 - t^* \quad , \end{aligned}\right\} \tag{4.36}$$

where

$$t^* = \frac{1}{2\pi}\cos^{-1}\left(\frac{(1/K-1)^{1/2}}{A}\right) \quad , \quad \text{for } 0 < t^* < 1/4 \quad .$$

In obtaining (4.36), it is again expected that K will be less than one.

(4.36) comprises the set Φ. However, not all of the points in Φ generate different values of $|\epsilon(t_j)|$. There exists only two distinct maximums, and these are

$$|\epsilon(t_1)| = \left(\frac{2\pi}{\omega}\right)^2 \left|\tan^{-1}(A) - KA\right| \quad , \tag{4.37}$$

and

$$|\epsilon(t_2)| = \left(\frac{2\pi}{\omega}\right)^2 \left|\tan^{-1}\left((1/K-1)^{1/2}\right) - K(1/K-1)^{1/2}\right| \quad . \tag{4.38}$$

K is selected such that

$$\text{Max}\left(|\epsilon(t_1)|, |\epsilon(t_2)|\right) = \text{minimum} \tag{4.39}$$

It can be shown, by considering the behavior of (4.37) and (4.38) as functions of K, that (4.39) is satisfied whenever $|\epsilon(t_1)| = |\epsilon(t_2)|$. Furthermore, for (4.23) to be a solution of (4.22), K must equal ω^2. Therefore, equating (4.37) and (4.38), the amplitude-frequency relation generated using MAVE is

$$A\omega^2 - \tan^{-1}(A) = \tan^{-1}(1/\omega^2 - 1)^{1/2} - \omega^2(1/\omega^2 - 1)^{1/2} \quad . \tag{4.40}$$

Discussion

The three approximate amplitude-frequency relations (4.29), (4.34), and (4.40) are given in Figure 10. It is convenient to plot the variable T^*, defined as

$$T^* = 1/\omega - 1 \quad , \tag{4.41}$$

as a function of initial amplitude A. Also included in the figure is the exact amplitude-frequency relation obtained by numerically integrating (4.20).

It is clear from Figure 10 that ASE gives the closest approximation of the three considered. All of the schemes give amplitude-frequency relations which possess qualitative behavior similar to the exact relation. Since the approximations appear to be diverging for A between 4 and 5, it seems unlikely that AAVE or MAVE would give an approximation better than ASE for some larger value of A.

In addition to providing the best results in this particular example, ASE is the simplest approach to apply. The manipulations necessary to obtain (4.29) involve only a simple integration. Whereas, AAVE and MAVE require the location of the zeros of $\epsilon(t)$ and $\frac{\partial \epsilon(t)}{\partial t}$ respectively. Furthermore, except for a very small number of minimizing parameters, MAVE becomes exceedingly laborious.

4.4 Example 2.

In this section, another conservative autonomous system is considered, namely,

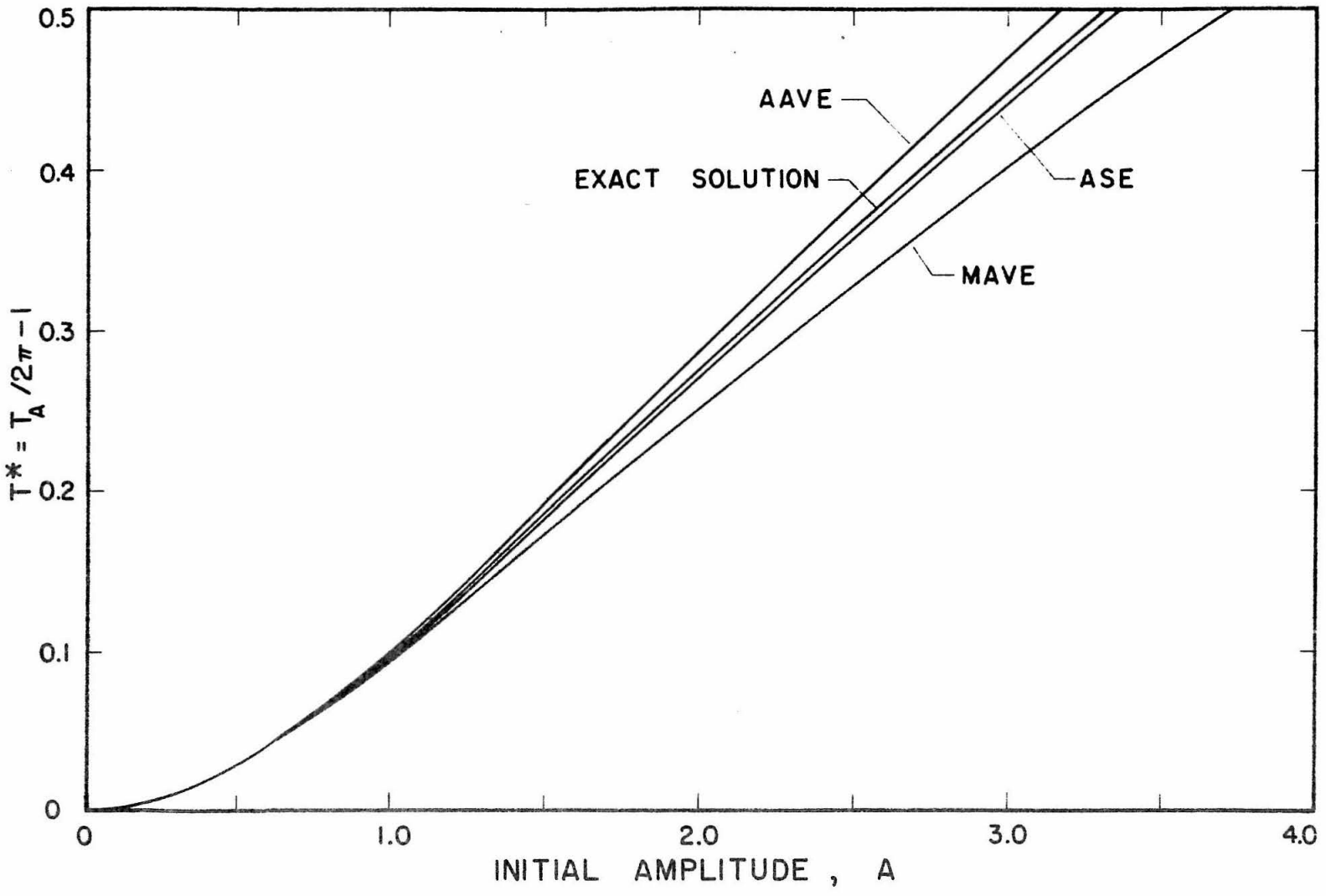

Figure 10: Approximations for $\ddot{x} + \tan^{-1}(x) = 0$

$$\frac{d^2x}{d\tau^2} + a_1x + a_3x^3 + a_5x^5 = 0 \quad , \tag{4.42}$$

where a_1, a_3, and a_5 are constants satisfying $a_1>0$ and $a_5\neq 0$. (4.42) is subject to the initial conditions

$$x(0) = A \quad , \quad \frac{dx(0)}{d\tau} = 0 \quad . \tag{4.43}$$

It is again of interest to determine approximate amplitude-frequency relation for (4.42) for the periodic solution symmetric about x=0. The auxiliary system used is

$$\frac{d^2y}{d\tau^2} + Ky = 0 \quad , \tag{4.44}$$

which possesses periodic solutions satisfying (4.43) of the form

$$y = A\cos(\omega\tau) \quad . \tag{4.45}$$

Again normalizing τ, using $\tau = 2\pi t/\omega$, equations (4.42), (4.44), and (4.45) become

$$\frac{d^2x}{dt^2} + \left(\frac{2\pi}{\omega}\right)^2 a_1x + \left(\frac{2\pi}{\omega}\right)^2 a_3x^3 + \left(\frac{2\pi}{\omega}\right)^2 a_5x^5 = 0 \quad , \tag{4.46}$$

$$\frac{d^2y}{dt^2} + \left(\frac{2\pi}{\omega}\right)^2 y = 0 \quad , \tag{4.47}$$

and

$$y = A\cos(2\pi t) \quad . \tag{4.48}$$

Using (4.3), the differential equation error is

$$\epsilon(t) = \left(\frac{2\pi}{\omega}\right)^2\left[(a_1-K)A\cos(2\pi t) + a_3A^3\cos^3(2\pi t) + a_5A^5\cos^5(2\pi t)\right] \quad . \tag{4.49}$$

ASE, AAVE, MAVE are now used to determine the parameter K, generating approximate amplitude-frequency relations. These

relations are then compared to the exact relation obtained by quadratures.

ASE

Using (4.49), the minimization condition (4.13) reduces to

$$\omega^2 = \frac{2}{A}\int_0^1 \left[a_1 A\cos(2\pi t) + a_3 A^3\cos^3(2\pi t) + a_5 A^5\cos^5(2\pi t)\right]\cos(2\pi t)dt \quad , \tag{4.50}$$

where the fact that K equals ω^2 has been utilized after performing the minimization. The integral is easily evaluated to give

$$\omega^2 = a_1 + \frac{3}{4}a_3 A^2 + \frac{5}{8}a_5 A^4 \quad , \tag{4.51}$$

where ω^2 must be positive so that periodic solutions do exist.

It is convenient to define the following dimensionless variables for the present example. For $a_3 \neq 0$, let

$$\Omega \equiv \frac{\omega^2 - a_1}{a_3^2/a_5} \quad , \tag{4.52}$$

$$\lambda \equiv \frac{a_5 A^2}{a_3} \quad , \tag{4.53}$$

and

$$\mu \equiv \frac{a_3^2}{a_1 a_5} \quad . \tag{4.54}$$

Using these definitions, (4.51) becomes, for $a_3 \neq 0$,

$$\Omega = \frac{3}{4}\lambda + \frac{5}{8}\lambda^2 \quad , \tag{4.55}$$

where

$$\Omega > -\frac{1}{\mu} \quad , \quad \text{if } a_5 > 0$$

and

$$\Omega < -\frac{1}{\mu} \quad , \quad \text{if } a_5 < 0 \quad .$$

If $a_3=0$, it is convenient to define

$$\Omega^* = \frac{\omega^2 - a_1}{a_1} \quad , \tag{4.56}$$

and

$$\lambda^* = \frac{a_5 A^4}{a_1} \quad . \tag{4.57}$$

(4.51) becomes, for $a_3=0$,

$$\Omega^* = \frac{5}{8}\lambda^* \quad , \tag{4.58}$$

where $\Omega^* > -1$ for periodic solutions to exist. Equations (4.55) and (4.58) give the approximate amplitude-frequency relations generated using ASE.

AAVE

To apply AAVE, it is necessary to locate the zeros of $\epsilon(t)$. From (4.49), $\epsilon(t)$ is zero when

$$\cos(2\pi t) = 0$$

or

$$a_1 - K + a_3\xi + a_5\xi^2 = 0 \quad , \tag{4.59}$$

where $\xi = A^2\cos^2(2\pi t)$. The first relation is satisfied for $t=1/4$ and $t=3/4$. The second relation is satisfied for

$$\xi_1 = \frac{-a_3 + \left[a_3^2 - 4a_5(a_1 - K)\right]^{1/2}}{2a_5} \quad ,$$

and (4.60)

$$\xi_2 = \frac{-a_3 - \left[a_3^2 - 4a_5(a_1 - K)\right]^{1/2}}{2a_5} \quad .$$

From the definition given in (4.59), ξ must satisfy

$$0 \le \xi \le A^2 \quad . \tag{4.61}$$

Depending on the specific values of the parameters a_1, a_3, a_5, and A, one or both of the roots in (4.60) may satisfy (4.61). It can be shown that, for $a_5 > 0$, if $A^2 \le -4/7\,(a_3/a_5)$, only ξ_2 satisfies (4.61), if $-4/7\,(a_3/a_5) \le A^2 \le -4/3\,(a_3/a_5)$, both ξ_1 and ξ_2 satisfy (4.61), if $-4/3\,(a_3/a_5) \le A^2$, only ξ_1 satisfies (4.61). For $a_5 < 0$, the above statements are still valid except that ξ_1 and ξ_2 are interchanged. The approximate amplitude-frequency relation has to be determined in parts.

If ξ_1 is the only root satisfying (4.61), the zeros of $\epsilon(t)$ occur at

$$\begin{aligned} &t_1 = t^* \quad , \quad t_2 = 1/4 \quad , \quad t_3 = 1/2 - t^* \quad , \\ &t_4 = 1/2 + t^* \quad , \quad t_5 = 3/4 \quad , \quad t_6 = 1 - t^* \quad , \end{aligned} \tag{4.62}$$

where

$$t^* = \frac{1}{2\pi} \cos^{-1}\left[\frac{-a_3 + \left[a_3^2 - 4a_5(a_1 - K)\right]^{1/2}}{2A^2 a_5}\right]^{1/2} \quad .$$

If ξ_2 is the only root satisfying (4.61), the zeros of $\varepsilon(t)$ occur at

$$t_1 = t^{**} \quad , \quad t_2 = 1/4 \quad , \quad t_3 = 1/2 - t^{**} \quad ,$$
$$t_4 = 1/2 + t^{**} \quad , \quad t_5 = 3/4 \quad , \quad t_6 = 1 - t^{**} \quad , \tag{4.63}$$

where

$$t^{**} = \frac{1}{2\pi} \cos^{-1} \left[\frac{-a_3 - \left[a_3^2 - 4a_5(a_1 - K) \right]^{1/2}}{2A^2 a_5} \right]^{1/2} \quad .$$

If both ξ_1 and ξ_2 satisfy (4.61), the zeros of $\varepsilon(t)$ occur at, for $a_5 > 0$,

$$t_1 = t^* \quad , \quad t_2 = t^{**} \quad , \quad t_3 = 1/4 \quad , \quad t_4 = 1/2 - t^{**} \quad , \quad t_5 = 1/2 - t^* \quad ,$$
$$t_6 = 1/2 + t^* \quad , \quad t_7 = 1/2 + t^{**} \quad , \quad t_8 = 3/4 \quad , \quad t_9 = 1 - t^{**} \quad , \quad t_{10} = 1 - t^* \quad , \tag{4.64}$$

where t^* and t^{**} are defined in (4.62) and (4.63). For $a_5 < 0$, the zeros of $\varepsilon(t)$ are the same as (4.64) except that t^* is replaced by t^{**} and vice versa.

The amplitude-frequency relation can now be obtained for each of the above cases using equation (4.18). Again there is only one minimizing parameter, i.e., K. It is easily shown that

$$\int_0^1 \frac{\partial \varepsilon}{\partial K} dt = 0 \quad ,$$

so that (4.18) reduces to

$$\int_{t_1}^{t_2} \frac{\partial \varepsilon}{\partial K} dt + \int_{t_3}^{t_4} \frac{\partial \varepsilon}{\partial K} dt + \ldots + \int_{t_{N-1}}^{t_N} \frac{\partial \varepsilon}{\partial K} dt = 0 \quad , \tag{4.65}$$

where t_i are given in (4.62), (4.63), and (4.64).

For the case where ξ_1 is the only root satisfying (4.61), (4.65) reduces to

$$\sin(2\pi t^*) = 1/2 \quad ,$$

which implies that $t^* = 1/12$. The corresponding value of ξ_1 is $\xi_1 = 3/4\, A^2$. Substituting ξ_1 into the first of (4.60) the amplitude-frequency relation is

$$\omega^2 = a_1 + \frac{3}{4} a_3 A^2 + \frac{9}{16} a_5 A^4 \quad , \tag{4.66}$$

where the fact that K must equal ω^2 has been used.

For the case where ξ_2 is the only root satisfying (4.61), (4.65) reduces to

$$\sin(2\pi t^{**}) = 1/2 \quad ,$$

implying that $t^{**} = 1/12$. ξ_2 then has the value $\xi_2 = 3/4A^2$. Using the second expression in (4.60) and the fact that K equals ω^2, the amplitude-frequency relation is

$$\omega^2 = a_1 + \frac{3}{4} a_3 A^2 + \frac{9}{16} a_5 A^4 \quad .$$

This is the same relation as (4.66). Therefore, the amplitude-frequency relations for $A^2 \leq -4/7(a_3/a_5)$ and for $A^2 \geq -4/3(a_3/a_5)$ is given by (4.66).

For the case where ξ_1 and ξ_2 both satisfy (4.61), for $a_5 > 0$, (4.65) reduces to

$$\sin(2\pi t^{**}) - \sin(2\pi t^*) = 1/2 \quad . \tag{4.67}$$

For $a_5 < 0$, t^* is replaced by t^{**} and vice versa. Using the definitions of t^* and t^{**} and trigonometric identities, (4.67) can be simplified yielding

$$\omega^2 = a_1 - \frac{a_3^2}{4a_5} + \frac{1}{8}a_3A^2 + \frac{15}{64}a_5A^4 \quad , \tag{4.68}$$

where K has been set equal to ω^2. (4.68) is valid for all non-zero a_5 and is the amplitude-frequency relation for $-\frac{4}{7}(a_3/a_5) \leq A^2 \leq -\frac{4}{3}(a_3/a_5)$. Note that (4.66) and (4.68) are continuous at the boundaries of A^2.

It is convenient to rewrite (4.66) and (4.68) in terms of the dimensionless variables defined in (4.52), (4.53), and (4.54). For $a_3 \neq 0$, the amplitude-frequency relation becomes

$$\Omega = \frac{3}{4}\lambda + \frac{9}{16}\lambda^2 \quad , \quad \text{for } \lambda \leq -\frac{4}{3} \text{ and } \lambda \geq -\frac{4}{7}$$

and (4.69)

$$\Omega = \frac{15}{64}\left(\lambda + \frac{4}{3}\right)\left(\lambda - \frac{4}{5}\right) \quad , \quad \text{for } -\frac{4}{3} \leq \lambda \leq -\frac{4}{7} \quad ,$$

where

$$\Omega > -\frac{1}{\mu} \quad , \quad \text{for } a_5 > 0 \quad ,$$

and

$$\Omega < -\frac{1}{\mu} \quad , \quad \text{if } a_5 < 0 \quad .$$

If a_3 is zero, ξ_1 and ξ_2 can never both satisfy (4.61). Consequently, the amplitude-frequency relation is (4.66) with $a_3 = 0$. Using the variables defined in (4.56), and (4.57), the amplitude-frequency relation for $a_3 = 0$ is

$$\Omega^* = \frac{9}{16}\lambda^* \quad , \tag{4.70}$$

where $\Omega^* > -1$ for a periodic solution to exist. Equations (4.69) and (4.70) are the appropriate expressions generated using AAVE.

MAVE

To apply MAVE, it is first necessary to determine the members of the set Φ. $\epsilon(t)$ has relative extrema at t satisfying

$$2\pi A \sin(2\pi t)\left[K - a_1 - 3a_3\xi + 5a_5\xi^2\right] = 0 \quad , \tag{4.71}$$

where $\xi^2 = A^2\cos^2(2\pi t)$. Two roots of (4.71) are always t=0 and t=1/2. Depending on the particular values of a_1, a_3, a_5, and A, the bracketed term may contain one or two roots in ξ. For a root ξ_i to be valid, it must satisfy

$$0 \le \xi_i \le A^2 \quad . \tag{4.72}$$

The bracketed term in (4.71) vanishes for

$$\xi_1 = -\frac{3a_3}{10a_5} + \left(\frac{9a_3^2}{100a_5^2} + \frac{K - a_1}{5a_5}\right)^{1/2} \quad , \tag{4.73}$$

and

$$\xi_2 = -\frac{3a_3}{10a_5} - \left(\frac{9a_3^2}{100a_5^2} + \frac{K - a_1}{5a_5}\right)^{1/2} \quad . \tag{4.74}$$

The set Φ will consist of the points

$$\begin{aligned} &t_1 = 0 \ , \ t_2 = t^* \ , \ t_3 = t^{**} \ , \ t_4 = 1/2 - t^{**} \\ &t_5 = 1/2 - t^* \ , \ t_6 = 1/2 \ , \ t_7 = 1/2 + t^* \ , \ t_8 = 1/2 + t^{**} \\ &t_9 = 1 - t^{**} \ , \ t_{10} = 1 - t^* \ , \end{aligned} \tag{4.75}$$

where

$$t^* = \frac{1}{2\pi}\cos^{-1}\left(\frac{\xi_1^{1/2}}{A}\right) \text{ and } t^{**} = \frac{1}{2\pi}\cos^{-1}\left(\frac{\xi_2^{1/2}}{A}\right) \quad .$$

Corresponding to the above set of points, there exist only three distinct values of $|\epsilon(t_j)|$, and these may be taken to occur at t_1, t_2 and t_3. From considerations of $|\epsilon(t_j)|$ (j=1, 2, 3) as functions of K, it can be shown that the maximum error is minimized whenever two of the above three errors are equal, and the third error is less than or equal to the two equal errors. Therefore, there exists only three possibilities, and these are

$$|\epsilon(0)| = |\epsilon(t^*)| \quad , \tag{4.76}$$

$$|\epsilon(0)| = |\epsilon(t^{**})| \quad , \tag{4.77}$$

or

$$|\epsilon(t^*)| = |\epsilon(t^{**})| \quad . \tag{4.78}$$

For various values of the parameters a_1, a_3, a_5, and A, one of the above possibilities will hold, and the appropriate K will be determined from that relation.

If a_3/a_5 is positive, ξ_2 will never satisfy (4.72). In this case, the approximation will be given by (4.76). If a_3/a_5 is negative, the situation becomes very complicated. Performing a very lengthy analysis of $|\epsilon(0)|$, $|\epsilon(t^*)|$, and $|\epsilon(t^{**})|$ as functions of K, the following results can be obtained. For $-0.52364 \le a_3A^2/a_5 \le 0$, the approximation is generated from (4.77). For $-0.8 \le a_3A^2/a_5 \le -0.52364$, the approximation is given by (4.78). For $a_3A^2/a_5 \le -0.8$, the approximation is given by (4.76).

Using expression (4.49) for $\epsilon(t)$ and the definitions of t^* and t^{**}, equations (4.76) through (4.78) may be solved for K. If the dimensionless variables defined in (4.52), (4.53), and (4.54) are now employed, the following approximation is obtained, for $a_3 \neq 0$.

$$\left.\begin{aligned} \Omega &= \frac{3}{4}\lambda + \Theta(\lambda)\lambda^2 \quad , \quad \text{for } \lambda \geq -0.52364 \quad , \\ \Omega &= -\frac{1}{5} \quad , \quad \text{for } -0.8 \leq \lambda \leq -0.52364 \quad , \\ \Omega &= \frac{3}{4}\lambda + \Theta(\lambda)\lambda^2 \quad , \quad \text{for } \lambda \leq -0.8 \quad , \end{aligned}\right\} \tag{4.79}$$

where

$$\Omega > -\frac{1}{\mu} \quad , \quad \text{for } a_5 > 0 \quad ,$$

$$\Omega \leq -\frac{1}{\mu} \quad , \quad \text{for } a_5 < 0 \quad .$$

$\Theta(\lambda)$ is a root of the following expression.

$$\frac{1}{4\lambda} + 1 - \Theta = \left[-\frac{3}{10\lambda} + \frac{\lambda}{|\lambda|}\left(\frac{9}{100\lambda^2} + \frac{3}{20\lambda} + \frac{\Theta}{5}\right)^{1/2} \right]^{1/2} *$$

$$\left[\frac{3}{5\lambda} + \frac{4}{5}\Theta + \frac{3}{25\lambda^2} - \frac{2}{5|\lambda|}\left(\frac{9}{100\lambda^2} + \frac{3}{20\lambda} + \frac{\Theta}{5}\right)^{1/2}\right] \quad . \tag{4.80}$$

(4.80) can be reduced to a fifth order polynomial in Θ. The appropriate root has to be real and has to have the proper limiting value since $\Theta(\lambda)$ must be continuous in the range of λ where $\Theta(\lambda)$ is defined. As $\lambda \to \pm\infty$, (4.80) reduces to

$$1 - \Theta = \frac{4}{5}\Theta\left(\frac{\Theta}{5}\right)^{1/4} \quad . \tag{4.81}$$

Θ is easily found to be equal to 0.67355356. It is well to note that $|\lambda|$

going to infinity corresponds to a_3 approaching zero. Therefore, the above value of Θ is the appropriate one to be used in the following special case.

If a_3 vanishes, ξ_1 is the only root satisfying (4.72). Therefore equation (4.76) generates the approximation. Simplifying (4.76) and utilizing the variables defined in (4.56) and (4.57), the approximation, for $a_3=0$, is

$$\Omega^* = 0.67355356\,\lambda^* \quad , \tag{4.82}$$

where $\Omega^* > -1$. Equations (4.79) and (4.82) are the approximate amplitude-frequency relations generated using MAVE. It is impractical to include all of the details involved in obtaining these relations. They are simply too laborious to describe.

Exact Solution

In order to have a basis for comparing the above approximations, the exact solution of (4.42) is developed. (4.42) can be reduced to quadratures using the conditions (4.43). Since the periodic solution, symmetric with respect to x=0, is of interest, the period is

$$T = 4\int_0^A \left(a_1(A^2-z^2)+\frac{a_3}{2}(A^4-z^4)+\frac{a_5}{3}(A^6-z^6)\right)^{-1/2} dz \quad . \tag{4.83}$$

Using the substitution $z = A/\sqrt{x}$, (4.83) reduces to

$$T = 2\left(a_1+\frac{a_3A^2}{2}+\frac{a_5A^4}{3}\right)^{-1/2}\int_1^\infty \left((x-1)(x^2+\alpha x+B)\right)^{-1/2} dx \quad , \tag{4.84}$$

where

$$\alpha = \frac{\frac{a_3A^2}{2}+\frac{a_5A^4}{3}}{a+\frac{a_3A^2}{2}+\frac{a_5A^4}{3}} \quad ; \quad \beta = \frac{\frac{a_5A^4}{3}}{a_1+\frac{a_3A^2}{2}+\frac{a_5A^4}{3}} \quad .$$

The integral in (4.84) can be evaluated in terms of elliptic integrals of the first kind[32]. Denote the roots of the denominator by

$$R_1=1 \quad , \quad R_2=-\frac{\alpha}{2}+\left(\frac{\alpha^2}{4}-\beta\right)^{1/2} \quad , \quad R_3=-\frac{\alpha}{2}-\left(\frac{\alpha^2}{4}-\beta\right)^{1/2} \quad . \tag{4.85}$$

If R_2 and R_3 are real and if $R_2<1$, the value of the integral is found to be

$$\int_1^{\infty}\left((x-1)(x^2+\alpha x+\beta)\right)^{-1/2}dx = \frac{2\overline{K}(\overline{k})}{(1-R_3)^{1/2}} \quad , \tag{4.86}$$

where $\overline{K}(\overline{k})$ is the complete elliptic integral of the first kind with modulus

$$\overline{k}=\left(\frac{R_2-R_3}{1-R_3}\right)^{1/2} \quad .$$

If R_2 and R_3 are real and if $R_2>1$, the value of the integral is complex. This indicates that no solution, symmetric with respect to the origin, exists.

If R_2 and R_3 are complex, the integral is evaluated using the following change of variables[32]. Let

$$y = \frac{(x-m)^2+n^2}{x-1} \quad , \tag{4.87}$$

where $m=-\frac{\alpha}{2}$ and $n=(\beta-\alpha^2/4)^{1/2}$. Define

$$\left.\begin{aligned} y_1 &= -2(m-1)+2\left((m-1)^2+n^2\right)^{1/2} \quad , \\ y_3 &= -2(m-1)-2\left((m-1)^2+n^2\right)^{1/2} \quad , \\ x_1 &= m+\frac{1}{2}y_1 \quad , \end{aligned}\right\} \tag{4.88}$$

and

$$x_3 = m+\frac{1}{2}y_3 \quad .$$

The value of the integral is

$$\int_1^{\infty}\left((x-1)(x^2+\alpha x+\beta)\right)^{-1/2} dx = \frac{\sqrt{2}\,F(\varphi,\bar{k})}{(x_1-x_3)^{1/2}} \quad , \tag{4.89}$$

where $F(\varphi,\bar{k})$ is the incomplete elliptic integral of the first kind with

$$\bar{k} = \left(\frac{-y_3}{y_1-y_3}\right)^{1/2}$$

and

$$\varphi = \cos^{-1}\left(\frac{1-x_1}{1-x_3}\right) \quad .$$

(4.86) and (4.89) give the value of the integral where it exists as a real quantity. Using the dimensionless variables defined in (4.52) through (4.54), the exact solution for $a_3 \neq 0$ may be written as

$$\Omega = \frac{\pi^2}{4I^2}\,\frac{1}{\mu}\left(1-\frac{4I^2}{\pi^2}\right)+\frac{\lambda}{2}+\frac{\lambda^2}{3} \quad , \tag{4.90}$$

where

$$I = \frac{1}{2}\int_1^{\infty}\left((x-1)(x^2+\alpha x+\beta)\right)^{-1/2} dx \quad .$$

It is of interest to note that as μ tends to zero, the exact amplitude-frequency relation becomes identical to the approximation obtained using ASE.

If a_3 equals zero, the exact solution can be written as

$$\Omega^* = \frac{\pi^2}{4I^2}\left(1+\frac{\lambda^*}{2}\right) - 1 \quad , \tag{4.91}$$

where $\Omega^* > -1$ and I is given in (4.90). For the special case of a_3 equals zero, I reduces to

$$I = \begin{cases} \dfrac{1}{\sqrt{2}}\left(\dfrac{2+\lambda^*}{1+\lambda^*}\right)^{1/2} \overline{K}\left(\dfrac{1}{\sqrt{2}}\ \dfrac{\lambda^*}{1+\lambda^*}\right)^{1/2} & , \text{ for } \lambda^* > 0 \\ \dfrac{1}{\sqrt{2}}\left(\dfrac{2+\lambda^*}{1+\lambda^*}\right)^{1/2} \overline{K}\left(\dfrac{\zeta}{(1+\zeta^2)^{1/2}}\right) & , \text{ for } -1 < \lambda^* < 0 \quad , \end{cases} \tag{4.92}$$

where $\overline{K}$ is the complete elliptic integral of the first kind and

$$\zeta = \frac{1}{\sqrt{2}}\left(-\frac{\lambda^*}{1+\lambda^*}\right)^{1/2} \quad .$$

With a_1 positive, no real solution exists for $\lambda^* < -1$.

Equations (4.90) and (4.91) are the exact amplitude-frequency relations for the equation (4.42).

Discussion

The approximate amplitude-frequency relations obtained from the three minimization techniques for a_3 non-zero are plotted on Figure 11. Specifically, equations (4.55), (4.69), and (4.79) are given. It is to be remembered that only certain portions of these curves may be valid depending on the particular value of μ. The value of μ dictates

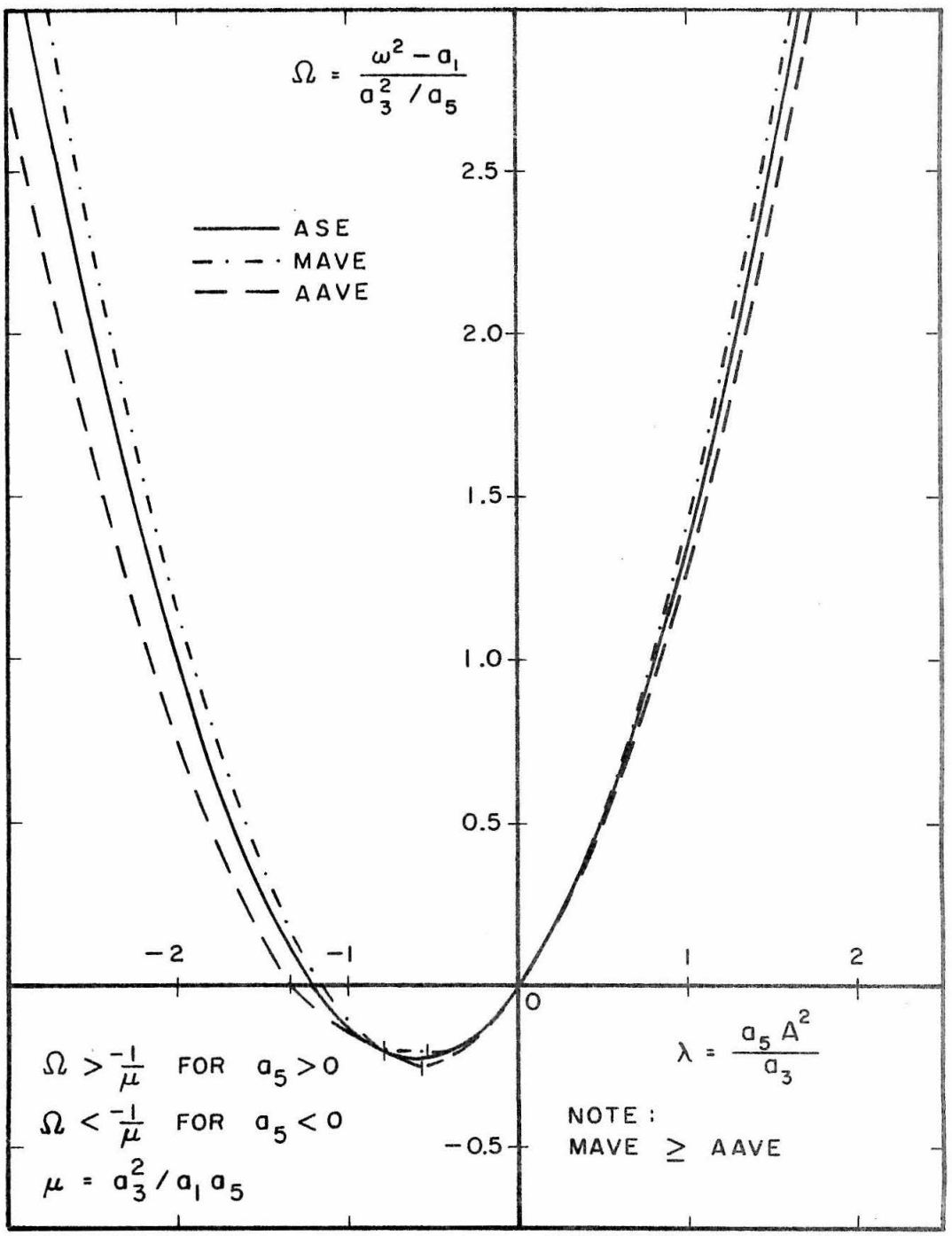

Figure 11: Approximations for $\ddot{x} + a_1 x + a_3 x^3 + a_5 x^5 = 0$

the regions where the approximations yield periodic (i.e., $K>0$) solutions symmetric with respect to $x=0$.

The exact solution is presented in Figure 12 for a_3 non-zero and for various values of μ. Throughout this section, a_1 is assumed positive. This also applies to the exact solution.

One fact which is immediately apparent in Figures 11 and 12 is that the behavior of the exact solution depends on μ whereas the behavior of the approximation does not. The only effect on the approximations is to terminate the curves at various values of Ω to insure that the equivalent spring constant K never becomes negative.

It is difficult to compare the accuracy of the various approximations by considering Figures 11 and 12. Consequently, differences between each approximation and the exact solution ($\Omega_{approx}-\Omega_{exact}$) are given as a function of λ for various μ in Figures 13, 14 and 15. Figure 13 is the error associated with ASE, Figure 14 is the error associated with AAVE, and Figure 15 is the error associated with MAVE. The dashed lines appearing in the figures indicate intervals of λ where the approximate techniques generated periodic solutions, but where, in point of fact, no exact symmetric periodic solutions existed.

By considering Figures 13, 14 and 15, certain qualitative conclusions may be reached. In deciding which minimization scheme seems most appropriate, it is necessary to compare the schemes for all values of λ and μ. Each of the schemes provides better results than the other two for certain specific combinations of λ and μ. However, Figures 13, 14, and 15 indicate that, on the average, the error

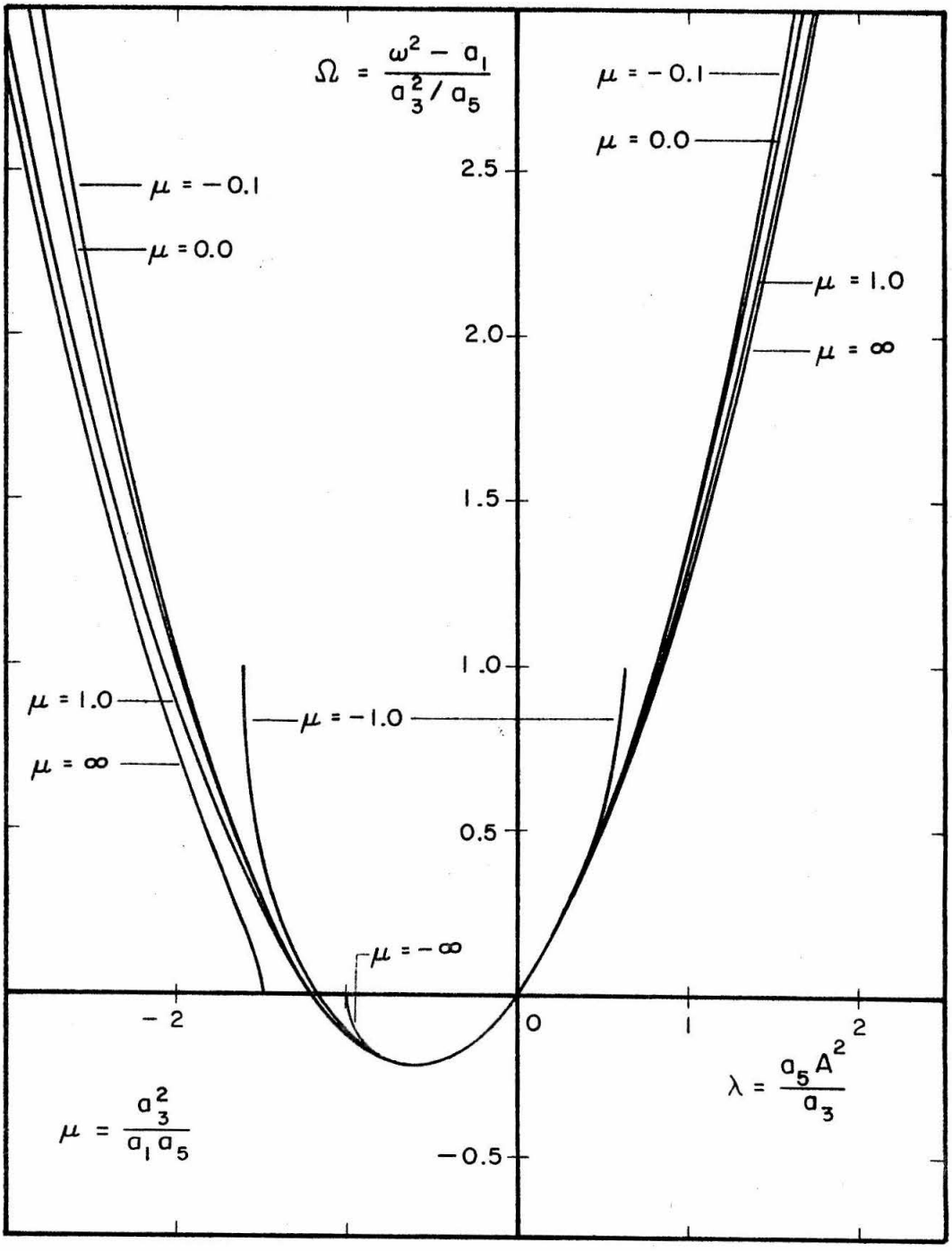

Figure 12: Exact Solution for $\ddot{x} + a_1 x + a_3 x^3 + a_5 x^5 = 0$

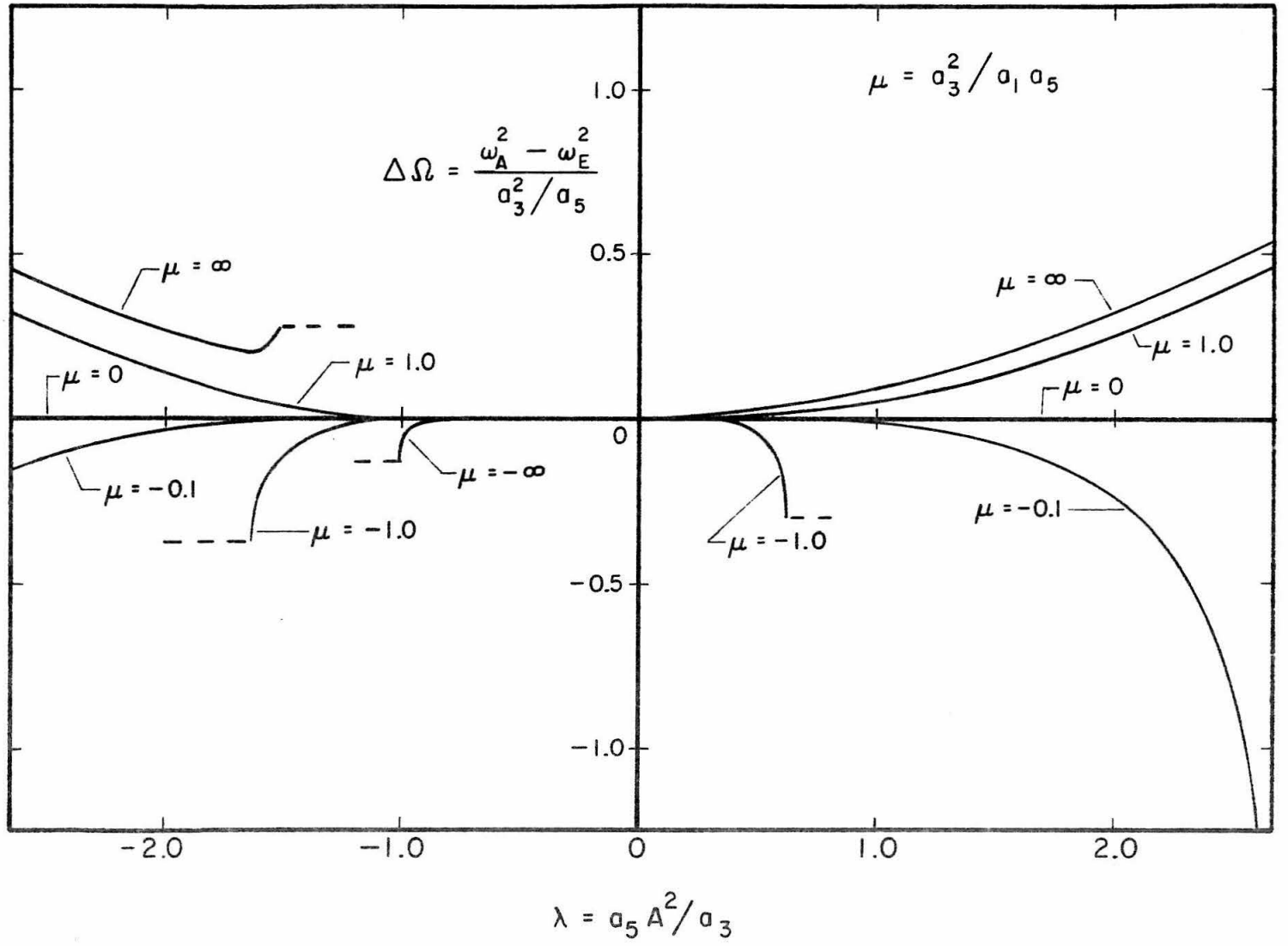

Figure 13: Error with ASE for $\ddot{x} + a_1 x + a_3 x^3 + a_5 x^5 = 0$

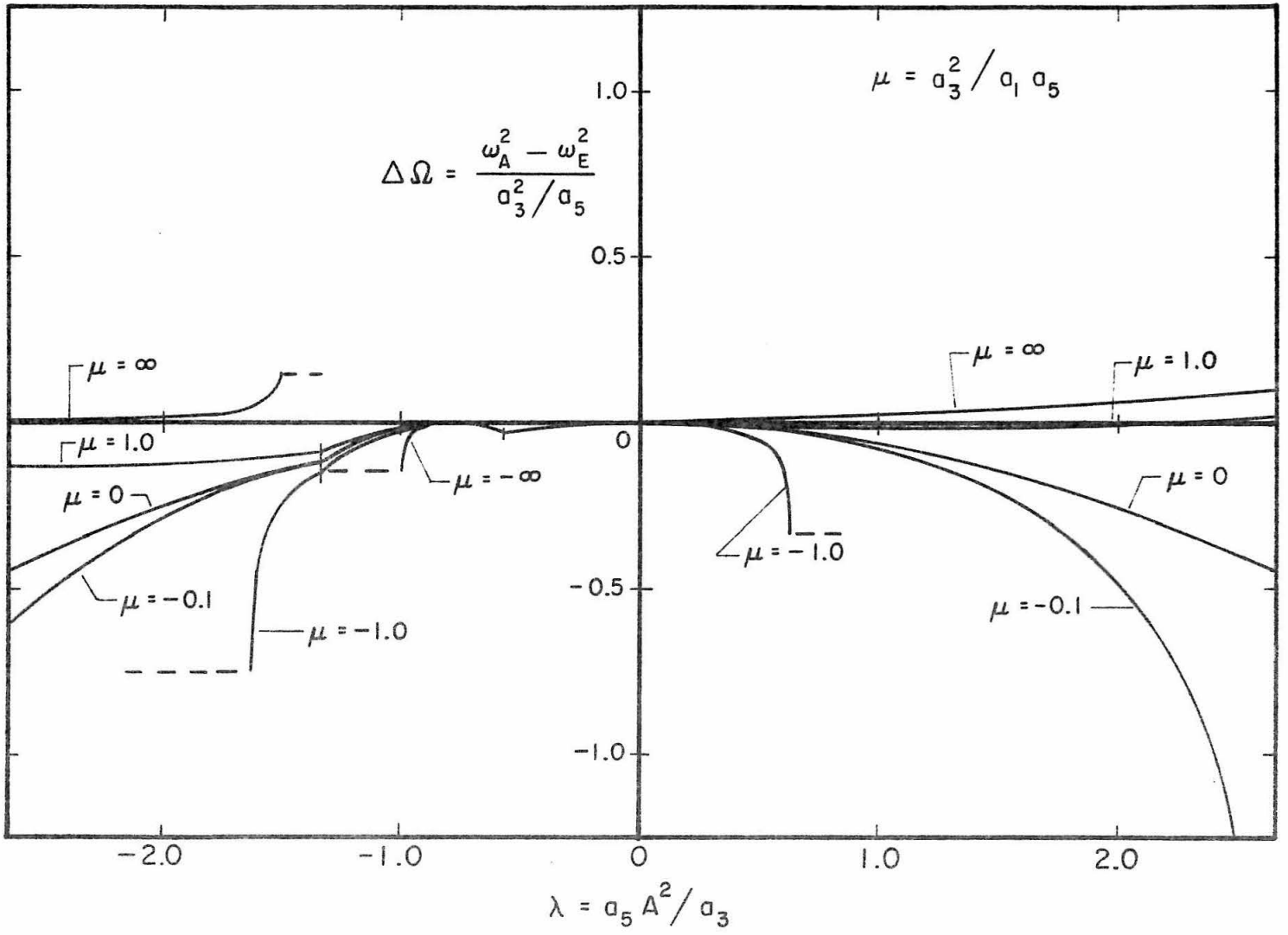

Figure 14: Error with AAVE for $\ddot{x} + a_1 x + a_3 x^3 + a_5 x^5 = 0$

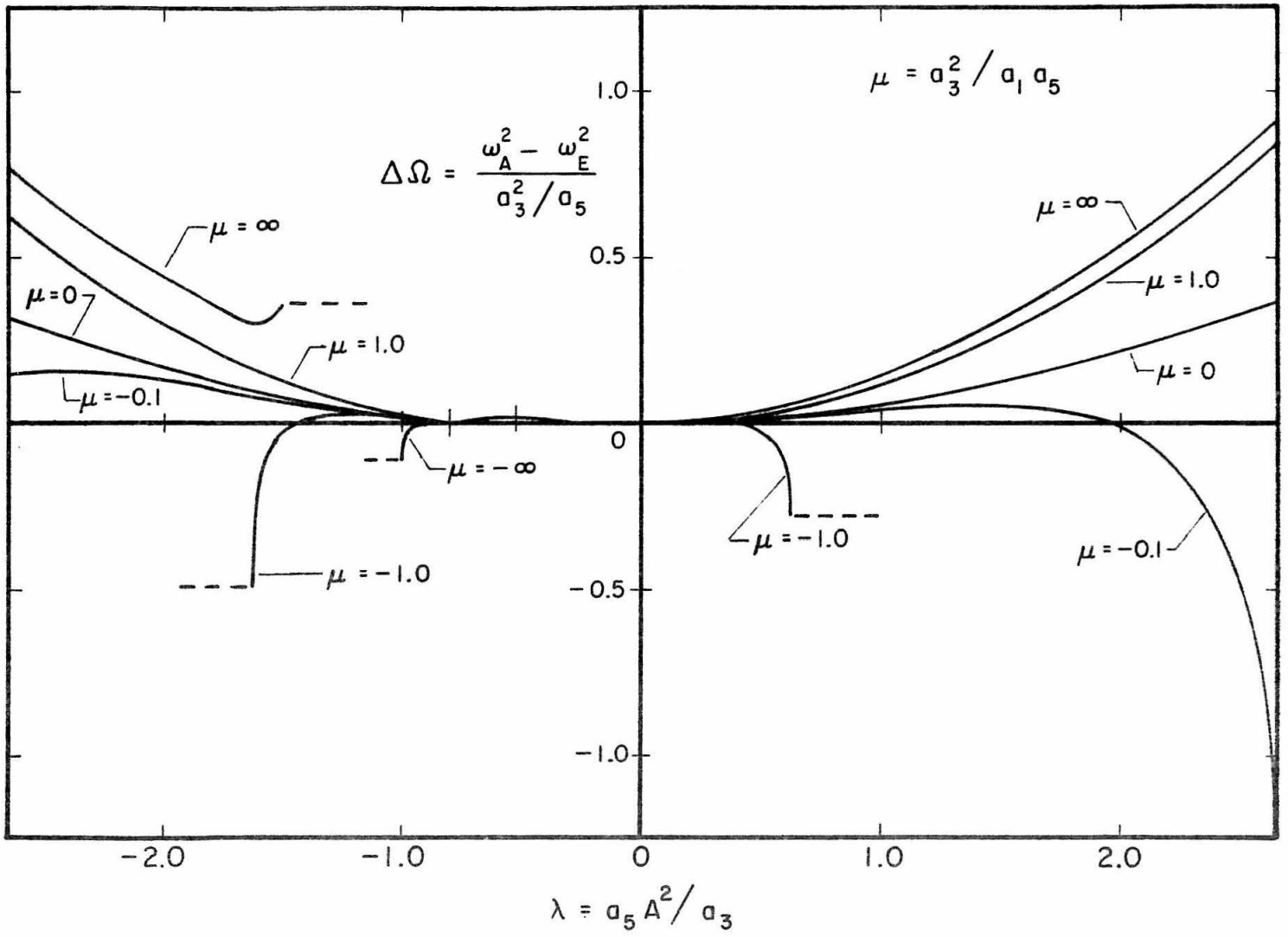

Figure 15: Error with MAVE for $\ddot{x} + a_1 x + a_3 x^3 + a_5 x^5 = 0$

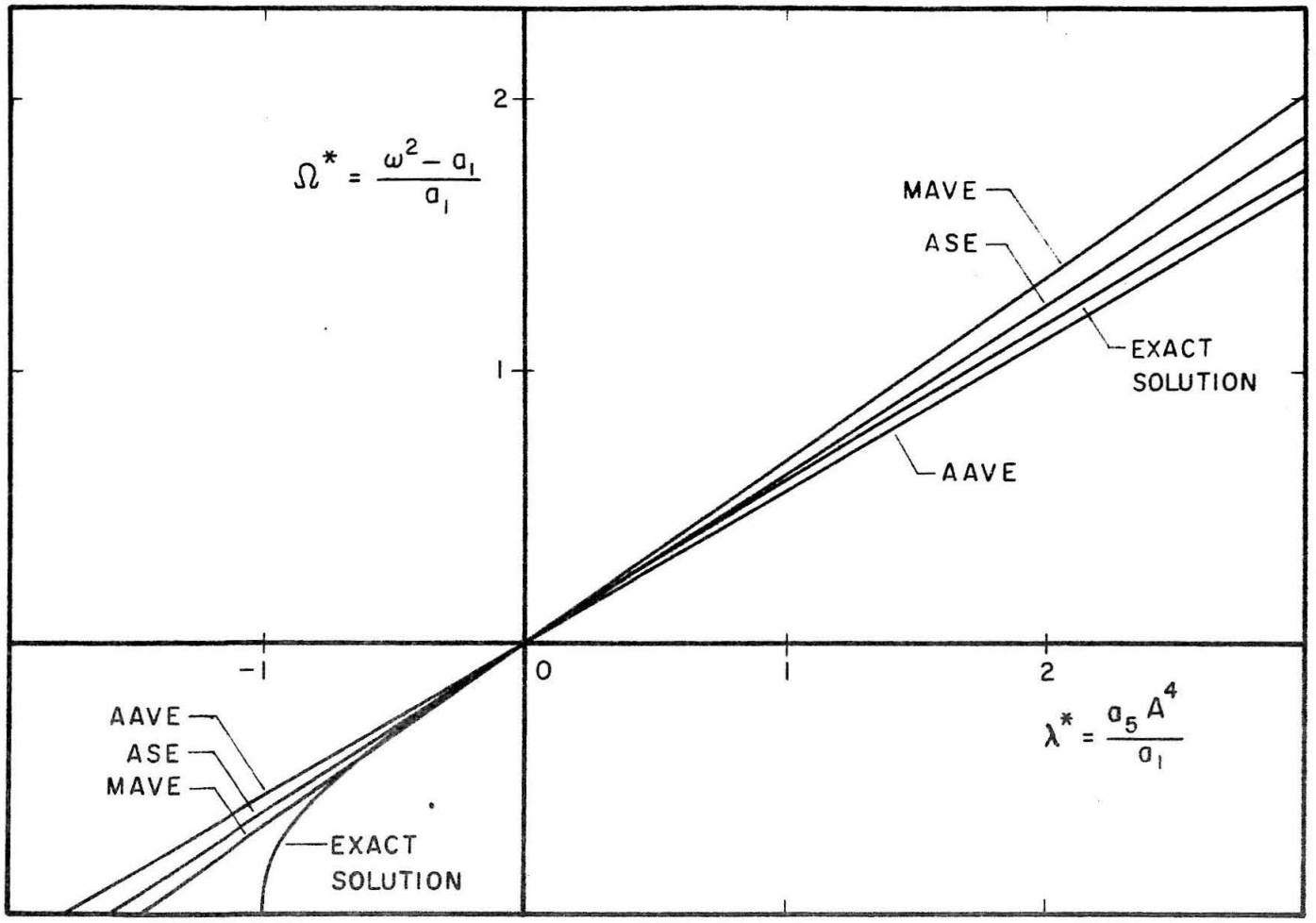

Figure 16: Approximations for $\ddot{x} + a_1 x + a_5 x^5 = 0$

associated with ASE is smaller than the errors associated with AAVE or MAVE. Furthermore, ASE provides an exact solution for the special case of μ equals zero. For no value of μ does AAVE or MAVE provide an exact solution.

Similar conclusions can be reached for the case of a_3 equals zero. Figure 16 illustrates the approximations generated using ASE, AAVE, and MAVE. Also included is the exact amplitude-frequency relation (4.91).

Figure 16 indicates that, for large positive values of λ^*, AAVE seems to provide the best results. However, AAVE is the worst approximation for λ^* negative. Similarly, MAVE seems to give the best results for λ^* less than -1/2 but provides the poorest results when λ^* is positive. ASE gives the best results for $|\lambda^*| < 1/2$. Furthermore, for $|\lambda^*| > 1/2$, ASE lies between AAVE and MAVE. Therefore, if one technique were to be selected as best for all λ^*, it seems that ASE would be the one chosen. In addition, ASE is, by far, the easiest of the three techniques to implement.

4.5 Example 3.

The previous two sections are concerned with conservative autonomous systems. In the present section, a non-conservative, non-autonomous system is considered, namely

$$\ddot{x} + \beta\dot{x} + \alpha x + \mu x^3 = B \cos(2\pi t) \quad , \tag{4.93}$$

where α, β, μ, and B are constants with α, β, and B positive. The symmetric periodic solution with period 1 is of interest.

The equivalent system approach is used with the auxiliary system

$$\ddot{y} + C\dot{y} + Ky = B\cos(2\pi t) \quad , \tag{4.94}$$

where B has the same value as in (4.93) and C and K are to be determined using the various equivalence criteria. (4.94) possesses periodic solutions of the form

$$y = A\cos(2\pi t - \varphi) \quad , \tag{4.95}$$

where A and φ satisfy

$$A = \frac{B}{\left((K-4\pi^2)^2 + 4\pi^2 C^2\right)^{1/2}} \quad \text{and} \quad \tan\varphi = \frac{2\pi C}{K-4\pi^2} \quad . \tag{4.96}$$

Using (4.3), the differential equation error is

$$\epsilon(t) = 2\pi A(C-\beta)\sin(2\pi t-\varphi) + A(\alpha-K)\cos(2\pi t-\varphi) + \mu A^3\cos^3(2\pi t-\varphi) \, . \tag{4.97}$$

Since the periodic solution is of interest, $\epsilon(t)$ is periodic in t. Therefore, a constant can be added to t without affecting the values of C and K. Replacing t by $t+\varphi/2\pi$, $\epsilon(t)$ can be taken to be

$$\epsilon(t) = 2\pi A(C-\beta)\sin(2\pi t) + A(\alpha-K)\cos(2\pi t) + \mu A^3\cos^3(2\pi t) \quad . \tag{4.98}$$

The three minimization schemes are now used to develop the approximate amplitude-frequency relations.

ASE

The general form of the minimization condition generated using ASE is given in (4.13). In the present example there are two minimizing parameters K and C. Minimizing with respect to K first, (4.13) reduces to

$$K = 2\int_0^1 \left[-\beta 2\pi\sin(2\pi t) + \alpha\cos(2\pi t) + \mu A^2\cos^3(2\pi t)\right]\cos(2\pi t)dt \quad .$$

The integrations are easily performed to give

$$K = \alpha + \frac{3}{4}\mu A^2 \quad . \tag{4.99}$$

Minimizing with respect to C, (4.13) reduces to

$$C = -\frac{1}{\pi}\int_0^1 \left[-2\pi\beta \sin(2\pi t) + \alpha \cos(2\pi t) + \mu A^2 \cos^3(2\pi t)\right] \sin(2\pi t) dt \quad .$$

Evaluating the integral gives

$$C = \beta \quad . \tag{4.100}$$

Equations (4.99) and (4.100) give the values of K and C for the approximation obtained using ASE.

AAVE

AAVE requires the location of the zeros of $\epsilon(t)$. From (4.98), $\epsilon(t)$ vanishes whenever,

$$2\pi(C-\beta)(1-u)^{1/2} + (\alpha - K)u^{1/2} + \mu A^2 u^{3/2} = 0 \quad , \tag{4.101}$$

where $u = \cos^2(2\pi t)$. Squaring (4.101) gives the following cubic

$$u^3 + \frac{2(\alpha-K)}{\mu A^2}u^2 + \left(\frac{(\alpha-K)^2 + 4\pi^2(C-\beta)^2}{\mu^2 A^4}\right)u - \frac{4\pi^2(C-\beta)^2}{\mu^2 A^4} = 0 \quad . \tag{4.102}$$

From physical considerations, it is reasonable to expect that the coefficient of the second term in (4.102) will be negative. Therefore, the possibility exists for (4.102) to possess three positive real roots. From the definition of u, a root of (4.102) is physically meaningful only if it satisfies

$$0 \leq u \leq 1 \quad . \tag{4.103}$$

Since C and K are unspecified in (4.102), it is impossible to determine,

at this time, how many of the roots satisfy (4. 103). Therefore, (4. 102) is assumed to possess three roots satisfying (4. 103). From the definition of u, this provides twelve possible values for zeros of $\epsilon(t)$. However, the definition of u involves squaring the cosine function which introduces six extraneous roots. Consequently, there are six true zeros of $\epsilon(t)$. Denoting the roots of (4. 102) by decreasing numerical value as u_1, u_2, and u_3, the zeros of $\epsilon(t)$ are

$$\left.\begin{aligned} &t_1 = t^* \quad ; \quad t_2 = t^{**} \quad ; \quad t_3 = t^{***} \\ &t_4 = \frac{1}{2} + t^* \quad ; \quad t_5 = \frac{1}{2} + t^{**} \quad ; \quad t_6 = \frac{1}{2} + t^{***} \quad , \end{aligned}\right\} \tag{4. 104}$$

where $t^* = \cos^{-1}\left(u_1^{1/2}\right)$, $t^{**} = \cos^{-1}\left(u_2^{1/2}\right)$, and $t^{***} = \cos^{-1}\left(u_3^{1/2}\right)$.

Using (4. 104), the minimization condition (4. 18) may now be employed to determine K. It is easily shown that $\int_0^1 \frac{\partial \epsilon(t)}{\partial K} dt$ vanishes. Therefore, performing the remaining integrations, (4. 18) reduces to

$$\cos(2\pi t^{**}) - \cos(2\pi t^*) - \cos(2\pi t^{***}) = 0 \quad . \tag{4. 105}$$

The condition minimizing C may be determined in a similar manner. It is clear that $\int_0^1 \frac{\partial \epsilon(t)}{\partial C} dt$ vanishes. Performing the remaining integrations, (4. 18) reduces to

$$\sin(2\pi t^{**}) - \sin(2\pi t^*) - \sin(2\pi t^{***}) = 0 \quad . \tag{4. 106}$$

Equations (4. 105) and (4. 106) determine the appropriate K and C.

To determine K and C specifically, it is convenient to reformulate the problem. Using the definition of u, (4. 105) and (4. 106) can be written as

$$\cos(2\pi t^{**}) = u_1^{1/2} - u_3^{1/2} \quad , \tag{4. 107}$$

and

$$\sin(2\pi t^{**}) = (1-u_1)^{1/2} + (1-u_3)^{1/2} \quad , \tag{4.108}$$

where the appropriate sign of the cosine function has been used. From the theory of cubic equations, the three roots u_1, u_2, and u_3 must satisfy

$$u_1 + u_2 + u_3 = -\frac{2(\alpha-K)}{\mu A^2} \quad , \tag{4.109}$$

$$u_1u_2 + u_2u_3 + u_1u_3 = \frac{(\alpha-K)^2 + 4\pi^2(C-\beta)^2}{\mu^2 A^4} \quad , \tag{4.110}$$

and

$$u_1u_2u_3 = \frac{4\pi^2(C-\beta)^2}{\mu^2 A^4} \quad . \tag{4.111}$$

Equations (4.107) through (4.111) and the definition of t^{**} in (4.104) are six relations for determining t^{**}, u_1, u_2, u_3, K, and C. Unfortunately, these relations are highly transcendental and cannot be solved by direct elimination of variables.

Assuming that the above set of equations possesses a unique solution, it is possible to determine K and C by a fortuitous guess. The solution is found to be

$$t^{**} = 0 \ , \quad u_1 = u_3 = \frac{3}{4} = \frac{K-\alpha}{\mu A^2} \ , \quad \text{and } C = \beta \quad . \tag{4.112}$$

Solution (4.112) can be verified by direct substitution into (4.107) through (4.111) and using the definition of t^{**}. Solving (4.112) for K and C yields

$$C = \beta \quad , \tag{4.113}$$

and

$$K = \alpha + \frac{3}{4}\mu A^2 \quad . \tag{4.114}$$

Equations (4.113) and (4.114) are identical to equations (4.99) and (4.100). Therefore, the approximations generated by ASE and AAVE are identical. The only difference is in the amount of labor involved in obtaining the approximation, ASE being much simpler.

MAVE

The minimization condition generated by MAVE involves the extrema of $|\epsilon(t)|$. In the present example there is a much more direct way of obtaining the approximation rather than using equation (4.19).

MAVE minimizes the maximum of the absolute value of $\epsilon(t)$ for $0 \le t \le 1$. Using (4.98) and trigonometric identities, $\epsilon(t)$ can be written as

$$\left|\frac{\epsilon(t)}{\mu A^3}\right| = \left|R(K_1, K_2)\cos(\theta - \varphi) + \frac{1}{4}\cos(3\theta)\right| \quad , \tag{4.115}$$

where

$$\left.\begin{aligned} R(K_1, K_2) &= \left(9K_1^2 + \left(\frac{3}{4} - 3K_2\right)^2\right)^{1/2} \quad , \\ \theta &= 2\pi t \\ \varphi &= \tan^{-1}\left(\frac{K_1}{1/4 - K_2}\right) \quad , \\ K_1 &= \frac{2\pi(C-\beta)}{3\mu A^2} \quad , \end{aligned}\right\} \tag{4.116}$$

and

$$K_2 = \frac{K-\alpha}{3\mu A^2} \qquad \Big\} \qquad (4.116)\ \text{cont.}$$

Using (4.115), it is desirable to select $R(K_1, K_2)$ and φ so that the maximum of $|\epsilon(t)/\mu A^3|$ is as small as possible.

The appropriate values of R and φ are R equals zero and φ arbitrary. To see this, consider the following argument. (4.115) will always possess a component $\cos(3\theta)$. $\cos(3\theta)$ possesses three sign changes and four extrema of equal magnitude in one half a cycle. Assume that there exists a non-zero R and some φ such that all of the extrema of $1/4\cos(3\theta)$ are reduced. For this to be true, $\cos(\theta-\varphi)$ must have at least as many sign changes in one half a cycle as $\cos(3\theta)$. This is a contradiction since $\cos(\theta-\varphi)$ can have at most one sign change in one half a cycle. Therefore, for a non-zero R, at least one of the extrema of $\cos(3\theta)$ will be increased in absolute value. Therefore, the value of R which gives the minimum $|\epsilon(t)/\mu A^3|$ is R equal zero. From (4.116), R equals zero implies K_1 equals zero and K_2 equals 1/4. K_1 vanishing implies that

$$C = \beta \quad . \qquad (4.117)$$

K_2 equals 1/4 implies that

$$K = \alpha + \frac{3}{4}\mu A^2 \quad . \qquad (4.118)$$

Hence, the approximation generated using MAVE is identical to the one generated by both ASE and AAVE.

Discussion

The present example is a very special case where all three minimization schemes generate the same approximation. It is, therefore, meaningless to compare accuracy. The only basis for comparison is the effort involved in obtaining each approximation.

Again the minimization condition obtained using ASE is reduced to simple integrations. However, AAVE generates a very complicated and highly transcendental set of equations for determining the approximation, which ultimately is solved by a lucky guess. Also, the uniqueness of the solution must be assumed, which is in general not a justifiable assumption. MAVE leads to a very simple derivation of the approximation in the present example. However, if the original statement of MAVE were used, the amount of labor involved would be increased. Even when the simple derivation is employed, the amount of labor involved is still comparable to that to that involved in using ASE.

4.6 Example 4.

The system of interest in the present section is the one considered in Chapter II. This particular system was selected in order to study the effect of the various minimization schemes whenever a nonlinear auxiliary system is used. The original system is equation (2.17) and is

$$\frac{d^2x}{d\tau^2} + ax + bx^3 = B\cos(\omega\tau) \quad , \tag{4.119}$$

where a, b, B and ω are constants. The steady-state periodic solution

is of interest. The auxiliary system to be used is equation (2.21) which is

$$\frac{d^2y}{d\tau^2} + ay + by^3 = \alpha\, cn(\eta\tau, k) \quad , \tag{4.120}$$

where a and b have the same values as in (4.119), and $cn(\eta\tau, k)$ is the Jacobian elliptic cosine function.

In order that (4.120) have the same period as (4.119), η must satisfy (2.19), i.e.,

$$\eta = \frac{2K(k)\omega}{\pi} \quad . \tag{4.121}$$

(4.120) has known periodic solution of the form

$$y = \beta\, cn(\eta\tau, k) \quad , \tag{4.122}$$

where A, η, and k must satisfy

$$b\beta^3 + (1-\eta^2)B = \alpha \quad , \tag{4.123}$$

and

$$k^2 = \frac{b\beta^2}{2\eta^2} \quad . \tag{4.124}$$

It is convenient to normalize the independent variable τ before determining the differential equation error so that the motion is periodic with period 2π. Using $\tau = 2\pi t/\omega$, the differential equation error is

$$\epsilon(t) = B\cos(2\pi t) - \alpha\, cn(4K(k)t, k) \quad , \tag{4.125}$$

where equation (4.121) has been used. The various minimization schemes are now employed to obtain the approximations. There is only one minimizing parameter, i.e., α.

ASE

The approximation is obtained using ASE in Section 2.2. Consequently, the results can be used directly. From Section 2.2, equation (2.26) is the approximation obtained. The system to be solved for b positive is equation (2.19), (2.23), and (2.27). For b negative, the equations are (2.28), (2.19)′, and (2.23)′ (appearing after (2.28)). The above system can be solved directly by assuming values for k and calculating corresponding values of ω.

AAVE

To determine α using AAVE, it is first necessary to locate the zeros of $\epsilon(t)$. From (4.125), $\epsilon(t)$ vanishes whenever

$$\cos(2\pi t) = \frac{\alpha}{B} \operatorname{cn}(4K(k)t, k) \quad . \tag{4.126}$$

(4.126) is satisfied for the following values of t,

$$\left.\begin{aligned} &t_1 = t^* \quad , \quad t_2 = \frac{1}{4} \quad , \quad t_3 = \frac{1}{2} - t^* \\ &t_4 = \frac{1}{2} + t^* \quad ; \quad t_5 = \frac{3}{4} \quad ; \quad t_6 = 1 - t^* \quad , \end{aligned}\right\} \tag{4.127}$$

where t^* is the root of (4.126) satisfying $0 < t^* < \frac{1}{4}$. It is not necessary to determine t^* explicitly in order to obtain the approximation.

The minimization condition (4.18) can now be evaluated. Using reference (28), it can be shown that $\int_0^1 \frac{\partial \epsilon(t)}{\partial \alpha} dt$ vanishes. Therefore, (4.18) reduces to

$$\int_{t_1}^{t_2} \frac{\partial \epsilon}{\partial \alpha} dt + \int_{t_3}^{t_4} \frac{\partial \epsilon}{\partial \alpha} dt + \int_{t_5}^{t_6} \frac{\partial \epsilon}{\partial \alpha} dt = 0 \quad . \tag{4.128}$$

Using (4. 127) and reference (28), (4. 128) can be evaluated to give

$$\mathrm{sn}\,(4K(k)t^*, k) = \frac{1}{k}\sin\left(\frac{\sin^{-1}(k)}{2}\right) \quad , \tag{4. 129}$$

where sn(u, k) is the Jacobian elliptic sine function. Equations (4. 121), (4. 123), (4. 124), (4. 126), and (4. 129) are sufficient to determine the five unknowns η, k, α, A, and t^*.

If ω is considered unknown and k as known, the above equations can be combined to give the following set for determining the approximation

$$t^* = \frac{F(\varphi, k)}{4K(k)} \quad , \tag{4. 130}$$

where F is the incomplete elliptic integral of the first kind and

$$\varphi = \sin^{-1}\left[\frac{1}{k}\sin\left(\frac{\sin^{-1}(k)}{2}\right)\right] \quad ;$$

$$\xi\left(4K(k)\right)^2(2k^2-1)\omega^3 + \xi a 4\pi^2\omega = \frac{4\pi^2 B\cos(2\pi t^*)}{\left(1-\frac{1}{k^2}\sin^2\left(\frac{\sin^{-1}(k)}{2}\right)\right)^{1/2}} \quad , \tag{4. 131}$$

where

$$\xi = \frac{2kK(k)}{\pi}\left(\frac{2}{b}\right)^{1/2} \quad ;$$

$$\beta = \xi\omega \quad ; \tag{4. 132}$$

$$\eta = \frac{2K(k)\omega}{\pi} \tag{4. 133}$$

and

$$\alpha = \frac{B \cos(2\pi t^*)}{\left(1 - \frac{1}{k^2}\sin^2\left(\frac{\sin^{-1}(k)}{2}\right)\right)^{1/2}} \quad . \tag{4.134}$$

The above equations are convenient only for b positive. The solution procedure is to consider a, b, B, and k as known and use the above equations to calculate t^*, ω, β, η, and α. For b negative, k is pure imaginary, and the above equations must be modified prior to performing any numerical calculations.

MAVE

The approximation generated using MAVE is obtained by first determining the set Φ. From (4.125), the extrema of $\epsilon(t)$ occur whenever

$$\sin(2\pi t) = \frac{2K(k)\alpha}{\pi\beta} \operatorname{sn}(4K(k)t, k) \operatorname{dn}(4K(k)t, k) \quad , \tag{4.135}$$

where sn(u, k) and dn(u, k) are Jacobian elliptic functions. Equation (4.135) is satisfied for

$$\left.\begin{aligned} &t_1 = 0, \quad , \quad t_3 = \frac{1}{2} \quad , \quad t_3 = 1 \quad , \quad t_4 = t^* \\ &t_5 = \frac{1}{2} - t^* \quad , \quad t_6 = \frac{1}{2} + t^* \quad , \quad t_7 = 1 - t^* \quad , \end{aligned}\right\} \tag{4.136}$$

where t^* is the root of (4.135) satisfying $0 < t^* < 1/4$. For the set of points in (4.136), there exist only two distinct values of $|\epsilon(t)|$, and these may be taken to occur at $t = 0$ and $t = t^*$.

Applying (4.19), the approximation is generated by

$$\max\left(|\epsilon(0)|, |\epsilon(t^*)|\right) = \text{minimum} \quad . \tag{4.137}$$

By considering the properties of $|\epsilon(0)|$ and $|\epsilon(t^*)|$ as functions of α, it

may be shown that (4. 137) is satisfied whenever the two errors are equal, i. e.,

$$|\epsilon(0)| = |\epsilon(t^*)| \quad . \tag{4. 138}$$

Using (4. 125) and (4. 135), (4. 138) reduces to

$$\alpha = B \frac{1 + \cos(2\pi t^*)}{1 + \mathrm{cn}(4K(k)t^*, k)} \quad . \tag{4. 139}$$

(4. 139) represents the approximation generated using MAVE. Equations (4. 121), (4. 123), (4. 124), (4. 135), and (4. 139) are sufficient for determining t^*, k, β, α, and η. Combining the above equations, the following set of equations may be obtained.

$$\frac{\pi}{2K(k)} \tan(\pi t^*) = \frac{\mathrm{sn}(4K(k)t^*, k)\,\mathrm{dn}(4K(k)t^*, k)}{1 + \mathrm{cn}(4K(k)t^*, k)} \quad , \tag{4. 140}$$

where t^* satisfies $0 < t^* < 1/4$;

$$\xi\left(4K(k)\right)^2 (2k^2 - 1)\omega^3 + \xi a 4\pi^2 \omega = 4\pi^2 B \frac{1 + \cos(2\pi t^*)}{1 + \mathrm{cn}(4K(k)t^*, k)} \quad ; \tag{4. 141}$$

where

$$\xi = \frac{2kK(k)}{\pi} \left(\frac{2}{b}\right)^{1/2} \quad ;$$

$$\beta = \xi\omega \quad ; \tag{4. 142}$$

$$\eta = \frac{2K(k)\omega}{\pi} \quad ; \tag{4. 143}$$

and

$$\alpha = B \frac{1 + \cos(2\pi t^*)}{1 + \mathrm{cn}(4K(k)t^*, k)} \quad . \tag{4. 144}$$

Equations (4. 140) through (4. 144) determine t^*, k, β, α, and η. Unlike the previous approximations, considering k as known does not enable the above approximation to be obtained by direct substitution. (4. 140)

is a transcendental equation for t^*. Consequently, the approximation generated using MAVE must be determined numerically. The above set of equations is convenient only for b positive. For b negative, they must again be modified.

Discussion

In order to compare the accuracy of the various approximations, it is necessary to consider specific numerical examples. Because of the large number of parameters, any comparison involving a variation in all of them would become too lengthy. Consequently, certain parameters were arbitrarily fixed. Specifically, a is chosen to be 1, and ω is chosen to be 0.6. Furthermore, b is restricted to be negative. Since b is negative, the response curve will lean to the left, and, for the value of ω chosen, there is a possiblity of the system possessing three periodic solutions. The comparisons will be based on the amplitudes associated with the upper branch of the response curve. To provide a basis for comparison, equation (4.119) was integrated numerically to obtain the exact steady-state amplitude.

Comparisons were made for various b and B. However, only the case which exhibited the largest differences will be presented. This occurred for B equal 1.0 and for b varying from -0.01 to -10.0. Figure 17 shows the normalized error in steady-state amplitude for the various approximations as a function of b. All of the other comparisons for different values of B possess the same qualitative behavior.

Figure 17 indicates that the best numerical accuracy is obtained using AAVE. However, AAVE and ASE are so close that a question arises whether the slight increase in accuracy provided by AAVE compensates for the additional labor involved in its use. The maximum difference between ASE and AAVE is on the order of 10^{-3}.

4.7 Conclusions.

In the previous four sections, certain examples are presented where approximations are obtained using three specific equivalence criteria, ASE, AAVE, and MAVE, described in Section 4.2. In each of the examples it may be concluded that ASE is, in some sense, the most appropriate equivalence criterion to use.

In each example, ASE is the easiest technique to implement. It involves only simple integrations, while both AAVE and MAVE require the location of the zeros of certain functions of t. This location can become rather involved as evidenced in Sections 4.4 and 4.6. Furthermore, MAVE usually requires some analysis of the functional behavior of certain expressions which can be quite tedious (cf. Section 4.4).

The most useful and desirable aspect of most approximating techniques is that they are much easier to use than are exact solution techniques. However, this advantage is essentially nullified in the case of AAVE and MAVE, and for this reason alone it would seem justified to label AAVE and MAVE as impractical unless they provided a substantial increase in accuracy over ASE.

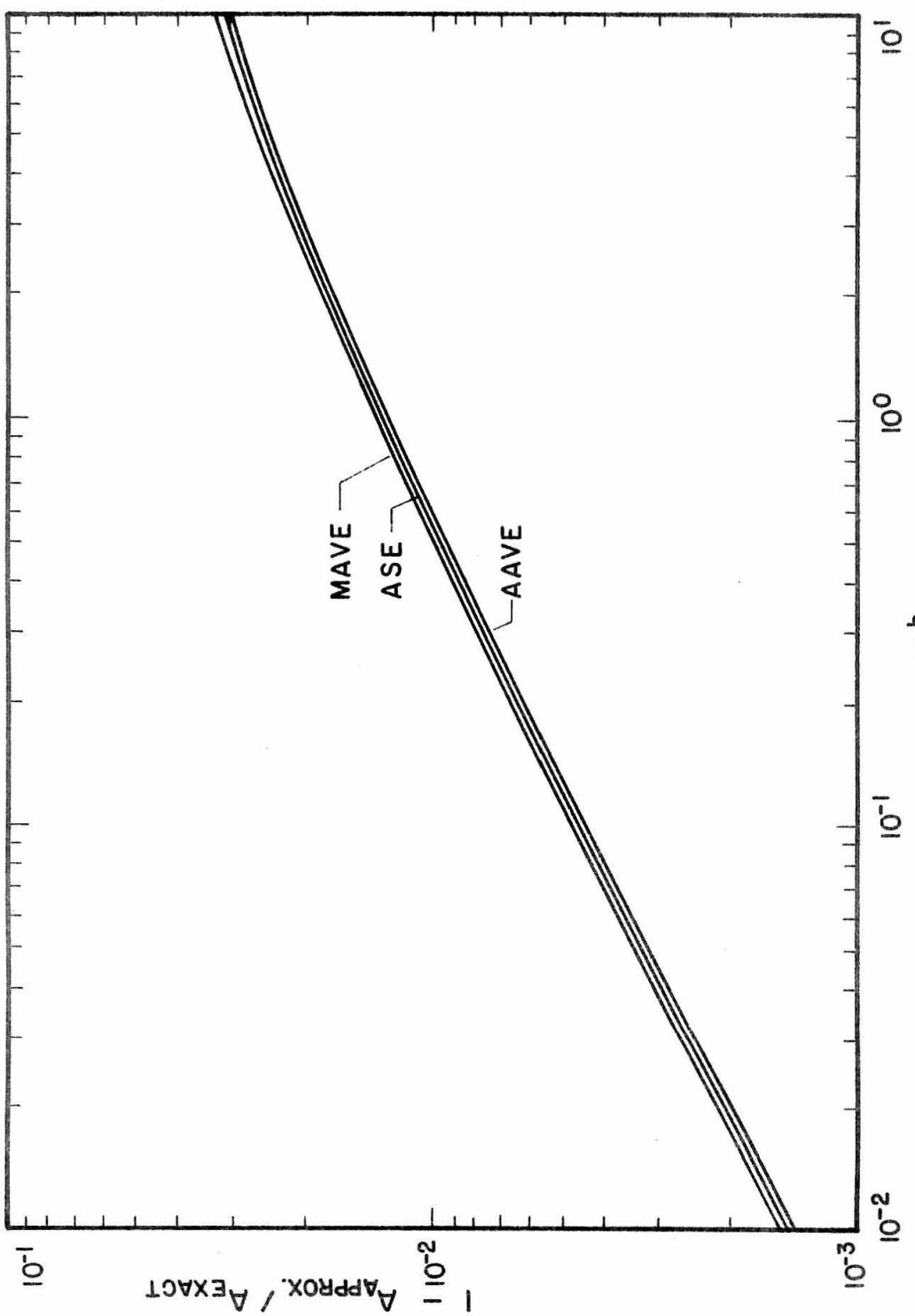

Figure 17: Approximations for $\ddot{x} + x + bx^3 = 1.0 \cos(0.6t)$

The examples did not show this to be the case. In every example, except Section 4.6, ASE provided as good or better results, on the average, than either AAVE or MAVE. In the one case where AAVE does provide better results, the increase in accuracy is on the order of 0.1 percent. This increase hardly compensates for the additional effort required by the AAVE technique.

It must be pointed out that the above conclusions are based on a very limited analysis. It is sometimes dangerous to draw conclusions based on certain specific examples, but an attempt was made to make the examples representative. Both autonomous and nonautonomous systems were considered, and both linear and cubic auxiliary systems are employed.

It is also realized that there exist many possible equivalence criteria other than the three considered in the present analysis. It would be impossible to compare all of them. ASE, AAVE, and MAVE were chosen for their physical and mathematical relevance in addition to their connection with the error bound as indicated in Section 4.1.

V. A COMPARISON OF LINEAR AND CUBIC APPROXIMATIONS FOR SECOND ORDER SCALAR SYSTEMS

In the previous chapter, two auxiliary systems are used for obtaining approximate periodic solutions for some specific second order scalar equations. These two systems are: 1) the linear system (4.94) and 2) the cubic system (4.120). In the present chapter, the above two systems are compared for some further examples. It is shown that the cubic approximation is potentially more accurate than the standard linear system in predicting steady-state response of nonlinear systems.

Section 5.1 presents the linear approximation to a general second order scalar system. The approximation is obtained using the equivalent equation approach. Section 5.2 deals with the cubic approximation to a general second order scalar system. The cubic approximation is also obtained using the equivalent equation approach. In Chapter IV some arguments are presented which indicate that minimizing the mean square differential equation error is the most appropriate equivalence criterion to utilize. Consequently, the above two approximations are obtained by minimizing the mean square error (4.13). In Sections 5.3 and 5.4 certain examples are considered, and the linear and cubic approximations are compared.

5.1. General Linear Approximation.

Let the system of interest be written as

$$\ddot{x} + f(x, \dot{x}, t) = F(t) \tag{5.1}$$

where $F(t)$ is continuous in t and periodic in t with period 1. Furthermore, let $f(x, \dot{x}, t)$ be continuous in x, $\dot{x}$, and t, and periodic in t with period 1. Assuming that (5.1) possesses periodic solutions with period 1, the equivalent equation approach may be employed to obtain an approximation.

In the present section, the auxiliary system is taken to be the linear equation

$$\ddot{y} + c\dot{y} + Ky = B \cos(2\pi t) \quad , \tag{5.2}$$

where the period of excitation has been taken to be the same as the period of the original system. (5.2) possesses exact periodic solutions of the form

$$y = A \cos(2\pi t - \varphi) \quad , \tag{5.3}$$

where A and φ must satisfy

$$A = \frac{B}{\left((K - 4\pi^2)^2 + 4\pi^2 c^2\right)^{1/2}} \quad , \tag{5.4}$$

and

$$\varphi = \tan^{-1}\left(\frac{2\pi c}{K - 4\pi^2}\right) \quad . \tag{5.5}$$

In Chapter II, the equivalent equation approach is described for two general equations. It is stated that certain of the differential equation parameters may be selected arbitrarily so that portions of the auxiliary system would be similar to the original system. However, here, it is desirable to leave all of the parameters (c, K, and B) in (5.2) unspecified, and to determine them using the minimization condition.

Using (2.7), the differential equation error is

$$\epsilon(t) = F(t) - B\cos(2\pi t) + c\dot{y} + Ky - f(y, \dot{y}, t) \quad . \tag{5.6}$$

c, K, and B may be determined by minimizing

$$\int_0^1 \epsilon^2(t)\, dt \quad . \tag{5.7}$$

Minimizing (5.7) with respect to c, substituting (5.3) for the approximate solution y, and evaluating the resulting trigonometric integrals, the following result is easily obtained.

$$B\sin\varphi - 2\pi cA + 2\int_0^1 F(t)\sin(2\pi t - \varphi)\, dt + \\ - 2\int_0^1 f(y, \dot{y}, t)\sin(2\pi t - \varphi)\, dt = 0 \quad , \tag{5.8}$$

where $y = A\cos(2\pi t - \varphi)$. A second relation is generated by minimizing (5.7) with respect to K. The result is

$$KA - B\cos\varphi + 2\int_0^1 F(t)\cos(2\pi t - \varphi)\, dt \\ - 2\int_0^1 f(y, \dot{y}, t)\cos(2\pi t - \varphi)\, dt = 0 \quad , \tag{5.9}$$

where $y = A\cos(2\pi t - \varphi)$.

If (5.7) is now minimized with respect to B, the relation generated is

$$2\pi Ac\sin\varphi + KA\cos\varphi - B + 2\int_0^1 F(t)\cos(2\pi t)\, dt \\ - 2\int_0^1 f(y, \dot{y}, t)\cos 2\pi t\, dt = 0, \tag{5.10}$$

where $y = A\cos(2\pi t - \varphi)$.

Upon close inspection, it is seen that (5.10) is not independent of (5.8) and (5.9). Consequently, minimizing (5.7) with respect to c, K and B yields only two independent relations. As discussed in Chapter II there are various alternatives for eliminating this underspecification. One possibility is to arbitrarily fix one parameter initially, and then determine the remaining two parameters using two of the above three relations. The major disadvantage of this approach is that it may not be clear which parameter to fix or what value to prescribe to it. This approach is the one normally utilized in the standard method of equivalent linearization[8]. It is assumed that F(t) is trigonometric, and consequently B is taken to be equal to the amplitude of the excitation F(t). If F(t) is not trigonometric, there is some question as to what to do in the method of equivalent linearization.

An alternative approach to arbitrarily selecting certain parameters is to divide one of the independent relations so that a third relation is generated. The particular relation that is separated and the manner in which it is separated are arbitrary. However, some physical arguments may exist for making the above decisions.

It is important to note that, in the present case, lack of specification occurs only in the differential equation parameters c, K, and B. The solution parameters A and φ are determined uniquely by the equations (5.8) and (5.9). This happens because the differential equation error is linear in the minimizing parameters, and after the minimization is performed, the resulting equations can be written entirely in terms of the solution parameters A and φ.

Consequently, only two relations are required. However, if a unique equivalent system is desired, it is necessary to initially specify one of the differential equation parameters or to separate one of the above equations to a generate a third relation.

Since in many cases it is desirable to determine a unique system as well as a unique solution, the alternative of separating certain equations is utilized in the present situation. Equations (5.8) and (5.9) are chosen as the independent equations. Furthermore, it seems reasonable to separate (5.9) in the following manner. Let

$$B \cos \varphi = 2 \int_0^1 F(t) \cos (2\pi t - \varphi) dt \quad , \tag{5.11}$$

and

$$KA = 2 \int_0^1 f(y, \dot{y}, t) \cos (2\pi t - \varphi)\, dt \quad , \tag{5.12}$$

where $y = A \cos (2\pi t - \varphi)$.

The above separation is an attempt to make B model the excitation F(t) and K model the restoring force $f(x, \dot{x}, t)$. As stated previously, (5.11) and (5.12) are obtained in an arbitrary manner. It would be equally as valid to separate (5.8).

The three relations for determining the auxiliary equation parameters are (5.8), (5.11), and (5.12). These relations combined with equations (5.4) and (5.5) are sufficient for determining A, φ, c, K, and B. If for physical reasons, some of the parameters are prescribed initially, then certain of the above relations no longer apply.

For example, if c is prescribed initially, equation (5.8) is no longer valid. K and F may then be determined using (5.9) and (5.10).

As stated previously, the present approximation is a generalization of the method of equivalent linearization. The additional flexibility of being able to permit all the parameter to participate in the minimization process would seem to indicate a capacity for greater accuracy in the approximation. Another advantage of the above formulation is that the auxiliary system is linear; consequently, the solution form is algebraically uncomplicated, and the integrals required in the approximation usually are trigonometric.

5.2. General Cubic Approximation.

In addition to the linear auxiliary system another system has received some consideration in the literature recently[6]. The system is

$$\ddot{y} + c\dot{y} + ay + by^3 = B\,\mathrm{cn}(\eta t, k)\left[1 - k^2\,\mathrm{sn}(\varphi, k)\,\mathrm{sn}^2(\eta t - \varphi), k\right] , \qquad (5.13)$$

where c, a, b, B are constants, cn(u, k) and sn(u, k) are Jacobian elliptic functions, and φ is defined shortly. It is desirable that (5.13) possess the same period of excitation as the original system (5.1); consequently, η satisfies

$$\eta = 4K(k) , \qquad (5.14)$$

where K(k) is the complete elliptic integral of the first kind with modulus k. (5.13) possesses exact periodic solutions of the form

$$y = A\,\mathrm{cn}(\eta t - \varphi, k) , \qquad (5.15)$$

where A, φ, and η must satisfy

$$k^2 = \frac{bA^2}{2\eta^2} \quad , \tag{5.16}$$

$$bA^3 + A(a-\eta^2) = B\,cn(\varphi, k) \quad , \tag{5.17}$$

and

$$cA\eta = B\,sn(\varphi, k)\,dn(\varphi, k) \quad , \tag{5.18}$$

where dn(u, k) is a Jacobian elliptic function.

As in Section (5.1) all of the differential equation parameters (c, a, b, B) are considered unspecified and are determined using the mean square equivalence criteria. Using (2.7), the differential equation error is

$$\epsilon(t) = F(t) - f(y, \dot{y}, t) + c\dot{y} + ay + by^3$$

$$- B\,cn(\eta t, k)\left(1 - k^2\,sn^2(\varphi, k)\,sn^2(\eta t - \varphi, k)\right) \tag{5.19}$$

where $y = A\,cn(\eta t, k)$. c, a, b, and B are determined by minimizing (5.7).

Minimizing (5.7) with respect to c leads to the following result.

$$\int_0^1 \left(F(t) - f(y, \dot{y}, t)\right) sn(\eta t - \varphi)\,dn(\eta t - \varphi)\,dt +$$

$$4\left(B\,\frac{sn(\varphi, k)\,dn(\varphi, k)}{3k^2\eta} - \frac{cA}{3k^2}\right)\left((2k^2-1)E(k) + (1-k^2)K(k)\right) = 0. \tag{5.20}$$

In obtaining (5.20), reference (28) is used to evaluate the integrals of the various combinations of elliptic functions.

Minimizing (5.7) with respect to a, and again using reference (28) to evaluate the resulting elliptic integrals, the following is obtained.

$$\int_0^1 \Big(F(t) - f(y, \dot{y}, t)\Big) cn(\eta t - \varphi, k) dt +$$

$$4\,\frac{\big(aA - B\,cn(\varphi, k)\big)}{\eta} C_2 + \frac{4bA^3}{\eta} C_4 = 0 \quad , \tag{5.21}$$

where

$$C_2 = \int_{-\varphi}^{K-\varphi} cn^2(u, k)du = \frac{E(k) - (1-k^2)K(k)}{k^2}$$

and

$$C_4 = \int_{-\varphi}^{K-\varphi} cn^4(u, k)du = \frac{(2-3k^2)(1-k^2)K(k) + 2(2k^2-1)E(k)}{3k^4} \quad .$$

Minimizing (5.7) with respect to b, and using reference (28), the following relation is obtained.

$$\int_0^1 \Big(F(t) - f(y, y, t)\Big) cn^3(\eta t - \varphi, k)\, dt +$$

$$\big(Aa - B\,cn(\varphi, k)\big) \frac{4C_4}{\eta} + \frac{4bA^3}{\eta} C_6 = 0, \tag{5.22}$$

where C_4 is defined in (5.21) and

$$C_6 = \int_{-\varphi}^{K-\varphi} cn^6(u, k)du = \frac{4(2k^2-1)C_4 + 3(1-k^2)C_2}{5k^2} \quad .$$

Minimizing (5.7) with respect to c, a, and b yields three independent relations (5.20), (5.21), and (5.22). However, minimizing (5.7) with respect to B does not lead to an independent relation. If the manipulation is carried out, the expression obtained is

$$cn(\varphi, k) \Big\{ \int_0^1 \Big(F(t) - f(y, \dot{y}, t)\Big) cn(\eta t - \varphi, k) dt + \frac{4C_2}{\eta}\Big(aA - B\,cn(\varphi, k)\Big) + \frac{4C_4}{\eta} bA^3 \Big\} - sn(\varphi, k)dn(\varphi, k) \Big\} \tag{5.23}$$

$$\left\{ \int_0^1 \left(F(t) - f(y, \dot{y}, t)\right) sn(\eta t - \varphi k)\, dn(\eta t - \varphi k)\, dt + 4\left(B \frac{sn(\varphi, k) dn(\varphi, k)}{3k^2 \eta} - \frac{cA}{3k^2}\right)\left((2k^2 - 1)E(k) + (1 - k^2)K(k)\right)\right\} = 0. \qquad (5.23)\ \text{cont.}$$

(5.23) is identically satisfied, if (5.20) and (5.21) are satisfied. Consequently, the system to determine c, a, b, and B is underspecified. However, as in the previous section, the underspecification occurs only in the differential equation parameters. Since the differential equation parameters appear linearly in the differential equation error, the relations resulting from the minimization process may be written entirely in terms of the solution parameters. Therefore, equations (5.20), (5.21), (5.22), and (5.14) are sufficient to uniquely determine the solution parameters A, φ, η, and k.

To determine the entire system uniquely, it is possible to eliminate the underspecification by 1) prescribing one of the parameters c, a, or B initially, and then using the appropriate remaining three equations of (5.20) through (5.23) to determine the other three parameters; or 2) separate one of the above three independent relations to generate a fourth relation. The latter alternative is the one used. Before separating any relation, it is convenient to rewrite equations (5.21)and(5.22). It is possible to solve these equations for the quantities $\eta^{-1}(aA - B\, cn(\varphi, k))$ and $\eta^{-1} bA^3$, yielding

$$\frac{aA - B\, cn(\varphi, k)}{\eta} = \frac{C_6 \int_0^1 \left(f(y, \dot{y}, t) - F(t)\right) cn(\eta t - \varphi, k)\, dt - \qquad (5.24)$$

$$4(C_2 C_6 - C_4^2)$$

$$- \frac{C_4 \int_0^1 \left(f(y, \dot{y}, t) - F(t)\right) cn^3(\eta t - \varphi, k)dt}{4(C_2 C_6 - C_4^2)} , \qquad (5.24) \text{ cont.}$$

$$\frac{bA^3}{\eta} = \frac{C_2 \int_0^1 \left(f(y, \dot{y}, t) - F(t)\right) cn^3(\eta t - \varphi, k)dt - C_4 \int_0^1 \left(F(y, y) - F(t)\right) cn(\eta t - \varphi)dt}{4(C_2 C_6 - C_4^2)} . \qquad (5.25)$$

where C_2, C_4, and C_6 are given in (5.21) and (5.22). Equation (5.24) is the one that is separated. By requiring that the linear coefficient a model the restoring force only, and that B model the excitation only, (5.24) separates into

$$\frac{aA}{\eta} = \frac{C_6 \int_0^1 f(y, \dot{y}, t) cn(\eta t - \varphi, k)dt - C_4 \int_0^1 f(y, \dot{y}, t)\, cn^3(\eta t - \varphi, k)dt}{4(C_2 C_6 - C_4^2)} , \qquad (5.26)$$

and

$$\frac{B\, cn(\varphi, k)}{\eta} = \frac{C_6 \int_0^1 F(t)\, cn(\eta t - \varphi, k)dt - C_4 \int_0^1 F(t)\, cn^3(\eta t - \varphi, k)dt}{4(C_2 C_6 - C_4^2)} , \qquad (5.27)$$

As stated in Section 5.1, the above separation is arbitrary, although the particular one chosen seems reasonable.

Equations (5.14), (5.16), (5.17), (5.18), (5.20), (5.25), (5.26), and (5.27) are the equations used to determine c, a, b, B, A, φ, k, and η. If some of the differential equation parameters are prescribed initially, then certain appropriate equations must be eliminated from the above list.

The cubic approximation is a generalization of the linear approximation described previously. In the limit as k approaches zero, the cubic approximation reduces to the linear approximation. Since the cubic approximation contains a larger number of differential equation parameters than the linear system and since it contains the linear system as a limiting case, it is clear that the cubic approximation cannot be worse than the linear. The degree of improvement provided by the cubic system depends on the particular original system being considered.

5.3. Example 1.

The original system to be considered is the same system considered in Chapter II, namely

$$\ddot{x} + ax + bx^3 = F\cos(\omega t) \quad . \tag{5.28}$$

The steady-state periodic motions of (5.28) are of interest.

The linear approximation uses the following system,

$$\ddot{y} + Ky = F\cos(\omega t) \quad , \tag{5.29}$$

where F_1 in (5.29) is set equal to F in (5.28) and K is the minimizing parameter. This approximation has been obtained previously in Section 4.5. The amplitude-frequency relation is given by equation (4.99) and (4.100) with B and c equal to zero. The entire approximation is obtained by solving equations (4.96) and (4.99), with c equal to zero, for the unknowns K and A.

The cubic approximation utilizes the following auxiliary system,

$$\ddot{y} + ay + by^3 = B\,cn(\eta t, k) \quad ,$$

where a and b are set equal to a and b in (5.28) and B is the minimizing parameter. This approximation has also been obtained in Section 2.2. The amplitude-frequency relation is given in equation (2.26). The entire approximation is solved by solving equations (2.19), (2.23), and (2.27) for the unknowns B, A, η, and k. The above set of equations are convenient if b is positive. For b negative the appropriate set of equations are (2.28), (2.19)′, and (2.23)′.

A comparison between the linear approximation and the cubic approximation for the present example has already been made in Chapter II. Certain values of B, b, and a were selected, and both approximations were presented as functions of the frequency ω. It was shown that, for certain ranges in frequency, the cubic approximation leads to more accurate results. This conclusion is valid for both positive and negative b.

In order to make the present example complementary to the earlier study, a comparison is made for various values of B and b, while a and ω are fixed. Prescribing a to have the value one represents no loss in generality, since the original equation (5.28) can always be scaled to meet this requirement. Since in Chapter II and reference (6) the relative accuracy of the linear and cubic approximations are roughly the same for b positive (for an appropriate choice of ω) and for b negative (again, for an appropriate choice of ω), only the case for b negative is considered. The value of frequency is arbitrarily selected to be 0.6.

Figure 18 presents the results of the comparison. The quantity compared in both approximations is the maximum response. As a base for comparison, the exact periodic solution is obtained by numerically integrating equation (5.28). Plotted on the ordinate of Figure 18 is the value of the difference between the approximate amplitude and the exact amplitude divided by the exact amplitude. The ordinate, therefore, is the fractional error in the maximum response. This quantity is plotted as a function of b for various B. The exact and approximate solutions that are compared, correspond to the upper branch of the response curve; (cf. Figure 1) for ω equal 0.6.

Figure 18 illustrates again that the cubic approximation leads to much more accurate results than does the linear approximation. It is also noticed that the linear approximation seems to be insensitive to the particular value of b or B chosen, at least for the range considered. The linear error is on the order of 10^{-1}. On the other hand, the cubic error is very much dependent on the values of b and B selected. As pointed out in Section 2.2, the accuracy of the cubic approximation is primarily influenced by the value of B. As B approaches zero, the cubic approximation becomes exact. This fact is illustrated in Figure 18. For B equal to 0.01, the cubic error is smaller than 10^{-3}, even for quite large values of b. For B equal to 1.0, the cubic error is on the order of 10^{-1} for the larger values of b. The straight line behavior of the cubic error suggests that, for the range of parameters considered, the cubic error is proportional to $B(-b)^{1/2}$.

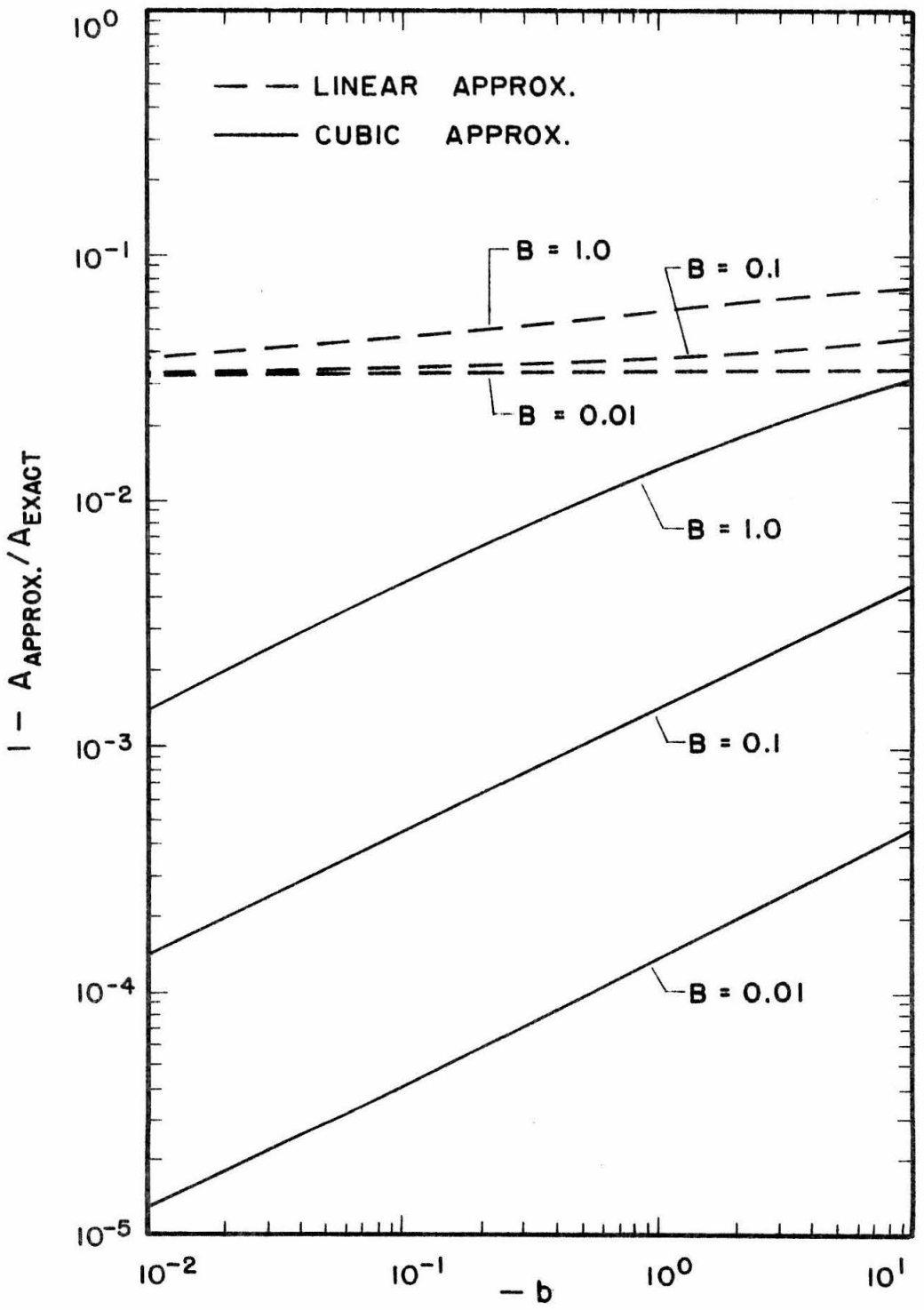

Figure 18: Response Error for $\ddot{x} + x + bx^3 = B \cos(0.6t)$

5.4. Example 2.

The system to be considered in this section is the following:

$$\frac{d^2x}{d\tau^2} + \frac{\gamma x^n}{1+\alpha|x|} = F\cos(\omega\tau) \quad , \tag{5.30}$$

where γ, α, F, and ω are constant and n is an odd positive integer so that the restoring force is symmetric with respect to the origin.* Equation (5.30) has considerable physical interest especially for the case when n equals 1. If α is positive, the restoring force has the property of saturation. For x small, the restoring force is nearly linear in x. For x large, the restoring force approaches a constant value. If α is negative, the restoring force becomes infinite as x approaches α^{-1}. The restoring force then resembles the force exerted on an elastically restrained particle moving in a one-dimensional rigid box. The equivalent equation approach is used to obtain linear and cubic approximations for the steady-state periodic oscillations of (5.30).

Linear Approximation

The linear approximation is obtained by utilizing the equations developed in Section (5.1). First, normalize τ in (5.30) so that the period of the solution is one. Letting $\frac{\omega\tau}{2\pi} = t$, (5.30) becomes

$$\frac{d^2x}{dt^2} + \left(\frac{2\pi}{\omega}\right)^2 \frac{\gamma x^n}{1+\alpha|x|} = \left(\frac{2\pi}{\omega}\right)^2 F\cos(2\pi t) \quad . \tag{5.31}$$

---

* The case for n an even positive integer can also be analyzed. Some specific details change, but in general, the procedure is identical.

From (5.31), it is clear that

$$f(x, \dot{x}) = \left(\frac{2\pi}{\omega}\right)^2 \frac{\gamma x^n}{1+\alpha|x|} \quad , \tag{5.32}$$

and

$$F(t) = \left(\frac{2\pi}{\omega}\right)^2 F \cos(2\pi t) \quad . \tag{5.33}$$

Since (5.31) contains no dissipation, it seems reasonable to set c equal zero in the linear approximation. The equations for determining the approximation are (5.4), (5.11), and (5.12). Since c equals zero, (5.5) implies that φ equals zero also. (5.4) reduces to

$$A = \frac{B}{K - 4\pi^2} \quad . \tag{5.34}$$

Using (5.33) and evaluating the trigonometric integrals, (5.11) reduces to

$$B = \left(\frac{2\pi}{\omega}\right)^2 F \quad . \tag{5.35}$$

Using (5.32) and standard integral tables, (5.12) reduces to

$$K = \frac{4\gamma}{\pi A^2 \alpha^{n+1}} \left\{ I - \frac{\pi}{2} \sum_{j=0}^{\frac{n-1}{2}} \frac{(A\alpha)^{2j}(2j)!}{2^{2j}(j!)^2} + \sum_{j=0}^{\frac{n-1}{2}} (A\alpha)^{2j+1} \frac{2^{2j}(j!)^2}{(2j+1)!} \right\} \quad , \tag{5.36}$$

where

$$I = \begin{cases} \dfrac{2}{\sqrt{1-A^2\alpha^2}} \tan^{-1}\left(\dfrac{1-A\alpha}{1+A\alpha}\right)^{1/2} , & |A\alpha| < 1 \\ \dfrac{1}{\sqrt{A^2\alpha^2-1}} \ln\left(\dfrac{A\alpha+1+(A^2\alpha^2-1)^{1/2}}{A\alpha+1-(A^2\alpha^2-1)^{1/2}}\right) , & A\alpha > 1 \end{cases} \quad .$$

The case of $A\alpha$ less than minus one is unphysical, since this corresponds to the particle penetrating the rigid walls of the box.

Equations (5.34), (5.35), and (5.36) constitute the determining equations for the linear approximation. For the numerical example to be considered later, n is chosen to be the one. In this case, (5.36) reduces to

$$K = \frac{4\gamma}{\pi A^2 \alpha^2} \left\{ I - \frac{\pi}{2} + A\alpha \right\} , \tag{5.37}$$

where I is given in (5.36).

Cubic Approximation

The cubic approximation is determined by utilizing the relations developed in Section 5.2. Since the original system (5.30) contains no dissipation, c is set to zero in (5.13). c being zero implies that φ vanishes through equation (5.18). The equations determining the cubic approximations are (5.14), (5.16), (5.17), (5.25), (5.26), and (5.27), with c and φ set equal to zero. These equations become:

$$\eta = 4K(k) , \tag{5.38}$$

$$k^2 = \frac{bA^2}{2\eta^2} , \tag{5.39}$$

$$bA^3 + A(a - \eta^2) = B , \tag{5.40}$$

$$\frac{bA^3}{\eta} = \frac{C_2 \int_0^1 \left(f(y, \dot{y}, t) - F(t)\right) cn^3(\eta t, k) dt - C_4 \int_0^1 \left(f(y, \dot{y}, t) - F(t)\right) cn(\eta t, k) dt}{4(C_2 C_6 - C_4^2)} , \tag{5.41}$$

$$\frac{aA}{\eta} = \frac{C_6 \int_0^1 f(y, \dot{y}, t)\, cn(\eta t, k)dt - C_4 \int_0^1 f(y, \dot{y}, t)\, cn^3(\eta t, k)dt}{4(C_2 C_6 - C_4^2)} \quad , \qquad (5.42)$$

and

$$\frac{B}{\eta} = \frac{C_6 \int_0^1 F(t)\, cn(\eta t, k)dt - C_4 \int_0^1 F(t)\, cn^3(\eta t, k)dt}{4(C_2 C_6 - C_4^2)} \quad . \qquad (5.43)$$

In equations (5.41), (5.42), and (5.43) there are four integrals that must be evaluated. The first one is

$$I_1 = \int_0^1 F(t)\, cn(\eta t, k)\, dt \quad . \qquad (5.44)$$

This integral is evaluated by first substituting F(t) given in (5.33) and then expanding the elliptic function in a Fourier series[36]. Using the orthogonality of the trigonometric functions, the only contribution comes from the first term in the expansion. Performing the algebra yields the following

$$I_1 = \frac{2\pi^3}{\omega^2} \frac{F}{kK(k)} \operatorname{sech}\left(\frac{\pi K(k')}{2K(k)}\right) \quad , \qquad (5.45)$$

where $k' = (1-k^2)^{1/2}$.

The second integral is

$$I_2 = \int_0^1 F(t)\, cn^3(\eta t, k)\, dt \quad . \qquad (5.46)$$

(5.46) is evaluated using techniques similar to the the above. It may

be shown that

$$I_2 = \left(1 - \frac{1}{k^2} + \frac{E(k)}{k^2 K(k)}\right) I_1 + \frac{2\pi^5 F}{\omega^2 k^3 K^3(k)} \sum (k) \quad , \tag{5.47}$$

where I_1 is given in (5.45) and

$$\sum (k) = +\frac{1}{2} \sum_{m=1}^{\infty} \frac{m}{\sinh(m\xi)} \left(\frac{1}{\cosh(m-1/2)\xi} + \frac{1}{\cosh(m+1/2)\xi)} \right) \quad , \tag{5.48}$$

where $\xi = \frac{\pi K(k')}{K(k)}$.

The remaining two integrals involve the restoring force f(y). The third integral is

$$I_3 = \int_0^1 f(y, \dot{y}, t)\, cn(\eta t, k)\, dt \quad . \tag{5.49}$$

Substituting (5.32) for f(y), multiplying by appropriate constants, adding and subtracting one in the numerator of the integral, and making the obvious change of variables, (5.49) becomes

$$I_3 = \left(\frac{2\pi}{\omega}\right)^2 \frac{4\gamma}{\eta A \alpha^{n+1}} \left\{ \int_0^{K(k)} \frac{du}{1 + A\alpha\, cn(u, k)} - \int_0^{K(k)} \sum_{m=0}^{n} \left(-A\alpha\right)^m \left(cn(u, k)\right)^m du \right\} \quad , \tag{5.50}$$

where the integration is over one quarter period because of the symmetry of the integrals. The value of the first integral in the brackets is [34]

$$I_5 = \int_0^{K(k)} \frac{du}{1 + \delta\, cn(u, k)} = \begin{cases} K(k) - E(k) + k' \quad , & \delta = 1 \\ \left(1 - \delta^2\right)^{-1} \left(\pi \left(\frac{\delta^2}{\delta^2 - 1}, k \right) - \delta f_1 \right) \quad , & \delta^2 \neq 1 \end{cases} \tag{5.51}$$

where $\pi(n^2, k)$ is the complete elliptic integral of the third kind and

$$f_1 = \begin{cases} \left(\frac{1-\delta^2}{k^2+k'^2\delta^2}\right)^{1/2} \tan^{-1}\left(\left(\frac{k^2+k'^2\delta^2}{1-\delta^2}\right)\frac{1}{k'}\right), & \frac{\delta^2}{\delta^2-1} < k^2 \\ \frac{1}{k}, & \frac{\delta^2}{\delta^2-1} = k^2 \\ \frac{1}{2}\left(\frac{\delta^2-1}{k^2+k'^2}\right)^{1/2} \ln\left(\frac{(k^2+k'^2\delta^2)^{1/2}+(\delta^2-1)^{1/2}k'}{(k^2+k'^2\delta^2)^{1/2}-(\delta^2-1)^{1/2}k'}\right), & \frac{\delta^2}{\delta^2-1} > k^2 \end{cases} .$$

It is to be remembered that I_5 has no physical meaning for δ less than minus one.

The remaining integrals in (5.50) are also evaluated using reference (28). Denoting C_m as

$$C_m = \int_0^{K(k)} cn^m(u,k)du , \tag{5.52}$$

the values of the integrals are

$$\left.\begin{aligned} &C_0 = K(k) , \quad C_1 = \frac{1}{k}\sin^{-1}(k) , \\ &C_2 = \frac{1}{k^2}\left(E(k) - k'^2K(k)\right) , \quad C_3 = \frac{1}{2k^2}\left((2k^2-1)C_1 + k'\right) \\ &C_{2m+2} = \frac{2m(2k^2-1)C_{2m} + (2m-1)k'^2C_{2m-2}}{(2m+1)k^2} \\ &C_{2m+3} = \frac{(2m+1)(2k^2-1)C_{2m+1} + 2mk'^2C_{2m-1}}{2(m+1)k^2} \end{aligned}\right\} \tag{5.53}$$

Combining the above values, I_3 may be written as

$$I_3 = \left(\frac{2\pi}{\omega}\right)^2 \frac{4\gamma}{\eta A \alpha^{n+1}} \left\{ I_5 - \sum_{m=0}^{n} (-A\alpha)^m C_m \right\} \quad , \tag{5.54}$$

where I_5 and C_m are given in (5.51) and (5.53).

The final integral to evaluate to complete the approximation is

$$I_4 = \int_0^1 f(y, \dot{y}, t)\, cn^3(\eta t, k) dt \quad . \tag{5.55}$$

Utilizing the same techniques as those employed in evaluating I_3, the value of I_4, is found to be

$$I_4 = \left(\frac{2\pi}{\omega}\right)^2 \frac{4\gamma}{\eta A^3 \alpha^{n+3}} \left\{ I_5 - \sum_{m=0}^{n+2} (-A\alpha)^m C_m \right\} \quad , \tag{5.56}$$

where I_5 and C_m are given in (5.51) and (5.53).

Equations (5.38) through (5.43) may be used to determine the unknowns A, η, k, a, b, and B. However, for numerical evaluation, it is more convenient to consider either A or k as known and to let ω be an unknown. Substituting the values of I_1 through I_4, and using equation (5.39), (5.40), (5.41), (5.42), and (5.43), it is possible to eliminate four of the unknowns and to obtain a single relation between A and k. If either A or k is considered known, the problem reduces to determining the roots of a transcendental equation for the other variable.

Discussion

In order to compare the approximations, it is necessary to consider specific examples. As indicated earlier, the restoring

force in (5.30) possesses two general types of behavior. If α is positive, the system is softening, and without loss in generality, it is convenient to select γ and α equal. In this situation, the restoring force is asymptotic to the value $|x|^{n-1}$ for large $|x|$. If α is negative, the system is hardening, and without loss in generality, it is convenient to select α equal to minus one. In this case, the restoring force becomes infinite as $|x|$ approaches one. The value selected for n is unity.

For the hardening system, the values selected for F and γ are $F=0.1$ and $\gamma=0.2$. Since the system is hardening, it is expected that the cubic approximation will also be hardening and, consequently, the equations (5.38) through (5.43) are in the appropriate form (i.e. k is positive and less than one). Both approximations for the periodic solution of (5.30) for the above values of the parameters are given in Figure 19. Also included is the exact solution (exact maximum amplitude of response) obtained by numerically integrating equation (5.30). The exact solution possesses different characteristics for various ranges in ω, and these are indicated in the figure.

One fact which is immediately apparent in Figure 19 is that the cubic approximation provides significantly better results than the linear system for $\omega > 0.3$. As pointed out previously, this is not surprising since the cubic system is a larger parameter system which includes the linear approximation as a limiting case. What might be surprising is the amount of improvement. For ω large and A near one, the cubic system yields very good results. For these

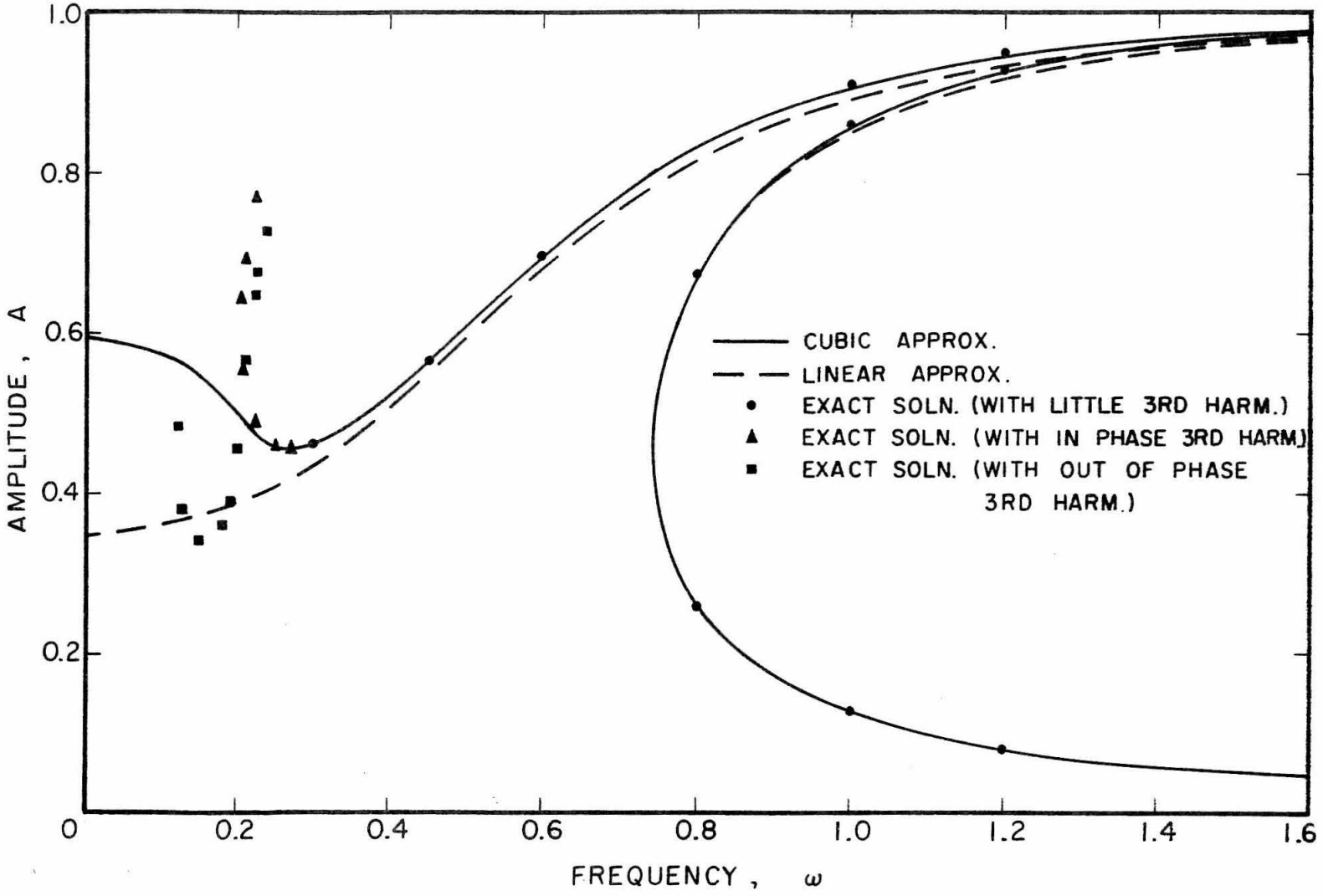

Figure 19: Response of $\ddot{x} + \frac{0.2x}{1-|x|} = 0.1 \cos(\omega t)$

values of amplitudes, equation (5.30) becomes highly nonlinear. However, the cubic system seems capable of representing the large nonlinearity quite well. For a frequency of 1.2, the error in amplitude for the upper branch of the response curve is less than one percent.

Another interesting aspect of the cubic approximation occurs for ω near 0.25. At this point, the approximate amplitude begins to increase. No such behavior is exhibited by the linear system. Furthermore, the exact solution also begins to increase for ω near 0.25 because of the influence of the ultraharmonic response of order 3. Although it is pure conjecture at the present time, it appears that the cubic system could be trying to model the ultraharmonic behavior of the exact solution. As the cubic system approaches ω equal 0.25, the modulus k is approaching 1. The limiting value of k equal 1 corresponds to ω equal 0. For k large, the wave form of the approximate solution (5.15) is very similar to a superposition of in phase $\cos(\omega t)$ and $\cos(3\omega t)$ terms. Furthermore, as the ultraharmonic response is approached by decreasing ω, the exact solution possesses a third harmonic component which is in phase with the primary harmonic component. Therefore, it seems that the cubic system is following the branch of the ultraharmonic where the third harmonic is in phase with the first harmonic. If the above reasoning is correct, the cubic approximation, which is essentially a primary response approximation, possesses the capability of yielding some information about the ultraharmonic response of the system. Although the approximation

rapidly deteriorates in the region of the ultraharmonic, it at least indicates the presence of a different phenomenon. The linear approximation does not possess such a capability. The cubic approximation is capable of modeling the exact solution until the magnitude of the in phase third harmonic is so large that the exact solution possesses six sign changes in one period of the motion. The value of ω for which this occurs is approximately $\omega=0.22$. For ω larger than 0.22, the cubic approximation is fairly accurate. For ω less than 0.22 the Jacobian elliptic cosine function is incapable of representing the exact solution wave form since the cn function possesses only two sign changes in one cycle for $0 \leq k < 1$. For $\omega < 0.22$, the cubic approximation is meaningless, but this is not surprising since there probably exists an infinite number of ultraharmonic responses for ω between 0 and 0.22. In addition, as ω goes to zero, the cubic approximation becomes ambiguous and essentially undefined.

An interesting aspect of the exact solution is that the two branches for the ultraharmonic of order 3 cross. This behavior occurs frequently, especially for equations possessing large nonlinearities.

For the softening system, the values selected for F and γ are $F=0.5$ and $\gamma=10$. In this case, it is expected that the cubic approximation will also be softening, and consequently equations (5.38) through (5.43) require modification. The modulus k becomes pure imaginary, and the above equations may all be transformed so that they involve only real quantities. This manipulation is purely

algebraic and is omitted for the sake of brevity. The final forms are easily obtainable from equations (5.38) through (5.43). Both approximate periodic solution amplitudes for the present values of the parameters are given in Figure 20. The exact maximum amplitude, obtained by numerical integration, is also included.

From Figure 20, it is clear that the cubic system again yields better results than the linear system. One interesting fact is that the major difference occurs at the "knee" of the response curve instead of for the larger values of amplitude.

Unlike the hardening case, the amplitude of the cubic approximation in Figure 20 does not seem to increase in the region of the ultraharmonic response. The main reason is that the location where the cubic approximation is no longer capable of representing the exact solution wave form occurs much sooner in the softening system. As ω is decreased and the ultraharmonic of order 3 is approached, the third harmonic component of the exact solution is out of phase with respect to the first harmonic component. Consequently, the exact solution wave form looks more like a square wave. The cubic approximation also yields a solution form which is capable of modeling a square wave. Since the present system is softening, the modulus k is pure imaginary, and the solution (5.15) takes the form $cn(\xi t, k_1)$ $dn(\xi t, k_1)$ where ξ and k_1 are real. The above combination may be written as $sn(\xi t + K, k_1)$ which approaches a square wave in the limit as k_1 approaches one. Therefore, it seems reasonable to expect that the cubic approximation could model the exact solution so long

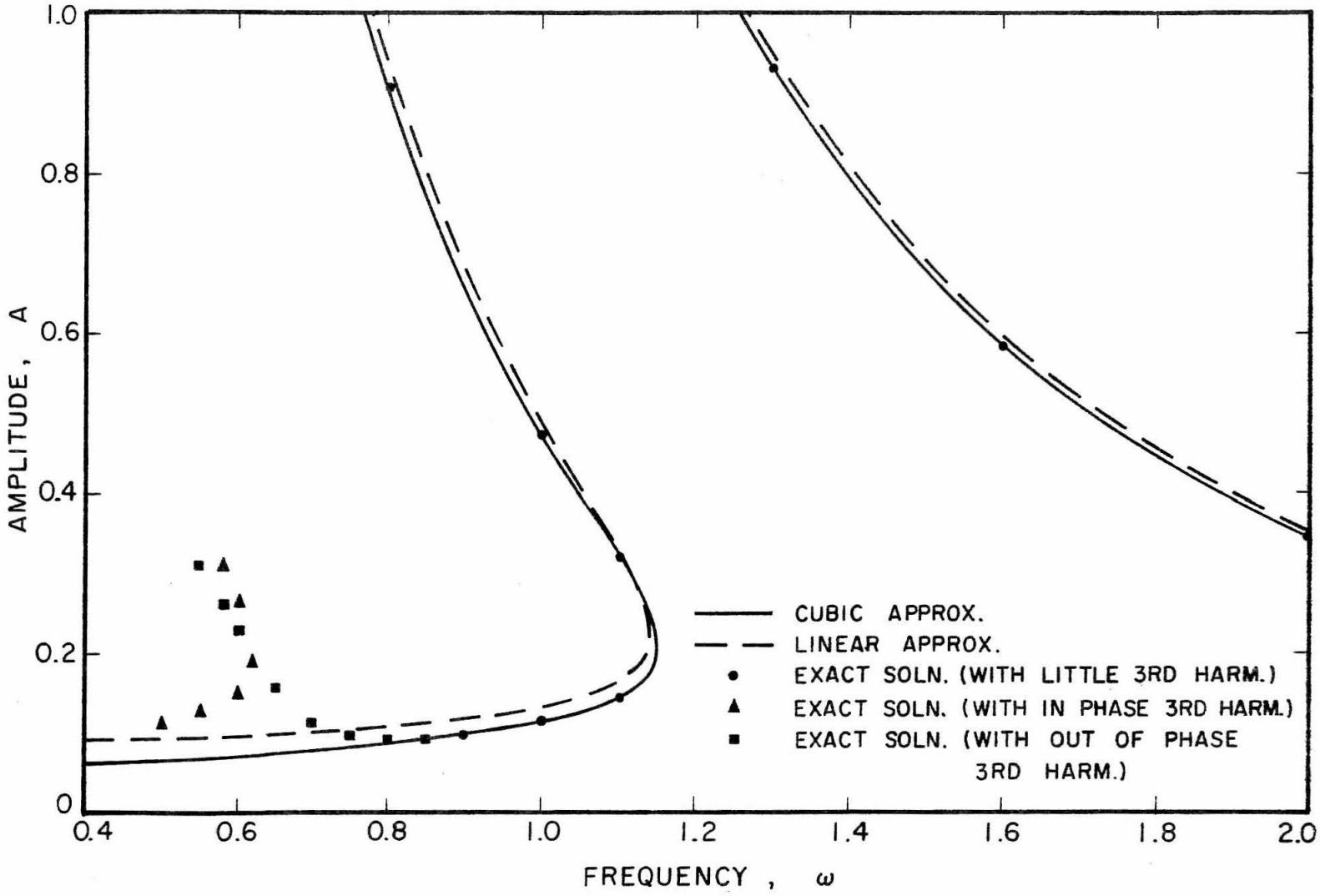

Figure 20: Response of $\ddot{x} + \frac{10x}{1+10|x|} = 0.5\cos(\omega t)$

as the third harmonic component does not become too large. If the third harmonic does get too large, the exact solution wave form becomes double peaked, and the cubic approximation is incapable of representing it. In Figure 20, the value of ω where the exact solution wave form becomes double peaked (in the region of the third harmonic resonance) is $\omega \sim 0.9$. Consequently, if $\omega > 0.9$, it seems reasonable to expect that the cubic approximate would be fairly accurate. However, for $\omega < 0.9$, the cubic system is no longer capable of representing the exact solution. At the value $\omega = 0.9$, the exact solution amplitude is still decreasing. It doesn't start to increase until ω is approximately 0.8. Therefore, the ultraharmonic resonance occurs outside the region where the cubic system is applicable, and. consequently, the cubic approximation does not increase. The softening case illustrates that the cubic approximation need not necessarily indicate the existence of a third ultraharmonic resonance.

The significance of the present example is threefold. It first shows that the cubic approximation can lead to a noticeable improvement over the linear approximation for systems possessing nonlinearities other than cubic nonlinearities. The example in Section 5.3 illustrates the superiority of the cubic system for modeling Duffing's equation. This is not surprising, since the original system possesses a cubic nonlinearity. However in the present example, it is not obvious, initially, that the cubic system would provide noticeable improvement.

Another aspect of the example is that, for the cubic approximation, the coefficient of the cubic term in the restoring force is allowed to vary. This tends to make the algebra somewhat more complicated but not unmanageable. The increased flexibility of the cubic modeling system seems to provide a more accurate approximation and allows for the possibility of describing some second order effects. (i.e. ultraharmonic response).

The third aspect is the behavior of the cubic approximation in the region of the ultraharmonic of order 3. In the hardening case, the cubic system seems capable of yielding some information concerning the ultraharmonic response. In the softening case, the cubic approximation seems to ignore the ultraharmonic response entirely. No explanation exists at the present time for the increase in the response curve for the hardening case in the region of the third ultraharmonic other than the one given above.

VI. RELATION OF THE EQUIVALENT EQUATION APPROACH TO OTHER APPROXIMATE TECHNIQUES

The equivalent equation approach is based on the concept of relating one differential equation, whose solution is known, to another differential equation, whose solution is desired. By making the two systems equivalent, it is assumed that the known solution of the first provides an accurate approximation to the solution of the second. Most other approximate techniques are based on assuming a certain solution form directly. The solution form contains some unspecified parameters which are selected by minimizing the residual obtained by substituting the assumed solution into the differential equation of interest.

In the present chapter, the relationship between the equivalent equation approach and some of the more common approximate techniques is examined. Section 6.1 presents the various approaches: collocation, subdomain, least squares, Galerkin's, and equivalent equation. The relationship between these techniques and the general method of weighted residuals is shown. In Section 6.2, some peculiarities associated with the method of least squares and the equivalent equation approach are illustrated. It is shown that the method of least squares may yield extraneous solutions when applied to nonlinear systems.

6.1. Method of Weighted Residuals.

The method of weighted residuals is a unification of all approximate averaging techniques which was introduced by Crandall [27].

He showed that many of the classical approximate techniques are related within the context of a weighted residual technique. It is also possible to include the equivalent equation approach in this classification.

Let the differential equation of interest be written as

$$\ddot{x} = f(x, \dot{x}, t) \quad , \tag{6.1}$$

where $f(x, \dot{x}, t)$ possesses the smoothness properties indicated in (3.56) and $f(x, \dot{x}, t)$ is periodic in t with period 1.* Most classical approximate techniques are based on assuming a certain solution form

$$x(t) = y(t, \beta_1, \ldots, \beta_s) \quad , \tag{6.2}$$

where y is periodic in t with period 1 and $\beta_j (j=1, \ldots, s)$ are undetermined parameters which are selected so that (6.2) represents an approximate solution of (6.1). The usual procedure is to obtain the error residual by substituting (6.2) into (6.1) yielding

$$\epsilon(t, \beta_1, \ldots, \beta_s) = f(y(t, \beta_1, \ldots, \beta_s), \dot{y}(t, \beta_1, \ldots, \beta_s), t) - \ddot{y}(t, \beta_1, \ldots, \beta_s) \quad . \tag{6.3}$$

The solution parameters $\beta_j (j=1, \ldots, s)$ are selected so as to minimize $\epsilon(t, \beta_1, \ldots, \beta_s)$. However, one difficulty arises in that $\epsilon(t, \beta_1, \ldots, \beta_s)$ is a function of t, which means that the β_j's obtained by a direct minimization would also depend on t. To eliminate this problem, the concept of a weighted average is introduced. The relations determining the $\beta_j (j=1, \ldots, s)$ may be written as

---

*This analysis could be performed for vector systems, but the second order scalar case was chosen for brevity and clarity.

$$\int_0^1 W_j(t)\,\epsilon(t,\beta_1,\ldots,\beta_s)dt = 0 \;;\; j=1,\ldots,s \quad , \tag{6.4}$$

where the average is over one cycle of the motion, and $W_j(t)$ are certain weight functions whose purpose is to eliminate the t dependence in $\epsilon(t,\beta_1,\ldots,\beta_s)$. The determining equations for many of the more common approximate techniques based on the averaging principle can be expressed in the form of (6.4) where the $W_j(t)$ depend on the particular approximate technique utilized.

Collocation

The method of collocation makes the error residual small for $t\in[0,1]$ by requiring that it be identically zero at certain arbitrarily prescribed points in [0, 1]. For the method of collocation, the weight functions take the form

$$W_j(t) = \delta(t-t_j) \;;\; j=1,\ldots,s \quad , \tag{6.5}$$

where $t_j(j=1,\ldots,s)$ are the joints in [0, 1] where $\epsilon(t,\beta_1,\ldots,\beta_s)$ vanishes identically. This method is particularly convenient because the relations generated by (6.4) using (6.5) are immediately algebraic in form. No further integration is necessary. The one major disadvantage of collocation is that, in general, the approximation obtained is not as accurate as some of the other techniques.

Subdomain

The method of subdomain consists of requiring that integrals of the error residual over certain arbitrarily selected intervals in t vanish identically. For this method, $W_j(t)$ are of the form

$$W_j(t) = H(t-t_{j-1}) - H(t-t_j) \ , \quad j=1,\ldots,s \ , \tag{6.6}$$

and

$$t_0 = 0 \ ; \ t_s = 1 \ ,$$

where H(u) is the Heaviside step function. The method of subdomain is more complicated than collocation in that the relations obtained from (6.4) still require an integration before determining $\beta_j (j=1,\ldots,s)$. However, if the solution form (6.2) is a truncated Fourier series with the $\beta_j (j=1,\ldots,s)$ as the undetermined coefficients, the resulting integrals are usually trigonometric.

Least Squares

The method of least square is based on minimizing the mean square of the error residual with respect to the $\beta_j (j=1,\ldots,s)$. The quantity minimized is

$$\int_0^1 \epsilon^2(t,\beta_1,\ldots,\beta_s)dt = \text{minimum} \ . \tag{6.7}$$

A necessary condition for a relative minimum is that the first derivative vanish. Therefore, (6.7) becomes

$$\int_0^1 \frac{\partial \epsilon}{\partial \beta_j}(t,\beta_1,\ldots,\beta_s)\,\epsilon(t,\beta_1,\ldots,\beta_s)dt = 0 \ , \quad j=1,\ldots,s \ , \tag{6.8}$$

where the weight functions are

$$W_j(t) = \frac{\partial \epsilon}{\partial \beta_j}(t,\beta_1,\ldots,\beta_s) \ ; \ j=1,\ldots,s \ . \tag{6.9}$$

(At first glance, it would seem that (6.8) is identical to the relations generated using the equivalent equation approach, but this is not so

as will be seen shortly.) Although the weight functions for collocation and subdomain are independent of the parameters $\beta_j (j=1,\ldots,s)$, the weight function (6.9) for the method of least square will in general depend on the $\beta_j (j=1,\ldots,s)$. This feature may result in undesirable consequences in that the method of least squares is capable of yielding extraneous approximate solutions and/or eliminating true approximate solutions. This aspect is considered further in Section 6.2. It may also be noted that the amount of labor involved in applying (6.8) is increased as compared to collocation or subdomain since the resulting integrations are usually more difficult.

Galerkin's Method

Galerkin's method involves making the error residual orthogonal to a set of trial functions on the interval $[0, 1]$. If the assumed solution (6.2) is of the form

$$y(t, \beta_1, \ldots, \beta_s) = \sum_{j=1}^{s} \beta_j \psi_j(t) \quad , \tag{6.10}$$

where $\psi_j(t)\ (j=1,\ldots,s)$ is an arbitrarily chosen set of trial functions depending on t only, Galerkin's procedure is straightforward. The relations determining $\beta_j (j=1,\ldots,s)$ are (6.4) where the weight functions are

$$W_j(t) = \psi_j(t) \ , \ j=1,\ldots,s \quad . \tag{6.11}$$

Higher order approximate solutions are easily obtained by simply taking the number of trial functions as large as desired.

If the assumed solution form is nonlinear in $\beta_j (j=1,\ldots,s)$, Galerkin's procedure is more difficult. One suggestion is to use the following as weight functions [11]

$$W_j(t) = \frac{\partial y(t,\beta_1,\ldots,\beta_s)}{\partial \beta_j} \quad , \quad j=1,\ldots,s \quad , \tag{6.12}$$

where $y(t,\beta_1,\ldots,\beta_s)$ is the assumed solution form. In general, the $W_j(t)$ in (6.12) will depend on $\beta_j (j=1,\ldots,s)$ as well as t, and the possibility arises for the modified Galerkin's procedure to yield meaningless approximations as is indicated in Section 6.2.

Equivalent Equation Approach

The equivalent equation approach differs from the above techniques in that it is primarily concerned with equivalent or approximate differential equations rather than equivalent solutions. The motivation for this approach is that the original differential equation of interest is immediately available, whereas the nature of the desired solution may not be known.

For the present approach it is necessary to define an auxiliary system as

$$\ddot{y} = g(y, \dot{y}, t \; ; \; \alpha_1,\ldots,\alpha_r) \quad , \tag{6.13}$$

where $g(y, \dot{y}, t;\alpha_1,\ldots,\alpha_r)$ is continuously differentiable in $y, \dot{y}, \alpha_1,\ldots,\alpha_r$ and continuous and periodic in t with period 1. The differential equation parameters $\alpha_i (i=1,\ldots,r)$ are selected so that the difference between (6.1) and (6.13) is minimized. (6.13) is chosen so that it is similar to (6.1) and so that it possesses known periodic solutions of the form (6.2) with period 1. In order that (6.2) be a solution of (6.13), there

will exist s relations between the $\alpha_i (i=1, \ldots, r)$ and the $\beta_j (j=1, \ldots, s)$. However, these s relations are not utilized until after the minimization of the differential equation error $\epsilon(t, \alpha_1, \ldots, \alpha_r, \beta_1, \ldots, \beta_s)$,

$$\epsilon(t, \alpha_1, \ldots, \alpha_r, \beta_1, \ldots, \beta_s) = f(y(t, \beta_1, \ldots, \beta_s), \dot{y}(\beta_1, \ldots, \beta_s), t) - g(y(t, \beta_1, \ldots, \beta_s), \dot{y}(t, \beta_1, \ldots, \beta_s), t; \alpha_1, \ldots, \alpha_r) \quad , \tag{6.14}$$

has been perofrmed. The $\alpha_i (i=1, \ldots, r)$ are selected so that (6.13) is close to (6.1) for all values of $\beta_j (j=1, \ldots, s)$.

The most appropriate equivalence criterion (cf. Chapter IV) is

$$\int_0^1 \epsilon^2(t, \alpha_1, \ldots, \alpha_r, \beta_1, \ldots, \beta_s) dt = \text{minimum} \quad . \tag{6.15}$$

Minimizing with respect to $\alpha_i (i=1, \ldots, r)$, (6.15) becomes

$$\int_0^1 \frac{\partial \epsilon(t, \alpha_1, \ldots, \alpha_r; \beta_1, \ldots, \beta_s)}{\partial \alpha_i} \epsilon(t, \alpha_1, \ldots, \alpha_r; \beta_1, \ldots \beta_s) dt = 0 \; ; \; i=1, \ldots, r. \tag{6.16}$$

As shown in Chapter II, the s relation between the α_i's and the β_j's may now be employed to obtain various other forms of the resulting relations, but the particular form given in (6.16) is convenient in the present situation. Furthermore, as discussed in Chapter II, only p of the r relations in (6.16) may be independent, and consequently certain other measures must be taken to insure a unique determination of the $\alpha_i (i=1, \ldots, r)$. (It may be noted that it is not always necessary to uniquely determine all of the $\alpha_i (i=1, \ldots, r)$ in order to determine uniquely the $\beta_j (j=1, \ldots, s)$; cf. Chapter V.)

It is clear that (6.16) may be put into the form (6.4) where the weight functions are

$$W_i(t) = \frac{\partial \epsilon}{\partial \alpha_i}(t, \beta_1, \ldots, \beta_s; \alpha_1, \ldots, \alpha_r) \quad ; \quad i=1, \ldots, r \quad . \tag{6.17}$$

Therefore, it is possible to consider the equivalent equation approach as a special case of the method of weighted residuals.

Although there is a great similarity between the form of (6.9) and the form of (6.17), the method of least squares and the equivalent equation approach are in fact different approximate techniques. The weight functions for least squares are obtained by minimizing the error residual (6.3) with respect to the solution parameters $\beta_j (j=1, \ldots, s)$, whereas the weight functions for the equivalent equation approach are obtained by minimizing the differential equation error (6.14) with respect to the differential equation parameters $\alpha_i (i=1, \ldots, r)$. Another difference is that the α_i often appear linearly in the differential equation error, and, consequently, the weight functions are independent of the $\alpha_i (i=1, \ldots, r)$. On the other hand, the weight functions for the method of least squares are almost always nonlinear in the $\beta_j (j=1, \ldots, s)$.

Under certain circumstances, Galerkin's procedure and the equivalent equation approach may yield the same approximation. If in Galerkin's procedure the set of trial functions ψ_j contains a certain number of trigonometric functions, the weight functions in (6.4) will be the trigonometric functions ψ_j. If in the equivalent equation approach, the auxiliary system is chosen to be the linear approximation with an excitation proportional to a linear combination of the

ψ_j, the corresponding weight functions will also be ψ_j. Therefore, in the present situation, Galerkin's procedure and the equivalent equation approach give the same approximation. However, this fact is not true in general, especially when the assumed solution forms are other than trigonometric.

It should also be pointed out that the equivalent equation approach is severely limited by the fact that, at the present time, the class of nonlinear differential equations possessing known periodic solutions is relatively small. It would seem worthwhile to try to enlarge this class of equations.

6.2. Anomalies Associated with the Method of Least Squares and Other Averaging Techniques.

The equivalent equation approach and the method of least squares are similar in that they both determine unspecified parameters by minimizing an averaged error quantity. This aspect may lead to peculiar results if caution is not exercised. To illustrate this fact, consider the following example.

Let us obtain an approximate solution using the method of least squares for the Duffing's equation

$$\ddot{x} + ax + bx^3 = B\cos(\omega t) \quad . \tag{6.18}$$

Assume the solution form to be

$$y = A\cos(\omega t) \quad , \tag{6.19}$$

where ω in (6.18) and (6.19) are equal and A is to be determined. Using (6.3), the error residual is

$$\epsilon(t, A) = (-A(a-\omega^2) + B)\cos(\omega t) - bA^3\cos^3(\omega t) \quad . \tag{6.20}$$

Applying (6.8) and evaluating the resulting trigonometric integrals, the relation determining A is

$$\frac{15}{8}bA^5 + 3b(a-\omega^2)A^3 - \frac{9}{4}bBA^2 + (a-\omega^2)^2A - B(a-\omega^2) = 0 \quad . \qquad (6.21)$$

Immediately it is noticed that (6.21) is of fifth order in A, whereas the relation generated in Section 2.2 using equivalent linearization was only of third order. The value of A given by (6.21) is plotted as a function of ω in Figure 21. The values of a, b, and B used are a=1.0, b=0.1, and B=0.1. Also included on the figure is the approximation obtained using equivalent linearization. Some exact solution points obtained by numerical integration of (6.18) are also shown.

From the figure, it is clear that the method of least squares gives some erroneous results. It predicts the existence of five periodic solutions of the form (6.19) for various ranges in ω. The theory of Duffing's equation is well known, and it is generally accepted that there exist only three solutions of the form (6.19) in the region of primary response[33]. Furthermore, the approximation predicts the emergence of two solutions from the point ω=1.0 and A=0. This is completely contrary to the usual notion of the behavior of Duffing's equation.

To understand the reason for the multiplicity of solutions, consider the minimization condition (6.7). The method of least squares is based on minimizing (6.7) with respect to $\beta_j (j=1,\ldots,s)$. However, if $\epsilon(t,\beta_1,\ldots,\beta_s)$ is nonlinear in the β_j (which is usually the case for nonlinear differential equations), the determining equations (6.8) may

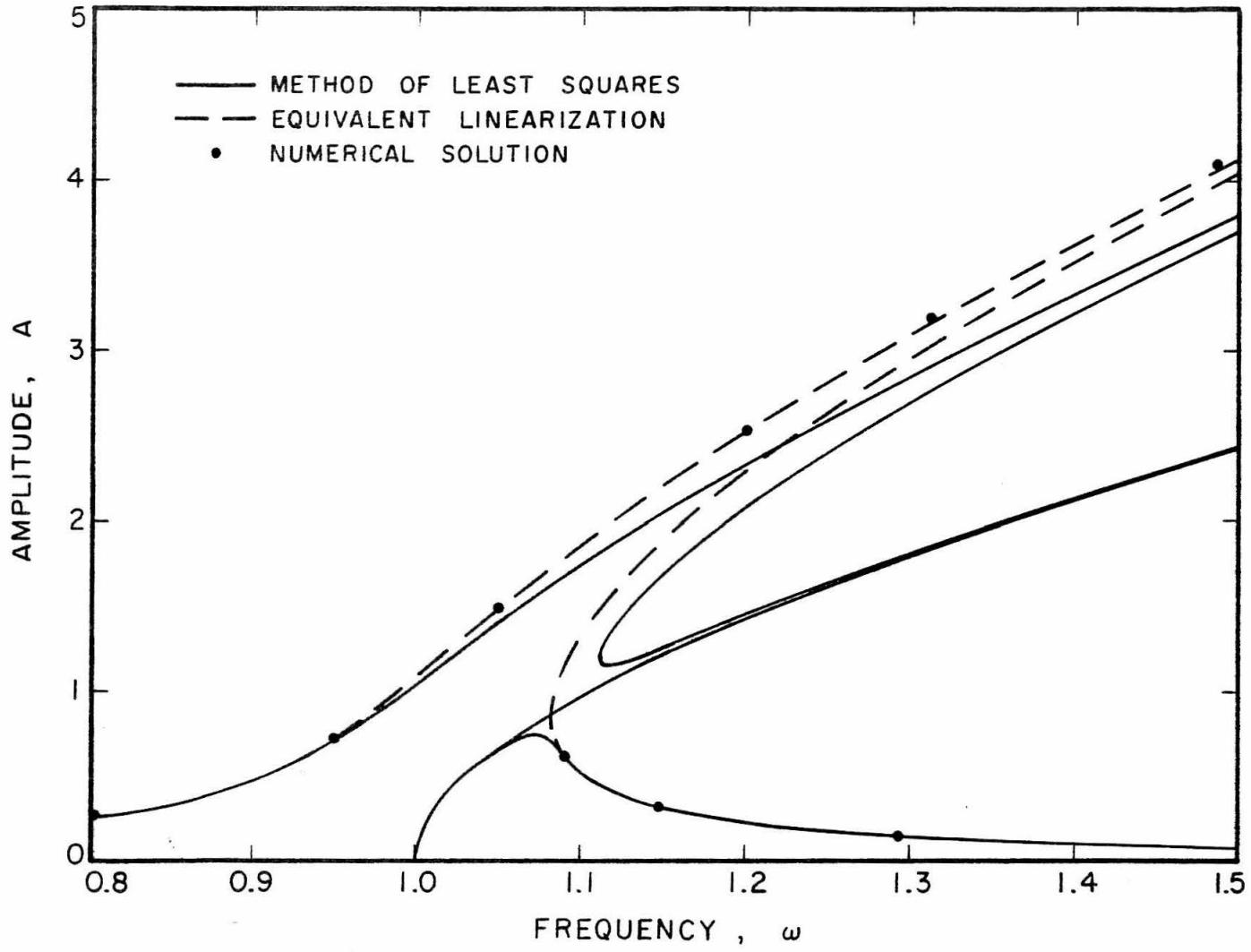

Figure 21: Response of $\ddot{x} + x + 0.1x^3 = 0.1\cos(\omega\tau)$

yield maximums as well as minimums of (6.7). To investigate the nature of the extremums, consider the sign of the second derivatives of (6.7) associated with the solutions generated by (6.8). Denoting the second derivatives of (6.7) with respect to β_i and β_j by κ_{ij}, the second derivatives are

$$\kappa_{ij} = \int_0^1 \frac{\partial W_i(t, \beta_1, \ldots, \beta_s)}{\partial \beta_j} \epsilon(t, \beta_1, \ldots, \beta_s)\, dt + \int_0^1 W_i(t, \beta_1, \ldots, \beta_s)\, W_j(t, \beta_1, \ldots, \beta_s)\, dt \quad , \qquad (6.22)$$

where $W_i(t, \beta_1, \ldots, \beta_s)$ are the weight functions given in (6.9). A necessary and sufficient condition for a solution to be a minimum is that the matrix κ_{ij} be positive definite[37]. It is clear that the matrix associated with the second integral in (6.22) is always positive definite. Therefore, the positive definiteness of κ_{ij} depends on the behavior of the matrix associated with the first integral in (6.22). In the method of least squares, the weight functions in general depend on the $\beta_j (j = 1, \ldots, s)$ and, therefore, the matrix κ_{ij} may be either positive definite, negative definite, or indefinite. Consequently, the solutions generated by (6.8) may correspond to maximums, minimums, or saddle points.

In the above example, the κ_{ij} matrix reduces to one term. If $\frac{\partial^2 E}{\partial A^2}$ is computed for the various solutions, it is seen that the extraneous ones correspond to maximums of the mean square error. In Figure 21 for a frequency of 1.05, the bottom two solutions correspond to maximums of the averaged error. The top solution

corresponds to a minimum. Similarly, for a frequency of 1.5, the bottom solution and the top two solutions correspond to minimums, whereas the middle two maximize the error. If the solutions generating maximums are eliminated, the response curve is more like the usual one associated with Duffing's equation except that the lower branch is discontinuous.

It is worthwhile to note that the approximation obtained from least squares using the solution form

$$y = A\,\mathrm{cn}\left(\frac{2K(k)\omega}{\pi}\,t, k\right) \tag{6.23}$$

possesses the same general behavior as the linear approximation obtained using least squares. The actual numerical coefficients in (6.21) are changed slightly, but the conclusions are similar.

As another example, consider Van der Pol's equation

$$\ddot{x} + \mu(x^2 - 1)\dot{x} + x = 0 \quad , \tag{6.24}$$

where it is required to obtain an approximate periodic solution using least squares and the solution form (6.19). Both A and ω are to be determined using (6.8). The relations obtained by minimizing with respect to A and ω are

$$\mu A\omega^2\left(\frac{3}{8}A^4 - A^2 + 1 + \frac{(1-\omega^2)^2}{\mu^2\omega^2}\right) = 0 \tag{6.25}$$

and

$$\frac{\mu^2 A^2 \omega}{8}\left(A^4 - 4A^2 + 8 - 16\,\frac{(1-\omega^2)^2}{\mu^2}\right) = 0 \quad . \tag{6.26}$$

The only solution of these two equations is $A=0$ and ω arbitrary. But it is well known that for $|\mu|$ sufficiently small, (6.24) possesses a limit cycle near $x=2\cos t$. In fact, for all μ, (6.24) possesses one

limit cycle [35]. Again, the method of least squares yields a meaningless approximate solution.

Another interesting situation is encountered if the method of least squares is used to obtain the limit cycle behavior of Rayleigh's equation

$$\ddot{x} + \mu\left(1 - \frac{\dot{x}^2}{3}\right)\dot{x} + x = 0 \quad . \tag{6.27}$$

Since Rayleigh's equation may be transformed into Van der Pol's equation using a simple transformation, it might be expected that least squares would yield the same conclusions about the periodic motions of (6.27) as it did about equation (6.24). This is not the case, however. Least squares for (6.27) predicts the solution A=0 and two limit cycles: y=1.839 cos t and y=1.191 cos t. If the second derivative of the error is evaluated for the two nontrivial solutions, it is seen that the first solution corresponds to a minimum of the error, and the second corresponds to a maximum.

As illustrated by the above examples, it is clear that the method of least squares predicts some unusual results when applied to nonlinear systems.

Although no specific examples where anomalies arise are presented for the equivalent equation approach, it is clear that similar difficulties could occur. Analogous to the method of least squares, the equivalent equation approach determines unspecified parameters by minimizing an averaged error. Equation (6.22) applies also to the equivalent equation approach except that the solution parameters $\beta_j (j=1,\ldots,s)$ are replaced by the differential equation parameters

$\alpha_i (i=1,\ldots,r)$. If the weight functions (6.17) depend on the parameters $\alpha_i (i=1,\ldots,s)$ the matrix K_{ij} may be positive definite, negative definite, or indefinite. Therefore, equations (6.16) may generate maximums, minimums, or saddle points. Fortunately, the $\alpha_i (i=1,\ldots r)$ often occur linearly in the auxiliary equation and, therefore, also occur linearly in $\epsilon(t)$. This means that the weight functions are independent of the $\alpha_i (i=1,\ldots,r)$, and, therefore, the matrix K_{ij} is positive definite. In this situation, all of solutions generated by (6.16) correspond to minimums of (6.15). Consequently, once the solutions are obtained, they do not require further checking.

The above analysis indicates that unusual results may occur using the method of least squares and the equivalent equation approach when the weight functions depend on the unspecified parameters. Similar behavior may also exist for the modified Galerkin's procedure. Although Galerkin's procedure is not related to minimizing an error quantity, the final equations determining the approximation are very similar to those given by the method of least squares and the equivalent equation approach. Furthermore, the weight functions (6.12) will in general depend on the $\beta_j (j=1,\ldots,s)$ if nonlinear solution forms are used. Therefore, it seems reasonable to suspect that difficulties could arise in this situation. Equation (6.22) does not apply to Galerkin's procedure since the technique is not related directly to any minimization condition. Further investigation is necessary to determine if anomalies can occur in the modified Galerkin's procedure and, if they can, to determine how to eliminate them.

VII. SUMMARY AND CONCLUSIONS

The purpose of the present investigation was to study an approach, suggested by W. D. Iwan, for obtaining approximate periodic solutions to nonlinear ordinary differential equations of the type which arise in dynamical systems. The approach, called the equivalent equation approach, is a generalization of the method of equivalent linearization, and it is applicable to any differential system possessing periodic solutions. In the present formulation, the equivalent equation approach treats only periodic motions, although there seems to be no conceptual difficulty in adapting the technique to treat transient problems as well. The approach is based on defining a differential system (linear or nonlinear) which is equivalent to the original system of interest. The alternative (auxiliary) system is selected such that it is "close" or "similar" to the original system and such that it possesses known periodic solutions. By making the auxiliary system equivalent to the original system, it is assumed that the corresponding solution of the auxiliary system will represent an accurate approximation to the exact solution of the original system. The conditions under which the above assumption is justified and the manner in which the auxiliary system is made close to the original system are two of the main considerations in this investigation.

The equivalent equation approach is described in detail in Chapter II. In section 2.1, the specific example of the undamped

Duffing's equation with trigonometric excitation is considered. It is shown that a cubic modelling system can give noticeable improvement over a linear modelling system in predicting the steady-state response amplitude. This should not be surprising since the cubic auxiliary system can represent, exactly, the nonlinear restoring force in Duffing's equation, whereas, the linear system cannot.

In Chapter III, the relationship between the differential equation error (the difference between the original system and the equivalent system) and the solution error (the difference between the exact periodic solution and the solution of the equivalent system) is investigated. Under certain conditions, bounds are obtained on the solution error in terms of the differential equation error. The technique employed is to consider the system governing the exact solution error as a two point boundary value problem. Reformulating the problem in terms of an integral equation using the Green's function for the unique linear part, the method of successive approximation is used to obtain a bound on the exact solution error. The analysis indicates that if the original system possesses an exact unique periodic solution, it is always possible to select an auxiliary system such that the exact solution error is less than any arbitrarily prescribed bound. Conversely, if the original system satisfies certain continuity and Lipschitz conditions and if there exists an auxiliary system such that certain inequalities are valid, then the original system possesses an exact unique solution in a region defined by the above mentioned inequalities. Unfortunately, the

above analysis gives no prior indication as to which auxiliary system should be selected in order to obtain the smallest error bound. From the form of the bound obtained, it does show that, as the differential equation error approaches zero, the corresponding solution error also approaches zero.

The general analysis utilizes the Green's function for the unique linear part of differential equation describing the exact solution error. In general, the coefficients are functions of the independent variable, and, consequently, in practical applications it becomes exceedingly difficult to determine the Green's function. To avoid this problem, the integral equation is reformulated using a Green's function for a system with constant coefficients whose general form is well known. The particular values of these constants are selected so as to minimize the resulting error bound. Although this procedure leads to a less accurate error bound, the additional applicability gained seems well worth the price.

Error bounds are obtained for the undamped trigonometrically excited Duffing's equation for both the linear and cubic approximations. Where obtainable, the error bounds seem to describe, fairly well, the qualitative behavior of the exact error. It is also shown that the error bound associated with the cubic approximation is an order of magnitude smaller than the bound for the linear system. The same relation holds for the exact solution errors in the regions where bounds are obtainable. As expected, bounds are not obtainable using the present techniques for all ranges of the parameters. A comparison of the

present approach and approaches suggested by other investigators is also included.

The approach which is applied successively to non-autonomous systems does not provide much information for the case of autonomous systems. It is shown that the conditions required by the approach are never satisfied for autonomous original systems possessing non-trivial periodic solutions. Therefore, the above approach gives information concerning the trivial solution only. Consequently, an alternative bound is obtained for second order conservative autonomous systems. It involves estimating an integral for the exact period of the motion in terms of a known integral for the period of an auxiliary system. An example of the autonomous Duffing's equation modelled by the linear system is included to illustrate the application and the accuracy of the technique.

In Chapter IV, the manner in which an auxiliary system is made equivalent to the original system is considered. Various equivalence criteria for minimizing the differential equation error are compared, namely, mean square error minimization, mean absolute value error minimization, and maximum absolute value error minimization. The differential equation error is defined as the difference between the original system and the auxiliary system when both are evaluated at the auxiliary system solution. The minimization is performed with respect to parameters appearing in the auxiliary system. It is of interest to determine which of the above schemes yields the smallest <u>actual</u> solution error.

Since the actual error is in general unaccessible analytically, examples are used to illustrate the major results. Only second order scalar systems are considered. Four examples are presented comprising autonomous and non-autonomous systems and including both linear and cubic auxiliary systems.

The analysis indicates that, depending on the specific example and values of parameters considered, each of the above minimization schemes can yield the most accurate approximation in certain cases, but, on the average, the minimum mean square error seems to be the most appropriate criterion to use. Furthermore, it is, by far, the easiest of the three methods to apply.

It is realized that the above conclusions are based only on a relatively few number of examples and that there exist many alternative error minimization techniques other than the ones considered in the analysis. However, an attempt was made to make the examples representative, and the specific minimization techniques considered were chosen because of their physical relevance and their relation to the error bound analysis done previously.

In Chapter V, a comparison is made between a linear and a cubic auxiliary system. The general second order linear and cubic systems for modelling an arbitrary second order original system are developed. Several examples are presented, and the results of both approximations are compared. The first involves the trigonometrically excited Duffing's equation considered in Chapter II. A comparison between the linear and cubic approximations is made for

various values of nonlinearity and excitation level. This complements the comparison in Chapter II which is performed as a function of excitation frequency. The example illustrates the degree of superiority of the cubic system. Furthermore, it shows that the solution error associated with the linear approximation is rather insensitive to the value of the cubic coefficient and the excitation level, at least for the particular parameters considered. On the other hand, for the same range of parameters, the exact error for the cubic system seems to be directly proportional to the excitation level and proportional to the square root of the cubic coefficient.

A second example for comparing the linear and cubic approximations involves a saturating system described in section 5.4. Both linear and cubic approximations are obtained for the two cases of a hardening and softening restoring force, and in both cases the cubic system provides more accurate results.

In addition to being more accurate, the cubic system seems capable of providing some information concerning the ultraharmonic response of the saturating system. For the hardening case, the cubic system seems to follow the branch of the third ultraharmonic where the third harmonic is in phase with the primary harmonic component. Although the accuracy rapidly deteriorates, the cubic system at least indicates the presence of a different phenomenon. For the softening case, the cubic approximation gives no indication of an ultraharmonic response. The reason for this difference in behavior is suspected to be related to the ability of the cubic

approximation to represent the exact solution wave form. Additional investigation is required in this area in order to obtain a better understanding of the phenomenon.

A brief comparison of the equivalent equation approach to some of the more classical approximate techniques where specific solution forms are assumed is presented in Chapter VI. The techniques considered are collocation, subdomain, least squares, and Galerkin's. The relation of all the above techniques to the general method of weighted residuals is shown. Under certain conditions, the equivalent equation approach and Galerkin's procedure yield identical approximations, but in general they are different. Also, the equivalent equation approach and the method of least squares do not generate the same approximation in general.

In section 6.2, some peculiarities associated with the method of least squares is presented. For example, the method of least squares predicts, for certain ranges in frequency, that the undamped trigonometrically excited Duffing's equation possesses five solutions of the form $A\cos(\omega t)$. It is shown that the extraneous solutions are associated with maximums of the mean square error residual.

Similar results may possibly occur for the equivalent equation approach. However, if the differential equation parameters appear linearly in the differential equation error, the equivalent equation approach always generates solutions corresponding to minimums of the mean square error.

The main conclusion of the investigation is that the equivalent

equation approach seems capable of providing a substantial improvement over other low order approximate techniques in describing periodic motions. It allows for the possibility of using nonlinear systems to model other nonlinear systems thus incorporating some of the features peculiar to nonlinear systems in a very natural manner. This enables one to treat equations with moderately large nonlinearities which are poorly handled by most classical approximate techniques.

Areas for Further Investigation

The major areas for further investigation associated with the present analysis seem to be the following:

1. A detailed analysis of the behavior of the cubic system in the area of the ultraharmonic response of order three is needed. It is of interest to determine if and when the cubic system is capable of providing information concerning the ultraharmonic behavior of the original system.

2. A more comprehensive investigation of the anomalous behavior of some approximate techniques is necessary. Specifically, it seems worthwhile to determine if the modified Galerkin's procedure can yield meaningless results.

Other areas of possible investigation not considered in any detail in the present study are:

1. adaption of the equivalent equation approach to model subharmonic and ultraharmonic response;

2. determining the feasibility of obtaining (or generating) higher order approximations using the equivalent equation approach;

3. modification of the equivalent equation approach to treat transient problems; and

4. investigation of the merits of approximate stability analyses based on nonlinear solution forms arising from the equivalent equation approach.

REFERENCES

1) Coddington, E. A. and Levinson, N., Theory of Ordinary Differential Equations, McGraw-Hill, New York, 1955, Chapter 1.

2) Minorsky, N., Nonlinear Oscillations, D. Van Nostrand, Princeton, New Jersey, 1962, Chapter 1.

3) La Salle, J. and Lefschetz, S., Stability by Liapunov's Direct Method, Academic Press, New York, 1961.

4) Struble, R. A., Nonlinear Differential Equations, McGraw-Hill, New York, 1962, pp. 67-71.

5) Bogoliubov, N. and Mitropolsky, A., Asymptotic Methods in the Theory of Nonlinear Oscillations, 2nd Edition, Gordon and Breach, New York, 1961.

6) Iwan, W. D., "On Defining Equivalent Systems for Certain Ordinary Nonlinear Differential Equations," International Journal of Nonlinear Mechanics, Vol. 4, No. 4, 1969, pp. 325-334.

7) Iwan, W. D., "Application of an Equivalent Nonlinear System Approach to Dissipative Dynamical Systems", Journal of Applied Mechanics, Vol. 36, No. 3, Sept. 1969, pp. 412-416.

8) Reference 2, pp. 348-355.

9) Kantrovich, L. V. and Krylov, V. I., Approximate Methods of Higher Analysis, P. Noordhoff Ltd., Gröningen, 1958, pp. 261-262.

10) Crandall, S. H., Engineering Analysis, McGraw-Hill, New York, 1956, pp. 148-150.

11) Eringen, A. C., "Transverse Impact on Beams and Plates," Journal of Applied Mechanics, Vol. 20, No. 4, 1953, pp. 461-468.

12) Klotter, K. and Cobb, P. R., "On the Use of Nonsinusoidal Approximating Functions for Nonlinear Oscillation Problems", Journal of Applied Mechanics, Vol. 27, No. 3; Trans. ASME, Vol. 82, Series E, Sept. 1960, pp. 579-583.

13) Barkham, P. G. D. and Soudack, A. C., "An Extension of the Method of Kryloff and Bogoliuhoff", International Journal of Control, Vol. 10, No. 4, 1969, pp. 377-392.

14) Barkham, P. G. D. and Soudack, A. C., "Approximate Solutions of Non-linear Non-autonomous Second Order Differential Equations", International Journal of Control, Vol. 11, No. 1, 1970, pp. 101-114.

15) Denman, H. H. and Liu, Y. K., "Applications of Ultraspherical Polynomials to Nonlinear Oscillations. II - Free Oscillations", Quarterly of Applied Mathematics, Vol. 22, 1965, pp. 273-292.

16) Liu, Y. K., "Application of Ultraspherical Polynomials to Nonlinear Forced Oscillations," Journal of Applied Mechanics, Vol. 34; Trans. ASME, Vol. 89, Series E, 1967, pp. 225-226.

17) Helfenstein, H., Ueber eine spezielle Lamésche Differentialgleichung, mit Anwendung auf eine approximative Resonany formel der Duffingschen Schwingungagleichung, Eidgenössisohen Technischen Hochschule in Zurich, Promotionsarbeit No. 1985, 1950.

18) Cesari, L., "Functional Analysis and Periodic Solutions of Nonlinear Differential Equations", Contributions to Differential Equations, Vol. 1, 1963, pp. 149-187.

19) Cesari, L., "Functional Analysis and Galerkin's Method", Michigan Mathematics Journal, Vol. 11, 1964, pp. 385-414.

20) Urabe, M., "Galerkin's Procedure for Nonlinear Periodic Systems," Archives of Rational Mechanics and Analysis, Vol. 20, 1965, pp. 120-152.

21) Urabe, M., "Periodic Solutions of Differential Systems, Galerkin's Procedure and the Method of Averaging", Journal of Differential Equations, Vol. 2, 1966, pp. 265-280.

22) Urabe, M. and Reiter, A., "Numerical Computation of Non-Linear Forced Oscillations by Galerkin's Procedure", Journal of Mathematical Analysis and Applications, Vol. 14, 1966, pp. 107-140.

23) McLaughlin, R. J., "Error Bounds for Quasi-Harmonic Oscillations: Non-Resonant Case", Harvard University, Division of Engineering and Applied Physics, Technical Report No. 538, August 1967.

24) McLaughlin, R. J., "Error Bounds for Quasi-Harmonic Oscillations: Resonant Case", Harvard University Division of Engineering and Applied Physics, Technical Report No. 540, September 1967.

25) Holtzman, J. M., "Contraction Maps and Equivalent Linearization", The Bell System Technical Journal, December 1967, pp. 2405-2435.

26) Lazer, A. C., "On the Computation of Periodic Solutions of Weakly Nonlinear Differential Equations", SIAM Journal of Applied Mathematics, Vol. 15, No. 5, 1967, pp. 1158-1170.

27) Reference 10, pp. 151-152.

28) Byrd, P. F. and Friedman, M. D., Handbook of Elliptic Integrals for Engineers and Physicists, Springer-Verlag, Berlin, 1954, pp. 192-193.

29) Bellman, R., Introduction to Matrix Analysis, McGraw-Hill, New York, 1960, p. 162.

30) Hartman, P., Ordinary Differential Equations, John Wiley and Sons, New York, 1964, pp. 407-408.

31) Reference 1, p. 348.

32) Greenhill, A. G., The Applications of Elliptic Functions, Macmillan and Co., London, 1892, p. 51.

33) Stoker, J. J., Nonlinear Vibrations, Interscience Publishers, New York, 1950, pp. 83-90.

34) Reference 28, p. 215. (This equation has some numerical errors.)

35) Reference 4, p. 188.

36) Reference 28, p. 304.

37) Reference 29, pp. 2-9.