APPROXIMATIONS TO THE FREE RESPONSE OF A DAMPED NON-LINEAR SYSTEM

by

PAUL TSANG-LEUNG CHAN

B.A.Sc., University of British Columbia, 1962

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE

In the Department of

Electrical Engineering

We accept this thesis as conforming to the standards required from candidates for the degree of Master of Applied Science

Members of the Department
of Electrical Engineering

THE UNIVERSITY OF BRITISH COLUMBIA

March 1965

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the Head of my Department or by his representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Electrical Engineering

The University of British Columbia,
Vancouver 8, Canada

Date March 30, 1965

ABSTRACT

In the study of many engineering systems involving non-linear elements such as a saturating inductor in an electrical circuit or a hard spring in a mechanical system, we face the problem of solving the equation

$$\ddot{x} + 2\varepsilon\dot{x} + x + \mu x^3 = 0$$

which does not have an exact analytical solution. Because a consistent framework is desirable in the course of the study, we can assume that the initial conditions are $x(0) = 1$ and $\dot{x}(0) = 0$ without loss of generality. This equation is studied in detail by using numerical solutions obtained from a digital computer.

When ε and μ are small, classical methods such as the method of variation of parameters and averaging methods based on residuals provide analytical approximations to the equation and enable the engineer to gain useful insight into the system. However, when ε and μ are not small, these classical methods fail to yield acceptable results because they are all based on the assumption that the equation is quasi-linear. Therefore, two new analytical methods, namely: the parabolic phase approximation and the correction term approximation, are developed according to whether $\varepsilon < 1$ or $\varepsilon \geqslant 1$, and are proven to be applicable for values of ε and μ far beyond the limit of classical methods.

ACKNOWLEDGEMENT

Acknowledgement is due to all who have helped during the course of this work. In particular, the author wishes to thank Dr. A.C. Soudack, supervisor of the project, for his invaluable guidance and much-needed encouragement, Dr. F. Noakes, Head of the Department of Electrical Engineering, University of British Columbia, and the staff of the Computing Centre of the University of British Columbia, without whose help this work would have been impossible.

The author also wishes to express his gratitude to the National Research Council of Canada for a Bursary and a Studentship awarded him in 1962 and 1963 respectively.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

APPROXIMATIONS TO THE FREE RESPONSE OF A DAMPED NON-LINEAR SYSTEM

1. INTRODUCTION

1.1 Mathematical Models

In the study or analysis of physical systems, it is common practice to represent them by mathematical models. In order to make the mathematical models more tractable, certain simplifying assumptions are usually made. For example, the differential equations evolved are usually linearized, so that they may be solved by well established techniques used for linear differential equations. Most physical systems, however, behave in a manner which is far from linear, for example, a triode amplifier with large signal inputs or a mass, restrained by a non-linear spring, oscillating with large amplitudes. Therefore, non-linear analysis is required in order to yield results closer to reality.

1.2 Analytical Approximations

Exact solutions to non-linear differential equations are usually difficult, if not impossible, to find in closed form. Techniques for solving the equations vary according to the types of equations involved, and are very limited.

With the aid of digital computers, numerical solutions, to almost any degree of accuracy, to any non-linear differential equations are available. Using analogue computers, solutions to ordinary differential equations can be obtained. The

solutions obtained from these computers, however, do not furnish all the information concerning the physical system of interest, i.e. they usually reveal only the behaviour of the system under certain particular conditions. Exhaustive tests are needed if some insight into the system is required, and an engineer cannot necessarily predict from them how the system will behave if some parameters in the system are changed. The problem of cost and accessibility is another disadvantage in using computers as a means to solve a non-linear differential equation. For these reasons, analytical approximations to the solutions of ordinary non-linear differential equations are developed. These approximations are obtained in algebraic or transcendental form without the necessity of introducing numerical values for parameters or initial conditions during the process. Though some degree of accuracy is sacrificed, an over-all insight into the system is often obtained at a low cost. For instance, the dependence of the solution on a certain parameter may be explicit, thus yielding useful information for system design. A few well established approximate analytical methods[1] are

(a) Perturbation method,

(b) Variation of parameters,

(c) Averaging methods based on residuals, and

(d) Principle of harmonic balance.

Though these methods are developed to cover a very large class of non-linear differential equations, they have a common weakness in that they are incapable of dealing with equations exhibiting gross non-linearities. This limitation is due to the general approach to solving the equation, namely, making the grossly non-linear equation only slightly non-linear, or quasi-linear, in an

attempt to get more insight into the behaviour of the system, using linear theory. In order to break through this limitation, a bolder approach is in order, i.e. a direct attack on the non-linear equation in question. To this end, a study, for the purpose of obtaining approximate solutions to a certain type of grossly non-linear equation which arises from many engineering systems, was undertaken.

1.3 Derivation of the System Equation

A large class of physical systems contain a non-linear element whose characteristic is represented by an odd cubic polynomial with positive coefficients, for example, a hard spring characterized by

$$F(x) = a_1 x + a_3 x^3$$

where $F(x)$ = restoring force in spring,

x = displacement,

and a_1 and a_3 are positive coefficients. This odd cubic polynomial is often an approximation to a grosser non-linearity such as an odd polynomial of higher order, i.e.,

$$F(x) = \sum_{k \text{ odd}}^{k = n} a_k x^k, \qquad n > 3.$$

The general shapes of this odd polynomial and the "odd cubic" characteristic are shown in Fig. 1.1.

Now consider the following systems containing "odd cubic" elements:

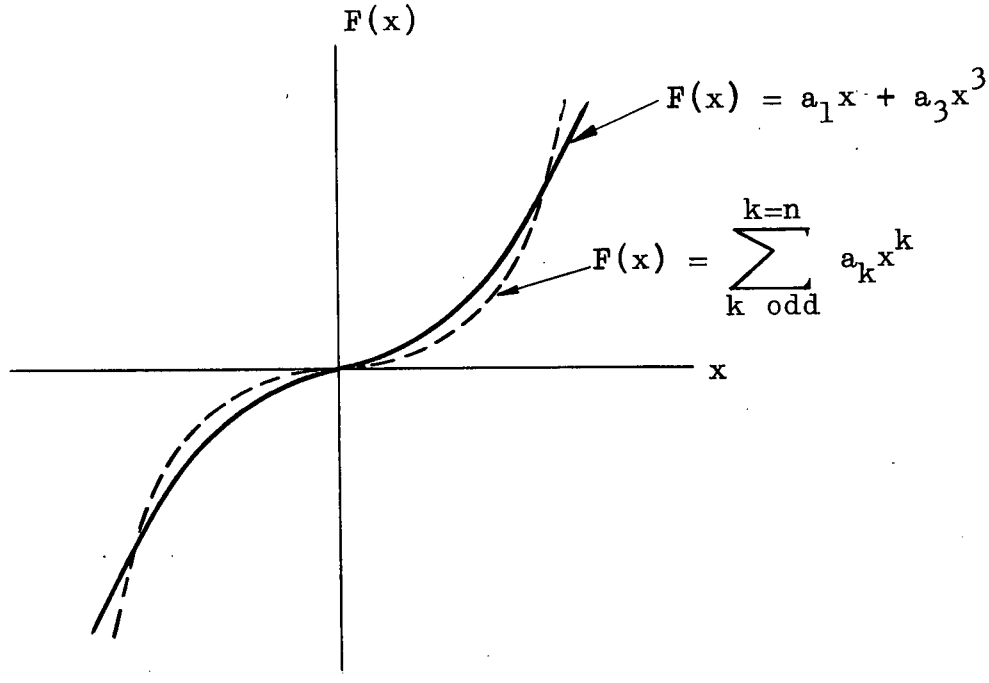


Fig. 1.1 Non-linear characteristics

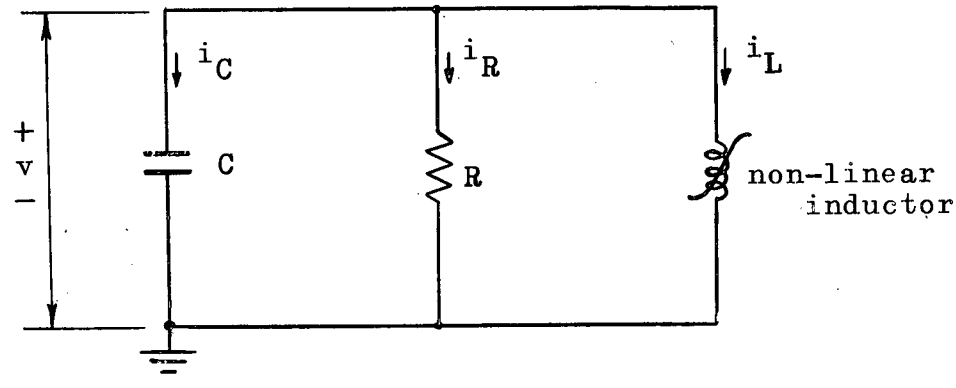


Fig. 1.2 Non-linear RLC circuit

(a) RLC circuit.

The parallel RLC circuit in Fig. 1.2 consists of a linear resistor R, a linear capacitor C, and an inductor which is non-linear because of saturation. Neglecting hysteresis, the inductor current can be approximated by

$$i_L = a_1\lambda + a_3\lambda^3$$

where λ = flux linkage, and a_1 and a_3 are positive coefficients. Applying Kirchhoff's current law, we obtain

$$C\dot{v} + \frac{v}{R} + a_1\lambda + a_3\lambda^3 = 0 \quad *$$

where v is the voltage across the parallel elements. Since, by Faraday's law, $v = \dot{\lambda}$, this equation can be re-written as

$$C\ddot{\lambda} + \frac{1}{R}\dot{\lambda} + a_1\lambda + a_3\lambda^3 = 0 \qquad (1.1)$$

which is, then, the equation of the RLC circuit.

(b) Hard spring with pure viscous damping.

Fig. 1.3(a) shows a simple mechanical system involving a mass sliding on a surface with pure viscous friction and restrained by a hard spring whose characteristic is given by

$$F = b_1x + b_3x^3$$

where F = restoring force in spring, and b_1 and b_3 are positive coefficients. From the free-body diagram shown in Fig. 1.3(b),

---

* $\dot{v} \triangleq \frac{dv}{dt}$

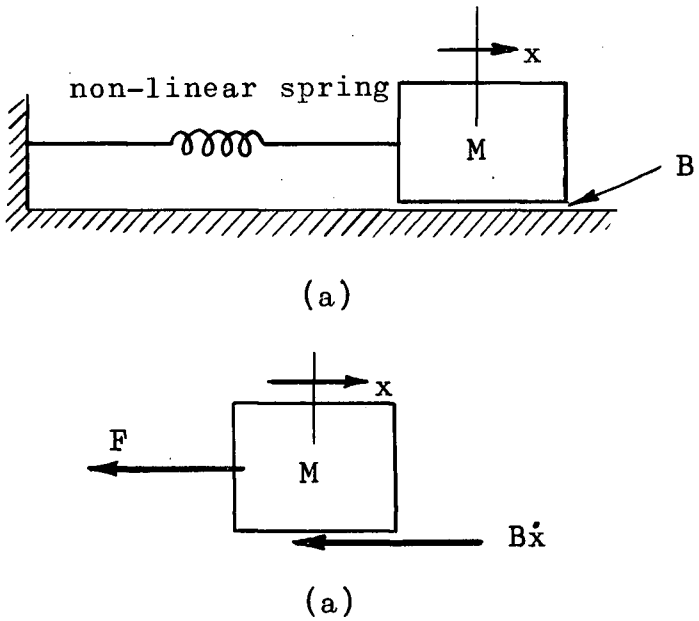


Fig. 1.3 Restrained sliding mass

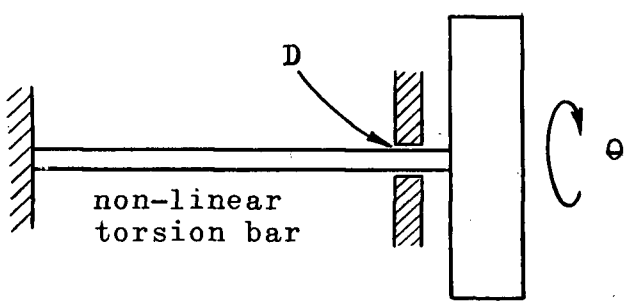


Fig. 1.4 Torsional pendulum

we obtain

$$F + B\dot{x} = -M\ddot{x}$$

where M = mass,

and B = coefficient of viscous friction, and hence

$$M\ddot{x} + B\dot{x} + b_1 x + b_3 x^3 = 0. \qquad (1.2)$$

(c) Torsional pendulum with pure viscous damping.

A simple torsional pendulum is illustrated in Fig. 1.4. It consists of a disc with moment of inertia J, a support with viscous frictional coefficient D, and a non-linear torsion bar whose characteristic is given by

$$T = c_1\theta + c_3\theta^3$$

where T = restoring torque in the torsion bar,

θ = angular deflection,

and c_1 and c_3 are positive coefficients. Consideration of torques gives

$$T + D\dot{\theta} = -J\ddot{\theta}$$

and the system equation becomes

$$J\ddot{\theta} + D\dot{\theta} + c_1\theta + c_3\theta^3 = 0. \qquad (1.3)$$

The systems described above are, in fact, analogous to each other, because their equations all have the following form

$$\alpha\ddot{y} + \beta\dot{y} + \gamma y + \delta y^3 = 0 \qquad (1.4)$$

where α corresponds to C, M, or J,

β corresponds to $\frac{1}{R}$, B, or D,

γ corresponds to a_1, b_1, or c_1,

δ corresponds to a_3, b_3, or c_3,

and y corresponds to ϕ, x, or θ.

Thus, a solution to equation (1.4) will provide a solution to all the above systems. If both the terms $\beta\dot{y}$ and δy^3 are relatively small, a classical method based on variation of parameters, such as the Kryloff-Bogoliuboff method, gives satisfactory analytical approximate solutions.[2],[3] However, if either $\beta\dot{y}$ or δy^3 is large, this method fails to yield good results. A detailed study of the equation where $\beta\dot{y}$ and δy^3 are <u>not</u> negligible and a direct approach to finding analytical approximate solutions was therefore attempted.

2. STUDY OF THE SYSTEM EQUATION

2.1 Normalization

Equation (1.4) ostensibly contains four arbitrary coefficients, which make it difficult to study. However, two of these coefficients can be made implicit if the following normalization is performed. Dividing through by α, equation (1.4) can be rewritten as

$$\ddot{y} + a\dot{y} + by + cy^3 = 0 \qquad (2.1)$$

where $a = \beta/\alpha$, $b = \gamma/\alpha$, and $c = \delta/\alpha$. Letting $\tau = \sqrt{b}\ t$, we obtain

$$\begin{aligned} \dot{y} &= \frac{dy}{dt} \\ &= \frac{dy}{d\tau} \cdot \frac{d\tau}{dt} \\ &= \sqrt{b}\ y' \end{aligned}$$

where $y' = \frac{dy}{d\tau}$,

and

$$\begin{aligned} \ddot{y} &= \frac{d^2y}{dt^2} \\ &= \sqrt{b}\ \frac{d}{d\tau}\ \left(\frac{dy}{dt}\right) \\ &= \sqrt{b}\ \frac{d}{d\tau}\ \left(\sqrt{b}\ y'\right) \\ &= b\ y'' , \end{aligned}$$

where $y'' = \frac{d^2y}{d\tau^2}$. Substituting y' and y'' into equation (2.1), we have

$$by'' + a\sqrt{b}\ y' + by + cy^3 = 0,$$

or

$$y'' + \frac{a}{\sqrt{b}}\ y' + y + \frac{c}{b}\ y^3 = 0.$$

In order to facilitate subsequent work, this equation is re-written in the form

$$y'' + 2\varepsilon y' + y + \mu y^3 = 0 \qquad (2.2)$$

where $\varepsilon = \frac{a}{2\sqrt{b}}$,

and $\mu = \frac{c}{b}$.

Moreover, since a consistent framework is desirable, this equation is assumed to have the following initial conditions:

$$y(0) = 1, \quad y'(0) = 0.$$

To show that this assumption does not affect the generality of the approach, let the initial conditions be

$$y(0) = Q_0, \quad y'(0) = 0.^*$$

Replacing y by $x = \frac{y}{Q_0}$ yields

$$x(0) = 1, \quad x'(0) = 0,$$

and equation (2.2) becomes

$$Q_0 x'' + 2\varepsilon Q_0 x' + Q_0 x + \mu Q_0^3 x^3 = 0$$

---

* $y'(0)$ is assumed to be zero in this work for physical reasons. For example, in the study of the sliding mass referred to in 1.3, one usually displaces it from its neutral position by the initial amount of Q_0, then releases it. Very seldom does one incorporate an initial velocity because it is difficult to obtain accurately. Moreover, in the case where the system is oscillatory, one can arbitrarily fix $t = 0$ at the peak of the oscillation, i.e. $y'(0) = 0$.

or
$$x'' + 2\varepsilon x' + x + \mu Q_o^2 x^3 = 0,$$
$$x(0) = 1, \quad x'(0) = 0 \tag{2.3}$$

A comparison between equations (2.2) and (2.3) reveals that the substitution of $x = \frac{y}{Q_o}$ leads to an equation with a different coefficient in the non-linear term if $Q_o \neq 1$; but since the method to be used for solving the equation is not altered by the values of this coefficient, no generality is lost.

2.2 Phase-plane Analysis

Consider the equation

$$\ddot{x} + 2\varepsilon\dot{x} + x + \mu x^3 = 0,$$
$$x(0) = 1, \quad \dot{x}(0) = 0. \tag{2.4}$$

If $\mu = 0$, it degenerates to the linear equation

$$\ddot{x} + 2\varepsilon\dot{x} + x = 0,$$
$$x(0) = 1, \quad \dot{x}(0) = 0. \tag{2.5}$$

which will be referred to as the complementary linear equation of (2.4). Exact solutions of this equation depend on the values of ε*, namely, if

(a) $\varepsilon < 1$ (underdamped), then

$$x = \frac{1}{\sqrt{1 - \varepsilon^2}} e^{-\varepsilon t} \cos\left(\sqrt{1 - \varepsilon^2}\, t + \phi_o\right), \tag{2.6}$$

---

* Because the system to be considered contains only passive elements, i.e. no energy sources, the value of ε will either be zero or positive.

where $\phi_o = \tan^{-1} \dfrac{\varepsilon}{\sqrt{1 - \varepsilon^2}}$,

(b) $\varepsilon = 1$ (critically damped), then

$$x = (1 + t)\, e^{-t}, \tag{2.7}$$

(c) $\varepsilon > 1$ (over-damped), then

$$x = \frac{1}{2}\left(1 + \frac{\varepsilon}{\sqrt{\varepsilon^2 - 1}}\right) e^{(-\varepsilon + \sqrt{\varepsilon^2 - 1})t} + \frac{1}{2}\left(1 + \frac{\varepsilon}{\sqrt{\varepsilon^2 + 1}}\right) e^{(-\varepsilon - \sqrt{\varepsilon^2 - 1})t} \quad . \tag{2.8}$$

If $\mu \neq 0$, solutions are also dependent on the damping factor ε. For example, if

(a) $\varepsilon = 0$ (conservative system), then equation (2.4) becomes

$$\ddot{x} + x + \mu x^3 = 0, \quad x(0) = 1, \quad \dot{x}(0) = 0.$$

The solution is the Jacobian elliptic cosine

$$x = Cn\ (k, \omega t)^{(4)}$$

where $k^2 = \dfrac{\mu}{2(1+\mu)}$,

and $\omega^2 = 1 + \mu$.

(b) $\varepsilon > 0$ (non-conservative system), then no exact solutions to equation (2.4) are available in closed form. Though the solutions are similar to those of equation (2.5), they exhibit a greater number of oscillations in the same length of time. In order to clarify the picture, a phase-plane analysis is most

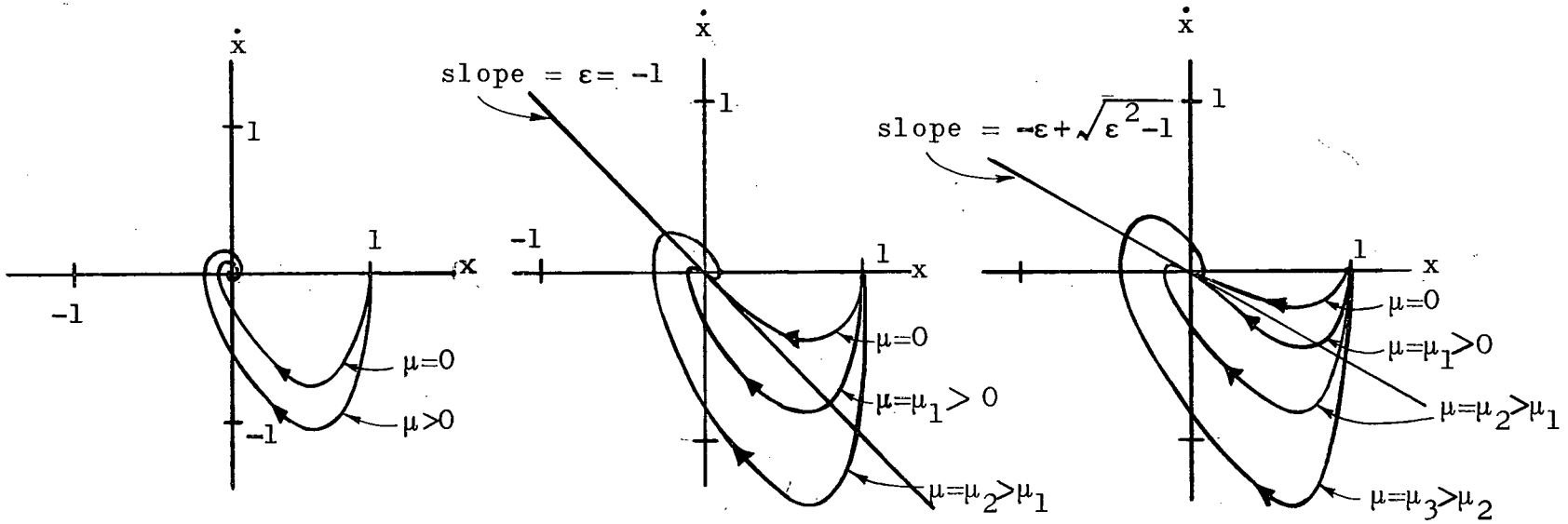
helpful. Using the method of isoclines[5], phase-plane diagrams of equations (2.4) and (2.5), as shown in Fig. 2.1, are obtained. In Fig. 2.1(a) where $\varepsilon < 1$, the two trajectories suggest damped oscillations. In the case where $\mu > 0$, the period of oscillation is shorter because x decreases with a greater slope. In Fig. 2.1(b) where $\varepsilon = 1$, the trajectory for $\mu = 0$ represents a solution without "overshoot", i.e. x never going negative, while trajectories for $\mu > 0$ show one or more overshoots, the number of which increases as μ increases. Similar results are observed in Fig. 2.1(c) where $\varepsilon > 1$. If under-damping and over-damping are defined respectively by the presence and absence of overshoots, one sees, therefore, that consideration of ε alone is not sufficient to predict whether the system is under-damped or over-damped, as in the linear case, for the value of μ is also an important factor. An extensive investigation of this aspect was undertaken, using the digital computer to provide numerical solutions.* The result, as illustrated in Fig. 2.2, is a curve in the ε-μ plane showing regions where the system has overshoot and where it has not. Contrary to linear theory, overshoots may be observed for $\varepsilon > 1$, if μ is high enough. This result is not surprising as equivalent linearization also predicts possible overshoots.

2.3 Investigations in the Time-domain

The independent variable, time, is implicit in phase-plane diagrams, and solutions to equation (2.4) are, therefore, not readily available as functions of time. Using the method of equivalent linearization[6] and numerical solutions obtained from the digital computer, the effect of μ on the solution is as follows:

---

* See Appendix A for computational details.



(a) $0<\varepsilon<1$ (b) $\varepsilon = 1$ (c) $\varepsilon>1$

Fig. 2.1 Phase-plane diagrams

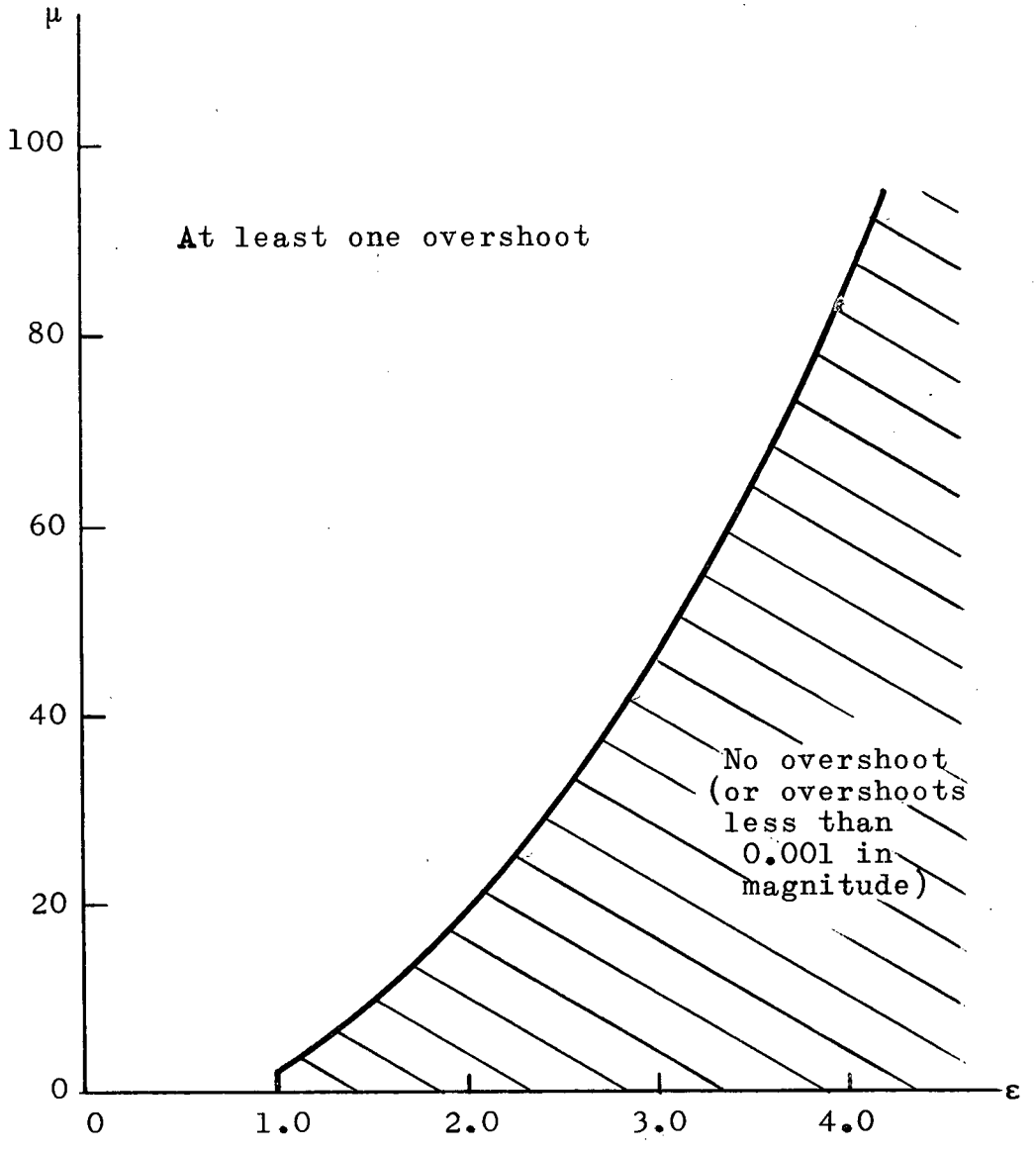


Fig. 2.2 Dependence of overshoot on ε and μ

(a) $0 < \varepsilon < 1$

Fig. 2.3 shows a typical example in this case. A numerical solution to the equation

$$\ddot{x} + 0.4\dot{x} + x + 2x^3 = 0,\ x(0) = 1,\ \dot{x}(0) = 0 \qquad (2.9)$$

is displayed together with the solution to its complementary linear equation, i.e.

$$\ddot{x} + 0.4\dot{x} + x = 0,\ x(0) = 1,\ \dot{x}(0) = 0 \qquad (2.10)$$

Any difference between these two curves reflects the effect of the non-linear term. The solution to equation (2.10) is

$$x = \frac{1}{\sqrt{0.96}}\ e^{-0.2t}\ \cos\ (\sqrt{0.96}t - \phi_o)$$

where $\phi_o = \tan^{-1} \frac{0.2}{\sqrt{0.96}}$. It represents a damped sinusoid having an envelope $\frac{1}{\sqrt{0.96}}\ e^{-0.2t}$ and a phase of $\sqrt{0.96}\ t - \phi_o$. Although the solution to equation (2.9) resembles a damped sinusoid, its amplitude decays at a slower rate than $\frac{1}{\sqrt{0.96}}\ e^{-0.2t}$ and its phase increases in a non-linear manner, i.e. the phase is retarded as time increases.

Equivalent linearization of equation (2.4), based on variation of parameters[6], yields

$$\ddot{x} + 2\varepsilon\dot{x} + (1 + \frac{3\mu A^2}{8})^2\ x = 0 \qquad (2.11)$$

where $A = \frac{-\dot{A}}{\varepsilon}$

= amplitude.

Here, the value of A at $t = 0$ must not be taken as $x(0)$, for it

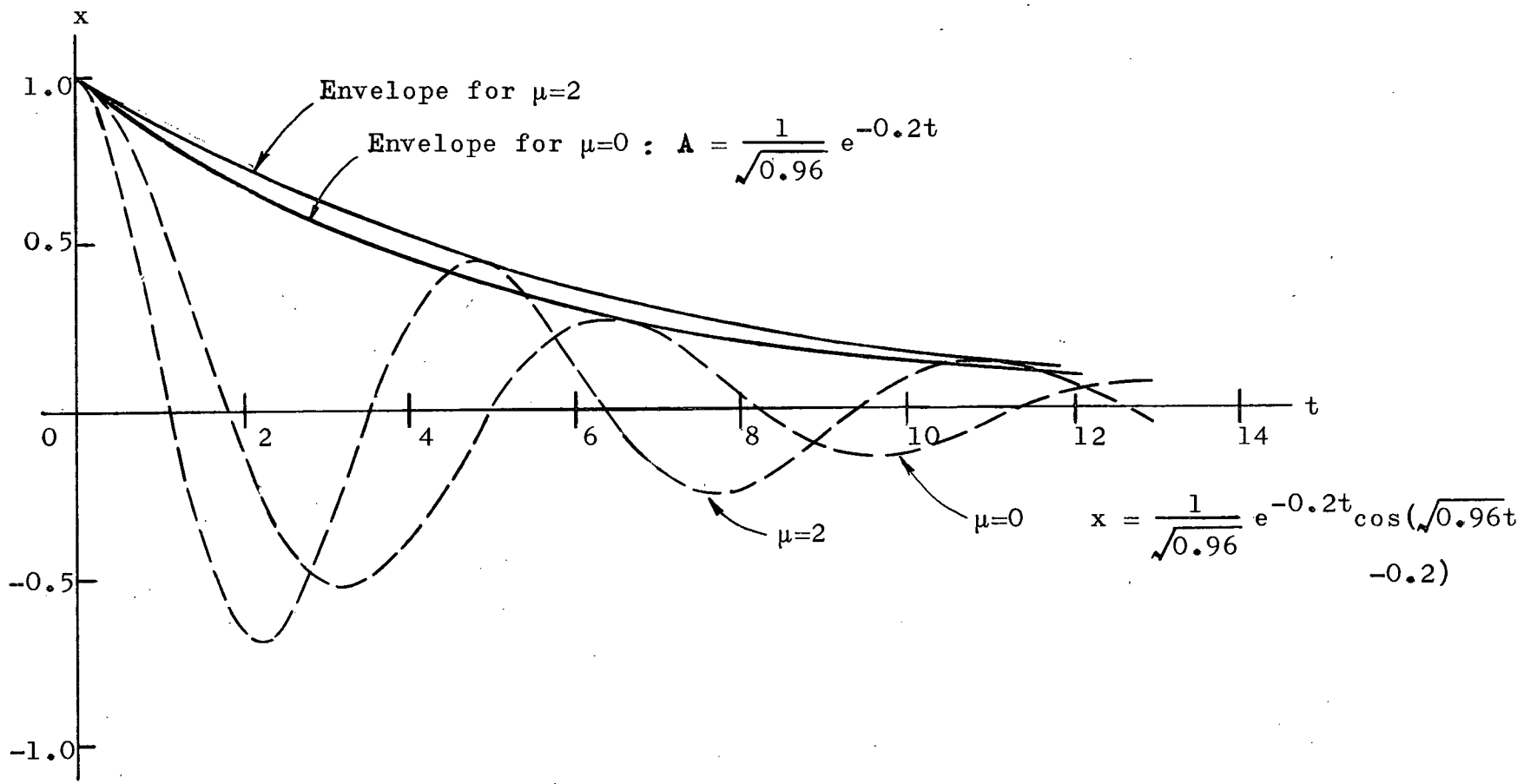


Fig. 2.3 Solution curves to the equation

$$\ddot{x} + 0.4\dot{x} + x + \mu x^3 = 0, \quad x(0) = 1, \quad \dot{x}(0) = 0$$

represents the envelope of the solution and is always greater than x(0), though in most cases the difference is small. Using linear theory, the effect of μ on the frequency is evident. This equation represents an oscillation with both amplitude and frequency varying with respect to time. The amplitude A decays exponentially according to $e^{-\varepsilon t}$, and the frequency similarly decreases as the amplitude decays with increasing time. As time progresses, A ultimately becomes small enough so that $\frac{3\mu A^2}{8}$ is negligible compared to unity. Then, the frequency becomes effectively $\sqrt{1 - \varepsilon^2}$, which is the frequency of oscillation of the complementary linear equation of equation (2.4). Therefore, equation (2.4) degenerates to its complementary linear equation as time increases.

(b) $\varepsilon \geqslant 1$

A typical example in this case is illustrated in Fig. 2.4 by

$$\ddot{x} + 2.4\dot{x} + x + 7x^3 = 0, \; x(0) = 1, \; \dot{x}(0) = 0.$$

Although this equation has an over-damped complementary linear equation, an overshoot is observed in the solution. The presence of the non-linear term is responsible for this overshoot as already shown in the phase-plane analysis. Consistent results are also predicted from equivalent linearization.

Consider now the equivalent linear equation (2.11) where $\varepsilon \geqslant 1$. If μ has a value such that

$$1 + \frac{3\mu A^2}{8} > \varepsilon,$$

oscillatory solutions may be obtained. It must be noted, however,

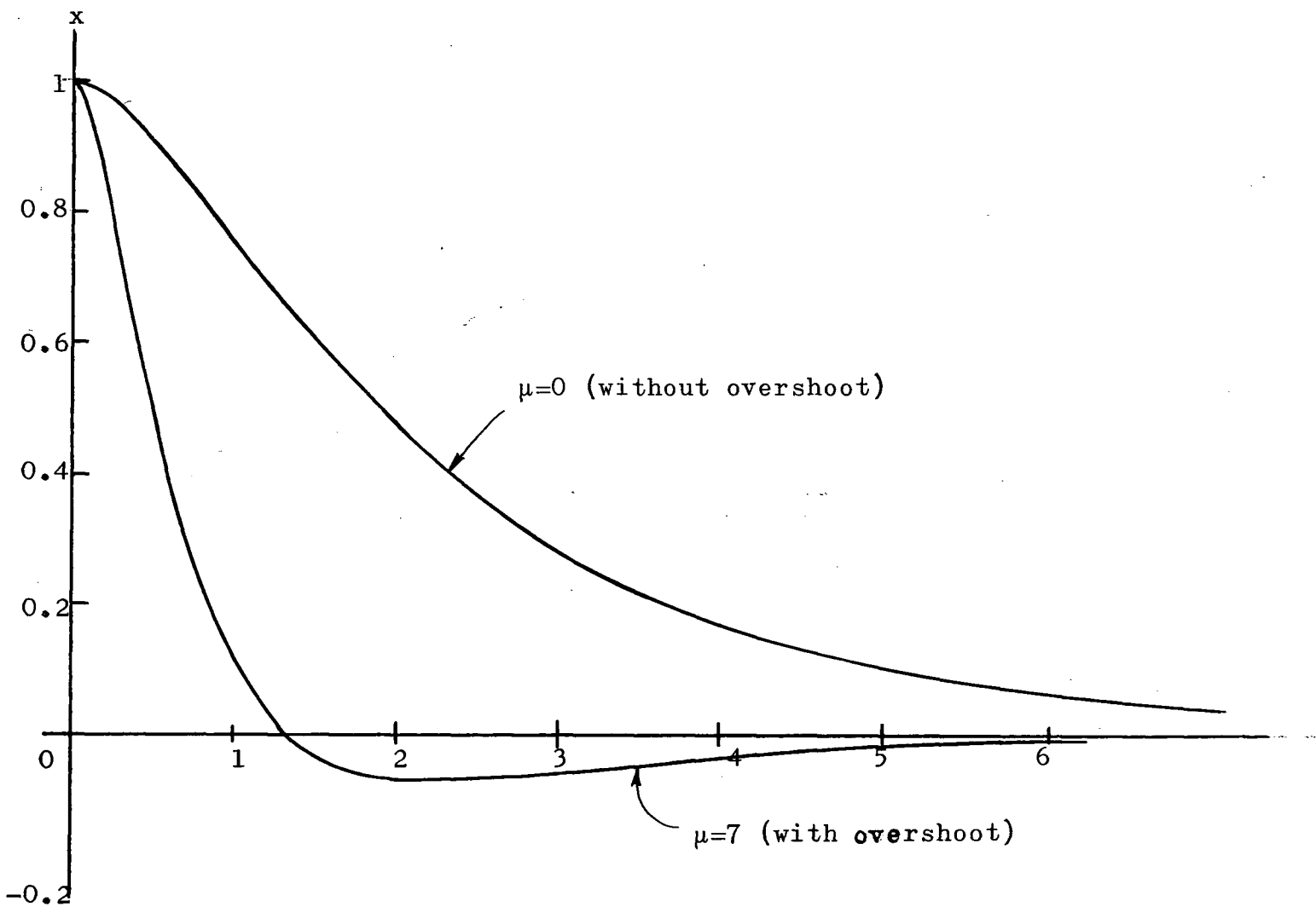


Fig. 2.4 Solution curves to the equation

$$\ddot{x} + 2.4\dot{x} + x + \mu x^3 = 0, \quad x(0) = 1, \quad \dot{x}(0) = 0$$

that this inequality only predicts qualitatively how overshoots may possibly occur, and is not necessarily capable of yielding accurate results, for small values of ε and μ have been assumed in the equivalent linearization. If accurate results are required, then the curve in Fig. 2.2 may be used.

2.4 Conclusions

In conclusion, the general behaviour of the system has been studied, using the digital computer and the method of phase plane analysis. Solutions to the non-linear equation are compared with solutions to its complementary linear equation. In damped oscillatory systems, the amplitude of the solution decreases more slowly in the non-linear case, and its frequency, being initially greater, approaches that of the complementary linear equation as time progresses. In the case where the complementary linear equation is critically or over-damped, the presence of the non-linear term may lead to over-shoots.

Because the system is damped, x will eventually disappear, and the non-linear term μx^3 will become insignificant compared to x, when x becomes small. Hence, an approach to finding the analytical solution is suggested by neglecting the non-linear term at a point where x has become sufficiently small.

3. APPROXIMATE SOLUTIONS TO THE SYSTEM EQUATION

3.1 Motivation

Consider equation (2.4)

$$\ddot{x} + 2\varepsilon\dot{x} + x + \mu x^3 = 0,\ x(0) = 1,\ \dot{x}(0) = 0.$$

As mentioned previously, the solution to this equation is not available in closed form. Therefore, attempts were made to approximate the solution.

In the case where $\varepsilon < 1$, the method of Kryloff and Bogoliuboff*(3) was used. From Appendix B, the solution is given by

$$\tilde{x}(t) = e^{-\varepsilon t} \cos\ (1 + \tfrac{3\mu}{8})t, \qquad (3.1)$$

where $\varepsilon, \mu \ll 1$.

By comparison with

$$x(t) = \frac{1}{\sqrt{1 - \varepsilon^2}}\ e^{-\varepsilon t} \cos\ (\sqrt{1 - \varepsilon^2}\,t + \phi_o) \qquad (3.2)$$

which is the solution to the complementary linear equation (2.4), one observes that solution (3.1) cannot be extended to higher values of ε and μ, because

(a) contrary to the result in the last chapter, the frequency in (3.1) remains constant as time progresses, and

(b) this frequency approaches unity if μ becomes zero, which does not agree with (3.2) if ε is not negligible.

For large values of ε and μ, therefore, a new method must be

---

*This method will hereafter be referred to as the K-B method. See Appendix B.

developed.

In the case where $\varepsilon \geqslant 1$, the K-B method is no longer applicable because the solution is not oscillatory in nature. Here, the Ritz method[7], which is an averaging method based on residuals is used in conjunction with initial condition matching. Since the solution of interest is either monotonically decreasing or exhibits one overshoot, the approximate solution is assumed to be of the following form:

$$\tilde{x}(t) = A\, e^{at} + B\, e^{bt}$$

where A and B are constants and both a and b are different negative numbers. Substituting this solution into the original differential equation, we have the residual given by

$$\begin{aligned}\sigma(t) &= \ddot{\tilde{x}} + 2\varepsilon\dot{\tilde{x}} + \tilde{x} + \mu\tilde{x}^3 \\ &= (a^2 + 2\varepsilon a + 1)Ae^{at} + (b^2 + 2\varepsilon b + 1)Be^{bt} + \mu A^3 e^{3at} \\ &\quad + \mu B^3 e^{3bt} + 3\mu A^2 B e^{(2a+b)t} + 3\mu AB^2 e^{(a+2b)t} \quad .\end{aligned}$$

The Ritz criteria are

$$\int_0^\infty \sigma(t) e^{at}\, dt = 0,$$

and

$$\int_0^\infty \sigma(t) e^{bt}\, dt = 0.$$

Now, from the initial conditions $\tilde{x}(0) = 1$, $\dot{\tilde{x}}(0) = 0$, we also have

$$A + B = 1$$

$$a\ \ A + b\ B = 0$$

Hence we obtain four equations in four unknowns. After integrating

and eliminating A and B from these equations, we have

$$1.\quad \frac{-b}{2a}(a^2+2\varepsilon a+1)+\frac{a}{a+b}(b^2+2\varepsilon b+1)-\frac{\mu b^3}{4a(a-b)^2}$$

$$+\frac{\mu a^3}{(3b+a)(a-b)^2}+\frac{3\mu b^2 a}{(3a+b)(a-b)^2}-\frac{3\mu a^2 b}{2(a+b)(a-b)^2}=0$$

$$2.\quad \frac{-a}{2b}(b^2+2\varepsilon b+1)+\frac{b}{a+b}(a^2+2\varepsilon a+1)-\frac{\mu a^3}{4b(a-b)^2}$$

$$+\frac{\mu b^3}{(3a+b)(a-b)^2}+\frac{3\mu a^2 b}{(3b+a)(a-b)^2}-\frac{3\mu a b^2}{2(a+b)(a-b)^2}=0\ .$$

Without the aid of a digital computer, solving the above two equations simultaneously is very laborious. Therefore, from an engineer's point of view, this method is highly impractical.

As a result, two new approximating methods were developed, depending on whether ε is less than unity or greater than unity. The rest of this work will be devoted to the development of these methods.

3.2 Case I - $\varepsilon < 1$

3.2.1 Choice of Approximant

In this case, where $\varepsilon < 1$, both phase-plane analysis and computer solutions from Sections (2.2) and (2.3) have indicated oscillatory solutions resembling damped sinusoids. The phase has also been shown to increase non-linearly with time. The approximant, therefore, will assume the form

$$x(t) = A(t)\cos\Omega(t),$$

where $A(t)$ = amplitude,

and $\Omega(t)$ = phase.

Consider equation (2.4) with $\varepsilon < 1$. Because the system is damped, x will eventually vanish and the non-linear term μx^3 will become negligible compared to x, when x is sufficiently small. Let this point of negligibility occur at $t = t_m$. Beyond this point, then, equation (2.4) essentially degenerates to its complementary linear equation (2.5) and will be treated as such. Therefore, the approximant will have the following form:

for $0 < t \leqslant t_m$ $\qquad x(t) = A(t) \cos \Omega(t)$ $\qquad (3.3)$

for $t \geqslant t_m$ $\qquad x(t) = P\, e^{-\varepsilon t} \cos \left(\sqrt{1 - \varepsilon^2}\, t + \phi_o \right)$ $\qquad (3.4)$

where P and ϕ_o are constants.

Here, it must be noted that $P \neq \frac{1}{\sqrt{1 - \varepsilon^2}}$ and $\phi_o \neq \tan^{-1} \frac{\varepsilon}{\sqrt{1 - \varepsilon^2}}$, as they are in equation (2.6), because initial conditions for (3.3) must be adjusted to match (3.2) at t_m. And so, the problem is now to find the functions $A(t)$ and $\Omega(t)$, and the constants P, ϕ_o, and t_m.

3.2.2 The Angle Criterion and the Determination of t_m

In order to determine t_m, some information about the point at which μx^3 may be neglected is necessary. Consider, again, the phase-plane diagram. Usually the first step in the construction of the phase-plane diagram is the construction of isoclines, i.e. curves of constant slope in the x-$\dot{x}$ plane. Therefore, if two systems have almost identical sets of

isoclines, they must have almost the same phase-trajectories. Furthermore, if two systems have almost identical phase-trajectories, it is reasonable to assume that their solutions as functions of time are almost identical. Hence a measure of "closeness" between the isoclines of two systems may be regarded as a measure of how close their solutions are to each other.

Now, consider the two systems represented by the equations (2.4) and (2.5). Typical isoclines of the same slope m for these systems are shown in Fig. 3.1. In order to have a measure of "closeness" between these isoclines, a circle of radius R is constructed, intersecting the linear isocline at point P. Through P, a vertical straight line is drawn, intersecting the non-linear isocline at point Q. Then, the angle $\delta\theta$ between the lines OP and OQ can be regarded as a measure of "closeness" between the two isoclines. If a slope different from m is chosen, and the same construction performed, the resulting angle $\delta\theta$ may be different. However, the maximum value $(\delta\theta)_{max}$ of these angles, as m varies, is a function of μR^2.* In particular, $(\delta\theta)_{max}$ decreases as μR^2 decreases, but since the amplitude A of the solution follows R quite closely,* $(\delta\theta)_{max}$ decreases as μA^2 decreases. Therefore, the value of μA^2 can also be regarded as a measure of "closeness" between the two isoclines, or a measure of the effect of the non-linear term. For example, if $\mu A^2 = 0.2$, $(\delta\theta)_{max}$ is approximately 3^o, or 0.05 radian. Therefore, as the amplitude A decays from its initial value and reaches a value such that $\mu A^2 = 0.2$, the two sets of isoclines may be considered coincident for all practical

---

* See Appendix C.

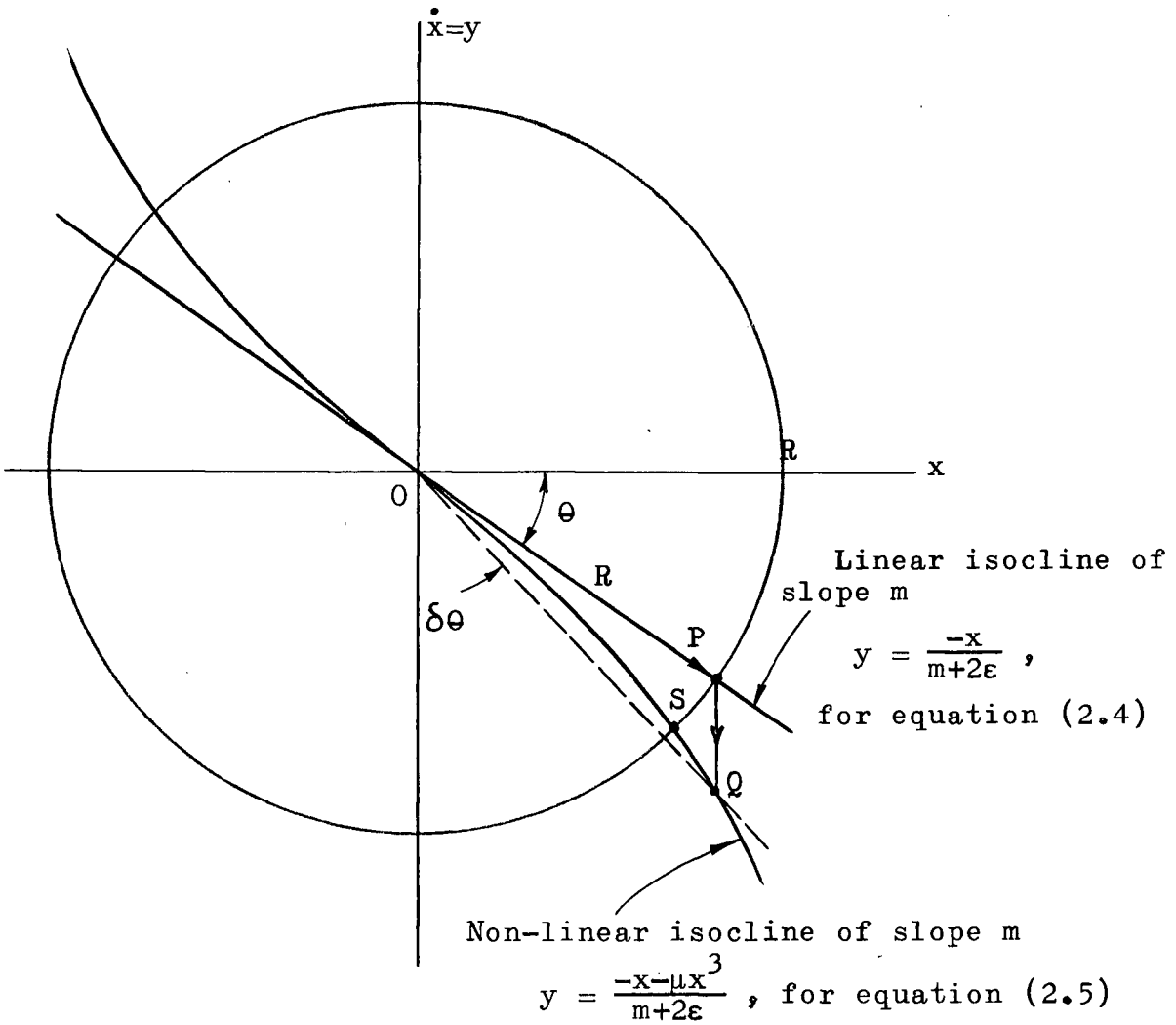


Fig. 3.1 Difference between the linear and non-linear isoclines of the same slope, m

purposes. This point then determines a point at which the term μx^3 may be neglected. Hence t_m may be chosen such that

$$\mu\left[A(t_m)\right]^2 = 0.2.$$

It must be remembered that this relation is an arbitrary criterion based on the consideration of the angle $(\delta\theta)_{max}$, and it serves only the purpose of obtaining a point where the non-linear equation can be replaced by the linear one. In choosing such a point of transistion, how the value of μA^2 varies with time must also be considered.

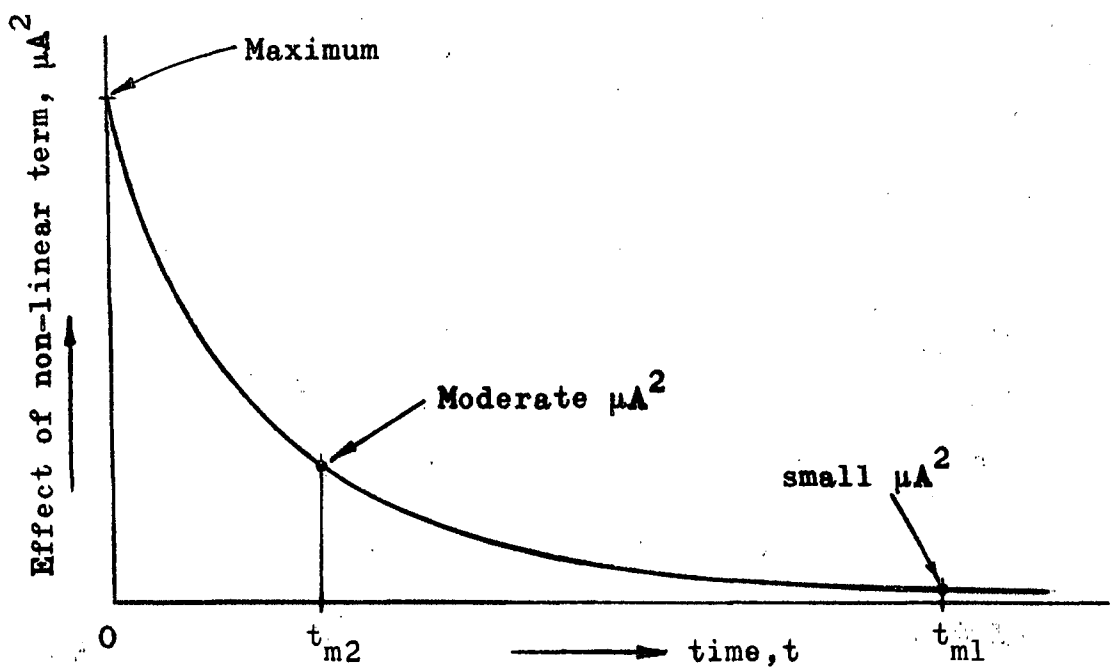


Fig. 3.2 Effect of the Non-linear Term

Fig. 3.2 shows the general shape of the value of μA^2 as time varies. Since the amplitude A decays almost exponentially, as mentioned before, μA^2 drops very quickly at the beginning and approaches zero asymptotically as time increases. This

curve also represents the effect of the non-linear term. In order to show that consideration of the value of μA^2 alone may not necessarily lead to a wise choice of the point of transition from the non-linear equation to a linear one, the following two choices of such a point are compared. First, let the transition occur at $t_m = t_{m1}$. The value of μA^2 is small as shown in Fig. 3.2, indicating that the equation is essentially linear for $t = t_m = t_{m1}$, and therefore the "linear" part of the approximant, i.e. equation (3.4):

$$\text{for } t \geqslant t_m \qquad \tilde{x}(t) = P\, e^{-\varepsilon t} \cos\left(\sqrt{1-\varepsilon^2}\, t + \phi_o\right),$$

is very close to the exact solution. As a second choice let the transition occur much earlier, at $t_m = t_{m2}$. The value of μA^2 is now larger, and the "linear" part of the approximant is therefore not as good as the first choice. This does not necessarily mean, however, that the second choice is poorer, for it may yield a better "non-linear" part of the approximant, i.e. equation (3.3):

$$\text{for } 0 < t \leqslant t_m \qquad \tilde{x}(t) = A(t)\ \mathrm{Cos}\ \Omega(t).$$

In fact, a better "non-linear" part is usually expected because its range of approximation is now greatly reduced. As a result, one must consider a compromise in choosing the point t_m, so that both the "linear" and "non-linear" parts of the approximant are reasonably accurate. To this end, one must choose t_m as small as possible and at the same time, avoid an unduly large value of $\mu\left[A(t_m)\right]^2$. Using Galerkin's method, which is an averaging method based on residuals[8], it seems that the

value of t_m may be optimized in the sense that the integral

$$J = \int_0^\infty \sigma^2(t)\, dt$$

is a minimum, where $\sigma(t) = \ddot{\tilde{x}} + 2\varepsilon\dot{\tilde{x}} + \tilde{x} + \mu\tilde{x}^3$. However, this is impractical because (a) the functions $A(t)$ and $\Omega(t)$ are not known, and (b) even if they are known, solving the set of equations

$$\frac{\partial}{\partial p}\int_0^\infty \sigma^2(t)\, dt = 0$$

$$\frac{\partial}{\partial \phi_o}\int_0^\infty \sigma^2(t)\, dt = 0$$

$$\frac{\partial}{\partial t_m}\int_0^\infty \sigma^2(t)\, dt = 0$$

will be a formidable task due to the presence of the non-linear term. Therefore, experimental results are used to obtain an empirical criterion for choosing an acceptable t_m. For example, numerical solutions of the equation

$$\ddot{x} + 0.4\dot{x} + x + \mu x^3 = 0, \quad x(0) = 1, \quad \dot{x}(0) = 0$$

with various values of μ have indicated that the phase-retardation becomes negligible when $\mu A^2 = 0.2$. This means that the equation behaves practically like a linear one when $\mu A^2 = 0.2$, and therefore t_m can be chosen such that

$$\mu\left[A(t_m)\right]^2 = 0.2. \qquad (3.5)$$

From this empirical criterion, as μ becomes larger, $A(t_m)$ becomes smaller, yielding a greater t_m. This is reasonable because with a larger μ, the non-linear term μx^3 must take a longer time to become negligible compared to x. It must also be noted here that the relation (3.5) is obtained empirically with $\varepsilon = 0.2$. If equations with a larger ε are considered, however, a different empirical criterion may be obtained. In fact, study of the numerical solution of the equation

$$\ddot{x} + 0.8\,\dot{x} + x + \mu x^3 = 0, \quad x(0) = 1, \quad \dot{x}(0) = 0$$

has revealed that the phase-retardation becomes negligible at t_m given by

$$\mu\left[A(t_m)\right]^2 = 0.4.$$

From Fig. 3.2, this indicates that the effect of the non-linear term is greater at t_m, but it should be noted that a larger ε results in a faster decay of the amplitude A. Because the effect of the non-linear term varies as A^2, this means that μx^3 becomes negligible compared to x in a much shorter time interval. Therefore it is conceivable to relax the criterion. Many examples with various values of ε have been solved numerically and the result has suggested that this criterion can be assumed to depend on ε in the following linear manner:

$$\mu\left[A(t_m)\right]^2 = \frac{\varepsilon}{2}\,. \tag{3.6}$$

Note that this relation is only an empirical criterion to be used as a "rule of thumb" for choosing t_m <u>wisely</u>, and is not necessarily the <u>best</u> criterion, if one exists at all. Now, the next step is to evaluate t_m, using this criterion.

Although this criterion gives a value of $A(t_m)$ when ε and μ are specified, it does not provide the value of t_m directly. It is necessary to find a relationship between $A(t_m)$ and t_m. Again, consider equation (2.4)

$$\ddot{x} + 2\varepsilon\dot{x} + x + \mu x^3 = 0, \quad x(0) = 1, \; \dot{x}(0) = 0.$$

It has been shown earlier in section 2.3(a) that the envelope $A(t)$ of the solution decreases at a slower rate with $\mu \neq 0$ than with $\mu = 0$, as illustrated in Fig. 3.3 by curves I and II. A horizontal line is drawn through $A(t) = A(t_m)$, intersecting curves I and II at F and G respectively. Therefore, the abscissa for G is t_m, and if the abscissa for F is denoted by t_{mo}, we have

$$A(t_m) = \frac{1}{\sqrt{1 - \varepsilon^2}} \, e^{-\varepsilon t_{mo}} \quad ,$$

which gives

$$t_m = \frac{-1}{\varepsilon} \log_e \left[\sqrt{1 - \varepsilon^2} \; A(t_m)\right] .$$

Letting t_o be the interval between F and G, then

$$\begin{aligned} t_m &= t_{mo} + t_o \\ &= \frac{-1}{\varepsilon} \log_e\left[\sqrt{1 - \varepsilon^2} \; A(t_m)\right] + t_o . \end{aligned} \qquad (3.7)$$

Hence the problem becomes finding t_o in terms of $A(t_m)$, μ and ε, which are all the known quantities. Here, it may seem that the introduction of t_o does not help in solving the problem at all, because the original problem was to find t_m also in terms of these three known quantities. But this is not the case, because by finding t_o, one is looking only for that part

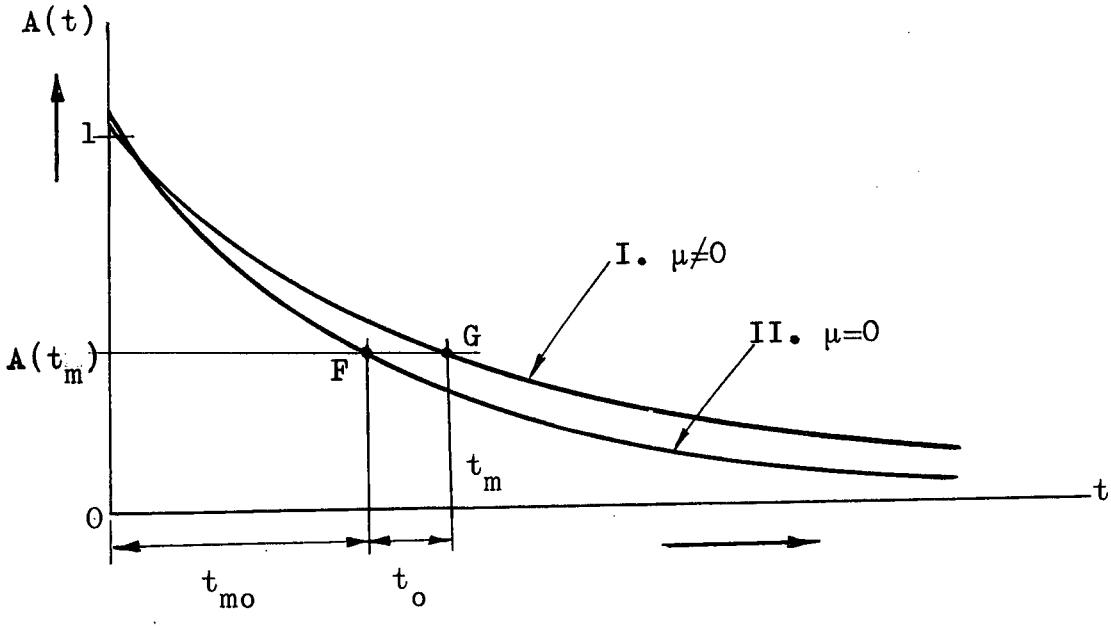


Fig. 3.3 Envelope of the solution to equation (2.4) i.e., $\ddot{x} + 2\varepsilon\dot{x} + x + \mu x^3 = 0$

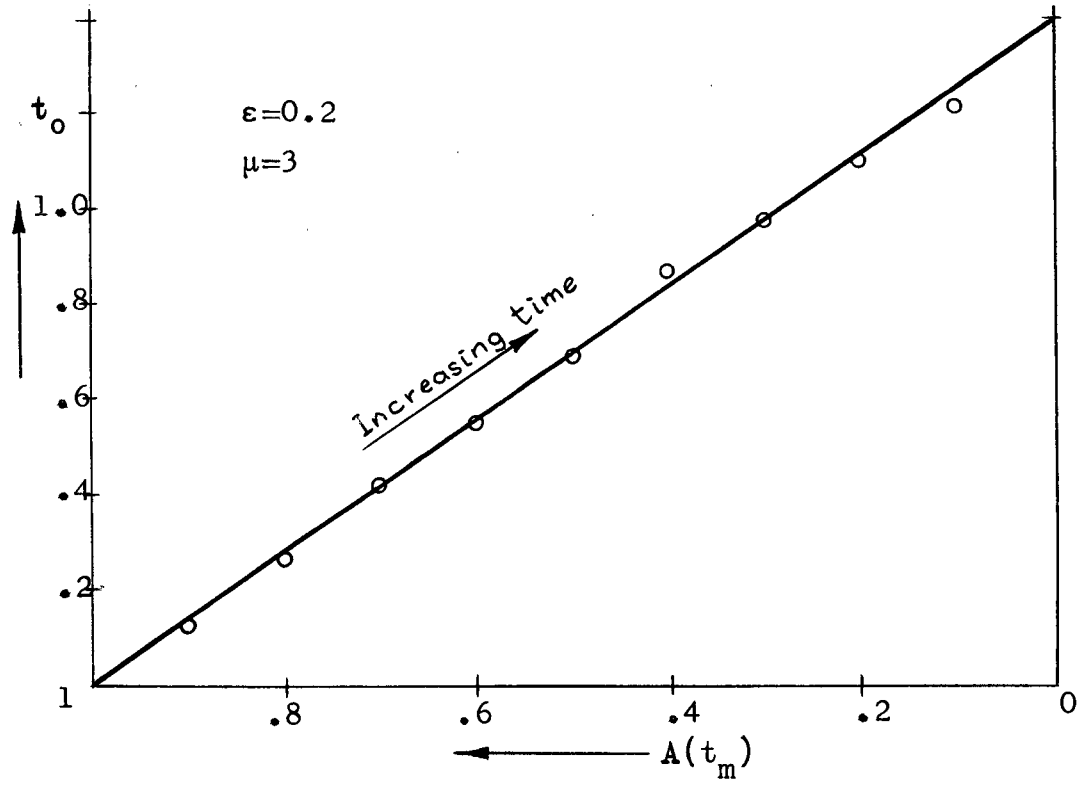


Fig. 3.4 Dependence of t_o on $A(t_m)$

of t_m due to the non-linear term, i.e. only the effect of the non-linear term on the envelope.

The approach in finding t_o is to investigate how t_o depends on the three quantities $A(t_m)$, μ and ε respectively. Here, experimental results are again used. First, the dependence of t_o on $A(t_m)$ is illustrated by Fig. 3.4, in which the numerical solutions to the equations

$$\ddot{x} + 0.4\,\dot{x} + x = 0, \qquad x(0) = 1, \qquad \dot{x}(0) = 0$$

and $$\ddot{x} + 0.4\,\dot{x} + x + 3x^3 = 0, \quad x(0) = 1, \quad \dot{x}(0) = 0$$

are used. As $A(t_m)$ decreases from its initial value,* t_o increases fairly linearly and reaches a maximum at $A(t_m) = 0$. Hence we have

$$t_o \propto 1 - A(t_m) \tag{3.8}$$

Secondly, in order to reveal the dependence of t_o on μ, the numerical solutions to the equation

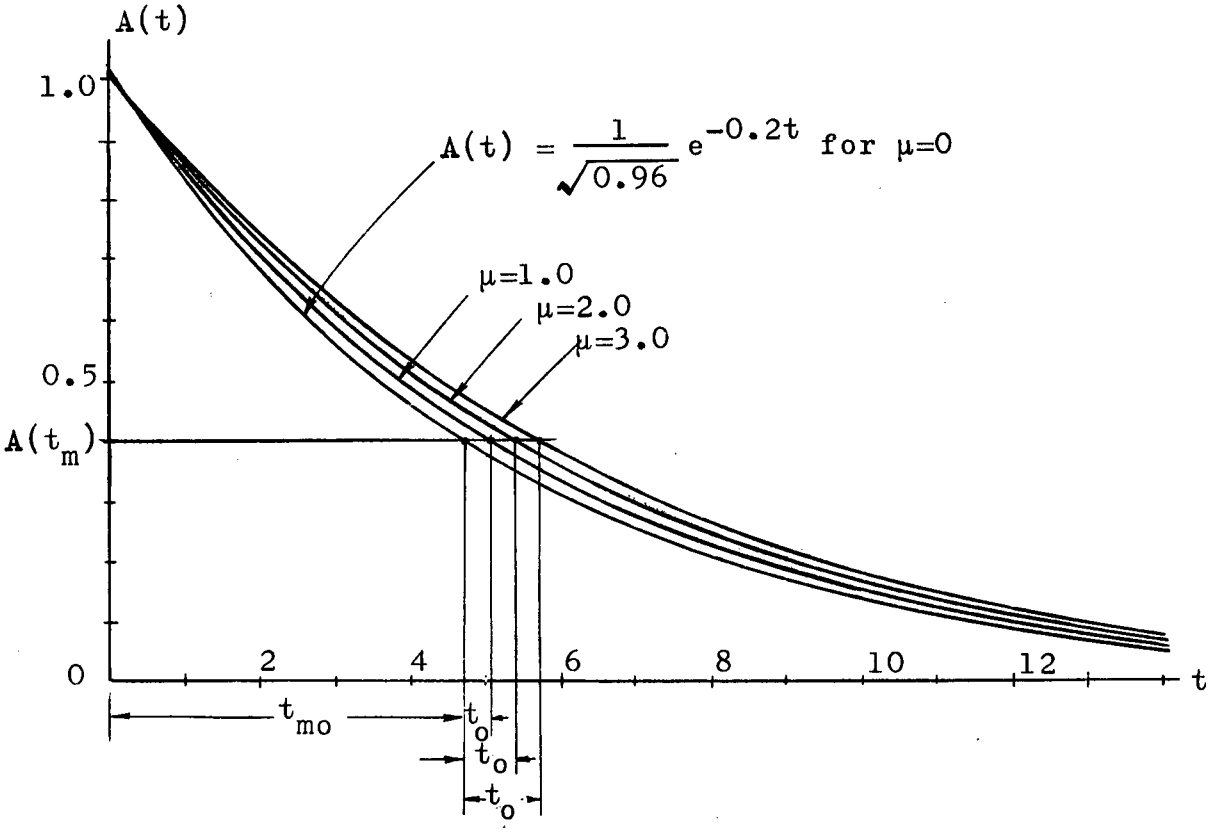
$$\ddot{x} + 0.4\,\dot{x} + x + \mu x^3 = 0, \quad x(0) = 1, \quad \dot{x}(0) = 0$$

are shown in Fig. 3.5(a) as μ varies from 0 to 3 in increments of 1.0. A fixed value of $A(t_m)$ is chosen, and values of t_o, corresponding to particular values of μ are found. If t_o is now plotted against μ, a straight line is obtained, as in Fig. 3.5(b). Thus, t_o varies approximately linearly as μ, or

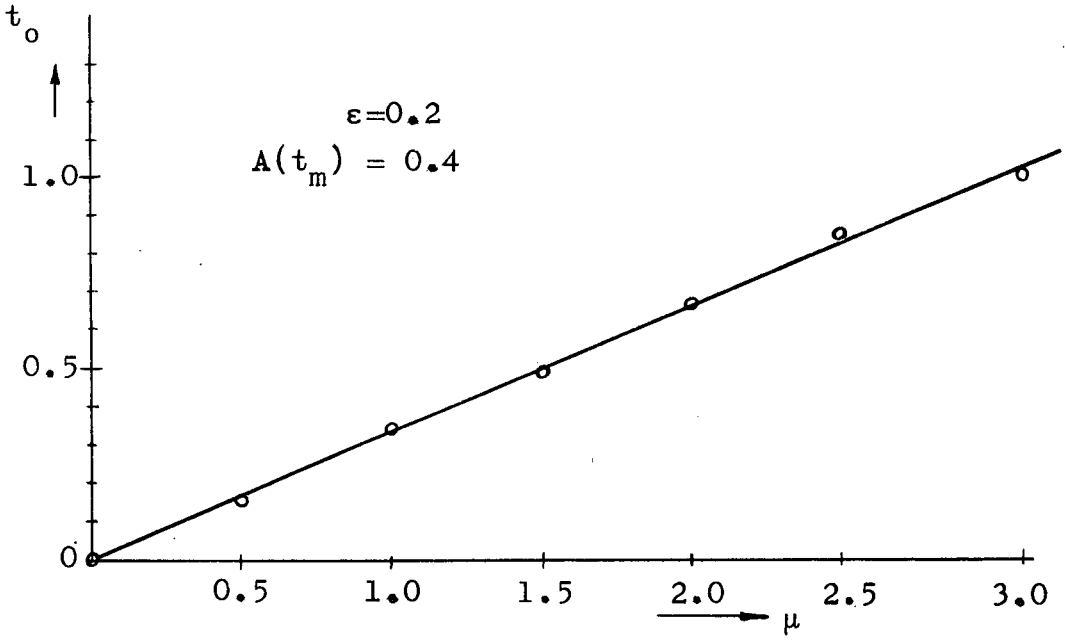
$$t_o \propto \mu \tag{3.9}$$

---

* The initial values of the envelopes are assumed to be unity here for the purpose of finding an approximation. Their true values, however, are greater than unity.



(a)



(b)

Fig. 3.5 Dependence of t_o on μ

Now, the dependence of t_o on ε can be seen from the numerical solutions to the equation

$$\ddot{x} + 2\varepsilon\dot{x} + x + 3x^3 = 0,\ x(0) = 1,\ \dot{x}(0) = 0$$

as ε varies from 0 to 0.9. Corresponding values of t_o, as ε increases, are obtained with $A(t_m)$ fixed, for example, at 0.4. Then t_o is plotted against ε as in Fig. 3.6. The curve obtained resembles a hyperbola, suggesting that t_o varies inversely as ε, or

$$t_o \propto \frac{1}{\varepsilon} \quad . \tag{3.10}$$

Therefore, from equations (3.8), (3.9) and (3.10), we have

$$t_o = k\,\frac{\mu\left[1 - A(t_m)\right]}{\varepsilon}$$

where k = constant. From numerous examples with various values

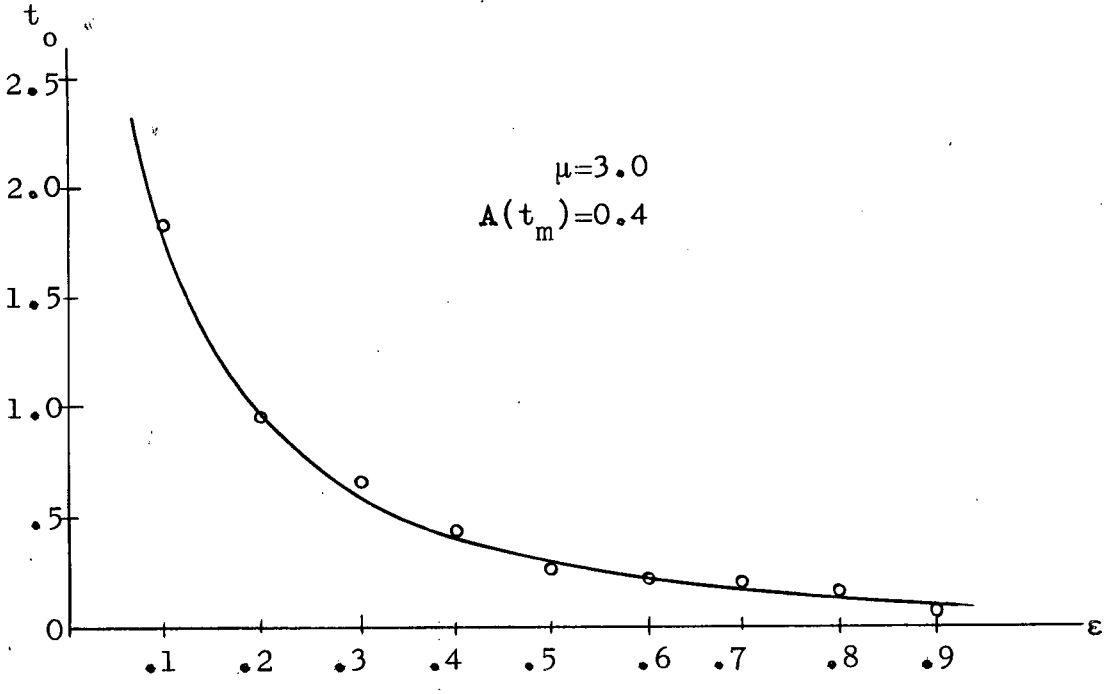


Fig. 3.6 Dependence of t_o on ε

of ε, μ and $A(t_m)$, k is empirically found to be about 1/10. We then arrive at

$$t_o = \frac{\mu\left[1 - A(t_m)\right]}{10\,\varepsilon}\,, \tag{3.11}$$

which enables us to calculate t_o when ε, μ, and $A(t_m)$ are known. Hence, from equation (3.7), we have

$$t_m = -\frac{1}{\varepsilon}\log_e\left[\sqrt{1-\varepsilon^2}\,A(t_m)\right] + \frac{\mu\left[1 - A(t_m)\right]}{10\,\varepsilon} \tag{3.12}$$

Note that t_m is always positive because both ε and $A(t_m)$ are less than unity.

In short, t_m can be calculated from equations (3.6) and (3.11) when ε and μ are specified. Equation (3.6) is an arbitrary criterion based on the consideration of the angle $(\delta\theta)_{max}$ between the linear and non-linear isoclines in the phase-plane, and equation (3.11) is obtained empirically using numerical solutions obtained on the digital computer. It must also be remembered that the t_m thus calculated is not necessarily an optimal choice of the point of transition from the non-linear equation to the linear one, but rather, is a judicious choice of such a point for the purpose of approximating the exact solution.

3.2.3 Determination of A(t) and $\Omega(t)$

In the determination of the amplitude function $A(t)$, consider first the amplitude of the solution to the complementary linear equation, i.e.

$$\ddot{x} + 2\varepsilon\dot{x} + x = 0, \quad x(0) = 1, \quad \dot{x}(0) = 0,$$

which is the case for $\mu = 0$. The amplitude $A(t)$ of the solution is given by

$$A(t) = \frac{1}{\sqrt{1 - \varepsilon^2}} e^{-\varepsilon t} \quad .$$

In the case where $\mu \neq 0$, graphs such as Fig. 3.5(a) have shown that the amplitude $A(t)$ also resembles an exponential but decreases at a slower rate. As suggested by Tuttle,[13] $A(t)$ can assume the following form:

$$A(t) = A_o \, e^{-pt} \quad ,$$

where A_o and p are constants.
Since $A(0)$ is assumed to be unity as mentioned previously, $A_o = 1$, and therefore

$$A(t) = e^{-pt} \quad . \tag{3.13}$$

But at the point where the non-linear term becomes negligible, $t = t_m$. Hence

$$A(t_m) = e^{-pt_m}$$

or

$$p = -(1/t_m)\log_e A(t_m) \tag{3.14}$$

Note that p is always positive for $A(t_m)$ less than 1. Having calculated t_m and $A(t_m)$ from equations (3.6) and (3.12), p is now easily obtained, and equation (3.13) becomes

$$A(t) = \exp\left[-\frac{\log_e A(t_m)}{t_m} t\right] \quad . \tag{3.15}$$

The next step is to determine the phase function $\Omega(t)$. Previous study of the equation with $\mu \neq 0$ has revealed that the phase increases in a non-linear manner, or the frequency of oscillation varies with time.* As a first approximation, we consider that the frequency varies linearly with time and therefore

$$\frac{d}{dt}\left[\Omega(t)\right] = 2\omega_2 t + \omega_1 \tag{3.16}$$

where ω_2 and ω_1 are constant.

Integrating once, we have

$$\Omega(t) = \omega_2 t^2 + \omega_1 t + \omega_o \tag{3.17}$$

where ω_o = constant.

To find ω_o, ω_1, and ω_2, consider equation (3.3) which now becomes

$$\tilde{x}(t) = e^{-pt} \cos\ (\omega_o + \omega_1 t + \omega_2 t^2) \quad .$$

Since $\tilde{x} = 1$ at $t = 0$, we have

$$1 = \cos\ \omega_o$$

Therefore $$\omega_o = 0. \tag{3.18}$$

In order to find ω_1 and ω_2, the method used by Soudack[9] is

---

* In the linear case, the frequency of oscillation is constant, and equals the first derivative of the phase w.r.t. time. Therefore, as a generalization to the non-linear case, the first time derivative of the phase is referred to as the frequency of oscillation.

adopted. They are obtained by considering that both the phase and the frequency in the approximants (3.3) and (3.4) should be matched at the point of transition. This means that at $t = t_m$,

$$\Omega(t_m) = \sqrt{1 - \varepsilon^2}\ t_m + \phi_o \ ,$$

and

$$\dot{\Omega}(t_m) = \sqrt{1 - \varepsilon^2} \ .$$

From equations (3.16), and (3.17), we have

$$\omega_o + \omega_1 t_m + \omega_2 t_m^2 = \sqrt{1 - \varepsilon^2}\ t_m + \phi_o \ , \qquad (3.17a)$$

and

$$\omega_1 + 2\omega_2 t_m = \sqrt{1 - \varepsilon^2} \ . \qquad (3.16a)$$

From the last equation, we then obtain

$$\omega_2 = \frac{\sqrt{1 - \varepsilon^2} - \omega_1}{2\ t_m} \ . \qquad (3.19)$$

Since ω_2 is now expressed explicitly in terms of ω_1, all that remains to do is to determine ω_1 independently. But before we do so, let us see whether the parameters p and ω_2 determined are consistent with the case where $\varepsilon \longrightarrow 0$. The limiting values for both p and ω_2 are expected to be zero, because if $\varepsilon = 0$, we have (a) $A(t) = 1$ and (b) frequency = constant, i.e. no phase retardation. Let us first consider equations (3.6) and (3.12) as $\varepsilon \longrightarrow 0$. We have

$$\lim_{\varepsilon \to 0} A(t_m) = \lim_{\varepsilon \to 0} \sqrt{\frac{\varepsilon}{2\mu}} = 0,$$

and

$$\lim_{\varepsilon \to 0} t_m = \lim_{\varepsilon \to 0} \left\{ - \frac{1}{\varepsilon} \log_e \left[\sqrt{1 - \varepsilon^2}\ A(t_m) \right] + \frac{\mu \left[1 - A(t_m) \right]}{10\varepsilon} \right\}$$

$$= \infty \quad .$$

These results could also be derived from the following argument: As ε becomes smaller, the envelope will decrease at a slower rate and it will take longer time to reach the point of transition to the linear equation, i.e. t_m will become larger. Eventually, as ε gets very close to zero, t_m will approach infinity and $A(t_m)$ will approach zero, for the envelope is always decreasing so long as $\varepsilon \neq 0$. Now, from equation (3.14),

$$\lim_{\varepsilon \to 0} p = \lim_{\varepsilon \to 0} \left[- \frac{1}{t_m} \log_e A(t_m) \right]$$

$$= - \frac{\lim_{\varepsilon \to 0} \log_e A(t_m)}{\lim_{\varepsilon \to 0} t_m}$$

$$= \frac{- \lim_{\varepsilon \to 0} \log_e A(t_m) \lim_{\varepsilon \to 0} \varepsilon}{- \lim_{\varepsilon \to 0} \left[\log_e A(t_m) - \frac{\mu}{10} \right]}$$

$$= \lim_{\varepsilon \to 0} \varepsilon$$

$$= 0 \ .$$

Finally, from equation (3.19)

$$\lim_{\varepsilon \to 0} \omega_2 = \lim_{\varepsilon \to 0,\ t_m \to \infty} \frac{\sqrt{1 - \varepsilon^2} - \omega_1}{2t_m}$$

$$= 0 \ ,$$

provided $\omega_1 \neq \infty$. However, we are guaranteed that $\omega_1 \neq \infty$, for if it is, we have an infinite frequency, which is not likely to occur in the physical systems with which we are dealing. Thus, the limiting values for both ω_2 and p are consistent.

Soudack then proposed a method of finding ω_1 by considering these limits as $\varepsilon \to 0$.[9] As $\varepsilon \to 0$, the differential equation becomes

$$\ddot{x} + x + \mu x^3 = 0, \quad x(0) = 1, \quad \dot{x}(0) = 0. \tag{3.20}$$

Now, since $t_m \to \infty$, $p \to 0$, and $\omega_2 \to 0$, the approximant (3.3) degenerates to the form

$$\tilde{x}(t) = \cos \omega_1 t$$

and the frequency of oscillation is ω_1. The exact solution to (3.20) is, however, given by the Jacobian elliptic cosine

$$x(t) = \mathrm{Cn}\ (k, \omega t)\ [4]$$

where $k^2 = \dfrac{\mu}{2(1 + \mu)}$,

and $\omega = \sqrt{1 + \mu}$.

From books of tables, for example, Jahnke and Emde,[10] the quarter period K(k) of the oscillation can be found. Now ω_1 can be chosen such that the degenerate sinusoidal case has this same quarter period. Since

$$K(k) = \omega \frac{T}{4} = \sqrt{1 + \mu}\ \frac{T}{4}$$

and

$$\omega_1 \frac{T}{4} = \frac{\pi}{2}$$

for the cosine to be zero at $\frac{T}{4}$, we require

$$\omega_1 = \frac{\pi}{2K(k)} \sqrt{1 + \mu}\ . \tag{3.21}$$

To complete the proof of consistency of the solution in the degenerate cases, we need to consider the limiting cases of the functions $A(t)$ and $\Omega(t)$ as $\mu \longrightarrow 0$. In order to do this, the angle criterion must be re-examined. From equation (3.6),

$$A(t_m) = \sqrt{\frac{\varepsilon}{2\mu}} \quad ,$$

which is not valid for very small μ, because then the value of $A(t_m)$ is very much higher than 1. However, since $A(t_m)$ is the amplitude at the point of transition to a linear solution, and this point occurs at $t = 0$ if $\mu \longrightarrow 0$, we have

$$\lim_{\mu \longrightarrow 0} A(t_m) = A(0) = 1.$$

Now, from equation (3.11), since

$$\lim_{\mu \longrightarrow 0} t_o = \lim_{\mu \longrightarrow 0,\ A(t_m) \longrightarrow 1} \frac{\mu \left[1 - A(t_m)\right]}{10\varepsilon}$$

$$= 0,$$

the amplitude curve coincides with that of the linear solution. As a result, the amplitude function $A(t)$ of the nonlinear solution degenerates properly to the amplitude of the linear solution as $\mu \longrightarrow 0$. Considering the phase, since $t_m = 0$ as $\mu \longrightarrow 0$, $\Omega(t) = \sqrt{1 - \varepsilon^2}\,t + \phi_o$ for all time. Therefore

$$\omega_2 t^2 + \omega_1 t + \omega_o = \sqrt{1 - \varepsilon^2}\; t + \phi_o \qquad (3.22)$$

for all time. Since t^o, t^1, and t^2 are linearly independent, the only consistent solution of equation (3.22) for all time is

$$\omega_2 = 0$$

$$\omega_1 = \sqrt{1 - \varepsilon^2}$$

$$\omega_o = \phi_o = 0$$

Hence as $\mu \rightarrow 0$ the parameters ω_1 and ω_2 degenerate properly to the linear solution. However, the proper value of ϕ_o is $-\tan^{-1} \frac{\varepsilon}{\sqrt{1 - \varepsilon^2}}$, and not 0. This discrepency arises from the assumption that the initial value of the amplitude is unity, which leads to $\omega_o = 0$.* Since the solution is only approximate and the critical parameters are ω_1 and ω_2, this discrepency will be tolerated. The error thus introduced will be small because for ω_o as high as 0.2 radians, $\cos\omega_o = 0.98$. This completes the proof of the consistency of the non-linear solution with the known solutions in the degenerate cases where $\varepsilon \rightarrow 0$, or $\mu \rightarrow 0$.

In conclusion, the functions $A(t)$ and $\Omega(t)$ are obtained in the forms

$$A(t) = e^{-pt},$$

and

$$\Omega(t) = \omega_o + \omega_1 t + \omega_2 t^2 .$$

The parameters p, ω_o, ω_1, and ω_2 can easily be calculated from equations (3.14), (3.18), (3.19) and (3.21) for specified values of ε and μ. The value of p is obtained by making the amplitude equal to $A(t_m)$ at $t = t_m$, and the value of ω_o is obtained from the initial condition $x(0) = 1$. Values of ω_1

---

* See section 3.2.6 for initial value correction of the amplitude.

and ω_2 are found by letting $\varepsilon \to 0$, and by matching the phase to the first derivative, which can be regarded as the frequency. Consistency with the known solutions of the degenerate cases where $\varepsilon \to 0$ or $\mu \to 0$ has been shown from the limits of $A(t)$ and $\Omega(t)$, except for a small error in ω_o, which arises from the assumption that the initial value of the amplitude $A(t)$ is unity.

3.2.4 Determination of P and ϕ_o

After $A(t)$ and $\Omega(t)$ are determined, it is a relatively simple matter to find P and ϕ_o. In fact, ϕ_o can be calculated directly from equation (3.17a), i.e.

$$\phi_o = \omega_1 t_m + \omega_2 t_m^2 - \sqrt{1-\varepsilon^2}\, t_m + \omega_o$$

$$= (\omega_1 - \sqrt{1-\varepsilon^2})\, t_m + \omega_2 t_m^2 + \omega_o$$

But from equation (3.16a),

$$\omega_1 = \sqrt{1-\varepsilon^2} - 2\,\omega_2 t_m \quad .$$

Therefore

$$\phi_o = (\sqrt{1-\varepsilon^2} - 2\omega_2 t_m - \sqrt{1-\varepsilon^2})t_m + \omega_2 t_m^2 + \omega_o$$

$$= -\omega_2 t_m^2 + \omega_o \qquad (3.23)$$

P is found by matching the amplitude of (3.3) and (3.4) at $t = t_m$. Thus,

$$A(t_m) = P\, e^{-\varepsilon t_m} \quad .$$

Hence

$$P = A(t_m)\, e^{\varepsilon t_m} \quad . \qquad (3.24)$$

This completes the determination of the parameters of the approximant to the solution of the non-linear equation (2.4) where $\varepsilon < 1$. An example using this approximating scheme is worked out in the next section.

3.2.5 Example

In order to illustrate the approximating scheme just developed and to see how good it is, an example is worked out. Because this approximating scheme has a total phase which is a quadratic in t, it will hereafter be referred to as the parabolic phase approximation, a notation first used by Soudack.[11] Consider, then, the equation

$$\ddot{x} + 0.8\,\dot{x} + x + 3x^3 = 0, \quad x(0) = 1, \quad \dot{x}(0) = 0$$

which has $\varepsilon = 0.4$ and $\mu = 3.0$. Both the magnitudes of ε and μ are inadmissible in the K-B method as we shall see when we compare the solution with the true numerical solution. Using the parabolic phase approximation, however, a much better solution is obtained. These three solutions, i.e., the true numerical solution, the K-B approximation, and the solution obtained from the parabolic phase approximation, are shown in Fig. 3.7.

First, from Appendix B, the K-B method yields the following approximation:

$$\tilde{x}(t) = e^{-\varepsilon t} \cos\left(1 + \frac{3\mu}{8}\right)t \qquad \text{(B.7)}$$

or

$$\tilde{x}(t) = e^{-0.4t} \cos(2.125\,t)\ .$$

Using the parabolic phase approximation, on the other

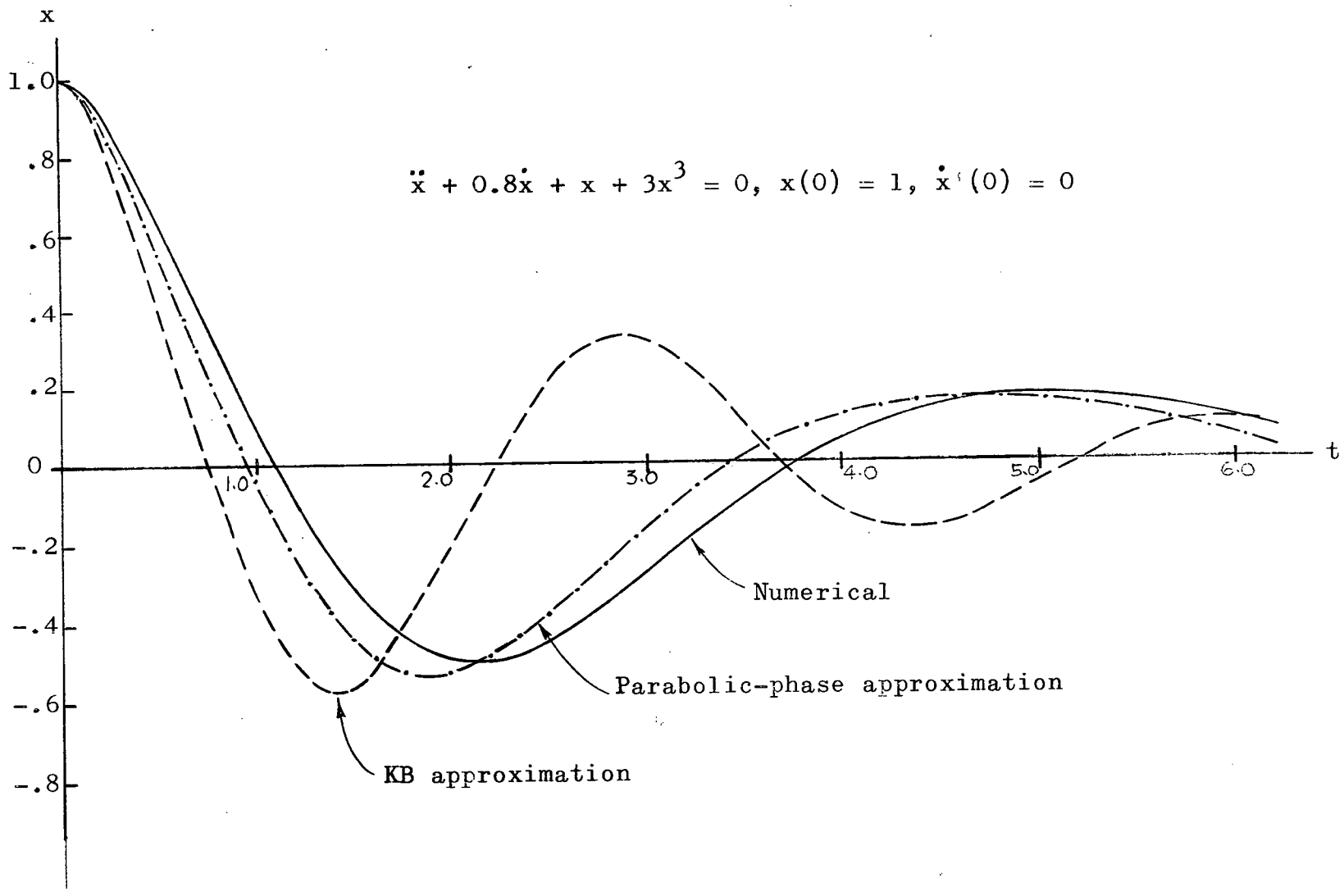


Fig. 3.7 Comparison between approximating methods

hand, we have

for $0 < t \leqslant t_m$ $\qquad \tilde{x}(t) = e^{-pt} \cos(\omega_1 t + \omega_2 t^2)$

for $t \geqslant t_m$ $\qquad \tilde{x}(t) = p \cos(\sqrt{1 - \varepsilon^2}\, t + \phi_o)$.

The unknown parameters are then obtained from the equations developed in the last four sections as follows:

$$A(t_m) = \sqrt{\frac{\varepsilon}{2\mu}} \tag{3.6}$$

$$= \sqrt{\frac{0.4}{2(3)}}$$

$$= 0.258$$

$$t_m = -\frac{1}{\varepsilon} \log_e \left[\sqrt{1 - \varepsilon^2}\, A(t_m)\right] + \frac{\mu\left[1 - A(t_m)\right]}{10\varepsilon} \tag{3.12}$$

$$= -\frac{1}{0.4} \log_e \left[\sqrt{0.84}\,(0.258)\right] + \frac{3(0.742)}{4}$$

$$= 4.17$$

$$p = -\frac{1}{t_m} \log_e A(t_m) \tag{3.14}$$

$$= -\frac{1}{4.17} \log_e (0.258)$$

$$= 0.325$$

$$\omega_o = 0 \tag{3.18}$$

$$\omega_1 = \frac{\pi}{2K(k)} \sqrt{1 + \mu} \tag{3.22}$$

where $k^2 = \frac{\mu}{2(1+\mu)} = \frac{3}{2(1+3)} = 0.375$, and from Jahnke and Emde,[10]

K(k) = 1.761. Therefore

$$\omega_1 = \frac{3.142}{2(1.761)}\sqrt{4} = 1.783 .$$

$$\omega_2 = \frac{\sqrt{1-\varepsilon^2}-\omega_1}{2\,t_m} \qquad (3.19)$$

$$= \frac{\sqrt{0.84}-1.783}{2(4.17)}$$

$$= -0.104$$

$$P = A(t_m)\, e^{\varepsilon t_m} \qquad (3.24)$$

$$= 0.258\, e^{0.4(4.17)}$$

$$= 1.37$$

$$\phi_o = -\omega_2 t_m^{\,2} + \omega_o \qquad (3.23)$$

$$= 0.104(4.17)^2$$

$$= 1.81 .$$

Finally, the complete approximation becomes

for $0 < t \leqslant 4.17$ $\quad \tilde{x}(t) = e^{-0.325t}\cos(1.783t - 0.104t^2)$

for $t \geqslant 4.17$ $\quad \tilde{x}(t) = 1.37\, e^{-0.4t}\cos(0.916t + 1.81)$

Now, a comparison between the two approximations, as shown in Fig. 3.7, indicates that the K-B method is exceedingly simple to carry out, but the result is poor. The frequency of oscillation is too large, and there is no phase retardation. Also, the amplitude of oscillation decays too rapidly. On the other hand, the parabolic phase approximation requires a few more simple computational steps but the extra effort is well rewarded. The frequency of oscillation is now smaller than

that obtained by the K-B method and there is phase retardation in the first part of the solution, where the effect of the non-linear term cannot be neglected. Also, in this first part, the amplitude decays more slowly than $e^{-0.4t}$, as already predicted from previous studies of the equation. Moreover, examples with still higher ε and/or μ will show that the parabolic phase approximation is far superior to the K-B method in dealing with these types of non-linear equations.

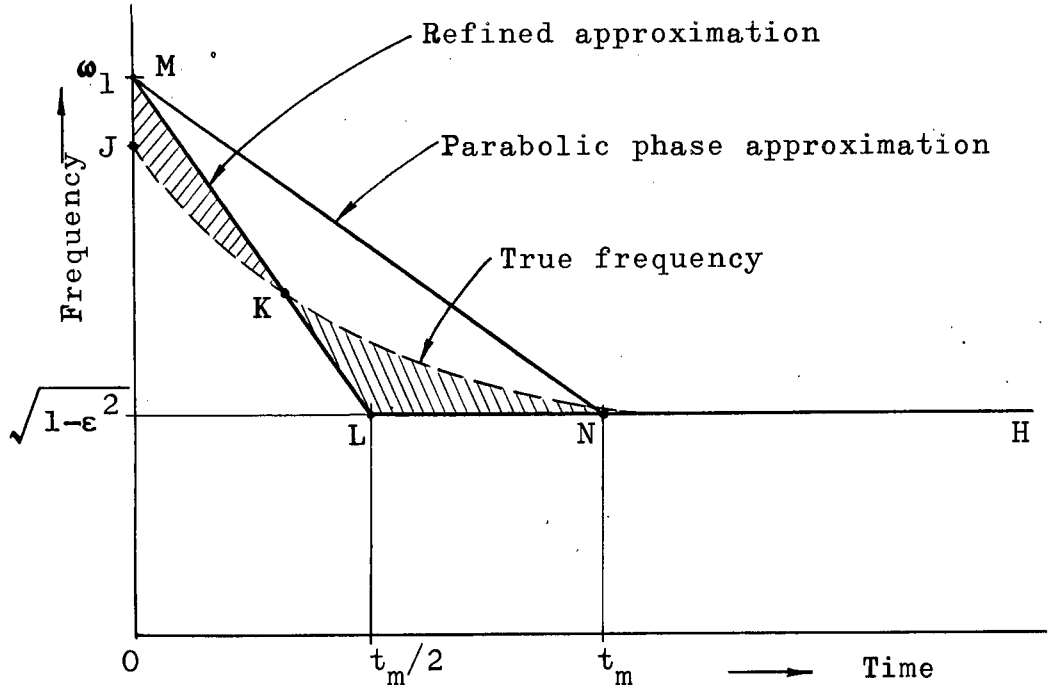
3.2.6 Refinements in the Approximation

Although examples such as the one considered in the last section have indicated that the parabolic phase approximation yields far better results than the K-B method, close examination of these examples suggests that some refinements in the method would make the results even better. The first refinement involves no extra labour and is essentially a modification based on the consideration of the phase term. The second one is a correction of the initial value of the amplitude.

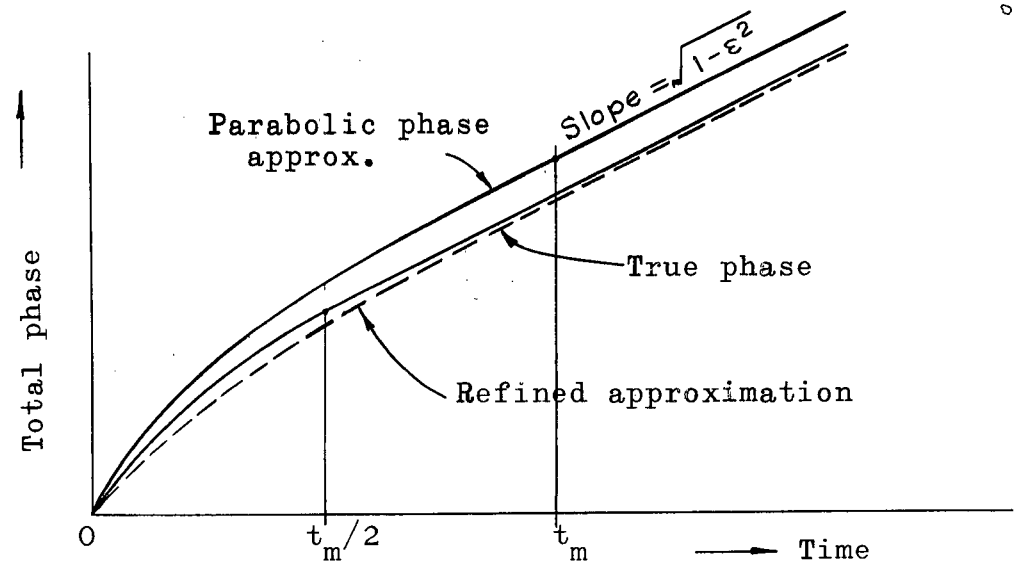
From various examples, the amplitude function $A(t)$ obtained is in fairly good agreement with the numerical solution. It is the phase function $\Omega(t)$ that contributes mainly to the discrepency in the solution. Examination of many examples reveals that the phase $\Omega(t)$ always leads that of the true solution, indicating that either ω_1 is too large or the phase retardation too small. Before we attempt to improve the phase function, let us review how it is obtained and compare it to its true value. First, the angle criterion

is used to determine $A(t_m)$ and consequently t_m, at which the phase retardation is considered negligible. Then, the first time derivative of the phase, i.e. the frequency, is assumed to decrease linearly as time increases for $t < t_m$. For $t = t_m$, the frequency is assumed to be $\sqrt{1 - \varepsilon^2}$ which is the frequency of oscillation of the solution to the complementary linear equation. Thus, as shown in Fig. 3.8(a), the graph representing the frequency begins at the point $M(0, \omega_1)$, drops linearly to the point $N(t_m, \sqrt{1 - \varepsilon^2})$ and becomes level thereafter. The area under this graph then represents the approximate total phase,* as shown in Fig. 3.8(b). But since the approximate phase is always leading the true phase as already pointed out, the true phase may be represented by the dotted curve in Fig. 3.8(b). It is always below the approximate phase and approaches an asymptote with a slope of $\sqrt{1 - \varepsilon^2}$. The true frequency may, therefore, be represented by the dotted curve in Fig. 3.8(a). It becomes clear now how the discrepency arises. Apparently, the straight lines MN and NH do not approximate the true frequency too well for $t < t_m$, and consequently do not give a particularly good approximation to the total phase. In an attempt to improve the approximation, let us consider the point $L(t_m/2, \sqrt{1 - \varepsilon^2})$ as shown in Fig. 3.8(a). The lines ML and LH would give a better approximation to the true total phase because the area KLN under the dotted curve would compensate for the area JMK

* A small error is present here, because the initial value of the phase is slightly different from zero. However, for the purpose of finding an approximation, we have assumed that $A(0) = 1$, which leads to $\omega_o = 0$, or $\Omega(0) = 0$.



(a) Frequency



(b) Phase

Fig. 3.8 Comparison of the true frequency and phase with their approximations

over it. The point L is so chosen partly because the approximating scheme already developed can be adopted with practically no change. We need only match the linear and the non-linear parts of the approximant at $t_m/2$ instead of at t_m. Following the above argument, it might be noted that this new matching point could have been different from $t_m/2$, such as $\frac{1}{3}\ t_m$, $\frac{3}{4}\ t_m$, or $\frac{2}{5}\ t_m$, etc, and that matching at $t_m/2$ does not necessarily give the best result. However, we must not forget that the objective of this section is to modify the parabolic phase method in order to give improved results in general, and not optimum results in particular cases. Since various numerical examples have shown better results by matching at $t_m/2$, we now replace t_m in the equations previously developed by $t_m/2$ and obtain the following equations:

From equation (3.19), $$\omega_2 = \frac{\sqrt{1 - \varepsilon^2} - \omega_1}{t_m} \quad . \tag{3.19a}$$

From equation (3.23), $$\phi_o = -\frac{\omega_2}{4}\ t_m^{\,2} + \omega_o \quad . \tag{3.23a}$$

From equation (3.24), $$P = A\ (\tfrac{t_m}{2})\ e^{\frac{\varepsilon}{2}\ t_m}$$
$$= e^{\frac{\varepsilon - p}{2}\ t_m} \quad . \tag{3.24a}$$

Therefore, the complete approximation becomes

for $0 < t \leqslant t_m/2$ $\quad \tilde{x}(t) = e^{-pt} \cos(\omega_o + \omega_1 t + \omega_2 t^2)$.

for $t \geqslant t_m/2$ $\quad \tilde{x}(t) = P\ e^{-\varepsilon t} \cos\ (\sqrt{1 - \varepsilon^2}\ t + \phi_o)$.

The other refinement is to improve the amplitude function $A(t)$. So far, the initial value of the amplitude has been assumed to be unity, which is not quite correct. From various examples, it has been observed that the true initial amplitude is greater than unity and the difference between the true initial amplitude and unity decreases as μ increases. For most cases where ε is not too large or μ is greater than 3, this difference is negligible. However, if ε becomes close to unity or μ becomes smaller than 3, this difference will be appreciable and a correction added onto the assumed initial amplitude of the approximant will definitely improve the result. Since this difference is greatest for $\mu = 0$, and becomes negligible for $\mu = 3$, we may assume as a first approximation, that it drops linearly as μ increases from 0 to 3. Knowing that the true initial amplitude is $\frac{1}{\sqrt{1 - \varepsilon^2}}$ if $\mu = 0$, we therefore obtain the following relation:

$$\text{Initial Amplitude Correction} = \frac{3-\mu}{3}\left[\frac{1}{\sqrt{1 - \varepsilon^2}} - 1\right] .$$

Thus, for $\mu \leqslant 3$,

$$A(0) = 1 + \frac{3-\mu}{3}\left[\frac{1}{\sqrt{1 - \varepsilon^2}} - 1\right] . \qquad (3.25)$$

Denoting this value of the initial amplitude by A_o, we have

$$A(t) = A_o \, e^{-pt} ,$$

which from equation (3.14) leads to

$$p = - \frac{1}{t_m} \log_e \frac{A(t_m)}{A_o} . \qquad (3.14a)$$

From the initial condition that $\tilde{x}(0) = 1$, we also have

$$1 = A_o \cos \omega_o \ ,$$

or
$$\omega_o = \cos^{-1}(1/A_o), \ \omega_o \leqslant 0 \ ^{*} \qquad (3.18a)$$

Finally the non-linear part of the approximant becomes

for $0 < t \leqslant t_m/2 \quad \tilde{x}(t) = A_o e^{-pt} \cos(\omega_o + \omega_1 t + \omega_2 t^2)$.

As a whole, these refinements in the approximation will yield better results and can be made with almost no extra effort, for the general development of the procedure is not changed. Two examples will be worked out in the next section to illustrate this refined parabolic phase approximation.

3.2.7 Summary and Examples of the Refined Parabolic Phase Approximation

In order to facilitate the application of the refined parabolic phase approximation, a summary of the computational procedure is given as follows:

(1) Normalize the equation into the form

$$\ddot{x} + 2\varepsilon\dot{x} + x + \mu x^3 = 0.$$

(2) Compute $A(t_m)$ from the angle criterion, i.e. equation (3.6)

$$A(t_m) = \sqrt{\frac{\varepsilon}{2\mu}} \ .$$

(3) Compute t_m from equation (3.12)

$$t_m = -\frac{1}{\varepsilon} \log_e\left[\sqrt{1 - \varepsilon^2}\, A(t_m)\right] + \frac{\mu\left[1 - A(t_m)\right]}{10\varepsilon} \ .$$

---

* Positive values of ω_o are not used, as will be explained in section 3.2.8.

(4) If $\mu < 3$, compute A_o from equation (3.25)

$$A_o = 1 + \frac{3-\mu}{3}\left[\frac{1}{\sqrt{1-\varepsilon^2}} - 1\right] ,$$

and assume $A_o = 1$ if $\mu \geqslant 3$.

(5) Compute p from equation (3.14a)

$$p = -\frac{1}{t_m} \log_e \frac{A(t_m)}{A_o} .$$

(6) Compute ω_o from equation (3.18a)

$$\omega_o = \cos^{-1}\left(\frac{1}{A_o}\right) , \qquad \omega_o \leqslant 0.$$

(7) Compute k from $k^2 = \frac{\mu}{2(1+\mu)}$, and then obtain the quarter period K(k) from tables of elliptic functions.

(8) Compute ω_1 from equation (3.22)

$$\omega_1 = \frac{\pi}{2K(k)} \sqrt{1+\mu} .$$

(9) Compute from equation (3.19a)

$$\omega_2 = \frac{\sqrt{1-\varepsilon^2} - \omega_1}{t_m} .$$

(10) Compute P from equation (3.24a)

$$P = A\left(\frac{t_m}{2}\right) e^{\frac{\varepsilon}{2} t_m} .$$

(11) Compute ϕ_o from equation (3.23a)

$$\phi_o = -\frac{\omega_2}{4} t_m^2 + \omega_o .$$

The complete approximation finally becomes

for $0 < t \leqslant t_m/2$ $\quad \tilde{x}(t) = A_o e^{-pt} \cos(\omega_o + \omega_1 t + \omega_2 t^2)$

for $t \geqslant t_m/2$ $\quad \tilde{x}(t) = P\, e^{-\varepsilon t} \cos(\sqrt{1 - \varepsilon^2}\, t + \phi_o)$.

Since one can quickly arrive at answers within slide-rule accuracy, the method seems to be very suitable for preliminary engineering analyses, and as an added benefit it will give the engineer some useful insight into the behaviour of the system.

Example

The same equation considered in section 3.2.5 is again used so that the advantage of the refined method can be illustrated. The equation

$$\ddot{x} + 0.8\,\dot{x} + x + 3x^3 = 0, \quad x(0) = 1, \quad \dot{x}(0) = 0,$$

has $\varepsilon = 0.4$ and $\mu = 3$. Following the steps just outlined, we obtain

$$A(t_m) = \sqrt{\frac{0.4}{2(3)}} = 0.258$$

$$t_m = -\frac{1}{0.4} \log_e \left[\sqrt{0.84}\,(0.258)\right] + \frac{3(0.742)}{4} = 4.17$$

$$A_o = 1$$

$$p = -\frac{1}{4.17} \log_e (0.258) = 0.325$$

$$\omega_o = \cos^{-1}(1) = 0$$

$$k = \sqrt{\frac{3}{2(1+3)}} = 0.612$$

$$K(k) = 1.761$$

$$\omega_1 = \frac{3.142}{2(1.761)}\sqrt{1 + 3} = 1.783$$

$$\omega_2 = \frac{\sqrt{0.84 - 1.783}}{4.17} = -0.208$$

$$P = e^{(0.4 - 0.325)(2.085)} = 1.17$$

$$\phi_o = \frac{0.208}{4} (4.17)^2 = 0.905$$

Hence the complete approximation is as follows:

for $0 < t \leqslant 2.09$ $\quad \tilde{x}(t) = e^{-0.325t} \cos(1.783t - 0.208t^2)$

for $t \geqslant 2.09$ $\quad \tilde{x}(t) = 1.17\, e^{-0.4t} \cos(0.916t + 0.905)$

This approximate solution is plotted in Fig. 3.9 together with the numerical solution, the K-B approximation, and the unrefined parabolic phase approximation, which are obtained in section 3.2.5. It is observed that the refined parabolic phase approximation is the closest to the numerical solution.

Example

As another example of the refined parabolic phase approximation, consider the equation

$$\ddot{x} + \dot{x} + 5x + 10\, x^3 = 0, \quad x(0) = 1, \quad \dot{x}(0) = 0.$$

Since normalization is required in this example, let $\tau = \sqrt{5}t$, to obtain

$$\dot{x} = 2.236\, x'$$

$$\ddot{x} = 5\, x''$$

where $x' = \frac{dx}{d\tau}$ and $x'' = \frac{d^2x}{d\tau^2}$.

Substituting x' and x'' into the original equation and

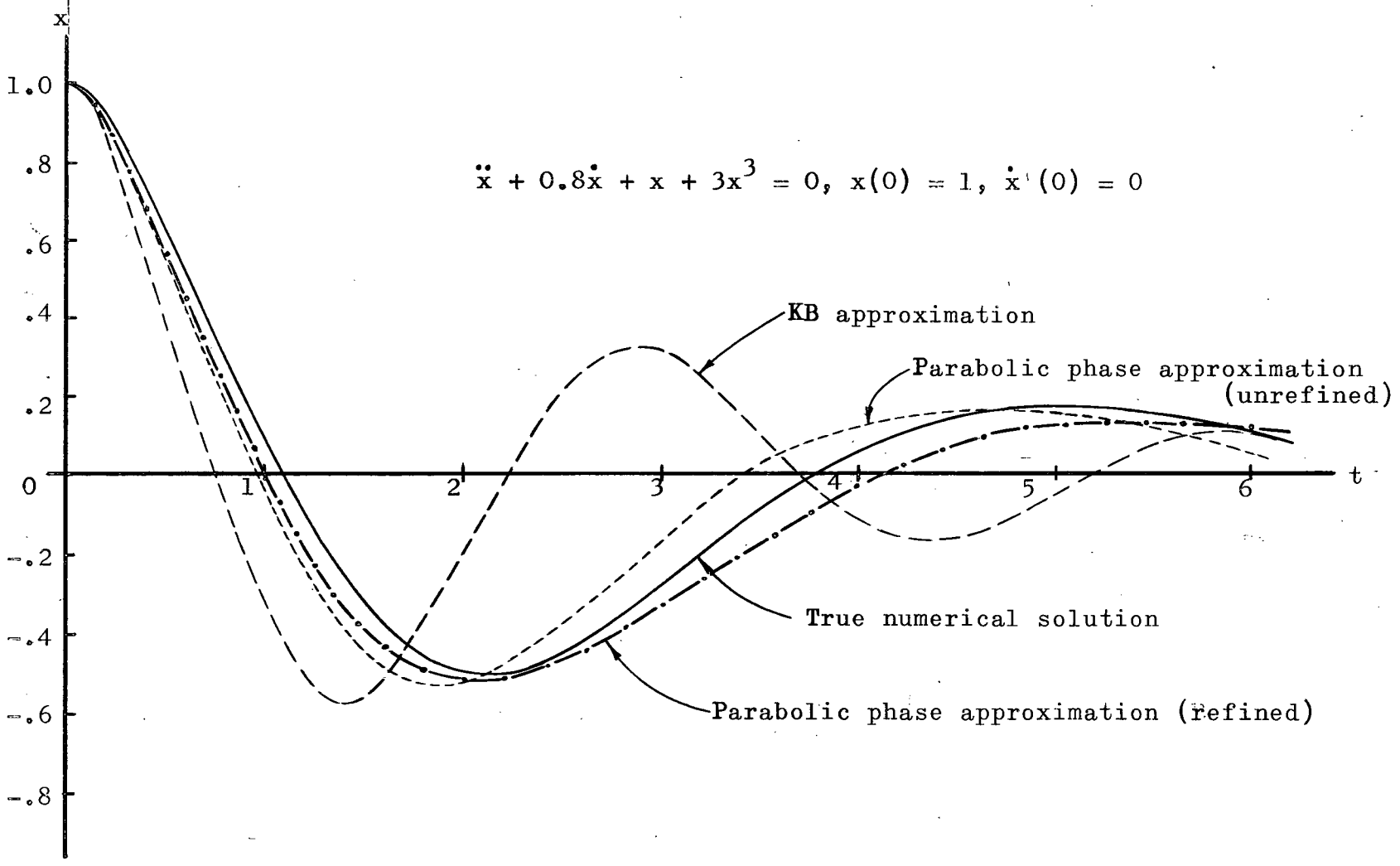


Fig. 9 Approximation by the refined parabolic phase approximation for the equation

$$\ddot{x} + 0.8\dot{x} + x + 3x^3 = 0,\ x(0) = 1,\ \dot{x}(0) = 0$$

dividing through by 5, we have

$$x'' + 0.4472\, x' + x + 2x^3 = 0,$$

which has $\varepsilon = 0.2236$ and $\mu = 2$. Carrying out the computational steps, we obtain

$$A(t_m) = \sqrt{\frac{0.2236}{2(2)}} = 0.236$$

$$t_m = -\frac{1}{0.2236} \log_e\left[\sqrt{1 - 0.05}\ (0.236)\right] + \frac{2(0.774)}{2.236}$$

$$= 7.26$$

$$A_o = 1 + \frac{3-2}{3}\left[\frac{1}{\sqrt{0.95}} - 1\right] = 1.008$$

$$p = -\frac{1}{7.26} \log_e\left[\frac{0.236}{1.008}\right] = 0.200$$

$$\omega_o = \cos^{-1}\frac{1}{1.008} = -0.13$$

$$k = \sqrt{\frac{2}{2(1+2)}} = 0.577$$

$$K(k) = 1.734$$

$$\omega_1 = \frac{3.142}{2(1.734)}\sqrt{1 + 2} = 1.568$$

$$\omega_2 = \frac{\sqrt{0.95} - 1.568}{7.26} = -0.0817$$

$$P = e^{(0.2236 - 0.2)(3.63)} = 1.09$$

$$\phi_o = \frac{0.0817}{4}(7.26)^2 - 0.13 = 0.945$$

Hence the complete approximation is

for $0 < t \leqslant 3.63$ $\quad \tilde{x}(\tau) = 1.008\ e^{-0.2\tau}\cos(1.568\tau - 0.0817\tau^2 - 0.13)$

for $t \geqslant 3.63$ $\quad \tilde{x}(\tau) = 1.09\ e^{-.2236\tau}\cos(0.975\tau + 0.945)$

But since $\tau = 2.236t$, the final approximation becomes

for $0 < t \leqslant 1.62$ $\quad \tilde{x}(t) = 1.008\ e^{-0.447t}\cos(3.508t - 0.409t^2 - 0.13)$

for $t \geqslant 1.62$ $\quad \tilde{x}(t) = 1.09\ e^{-0.5t}\cos(2.18t + 0.945)$

Now, let us compare this solution with the K-B approximation. From Appendix B, the K-B approximation is given by

$$\tilde{x}(t) = e^{-\varepsilon t}\cos\left(1 + \frac{3\mu}{8}\right)t \qquad (B.7)$$

or $\quad \tilde{x}(t) = e^{-0.5t}\cos(3.913t).$

These two solutions are displayed in Fig. 3.10 together with the numerical solution. Results are, again, in favour of the refined parabolic phase approximation.

3.2.8 Errors and Limitations

An important uncertainty inherent in analytical approximating methods is the error in the solution obtained.[12] Because of the presence of the non-linear term, it is usually not a simple matter to make an error analysis. After the approximate solution $\tilde{x}$ is obtained by using the parabolic phase approximation, a common criterion[8] for the system error is given by the integral

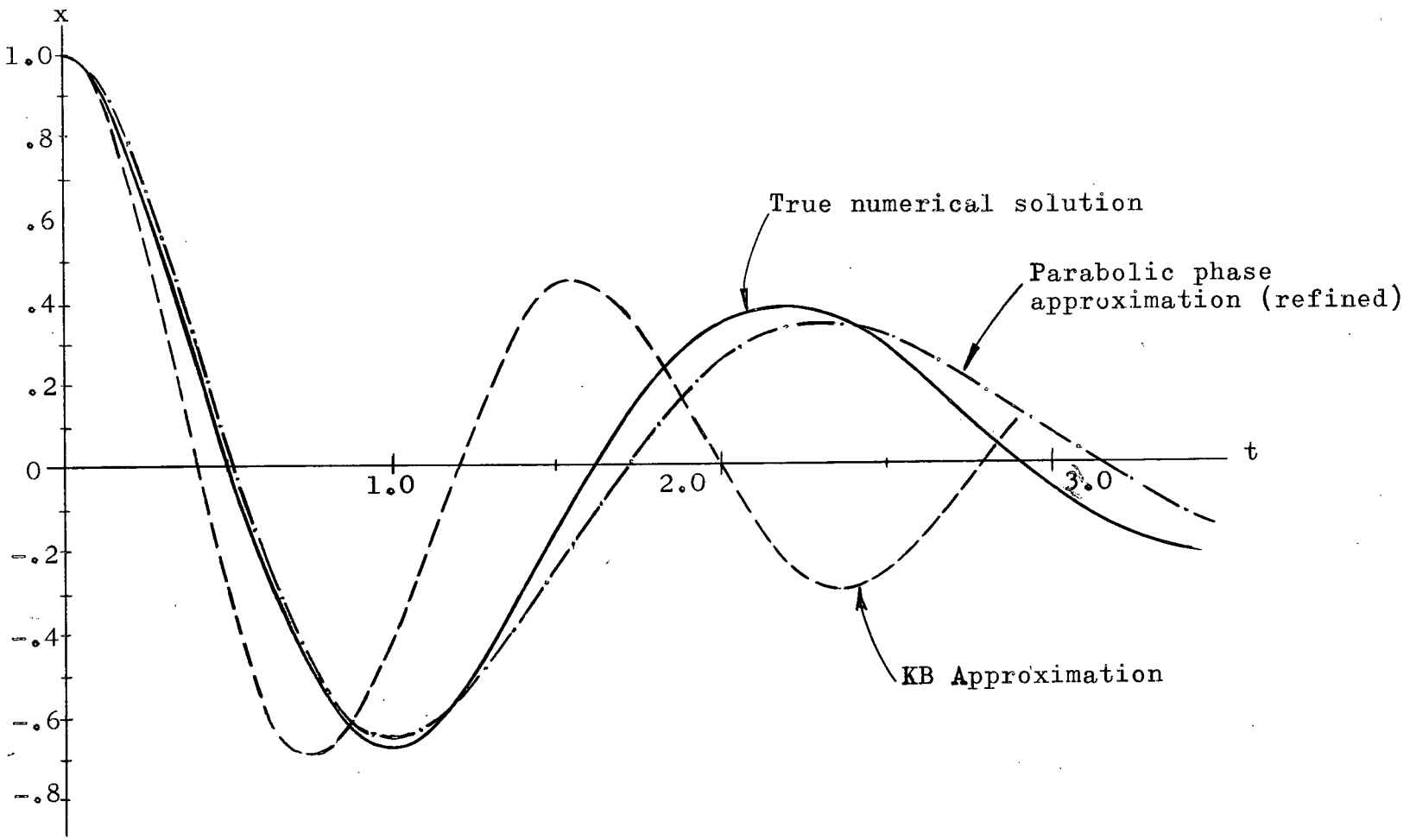


Fig. 3.10 Approximation by the refined parabolic approximation for equation

$$\ddot{x} + \dot{x} + 5x + 10x^3 = 0,\ x(0) = 1,\ \dot{x}(0) = 0$$

$$J = \int_0^{t_m} \sigma^2(t)dt + \int_{t_m}^{\infty} \sigma^2(t)dt$$

where $\sigma(t) = \ddot{\tilde{x}} + 2\varepsilon\dot{\tilde{x}} + \tilde{x} + \mu\tilde{x}^3$.

As already discussed in section 3.2.2, the evaluation of this integral is a formidable job, and leads to no immediate insight into the accuracy of the solution. It must be noted that this type of error analysis does not require the knowledge of the true solution. If, on the other hand, we knew the true solution, it would be very simple to measure the error. The absolute deviation, which is defined as the magnitude of the difference between the approximate and true solutions, could be computed and plotted against the independent variable. We could then have at our disposal a number of quantities as measures of error, for example: (1) the maximum deviation or (2) the area under the deviation curve. In this work, the numerical solution obtained from the digital computer is considered to be the true solution* and is compared with the refined parabolic phase approximation. The maximum deviation is chosen as the measure of error and is evaluated for a large number of examples with a wide range of ε and μ. If the maximum deviation is found to be sufficiently small, we can conclude, because of continuity, that the approximating method is satisfactory for every ε and μ within the range. The following table shows a few numerical results as compared to those obtained from the K-B method.

---

* The machine solution is accurate to three decimal places. See Appendix A.

Maximum Deviation

ε	μ	K-B method	Refined Parabolic Phase Approximation
0.2236	2	0.64	0.13
0.3	9	1.04	0.23
0.4	3	0.66	0.10
0.6	3	0.44	0.07
0.6	9	0.80	0.16
0.8	3	0.48	0.03
0.9	3	0.50	0.10

As suggested from the above table, the limits of ε and μ are not very definite, because they depend on each other as well as on the specified accuracy. For example, higher values of μ may be accepted if the value of ε is higher, while smaller values of ε make the allowable value of μ lower. This is reasonable because as the damping becomes lighter, the non-linear effect takes a longer time to become negligible, and the parabolic phase is not sufficient to ensure a good phase fit. For a maximum deviation of about 0.1, various experimental results have shown that μ may be as high as 5 if $\varepsilon < 0.5$ and as high as 10 if $\varepsilon > 0.5$. In most cases, better accuracy can be expected if μ is not so large. Thus, the above error consideration gives us an idea of the upper limit of μ. The lower limit of μ, however, is determined by one of the approximating steps. Early in the approximating procedure, the amplitude at the point of transition from the non-linear to the linear solution is determined by equation (3.6), i.e.

$$A(t_m) = \sqrt{\frac{\varepsilon}{2\mu}} \quad .$$

Since the amplitude cannot be greater than 1, the relative value of ε and μ should be such that

$$\sqrt{\frac{\varepsilon}{2\mu}} \leqslant 1,$$

or

$$\mu \geqslant \frac{\varepsilon}{2} .$$

Therefore, in the case where $\mu < \frac{\varepsilon}{2}$, the parabolic phase approximation becomes inapplicable and we can consider that the equation is essentially linear because the angle criterion has indicated that the non-linear effect is negligible from the very start. Thus, we conclude that an acceptable lower limit of μ is $\frac{\varepsilon}{2}$.

Another error that is also inherent in the refined parabolic phase approximation is that the initial condition $\dot{x}(0) = 0$ is usually not met. If $\tilde{x}$ is differentiated we have

$$\dot{\tilde{x}}(t) = -pA_o e^{-pt}\cos(\omega_o + \omega_1 t + \omega_2 t^2) - (\omega_1 + 2\omega_2 t)A_o e^{-pt}\sin(\omega_o + \omega_1 t + \omega_2 t^2)$$

Therefore, $$\dot{\tilde{x}}(0) = -pA_o \cos\omega_o - \omega_1 A_o \sin\omega_o$$

$$= -p - \omega_1 A_o \sin\omega_o .$$

Assuming $p \geqslant \left|\omega_1 A_o \sin\omega_o\right|$, we know that the maximum error in the initial slope will never exceed $-p$ if we do not allow positive values of ω_o from equation (3.18a). This is the reason why ω_o is either negative or zero. Thus, we see that an error in the initial slope is inherent in the method, but this error can be ignored because we are primarily interested

in the solution of $x(t)$ for a wide time range. This completes the development of the approximation for the case where $\varepsilon < 1$.

3.3 Case II - $\varepsilon \geqslant 1$

3.3.1 Choice of Approximant

In this case, where $\varepsilon \geqslant 1$, the solution may or may not have overshoots depending on the relative values of ε and μ, as already discussed in section 2.3. Since numerical solutions obtained from the digital computer for $\varepsilon = 1.1$ have shown that μ has to be higher than 25 before a second overshoot appears, and a still higher μ will be required for a second overshoot if ε is larger than 1.1, we need only consider the case with at most one overshoot if we limit our interest to $\mu \leqslant 10$. Because the solution is not oscillatory in general, we can no longer assume an approximant of the form

$$\tilde{x}(t) = A(t) \cos \Omega(t),$$

and consequently, both the parabolic phase approximation and the classical K-B method are not applicable. Therefore, a new method of approximating the solution must be developed.

To this end, consider the "complementary" linear equation

$$\ddot{x} + 2\varepsilon\dot{x} + x = 0, \quad x(0) = 1, \quad \dot{x}(0) = 0$$

where $\varepsilon \geqslant 1$. As indicated in chapter 2, if $\varepsilon = 1$, the solution of this equation is given by

$$x(t) = (1 + t)e^{-t} , \tag{2.7}$$

and if $\varepsilon > 1$, the solution becomes

$$x(t) = \frac{1}{2}\left(1 + \frac{\varepsilon}{\sqrt{\varepsilon^2 - 1}}\right)e^{(-\varepsilon + \sqrt{\varepsilon^2 - 1})t} + \frac{1}{2}\left(1 - \frac{\varepsilon}{\sqrt{\varepsilon^2 - 1}}\right)e^{(-\varepsilon - \sqrt{\varepsilon^2 - 1})t} . \tag{2.8}$$

Let us denote, hereafter, the solution to the complementary linear equation by $x_c(t)$, which is given by equations (2.7) and (2.8), and examine the effect of introducing a non-linear term μx^3 to the equation. Consider then the equation

$$\ddot{x} + 2\varepsilon\dot{x} + x + \mu x^3 = 0, \quad x(0) = 1, \quad \dot{x}(0) = 0$$

whose solution is

$$x(t) = x_c(t) - z(t), \tag{3.26}$$

where $z(t)$ is a correction term to account for the effect of the non-linearity. Thus, the problem now is to approximate $z(t)$.

In order to determine the form of $z(t)$ for the approximation, solutions $x(t)$ to various examples have been obtained numerically from the digital computer and $z(t)$ is then calculated from equation (3.26), i.e.

$$z(t) = x_c(t) - x(t).$$

Fig. 3.11 shows a typical example of $z(t)$. Close examination of the general shape of $z(t)$ has revealed its characteristics from which possible approximants are suggested as follows:

(1) Before $z(t)$ reaches its maximum, i.e. for $t < t_p$, it may be approximated by the function t^n, where $n > 1$.

(2) For $t > t_p$, since $z(t)$ decreases as t increases, the possible approximants are t^m or e^{mt}, where $m < 0$.

(3) $z(t)$ is always positive and vanishes at $t = 0$ and $t = \infty$.

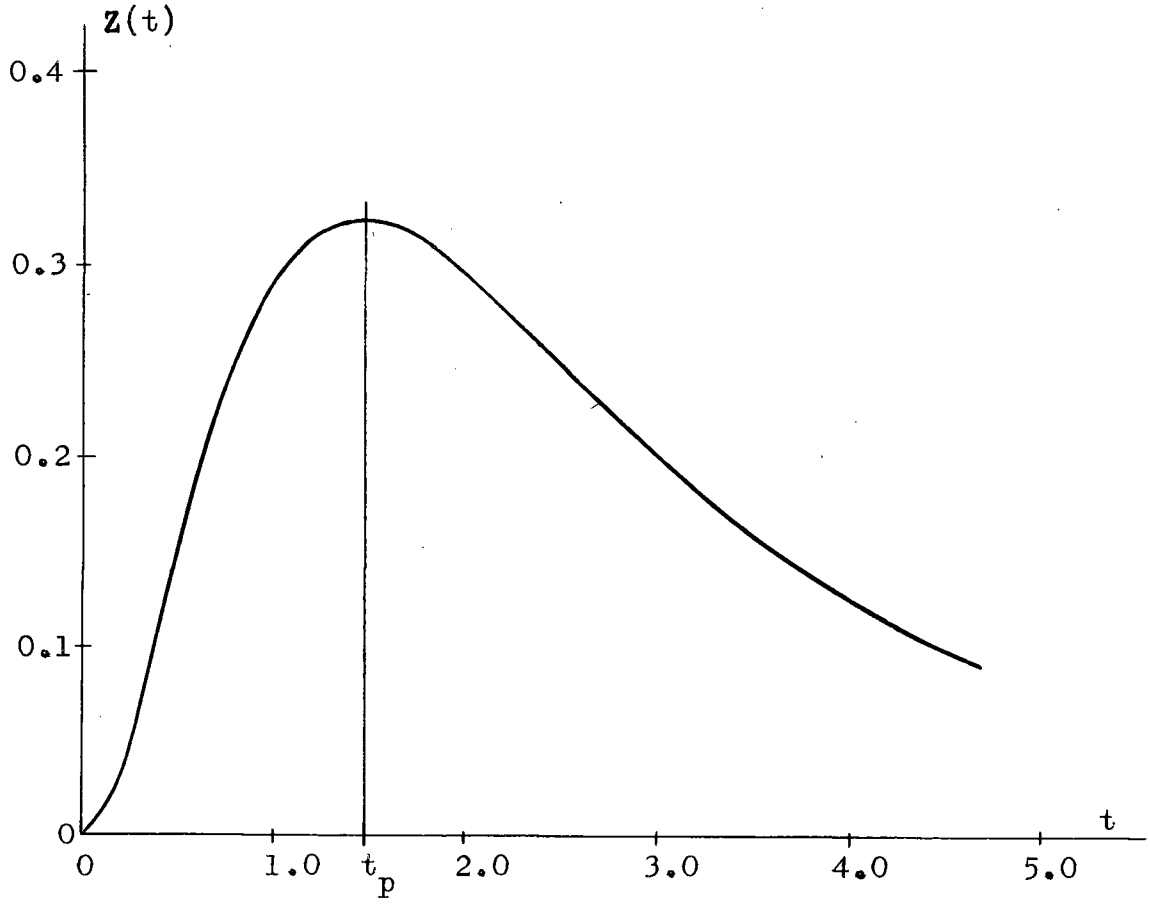


Fig. 3.11 Correction term $z(t)$ for $\varepsilon = 1.2$ and $\mu = 2$

Let us therefore examine the function

$$F(t) = t^n e^{-t}, \qquad n > 1,$$

to see if it has all the above features:

(1) $F(t)$ has only one maximum, i.e. at $t = -n$. For small t, $F(t) \simeq t^n$.

(2) For large t, e^{-t} is the dominating factor.

(3) $F(t)$ is always positive and vanishes at $t = 0$ and $t = \infty$.

Since this function has all the features $z(t)$ has, we will assume that $z(t)$ takes the form

$$z(t) = g\ t^n\ e^{-t}$$

where g is a constant.

Our problem now is to determine the constant parameters g and n in terms of ε and μ. Discrepencies may arise from the assumption that $m = -1$, but from the accuracy of the results which will be developed later, they may be either too small to be of significance or else may have been taken up by the other factor t^n. As a result, the form of the approximant of $z(t)$ becomes

$$\tilde{z}(t) = g\ t^n\ e^{-t} \qquad (3.27)$$

and the rest of this work will be devoted to the determination of g and n as functions of ε and μ.

3.3.2. Determination of n

Empirical results are used to determine both the parameters n and g. In order to determine n, consider the equation (3.27). The maximum value of $\tilde{z}(t)$ is given by

$$\frac{d\tilde{z}(t)}{dt} = 0,$$

or $$g\, n\, t^{n-1}\, e^{-t} - g\, t^{n}\, e^{-t} = 0.$$

Since g, t, and e^{-t} are not zero for finite t, we obtain

$$t = n, \tag{3.28}$$

i.e., n is numerically equal to the time at which $\tilde{z}(t)$ is a maximum. Therefore, if we insist that $\tilde{z}(t)$ has a maximum at the same time as z(t), we can find n from experimental results by simply noting the time at which the maximum of z(t) occurs, as shown in Fig. 3.11. From examples with various values of ε and μ, the following table is obtained:

Numerical Values of n

ε \ μ	0.5	1.0	1.5	2.0	2.5	3.0
1.0	1.62	1.55	1.49	1.44	1.40	1.37
1.2	1.71	1.58	1.50	1.47	1.40	1.38
1.4	1.72	1.62	1.55	1.49	1.42	1.40
1.6	1.87	1.73	1.60	1.54	1.43	1.40
1.8	2.02	1.78	1.68	1.59	1.53	1.42
2.0	2.13	1.90	1.76	1.69	1.60	1.49
2.2	2.39	2.10	1.87	1.74	1.66	1.55
2.4	2.50	2.24	1.97	1.87	1.81	1.71

An error of the order of 0.05 may be expected in some of these figures because some curves of z(t) have a rather flat peak and it is difficult to locate the maximum accurately. At any rate, these figures give a good picture of how n varies with both ε and μ. If the contours of constant n are plotted in the ε-μ plane, Fig. 3.12 is obtained. Taking the possible error of n into consideration, we may now approximate these contours by a set of parallel straight lines as shown in Fig. 3.13. The slope of these straight lines is found to be 0.472. Therefore,

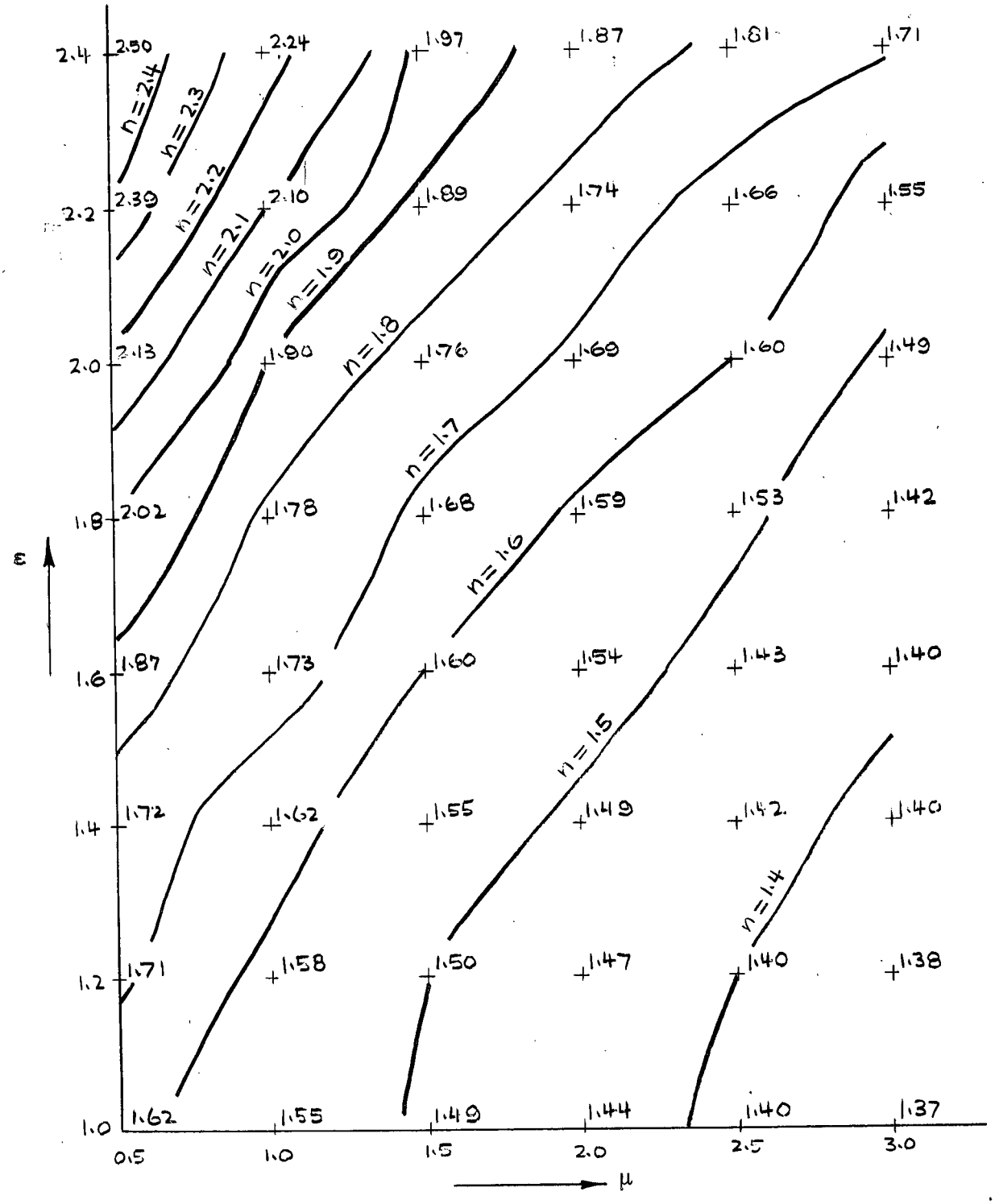


Fig. 3.12 Contours of constant n from experimental results

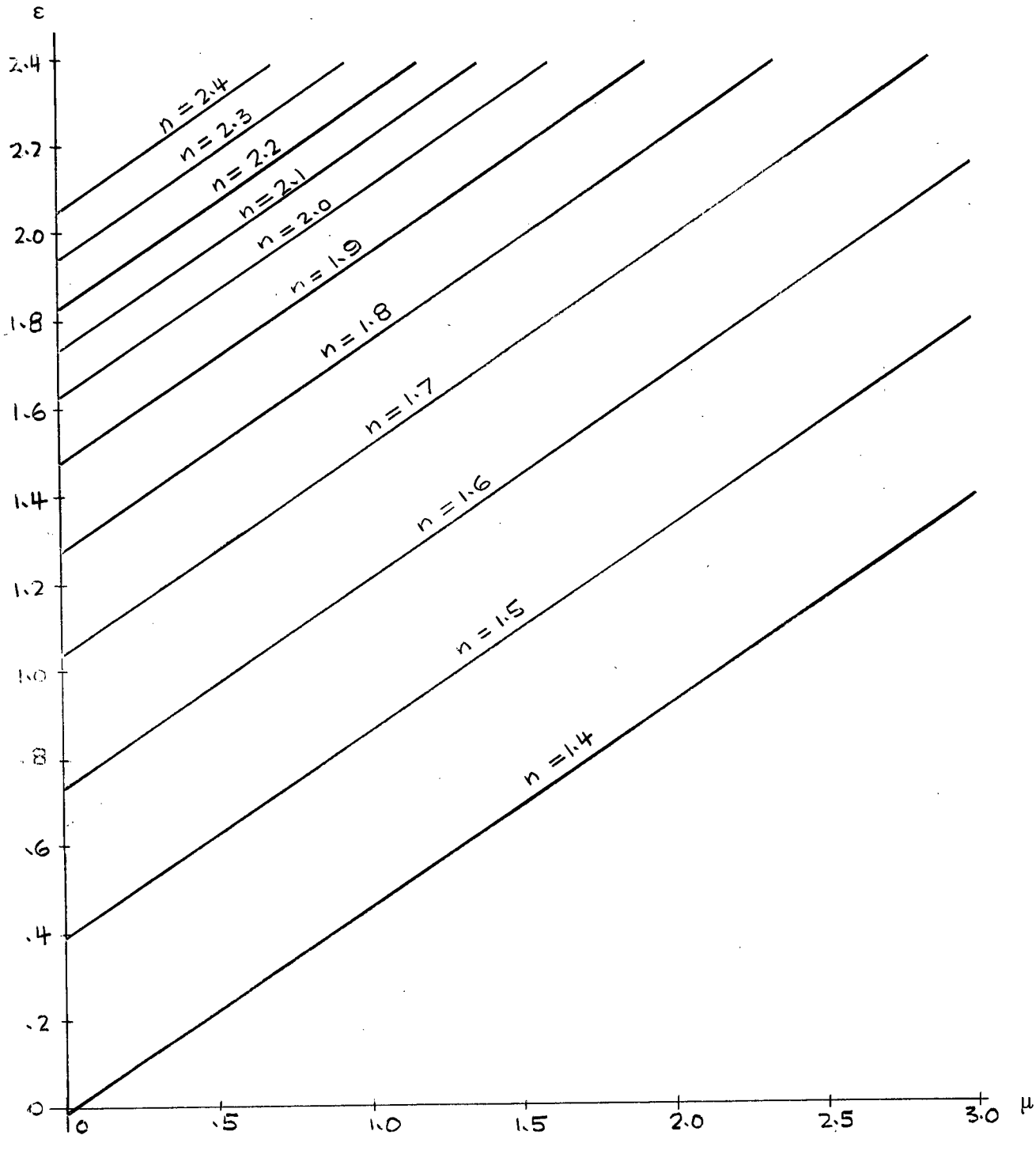


Fig. 3.13 Approximation to the contours of constant n

they can be represented analytically by the simple relation

$$\varepsilon = 0.472\mu + c \qquad (3.29a)$$

where c is the intercept and depends on n. Now, the intercept c is plotted against the corresponding n as shown in Fig. 3.14.

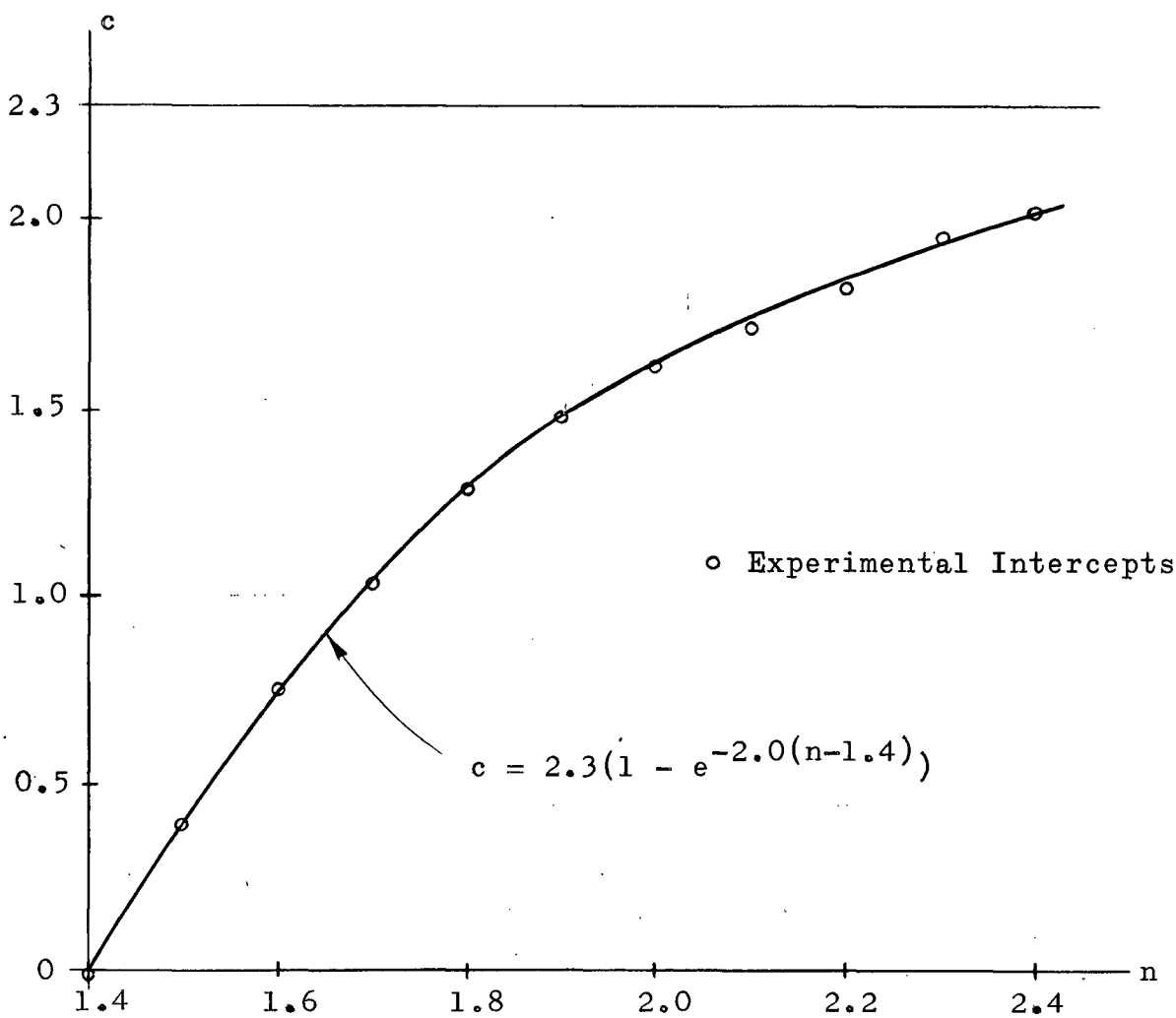


Fig. 3.14 Determination of c as function of n

From this diagram, we observe that the curve exhibits saturation at about c = 2.3 and intercepts the abscissa at n = 1.4. Considering also the shape of the curve, we are led to believe that c can be expressed in terms of n as

follows:

$$c = 2.3\ (1 - e^{q(n - 1.40)})$$

where q is a constant. By trial and error, a q of -2.0 has been found to give good results as illustrated in Fig. 3.14 where the function

$$c = 2.3\ (1 - e^{-2.0(n - 1.40)})$$

is also plotted. Now substituting c into equation (3.29a), we have

$$\varepsilon = 0.472\mu + 2.3\ (1 - e^{-2.0(n-1.4)})$$

which yields

$$n = 1.4 + (0.5)\ \log_e \frac{2.3}{2.3 + 0.472\mu - \varepsilon} \qquad (3.29)$$

Hence n can be computed when ε and μ are specified.

3.3.3 Determination of g

After n has been obtained, g can be determined by making $\tilde{z}(t)$ equal $z(t)$ at the maximum, i.e. at $t = n$, from equation (3.27). The maximum is chosen as the matching point because we have determined the parameter n, such that the peak of the approximant $\tilde{z}(t)$ occurs at the same time as that of $z(t)$. Since the maximum occurs at $t = n$, we have

$$g = z(n)\ n^{-n}\ e^{n} \qquad (3.30)$$

Using the same example as in Fig. 3.11, where $\varepsilon = 1.2$ and $\mu = 2$, we have

$$n = 1.4 + 0.5\ \log_e \frac{2.3}{2.3 + 0.944 - 1.2}$$

$$= 1.46$$

and $$g = 0.325\ (1.46^{-1.46})e^{1.46}$$

$$= 0.805\ .$$

Hence $$\tilde{z}(t) = 0.805\ t^{1.46}\ e^{-t}\ .$$

This approximation is now shown in Fig. 3.15 together with the true z(t) obtained numerically. The result is encouraging

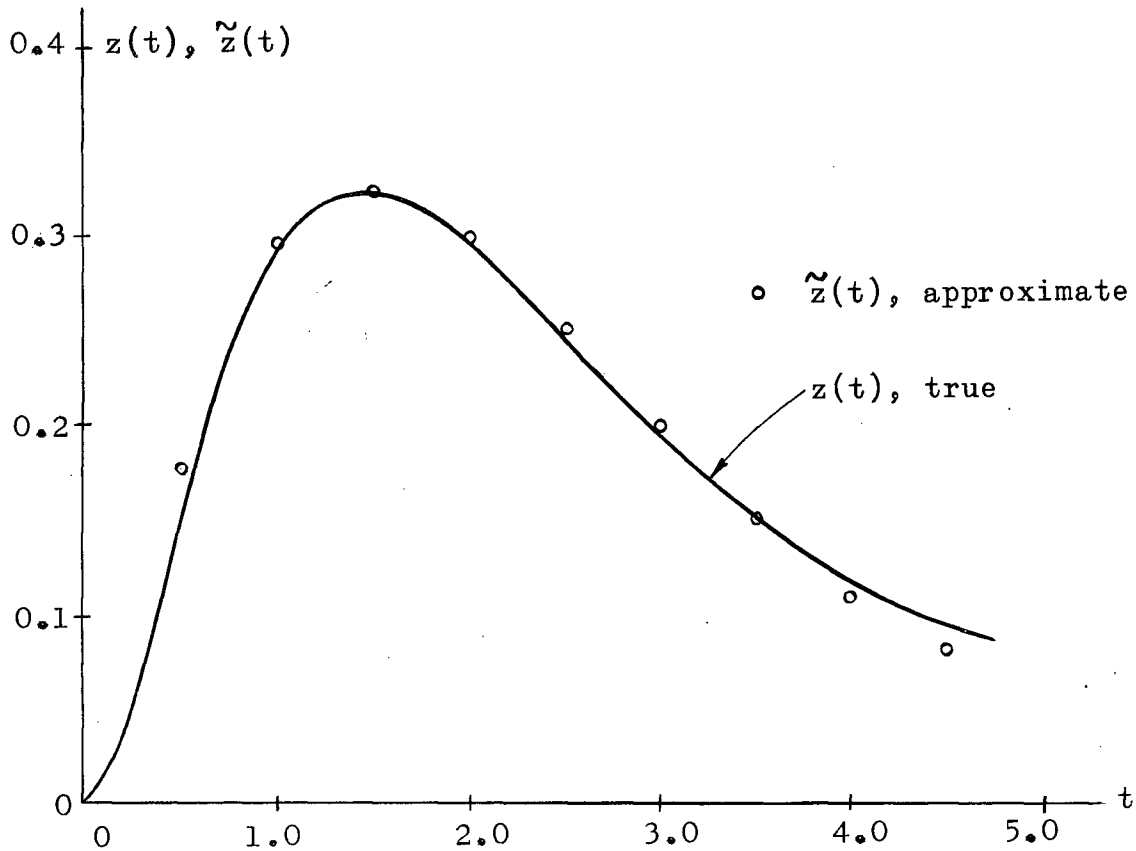


Fig. 3.15 Comparison between z(t) and $\tilde{z}$(t) for $\varepsilon = 1.2$, $\mu = 2$

because very good agreement between z(t) and $\tilde{z}$(t) is observed, indicating that the form assumed for $\tilde{z}$(t) is a good one.

Examples with various values of ε and μ are then investigated in a similar manner, and the corresponding values of g are shown in the following table:

Numerical Values of g

ε \ μ	0.5	1.0	1.5	2.0	2.5	3.0
1.0	.275	.500	.680	.840	1.000	1.14
1.2	.260	.455	.651	.805	.946	1.07
1.4	.238	.420	.584	.731	.862	.978
1.6	.209	.385	.541	.681	.802	.912
1.8	.184	.351	.493	.629	.742	.848
2.0	.159	.314	.454	.578	.691	.793
2.2	.131	.274	.408	.533	.640	.743

Our next task is to determine g as a function of ε and μ from this table. The first attempt was to use the contours of constant g in the ε-μ plane as we did before in the determination of n. The contours of constant g were then plotted as shown in Fig. 3.16. From this diagram, we observed that the contours could not be represented by a set of simple functions such as parallel straight lines because there is a definite trend showing that the slope of each line is different from the rest.

In another attempt, g is now plotted against μ for constant values of ε, as in Fig. 3.17. From the shape of the curves obtained, g is seen to have the form

$$g = a\,\mu^{b}$$

for constant ε, where a and b are constants. Therefore, for constant ε, the plot of $\log_{10} g$ against $\log_{10}\mu$ is a set of straight lines as shown in Fig. 3.18. Hence we can write

$$\log_{10} g = b \log_{10}\mu + \log_{10} a \qquad (3.31)$$

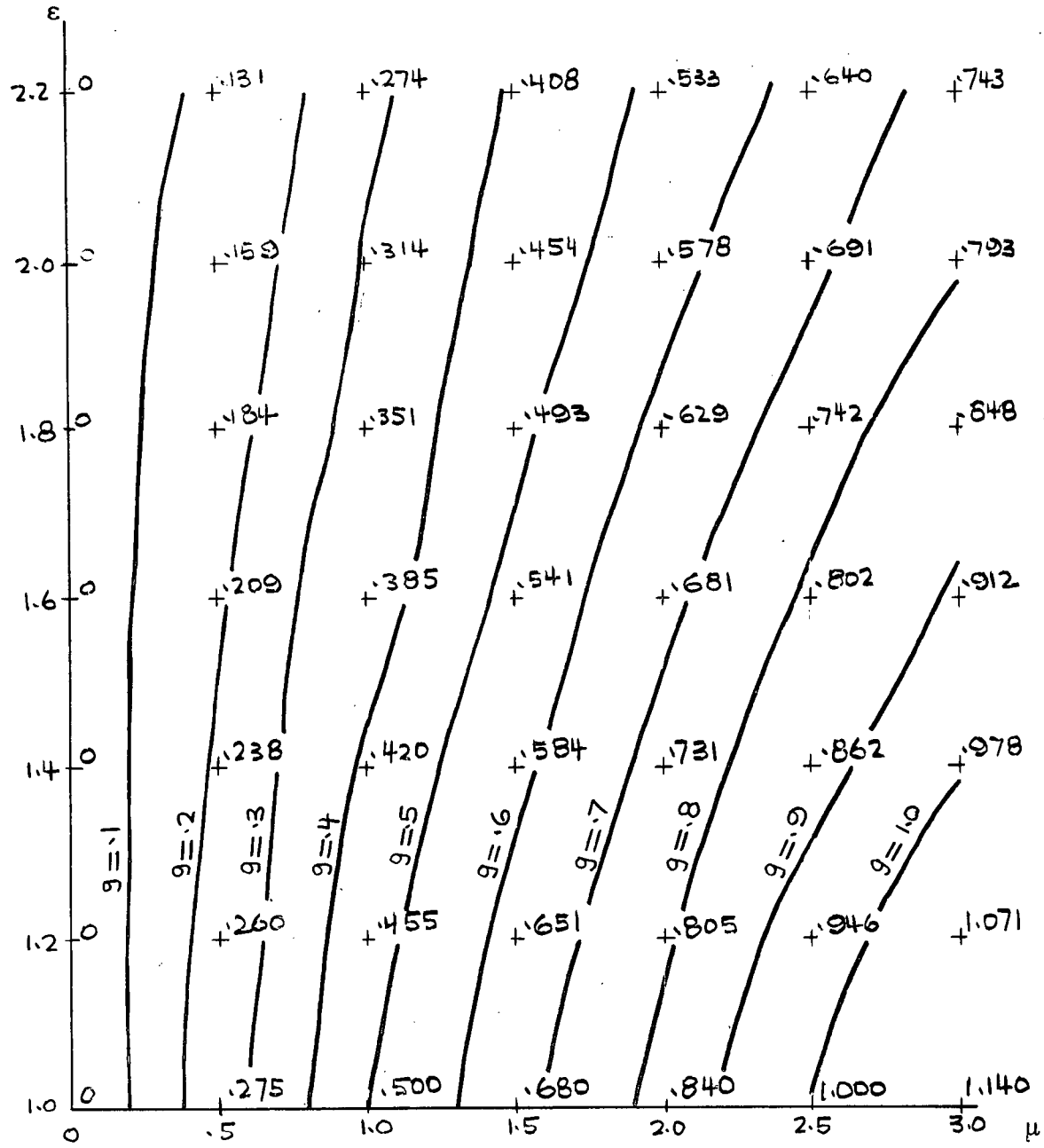


Fig. 3.16 Contours for constant g

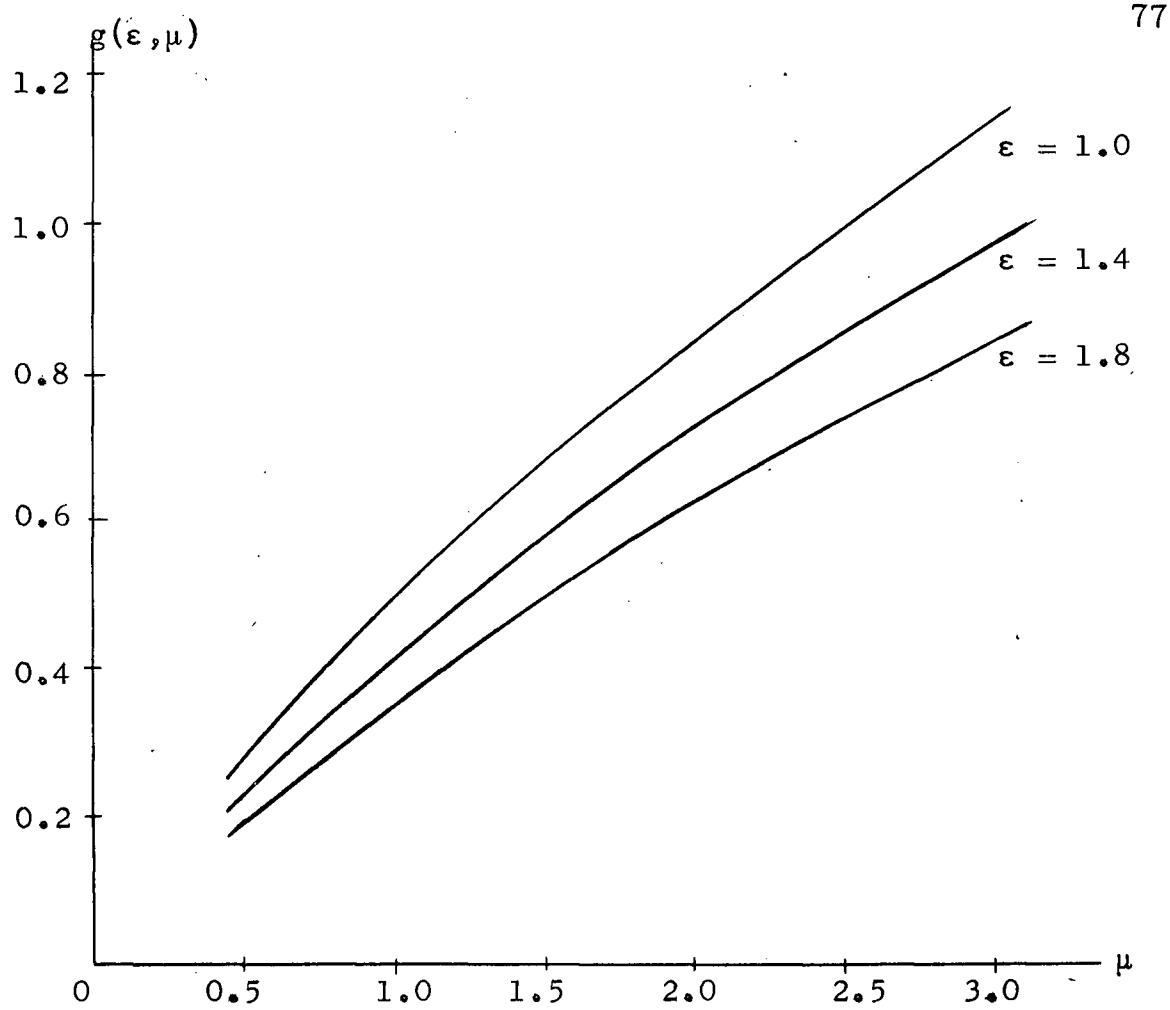


Fig. 3.17 g as a function of μ for constant ε

Since these straight lines are almost parallel and equally spaced, we may approximate this set of lines by Fig. 3.19 in which the straight lines are parallel and equally spaced. Therefore, the constant b is the common slope of all these straight lines and is found to be 0.794. Since the vertical distance between the straight lines for

$\varepsilon = 1.0$ and $\varepsilon = 2.0$ is 0.15, and the intercept for

$\varepsilon = 1.0$ is -0.325, equation (3.31) becomes

$$\log_{10} g = 0.794 \log_{10} \mu - \left[0.325 + 0.15(\varepsilon - 1)\right]. \tag{3.32}$$

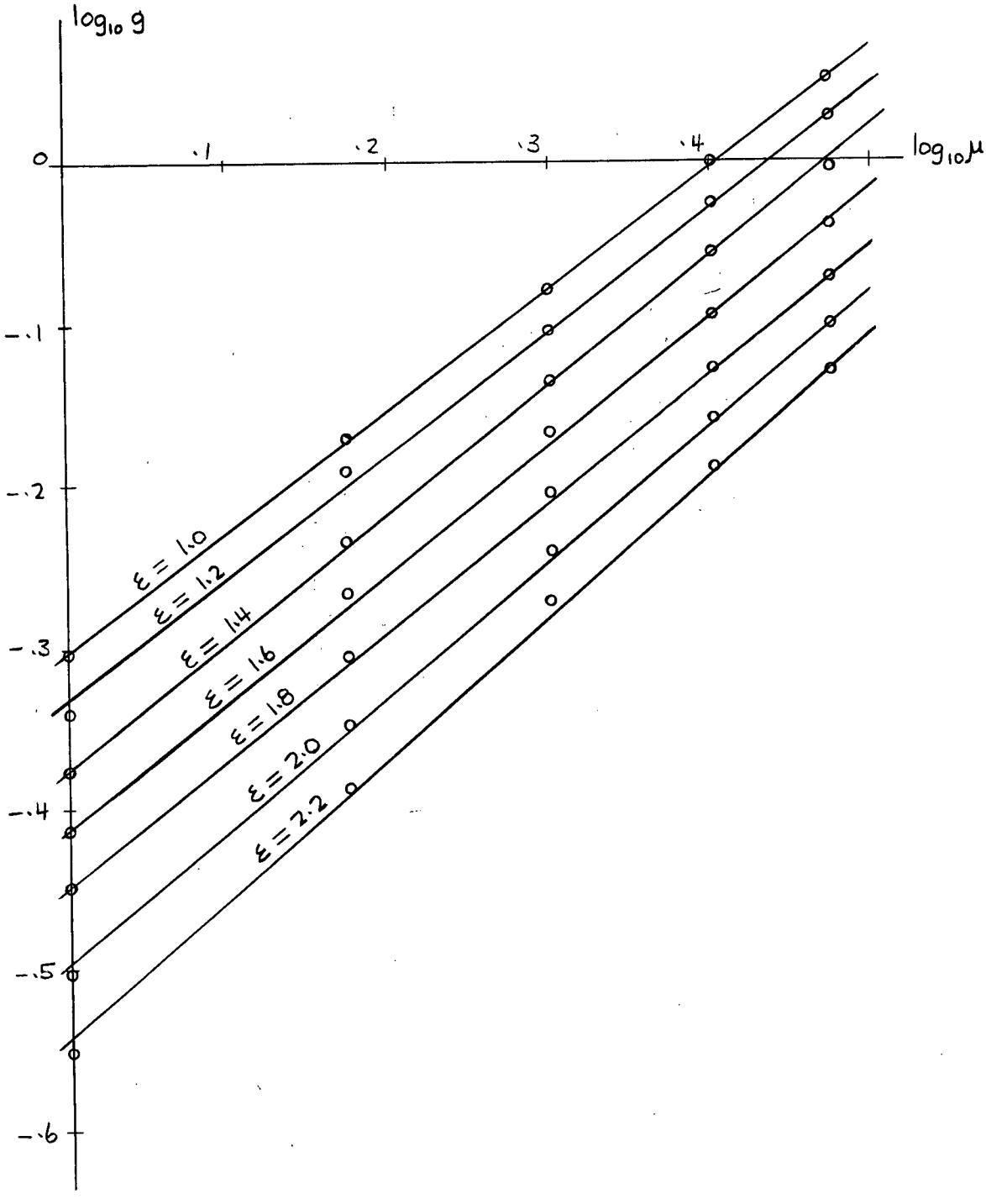


Fig. 3.18 $\log_{10} g$ vs $\log_{10} \mu$ for constant ε

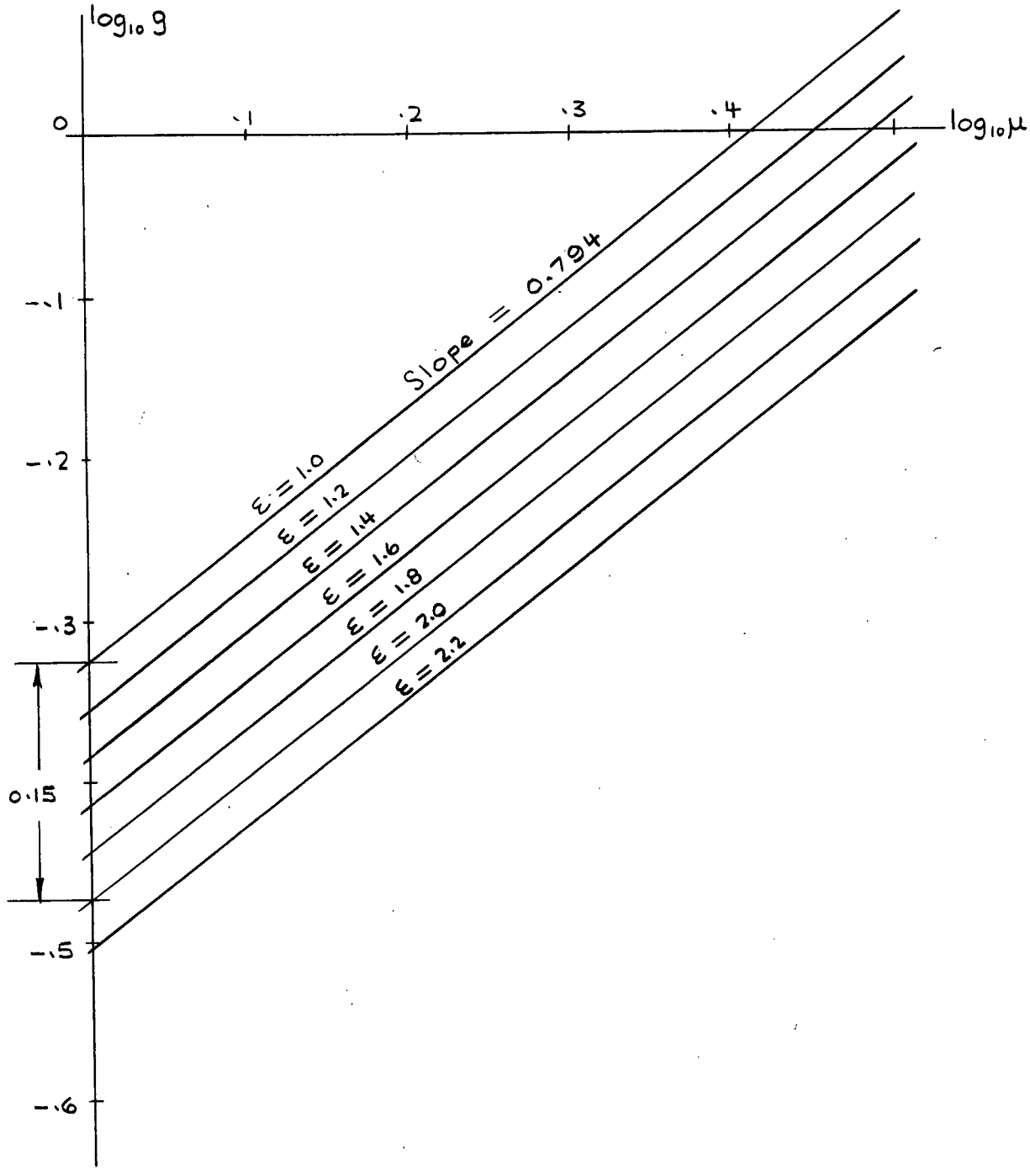


Fig. 3.19 Approximation to Fig. 3.18

Thus, g can be computed when ε and μ are given.

Let us now show that the result is consistent in the limiting case where $\mu \to 0$. As $\mu \to 0$, the non-linear equation approaches its complementary linear equation whose solution is $x_c(t)$. From equation (3.32), we have

$$\lim_{\mu \to 0} (\log_{10} g) = -\infty$$

or $$\lim_{\mu \to 0} g = 0 \ .$$

Therefore $$\lim_{\mu \to 0} \tilde{z}(t) = 0,$$

which yields

$$\lim_{\mu \to 0} \tilde{x}(t) = x_c(t) \ .$$

Hence, our approximate solution degenerates to the correct solution for the complementary equation. An example will be worked out in the next section to illustrate method.

3.3.4 Summary and Example of the Correction Term Approximation

This approximating method, then, is essentially the determination of the solution $x_c(t)$ to the complementary linear equation and the approximation of the correction term $z(t)$ due to the presence of the non-linear term μx^3. The procedure is summarized as follows:

(1) Determine the solution to the complementary linear equation by equations (2.7) and (2.8), i.e.

$$\text{if } \varepsilon = 1, \quad x_c(t) = (1 + t)e^{-t} \ ,$$

and if $\varepsilon > 1$,

$$x_c(t) = \frac{1}{2}\left(1 + \frac{\varepsilon}{\sqrt{\varepsilon^2 - 1}}\right)e^{(-\varepsilon+\sqrt{\varepsilon^2-1})t} + \frac{1}{2}\left(1 - \frac{\varepsilon}{\sqrt{\varepsilon^2 - 1}}\right)e^{(-\varepsilon-\sqrt{\varepsilon^2-1})t} .$$

(2) Approximate $z(t)$ by

$$\tilde{z}(t) = gt^n e^{-t} ,$$

where

$$n = 1.40 + 0.5 \log_e \frac{2.3}{2.3 + 0.472\mu - \varepsilon}$$

$$\log_{10} g = 0.794 \log_{10}\mu - \left[0.325 + 0.15(\varepsilon - 1)\right]$$

The complete approximate solution is then given by

$$\tilde{x}(t) = x_c(t) - g\, t^n e^{-t} .$$

An example is now worked out to illustrate the method.

Example

Consider the equation

$$\ddot{x} + 2.8\,\dot{x} + x + 3x^3 = 0, \quad x(0) = 1, \quad \dot{x}(0) = 0,$$

in which $\varepsilon = 1.4$ and $\mu = 8$. Following the steps just outlined, we obtained

$$x_c(t) = \frac{1}{2}\left(1 + \frac{1.4}{\sqrt{0.96}}\right)e^{(-1.4+\sqrt{0.96})t} + \frac{1}{2}\left(1 - \frac{1.4}{\sqrt{0.96}}\right)e^{(-1.4-\sqrt{0.96})t}$$

$$= 1.21\, e^{-0.42t} - 0.21 e^{-2.38t}$$

$$n = 1.40 + 0.5 \log_e \frac{2.3}{2.3+0.472(3)- 1.4}$$

$$= 1.4$$

$$\log_{10} g = 0.794 \log_{10} 3 - \left[0.325 + 0.15(1.4 - 1)\right]$$

$$= -0.006$$

$$g = 0.987 .$$

Finally, the complete approximation is

$$\tilde{x}(t) = 1.21\ e^{-0.42t} - 0.21\ e^{-2.38t} - 0.987\ t^{1.4} e^{-t} .$$

This approximate solution is compared with the numerical solution in Fig. 3.20 and is seen to be quite satisfactory, the maximum deviation being 0.04. In the same Figure, the approximation by using the Ritz and initial condition matching method is also shown. Referring to Section 3.1, this solution is obtained by solving numerically the equations

$$-\frac{b}{a}(a^2+ 2\varepsilon a+1) + \frac{a}{a+b}(b^2+ 2\varepsilon b+1) - \frac{\mu b^3}{4a(a-b)^2} + \frac{\mu a^3}{(3b+a)(a-b)^2}$$

$$+ \frac{3\mu b^2 a}{(3a+b)(a-b)^2} - \frac{3\mu ab^2}{2(a+b)(a-b)^2} = 0$$

$$-\frac{a}{b}(b^2+ 2\varepsilon b+1) + \frac{b}{a+b}(a^2+ 2\varepsilon a+1) - \frac{\mu a^3}{4b(a-b)^2} + \frac{\mu b^3}{(3a+b)(a-b)^2}$$

$$+ \frac{3\mu a^2 b}{(3b+a)(a-b)^2} - \frac{3\mu ba^2}{2(a+b)(a-b)^2} = 0$$

where $\varepsilon = 1.4$ and $\mu = 3.0$.

By extensive trials, a digital computer produced the following roots:

$$a = -1.11$$

$$b = -4.44$$

Hence the solution is

$$\tilde{x}(t) = A\, e^{-1.11t} + B\, e^{-4.44t} \quad .$$

Applying initial conditions of $\tilde{x}(0) = 1$, and $\dot{\tilde{x}}(0) = 0$, we have

$$A = \frac{-b}{a-b} = 1.333$$

$$B = \frac{a}{a-b} = -0.333$$

Therefore, the Ritz method and the initial condition matching gives

$$\tilde{x}(t) = 1.333\, e^{-1.11t} - 0.333\, e^{-4.44t}$$

As illustrated in Fig. 3.20, this solution has a maximum deviation from the numerical solution of 0.05, which is not as good as that obtained by the correction term method just developed. If we consider the practicability and the effort required in applying the Ritz and initial condition matching method, it is evident that the correction term method is much more tractable.

3.3.5 Errors and Limitations

Following the same reasons as in the case where $\varepsilon < 1$, we again use the deviation between the true and the approximate solutions as a measure of accuracy to justify the validity of the correction term method. The following

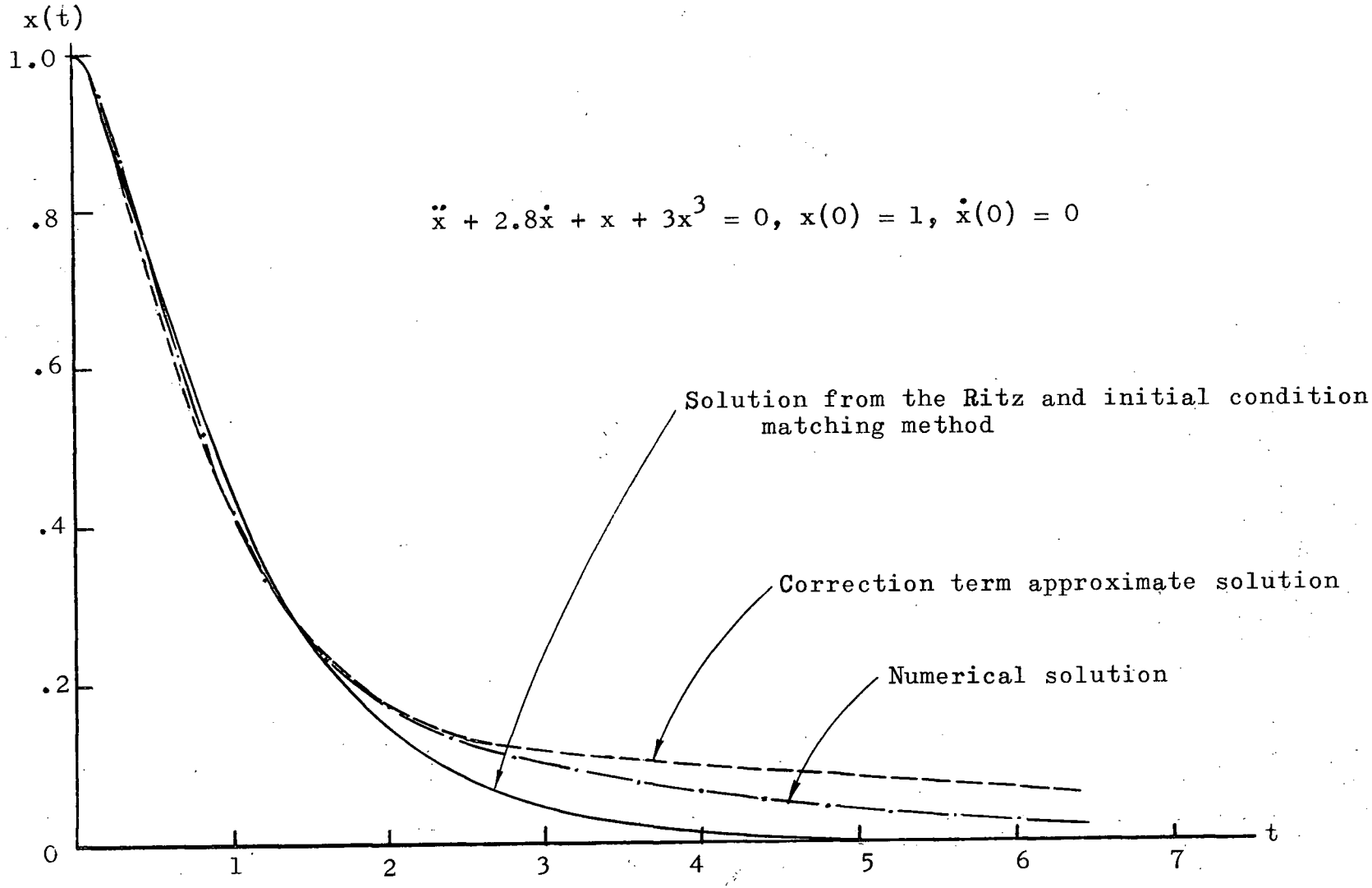


Fig. 3.20 Approximation by the Correction Term Method for $\varepsilon = 1.4$, $\mu = 3$

table is the result obtained from a large number of examples:

ε	μ	Maximum Deviation
1.0	1.0	0.02
1.0	3.0	0.07
1.0	7.0	0.24
1.4	1.0	0.02
1.4	3.0	0.04
1.4	8.0	0.21
1.8	1.0	0.04
1.8	2.0	0.06
1.8	3.0	0.08
2.2	1.0	0.05
2.2	2.0	0.09
2.2	3.0	0.11

From this table, we see that the maximum deviation is rather high for $\mu = 8$. But it must not be forgotten that the deviation is generally much smaller than its maximum. An example will help clarify this point. The magnitude of the deviation between the approximate and true solutions of the equation

$$\ddot{x} + 2.8\,\dot{x} + x + 8\,x^3 = 0, \quad x(0) = 1, \quad \dot{x}(0) = 0$$

is shown in Fig. 3.21. Thus, we see that the deviation is fairly large for small t, then drops off quickly and remains well under 0.08. As a result, we may allow μ to be as high as 8 for $\varepsilon = 1.4$. It is also suggested from the above table that the allowable μ is lowered if ε becomes larger. This is reasonable because as ε increases, one of the exponents involved in $x_c(t)$, i.e. $e^{(-\varepsilon - \sqrt{\varepsilon^2 - 1})t}$, becomes negligible compared to the other and the choice of e^{-t} in the correction term will not be a very good one. At any rate, for ε as high as 2.2, a μ of 5 may still be allowed. Thus, we obtain an idea as to the upper limit of μ from the above error con-

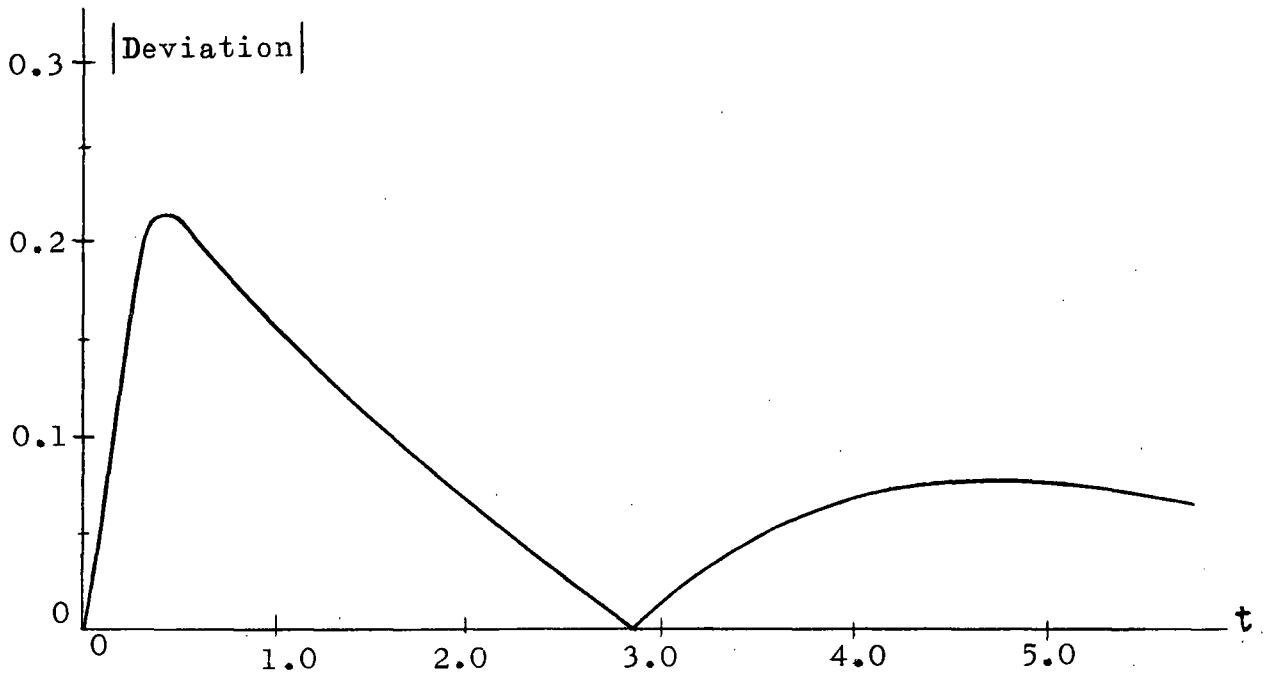


Fig. 3.21 Magnitude of the Deviation for $\varepsilon = 1.4$, $\mu = 8$

sideration. The lower limit of μ is zero of course, because we have shown in Section 3.3.4 (page 80) that the approximation $\tilde{x}(t)$ degenerates to the solution $x_c(t)$ of the complementary equation.

Regarding the value of ε, however, there is an inherent limit in the method. Consider equation (3.29):

$$n = 1.40 + 0.5 \log_e \frac{2.3}{2.3+0.472\mu-\varepsilon}$$

First, n must not be infinite. Therefore, we have the restriction that

$$2.3 + 0.472\mu - \varepsilon \neq 0 \quad .$$

Secondly, since the logarithm of a negative number is not allowable, the restriction then becomes

$$2.3 + 0.472\mu - \varepsilon > 0 \quad .$$

Knowing that the lowest value of μ is zero, we see that for $\mu = 0$, the upper limit of ε is 2.3. However, since we are dealing with non-linear equations, μ is always greater than zero and the upper limit of ε is, therefore, usually higher than 2.3.

Finally, it may be worth mentioning that in this approximation, the initial conditions $x(0) = 1$ and $\dot{x}(0) = 0$ are met because $\tilde{z}(0) = \dot{\tilde{z}}(0) = 0$ which lead to

$$\tilde{x}(0) = x_c(0) = 1$$

and

$$\dot{\tilde{x}}(0) = \dot{x}_c(0) = 0 \quad .$$

This concludes the correction term approximating method which has been developed for $\varepsilon \geqslant 1$.

3.4 Summary

In this chapter, two methods have been developed to approximate directly the solution to the equation

$$\ddot{x} + 2\varepsilon\dot{x} + x + \mu x^3 = 0, \quad x(0) = 1, \quad \dot{x}(0) = 0$$

where both ε and μ are not small numbers.

First, in the case where $\varepsilon < 1$, the parabolic phase approximation was developed and refined. A heuristic argument was given for the use of the form

for $0 < t \leqslant t_m$ $\quad \tilde{x}(t) = A(t) \cos \Omega(t)$

for $t \geqslant t_m$ $\quad \tilde{x}(t) = P\, e^{-\varepsilon t} \cos(\sqrt{1 - \varepsilon^2}\, t + \phi_o)$

as the approximant. The value of t_m where the non-linear effect becomes negligible was determined after $A(t_m)$ had been obtained using an arbitrary criterion based on the consider-

ation of the angle between phase plane isoclines for the linear and non-linear equations. The amplitude $A(t)$ was assumed to have the form

$$A(t) = A_o\, e^{-pt} ,$$

and a parabolic phase of the form

$$\Omega(t) = \omega_o + \omega_1 t + \omega_2 t^2 .$$

The value of ω_o was determined from the initial condition $x(0) = 1$, then ω_1 and ω_2 were found by letting $\varepsilon \longrightarrow 0$ and by matching the phase to the first derivative. The value of A_o was first assumed to be unity and p was obtained by making the amplitude equal to $A(t_m)$ at $t = t_m$.

The method of obtaining these parameters was later refined by using $t = t_m/2$ as the point of transition instead of $t = t_m$, and by correcting the initial amplitude. The approximation then became

$$\text{for } 0 < t \leqslant t_m/2 \quad \tilde{x}(t) = A_o\, e^{-pt} \cos(\omega_o + \omega_1 t + \omega_2 t^2)$$

$$\text{for } t \geqslant t_m/2 \quad \tilde{x}(t) = P\, e^{-\varepsilon t} \cos(\sqrt{1-\varepsilon^2}\, t + \phi_o) .$$

Hence, the parameters P and ϕ_o were determined by matching the two parts of the approximate solution at $t = t_m/2$. Consistency with the known solutions of the degenerate cases where $\varepsilon \longrightarrow 0$ and $\mu \longrightarrow 0$ was also shown from the limit of t_m and the limits of all the parameters in the approximant.

Examples using this refined method of parabolic phase approximation were worked out, and compared with the true numerical solution obtained from the digital computer, as

shown in Fig. 3.9 and Fig. 3.10. The approximations were very close to the true solutions but could be improved by further reducing the value of ω_1. This suggests a project for future research, since this work has illustrated the validity of the approach. Finally, the K-B approximations were also plotted in Fig. 3.9 and Fig. 3.10 for comparison. It is obvious by inspection that the K-B approximations were not as good as the parabolic phase approximations for this type of equation because the phase retardation appeared in the parabolic phase approximations and did not in the K-B approximations.

In the case where $\varepsilon \geqslant 1$, both the parabolic phase approximation and the K-B approximation fail to yield acceptable results because the solution is no longer oscillatory. Therefore, an entirely different method was developed. The solution $x_c(t)$ to the complementary linear equation was first computed, and a correction term $z(t)$ was then defined by

$$z(t) = x_c(t) - x(t)$$

where $x(t)$ was the solution to the non-linear equation. Thus, the problem of approximating $x(t)$ was reduced to approximating $z(t)$. From various numerical examples, it was suggested that $z(t)$ could be approximated by

$$\tilde{z}(t) = g\, t^n\, e^{-t}$$

where g and n are constants depending on ε and μ. Using Fig. 3.12 and Fig. 3.13, which were contour diagrams of n in the ε-μ plane, n was empirically determined to be given by

$$n = 1.40 + 0.5 \log_e \frac{2.3}{2.3+0.472\mu\mu - \varepsilon} \quad .$$

Plots of $\log_{10} g$ against $\log_{10}\mu$ for constant values of ε were used, and g was then found to be given by the empirical formula

$$\log_{10} g = 0.794 \log_{10}\mu - \left[0.325+0.15(\varepsilon-1)\right]$$

Hence we were able to compute $\tilde{z}(t)$ when ε and μ were specified, and the approximation to the solution x(t) was

$$\tilde{x}(t) = x_c(t) - g\, t^n\, e^{-t} \quad .$$

Consistency of this approximant with the known solution of the degenerate case where $\mu \longrightarrow 0$ was also shown.

An example was worked out to illustrate this correction term method, and the result was compared with the numerical solution in Fig. 3.20. With the aid of a digital computer, the Ritz method in conjunction with initial condition matching was used to obtain another approximate solution which was also shown in Fig. 3.20. When accuracy, effort, and practicability were considered, the correction term method was much preferred.

Finally, we see that by using these two approximating methods, values of μ up to 10 and ε and high as 2 may be accepted. Therefore, unlike all the classical methods, they are good for fairly gross non-linearities. The essential difference between these methods and classical methods is their <u>direct</u> approach in attacking systems which are not quasi-linear. In conclusion, both the parabolic phase approximation and the correction term approximation have strong

potential in approximating the solutions to non-linear equations with fairly large non-linearities whose characteristics can be represented by odd cubics such as the flux-current relation of a saturating indicator, or the force-displacement relation of a hard spring.

4. CONCLUSION

As stated in the Introduction, the purpose of this work has been to find a direct method of approximating the solution of the non-linear differential equations of the type

$$\ddot{x} + 2\varepsilon\dot{x} + F(x) = 0$$

where $F(x)$ is, or may be approximated by, an odd cubic with positive coefficients. Without loss of generality, we have studied in detail the equation

$$\ddot{x} + 2\varepsilon\dot{x} + x + \mu x^3 = 0, \; x(0) = 1, \; \dot{x}(0) = 0,$$

and then two methods have been developed to approximate the solution, according to whether $\varepsilon < 1$, or $\varepsilon \geqslant 1$.

In the case where $\varepsilon < 1$, the parabolic phase approximation was developed. The approximant was first derived to be of the following form

for $0 < t \leqslant t_m$ $\qquad \tilde{x}(t) = e^{-pt} \cos(\omega_1 t + \omega_2 t^2)$

for $t \geqslant t_m$ $\qquad \tilde{x}(t) = Pe^{-\varepsilon t} \cos(\sqrt{1 - \varepsilon^2}\, t + \phi_o)$

where all the parameters were determined in terms of ε and μ. Then, a refinement of the method changed the approximant to

for $0 < t \leqslant t_m/2$ $\qquad \tilde{x}(t) = A_o e^{-pt} \cos(\omega_o + \omega_1 t + \omega_2 t^2)$

for $t \geqslant t_m/2$ $\qquad \tilde{x}(t) = Pe^{-\varepsilon t} \cos(\sqrt{1 - \varepsilon^2}\, t + \phi_o)$,

which yielded better results.

In the case where $\varepsilon \geqslant 1$, the solution was approximated by subtracting a correction term $\tilde{z}(t)$ from the solution $x_c(t)$ of the complementary linear equation. The correction term $\tilde{z}(t)$

was of the form

$$\tilde{z}(t) = g\ t^n\ e^{-t},$$

where g and n were computed from formulae involving ε and μ only. Therefore,

$$\tilde{x}(t) = x_c(t) - g\ t^n\ e^{-t}\ .$$

Since the values of ε and μ are not limited to small values, we have found a direct method of approximating the solution without resorting to quasi-linearization of the equation. The limit of ε is slightly above 2 and the limit of μ is close to 10. These values are far too large to be handled by any classical method. Although this method cannot handle ε and μ beyond their limits, it has illustrated the validity of the approach, and further research along this line is encouraging. For example, similar methods may be developed for more general types of grossly non-linear equations.

Finally, the goal of finding directly an approximate solution to the type of grossly non-linear equation has been achieved, and valuable insight into the free response of many engineering systems with odd cubic non-linear characteristics, such as the hard spring and the saturating inductor, can readily be obtained.

APPENDIX A ON COMPUTATION

The 4-th order Runge-Kutta method was used to obtain the numerical solutions used throughout this work. In our case, the formulae for the computation of x and $\dot{x}$ are as follows:[14]

$$x_{n+1} = x_n + h\dot{x}_n + 1/6\ h^2(k_1 + k_2 + k_3)$$

$$\dot{x}_{n+1} = \dot{x}_n + 1/6\ h(k_1 + 2k_2 + 2k_3 + k_4)$$

where

$$k_1 = -(2\varepsilon\dot{x}_n + x_n + \mu x_n^3)$$

$$k_2 = -\left[2\varepsilon(\dot{x}_n + 1/2\ h\ k_1) + (x_n + \tfrac{1}{2}\ h\ \dot{x}_n) + \mu(x_n + \tfrac{1}{2}\ h\dot{x}_n)^3\right]$$

$$k_3 = -\left[2\varepsilon(\dot{x}_n + \tfrac{1}{2}\ hk_2) + (x_n + \tfrac{1}{2}\ h\dot{x}_n + \tfrac{1}{4}\ h^2\ k_1) + \mu(x_n + \tfrac{1}{2}\ h\dot{x}_n + \tfrac{1}{4}\ h^2 k_1)^3\right]$$

$$k_4 = -\left[2\varepsilon(\dot{x}_n + hk_3) + (x_n + h\dot{x}_n + \tfrac{1}{2}\ h^2 k_2) + \mu\ (x_n + h\dot{x}_n + \tfrac{1}{2}\ h^2 k_2)^3\right]$$

and

$$h = t_{n+1} - t_n.$$

The University of British Columbia's IBM 1620 computer was used to carry out the computations to eight significant figures. The program was written in Fortran II.

Since the error of this method is of the order of h^5, and h = 0.2 was used, the error expected was of the order of $(0.2)^5$, or 0.0003, which is negligible for all practical purposes.

APPENDIX B THE KRYLOFF AND BOGOLIUBOFF APPROXIMATION

The Kryloff-Bogoliuboff, or K-B method is concerned with the transient solution of equations of the type

$$\ddot{x} + \omega^2 x + k\, g(x,\dot{x}) = 0$$

where $g(x,\dot{x})$ is arbitrary and K is small. An approximate solution is developed by Kryloff and Bogoliuboff[3] and is essentially a variation of parameters technique.

The solution is assumed to have the form

$$\tilde{x}(t) = A(t) \cos \Theta(t) \qquad (B.1)$$

where $\Theta(t) = \omega t + \phi(t)$.

Differentiating once, we have

$$\dot{\tilde{x}}(t) = \dot{A}(t) \cos\Theta(t) - A(t)\left[\omega+\dot{\phi}(t)\right] \sin\Theta(t).$$

Since we have introduced one more variable, we can impose a constraint such as

$$\dot{A}(t) \cos\Theta(t) - A(t)\, \dot{\phi}(t) \sin\Theta(t) = 0 \qquad (B.2)$$

so that $$\dot{\tilde{x}}(t) = -A(t)\, \omega \sin\Theta(t), \qquad (B.3)$$

and therefore $\ddot{\tilde{x}}(t) = -A(t)\, \omega \sin\Theta(t) - \omega A(t)\, \dot{\Theta}(t) \cos\Theta(t)$.

Substituting $\tilde{x}(t)$, $\dot{\tilde{x}}(t)$ and $\ddot{\tilde{x}}(t)$ into the original equation, we obtain

$$-\dot{A}(t)\, \omega \sin\Theta(t) - \omega A(t)\, \dot{\phi}(t) \cos\Theta(t) + K\, g(\tilde{x},\dot{\tilde{x}}) = 0 \qquad (B.4)$$

Solving equations (B.2) and (B.4) simultaneously, we have

$$\dot{A}(t) = \frac{K\, g(\tilde{x},\dot{\tilde{x}})}{\omega} \sin\Theta(t) \qquad (B.5)$$

and $$A(t)\,\dot{\phi}(t) = \frac{K\,g(\tilde{x},\dot{\tilde{x}})}{\omega}\cos\theta(t) \quad . \qquad (B.6)$$

The approximation is made by averaging (B.5) and (B.6) over one period of oscillation, assuming that A(t) is constant over this period and can be taken out from under the integral sign. This means that if A(t) is slowly varying, the approximation is a good one.

In our case,

$$g(\tilde{x},\dot{\tilde{x}}) = 2\varepsilon\dot{\tilde{x}} + \mu\tilde{x}^3$$
$$= -2\varepsilon A(t)\omega\sin\theta(t) + \mu A^3\cos^3\theta(t),$$
$$\omega = 1,$$
$$K = 1.$$

Substituting in (B.6), and averaging over a period of 2π, we obtain

$$\dot{A}(t) = -\varepsilon A(t),$$

and

$$\dot{\phi}(t) = \frac{3\mu A^2(t)}{8} \quad .$$

But A(0) = 1, from our framework of initial conditions, and since A(t) is assumed to be constant over the period of oscillation,

$$A(t) \cong A(0) = 1.$$

Therefore, we have

$$A(t) = e^{-\varepsilon t},$$

and

$$\phi(t) = \frac{3\mu t}{8} \quad .$$

Finally, (B.1) becomes

$$\tilde{x}(t) = e^{-\varepsilon t}\cos\left(1 + \frac{3\mu}{8}\right)t, \qquad (B.7)$$

which is the approximant used in this work as a comparison to the methods developed in Chapter 3.

Note that the K-B method fails to yield acceptable results in the case where ε and μ are not small, because in such cases, $A(t)$ does vary considerably over one period of oscillation and the assumption required in the averaging procedure is not a reasonable one. However, if we do not make this assumption removing $A(t)$ from under the integral signs, the integrations become very difficult, if not impossible, to handle. Thus, we may not expect good results from equation (B.7) when ε and μ are relatively large, as already illustrated by various examples.

APPENDIX C A MEASURE OF CLOSENESS BETWEEN LINEAR AND NON-LINEAR ISOCLINES

The method of isoclines is often used to construct phase-plane diagrams of 2-nd order differential equations. Consider the equation

$$\ddot{x} + 2\varepsilon\dot{x} + x + \mu x^3 = 0 \qquad (C.1)$$

and its complementary linear equation

$$\ddot{x} + 2\varepsilon\dot{x} + x = 0 \ . \qquad (C.2)$$

The isoclines for (C.1) and (C.2) are respectively given by

$$y = \frac{-x - \mu x^3}{m + 2\varepsilon} = \frac{-x - \mu x^3}{a} \qquad (C.3)$$

and

$$y = \frac{-x}{m + 2\varepsilon} = \frac{-x}{a} \qquad (C.4)$$

where $y = \dot{x}$

$m = \frac{d\dot{x}}{dx}$ = Slope of trajectory,

and $a = m + 2\varepsilon$.

If these two sets of isoclines are close to each other in the phase-plane, then equation (C.2) is a good approximation to equation (C.1). This means that the non-linear effect in equation (C.1) may be neglected. In order to obtain a measure of closeness between the two sets of isoclines, Fig. 3.1 has been constructed, part of which is reproduced in Fig. C.1 for convenient reference. In Fig. C.1, the circular arc of radius R intersects the linear and non-linear isoclines, for the same slope m, at points P and S respectively. The angle subtended by

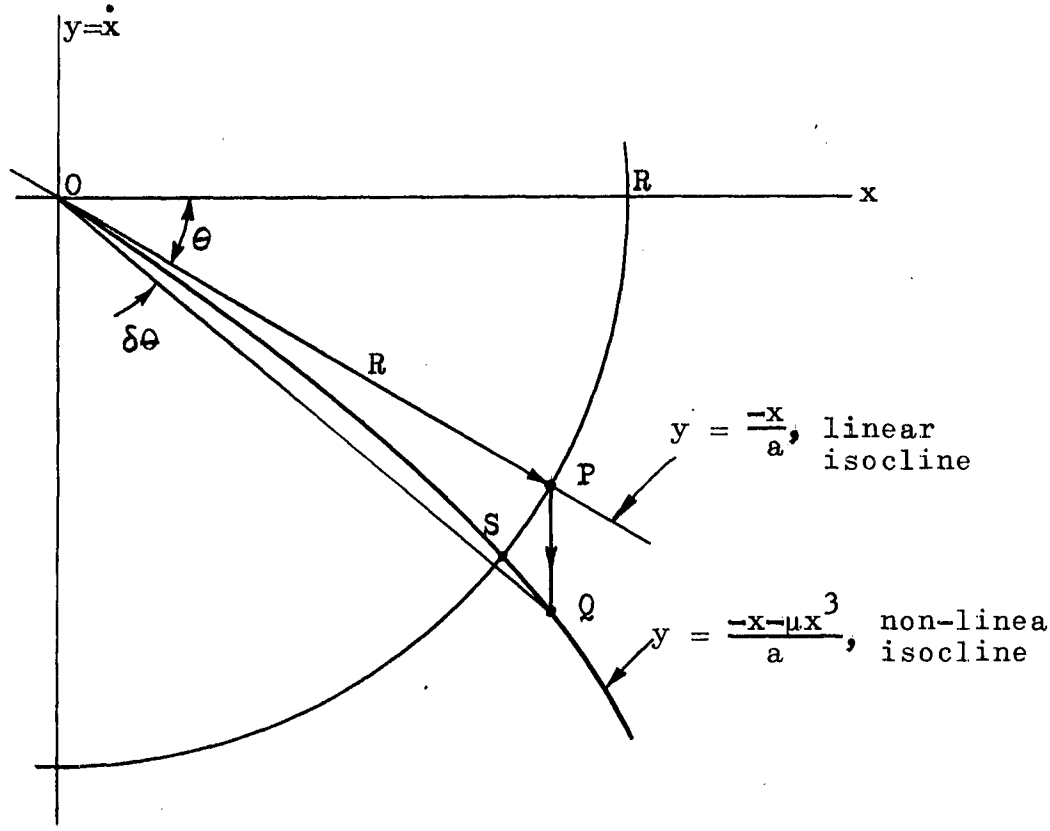


Fig. C.1 A Measure of Closeness between Linear and Non-linear Isoclines of the Same Slope m

the arc PS is therefore a measure of the closeness between the isoclines. However, because the co-ordinates of the point S are difficult to obtain, the point Q is chosen instead, by dropping a vertical line from P to meet the non-linear isocline. Then, the angle $\delta\theta$ can be regarded as a measure of the closeness between the isoclines. From equations (C.3) and (C.4), the co-ordinates of the point P are

$$x_p = R \cos \theta$$

$$y_p = \frac{-R \cos\theta}{a}$$

and those of the point Q are

$$x_Q = R \cos\theta$$

$$y_Q = \frac{-R\cos\theta}{a}\,(1 + \mu R^2\cos^2\theta).$$

Therefore, we have

$$\tan\theta = \frac{y_p}{x_p} = \frac{-1}{a},$$

and $$\tan(\theta + \delta\theta) = \frac{y_Q}{x_Q} = \frac{-1 - \mu R^2\cos^2\theta}{a}$$

$$= -\frac{1}{a}\left(1 + \frac{\mu R^2 a^2}{1 + a^2}\right).$$

Hence $\delta\theta$ is given by

$$\tan\delta\theta = \tan\left[(\theta + \delta\theta) - \delta\theta\right]$$

$$= \frac{-\mu R^2 a^3}{(1+a^2)^2 + \mu R^2 a^2}. \qquad (C.5)$$

This shows that the angle $\delta\theta$ is a function of R and a. As a varies, the angle $\delta\theta$ varies, and its maximum value, for a constant R, is given by

$$\frac{\partial}{\partial a}\left[\tan\delta\theta\right] = 0$$

which can be reduced to

$$a^4 - (\mu R^2 + 2)a^2 - 3 = 0.$$

Therefore, we have

$$a^2 = \frac{1}{2}(2 + \mu R^2) + \sqrt{\frac{1}{4}(2+\mu R^2)^2 + 3}. \qquad (C.6)$$

This is, then, the value of a^2 that will give the maximum $\delta\theta$

for a given R. From equations (C.5) and (C.6), therefore, the maximum value of $\delta\theta$, i.e. $(\delta\theta)_{max}$, is dependent on μR^2 only, and μR^2 becomes a measure of the closeness between the linear and non-linear isoclines. For example, if $\mu R^2 = 0.2$, $(\delta\theta)_{max}$ is approximately 3^o, or 0.05 radian. Thus, for $\mu R^2 = 0.2$, the two isoclines are almost coincident, suggesting that the effect of the non-linear term may be neglected at this point.

Finally, since for equation (C.2),

$$x(t) = \frac{1}{\sqrt{1 - \varepsilon^2}} e^{-\varepsilon t} \cos(\sqrt{1 - \varepsilon^2}\, t - \phi_o)$$

where $\phi_o = \tan^{-1} \frac{\varepsilon}{\sqrt{1 - \varepsilon^2}}$, we have

$$A(t) = \frac{1}{\sqrt{1 - \varepsilon^2}} e^{-\varepsilon t}$$

and

$$R = \sqrt{x^2 + \dot{x}^2} = \frac{e^{-\varepsilon t}}{\sqrt{1 - \varepsilon^2}} \left[1 + \varepsilon \sin (2\sqrt{1-\varepsilon^2}\, t - \phi_o)\right]^{\frac{1}{2}} .$$

Thus, R oscillates about A(t) with a smaller and smaller amplitude as t increases. In other words, A(t) is a very good approximation to R, and therefore, $\mu A^2(t)$ is also a measure of the closeness between the two sets of isoclines. This has enabled us to establish the angle criterion in the development of the parabolic phase approximation.

REFERENCES

1. Cunningham, W.J., "Introduction to Non-linear Analysis", McGraw-Hill Book Co., New York, 1958, pp. 121-170.

2. McLachlan, N.W., "Ordinary Non-linear Differential Equations in Engineering and Physical Science", Oxford, at the Clarendon Press, Second Edition, 1958, pp. 91-92.

3. Kryloff, N. and Bogoliuboff, N., "Introduction to Non-linear Mechanics", Princeton University Press, Princeton, 1947.

4. Cunningham, W.J., op. cit., pp. 75-78.

5. Cunningham, W.J., op. cit., pp. 32-36.

6. Cunningham, W.J., op. cit., p. 168.

7. Cunningham, W.J., op. cit., p. 157.

8. Cunningham, W.J., op. cit., pp. 151-154.

9. Soudack, A.C., "Jacobian Elliptic and Other Functions as Approximate Solutions to a Class of Grossly Non-linear Differential Equations", Technical Report No. 2054-1, April 24, 1961, Stanford Electronics Laboratory, pp. 73-76.

10. Jahnke, E. and Emde, F., "Tables of Functions", Dover Publications, Fourth Edition, New York, 1945, p. 95.

11. Soudack, A.C., op. cit., pp. 75-76.

12. Cunningham, W.J., op. cit., p. 123.

13. Tuttle, D.F., Unpublished notes on Nonlinear Analysis, Stanford University, 1960.

14. Antosiewicz, H.A. and Gautschi, W., "Numerical Methods in Ordinary Differential Equations", A Survey of Numerical Analysis, edited by J. Todd, McGraw Hill Book Co. Inc., New York, 1962, p. 321.