CHAPTER 5

MULTIPLE SCALES

5.1. OVERVIEW OF MULTIPLE SCALES AND AVERAGING

This chapter and the next concern initial value problems of oscillatory type on long intervals of time. Until Section 5.4, we will study autonomous oscillatory second order initial value problems of the form

\[ \ddot{y} + k^2 y = \varepsilon f(y, \dot{y}, \varepsilon), \]
\[ y(0) = \alpha, \]
\[ \dot{y}(0) = \beta. \]  

(5.1.1)

In Section 5.4 it will be shown that this and many other systems, including periodically forced oscillators and systems of several coupled oscillators, can be put into periodic standard form

\[ \dot{u} = \varepsilon f(u, t, \varepsilon), \]  

(5.1.2)

where \( u = (u_1, \ldots, u_N) \) is an \( N \)-dimensional vector variable and \( f \) is periodic in \( t \). Because of its generality and simplicity, periodic standard form will be used in the rest of this chapter and most of the next. It may appear that changing an equation into standard form is an unnecessary step requiring extra work, but in fact the variables which form the components of \( u \) are often exactly those of greatest interest from the standpoint of applications. In addition, theoretical discussions can be given once for equations in standard form and then applied to a wide variety of problems, each of which would require separate treatment if standard form were not used.
For these reasons we consider standard form to be an advantage rather than a disadvantage. On the other hand, beginning with (5.1.1) makes for a smooth transition from previous chapters.

Of course, equation (5.1.1) has been studied in Section 4.4 by the Lindstedt method. However, due to the inherent limitations of that method, only periodic solutions could be found at that time. For instance, the limit cycle of a Van der Pol equation could be located, but it was not possible to find approximations of the nonperiodic orbits that approach the limit cycle. It is the purpose of the method of multiple scales and the method of averaging to remove this limitation. These methods enable the construction of an approximate solution with arbitrary initial conditions which no longer need to satisfy a determining equation. One feature of the Lindstedt method carries over to these new methods: the results are usually valid on an "expanding interval" of time. In most cases this expanding interval has length $O(1/\epsilon)$. Error estimates on longer intervals, such as $O(1/\epsilon^2)$ or even for all time, can be obtained under various special circumstances, but there are no general results of this type. In particular, the "trade-off" property of Lindstedt expansions expressed in equation (4.2.15) is definitely false for multiple scale and averaging methods in general.

The solution of the reduced problem of (5.1.1) can be written in several ways, such as

$$y = A \cos kt + B \sin kt$$

and

$$y = \rho \cos(\psi - kt),$$

where $A$, $B$, $\rho$, and $\psi$ are constants of integration. The first of these, (5.1.3), will be called Cartesian form, and (5.1.4) will be called polar form, because the relations $A = \rho \cos \psi$ and $B = \rho \sin \psi$, which follow from (5.1.3) and (5.1.4), are the same as the relations between Cartesian and polar coordinates in the plane. (This will become more explicit in Section 5.4 and explains our preference for (5.1.4) over forms such as $y = a \sin(\delta + kt)$.) The solution of the reduced problem can be regarded as the "zeroth" approximation to the solution of the perturbed problem (5.1.1). According to the results of Chapter 2, this zeroth approximation has an error $O(\epsilon)$ uniformly for $t$ in a finite interval $0 \leq t \leq T$, and is (in most cases) useless on longer intervals. The first task of both the method of multiple scales and the method of averaging is to improve this zeroth approximation to produce a "first approximation" having error $O(\epsilon)$ uniformly for $t$ in an expanding interval of the form $0 \leq t \leq L/\epsilon$. This is an essentially different problem from that of improving the accuracy to $O(\epsilon^2)$ on a finite interval, a task which is accomplished quite satisfactorily by regular perturbation theory.
The terms "zeroth" and "first" approximation, used in the last paragraph, have a slightly different significance here than in previous chapters. The phrase zeroth approximation always refers to the solution of the reduced problem. In earlier chapters, the solution of the reduced problem was also the leading term in the asymptotic series being constructed. However, in the methods of multiple scales and averaging, the leading term of the asymptotic solution is not the solution of the reduced problem, but is already an improvement over that solution. The phrase first approximation will always be used to refer to the first improvement of the zeroth approximation. Therefore, in regular perturbation theory, the "first approximation" consists of two terms of the asymptotic solution (the zeroth and first order terms, as measured by powers of \( \varepsilon \)); in Chapters 5 and 6, the "first approximation" will consist only of the leading term of the asymptotic series.

The method of multiple scales and the method of averaging both give the same answer to the problem of finding a first approximation to the solution of (5.1.1) on an expanding interval, although they obtain this answer by quite different reasoning. The first approximation given by both methods takes the same form as the zeroth approximation (5.1.3) or (5.1.4), except that the quantities \( A, B, \rho, \) and \( \psi \), which were constants of integration, become slowly varying functions of time. More precisely, the approximations take the form

\[
y \approx A(\varepsilon t) \cos kt + B(\varepsilon t) \sin kt
\]

or

\[
y \approx \rho(\varepsilon t) \cos(\psi(\varepsilon t) - kt),
\]

where the functions of \( \varepsilon t \) are the solutions of certain differential equations which will be given in the next section. Any function of \( \varepsilon t \) can be considered as a slowly varying function of time, since the derivative of such a function contains a factor of \( \varepsilon \) due to the chain rule. (Another way to look at this is that \( t \) must increase a great deal before \( \varepsilon t \) will change very much, when \( \varepsilon \) is small.) It is customary to write either \( \tau := \varepsilon t \), or else \( T_0 = t \) and \( T_1 = \varepsilon t \), so that the approximations take the form

\[
y \approx A(\tau) \cos kt + B(\tau) \sin kt = A(T_1) \cos kT_0 + B(T_1) \sin kT_0
\]

or

\[
y \approx \rho(\tau) \cos(\psi(\tau) - kt) = \rho(T_1) \cos(\psi(T_1) - kT_0).
\]

Here \( \tau \) is called slow time, and the variables \( t \) and \( \tau \) (or \( T_0 \) and \( T_1 \)) are referred to as two time scales. When form (5.1.8) is used, one speaks of
the method of slowly varying amplitude and phase, since $\rho$ and $\psi$ are the amplitude and phase of the simple harmonic motion (5.1.4) in the reduced case. (The "method" of slowly varying amplitude and phase is not actually a method in itself, but only a notation which can be used with several methods such as averaging and multiple scales.)

In most applications of the methods of multiple scales or averaging, the first approximation is sufficient. Therefore this approximation will be discussed at length, in both this chapter (Section 5.2) and the next (Sections 6.1, 6.2, and 6.3). To go beyond the first approximation can be quite difficult. Beyond the first approximation, the methods of averaging and of multiple scales do not always give the same results; in fact, each of these methods (multiple scales and averaging) has several forms, and these different forms do not give the same results beyond the first approximation. (This is possible because these are generalized asymptotic expansions and do not satisfy a uniqueness theorem like Theorem 1.8.3. The different results do not conflict with each other; each is a legitimate asymptotic approximation of the exact solution.) The various multiple scale methods have the reputation of being computationally simpler than the method of averaging after the first approximation, and this is no doubt true in many specific problems. On the other hand, the theoretical foundations of the method of averaging are much better understood than those of multiple scales, and it is possible to draw many conclusions about the behavior of the solutions, such as stability and periodicity, from the averaging calculations.

Higher order multiple scale methods for oscillatory problems can be grouped into three types:

1. Two-scale methods using $T_0 := t$ and $T_1 = \tau := \varepsilon t$. The solutions are written in the form $y_0(t, \tau) + \varepsilon y_1(t, \tau) + \cdots$.

2. Two-scale methods using a strained time $t^+ := (\nu_0 + \varepsilon \nu_1 + \varepsilon^2 \nu_2 + \cdots + \varepsilon^t \nu_t) t$ and a slow time $\tau := \varepsilon t$, with a suitable choice of $\nu_1, \ldots, \nu_t$. The solutions appear as $y_0(t^+, \tau) + \varepsilon y_1(t^+, \tau) + \cdots$.

3. Multiple scale methods using $M$ scales $T_m := \varepsilon^m t$ for $m = 1, \ldots, M$. The solutions are written $y_0(T_0, T_1, \ldots, T_M) + \varepsilon y_1(T_0, T_1, \ldots, T_M) + \cdots$. A variation of this method (the "short form") omits one time scale in each successive term; for instance, a three-scale three-term solution would look like $y_0(T_0, T_1, T_2) + \varepsilon y_1(T_0, T_1) + \varepsilon^2 y_2(T_0)$.

The theory of higher order approximations by the first of these methods is fairly well understood. This form is applicable to a wide variety of problems and gives approximations to any order of accuracy which are valid on expanding intervals of length $O(1/\varepsilon)$ but not (except under special circumstances) on longer intervals. The second and third forms are less satisfactory; they are intended to give approximations on expanding intervals longer than $O(1/\varepsilon)$, but they do not always work (even formally),
and why. The second form is motivated by the Lindstedt method; the idea is that many nonperiodic solutions have the form of damped oscillations in which the "periodic part" can be described using $t^+$ and the damping takes place on the time scale $\tau$. The third form is the most general, since it includes the first two, the first by taking $M = 1$, and the second since $t^+$ (taken to $\ell$ terms) can be written as a function of the scales $T_0, \ldots, T_\ell$.

In Section 5.3 (which may be omitted without loss of continuity) the first and third of these methods will be illustrated.

To conclude this introductory section we will study a certain linear problem of the form (5.1.1) "in reverse." That is, we will write down the exact solution, and then expand this solution into three different forms of multiple scale approximations and examine the error. This is, of course, "cheating," and is intended only to give insight; we are not actually using the method of multiple scales. (It is rather like "cheating" by using the quadratic formula in Chapter 1.) The example to be studied is

$$\dot{y} + 2\varepsilon \dot{y} + (1 + \varepsilon)y = 0,$$

$$y(0) = \alpha,$$

$$\dot{y}(0) = 0,$$

with exact solution

$$y = \alpha e^{-\varepsilon t} \cos \sqrt{1 + \varepsilon - \varepsilon^2} t + \frac{\varepsilon \alpha}{\sqrt{1 + \varepsilon - \varepsilon^2}} e^{-\varepsilon t} \sin \sqrt{1 + \varepsilon - \varepsilon^2} t.$$  (5.1.10)

There are two effects apparent here: The damping term in (5.1.9) causes an exponential decay of the solution "on the time scale $\varepsilon t$," while both perturbation terms interact to produce a frequency shift from the free frequency 1. The shifted frequency can be expanded as

$$\sqrt{1 + \varepsilon - \varepsilon^2} \sim 1 + \frac{1}{2} \varepsilon - \frac{5}{8} \varepsilon^2 + \frac{15}{48} \varepsilon^3 + \cdots.$$  (5.1.11)

The first form of multiple scale solution which will be developed is a two-time-scale expansion using $t$ and $\tau$. The technique for converting (5.1.10) into a two-time expansion is the same as that used in Section 4.7 for changing a Lindstedt expansion into a two-time expansion. First, using (5.1.10) and (5.1.11), and the time scales $t$ and $\tau$, write

$$y = \alpha e^{-\tau} \cos \left( t + \frac{1}{2} \tau - \frac{5}{8} \varepsilon \tau + \cdots \right)$$

$$+ \frac{\varepsilon \alpha}{\sqrt{1 + \varepsilon - \varepsilon^2}} e^{-\tau} \sin \left( t + \frac{1}{2} \tau + \cdots \right).$$  (5.1.12)

Next, expand each term in $\varepsilon$ holding $t$ and $\tau$ constant as though $\tau$ were
an independent variable rather than a function of $\varepsilon$. The result, through order $\varepsilon$, is

$$y \simeq a e^{-t} \cos \left(t + \frac{1}{2}\tau\right) + \varepsilon a e^{-t} \left(\frac{5}{8}\tau + 1\right) \sin \left(t + \frac{1}{2}\tau\right). \quad (5.1.13)$$

The same argument as in Section 4.7 shows that the error committed in passing from (5.1.10) to (5.1.13) is of order $\varepsilon^2$ on expanding intervals of order $1/\varepsilon$.

In fact, it is not hard to show that for this example the error is actually of order $\varepsilon^2$ for all $t > 0$. Consider the first term of (5.1.12). When this term is expanded in $\varepsilon$ and two terms are retained, the error committed is bounded by a constant times $\varepsilon^2 e^{-\varepsilon t}$. This expression reaches a maximum of $\varepsilon^2 e^{-\varepsilon}$ at $t = 1$, and since this maximum is less than $\varepsilon^2$, the error is of order $\varepsilon^2$ as claimed. A similar analysis can be done for the second term of (5.1.12). The reason that a solution using only two time scales is able to remain valid for all time, rather than only on an expanding interval of order $1/\varepsilon$, is that the exponential damping acts to reduce not only the size of the solution but also of the error as $t \to \infty$. The mechanism is similar to that by which the regular perturbation solution of a damped oscillation was found to be valid for all time in Chapter 4, the difference being that in Chapter 4 the damping was of order $O(1)$ and here it is $O(\varepsilon)$; such damping is too weak to rescue the regular expansion from the ravages of its secular terms.

The second type of multiple scale approximation to be derived from (5.1.10) is a two-scale approximation using the scales

$$t^+ = \left(1 - \frac{5}{8}\varepsilon^2\right)t,$$

$$\tau = \varepsilon t. \quad (5.1.14)$$

Such an approximation can be obtained from (5.1.12) by dropping the dotted terms inside the trig functions:

$$y \simeq a e^{-t} \cos \left(t^+ + \frac{\tau}{2}\right) + \varepsilon a e^{-t} \sin \left(t^+ + \frac{\tau}{2}\right). \quad (5.1.15)$$

The final type of multiple scale expansion to be derived from (5.1.10) is a three-scale solution using $t$, $\tau$, and $\sigma := e^2 t$ (or $T_0$, $T_1$, and $T_2$). We leave it to the reader (Exercise 5.1.1) to check that the complete two-term, three-scale expansion of (5.1.10) is

$$y \simeq a e^{-t} \cos \left(t + \frac{1}{2}\tau - \frac{5}{8}\sigma\right)$$

$$+ \varepsilon a e^{-t} \left(1 - \frac{15}{48}\sigma\right) \sin \left(t + \frac{1}{2}\tau - \frac{5}{8}\sigma\right). \quad (5.1.16)$$
The "short form" of this three-scale expansion is obtained by dropping $\sigma$ from the term of order $\varepsilon$; this is equivalent to (5.1.15). We have said that adding a third time scale is usually an attempt to gain validity on a longer time interval. Here it is not needed for that purpose, since as noted above, (5.1.13) is already uniformly valid for all time because of the damping.

Exercises 5.1

1. Introduce $t$, $\tau$, and $\sigma$ as "independent" variables in (5.1.10) and expand in $\varepsilon$ to obtain (5.1.16).

5.2. THE FIRST ORDER TWO-SCALE APPROXIMATION

The idea of the two-scale method is to seek an approximate solution of the initial value problem

$$y + k^2 y = \varepsilon f(y, \dot{y}),$$

$$y(0) = \alpha,$$

$$\dot{y}(0) = \beta$$

(5.2.1)

in the form

$$y \sim y_0(t, \tau) + \varepsilon y_1(t, \tau) + \varepsilon^2 y_2(t, \tau) + \cdots ,$$

(5.2.2)

where $\tau = \varepsilon t$. In this section, only the first term of (5.2.2) will be determined. (Recall from the last section that $y_0$ is called the first approximation, rather than the zeroth, because it is not equal to the solution of the reduced problem.) As in the Lindstedt method, finding $y_0$ requires that the equation for $y_1$ be written down as well; the equation for $y_2$ will also be found, since it is needed in the next section.

It must be understood that (5.2.2) is a rather peculiar way to write the solution of an ordinary differential equation. There are actually only two independent variables, $t$ and $\varepsilon$, in (5.2.2); $\tau$ is a function of these two, and so is not independent. Nevertheless, the principal steps in finding the coefficients $y_n$ are carried out as though $t$, $\tau$, and $\varepsilon$ were independent variables. This is one reason why these steps cannot be justified rigorously in advance, but are merely heuristic. Secondly, it must be remarked that (5.2.2) is a generalized asymptotic expansion, since $\varepsilon$ enters both through the gauges (which are just the powers of $\varepsilon$) and also through the coefficients $y_n$ by way of $\tau$. Although there is no general theorem allowing the differentiation of a generalized asymptotic expansion term by term (see Section 1.8), it is nevertheless reasonable to construct the coefficients of (5.2.2) on the assumption that such differentiation is possible, and then to justify the resulting series by direct error estimation afterwards. The procedure is to take the total derivative of (5.2.2) with respect to $t$ twice, termwise, and substitute the result into (5.2.1). At the time of taking these derivatives, $\tau$ is regarded as equal to $\varepsilon t$, but afterwards, $\tau$ is treated as an independent
variable. Since the total dependence of each term \( y_n(t, \tau) \) upon \( t \) is a combination of its explicit dependence on \( t \) and its indirect dependence on \( t \) by way of \( \tau \), the first and second total derivatives of each term are given by the following formulas based on the chain rule:

\[
\begin{align*}
\frac{d}{dt} y_n(t, \tau) &= \frac{d}{dt} y_n(t, \varepsilon t) = \frac{\partial}{\partial t} y_n(t, \varepsilon t) + \varepsilon \frac{\partial}{\partial \tau} y_n(t, \tau) \\
&= y_{nt}(t, \tau) + \varepsilon y_{nt}(t, \tau), \\
\frac{d^2}{dt^2} y_n(t, \tau) &= y_{ntt}(t, \tau) + 2\varepsilon y_{ntt}(t, \tau) + \varepsilon^2 y_{ntt}(t, \tau).
\end{align*}
\] (5.2.3)

These equations are easiest to remember in operator form:

\[
\begin{align*}
\frac{d}{dt} &= \frac{\partial}{\partial t} + \varepsilon \frac{\partial}{\partial \tau}, \\
\frac{d^2}{dt^2} &= \frac{\partial^2}{\partial t^2} + 2\varepsilon \frac{\partial^2}{\partial t \partial \tau} + \varepsilon^2 \frac{\partial^2}{\partial \tau^2}.
\end{align*}
\] (5.2.4)

Substituting the derivatives of (5.2.2), computed in this way, into (5.2.1) gives the following expressions for the differential equation and initial conditions:

\[
\begin{align*}
(y_{0tt} + 2\varepsilon y_{0tt} + \varepsilon^2 y_{0t\tau}) + \varepsilon (y_{1tt} + 2\varepsilon y_{1tt} + \varepsilon^2 y_{1\tau}) \\
+ \varepsilon^2 (y_{2tt} + 2\varepsilon y_{2tt} + \varepsilon^2 y_{2\tau}) + \cdots + k^2 (y_0 + \varepsilon y_1 + \varepsilon^2 y_2 + \cdots), \\
= \varepsilon f(y_0 + \varepsilon y_1 + \cdots, y_0 + \varepsilon y_{0t} + \varepsilon^2 y_{1t} + \cdots),
\end{align*}
\] (5.2.5)

\[
\begin{align*}
y_0(0, 0) + \varepsilon y_1(0, 0) + \varepsilon^2 y_2(0, 0) + \cdots &= \alpha, \\
y_0(0, 0) + \varepsilon \{y_{0t}(0, 0) + y_{1t}(0, 0)\} + \varepsilon^2 \{y_{1t}(0, 0) + y_{2t}(0, 0)\} + \cdots &= \beta.
\end{align*}
\]

Expanding the right hand side of the differential equation in powers of \( \varepsilon \) and identifying coefficients of equal powers leads to the following sequence of initial value problems for the coefficients:

\[
\begin{align*}
y_{0tt} + k^2 y_0 &= 0, \\
y_0(0, 0) &= \alpha, \\
y_{0t}(0, 0) &= \beta; \\
y_{1tt} + k^2 y_1 &= f(y_0, y_{0t}) - 2y_{0t}, \\
y_1(0, 0) &= 0, \\
y_{1t}(0, 0) &= -y_{0t}(0, 0); \\
y_{2tt} + k^2 y_2 &= f_y(y_0, y_{0t})y_1 + f_y(y_0, y_{0t})(y_{0t} + y_{1t}) \\
&- y_{0tt} - 2y_{1tt}, \\
y_2(0, 0) &= 0, \\
y_{2t}(0, 0) &= -y_{1t}(0, 0).
\end{align*}
\] (5.2.6a, 5.2.6b, 5.2.6c)
It should be remembered that even the step of equating coefficients of equal powers of \( \varepsilon \), used in passing from (5.2.5) to (5.2.6), is not justified by any theorem about generalized asymptotic expansions (since there is no uniqueness theorem for such expansions). It is instead a heuristic assumption used to arrive at a candidate for an approximate solution, whose validity is to be determined afterwards by error analysis.

Since \( t \) and \( \tau \) are being treated (temporarily) as independent, the differential equation in (5.2.6a) is actually a partial differential equation for a function \( y_0 \) of two variables \( t \) and \( \tau \). However, since no derivatives with respect to \( \tau \) appear in (5.2.6a), it may be regarded instead as an ordinary differential equation for a function of \( t \), regarding \( \tau \) as merely an auxiliary parameter. Therefore the general solution of (5.2.6a) may be obtained from the general solution of the corresponding ordinary differential equation just by letting the arbitrary constants become arbitrary functions of \( \tau \):

\[
y_0(t, \tau) = A(\tau) \cos kt + B(\tau) \sin kt = \rho(\tau) \cos (\psi(\tau) - kt).
\]

The initial conditions of (5.2.6a) impose the following restrictions upon the otherwise arbitrary functions in (5.2.7):

\[
A(0) = \alpha, \\
B(0) = \beta/k, \\
\rho(0) = \sqrt{\alpha^2 + (\beta/k)^2}, \\
\psi(0) = \arctan(\beta/k\alpha).
\]

Up to this point, the calculations are routine, although they require taking careful thought for the meaning of each step, so that, for instance, the total and partial derivatives with respect to \( t \) are not confused. Now, however, we have reached a crucial point, because it becomes clear that the initial value problem (5.2.6a) for \( y_0 \) does not completely determine \( y_0 \). We have used all of the information contained in (5.2.6a), and the functions \( A \) and \( B \), or (alternatively) \( \rho \) and \( \psi \), are still undetermined except for their initial values (5.2.8). Some new idea is necessary in order to complete the determination of these functions, and hence of \( y_0 \).

The new idea needed is actually not so new—it is the concept of eliminating secular terms, used repeatedly in Chapter 4. However, the notion of secular term required here is slightly different from that of Chapter 4. It can be motivated as follows. Our aim is that (5.2.2) should be a uniformly valid asymptotic expansion, on expanding intervals of (at least) order \( 1/\varepsilon \), of the exact solution of (5.2.1). In an asymptotic expansion, the error after the first term (that is, the difference between \( y_0 \) and the exact solution) is expected to be of the order of the second term, \( \varepsilon y_1 \). Therefore to try to make the error be of order \( \varepsilon \) on expanding intervals of order \( 1/\varepsilon \), we should arrange things so that \( \varepsilon y_1 \) is of order \( \varepsilon \) on such intervals, or in other words, that \( y_1 \) is bounded on these intervals. It is not guaranteed in advance that
this will produce a valid asymptotic solution, but the plan we are adopting is the most likely to lead to the desired result. (In the language of Section 1.8, we know that in order to be \textit{uniformly valid} on expanding intervals of order $1/e$, the series must be \textit{uniformly ordered} on these intervals, and that in turn requires that the coefficients must be bounded. On the other hand, achieving this uniform ordering does not automatically guarantee that the series is uniformly valid.) Since $r$ is bounded on expanding intervals of order $1/e$, it is permissible for $y_1$ to contain so-called $\tau$-secular terms (or $T_1$-secular terms) which involve factors of $r$, but not $t$-secular terms which involve factors of $t$. (Both kinds of terms are secular in the sense of being unbounded on the positive $t$ axis, but only the latter are unbounded on expanding intervals of order $1/e$.)

With this idea in mind, we shall examine the differential equation (5.2.6b) for $y_1$ in order to see whether $y_1$ will contain $t$-secular terms. In view of (5.2.7), this differential equation takes the form

$$y_{1tt} + k^2y_1 = f(A(r) \cos kt + B(r) \sin kt, -kA(r) \sin kt + kB(r) \cos kt)$$

$$+ 2kA'(r) \sin kt - 2kB'(r) \cos kt$$

(5.2.9)

in Cartesian variables. Since the right hand side of (5.2.9) is periodic in $t$ with period $2\pi/k$, it can be expanded in a Fourier series in $t$ (for fixed $r$). Then $y_1$ will be free of $t$-secular terms if and only if this Fourier series has no terms in $\sin kt$ and $\cos kt$; these terms, if present, would be linearly resonant with the free frequency $k$. Setting the coefficients of these terms equal to zero imposes two conditions on $A(r)$ and $B(r)$, namely, the following differential equations:

$$\frac{dA}{dr} = P(A, B) :=$$

$$- \frac{1}{2\pi} \int_0^{2\pi/k} f(A \cos kt + B \sin kt, -kA \sin kt + kB \cos kt) \sin kt \, dt,$$

$$\frac{dB}{dr} = Q(A, B) :=$$

$$+ \frac{1}{2\pi} \int_0^{2\pi/k} f(A \cos kt + B \sin kt, -kA \sin kt + kB \cos kt) \cos kt \, dt.$$  

(5.2.10)

For any specific function $f$, the integrals on the right hand side can (at least in principle, and often in fact) be evaluated, so that $P(A, B)$ and $Q(A, B)$ become known functions and (5.2.10) becomes an explicit set of differential equations for $A$ and $B$. These equations are to be solved with the initial conditions given by (5.2.8); this completes the determination of $A$ and $B$, and hence of $y_0$ (see (5.2.7)).

Since our purpose in this section is only to find the first approximation, we will not solve (5.2.9) for $y_1$. It was necessary to make use of the equation
5.2. THE FIRST ORDER TWO-SCALE APPROXIMATION

for \( y_1 \), but only in order to arrive at equations (5.2.10) for \( A \) and \( B \). Briefly, the way to continue is as follows. Now that \( A \) and \( B \) are fixed, the equation (5.2.9) can be solved for \( y_1 \); but this equation does not completely determine \( y_1 \), rather it introduces further unknown functions that must be determined by eliminating resonant terms from the equation for \( y_2 \). This idea will be developed further in the next section.

In general, the functions \( P(A, B) \) and \( Q(A, B) \) in (5.2.10) are nonlinear, and at first it is not apparent how to solve these differential equations for \( A \) and \( B \). In terms of the polar variables \( \rho \) and \( \psi \), the corresponding equations are easier to solve. Therefore we will repeat the discussion of how to compute \( y_0 \), using these polar variables. When the polar form of (5.2.7) is substituted into (5.2.6b), the following equation is obtained, corresponding to (5.2.9):

\[
y_{1tt} + k^2 y_1 = f(\rho(\tau) \cos(\psi(\tau) - kt), \rho(\tau) \sin(\psi(\tau) - kt))
- 2\rho'(\tau) \sin(\psi(\tau) - kt) - 2\rho(\tau) \psi'(\tau) \cos(\psi(\tau) - kt).
\]

(5.2.11)

Recall that this equation is treated as an ordinary differential equation for \( y_1 \) as a function of \( t \), with \( \tau \) appearing as a parameter (and therefore treated as constant); the condition that \( y_1 \) have no \( t \)-secular terms is that the right hand side have no terms in \( \sin kt \) and \( \cos kt \) when expanded in a Fourier series. It is not difficult to show that this condition is equivalent to

\[
\rho' = \frac{1}{2\pi} \int_0^{2\pi/k} f(\rho \cos(\psi - kt), \rho \sin(\psi - kt)) \sin(\psi - kt) \, dt,
\]

\[
\psi' = \frac{1}{2\pi \rho} \int_0^{2\pi/k} f(\rho \cos(\psi - kt), \rho \sin(\psi - kt)) \cos(\psi - kt) \, dt.
\]

(5.2.12)

One way to see this is shown in Exercise 5.2.1.

Another way, which makes the result obvious but depends on theory that some readers may not know, is to observe that (for fixed \( \tau \)) the functions \( \sin(\psi(\tau) - kt) \) and \( \cos(\psi(\tau) - kt) \) span the same vector space of functions as \( \sin kt \) and \( \cos kt \). Then (5.2.12) merely says that the projection of the right hand side of (5.2.11) onto these new basis vectors is zero, using the standard inner product for the space of functions of period \( 2\pi/k \).

Next, (5.2.12) can be simplified by the substitution \( \theta = \phi - kt \) to the form

\[
\rho' = F(\rho) := \frac{1}{2\pi k} \int_0^{2\pi} f(\rho \cos \theta, k \rho \sin \theta) \sin \theta \, d\theta,
\]

\[
\psi' = G(\rho) := \frac{1}{2\pi k \rho} \int_0^{2\pi} f(\rho \cos \theta, k \rho \sin \theta) \cos \theta \, d\theta.
\]

(5.2.13)
These are the promised equations, which are simpler than the corresponding equations (5.2.10) in Cartesian variables since $F$ and $G$ depend upon only one variable $p$, whereas $P$ and $Q$ depend upon both $A$ and $B$. Although both (5.2.10) and (5.2.13) are in general nonlinear, the equations in (5.2.13) are "solvable by quadrature," that is, the problem of solving them is reducible to a problem of integration. Namely, the first equation can be written in separated form $d\rho/F(\rho) = d\tau$ and integrated, using the initial condition from (5.2.8); after the solution $\rho(\tau)$ is found, the second equation becomes $d\psi = G(\rho(\tau)) d\tau$, where the right hand side is a known function of $\tau$ to be integrated. Provided these two integrations can be done, the solution is complete.

Of course, since (5.2.10) is equivalent to (5.2.13) by a simple change of variables, it also is solvable by quadrature although that is not obvious from its form. This observation is useful, since the Cartesian form has a definite advantage when it comes to computing higher approximations (Section 5.3). The procedure for solving (5.2.10) by quadrature will be developed in the examples below.

This completes the heuristic derivation of the first order two-scale approximation. It has already been mentioned that the same approximation will be derived again, from an entirely different point of view, in Chapter 6. The question remains: Does this approximation accomplish its purpose? The answer is yes. The following theorem is proved at the end of Section 5.5:

**Theorem 5.2.1.** For fixed $\alpha$ and $\beta$ there exist constants $\epsilon_1$, $c$, and $T$ such that the exact solution $y(t, \epsilon)$ of (5.2.1) satisfies

$$|y(t, \epsilon) - y_0(t, \epsilon t)| \leq ce \quad \text{for} \quad 0 \leq t \leq T/\epsilon, \quad 0 \leq \epsilon \leq \epsilon_1. \quad (5.2.14)$$

A similar theorem giving more information about $T$ is proved in Chapter 6 (Theorem 6.2.2).

**Example 5.2.1. A Linear Problem**

As a first illustration of the method of multiple scales, we will compute the first approximation to the (exactly solvable) linear problem discussed in Section 5.1:

$$\dot{y} + 2\epsilon \dot{y} + (1 + \epsilon)y = 0,$$

$$y(0) = \alpha,$$

$$\dot{y}(0) = 0. \quad (5.2.15)$$

The desired approximation has already been derived from the exact solution; it is the first term of (5.1.13). Here it will be obtained by following the procedures described above. As always, it is more instructive to work
out each step of the procedure for each application, rather than to apply formulas such as (5.2.10) or (5.2.13) mechanically. Substitution of (5.2.2) into (5.2.15), using (5.2.4), leads to the following sequence of linear initial value problems (compare (5.2.6)):

\[
\begin{align*}
\dot{y}_0 + y_0 & = 0, \\
y_0(0, 0) & = \alpha, \\
\dot{y}_0(0, 0) & = 0; \\
\dot{y}_1 + y_1 & = -(y_0 + 2y_0 + 2y_0), \\
y_1(0, 0) & = 0, \\
\dot{y}_1(0, 0) & = -y_0(0, 0); \\
\dot{y}_2 + y_2 & = -(2y_0 + y_0 + y_1 + 2y_1 + 2y_1), \\
y_2(0, 0) & = 0, \\
\dot{y}_2(0, 0) & = -y_1(0, 0).
\end{align*}
\]

We will need only the first two of these in this section, but the third will be used later. The solution of (5.2.16a) in “amplitude/phase” form is

\[
\begin{align*}
y_0 = \rho(\tau) \cos(\psi(\tau) - t), \\
\rho(0) & = \alpha, \\
\psi(0) & = 0,
\end{align*}
\]

where \(\rho(\tau)\) and \(\psi(\tau)\) are as yet undetermined except for their initial conditions. The equation for \(y_1\) now becomes

\[
\dot{y}_1 + y_1 = -\rho(1 + 2\psi') \cos(\psi - t) - 2(\rho' + \rho) \sin(\psi - t).
\]

Since \(\psi\) is a function of \(\tau\) only, it is treated as a constant while solving the equation for \(y_1\). Rather than expand the right hand side in a Fourier series in \(t\), it is more convenient to expand it in \(\psi - t\); in fact, it is already expressed as a Fourier series in \(\psi - t\), which contains only two terms, both of them resonant with the free frequency 1. Therefore the heuristic rule that \(t\)-secular terms should be eliminated compels us to choose \(\rho(\tau)\) and \(\psi(\tau)\) to satisfy

\[
\begin{align*}
\rho' & = -\rho, \\
\psi' & = -1/2.
\end{align*}
\]

The solution of these equations, with the initial conditions given in (5.2.17), are

\[
\begin{align*}
\rho(\tau) & = \alpha e^{-\tau}, \\
\psi(\tau) & = -\tau/2,
\end{align*}
\]
leading to the complete first approximation,

$$y_0(t, \tau) = a e^{-\tau} \cos(-\tau/2 - t) = a e^{-\tau} \cos(t + \tau/2).$$  \hfill (5.2.21)

Now we will repeat these calculations in Cartesian variables. As remarked above, it is more difficult to compute the first approximation in Cartesian variables than in polar, but for higher approximations the reverse is true. Therefore it is worthwhile to understand how to compute the first approximation in Cartesian variables. Beginning with (5.2.16a), the solution can be written

$$y_0 = A(t) \cos t + B(t) \sin t,$$

$$A(0) = a,$$

$$B(0) = 0.$$  \hfill (5.2.22)

The equation for $y_1$ is

$$y_{1tt} + y_1 = -(A + 2B + 2B') \cos t + (-B + 2A + 2A') \sin t.$$  \hfill (5.2.23)

The right hand side is, once again, already expanded in a Fourier series (this time in $t$), having two terms, both of which are resonant. The non-secularity conditions are

$$A' = -A + B/2,$$

$$B' = -A/2 - B.$$  \hfill (5.2.24)

This is a linear system of two differential equations in two unknown functions, and can be solved as such, for instance by matrix methods. But it is linear only because the original equation (5.2.15) is linear; so to solve (5.2.24) by linear methods will not help with more general problems. Instead, observe that by the chain rule and by (5.2.24),

$$\frac{d}{d\tau} (A^2 + B^2) = 2AA' + 2BB' = -2(A^2 + B^2).$$  \hfill (5.2.25)

This equation can be solved for the quantity $A^2 + B^2$:

$$A^2 + B^2 = [A(0)^2 + B(0)^2] e^{-2\tau} = (ae^{-\tau})^2.$$  \hfill (5.2.26)

(Notice that since $A^2 + B^2 = \rho^2$, we have actually introduced polar coordinates "on the sly" here. This is the secret of how to solve (5.2.10) by quadrature.) It follows that there exists $\xi(\tau)$ such that

$$A(\tau) = ae^{-\tau} \cos \xi(\tau),$$

$$B(\tau) = ae^{-\tau} \sin \xi(\tau).$$  \hfill (5.2.27)
To compute $\xi$, substitute the first equation of (5.2.27) into the first equation of (5.2.24) to obtain

$$2A' + 2A - B = -\alpha(2\xi' + 1) \sin \xi = 0$$

or

$$\xi' = -1/2,$$  \hspace{1cm} (5.2.28)

so that $\xi(\tau) = \xi(0) - \tau/2$. (Compare (5.2.28) with the second equation of (5.2.19): $\xi$ is another "secret" polar coordinate.) The initial value $\xi(0) = 0$ follows from (5.2.22) and (5.2.27), from which

$$A(\tau) = \alpha e^{-\tau} \cos \frac{\tau}{2},$$

$$B(\tau) = -\alpha e^{-\tau} \sin \frac{\tau}{2},$$

giving

$$y_0(t, \tau) = \alpha e^{-\tau} \left\{ \cos \frac{\tau}{2} \cos t - \sin \frac{\tau}{2} \sin t \right\},$$  \hspace{1cm} (5.2.30)

in agreement with (5.2.21).

**Example 5.2.2. The Van der Pol Equation**

For the next example, consider the Van der Pol equation

$$\dot{y} + \epsilon(y^2 - 1)\dot{y} + y = 0,$$

$$y(0) = \alpha,$$

$$\dot{y}(0) = 0,$$  \hspace{1cm} (5.2.31)

which has been studied in Section 4.4. Since this is a nonlinear equation, the exact solution is not available for comparison with the approximation found by multiple scale methods. However, it has been seen that for small $\epsilon$ there is a unique limit cycle and it is stable. These features can be recognized in the approximate solution by two-timing, which is (see Exercise 5.2.2):

$$y \approx \frac{2\alpha}{\sqrt{\alpha^2 + (4 - \alpha^2)e^{-\epsilon t}}} \cos t.$$  \hspace{1cm} (5.2.32)

It is possible to recognize some of the properties of the exact solutions of Van der Pol's equation in this approximate solution. For instance, as $t \to \infty$ in (5.2.32), the transient effect (on the time scale $\tau$) dies out, leaving the periodic solution

$$y \approx 2 \cos t$$  \hspace{1cm} (5.2.33)
which coincides with the Lindstedt approximation to the limit cycle found in Exercise 4.4.1. Thus the approximate solution exhibits a stable limit cycle close to the exact limit cycle of the Van der Pol equation. One might wonder why the two-timing approximation is this good; after all, it is only intended to be uniformly valid on an expanding interval, not for all time. It turns out that a careful analysis of the error in (5.2.33) shows that while the phase $\psi$ is only approximated well on an expanding interval, the amplitude $\rho$ is approximated well for all time. As in the previous example, this is due to damping (or more precisely, to the attracting character of the limit cycle). In this problem, damping only affects the error in a direction normal to the limit cycle, and “in-track error” continues to accumulate in the tangential direction just as it does for the Lindstedt approximation to the limit cycle. (See Section 4.3 for the notion of “in-track error.”) It is not always permissible to draw conclusions about the behavior of a system as time approaches infinity from approximate solutions, but when the behavior in question is sufficiently “strong” in some sense, it will survive in the approximations. The further exploration of this topic is beyond this book; see Notes and References.

Exercises 5.2

1. Derive (5.2.12) from the nonsecularity condition on (5.2.11) by another method than that indicated in the text. Hint: Multiply the right hand side of (5.2.11) by $\sin kt$ and by $\cos kt$, and integrate from 0 to $2\pi/k$. Set the results equal to zero, and use trig identities.

2. Find the first order two-scale approximation for the Van der Pol equation (5.2.30). Do not use the general formulas, but carry out the procedure of substituting (5.2.2) into (5.2.30) and obtaining equations for $y_0$ and $y_1$. Use polar form for the solution for $y_0$, and fix $\rho(\tau)$ and $\psi(\tau)$ by eliminating resonant terms from the equation for $y_1$. After you have obtained your approximate solution, compare it with the solution to Exercise 2.4.4b. Repeat the calculation using Cartesian variables.

3. Find the first order two-scale approximation for the autonomous Duffing equation $\ddot{y} + y + \varepsilon y^3 = 0$, $y(0) = \alpha$, $\dot{y}(0) = 0$. Do the calculation in both polar and Cartesian variables.

4. Find a leading order approximation (by the multiple scale method) for $\ddot{y} + \varepsilon y \dot{y} + y = 0$, $y(0) = 1$, $\dot{y}(0) = 0$. Compare the result with Exercise 4.3.6.

5.3* HIGHER ORDER APPROXIMATIONS

In this section (which may be omitted without loss of continuity) we will investigate three strategies for finding higher order multiple scale approximations. These three strategies have already been mentioned at the end
of Section 5.1, where three types of second order approximations were given for the linear initial value problem (5.1.9). Briefly stated, the three approaches are: (1) to continue as in Section 5.2 with the two time scales $t$ and $\tau$, computing higher order terms in the series (5.2.2); (2) to use two time scales, but replace $t$ by a “strained” time scale similar to that used in the Lindstedt method; and (3) to add more time scales besides $t$ and $\tau$.

The purpose of the first strategy is to improve the accuracy of the first approximation, without attempting to increase the length of time $O(1/\varepsilon)$ for which the approximation is valid. For problems of the form (5.1.1) it can be shown that the two-time-scale expansion using $t$ and $\tau$ which was begun in the last section can be carried out to any order, and that the solution (5.2.2) taken through the term $y_{k-1}$ has error $O(\varepsilon^k)$ on expanding intervals of length $O(1/\varepsilon)$. Thus, at least in theory, the first strategy is always successful at achieving its goal. However, to carry out the solution in practice requires solving certain differential equations in order to eliminate secular terms; these differential equations are in general nonlinear, and therefore may not have “closed form” solutions (that is, explicit solutions in terms of elementary functions). So the fact that the solutions are possible “in theory” does not always guarantee that the calculations are possible in practice, although in many cases they are.

The second and third strategies are more ambitious. Their aim is not only to improve the asymptotic order of the error estimate, but also to extend the validity of the approximations to “longer” intervals of time, that is, expanding intervals of length $O(1/\varepsilon^2)$ or longer. In other words, these methods are an attempt to recapture some of the “trade-off” property enjoyed by Lindstedt expansions in the periodic case. These methods were originally developed by heuristic reasoning only, and there does not yet exist a fully adequate rigorous theory explaining their range of validity. There are a number of specific problems of the form (5.1.1) for which these methods work, or seem to work. There are others for which they definitely do not work. Because the situation is not yet fully understood, we will not attempt to state any general results for these strategies, but will illustrate the third strategy briefly and comment on the difficulties.

It should be mentioned that there is another approach to finding higher order approximations for these problems, namely the method of “higher order averaging” which will be developed in the next chapter. This method does not require specifying a set of time scales in advance. Instead, the necessary time scales appear automatically in the course of solving the problem. The approximations constructed by averaging are always valid at least on expanding intervals of order $1/\varepsilon$, and sometimes on longer intervals. It would appear that this method is the most powerful.

To explain the higher order two-time-scale method using $t$ and $\tau$, we will continue the solution of the linear example (5.2.15) from where it was left off in Section 5.2. At that time we had found $y_0$, either in the form...
(5.2.21) or (5.2.30). In doing so, we had eliminated resonant terms from the right hand side of (5.2.16b) and had in fact found that the entire right hand side was resonant. Therefore we begin with (5.2.16b) in the form

\begin{align}
    y_{1tt} + y_1 &= 0, \\
    y_1(0,0) &= 0, \\
    y_{1t}(0,0) &= \alpha.
\end{align}

(5.3.1)

Although the polar form was simplest for finding \( y_0 \), the computational advantage seems to lie with the Cartesian form this time, so we write the solution of (5.3.1) as

\[ y_1(t, \tau) = C(\tau) \cos t + D(\tau) \sin t, \]

(5.3.2)

As before, the solution is not uniquely determined by the information available, and it is necessary to adopt heuristic reasoning to fix the functions \( C \) and \( D \) completely. Since the error in the approximation \( y \approx y_0 + \varepsilon y_1 \) is expected to behave like \( \varepsilon^2 y_2 \), the procedure is to eliminate \( t \)-secular terms from \( y_2 \) so that it remains bounded on expanding intervals of order \( 1/\varepsilon \). To this end we examine equation (5.2.16c), which can be written out as

\[ y_{2tt} + y_2 = -\{A'' + 2A' + 2D' + C + 2D\} \cos t \]

\[ - \{B'' + 2B' - 2C' + D - 2C\} \sin t, \]

(5.3.3)

where \( A \) and \( B \) are as in (5.2.29), and \( C \) and \( D \) are as yet unknown. Both terms on the right hand side of (5.3.3) are resonant and must be set equal to zero. Making use of the expressions for \( A \) and \( B \), the resulting nonsecularity conditions are

\[ C' = -C + \frac{1}{2} D + \frac{5}{8} \alpha e^{-\tau} \sin \frac{\tau}{2}, \]

\[ D' = -\frac{1}{2} C - D + \frac{5}{8} \alpha e^{-\tau} \cos \frac{\tau}{2}. \]

(5.3.4)

This inhomogeneous linear system of ordinary differential equations is solvable, with the initial conditions given in (5.3.2):

\[ C(\tau) = \alpha e^{-\tau} \left( \frac{5}{8} \tau + 1 \right) \sin \frac{\tau}{2}, \]

\[ D(\tau) = \alpha e^{-\tau} \left( \frac{5}{8} \tau + 1 \right) \cos \frac{\tau}{2}. \]

(5.3.5)
When this is substituted into (5.3.2), it is easy to see that $y_0 + \varepsilon y_1$ coincides with the approximation (5.1.13) developed from the exact solution.

If we had solved (5.3.1) in polar form $y = \rho_1 \cos(\psi_1 - t)$, the nonsecularity conditions and initial conditions to be satisfied by $\rho_1(\tau)$ and $\psi_1(\tau)$ would have been (after some calculation)

\[
\rho' + \rho = \frac{5}{8} \alpha e^{-\tau} \sin \left( \psi + \frac{\tau}{2} \right),
\]

\[
\rho(2\psi' + 1) = \frac{5}{4} \alpha e^{-\tau} \cos \left( \psi + \frac{\tau}{2} \right),
\]

\[
\rho(0) = \alpha,
\]

\[
\psi(0) = \frac{\pi}{2}.
\]

It is possible to solve these by clever inspection; the solution is

\[
\rho = \alpha e^{-\tau} \left( \frac{5}{8} \tau + 1 \right),
\]

\[
\psi = \frac{\pi}{2} - \frac{\tau}{2}.
\]

However there does not seem to be any technique behind this, and if one attempts to solve the original problem (5.2.15) with initial velocity $\beta$ instead of zero, the polar equation becomes intractable, whereas the analog of (5.3.4) is still an inhomogeneous linear system. This is why we said that the Cartesian form seems preferable for this problem. Nevertheless, it must be emphasized that (5.3.4) is linear only because the original problem (5.3.4) is linear, and there is no general procedure for solving the nonsecularity conditions that arise in the course of the two-timing method. As mentioned before, the problem is a practical one, not a theoretical one: The nonsecularity conditions in the two-timing method (using "unstrained" fast time $t$ and slow time $\tau$) are always initial value problems of a sort which, in theory, have solutions, although they may not be solvable in "closed form" using elementary functions.

The method to be investigated next has been used to solve a number of applied problems and is quite popular in the engineering literature. However, it rests on quite shaky foundations from a mathematical point of view. The purpose of the following example is not so much to teach the method, as to discuss it and point out some of the difficulties with it. Anyone wishing to learn the method for practical use will have to study many examples from the literature to see how these difficulties are dealt with on an ad hoc basis in different situations.

The problem we will study is once again the linear example (5.2.15). The solution will be sought in the form

\[
y \cong y_0(t, \tau, \sigma) + \varepsilon y_1(t, \tau, \sigma) + \varepsilon^2 y_2(t, \tau, \sigma),
\]

(5.3.6)
where \( \tau = \varepsilon t \) and \( \sigma = \varepsilon^2 t \). Working with the operator equation

\[
\frac{d}{dt} = \frac{\partial}{\partial t} + \varepsilon \frac{\partial}{\partial \tau} + \varepsilon^2 \frac{\partial}{\partial \sigma},
\]

one obtains the following sequence of differential equations and initial conditions:

\begin{align*}
Y_{0tt} + Y_0 &= 0, \\
Y_0(0, 0, 0) &= \alpha, \\
Y_{0t}(0, 0, 0) &= 0; \\
Y_{1tt} + Y_1 &= -(Y_0 + 2Y_{0t} + 2Y_{0tt}), \\
Y_1(0, 0, 0) &= 0, \\
Y_{1t}(0, 0, 0) &= -Y_{0t}(0, 0, 0); \\
Y_{2tt} + Y_2 &= -(2Y_{0r} + Y_{0tt} + 2Y_{0t\sigma} + Y_1 + 2Y_{1t} + 2Y_{1tt}), \\
Y_2(0, 0, 0) &= 0, \\
Y_{2t}(0, 0, 0) &= -Y_{1t}(0, 0, 0) - Y_{0t}(0, 0, 0).
\end{align*}

These equations are very similar to (5.2.16), the only differences being that the initial conditions occur at \((0, 0, 0)\) rather than \((0, 0)\) and that (5.3.7c) has additional terms involving derivatives with respect to \(\sigma\). Therefore the solution of these equations proceeds very much like that of (5.2.16) until (5.3.7c) is reached. It is only necessary to remember that each function is allowed to depend on \(\sigma\). Thus, the solution of (5.3.7a) is

\[
y_0 = A(\tau, \sigma) \cos t + B(\tau, \sigma) \sin t,
\]

with

\[
A(0, 0) = \alpha, \\
B(0, 0) = 0;
\]

compare (5.2.22). Then it is possible to follow through the steps from (5.2.23) to (5.2.26) almost exactly, except that \(A'\) and \(B'\) must be written \(A_\tau\) and \(B_\tau\) because they depend also on \(\sigma\), and in (5.2.25) one has \(\frac{\partial}{\partial \tau}\). In place of (5.2.27), one gets

\[
A(\tau, \sigma) = K(\sigma)e^{-\tau} \cos \xi(\tau, \sigma), \\
B(\tau, \sigma) = K(\sigma)e^{-\tau} \sin \xi(\tau, \sigma),
\]

where

\[
K(0) = \alpha, \\
\xi(0, 0) = 0.
\]
As in (5.2.28), $\zeta_\tau = -1/2$; therefore

$$\zeta(\tau, \sigma) = \eta(\sigma) - \frac{\tau}{2}$$

for some function $\eta(\sigma)$ satisfying $\eta(0) = 0$. Putting (5.3.8) together with the results just obtained, one finds that at this point $y_0$ is given by

$$y_0(t, \tau, \sigma) = K(\sigma)e^{-\tau} \cos \left(t + \frac{\tau}{2} - \eta(\sigma)\right), \quad (5.3.9)$$

where $K$ and $\eta$ are as yet unspecified functions satisfying

$$K(0) = \alpha,$$
$$\eta(0) = 0.$$

Here we see the first serious difference between this method and the two-scale method: in the latter, $y_0$ is completely specified when the resonant terms are eliminated from the $y_1$ equation, while in the present method there is a dependence on $\sigma$ which is not yet determined.

Having eliminated the resonant terms from the $y_1$ equation, one is left with

$$y_{1tt} + y_1 = 0,$$
$$y_1(0, 0, 0) = 0,$$
$$y_{1tt}(0, 0, 0) = \alpha,$$

resembling (5.3.1). In solving (5.3.1) by (5.3.2), we used $\cos t$ and $\sin t$ as a fundamental set of solutions, but in the present case later calculations will be simplified if we use the cosine and sine of the quantity $t + \tau/2 - \eta(\sigma)$ appearing in (5.3.9); this is permissible since for fixed $\tau$ and $\sigma$ these are linearly independent solutions of the differential equation. Therefore we write

$$y_1(t, \tau, \sigma) = C(\tau, \sigma) \cos(t + \frac{\tau}{2} - \eta(\sigma)) + D(\tau, \sigma) \sin(t + \frac{\tau}{2} - \eta(\sigma)). \quad (5.3.10)$$

It is easy to check (using the fact that $\eta(0) = 0$ and the initial conditions for $y_1$) that

$$C(0, 0) = 0,$$
$$D(0, 0) = \alpha.$$

Until now the calculations have been routine. We are now rapidly approaching the point at which the fundamental nature of the method (and
its difficulties) will become clear. Substituting (5.3.9) and (5.3.10) into the
right hand side of (5.3.7c), one obtains the following equation for \( y_2 \):

\[
y_{2tt} + y_2 = \left[ \left( \frac{5}{4} - 2\eta' \right) Ke^{-\tau} - 2(D + D_t) \right] \cos(t + \frac{\tau}{2} - \eta) \\
+ \left[ 2K'e^{-\tau} + 2(C + C_t) \right] \sin(t + \frac{\tau}{2} - \eta).
\]

(5.3.11)

Here, of course, the prime denotes a derivative with respect to \( \sigma \), since it is
applied to functions of \( \sigma \) alone. In order that \( y_2 \) will contain no \( t \)-secular
terms (terms proportional to \( t \)), the resonant terms must be eliminated
from the right hand side of (5.3.11), as usual. Since the entire right hand
side is resonant, the coefficients of the cosine and sine terms must vanish,
leading to the equations

\[
C_t = -C - K'e^{-\tau}, \\
D_t = -D + (\frac{5}{8} - \eta') Ke^{-\tau}.
\]

(5.3.12)

This provides differential equations for \( C \) and \( D \) once \( K \) and \( \eta \) are known,
but is of no help in determining the latter functions which are still un-
known. It is clear that a new principle is required, besides the avoidance
of \( t \)-secular terms, which will serve to determine \( K \) and \( \eta \).

Recall the reason for avoiding \( t \)-secular terms: A term proportional to \( t \)
will grow with time (assuming the other factors do not approach zero) and
will become of order \( 1/e \) at the end of an expanding interval of length \( 1/e \).
If a term of formal order \( e^n \) is \( t \)-secular, its actual order on the expanding
interval is \( e^{n-1} \), and the series fails to be uniformly ordered. Now the aim
of a three-scale method is to achieve uniform validity on an expanding
interval of length \( 1/e^2 \). On an interval of this length, a \( \tau \)-secular term will
produce the same type of disordering that a \( t \)-secular term produces on
the shorter expanding interval. Now it is clear from (5.3.9) that \( \tau \)-secular
terms cannot arise in \( y_0 \). (And in any case, according to the definition of
uniform ordering in Section 1.8, the leading term of a perturbation series
does not matter.) So the first possibility for such a term occurs in \( y_1 \); and
in fact \( y_1 \) will contain a \( \tau \)-secular term if and only if \( C \) or \( D \) contains such
a term. So we propose the following rule: \( K(\sigma) \) and \( \eta(\sigma) \) should be chosen
in such a way that the solutions \( C \) and \( D \) of (5.3.12) contain no \( \tau \)-secular
terms.

We will now test this proposed rule by examining the solutions of
(5.3.12). Remember that the coefficients \( K' \) and \( (5/8 - \eta') \) are functions
of \( \sigma \) alone and therefore may be treated as constants in solving (5.3.12)
for \( C \) and \( D \) as functions of \( \tau \). Now it is apparent that the "forcing" terms
proportional to \( e^{-\tau} \) produce a response proportional to \( e^{-\tau} \); this can be
seen by explicitly solving the equations, or by the fact that the associated
homogeneous equations have a solution proportional to \( e^{-\tau} \) and so the
forced equations respond with an additional factor of $\tau$. Now a term of the form $te^{-t}$ is often considered as a $\tau$-secular term. In fact, we shall show below that this is not the case; but for the moment let us regard this term as something to be avoided, because it contains a factor of $\tau$. In order to avoid this term it is necessary to take

$$K' = 0, \quad \eta' = \frac{5}{8}. \quad (5.3.13)$$

With the initial conditions, this implies

$$K(\sigma) = \alpha, \quad \eta(\sigma) = \frac{5}{8} \sigma.$$ 

Finally, then, $y_0$ is completely determined:

$$y_0 = a e^{-t} \cos \left( t + \frac{\tau}{2} - \frac{5\sigma}{8} \right), \quad (5.3.14)$$

in agreement with the first term of the expansion (5.1.16) derived from the exact solution.

Although it appears that we have solved this problem successfully, there is a difficulty: As hinted above, the term $te^{-t}$ is not actually a $\tau$-secular term. To understand why, recall the precise meaning of a secular term: a term that is unbounded (on the domain of interest) and therefore causes the perturbation series to be not uniformly ordered. Now $te^{-t}$ attains a maximum and then decreases to zero as $t$ approaches infinity. Therefore it is not secular, either on an expanding interval or indeed on the entire real line. It follows that in fact there is no need to impose the conditions (5.3.13) which we used to determine $K$ and $\eta$. It is convenient to do so, because these conditions simplify the equations (5.3.12). But we would have obtained just as good a solution (as far as the presence of secular terms is concerned) had we made any choice at all for $K$ and $\eta$. The ultimate reason that $K$ and $\eta$ are not unique in this problem is that three scales are not actually needed to solve it: We have already pointed out in Section 5.1 that the two-scale solution for this problem is in fact valid for all time, because of the damping.

In the literature, one finds at least three justifications given for requiring (5.3.13) in this problem. One is that it simplifies the equations (5.3.12). Another is that it eliminates the term $te^{-t}$, which is considered as a bad term merely because it looks secular. The third is that it makes the ratio $y_1/y_0$ bounded. If $y_1$ contains a term proportional to $te^{-t}$, then indeed this ratio is unbounded; but this is not harmful. The demand that $y_{n+1}/y_n$ be bounded comes from the idea that the $n + 1$st term of a perturbation series should be small compared to the $n$th term. But a careful study of the requirements for
uniform asymptotic validity (see the discussion surrounding Theorem 1.8.1) shows that the correct concept of uniform ordering involves the boundedness of the coefficients and not of their ratios. So in the present problem the only valid justification of (5.3.12) is the first: that it simplifies the equations.

In the foregoing discussion we have emphasized repeatedly that in this problem the requirement to eliminate \( r \)-secular terms does not determine \( K \) and \( \eta \) uniquely. In general, there are three \textit{a priori} possibilities, all of which do in fact occur in specific problems: There may be a unique solution that is free of \( r \)-secular terms; there may be many solutions that are free of these terms (as in the problem treated above); or there may be no solution that is free of these terms. Clearly nonuniqueness of the solution (as in our case) is not a serious drawback for a method. What is most disconcerting about the three-scale method is the existence of problems for which the equations corresponding to (5.3.12) have no solutions that are free of \( r \)-secular terms. When this is encountered, the method simply fails. An example of such a problem is the following equation, which combines a "Duffing term" of order \( \varepsilon \) with a "Van der Pol term" of order \( \varepsilon^2 \):

\[
y + \varepsilon^2(y^2 - 1)\dot{y} + y + \varepsilon y^3 = 0.
\] \hspace{1cm} (5.3.15)

The ambitious reader may wish to attempt a three-scale solution of this problem, although the calculations are formidable (and are best handled with a notation that uses complex variables). But the fact is that at the end one finds there is no way to eliminate secular terms. (A reference is given in Section 5.7.) At this point there is no general theory that would explain why the method fails for some problems and succeeds with others.

In cases for which the three-scale method works (formally), the resulting series is uniformly ordered on expanding intervals of length \( 1/\varepsilon^2 \). This in itself does not imply that the series is uniformly valid on these intervals; as usual, uniform ordering of the terms in the series says nothing about the error. But it is at least plausible that such solutions are uniformly valid. We will not address the question of error estimation for the methods discussed in this section, but will prove the validity of the first order two-scale approximation in Section 5.5.

### 5.4. PERIODIC STANDARD FORM

All of the differential equations considered in Chapter 4 and in this chapter, and a great many others, can be put into the form

\[
x = \varepsilon f(x, t, \varepsilon) = \varepsilon f_1(x, t) + \varepsilon^2 f_2(x, t) + \cdots,
\] \hspace{1cm} (5.4.1)

where \( x \) is an \( N \)-dimensional vector and \( f \) is an \( N \)-vector-valued function that is periodic in \( t \) with period independent of \( \varepsilon \). Usually the period is assumed to be \( 2\pi \); this can always be achieved by changing the unit of time.
This section is entirely concerned with the procedure for achieving the form (5.4.1), which we call periodic standard form, and which is frequently referred to in the literature as standard form for the method of averaging. (There are actually several standard forms used in averaging and multiple scale methods, some of which will appear later, and we distinguish these by calling them periodic standard form, quasiperiodic standard form, angular standard form, and so forth.) The section is organized into examples; each example is a general class of equations, rather than a concrete instance. The first few examples are types of equations which have been considered before. After these, new types of problems are introduced, including some (for illustrative purposes) which cannot be put into standard form.

**Example 5.4.1. The Autonomous Oscillator**

There are several ways to put an autonomous oscillator

\[ \ddot{y} + k^2 y = \varepsilon f(y, \dot{y}) \]  

(5.4.2)

into periodic standard form. Begin by introducing phase plane variables

\[ u = y, \quad v = \dot{y}/k \]  

(5.4.3)

to obtain

\[ \dot{u} = kv, \]

\[ \dot{v} = -ku + \frac{\varepsilon}{k} f(u, kv). \]  

(5.4.4)

When \( \varepsilon = 0 \), the solutions of this system rotate clockwise at angular velocity \( k \) on circles centered at the origin. If a new set of coordinate axes is introduced which rotates clockwise at this same rate, the unperturbed solutions will appear in the new coordinate system to be at rest. Letting \( a \) and \( b \) be the new rotating coordinates, the transformation from \((u, v)\) to \((a, b)\) is given by

\[ u = a \cos kt + b \sin kt, \]

\[ v = -a \sin kt + b \cos kt. \]  

(5.4.5)

When this transformation is applied to (5.4.4), the result will be in periodic standard form. To carry out the calculation, differentiate (5.4.5) with respect to time and substitute into (5.4.4), obtaining

\[ \dot{a} \cos kt + \dot{b} \sin kt = 0, \]  

(5.4.6)

\[ -\dot{a} \sin kt + \dot{b} \cos kt = \frac{\varepsilon}{k} f(a \cos kt + b \sin kt, -ak \sin kt + bk \cos kt). \]
Now solve this system for \( \dot{a} \) and \( \dot{b} \) (by, for example, multiplying the first equation by \( \cos k t \) and the second by \( -\sin k t \) and adding) to obtain

\[
\dot{a} = -\frac{\varepsilon}{k} f(a \cos k t + b \sin k t, -ak \sin k t + bk \cos k t) \sin k t,
\]
\[
\dot{b} = +\frac{\varepsilon}{k} f(a \cos k t + b \sin k t, -ak \sin k t + bk \cos k t) \cos k t.
\]

The right hand side of this system is periodic in \( t \) with period \( 2\pi/k \), and contains a factor of \( \varepsilon \); it is therefore in periodic standard form with \( x = (a, b) \). Note that the function \( f \) is constructed from \( f \) as specified on the right hand side of (5.4.7); in particular, the lightface letter \( f \) does not denote the magnitude (norm or length) of the boldface vector \( f \), a convention common in physics. Instead we regard \( f \) as an entirely separate letter from \( f \), and define it as needed in each example. (In this book the norm of a vector \( v \) is denoted \( \|v\| \), as usual in mathematics.)

Another way to put (5.4.2) into standard form is to introduce polar coordinates into the phase plane before rotating. Setting

\[
u = r \cos \theta,
\]
\[
(5.4.8)
\]
\[
v = r \sin \theta
\]

in (5.4.4) leads to

\[
\dot{r} = \frac{\varepsilon}{k} f(r \cos \theta, kr \sin \theta) \sin \theta,
\]
\[
\dot{\theta} = \frac{\varepsilon}{kr} f(r \cos \theta, kr \sin \theta) \cos \theta.
\]

(A specific example of this computation is given below; see the steps from (5.4.17) to (5.4.18).) To rotate polar coordinates at angular velocity \( k \) in the clockwise direction it is only necessary to change the angular variable, leaving the radial variable as it is: setting

\[
\theta = \varphi - kt
\]

yields

\[
\dot{r} = \frac{\varepsilon}{k} f(r \cos(\varphi - kt), kr \sin(\varphi - kt)) \sin(\varphi - kt),
\]
\[
\dot{\varphi} = \frac{\varepsilon}{kr} f(r \cos(\varphi - kt), kr \sin(\varphi - kt)) \cos(\varphi - kt),
\]

which is in periodic standard form with \( x = (r, \varphi) \). (Do not forget, in checking that an equation is in periodic standard form, to check both that it is periodic and that it contains the factor \( \varepsilon \).)

The crucial step in arriving at the periodic standard forms (5.4.7) and (5.4.11) is to make a change of variables which "removes" the unperturbed
solution, that is, renders the unperturbed solution constant (in the new variables). In the present example, it is easy to see how to remove the unperturbed solution: Since that solution is a rotation, it is only necessary to rotate the coordinates at the same rate. When it is not obvious how to proceed, the following strategy is useful: First find the general solution of the reduced problem, which will contain arbitrary constants, and then reinterpret this solution as a change of variables (the new variables being the "constants"). In the present case, the solution of (5.4.4) with \( \varepsilon = 0 \) is (5.4.5) with \( a \) and \( b \) constant, or else

\[
\begin{align*}
    u &= r \cos(\varphi - kt), \\
    v &= r \sin(\varphi - kt),
\end{align*}
\]

(5.4.12)

with \( r \) and \( \varphi \) constant. Using (5.4.5) as a coordinate change (with \( a \) and \( b \) as the new variables) leads to (5.4.7), and using (5.4.12) leads to (5.4.11).

The idea of using the unperturbed solution to define a change of variables is closely related to the method of variation of constants (or variation of parameters). In Appendix C, it is shown that the usual method of variation of constants (for second order inhomogeneous linear differential equations) can be regarded as a change of variables. We will now show that the changes of variables used above can be regarded as a variation of constants, and that in this way it is possible to achieve both of the periodic standard forms (5.4.7) and (5.4.11) directly from the second order equation (5.4.2) without first introducing the phase plane variables \( u \) and \( v \).

We will not make much use of this method, since we regard the previous methods as more conceptually clear, but variation of parameters is commonly seen in the literature. One begins with the solution of (5.4.2) when \( \varepsilon = 0 \), which can be written in the form

\[
y = a \cos kt + b \sin kt,
\]

(5.4.13)

where \( a \) and \( b \) are constants. The derivative of (5.4.13) is, of course,

\[
\dot{y} = -ak \sin kt + bk \cos kt.
\]

(5.4.14)

The idea of variation of constants is to look for the solution of the perturbed problem (5.4.2) in the form (5.4.13) by allowing \( a \) and \( b \) to become functions of \( t \) rather than constants. Since it requires two conditions to determine the two unknown functions \( a \) and \( b \), it is permissible to impose (5.4.14) as a condition on \( a \) and \( b \) in addition to the requirement that (5.4.13) solve (5.4.2); this is equivalent to

\[
\dot{a} \cos kt + \dot{b} \sin kt = 0.
\]

(5.4.15)

Substituting (5.4.13) and (5.4.14) into (5.4.2) leads directly to (5.4.7), by algebraic steps similar to those carried out above. A similar analysis beginning with \( y = r \cos(\varphi - kt) \) leads to (5.4.11).
In order to understand the procedures leading to periodic standard form, it is important to follow the steps in some specific examples. We will therefore put the Van der Pol equation
\[ \dot{y} + \varepsilon(y^2 - 1)y + y = 0 \quad (5.4.16) \]
into periodic standard form. First write (5.4.16) as a system with \( u = y, \) \( v = \dot{y} : \)
\[ \begin{align*}
\dot{u} &= v, \\
\dot{v} &= -u + \varepsilon(1 - u^2)v.
\end{align*} \quad (5.4.17) \]
Then differentiate (5.4.8) and substitute into (5.4.17), obtaining
\[ \begin{align*}
\dot{r} \cos \theta - r \dot{\theta} \sin \theta &= r \sin \theta, \\
\dot{r} \sin \theta + r \dot{\theta} \cos \theta &= -r \cos \theta + \varepsilon(1 - r^2 \cos^2 \theta) r \sin \theta.
\end{align*} \]
Multiplying these equations by \( \cos \theta \) and \( \sin \theta, \) and adding or subtracting, gives
\[ \begin{align*}
\dot{r} &= \varepsilon(1 - r^2 \cos^2 \theta) r \sin^2 \theta, \\
\dot{\theta} &= -1 + \varepsilon(1 - r^2 \cos^2 \theta) \sin \theta \cos \theta.
\end{align*} \quad (5.4.18) \]
(Notice the division by \( r \) in obtaining the second equation.) Now (5.4.10) gives the periodic standard form (remember \( k = 1 \)):
\[ \begin{align*}
\dot{r} &= \varepsilon \left(1 - r^2 \cos^2 (\varphi - t)\right) r \sin^2 (\varphi - t), \\
\varphi &= \varepsilon \left(1 - r^2 \cos^2 (\varphi - t)\right) \sin(\varphi - t) \cos(\varphi - t).
\end{align*} \quad (5.4.19) \]
Additional practice with these ideas will be found in Exercises 5.4.1, 5.4.2, and 5.4.3.

**Example 5.4.2. Harmonic Resonance**

The periodically forced nearly linear oscillator is
\[ \ddot{y} + k^2 y = \varepsilon f(y, \dot{y}, \omega(\varepsilon)t), \quad (5.4.20) \]
where \( f \) is periodic of period \( 2\pi \) in its third argument and therefore of period \( 2\pi / \omega(\varepsilon) \) in \( t. \) Consider the harmonic resonance case (defined in Section 4.5) for equation (5.4.20), that is, the case in which the forcing frequency reduces to the free frequency when \( \varepsilon = 0: \)
\[ \omega(\varepsilon) \sim k + \varepsilon \omega_1 + \varepsilon^2 \omega_2 + \cdots. \quad (5.4.21) \]
Since the unperturbed form of (5.4.20) is the same as that of (5.4.2), it is tempting to try the same coordinate transformations (5.4.5) and (5.4.12) which lead to periodic standard form in the previous example. However, the result of applying (5.4.5) to (5.4.20) is

\[
\begin{align*}
\dot{a} &= -\frac{\varepsilon}{k} f(a \cos kt + b \sin kt, -ak \sin kt + bk \cos kt, \omega(e)t) \sin kt, \\
\dot{b} &= +\frac{\varepsilon}{k} f(a \cos kt + b \sin kt, -ak \sin kt + bk \cos kt, \omega(e)t) \cos kt,
\end{align*}
\]

which is not in periodic standard form because the right hand side is not periodic in \(t\) (unless \(\omega(e) \equiv k\)). This periodicity can be recaptured by modifying the angular velocity at which the coordinates are rotated when \(\varepsilon \neq 0\). Namely, if (5.4.5) is replaced by

\[
\begin{align*}
u &= a \cos \omega(e)t + b \sin \omega(e)t, \\
v &= -a \sin \omega(e)t + b \cos \omega(e)t,
\end{align*}
\]

which reduces to (5.4.5) when \(\varepsilon = 0\), then the result of transforming (5.4.20) is

\[
\begin{align*}
\dot{a} &= -b(\omega(e) - k) - \frac{\varepsilon}{k} f(u, kv, \omega(e)t) \sin \omega(e)t, \\
\dot{b} &= +a(\omega(e) - k) + \frac{\varepsilon}{k} f(u, kv, \omega(e)t) \cos \omega(e)t,
\end{align*}
\]

where \(u\) and \(v\) are being used not as separate variables but as "stand-ins" for the expressions given by (5.4.22), in order to shorten the equations. This system is nearly in periodic standard form, except that the period of the right hand side is \(2\pi/\omega(e)\), which depends upon \(\varepsilon\). For some purposes this is satisfactory, but in order to eliminate the \(\varepsilon\) dependence from the period it is only necessary to introduce the strained time

\[
t^+ = \omega(e)t.
\]

Then (5.4.23) becomes

\[
\begin{align*}
\frac{da}{dt^+} &= -b \frac{\omega(e) - k}{\omega(e)} \\
&\quad - \frac{\varepsilon}{k\omega(e)} f(a \cos t^+ + b \sin t^+, -ak \sin t^+ + bk \cos t^+, t^+) \sin t^+, \\
\frac{db}{dt^+} &= +a \frac{\omega(e) - k}{\omega(e)} \\
&\quad + \frac{\varepsilon}{k\omega(e)} f(a \cos t^+ + b \sin t^+, -ak \sin t^+ + bk \cos t^+, t^+) \cos t^+,
\end{align*}
\]
which can be expanded, using (5.4.21), into

$$\frac{da}{dt^+} =$$

$$-\frac{e}{k} \left\{ b\omega_1 + \frac{1}{k} f(a\cos t^+ + b\sin t^+, -ak\sin t^+ + bk\cos t^+, t^+) \sin t^+ \right\}$$

$$+ \mathcal{O}(e^2),$$

$$\frac{db}{dt^+} =$$

$$+\frac{e}{k} \left\{ a\omega_1 + \frac{1}{k} f(a\cos t^+ + b\sin t^+, -ak\sin t^+ + bk\cos t^+, t^+) \cos t^+ \right\}$$

$$+ \mathcal{O}(e^2).$$

This is in periodic standard form. It would not have been correct to expand in \( e \) prior to introducing \( t^+ \), because this would have destroyed the periodicity (since the period in \( t \) depends upon \( e \)). The essential point in achieving periodic standard form for this problem is to use (5.4.22). In Example 5.4.1, the coordinate change that leads to periodic standard form is the solution of the unperturbed equation. In the present example, (5.4.22) has built into it both the solution of the unperturbed equation (to which it reduces when \( e = 0 \)) and the perturbation frequency \( \omega(e) \). If variation of constants is used, it is necessary to modify the postulated form of the solution in the same way. (Exercise 5.4.4.)

Similarly, it is possible to obtain a periodic standard form in polar coordinates. Setting

$$u = r \cos (\varphi - \omega(e)t),$$

$$v = r \sin (\varphi - \omega(e)t)$$

results in

$$t = \frac{e}{k} f(u, kv, \omega(e)t) \sin (\varphi - \omega(e)t),$$

$$\varphi = (\omega(e) - k) + \frac{e}{kr} f(u, kv, \omega(e)t) \cos (\varphi - \omega(e)t) + \mathcal{O}(e^2),$$

where this time \( u \) and \( v \) stand for the expressions given in (5.4.27). (See Exercise 5.4.5.) Upon introducing \( t^+ \) and expanding in \( e \), this gives

$$\frac{dr}{dt^+} = \frac{e}{k^2} f(r \cos(\varphi - t^+), kr \sin(\varphi - t^+), t) \sin(\varphi - t^+) + \mathcal{O}(e^2),$$

$$\frac{d\varphi}{dt^+} = \frac{e}{k} \left\{ \omega_1 + \frac{1}{kr} f(r \cos(\varphi - t^+), kr \sin(\varphi - t^+), t) \cos(\varphi - t^+) \right\}$$

$$+ \mathcal{O}(e^2),$$

which is in periodic standard form.
As an illustration of periodic standard form for a forced oscillator in harmonic resonance, consider the weakly forced and damped Duffing equation

\[ \ddot{y} + \varepsilon \delta_1 \dot{y} + y + \varepsilon \lambda_1 y^3 = \varepsilon \gamma_1 \cos \omega(t). \quad (5.4.30) \]

Here \( k = 1 \), so \( u = y \) and \( v = \dot{y} \), and by comparison with (5.4.20),

\[ f(u, v, t^+) = \gamma_1 \cos t^+ - \lambda_1 u^3 - \delta_1 v. \]

Therefore the periodic standard form in polar coordinates (5.4.24) is in this case

\[ \frac{dr}{dt^+} = \varepsilon \left\{ \gamma_1 \cos t^+ \sin(\phi - t^+) - \lambda_1 r^3 \cos^3(\phi - t^+) \sin(\phi - t^+) - \delta_1 r \sin^2(\phi - t^+) \right\} + O(\varepsilon^2), \]

\[ \frac{d\phi}{dt^+} = \varepsilon \left\{ \omega_1 + \frac{\gamma_1}{r} \cos t^+ \cos(\phi - t^+) - \lambda_1 r^2 \cos^4(\phi - t^+) - \delta_1 \cos(\phi - t^+) \sin(\phi - t^+) \right\} + O(\varepsilon^2). \]

The reader should obtain these equations by working through the steps leading to (5.4.29) beginning with (5.4.30). See Exercise 5.4.6. Another way to obtain a periodic standard form for (5.4.20) is given in Exercise 5.4.7.

Similar periodic standard forms are possible for subharmonic, superharmonic, and supersubharmonic resonances. In nonresonant cases, when \( \omega(0)/k \) is not rational, it is not possible to achieve periodic standard form. In this case either quasiperiodic standard form or angular standard form may be used; these forms will be mentioned again in the next example, but serious consideration of them is deferred to Chapter 6.

**Example 5.4.3. Coupled Autonomous Oscillators**

Consider the following system of coupled autonomous oscillators:

\[ \ddot{y}_i + k_i^2 y_i = \varepsilon f_i(y_1, \ldots, y_N, \dot{y}_1, \ldots, \dot{y}_N) \quad (5.4.32) \]

for \( i = 1, \ldots, N \). When \( \varepsilon = 0 \), these oscillators are uncoupled, each having its own free frequency \( k_i \). Following the ideas in Example 5.4.1, phase plane variables

\[ u_i := y_i, \quad v_i := \dot{y}_i/k_i \quad (5.4.33) \]
may be introduced for each oscillator, and polar coordinates \( r_i \) and \( \theta_i \) defined by

\[
\begin{align*}
    u_i &= r_i \cos \theta_i, \\
    v_i &= r_i \sin \theta_i.
\end{align*}
\] (5.4.34)

In these coordinates the system looks like

\[
\begin{align*}
    \dot{r}_i &= \frac{\varepsilon}{k_i} f_i(r_1 \cos \theta_1, \ldots, r_N \cos \theta_N, k_1 r_1 \sin \theta_1, \ldots, k_N r_N \sin \theta_N) \sin \theta_i, \\
    \dot{\theta}_i &= -k_i + \frac{\varepsilon}{k_i r_i} f_i(r_1 \cos \theta_1, \ldots, k_N r_N \sin \theta_N) \cos \theta_i.
\end{align*}
\] (5.4.35)

This is a form which will appear again in Chapter 6, and will be called angular standard form. If we continue as in Example 5.4.1, rotating each phase plane at its own angular velocity \( k_i \) by introducing

\[
\varphi_i := \theta_i + k_i t,
\] (5.4.36)

the result is

\[
\begin{align*}
    \dot{r}_i &= \frac{\varepsilon}{k_i} f_i(r_1 \cos(\varphi_1 - k_1 t), \ldots, k_N r_N \sin(\varphi_N - k_N t)) \sin(\varphi_i - k_i t), \\
    \dot{\varphi}_i &= \frac{\varepsilon}{k_i r_i} f_i(r_1 \cos(\varphi_1 - k_1 t), \ldots, k_N r_N \sin(\varphi_N - k_N t)) \cos(\varphi_i - k_i t).
\end{align*}
\] (5.4.37)

The right hand side is not periodic unless all of the free frequencies \( k_i \) are equal, or at least are all integer multiples of a single frequency \( K \):

\[
k_i = m_i K.
\] (5.4.38)

In this case, (5.4.37) is in periodic standard form with period \( 2\pi/K \). Otherwise, the right hand side of (5.4.37) is what is called a quasiperiodic function, and (5.4.37) is said to be in quasiperiodic standard form. A quasiperiodic function is one which results from the combination of finitely many periodic components with different periods. (This is a special case of an almost periodic function, which allows for infinitely many different periods.) We will not discuss multiple scale methods for quasiperiodic or angular standard forms, but averaging methods for these cases will be introduced in Chapter 6.
Example 5.4.4. The Duffing Equation with Strong Forcing

Recall from Section 4.7 that the general form of Duffing's equation in natural (but already nondimensionalized) parameters is

$$\ddot{y} + \delta \dot{y} + y + \lambda y^3 = \gamma \cos \omega t,$$

with

$$y(0) = \alpha,$$
$$\dot{y}(0) = \beta.$$  \hspace{1cm} (5.4.39)

Until now, the perturbation families that we have considered for (5.4.39) have always involved weak forcing, that is, $y$ has been made into a function of $\varepsilon$ in such a way that $y(0) = 0$. Usually we have taken $y(\varepsilon) = \varepsilon \gamma_1$ for simplicity, although it is equally possible to take $y(\varepsilon) = \varepsilon \gamma_1 + \varepsilon^2 \gamma_2 + \cdots$, as in (4.7.2). Now we will consider what happens under the assumption of strong forcing, that is, $y(\varepsilon) = \gamma_0 + \varepsilon \gamma_1 + \cdots$ with $\gamma_0 \neq 0$; for simplicity, we will assume $y(\varepsilon) \equiv \gamma_0$. The parameters $\delta$ and $\lambda$ will be taken to be small as usual, so the differential equation to be considered is

$$\ddot{y} + \varepsilon \delta \dot{y} + y + \varepsilon \lambda y^3 = \gamma_0 \cos \omega(\varepsilon) t.$$  \hspace{1cm} (5.4.40)

The reduced problem for this equation is

$$\ddot{y} + y = \gamma_0 \cos \omega(0) t,$$  \hspace{1cm} (5.4.41)

which behaves quite differently according to whether $\omega(0) = 1$ (the harmonic resonance case) or $\omega(0) \neq 1$. In the former case, all solutions of (5.4.41) are unbounded (since they contain a term with a factor of $t$). It is not possible to attain periodic standard form in this case.

In the latter case, all solutions of (5.4.41) are bounded and are the sum of two periodic functions, one having frequency 1 and the other $\omega(0)$. Such solutions are periodic if these two periodic terms have a common period; this happens whenever $\omega(0) = p/q$ is a rational number other than 1. Under these circumstances, that is, in the nonharmonic resonance case, periodic standard form can be achieved. As an illustration, consider the third subharmonic problem

$$\omega(\varepsilon) = 3 + \varepsilon \omega_1.$$  \hspace{1cm} (5.4.42)

In this case the general solution of the reduced problem (5.4.41) can be written

$$y = a \cos t + b \sin t - \frac{\gamma_0}{8} \cos 3t,$$  \hspace{1cm} (5.4.43)

with $a$ and $b$ constant. Our previous experience suggests that something similar to (5.4.43) should be used as a variation-of-constants solution for the perturbed problem (5.4.40); more precisely, the correct form should
reduce to (5.4.43) when \( \varepsilon = 0 \) but should have period depending on \( \varepsilon \) in such a way as to be entrained to the forcing frequency \( \omega \) with entrainment index \( \ell = 3 \); in other words the period should equal \( 6\pi/\omega \). In order to find the correct form, we will use the strategy developed for the harmonic case in Exercise 5.4.7. Namely, write (5.4.40) in the form

\[
y + \frac{\omega^2}{9} y = \gamma_0 \cos \omega t + \frac{\omega^2 - 9}{9} y - \varepsilon \delta_1 \dot{y} - \varepsilon \lambda_1 y^3, \tag{5.4.44}
\]

where \( \omega \) is given by (5.4.42). All of the terms on the right hand side, except for the first, are small of order \( \varepsilon \), and the new "free frequency" on the left hand side matches the expected solution frequency. Now set \( \varepsilon = 0 \) to obtain the reduced equation in the form

\[
y + \frac{\omega^2}{9} y = \gamma_0 \cos \omega t. \tag{5.4.45}
\]

This is exactly the same as (5.4.41), considering that \( \omega = 3 \) when \( \varepsilon = 0 \); however, by retaining \( \omega \) (and not replacing it by 3) the general solution of (5.4.45) will still have the correct frequency when \( \varepsilon \neq 0 \). (If this all seems slightly mystical, remember that once the correct change of variables is found, it will justify itself.) The general solution of (5.4.45) is

\[
y = a \cos \frac{\omega}{3} t + b \sin \frac{\omega}{3} t - \frac{9\gamma_0}{8\omega^2} \cos \omega t. \tag{5.4.46}
\]

Again, this is actually the same as (5.5.43) when \( \varepsilon = 0 \), but now it has the correct form for a variation-of-constants solution when \( \varepsilon \neq 0 \). (Surprisingly, perhaps, one cannot get the right form just by modifying the frequencies in (5.5.43); there is also a change in the coefficient of the third term.)

From here the sailing is smooth, although tedious. One allows \( a \) and \( b \) to become variables, while imposing the usual condition that the first derivative of (5.4.46) be the same as if \( a \) and \( b \) were constants; that is to say, one sets

\[
\dot{a} \cos \frac{\omega}{3} t + \dot{b} \sin \frac{\omega}{3} t = 0
\]

and proceeds to differentiate (5.4.46) twice and use (5.4.44) to obtain a second equation for \( \dot{a} \) and \( \dot{b} \). The final result is that

\[
\dot{a} = -\varepsilon f(a, b, t, \varepsilon) \sin \frac{\omega(e)}{3} t + O(\varepsilon^2), \tag{5.4.47}
\]

\[
\dot{b} = +\varepsilon f(a, b, t, \varepsilon) \cos \frac{\omega(e)}{3} t + O(\varepsilon^2),
\]
where

\[ f(a, b, t, \varepsilon) = \frac{2}{3} \omega_1 \left( a \cos \frac{\omega}{3} t + b \sin \frac{\omega}{3} t - \frac{\gamma_0}{8} \cos \omega t \right) \]

\[ - \delta_1 \left( -a \sin \frac{\omega}{3} t + b \cos \frac{\omega}{3} t + \frac{\gamma_0}{8} \sin \omega t \right) \]

\[ - \lambda_1 \left( a \cos \frac{\omega}{3} t + b \sin \frac{\omega}{3} t - \frac{\gamma_0}{8} \cos \omega t \right)^3, \quad (5.4.48) \]

with (of course) \( \omega \) again given by (5.4.42). In obtaining (5.4.48) we have replaced \( \omega \) by \( 3 \) everywhere except inside the trig functions, since the error committed by doing so belongs to the \( \mathcal{O}(\varepsilon^2) \) terms in (5.4.47). Now (5.4.47) is in periodic standard form, except that one may still wish to introduce a strained time \( t^+ \) to make the period independent of \( \varepsilon \).

To understand the idea of a third subharmonic, consider what it means for the forcing to undergo three cycles while the solution goes through one. Imagine, for instance, pushing a child on a swing. One usually pushes the child with a harmonic resonance, pushing once for each period. But an eccentric swing-pusher might try a cycle of push–pull–push as the child goes forward, pull–push–pull as the child swings back. Since the push predominates during the forward phase, and the pull predominates during the return phase, the overall input of energy acts to sustain the swinging motion against air resistance or other damping forces. Therefore it seems reasonable that a third subharmonic problem such as (5.4.42) might have periodic solutions. It has been shown in Exercise 4.7.3 that there are no periodic solutions to the third subharmonic problem when the forcing is small, at least when damping is present. Using the periodic standard form presented above, together with the methods of Section 6.3 below, it can be shown that periodic solutions do exist when the forcing is strong.

Exercises 5.4

1. Put the Van der Pol equation (5.4.16) into the periodic standard form (5.4.7) in Cartesian coordinates. Do this by following through the steps leading to (5.4.7), as we did in (5.4.17)–(5.4.19) for the polar form.

2. Put the free Duffing equation \( \ddot{y} + y + \varepsilon y^3 = 0 \) into periodic standard form in both Cartesian and Polar coordinates.

3. Repeat the derivation of (5.4.19) from (5.4.16), this time using variation of constants, beginning with (5.4.13) and (5.4.14) with \( k = 1 \).

4. Derive (5.4.26) by variation of constants. Begin by looking for a solution of (5.4.20) in the form \( y = a \cos \omega(\varepsilon)t + b \sin \omega(\varepsilon)t \), where \( a \) and \( b \) are functions of \( t \). (When \( a \) and \( b \) are constants and \( \varepsilon = 0 \), this expression reduces to the solution of the unperturbed equation; it also has the
correct frequency when $\varepsilon \neq 0$, since $\omega(\varepsilon)$ has been used in place of $k$.) Remember that an additional condition must be imposed on $a$ and $b$ when they become functions of $t$.

5. Carry out the details of the derivations of (5.4.26) and (5.4.29), beginning from (5.4.20). If hints are needed beyond those in the text, follow the patterns of Example 5.4.1.

6. Put the forced Duffing equation (5.4.30) into periodic standard form in polar coordinates (5.4.31) by working through the steps in the derivation of (5.4.29). Also obtain the Cartesian form corresponding to (5.4.26).

7. An alternate way to obtain a periodic standard form for (5.4.20) is to add $\omega^2 y$ to both sides, obtaining

$$\dot{y} + \omega(\varepsilon)^2 y = \left(\omega(\varepsilon)^2 - k^2\right) y + \varepsilon f(y, \dot{y}, \omega(\varepsilon)t).$$

Introduce phase plane coordinates $U = y$, $V = \dot{y}/\omega(\varepsilon)$, then introduce rotating Cartesian coordinates $A$, $B$ by an equation similar to (5.4.22). Introduce $t^+$ as in (5.4.24), show from (5.4.21) that $\omega^2 - k^2 = 2\varepsilon k^2 + O(\varepsilon^2)$, and obtain a periodic standard form slightly different from (5.4.26). Then do the same thing in polar coordinates. (This is related to the discussion of "detuning" in Section 4.5; see (4.5.14).)

5.5. THE FIRST ORDER TWO-SCALE APPROXIMATION IN PERIODIC STANDARD FORM AND ITS JUSTIFICATION

This section has three objectives. The first order two-scale approximation will be obtained for an arbitrary system in periodic standard form. Then this approximation will be applied to an autonomous oscillator using the periodic standard form obtained in Example 5.4.1. Finally it will be shown that for any system in periodic standard form this first approximation has error $O(\varepsilon)$ on expanding intervals of order $1/\varepsilon$. The method of error estimation used is similar to that in Section 3.3, but takes advantage of special features of the periodic standard form and of the two-scale approximation. As a corollary of the general result, Theorem 5.2.1 will be proved at the end of this chapter; this theorem was stated without proof in Section 5.2, and justifies the direct application of the two-scale method to second order equations without finding the periodic standard form.

Consider the initial value problem

$$\begin{align*}
\dot{x} &= \varepsilon f(x, t, \varepsilon), \\
x(0) &= \alpha,
\end{align*} \quad (5.5.1)$$

Downloaded 02/11/16 to 128.61.77.45. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php
where \( x \) is an \( N \)-vector and \( f \) is periodic in \( t \) with period \( 2\pi \). This is to be solved by a two-time expansion

\[
x \sim x_0(t, \tau) + \varepsilon x_1(t, \tau) + \cdots,
\]

(5.5.2)

where \( \tau = \varepsilon t \). The total time derivative of (5.5.2) is (or, since this is heuristic, is supposed to be)

\[
\dot{x} \sim x_{0t} + \varepsilon \{x_{1t} + x_{0t}\} + \cdots.
\]

(5.5.3)

Substituting into (5.5.1), the equations for the first two coefficients are

\[
x_{0t} = 0, \\
x_0(0, 0) = a;
\]

(5.5.4a)

\[
x_{1t} = f(x_0, t, 0) - x_{0t},
\]

\[
x_1(0, 0) = 0.
\]

(5.5.4b)

The differential equation in (5.5.4a) tells us that \( x_0 \) is a function of \( \tau \) alone, and so we write simply \( x_0 = x_0(\tau) \) with \( x_0(0) = \alpha \). The function \( x_0(\tau) \) must be determined so that the solution \( x_1 \) of (5.5.4b) contains no \( t \)-secular terms.

The differential equation in (5.5.4b) now reads

\[
x_{1t}(t, \tau) = f(x_0(\tau), t, 0) - x_{0t}(\tau),
\]

(5.5.5)

and is to be thought of as an ordinary differential equation containing \( \tau \) as a parameter; in fact, it can be solved simply by integrating both sides with respect to \( t \). The right hand side is periodic in \( t \) and so can be expanded in a Fourier series. This is a first order equation, unlike the second order equations considered in Section 5.2, and for first order equations, secular terms do not arise from resonance between a Fourier component of the right hand side and a free frequency on the left. Instead, it is the constant term in the Fourier series that matters, since the integral of a constant is a linear function of time. In fact, all that is necessary to solve (5.5.5) is to expand \( f(x_0(\tau), t, 0) \) in a Fourier series in \( t \) having vector-valued coefficients which are functions of \( \tau \):

\[
f(x_0(\tau), t, 0) = \sum_{m=-\infty}^{\infty} a_m(\tau) e^{imt}.
\]

(5.5.6)

Then (5.5.5) becomes

\[
x_{1t}(t, \tau) = \{a_0(\tau) - x_0(\tau)\} + \sum_{m \neq 0} a_m(\tau) e^{imt}.
\]

(5.5.7)

Upon integrating with respect to \( t \), each term in the summation remains periodic and hence bounded, while the term in braces becomes multiplied...
by $t$, producing a secular term unless $x'_0(\tau) = a_0(\tau)$. Since $a_0(\tau)$, being the constant term of a Fourier series, is merely the average of the periodic function, this nonsecularity condition can be written

$$\frac{dx_0}{d\tau} = \frac{1}{2\pi} \int_0^{2\pi} f(x_0(\tau), t, 0) dt$$

or

$$\frac{dx_0}{d\tau} = \bar{f}(x_0), \quad (5.5.8)$$

where

$$\bar{f}(x) := \frac{1}{2\pi} \int_0^{2\pi} f(x, t, 0) dt. \quad (5.5.9)$$

(If $f(x, t, e)$ has a fixed period other than $2\pi$, (5.5.9) is replaced by the average over the appropriate period.) Since $x_0$ is in fact a function of only one time scale, we can just as well change (5.5.8) back into terms of $t$ and write

$$\frac{dx_0}{dt} = e\bar{f}(x_0). \quad (5.5.10)$$

This has the following interpretation: To find the differential equation (5.5.10) for the first order two time approximation for (5.5.1), it is only necessary to average the right hand side of (5.5.1) over one period in $t$, with $x$ held constant. This is just the recipe which will be given for the first approximation by the method of averaging, to be considered in the next chapter.

For the remainder of this section, we will work with this recipe for the first approximation. All of the heuristic reasoning leading from (5.5.1) to (5.5.10) may be set aside, including the notation $x_0$ and the existence of the series (5.5.2); we will merely consider the idea that a good approximation to the solution of (5.5.1) might happen to be given by the solution of

$$\dot{z} = e\bar{f}(z), \quad z(0) = \alpha. \quad (5.5.11)$$

The only equations appearing above that will be used in the remainder of this section are (5.5.1), (5.5.9), and (5.5.11). The reason for the letter $z$ is so that solutions of (5.5.1) and (5.5.11) can be compared without confusion. When one is only interested in the approximate solutions, it is more convenient to write (5.5.11) as $\dot{x} \equiv e\bar{f}(x)$, without introducing another variable; but then there is no notational distinction between the exact and approximate solutions.

Before examining the error in the approximation given by (5.5.11), we stop to consider what these equations look like in the familiar case of an
autonomous oscillator. The standard form for an autonomous oscillator is given by either (5.4.7) in Cartesian coordinates or by (5.4.11) in polar coordinates. We repeat these here as

\[
\begin{align*}
\dot{a} &= -\frac{\epsilon}{k} f(a \cos kt + b \sin kt, -ak \sin kt + bk \cos kt, \varepsilon) \sin kt, \\
\dot{b} &= +\frac{\epsilon}{k} f(a \cos kt + b \sin kt, -ak \sin kt + bk \cos kt, \varepsilon) \cos kt
\end{align*}
\]  

(5.5.12)

and

\[
\begin{align*}
\dot{r} &= \frac{\epsilon}{k} f(r \cos(\varphi - kt), kr \sin(\varphi - kt), \varepsilon) \sin(\varphi - kt), \\
\dot{\varphi} &= \frac{\epsilon}{kr} f(r \cos(\varphi - kt), kr \sin(\varphi - kt), \varepsilon) \cos(\varphi - kt).
\end{align*}
\]  

(5.5.13)

Averaging (5.5.12) over one period and changing the names of the variables from \(x = (a, b)\) to \(z = (A, B)\) gives

\[
\begin{align*}
\frac{dA}{dt} &= -\frac{\epsilon}{2\pi} \int_0^{2\pi/k} f(A \cos kt + B \sin kt, -Ak \sin kt + Bk \cos kt) \sin kt \, dt, \\
\frac{dB}{dt} &= +\frac{\epsilon}{2\pi} \int_0^{2\pi/k} f(A \cos kt + B \sin kt, -Ak \sin kt + Bk \cos kt) \cos kt \, dt.
\end{align*}
\]  

(5.5.14)

Recall that the original second order autonomous oscillator (5.4.2) was changed into periodic standard form by using (5.4.3) and (5.4.5). Therefore the relationship between the solution \(y\) of the second order equation and the solution \((a, b)\) of the first order system (5.5.12) is

\[
y = a \cos kt + b \sin kt.
\]  

(5.5.15)

According to our present ideas, the solution \((A, B)\) of (5.5.14) should give the first approximation to \((a, b)\), and therefore the first approximation to the solution of the autonomous oscillator in the original variables should be given by

\[
y \approx A \cos kt + B \sin kt.
\]  

(5.5.16)

When we applied the method of multiple scales directly to the autonomous oscillator, in Section 5.2, the first approximation \(y_0\) was found to have the form (5.5.16) with \(A\) and \(B\) being functions of \(\tau\) satisfying (5.2.10). But (5.2.10) is exactly the same as (5.5.14) when the latter is expressed in terms
of \( \tau \). Therefore our current procedure gives the same results as the method of Section 5.2.

A similar analysis can be made of the polar form. Namely, if (5.5.13) is averaged over \( t \), the variable names are changed from \( x = (r, \varphi) \) to \( z = (\rho, \psi) \), and the result is expressed in terms of \( \tau \), then the equations obtained are exactly (5.2.12).

Next we turn to the question of error estimation for the first order two-scale approximation. The theorem stated below will be proved again in Chapter 6 as Theorem 6.2.2, by quite different methods.

The solution of (5.5.1) will be denoted \( x(t, \alpha, \varepsilon) \). A similar expression could be used for the solution of (5.5.11), but in Chapter 6 it is more convenient to use a slightly different notation, and we will introduce it here to avoid conflict. Notice that if \( \tau := et \) is introduced in (5.5.11) the result is \( dz/d\tau = \tilde{f}(z) \); this is the same as (5.5.8) and does not contain the perturbation parameter \( \varepsilon \). The solution of this equation will be written \( z(\tau, \alpha) \); of course, it does not depend upon \( \varepsilon \). Then the solution of (5.5.11) is \( z(et, \alpha) \) (where \( z \) is the same function). Writing \( z(et, \alpha) \) for the solution of (5.5.11) reveals the manner of dependence upon \( \varepsilon \) more precisely than writing \( x(t, \alpha, \varepsilon) \) would do. Of course, \( x(t, \alpha, \varepsilon) \) cannot be written as \( x(et, \alpha) \) since it depends upon \( \varepsilon \) in a more complicated way than through the combination \( et \).

**Theorem 5.5.1.** Let \( f \) be smooth. There exist positive constants \( \varepsilon_0, T, \) and \( c \) such that the solutions of (5.5.1) and (5.5.11) satisfy

\[
\| x(t, \alpha, \varepsilon) - z(et, \alpha) \| < ce \quad \text{for} \quad 0 \leq t \leq \frac{T}{\varepsilon} \quad \text{and} \quad 0 \leq \varepsilon \leq \varepsilon_0. \quad (5.5.17)
\]

**Proof.** Choose a closed ball \( K \) of some positive radius \( \delta \) centered at \( \alpha \), and let \( \varepsilon_0 > 0 \). Let \( M \) be an upper bound for both \( \| \tilde{f}(z) \| \) and \( \| f(z, t, \varepsilon) \| \) for \( z \) in \( K \) and \( 0 \leq \varepsilon \leq \varepsilon_0 \); such a bound exists because \( f \) is periodic in \( t \). Then both solutions \( x \) and \( z \) begin at the center of \( K \) and reach at most a distance \( eMt \) from the center by time \( t \), as long as they remain in \( K \) (assuming \( 0 \leq \varepsilon \leq \varepsilon_0 \)). But on the other hand, they cannot leave \( K \) until they have travelled a distance \( \delta \). It follows that they must remain in \( K \) at least until time \( T/\varepsilon \), where \( T = \delta/M \). In the remainder of the proof, various bounds and Lipschitz constants will be defined using the set \( K \); these are now automatically known to hold for the interval \( 0 \leq t \leq T/\varepsilon \).

Since \( \alpha \) is fixed throughout the proof, it will be omitted from the expressions \( x(t, \alpha, \varepsilon) \) and \( z(et, \alpha) \) for convenience. Write \( x(t, \varepsilon) = z(et) + R(t, \varepsilon) \), so that \( R \) is the remainder (or error) to be estimated. Then (5.5.1) implies \( \dot{z} + \tilde{R} = ef(z + R, t, \varepsilon) \). Combining this with (5.5.11) gives the following equation satisfied by the remainder:

\[
R = \varepsilon \left[ f(z(et) + R, t, \varepsilon) - \tilde{f}(z(et)) \right], \quad (5.5.18)
\]

\[
R(0, \varepsilon) = 0.
\]
Integrating from 0 to \( t \), using the initial condition, yields
\[
R(t, \varepsilon) = \varepsilon \int_0^t \left[ f(z(\varepsilon s) + R(s, \varepsilon), s, \varepsilon) - \bar{f}(z(\varepsilon s)) \right] ds. \tag{5.5.19}
\]

Notice that the dummy variable of integration \( s \) (not \( t \)) now appears in the argument of \( z \) and \( R \) inside the integrals; from here on these arguments will be omitted for brevity. Adding and subtracting \( f(z, s, \varepsilon) \) and using the triangle inequality and the fact that the norm of an integral is less than or equal to the integral of the norm, we have
\[
\left\| R(t, \varepsilon) \right\| \leq \varepsilon \int_0^t \left\| f(z + R, s, \varepsilon) - f(z, s, \varepsilon) \right\| ds \tag{5.5.20}
\]
\[
+ \varepsilon \left\| \int_0^t \left[ f(z, s, \varepsilon) - f(z, s, 0) \right] ds \right\| + \varepsilon \left\| \int_0^t \left[ f(z, s, 0) - \bar{f}(z) \right] ds \right\|.
\]

Notice that only in the first term has the norm been taken inside the integral. Let \( L \) be a Lipschitz constant for \( f \) with respect to its first (vector) argument, valid in \( K \). Then the first term on the right in (5.5.20) is bounded by the integral of \( \varepsilon L \| R \| \). The second term is bounded by a constant times \( \varepsilon^2 t \) and the third term is estimated in Lemma 5.5.2 below. The key result is that there is a constant \( c_0 \) such that the last two terms in (5.5.20) are bounded by \( c_0 \varepsilon \) as long as \( 0 \leq t \leq T/\varepsilon \). Assuming this result for the moment, (5.5.20) implies
\[
\left\| R(t, \varepsilon) \right\| \leq \varepsilon L \int_0^t \| R(s, \varepsilon) \| ds + c_0 \varepsilon. \tag{5.5.21}
\]

From here the argument follows the familiar Gronwall pattern. Let
\[
S(t, \varepsilon) = \int_0^t \| R(s, \varepsilon) \| ds; \tag{5.5.22}
\]
then (5.5.21) can be written
\[
\frac{dS}{dt} \leq \varepsilon LS + c_0 \varepsilon, \tag{5.5.23}
\]
\[
S(0) = 0.
\]
Solving this differential inequality in the usual way (see, for instance, the solution of (4.2.13)) gives
\[
S(t, \varepsilon) \leq \frac{c_0}{L} (e^{\varepsilon L t} - 1) \leq \frac{c_0}{L} (e^{LT} - 1), \tag{5.5.24}
\]
where the last expression uses \( t \leq T/e \); recall that the entire argument only holds up to this time. Finally, putting (5.5.24) back into (5.5.23) gives

\[
\|R(t, \varepsilon)\| \leq \varepsilon c_0 e^{LT}.
\]  
(5.5.25)

This is the statement (5.5.17) of the theorem. \( \blacksquare \)

The missing step in the above proof is the estimate for the last term in (5.5.20). The crucial fact about the integrand \( f(z, s, 0) - f(z) \) is that for fixed \( z \), it has zero average (with respect to \( s \)). Therefore, if \( z \) were fixed during the integration, the integral would be bounded for all time. Of course, \( z \) is not fixed during the integration, but is a slowly varying function of \( s \); specifically, it is \( z(es) \). The crucial point is that because this is only slowly varying, the integral is still bounded, not for all time, but on the expanding interval \( 0 < t < T/e \). Namely:

**Lemma 5.5.2.** Let \( \varphi(z, t, \varepsilon) \) be a smooth function, periodic in \( t \) with period \( 2\pi \) for each fixed \( z \) and \( \varepsilon \), and having zero mean with respect to \( t \). Let \( z(et) \) be the solution of (5.5.11), which remains in \( K \) for \( 0 < t < T/e \). Then there exists a constant \( c_1 \) such that

\[
\left\| \int_0^t \varphi(z(es), s, \varepsilon) \, ds \right\| \leq c_1 \quad \text{for} \quad 0 \leq t \leq \frac{T}{\varepsilon} \quad \text{and} \quad 0 \leq \varepsilon \leq \varepsilon_0. 
\]  
(5.26)

**Proof.** The proof given here requires that \( \varphi \) have at least two continuous derivatives; it uses Fourier series estimates and integration by parts. There is another proof which requires only continuity (in fact even weaker hypotheses defined in terms of measure theory will suffice), but the computations used in that proof are less natural. (See Notes and References.)

Since \( \varphi \) is periodic with zero mean and has two continuous derivatives, it can be written as a uniformly convergent Fourier series

\[
\varphi(z, t, \varepsilon) = \sum_{n \neq 0} a_n(z, \varepsilon) e^{int} 
\]  
(5.5.27)

without a constant term, where the \( a_n \) are vector-valued coefficients satisfying \( \|a_n(z, t, \varepsilon)\| \leq k/n^2 \) for some constant \( k \) and for all \( z \) in \( K \), all \( t \), and for \( 0 \leq \varepsilon \leq \varepsilon_0 \). Now substitute \( z(et) \) into (5.5.27). To shorten the notation, we will omit \( \varepsilon \) as an argument from \( \varphi \) and \( a_n \) from here on. Since the convergence is uniform, the resulting series can be integrated termwise by parts to obtain

\[
\int_0^t \varphi(z(es), s) \, ds = \sum_{n \neq 0} \frac{a_n(z(et)) e^{int} - a_n(z(0))}{in} 
\]

\[
- \int_0^t \frac{e^{ins}}{in} a_n'(z(es)) \frac{d}{ds} z(es) \, ds. 
\]  
(5.5.28)
Here $a_n'$ is the matrix of partial derivatives of $a_n$ with respect to $z$; since these are the Fourier coefficients of $\partial \varphi / \partial z$, which has at least one continuous derivative, they satisfy $\|a_n'\| \leq \ell / n$ for $z$ in $K$, for some constant $\ell$, using a matrix norm. Since $dz(es)/ds = e\tilde{f}(z)$, its norm is bounded by $eM$ (where $M$ is as in the first paragraph of the proof of Theorem 5.5.1). Putting all of this together gives the conclusion:

$$\left\| \int_0^t \varphi(z(es, s)) \, ds \right\| \leq \sum_{n \neq 0} \frac{\|a_n(z(\ell t))\| + \|a_n(z(0))\|}{n}$$

$$+ \sum_{n \neq 0} \int_0^t \frac{1}{n} \|a_n(z(es))\| eM \, ds$$

$$\leq \sum_{n \neq 0} \frac{2k}{n^3} + \sum_{n \neq 0} \frac{\ell eM^2}{n^2} \frac{t}{\ell}$$

$$\leq 2k \left( \sum_{n \neq 0} \frac{1}{n^3} \right) + \ell M^2 T \left( \sum_{n \neq 0} \frac{1}{n^2} \right) =: c_0.$$  

(5.5.29)

In the next to the last step we have used $t \leq T/e$. 

Theorem 5.5.1 justifies the first order two-scale method for systems in periodic standard form. But it is easy to apply this theorem to obtain a justification also for the first order two-scale method applied directly to second order scalar equations without first changing them into periodic standard form. We will now illustrate how this is done by using Theorem 5.5.1 to prove Theorem 5.2.1, justifying the first order two-scale method for autonomous oscillators.

**Proof of Theorem 5.2.1.** It was shown above (in discussing equation (5.5.16)) that the approximate solution $y_0$ for an autonomous oscillator constructed in Section 5.2 by applying the two-scale method directly is the same as the one constructed in this section by first putting the oscillator into periodic standard form. For the periodic standard form, Theorem 5.5.1 implies that $a = A + O(e)$ and $b = B + O(e)$ for $0 \leq t \leq T/e$. It is only necessary to carry this estimate over to the original variable $y$. We have $y = a \cos kt + b \sin kt = (A + O(e)) \cos kt + (B + O(e)) \sin kt = A \cos kt + B \sin kt + O(e)$, for $0 \leq t \leq T/e$; the next-to-the-last equality in this sequence holds because $\cos kt$ and $\sin kt$ are bounded for all time. This proves Theorem 5.2.1. A similar estimate can be proved for the velocity; see Exercise 5.5.1.

**Exercises 5.5**

1. Let $y$ be the exact solution of the autonomous oscillator and let $y_0$ be the two-scale solution constructed in Section 5.2. Prove from Theorem
5.5.1 that $\dot{y} = y_0 + O(\varepsilon)$ on an expanding interval. Hint: $\dot{y} = v$; see (5.4.7). Remark: Taking the total derivative of $y \approx y_0$ with respect to $t$ suggests (heuristically) that $\dot{y} \approx y_0 + \varepsilon y_0 t$. The second term is formally of order $O(\varepsilon)$, the same as the expected order of the error, so it is reasonable to drop it. This exercise shows that this heuristic conclusion is in fact correct.

2. Find the first approximation for the free Duffing equation $\ddot{y} + y + \varepsilon y^3 = 0$, $y(0) = \alpha$, $\dot{y}(0) = 0$ using periodic standard form. Work the problem in both polar and Cartesian form, and express the solutions in the original variable $y$. See Exercise 5.4.2, and compare your result with Exercise 5.2.3.

5.6* PARTIAL DIFFERENTIAL EQUATIONS

The partial differential equation

$$u_{tt} - u_{xx} + u = 0$$

(5.6.1)

is called a linear dispersive wave equation and has as one of its particular solutions (this can be checked easily) the function

$$u = \sin(kx + \omega t),$$

(5.6.2)

where $k$ is arbitrary and

$$\omega := \sqrt{1 + k^2}.$$  

(5.6.3)

Throughout this section $\omega$ will be used as in (5.6.3); that is, it is not an independent constant but merely an abbreviation. The solution (5.6.2) satisfies the following initial conditions at $t = 0$:

$$u(x, 0) = \sin kx,$$
$$u_t(x, 0) = \omega \cos kx.$$  

(5.6.4)

As an example of the multiple scale method in partial differential equations, we will determine the first order two-scale approximation to the solution of the following perturbation of (5.6.1):

$$u_{tt} - u_{xx} + u + \varepsilon u^3 = 0.$$  

(5.6.5)

The solution to be found will be the one having the same initial conditions (5.6.4) as the solution (5.6.2) of (5.6.1).

At the very beginning, there is a pesky notational difficulty that does not arise in ordinary differential equations; it is a minor nuisance, but is
not serious. Based on our experience with ordinary differential equations in Section 5.2, we write

\[ \tau = \varepsilon t \]  

(5.6.6)

and represent the solution in the two-scale form

\[ u(x, t, \varepsilon) = u_0(x, t, \tau) + \varepsilon u_1(x, t, \tau) + \cdots. \]  

(5.6.7)

At this point we want to apply the usual differentiation rules

\[
\frac{d}{dt} = \frac{\partial}{\partial t} + \varepsilon \frac{\partial}{\partial \tau},
\]

(5.6.8)

\[
\frac{d^2}{dt^2} = \frac{\partial^2}{\partial t^2} + 2\varepsilon \frac{\partial^2}{\partial \tau \partial t} + \varepsilon^2 \frac{\partial^2}{\partial \tau^2},
\]

which were stated before in (5.2.4). The difficulty is that in order to find \( u_{tt} \), the left hand side of (5.6.7) must be partially differentiated twice with respect to \( t \), whereas the left hand sides of (5.6.8) involve ordinary derivatives. If we change the ordinary derivatives in (5.6.8) to partial derivatives, the result is nonsense, because the partial derivatives on the right hand side have a different meaning from those on the left. To resolve the difficulty it is only necessary to think of what (5.6.8) are really intended to mean. In the original application, there was no \( x \) in (5.6.7), and what was needed was to express a derivative with respect to the single time variable \( t \) in terms of derivatives with respect to two time variables \( t \) and \( \tau \). This is done by equations (5.6.8), with the ordinary derivatives on the left being derivatives with respect to \( t \) alone (when there is no \( \tau \) in the equation), and the partial derivatives with respect to \( t \) on the right being derivatives with respect to \( t \) when \( \tau \) is present and held constant. It is only necessary to preserve this interpretation even when the variable \( x \) is present. That is, equations (5.6.8) are still correct if we temporarily regard \( d/dt \) as a partial derivative with respect to \( t \) with only \( x \) held constant, and \( \partial/\partial t \) as a partial derivative with respect to \( t \) with both \( x \) and \( \tau \) held constant. With no further ado, then, we apply (5.6.8) to (5.6.7) and obtain the following sequence of problems, where the subscript \( t \) (when attached to one of the \( u_t \) rather than to \( u \) itself) means the derivative holding \( x \) and \( \tau \) constant:

\[
\begin{align*}
  u_{0tt} - u_{0xx} + u_0 &= 0, \\
  u_0(x, 0, 0) &= \sin kx, \\
  u_{0t}(x, 0, 0) &= \omega \cos kx;
\end{align*}
\]  

(5.6.9a)
\[ u_{1tt} - u_{1xx} + u_1 = -u_0^3 - 2u_{0tt}, \]
\[ u_1(x, 0, 0) = 0, \quad (5.6.9b) \]
\[ u_{1t}(x, 0, 0) = -u_{0t}(x, 0, 0), \]

Now the first of these problems, (5.6.9a), is the same as (5.6.1) and (5.6.4) except for the possibility of dependence on \( \tau \), so its solution is the same as (5.6.2) except for amplitude and phase factors that may depend on \( \tau \):

\[ u_0(x, t, \tau) = a(\tau) \sin(kx + \omega t + \phi(\tau)). \quad (5.6.10) \]

The initial conditions imply that

\[ a(0) = 1, \]
\[ \phi(0) = 0. \quad (5.6.11) \]

Substituting this (partial) solution for \( u_0 \) into (5.6.9b) and replacing \( \sin^3 \) by its expansion from Table E.1 gives

\[ u_{1tt} - u_{1xx} + u_1 = \left( -\frac{3}{4} a^3 + 2a \omega \phi' \right) \sin(kx + \omega t + \phi) \]
\[ - 2a' \omega \cos(kx + \omega t + \phi) \]
\[ + \frac{1}{4} a^3 \sin 3(kx + \omega t + \phi). \quad (5.6.12) \]

The first harmonics on the right hand side of (5.6.12) have the same form as the solutions of the associated homogeneous equation. Therefore they are in resonance with the free solutions and lead to \( t \)-secular terms. (Partial differential equations of this type behave similarly to ordinary differential equations in this respect.) Therefore the method of multiple scales calls for completing the determination of \( a(\tau) \) and \( \phi(\tau) \) by setting the coefficients of these first harmonics equal to zero, leading to

\[ a' = 0, \]
\[ \phi' = \frac{3a^2}{8\omega}. \quad (5.6.13) \]

Solving (5.6.13) with initial conditions (5.6.11) gives

\[ a(\tau) = 1, \]
\[ \phi(\tau) = \frac{3a^2}{8\omega} \tau, \]
so that the complete first order two-scale approximation for the desired solution $u(x, t, \varepsilon)$ is

$$u_0(x, t, \varepsilon t) = \sin \left( kx + \omega t + \frac{3a^2}{8\omega} \varepsilon t \right).$$

### 5.7. Notes and References

The method of multiple scales is emphasized in the books by Nayfeh and the one by Kevorkian and Cole which are listed in the Preface. Kevorkian and Cole prefer the version using slow time $\tau$ and strained fast time $t^+$, whereas Nayfeh gives more attention to the version using $T_0, T_1, \ldots, T_M$.

As explained in Sections 5.1 and 5.3, these two versions of multiple scales are only needed when attempting to gain validity for intervals longer than $1/\varepsilon$. The simpler version, using $t$ and $\tau$, is given in Chapters 3 and 4 of the book by Smith, along with error estimation (which is absent in the other books).

For the relationship between multiple scales and averaging, the essential references are


and


The proof of validity of the first order two-scale method for systems in periodic standard form, presented in Section 5.5, is due to J. G. Besjes, who presented it as a proof of the validity of first order averaging. (Actually his proof is the one mentioned in the first line of our proof and does not assume two continuous derivatives.) It is to be found in


For additional comments and references concerning this proof, see Section 6.7.

The example (5.3.15), along with some others for which the multiple scale method fails on time scales longer than $1/\varepsilon$, is given in

A different situation causing such a failure is discussed in Section 6.5 below.

A lengthy discussion of the nonlinear wave problem in Section 5.6, along with many variations, is given in Section 4.4.1 of the book by Kevorkian and Cole (listed in the Preface).

Since the publication of this book, some progress has been made on the questions addressed in section 5.3. There is a computable obstruction to introducing a third time scale $\varepsilon^2 t$; if the obstruction is not zero, the time scale cannot be introduced, while if it is zero, the time scale can be added and the solution is valid for time $O(1/\varepsilon^2)$. For details see James Murdock and Lih-Chyun Wang, Validity of the multiple scale method for very long intervals, *ZAMP (Journal of Applied Mathematics and Physics)* 47 (1996), 760-789.