MATHEMATICS OF COMPUTATION, VOLUME 32, NUMBER 142
APRIL 1978, PAGES 453–496

The Method of Envelopes

By W. L. Miranker and M. van Veldhuizen*

Abstract. The differential equation

$$\frac{dx}{dt} = \frac{A}{\epsilon}x + g(t, x)$$

where $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ and $\epsilon > 0$ is a small parameter is a model for the stiff highly oscillatory problem. In this paper we discuss a new method for obtaining numerical approximations to the solution of the initial value problem for this differential equation. As $\epsilon \to 0$, the asymptotic theory for this initial value problem yields an approximation to the solution which develops on two time scales, a fast time t and a slow time $\tau = t/\epsilon$. We redevelop this asymptotic theory in such a form that the approximation consists of a series of simple functions of τ, called carriers. (This series may be thought of as a Fourier series.) The coefficients of the terms of this series are functions of t. They are called envelopes and they modulate the carriers. Our computational method consists of determining numerical approximations to a finite collection of these envelopes. One of the principal merits of our method is its accuracy for the nonlinear problem.

Introduction. The differential equation

$$\frac{dx}{dt} = \frac{A}{\epsilon}x + g(t, x),$$

where $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ and $\epsilon > 0$ is a small parameter is a model for the stiff highly oscillatory problem. In this paper we discuss a new method for obtaining numerical approximations to the solution of the initial value problem for this differential equation.

As $\epsilon \longrightarrow 0$, the asymptotic theory for this initial value problem yields an approximation to the solution which develops on two time scales, a fast time t and a slow time $\tau = t/\epsilon$ (cf. Hoppensteadt-Miranker [11]). We redevelop this asymptotic theory in such a form that the approximation consists of a series of simple functions of τ, called carriers. (This series may be thought of as a Fourier series.) The coefficients of the terms of this series are functions of t. They are called envelopes and they modulate the carriers. Our computational method consists of determining numerical approximations to a finite collection of these envelopes.

This computational problem has been addressed by many others with a resulting variety of algorithms. A first class of algorithms consists of multistep methods which are exact for algebraic polynomials and/or trigonometric polynomials up to a certain degree (the degree may depend on the type of polynomial). We mention the work of

Received February 11, 1977.
AMS (MOS) subject classifications (1970). Primary 65L05, 34E99.
*On leave from Free University of Amsterdam.

Gautschi [6], of Bettis and Stiefel, cf. [21] and the references therein, and the work of Snider-Fleming [20]. In the latter paper the modified multistep methods are also made more efficient by aliasing high frequencies by lower ones. Other methods have been introduced by Amdursky-Ziv [1], Hoppensteadt-Miranker [10] and Miranker-Wahba [15]. Amdursky and Ziv deflate the system, removing its highest frequency after its determination by computational means. Hoppensteadt and Miranker use the asymptotic theory, as they derive it in terms of infinite averages to develop numerical evaluation schemes. Miranker and Wahba compute running averages of the solution. (The envelope of the carrier, which itself is a constant, is, of course, the average.) Thus our method is related to and generalizes several of the existing attacks on this computational problem. One of the principal merits of our method is its accuracy for the nonlinear problem.

In Section 1 we formulate the problem to be treated. We derive the asymptotic theory in the carrier-envelope form and we introduce and develop the notion of a smooth solution of a stiff differential equation. In Section 2 we show how the smooth solution of a differential equation may be approximated by a polynomial solution of an associated differential equation. We also introduce and give stability properties of backward differentiation formulae. These formulae are to be used later to determine the envelopes. Section 3 shows how complex valued carriers may be replaced by real ones and shows as well in what sense the smooth solution concept commutes with carrier changes. In Section 4 we describe two algorithms for generating envelopes. In Section 5 we demonstrate the existence and uniqueness of the solution produced by the computational scheme. We also obtain local and global error estimates as they depend on ϵ, the mesh width h, the aliasing error and on other related parameters. A superconvergence result as well as stability with respect to perturbations in initial data are also obtained. All of these results in Section 5 are obtained without restriction on the ratio ϵ/h. By restricting this ratio we are able to obtain an additional stability result, namely, stability with respect to roundoff errors. Finally in Section 6 we give results of computations with our methods on a sample nonlinear problem. We point out how the computations verify most of the theoretical behavior predicted in Section 5.

1. Exploration of the Problem. Consider the ordinary differential equation

$$\frac{dx}{dt} = \frac{A}{\epsilon}x + g(t, x). \tag{1.1}$$

Here $\epsilon > 0$ is a *small* parameter, g is a smooth mapping from a domain in $\mathbf{R} \times \mathbf{R}^2$ into $\mathbf{R}^2$, x is an $\mathbf{R}^2$-valued map and A is the skew-symmetric matrix

$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}. \tag{1.2}$$

A solution of (1.1) is sought for $t \in [0, T]$, subject to the initial condition

$$x(0) = \xi. \tag{1.3}$$

The existence of a unique solution $\tilde{x}$ is assumed, for $t \in [0, T]$. This solution $\tilde{x}$ is in general highly oscillatory, with approximately $2\pi/\epsilon$ oscillations per unit interval. Since $\epsilon > 0$ is a small parameter, it is quite natural to describe the solution $\tilde{x}$ by means of an asymptotic series ($\epsilon \longrightarrow 0$). This series is not so easily obtained; e.g. a zeroth order ($\epsilon \longrightarrow 0$) approximation is *not* obtained by neglecting $g(t, x)$ in (1.1).

Since the basis of the numerical algorithm to be proposed is closely related to the asymptotic series for the solution $\tilde{x}$, we give an interpretation of the asymptotic results, suitable for our purpose (i.e. the numerical algorithm). Our reference source for the asymptotic material is Hoppensteadt and Miranker [11].

Let $\Phi(t) = \exp(At)$. Hence

$$\frac{d}{dt}\Phi(t/\epsilon) - \frac{A}{\epsilon}\Phi(t/\epsilon) = 0. \tag{1.4}$$

Put $x(t) = \Phi(t/\epsilon)\mathbf{u}(t)$. By this change of variable we consider the solution relative to $\Phi(t/\epsilon)$. Using (1.4) and $\Phi(0) = I$, we find the following initial value problem for $\mathbf{u}$.

$$\frac{d}{dt}\mathbf{u} = \Phi^{-1}(t/\epsilon)g(t, \Phi(t/\epsilon)\mathbf{u}), \qquad \mathbf{u}(0) = \xi. \tag{1.5}$$

Now we introduce *two time scales* t and $\tau = t/\epsilon$. As long as the identification $\tau = t/\epsilon$ is maintained, no new features emerge; we simply introduce a new symbol and utilize it where convenient. The basically new element appears by *uncoupling* t and τ. I.e. we consider t and τ as completely unrelated variables. Then, by some abuse of notation, we are obliged to write

$$\frac{d}{dt} = \frac{\partial}{\partial t} + \frac{1}{\epsilon}\frac{\partial}{\partial \tau}, \tag{1.6}$$

where the t on the left-hand side is the old t, and where the t on the right-hand side is the new one, which replaces the old one everywhere where the old one does not appear as t/ϵ. The variable t/ϵ is replaced everywhere by τ.

For notational convenience we introduce the map G, which assigns to a function $\mathbf{u} = \mathbf{u}(t, \tau)$ a function $G(\mathbf{u})$, given for all t, τ by

$$G(\mathbf{u})(t, \tau) = \Phi^{-1}(\tau)g(t, \Phi(\tau)\mathbf{u}(t, \tau)). \tag{1.7}$$

Using this map G, we now have instead of (1.5)

$$\frac{\partial}{\partial t}\mathbf{u} + \frac{1}{\epsilon}\frac{\partial}{\partial \tau}\mathbf{u} = G(\mathbf{u}), \tag{1.8}$$

with unknown (vector-valued) function $\mathbf{u} = \mathbf{u}(t, \tau)$.

Consider (1.8) as a hyperbolic equation on the rectangle $[0, T] \times [0, T/\epsilon]$ in t, τ-space. We know that $\mathbf{u}(0, 0) = \xi$. This condition is clearly insufficient to guarantee unique solvability of (1.8). As we shall soon see, the asymptotic process supplies additional constraints in such a way that the uniqueness problem is circumvented. The manner in which this happens is the basis for the numerical algorithm to be proposed.

In order to find additional constraints, consider (1.8) in the form

$$\frac{\partial}{\partial \tau}\mathbf{u} = \epsilon G(\mathbf{u}) - \epsilon \frac{\partial}{\partial t}\mathbf{u}, \tag{1.9}$$

where t is a parameter. Then (1.9) is just an ordinary differential equation, independent variable τ, dependent variable $\mathbf{u}(t, \cdot)$. Thus, the ordinary differential equation is nonlinear, with the nonlinear term $\epsilon G(\mathbf{u})$. For all $u \in \mathbf{R}^2$ for which $G(u)(t, \tau)$ is well defined we have

$$G(u)(t, \tau + 2\pi) = G(u)(t, \tau), \quad \forall \tau,$$

for all values of the 'parameter' t. Thus, the ordinary differential equation (1.9), with t considered as a parameter, is a *forced oscillation* in the sense of Urabe [22]. The period of the nonlinear forcing term $G(\mathbf{u})$ is 2π. Therefore, we require $\mathbf{u}$ to be 2π-*periodic* in the τ-direction. This choice has to be justified later by the proof of the asymptotic character of the series obtained.

Because of the periodicity in τ, we introduce the Fourier series for $\mathbf{u}$,

$$\mathbf{u}(t, \tau) = \sum_{p \in \mathbf{Z}} e^{ip\tau} u_p(t). \tag{1.10}$$

u_p is defined by

$$u_p(t) = \frac{1}{2\pi}\int_0^{2\pi} e^{-ip\sigma}\mathbf{u}(t, \sigma)\, d\sigma, \quad \forall p. \tag{1.11}$$

Formula (1.10) expresses the idea of "separation of variables." The basis functions in the τ-direction (the $\{e^{ip\tau}\}$ in (1.10)) will be called *carriers* and the coefficients $\{u_p\}$ will be called *envelopes*. The $\{e^{ip\tau}\}$ are not the only carriers possible. E.g. the $\{\cos p\tau, \sin p\tau\}$ provide different carriers, and thus, different envelopes. We also need the Fourier coefficients of the right-hand side of (1.9). In particular, put

$$g_p(\mathbf{u})(t) = \frac{1}{2\pi}\int_0^{2\pi} e^{-ip\sigma} G(\mathbf{u})(t, \sigma)\, d\sigma, \quad \forall p \in \mathbf{Z}. \tag{1.12}$$

Then, under appropriate smoothness assumptions on $\mathbf{u}$ and g we may write

$$G(\mathbf{u})(t, \tau) - \frac{\partial}{\partial t}\mathbf{u}(t, \tau) = \sum_{p \in \mathbf{Z}} e^{ip\tau}\left[g_p(\mathbf{u})(t) - \frac{\partial}{\partial t}u_p(t)\right]. \tag{1.13}$$

Thus, from (1.9) we obtain an equation for each of the envelopes (i.e. Fourier coefficients) separately. For $p \neq 0$ and all $t \in [0, T]$ we obtain

$$ipu_p = \epsilon g_p(\mathbf{u}) - \epsilon \frac{\partial}{\partial t}u_p, \quad \forall p \neq 0, \tag{1.14}$$

which shows that the equations we obtain are ordinary differential equations with independent variable t. For $p = 0$, the left-hand side of (1.9) vanishes because of the periodicity. Then so should the right-hand. This results in

$$\frac{\partial}{\partial t}u_0 = g_0(\mathbf{u}). \tag{1.15}$$

Collecting these results, we have

$$\frac{\partial}{\partial t}u_p + \frac{ip}{\epsilon}u_p = g_p(\mathbf{u}), \quad \forall p \neq 0, \tag{1.16a}$$

$$\frac{\partial}{\partial t}u_0 = g_0(\mathbf{u}), \tag{1.16b}$$

$$\mathbf{u}(0, 0) = \sum_{p\in\mathbf{Z}} u_p(0) = \xi. \tag{1.16c}$$

We still lack enough initial conditions, or other constraints, to specify a unique solution. However, examine (1.16a), and imagine g_p to be independent of $\mathbf{u}$; thus g_p is a function of t only. Then, any solution of (1.16a) consists of a smooth solution (depending, together with its derivatives on t, and not on t/ϵ) plus a solution of the homogeneous equation. (This assertion will be clarified below.) This latter solution is highly oscillatory, and depends on $\tau = t/\epsilon$ rather than t. But u_p should be t-dependent, not τ-dependent. Hence we require *smooth solutions* of the equations (1.16a), i.e. for $p \neq 0$.

Unfortunately, the concept of a smooth solution is an imprecise notion. However, in many asymptotic results, and in many numerical considerations as well, it plays an important part, without being mentioned explicitly. The work of Karasalo [12] is an exception. Motivated by numerical considerations, he defines what is indicated here as a smooth solution. However, his approach requires strong assumptions (analyticity) and does not lead to a unique smooth solution. Here we adopt the following definition:

Definition 1.1. Consider the ordinary differential equation

$$\frac{dy}{dt} + \frac{B}{\epsilon}y = f(t) \tag{1.17}$$

in $\mathbf{C}^n$, B a nonsingular $n \times n$ matrix and f a $\mathbf{C}^n$-valued map of class C^∞. Then the *smooth solution* y_k *of order* k of (1.17) is defined by

$$y_k = \epsilon B^{-1}f - \epsilon^2 B^{-2}f' + \cdots + (-1)^k\epsilon^{k+1}B^{-(k+1)}f^{(k)}. \quad \square \tag{1.18}$$

Thus, a smooth solution of (1.17) is *not* a solution of (1.17), unless f is a polynomial and the order k is larger than or equal to the degree of this polynomial. However, y_k is close to a solution of (1.17) in an asymptotic sense as described by the following lemma:

LEMMA 1.2. *Let y be the solution of the initial value problem $dy/dt + By/\epsilon = f$, $y(0) = y_k(0)$. Then*

$$|y(t) - y_k(t)| = O(\epsilon^{k+2}),$$

uniformly on a bounded interval $[0, T]$.

Proof. We have

$$y(t) = \epsilon\exp(-Bt/\epsilon)y_k(0) + \int_0^t \exp(-B(t - s)/\epsilon)f(s)\,ds.$$

By k partial integration steps we get

$$y(t) - y_k(t) = (-1)^{k+1}\epsilon^{k+1}B^{-(k+1)}\int_0^t \exp(B(t - s)/\epsilon)f^{(k+1)}(s)\,ds.$$

Now one additional partial integration step on the right-hand side gives the result. $\square$

We return to (1.16a)–(1.16c). Our smooth solution concept as defined above seems ineffective, because the equations (1.16a) (of which a smooth solution should be obtained) are much more complicated than the simple linear ordinary differential equation (1.17). However, the smooth solution concept is basically an asymptotic concept (for $\epsilon \longrightarrow 0$), and in constructing an asymptotic series for the solution of (1.16a)–(1.16c), the above simple concept, applied repeatedly, suffices. We will show this by constructing the asymptotic series by means of a recursive process. Let us denote successive approximations for the envelopes by $[u_p]_k$ and for $\mathbf{u}$ itself by $[\mathbf{u}]_k$.

Zeroth Term. Put $[u_p]_0 = 0$, $\forall p \neq 0$. Let $[\mathbf{u}]_0 \equiv [u_0]_0$ where $[u_0]_0$ is defined as the solution of (essentially (1.16b)–(1.16c))

$$\frac{\partial}{\partial t} u_0 = g_0(u_0), \qquad u_0(0) = \xi. \tag{1.19}$$

This is an initial value problem for an ordinary differential equation which is not necessarily autonomous. According to (1.12), g_0 should be interpreted as a mapping acting on functions rather than function values, just as is G. Hence, if function value notation is preferred, (1.19) should read $\partial u_0(t)/\partial t = g_0(u_0)(t)$. This clearly shows that the notation is not the standard ordinary differential equation notation.

Approximation of Order k. Let $[\mathbf{u}]_{k-1}$ be given. Then define $[u_p]_k$ for all $p \neq 0$ as the smooth solution of order $k - 1$ of

$$\frac{\partial}{\partial t} u_p + \frac{ip}{\epsilon} u_p = g_p([\mathbf{u}]_{k-1}), \quad \forall p \neq 0. \tag{1.20}$$

See (1.12) for the g_p. Now define $[u_0]_k$ as the solution of

$$\frac{\partial}{\partial t} u_0 = g_0(u_0 + [\tilde{\mathbf{u}}]_k), \qquad u_0(0) = \xi - [\tilde{\mathbf{u}}]_k(0, 0), \tag{1.21}$$

where

$$[\tilde{\mathbf{u}}]_k(t, \tau) = \sum_{p \neq 0} e^{ip\tau} [u_p]_k(t). \tag{1.22}$$

Then we take $[\mathbf{u}]_k$ as $[\mathbf{u}]_k(t, \tau) = \Sigma_p e^{ip\tau} [u_p]_k(t)$. This is clearly a recursive definition.

We now cast this construction into terms which are independent of the explicit nature of the carriers and envelopes. We introduce the operators H, M and J_1.

$$(M\mathbf{u})(t, \tau) = \frac{1}{2\pi} \int_0^{2\pi} \mathbf{u}(t, \sigma)\, d\sigma, \qquad (J_2\mathbf{u})(t, \tau) = \int_0^{\tau} \mathbf{u}(t, \sigma)\, d\sigma, \tag{1.23}$$

$$H = (I - M)J_2(I - M), \qquad (J_1\mathbf{u})(t, \tau) = \int_0^t \mathbf{u}(s, \tau)\, ds. \tag{1.24}$$

We use the notation $[\tilde{\mathbf{u}}]_k = (I - M)[\mathbf{u}]_k$. Obviously, $[u_0]_k = M[\mathbf{u}]_k$. It is also easy to verify that

$$H(e^{ip\tau} u_p(t)) = \frac{e^{ip\tau}}{ip} u_p(t), \quad \forall p \neq 0. \tag{1.25}$$

In view of the definition of a smooth solution, we may replace the equations (1.20) by the following single equation

$$(1.26)\qquad [\widetilde{\mathbf{u}}]_k = \left[\epsilon H - \epsilon^2 H^2 \frac{\partial}{\partial t} + \cdots + (-1)^{k-1}\epsilon^k H^k \frac{\partial^{k-1}}{\partial t^{k-1}}\right] G([\mathbf{u}]_{k-1}).$$

Thus, if we define

$$(1.27)\qquad D_k = \epsilon H - \epsilon^2 H^2 \frac{\partial}{\partial t} + \cdots + (-1)^k \epsilon^{k+1} H^{k+1} \frac{\partial^k}{\partial t^k},$$

and if we use J_1 to write (1.21) as an integral equation, we have

$$(1.28a)\qquad [\widetilde{\mathbf{u}}]_k = D_{k-1} G([\mathbf{u}]_{k-1}),$$

$$(1.28b)\qquad [u_0]_k = \xi - [\widetilde{\mathbf{u}}]_k(0, 0) + J_1 g_0([\mathbf{u}]_k).$$

Using (1.28a)–(1.28b), we start anew to define the $[\mathbf{u}]_k$ recursively. For $k = 0$, put $[\widetilde{\mathbf{u}}]_0 = 0$ and use (1.28b) to define $[u_0]_0$. Then apply (1.28a)–(1.28b) with $k = 1$ to obtain $[\mathbf{u}]_1$, then with $k = 2$ to obtain $[\mathbf{u}]_2$, etc. If these $[\mathbf{u}]_k$ exist as sufficiently smooth functions, then this construction, by means of (1.28a)–(1.28b) is equivalent to the earlier one, i.e. (1.20)–(1.21), and at the same time it avoids Fourier series and related convergence questions.

The following assumption is necessary to make sure that $[\mathbf{u}]_k$, as defined recursively by (1.28a)–(1.28b) is well defined, for any given $k > 0$ and for $\epsilon > 0$ sufficiently small:

Assumption 1.3. (i) The initial value problem (cf. (1.19))

$$\frac{\partial}{\partial t} u_0 = g_0(u_0), \qquad u_0(0) = \xi,$$

has a unique solution u_0^* on $[0, T]$.

(ii) g is of class C^∞ on $[0, T] \times B(\delta)$, where

$$B(\delta) = \left\{ x \mid x \in \mathbf{R}^2, |x| \leqslant \sup_{t \in [0,T]} |u_0^*(t)| + \delta \right\}.$$

Here the vector norm $|\cdot|$ is the *euclidean* norm, and $\delta > 0$ is an appropriately chosen constant. □

Clearly, this assumption implies the unique solvability of (1.28b), provided that $[\widetilde{\mathbf{u}}]_k$ is sufficiently small in a supremum norm. The second condition implies that $[\mathbf{u}]_{k-1}$ and $G([\mathbf{u}]_{k-1})$ are of class C^∞ (by a recursive argument). Thus, for $\epsilon > 0$ sufficiently small it is indeed true that $[\widetilde{\mathbf{u}}]_k$ is sufficiently small, and unique existence of $[\mathbf{u}]_k$ follows.

For a more rigorous account we refer to Hoppensteadt-Miranker [11]. While our formalism differs from the one in [11], the relationship should be clear, in particular from (1.28a)–(1.28b). Thus, we may state the following theorem and corollary (without proof):

THEOREM 1.4 (HOPPENSTEADT-MIRANKER [11]). *Let Assumption* 1.3 *hold true, and let* $\epsilon > 0$ *be sufficiently small. Then* $[\mathbf{u}]_k$ *is well defined and we have*

$$|\widetilde{x}(t) - \Phi(t/\epsilon)[\mathbf{u}]_k(t, t/\epsilon)| = O(\epsilon^{k+1}), \qquad \epsilon \longrightarrow 0,$$

uniformly for $t \in [0, T]$. □

COROLLARY 1.5. *Under the above conditions*

$$\left|\frac{\partial^j}{\partial t^j}\{[\mathbf{u}]_k(t,\tau) - [\mathbf{u}]_{k-1}(t,\tau)\}\right| = O(\epsilon^k), \qquad \epsilon \longrightarrow 0,$$

for $j = 0, 1, 2, \ldots,$ *and uniformly in* t *and* τ, $t \in [0, T]$, $\tau \in \mathbf{R}$. □

We conclude this section with the following observations.

Remark 1.6. The relations (1.28a)–(1.28b) may be derived much more directly. However, we require the Fourier coefficients explicitly, to reveal the smooth solution concept behind the asymptotic series. The smooth solution concept forms the basis for the numerical algorithm.

Remark 1.7. The relations (1.28a)–(1.28b) demonstrate that the $[\mathbf{u}]_k$ are $\mathbf{R}^2$-valued. This is much less clear from the description in terms of Fourier coefficients. In Section 3 we will show how to choose a convenient basis in the τ-direction (i.e. set of carriers), which preserves the smooth solution concept, and which gives $\mathbf{R}^2$-valued envelopes. E.g. if we write

$$u(t,\tau) = \sum_0^\infty u_p(t)\cos p\tau + \sum_1^\infty v_p(t)\sin p\tau,$$

we have $\mathbf{R}^2$-valued envelopes, but the smooth solution concept is not applicable, because $\partial/\partial\tau$ maps cosines into sines.

The numerical algorithm to be proposed is closely related to (1.28a)–(1.28b). In place of $[\mathbf{u}]_k$ and $[\mathbf{u}]_{k-1}$ we introduce *one* discrete approximation (to $[\mathbf{u}]_k$). D_{k-1} is replaced by a discrete approximation and the equation (1.28b), essentially an ordinary differential equation is replaced by a collocation method. This set of discrete equations may be solved by an iterative process which is the analogue of the recursive process defining the $[\mathbf{u}]_k$. But equally well, a different solution method may be used, e.g. a Newton-type method. This is particularly important if $\epsilon > 0$ is not very small (in relation to problem dependent parameters or algorithm dependent parameters).

2. Approximation of a Smooth Solution. The smooth solution concept plays an important part in the construction of the $[\mathbf{u}]_k$ in Section 1. Indeed our aim is to approximate $\Sigma e^{ip\tau}[u_p]_k$. In this section we discuss the approximation of a smooth solution by a numerical method. The simple methods which will be proposed play an important part in the numerical algorithm for approximating an asymptotic expansion for the solution of (1.1).

Consider the ordinary differential equation (cf. Definition 1.1)

$$\frac{dy}{dt} + \frac{B}{\epsilon}y = f(t). \tag{2.1}$$

Here f is a $\mathbf{C}^n$-valued map of class C^∞, and B is a nonsingular matrix. Let $\pi_k f$ denote an interpolation polynomial of f on $k+1$ abscissae; the abscissae should all be different from each other, and they are supposed to belong to a segment $[0, h]$. Here h plays the role of the stepsize. Let $|\cdot|$ denote a norm on $\mathbf{C}^n$. As matrix-norm we use the corresponding l.u.b.-norm. The following lemma shows how to approximate a smooth solution of order k of (2.1).

LEMMA 2.1. *The equation*

$$\frac{dy^h}{dt} + \frac{B}{\epsilon} y^h = \pi_k f \tag{2.2}$$

in the unknown y^h, y^h a polynomial of degree $\leqslant k$, has a unique solution given by

$$y^h = \epsilon B^{-1}\pi_k f - \epsilon^2 B^{-2}(\pi_k f)' + \cdots + (-1)^k \epsilon^{k+1} B^{-(k+1)}(\pi_k f)^{(k)}. \tag{2.3}$$

Moreover, if y_k is the smooth solution of order k of (2.1), *then there exists a constant $c(k)$, depending on k, but not on ϵ, h and f, such that in the $L^\infty(0, h)$-norm $\| \cdot \|$*

$$\|y_k - y^h\| \leqslant \epsilon c(k) h^{k+1} \left(|B^{-1}| + \frac{\epsilon}{h}|B^{-2}| + \cdots + \frac{\epsilon^k}{h^k}|B^{-(k+1)}| \right) \|f^{(k+1)}\|. \tag{2.4}$$

Proof. Clearly, y^h as given by (2.3) is a solution of (2.2). Any other solution of (2.2) therefore differs from y^h by a solution of the homogeneous equation $dy/dt + By/\epsilon = 0$. Since B is nonsingular, a solution of the latter equation is never a polynomial, unless identically zero. This proves the uniqueness of y^h. The estimate (2.4) follows from a termwise application of the well-known estimate

$$\left\| \frac{d^j}{dt^j} \pi_k f - f^{(j)} \right\| \leqslant c_j h^{k-j+1} \|f^{(k+1)}\|$$

with constant c_j independent of h and f and the ensuing assignment $c(k) = \max_{j \leqslant k} c_j$. □

Clearly, formula (2.3) defines an *algorithm* for the approximation of the smooth solution y_k of order k. The approximation error is given by (2.4).

In general, the algorithm defined by (2.3) will be applied on consecutive subintervals $(0, h)$, $(h, 2h)$, $(2h, 3h)$ etc. On each subinterval, the approximation error is given by (2.4), with $\| \cdot \|$ interpreted as the L^∞-norm on that subinterval. Hence, if the total interval is $(0, T)$, then the approximation error again satisfies (2.4), but with $\| \cdot \|$ interpreted as the $L^\infty(0, T)$-norm.

The method defined by (2.3) is *self-starting.* No initial values are required. However, if y^h is an approximation for y_k, obtained by applying the method on consecutive subintervals, then y^h is in general *discontinuous,* with jump discontinuitites at the joins of the subintervals. This is caused by the $(\pi_k f)^{(j)}$-terms. The discontinuity is very mild, since the size of a jump is at most twice the size of the approximation error $y_k - y^h$.

The method defined by (2.3) should not be applied if $\epsilon|B^{-1}|/h$ is "large." Indeed, on a suitable basis in the linear space of all polynomials of degree $\leqslant k$, the terms $(\pi_k f)^{(j)}$ in (2.3) are simply the difference formulae for the approximation of $f^{(j)}$. Such formulae are well known for their instability with respect to roundoff. Since $y^h = O(\epsilon B^{-1})$ and the $(\pi_k f)^{(j)}$-terms in (2.3) are premultiplied by $\epsilon^{j+1}B^{-(j+1)}$, then in the scale of y^h, each such term is premultiplied by $\epsilon^j B^{-j}$. This premultiplication compensates for the numerical instability, or even annihilates it, if $\epsilon|B^{-1}|/h$ is sufficiently small.

If $f = f(t, y)$, formula (2.3) still applies, provided that $\pi_k f$ is interpreted as $(\pi_k f)(t) = \pi_k f(\cdot, y^h(\cdot))(t)$; thus $\pi_k f$ contains the unknown y^h. Clearly, the problem

$dy/dt + By/\epsilon = f(\cdot, y)$ is a model for the equations (1.16a), which are somewhat more complicated. In the numerical algorithm to be proposed, we make use of the equations (1.16a) and the method defined by (2.3), with nonlinear right-hand side. In this way, the recursive process for the $[u_p]_k$ is replaced by a single equation. In Section 5, the unique solvability and the approximation properties of this process will be discussed.

Practical experience indicates that the process described is quite costly if $f = f(t, y)$. In such a situation, the process has all the disadvantages of a Galerkin or collocation method, and the arguments of Keller [13] against such methods apply to the above process as well. Therefore, we look for an alternative procedure to couple with the one already described.

It will be shown for the simple equation (2.1), that the backward differentiation multistep formulae provide an alternative for the method defined by (2.3), as long as appropriate starting values are supplied (the latter being furnished by the method (2.3)). See Henrici [9], for general information about multistep methods, Gear [7], [8] about the backward differentiation formulae in particular, and Lambert [14] about stability domains in general.

The backward differentiation formulae are of the form

$$y_n + \alpha_1 y_{n-1} + \cdots + \alpha_r y_{n-r} = h\beta_0 f_n, \qquad \alpha_r \neq 0. \tag{2.5}$$

(We use the standard multistep notation; e.g. when applied to (2.1) one should read $f_n = -By_n/\epsilon + f(t_n)$ in (2.5).) The coefficients $\alpha_1, \ldots, \alpha_r, \beta_0$ with $\alpha_r \neq 0$, are determined by requiring (2.5) to be both stable in the sense of Dahlquist [5] and exact if applied to an ordinary differential equation with a solution which is a polynomial of degree $\leqslant r$. For $r = 1, 2, \ldots, 6$ these requirements are known to determine unique coefficients. See Gear [7]. Thus, we restrict ourselves to $r = 1, 2, \ldots, 6$. We now state and prove Lemmas 2.2–2.5 which describe the accuracy and stability with which the backward differentiation formulae (2.5) may be used to approximate smooth solutions.

LEMMA 2.2. *Let y_k be the smooth solution of order k of* (2.1). *For this y_k the local discretization error of the method* (2.5) *is given by*

$$\delta_n \equiv \frac{1}{h}\left\{y_k(t_n) + \cdots + \alpha_r y_k(t_{n-r}) + \beta_0 h \frac{B}{\epsilon} y_k(t_n) - \beta_0 h f(t_n)\right\}$$

$$= (-1)^k \beta_0 \epsilon^{k+1} B^{-(k+1)}(t_n) + h^r O(\|y_k^{(r+1)}\|),$$

with $\|\cdot\|$ the $L^\infty(t_{n-r}, t_n)$-norm.

Proof. The construction of the coefficients implies that for any smooth function, and thus for y_k in particular

$$\frac{1}{h}\{y_k(t_n) + \cdots + \alpha_r y_k(t_{n-r})\} = \beta_0 y_k'(t_n) + h^r O(\|y_k^{(r+1)}\|).$$

Hence, by simple substitution

$$\delta_n = \beta_0\left\{y_k'(t_n) + \frac{B}{\epsilon} y_k(t_n) - f(t_n)\right\} + h^r O(\|y_k^{(r+1)}\|).$$

For y_k we have the explicit formula (1.18). With this formula, the assertion readily follows. □

The following stability properties of the method (2.5) are well known, see Gear [7].

LEMMA 2.3. *Apply the method* (2.5) *to the scalar equation* $y' = \lambda y$. *Then, for all starting values* $y_0, y_1, \ldots, y_{r-1}$, *and for* $|h\lambda|$ *sufficiently large, we have* $y_n \longrightarrow 0$, $n \longrightarrow \infty$. *If* $\mathrm{Re}(\lambda) \leqslant 0$, *then* $y_n \longrightarrow 0$, $n \longrightarrow \infty$ *for all* $h > 0$ *and all starting values* $y_0, y_1, \ldots, y_{r-1}$, *provided that* $r = 1, 2$. □

This stability result is not enough; we also need a bound for the y_n. Indeed, the above lemma does not exclude that under the circumstances cited $\sup_n |y_n| \longrightarrow \infty$ for fixed starting values and $|h\lambda| \longrightarrow \infty$. We show:

LEMMA 2.4. *Apply the method* (2.5) *to the scalar equation* $y' = \lambda y$. *Then, for* $|h\lambda|$ *sufficiently large*

$$|y_n| \leqslant \max_{j=0,\ldots,r-1} |y_j|, \quad \forall n.$$

Proof. For $|h\lambda|$ large, the roots of the polynomial $(1 - \beta_0 h\lambda)x^r + \alpha_1 x^{r-1} + \cdots + \alpha_r$ are approximately given by

$$w_j \sim e^{2\pi ij/r} w, \qquad w = \left(\frac{\alpha_r}{\beta_0 h\lambda}\right)^{1/r}.$$

Write the solution y_n as $y_n = \Sigma_{j=0}^{n-1} a_j w_j^n$. The a_j are determined from the starting values by a linear equation, the matrix of which is the Vandermonde matrix

$$\begin{pmatrix} 1 & \cdots & 1 \\ w_0 & \cdots & w_{r-1} \\ \vdots & & \vdots \\ w_0^{r-1} & \cdots & w_{r-1}^{r-1} \end{pmatrix} \sim \begin{pmatrix} 1 & & \\ & w & \\ & & \ddots \\ & & & w^{r-1} \end{pmatrix} \begin{pmatrix} 1 & \cdots & 1 \\ 1 & \cdots & e^{-2\pi i/r} \\ \vdots & & \vdots \\ 1 & \cdots & e^{-2\pi i(r-1)/r} \end{pmatrix}.$$

Apart from a multiplicative factor $\sqrt{r}$, the last matrix here is unitary. Hence, $a_j = O(w^{1-r})$, for $j = 0, 1, \ldots, r-1$. Thus, $y_n = O(w^{1-r}w^n) = O(w^{n-r+1}) = O(w)$ for all $n \geqslant r$. But $w \longrightarrow 0$ for $|h\lambda| \longrightarrow \infty$. This proves the assertion. □

This result can be improved upon by carefully estimating the constants in the O-symbols in the above proof. Upon doing so we obtain the following lemma by standard techniques.

LEMMA 2.5. *Let* B *be diagonalizable, and let the method* (2.5) *be applied to* (2.1). *Then there exists a constant* c, *depending only on* B *and* r, *such that for* ϵ/h *sufficiently small and all* $n \geqslant r$

$$|y_k(t_n) - y_n| \leqslant \frac{ch}{1-\rho} \max_{p \leqslant n} |\delta_p| + c\rho^{n-r+1} \max_{0 \leqslant j \leqslant r-1} |y_k(t_j) - y_j|$$

with $\rho \sim O((\epsilon/h)^{1/r})$ *for* $\epsilon/h \longrightarrow 0$. □

Thus, because of the strong damping, the error in the starting values scarcely affects the y_n for $n \geqslant r$, and the global discretization error is mainly determined by

$h \times \max|\delta_p|$, with the δ_p given in Lemma 2.2. This shows that the backward differentiation multistep formulae may be used to approximate a smooth solution.

Because of the strong damping, starting values $y_j = 0, j = 0, 1, \ldots, r-1$, might be used. However, in a nonlinear situation, and if ϵ/h and ϵ are not extremely small, this should not be done. It is much better to use the method defined by (2.3) to generate the starting values. This is costly, but it has to be done only once. This process will be explained in more detail in defining the numerical algorithm for (1.1).

Finally, it should be observed that the problems and the algorithms of this section are closely related to earlier work of Miranker-Wahba [15]. Some ideas, connecting this section with their methods may be found in [16].

3. A Question of Formulation. In this section we show how to avoid complex, nonreal numbers in the formulation of the problem, and thus in the algorithm. In the next section we will describe the algorithm in a complex linear space, and thus the algorithm would require complex arithmetic in an actual computer program. It is possible to avoid the complex arithmetic, if the asymptotic series for the solution of (1.1) is real, and this is the common situation. One simply has to change the carriers (basis functions in the τ-direction, e.g. the $\{e^{ip\tau}\}$ in Section 1). In changing the carriers, one changes the equations (1.16a)–(1.16c). Thus the smooth solutions change. We need to make sure that the change in the carriers is reversible with respect to the smooth solution concept. Even more, we need to make sure that the smooth solution concept is preserved under the change of the carriers. E.g. replacing the $\{e^{ip\tau}\}$ by the basis functions $\{\cos p\tau, \sin p\tau\}$ destroys the smooth solution concept of Definition 1.1. It will be shown that a basis involving powers of $\Phi(\tau)$ (cf. (1.4)) is a good substitute for the basis of the $\{e^{ip\tau}\}$.

The result of this section is the following theorem:

THEOREM 3.1. (i) *If $f_1, f_{-1} \in \mathbf{C}^2$, then there exist unique $g_1, g_{-1} \in \mathbf{C}^2$ such that for all $\tau \in \mathbf{R}$, $f_1 e^{i\tau} + f_{-1} e^{-i\tau} = \Phi(\tau) g_1 + \Phi^{-1}(\tau) g_{-1}$.*

(ii) *For $j = 1, -1$ let u_j be the smooth solution of order k of*

$$\frac{d}{dt} u_j + \frac{ij}{\epsilon} u_j = f_j(t), \qquad f_j \in C^k.$$

Also, for $j = 1, -1$, let y_j be the smooth solution of order k of (cf. (1.2) for A)

$$\frac{d}{dt} y_j + j \frac{A}{\epsilon} y_j = g_j(t), \qquad g_j \in C^k.$$

Then $u_1 e^{i\tau} + u_{-1} e^{-i\tau} = \Phi(\tau) y_1 + \Phi^{-1}(\tau) y_{-1}$ iff $f_1 e^{i\tau} + f_{-1} e^{-i\tau} = \Phi(\tau) g_1 + \Phi^{-1}(\tau) g_{-1}$. □

Indeed, this is what we need. The first statement asserts that the basis functions $e^{i\tau}$ and $e^{-i\tau}$ have a representation in terms of the real functions $\Phi(\tau)$ and $\Phi^{-1}(\tau)$. The second statement asserts that this change in basis does not affect the smooth solution: the smooth solution concept commutes with this change in basis.

Proof of the Theorem. The first statement follows easily: First, we observe that $A = U\Lambda U^*$, U unitary and that

$$\Lambda = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}.$$

Second, we use $\Phi(\tau) = \exp A\tau = U(\exp \Lambda\tau)U^*$.

Making use of the definition of a smooth solution, it may be seen that the second assertion is implied by the equivalence relation

$$\{\forall p \in \mathbf{N},\ i^{-p} f_1 e^{i\tau} + (-i)^{-p} f_{-1} e^{-i\tau} = A^{-p}\Phi(\tau)g_1 + (-A)^{-p}\Phi^{-1}(\tau)g_{-1}\}$$
$$\Longleftrightarrow \{f_1 e^{i\tau} + f_{-1} e^{-i\tau} = \Phi(\tau)g_1 + \Phi^{-1}(\tau)g_{-1}\}.$$

So we prove this equivalence.

$\Leftarrow$: Differentiate $f_1 e^{i\tau} + f_{-1} e^{-i\tau}$ and $\Phi(\tau)g_1 + \Phi^{-1}(\tau)g_{-1}$ each p times with respect to τ. Observe $d\Phi^{-1}/d\tau = -A\Phi^{-1}$. Then use

$$i^{-2p} I = (-i)^{-2p} I = A^{-2p} = (-A)^{-2p} = (-1)^p I, \quad \forall p. \tag{3.1}$$

$\Rightarrow$: Integrate $i^{-p} f_1 e^{i\tau} + (-i)^{-p} f_{-1} e^{-i\tau}$ and $A^{-p}\Phi(\tau)g_1 + (-A)^{-p}\Phi^{-1}(\tau)g_{-1}$ each p times with respect to τ. Making use of (3.1), the result is,

$$f_1 e^{i\tau} + f_{-1} e^{-i\tau} = \Phi(\tau)g_1 + \Phi^{-1}(\tau)g_{-1} + \pi(\tau),$$

with $\pi(\tau)$ a polynomial in τ. Since $\Phi(\tau)$, $\Phi^{-1}(\tau)$ and $e^{\pm i\tau}$ are bounded uniformly in $\tau \in R$, $\pi(\tau)$ is a constant. However, in the τ-direction, $f_1 e^{i\tau} + f_{-1} e^{-i\tau}$ and $\Phi(\tau)g_1 + \Phi^{-1}(\tau)g_{-1}$ have mean value zero on $[0, 2\pi]$. Thus $\pi(\tau) \equiv 0$. □

4. The Algorithm. In this section the numerical algorithm is described.

The algorithm to be described consists of two parts: a starting method, and a method to be used once suitable starting values have been obtained. The latter is essentially Gear's multistep algorithm for stiff equations, see [7], and so we concentrate on the starting method.

The starting method is a discretization of the equations (1.16a)–(1.16c). First, the number of Fourier coefficients is made finite, and only Fourier coefficients with index p, $|p| \leqslant d$ are approximated. For $p \neq 0$, the approximation method in the t-direction is the self-starting smooth solution solver of Section 2. For $p = 0$, a collocation method is used, cf. Axelsson [2], Weiss [23] and Russell [18] for collocation methods. The use of a collocation method for $p = 0$ is quite natural, provided that the collocation points coincide with the abscissae of π_k, the interpolation polynomial projector used in the smooth solution solver.

It is also possible to view the algorithm as a discretization of (1.9). First a discretization in the t-direction and then a discretization in the τ-direction as well. In both cases, the discretization is a projection method. I.e., it is defined by a projection and (two) function spaces. We will now describe these discretizations.

Discretization in the τ-Direction. Consider a function $\mathbf{u} = \mathbf{u}(\tau)$. It is not important here whether $\mathbf{u}$ depends on τ and t rather than τ, because processing in the τ-direction is performed for t kept fixed. So we may as well omit t in the notation, at least in the description of the discretization in the τ-direction. We assume $\mathbf{u} = \mathbf{u}(\tau)$ to be periodic, period 2π. Its Fourier coefficients u_p are given by

$$u_p = \frac{1}{2\pi}\int_0^{2\pi} \mathbf{u}(\sigma)e^{-ip\sigma}\, d\sigma.$$

There exists a simple discrete version of this formula, which makes the approximation of the u_p, $|p| \leqslant d$, possible. Let $m \geqslant 2d + 1$ be an integer and set

$$\tau_j = \frac{2\pi j}{m}, \qquad j = 0, 1, 2, \ldots, m-1. \tag{4.1}$$

Then, u_p is approximated by $\tilde{u}_p$, where

$$\tilde{u}_p = \frac{1}{m}\sum_{j=0}^{m-1} \mathbf{u}(\tau_j)e^{-ip\tau_j}. \tag{4.2}$$

In (4.2) there is no distinction between $\tilde{u}_p$ and $\tilde{u}_q$ if $p = q \pmod m$. This is the well-known aliasing effect.

Let $\mathbf{u} = \mathbf{u}(\tau)$ (or $\mathbf{u} = \mathbf{u}(t, \tau)$, but then t is kept fixed) be continuous on $[0, 2\pi)$. Then the $\tilde{u}_p$ are well defined, because $\tau_j \in [0, 2\pi)$ for $j = 0, 1, \ldots, m-1$. Thus, we may define $\Pi_{d,m}$ by

$$(\Pi_{d,m}\mathbf{u})(\tau) = \sum_{|p|\leqslant d} e^{ip\tau}\tilde{u}_p. \tag{4.3}$$

Clearly, $\Pi_{d,m}$ assigns to $\mathbf{u}$ a trigonometric polynomial of degree $\leqslant d$. Since $m \geqslant 2d + 1$, $\tilde{u}_p = u_p$ for $\mathbf{u}$ a trigonometric polynomial of degree $\leqslant d$. Then $\Pi_{d,m}^2 = \Pi_{d,m}$. I.e., $\Pi_{d,m}$ is a projection.

As a special case, choose $d = 0$. Clearly, $\Pi_{0,m}\mathbf{u}$ approximates the mean value $(2\pi)^{-1}\int_0^{2\pi}\mathbf{u}(\sigma)d\sigma$ by the mean value of the data $\mathbf{u}(\tau_j), j = 0, 1, \ldots, m-1$. For ease of exposition we therefore use the notation

$$M_m = \Pi_{0,m}. \tag{4.4}$$

Discretization in the t-Direction. Consider a function $\mathbf{u} = \mathbf{u}(t)$. It does not matter whether or not $\mathbf{u} = \mathbf{u}(t, \tau)$ rather than $\mathbf{u} = \mathbf{u}(t)$, because all processing in the t-direction is performed for τ kept fixed. So we simply drop τ. We will make use of two discretizations in the t-direction.

The first discretization is the projection π_k, which assigns to $\mathbf{u}$ the Lagrange interpolation polynomial on $k + 1$ different abscissae in $[0, h]$.

The second discretization which we need is called $\tilde{\pi}_{k-1}$. $\tilde{\pi}_{k-1}$ is defined by: $\tilde{\pi}_{k-1}\mathbf{u}$ is a polynomial of degree $\leqslant k - 1$ and $\tilde{\pi}_{k-1}\mathbf{u} - \pi_k\mathbf{u}$ is orthogonal to all polynomials of degree $\leqslant k - 1$ in the $L^2(0, h)$ inner product. Thus, if P_i is the Legendre polynomial of degree i, shifted to $(0, h)$, and if $\pi_k\mathbf{u} = \alpha_0 P_0 + \cdots + \alpha_k P_k$, then $\tilde{\pi}_{k-1}\mathbf{u} = \alpha_0 P_0 + \cdots + \alpha_{k-1}P_{k-1}$. This shows that π_k and $\tilde{\pi}_{k-1}$ use function values evaluated at the same abscissae.

Discretization in the t, τ-Plane. Let $\mathbf{u} = \mathbf{u}(t, \tau)$ be continuous on $[0, h] \times \mathbf{R}$, and periodic with period 2π in the τ-direction for all t. By keeping t fixed, we may apply $\Pi_{d,m}$ to $\mathbf{u}(t, \cdot)$, and then, by keeping τ fixed, we may apply π_k (or $\tilde{\pi}_{k-1}$) to $(\Pi_{d,m}\mathbf{u})(\cdot, \tau)$. If $\mathbf{u}$ is sufficiently smooth in the τ-direction for all t, then $\mathbf{u}$ may be written as the convergent series

$$\mathbf{u}(t, \tau) = \sum_{p \in \mathbf{Z}} e^{ip\tau} u_p(t). \tag{4.5}$$

Obviously,

$$(\Pi_{d,m}\mathbf{u})(t, \tau) = \sum_{|p| \leqslant d} e^{ip\tau} \tilde{u}_p(t), \tag{4.6}$$

$$(\pi_k \Pi_{d,m}\mathbf{u})(t, \tau) = \sum_{|p| \leqslant d} e^{ip\tau} (\pi_k \tilde{u}_p)(t), \tag{4.7}$$

where the $\tilde{u}_p$ are given by (4.2) (t considered as a parameter).

On the other hand, $\Pi_{d,m}\pi_k\mathbf{u}$ is also well defined by the continuity of $\mathbf{u}$. Again, for $\mathbf{u}$ sufficiently smooth in the τ-direction, we have

$$(\pi_k\mathbf{u})(t, \tau) = \sum_{p \in \mathbf{Z}} e^{ip\tau} (\pi_k u_p)\dot{}(t), \tag{4.8}$$

while $\Pi_{d,m}\pi_k\mathbf{u} = \pi_k\Pi_{d,m}\mathbf{u}$ for $\mathbf{u}$ smooth. The following Lemma 4.1 and its Corollary 4.2 assert that continuity alone of $\mathbf{u}$ is sufficient for the commutativity of these projections.

LEMMA 4.1. *Let* $\mathbf{u}$ *be continuous on* $[0, h] \times (0, 2\pi)$. *Then* $\pi_k\Pi_{d,m}\mathbf{u} = \Pi_{d,m}\pi_k\mathbf{u}$.

Proof. Let $\mathbf{v}(t, \tau)$ be of the form $\Sigma\Sigma a_{pq} e^{ip\tau} t^q$, with *finite* summations, such that $\mathbf{v}(t_r, \tau_j) = \mathbf{u}(t_r, \tau_j)$ for all points τ_j, $j = 0, 1, \ldots, m-1$, and all $k+1$ abscissae t_r of π_k. Thus $\mathbf{v}$ is smooth in the above sense (no convergence problems because of the finiteness of the sums), and $\pi_k\Pi_{d,m}\mathbf{u} = \pi_k\Pi_{d,m}\mathbf{v} = \Pi_{d,m}\pi_k\mathbf{v} = \Pi_{d,m}\pi_k\mathbf{u}$. □

COROLLARY 4.2. *The projections* $\Pi_{d,m}$ *and* M_m *in the* τ*-direction commute with the projections* π_k *and* $\tilde{\pi}_{k-1}$ *in the* t*-direction if applied to continuous functions on* $[0, h] \times (0, 2\pi)$. □

The Discretization of (1.9) *in the* τ*-Direction.* The basic equation (1.9) is hyperbolic. Its solution is subjected to boundary conditions for $\tau = 0$, $\tau = 2\pi$ (the periodicity in τ), an initial (point) condition as well as the smooth solution concept in the t-direction. This situation is familiar in numerical analysis and one of the frequently used discretization methods is the "methods of lines". (This method is best known for parabolic equations.) The idea of this method is to discretize first in one direction (this yields the lines) and then in the other. We do exactly that. Following common practice, we discretize first in the variable for which boundary conditions are the constraints: in our case τ. As "line functions" we use envelopes, i.e., Fourier coefficients, and the discretization replaces an infinite number of them by a finite number. This is accomplished by $\Pi_{d,m}$. Thus, we construct a system of ordinary differential equations (with independent variable t) which approximates (1.9). The unknown will be of the form

$$\mathbf{u}_d(t, \tau) = \sum_{|p| \leqslant d} e^{ip\tau} u_p(t) \tag{4.9}$$

with differentiable u_p. Observe $(\Pi_{d,m}\mathbf{u}_d)(t, \tau) = \mathbf{u}_d(t, \tau)$ for all $(t, \tau) \in [0, h] \times (0, 2\pi)$. In Section 1 we used the Fourier coefficients $g_p(\mathbf{u})(t)$. Here we need their discrete counterparts,

$$\tilde{g}_p(\mathbf{u})(t) = \frac{1}{m}\sum_{j=0}^{m-1} e^{-ip\tau_j} G(\mathbf{u})(t, \tau_j). \tag{4.10}$$

This is just formula (4.2) for $G(\mathbf{u})$ instead of $\mathbf{u}$. Quite clearly

$$(\Pi_{d,m} G(\mathbf{u}))(t, \tau) = \sum_{|p|\leqslant d} e^{ip\tau}\tilde{g}_p(\mathbf{u})(t). \tag{4.11}$$

Since $\Pi_{d,m}\mathbf{u}_d = \mathbf{u}_d$, it follows that $\partial\Pi_{d,m}\mathbf{u}_d/\partial t = \partial\mathbf{u}_d/\partial t$ and that $\partial\Pi_{d,m}\mathbf{u}_d/\partial\tau = \partial\mathbf{u}_d/\partial\tau$. Thus we replace (1.9), written as

$$\frac{\partial}{\partial t}\mathbf{u} = -\frac{1}{\epsilon}\frac{\partial}{\partial\tau}\mathbf{u} + G(\mathbf{u}) \tag{4.12}$$

by

$$\frac{\partial}{\partial t}\mathbf{u}_d = -\frac{1}{\epsilon}\frac{\partial}{\partial\tau}\mathbf{u}_d + \Pi_{d,m} G(\mathbf{u}_d). \tag{4.13}$$

This leads to the following system of ordinary differential equations for the u_p (cf. (4.11)).

$$\frac{\partial}{\partial t}u_p = -\frac{ip}{\epsilon}u_p + \tilde{g}_p(\mathbf{u}_d), \qquad 0 < |p| \leqslant d. \tag{4.14a}$$

$$\frac{\partial}{\partial t}u_0 = \tilde{g}_0(\mathbf{u}_d). \tag{4.14b}$$

This defines $2d + 1$ ordinary differential equations in the $2d + 1$ unknown $\mathbf{C}^2$-valued functions u_p. This resembles (1.16a)–(1.16b), and it explains the nomenclature: *method of envelopes.*

The Self-Starting Method. Now we discretize (4.13), or equivalently (4.14a)–(4.14b), in the t-direction. We require smooth solutions for the equations (4.14a) and therefore, in view of Lemma 2.1 and the discussion following it, we replace the u_p in (4.9) by polynomials u_p^h. I.e., we replace $\mathbf{u}_d$ by $\mathbf{u}_{d,k}$,

$$\mathbf{u}_{d,k}(t, \tau) = \sum_{|p|\leqslant d} e^{ip\tau}u_p^h(t). \tag{4.15}$$

Each polynomial u_p^h has degree $\leqslant k$. Then (4.14a) itself is replaced by

$$\frac{\partial}{\partial t}u_p^h = -\frac{ip}{\epsilon}u_p^h + \pi_k\tilde{g}_p(\mathbf{u}_{d,k}), \qquad 0 < |p| \leqslant d, \tag{4.16a}$$

as suggested in Section 2. Since degree $(u_0^h) \leqslant k$, we replace the ordinary differential equation (4.14b) by the discretized ordinary differential equation (projection method)

$$\frac{\partial}{\partial t}u_0^h = \tilde{\pi}_{k-1}\tilde{g}_0(\mathbf{u}_{d,k}). \tag{4.16b}$$

Finally, we have the initial condition

$$\mathbf{u}_{d,k}(0, 0) = \sum_{|p|\leqslant d} u_p^h(0) = \xi. \tag{4.16c}$$

Thus, (4.16a)–(4.16c) is a direct discretization of (1.16a)–(1.16c).

The equations (4.16a)–(4.16b) are nonlinear in the unknown polynomials u_p^h (through the $\tilde{g}_p$-terms). If there is a unique solution for the u_p^h, it should be obtained by an iterative process. Picard iteration could be used for (4.16b). In (4.16a) we can make the troublesome $\partial/\partial t$-term more explicit, if we write this equation in the following equivalent form, hinted at in Section 2.

$$u_p^h = \frac{\epsilon}{ip}\pi_k \tilde{g}_p(\mathbf{u}_{d,k}) + \cdots + (-1)^k \left(\frac{\epsilon}{ip}\right)^{\kappa+1} \frac{\partial^k}{\partial t^k} \pi_k \tilde{g}_p(\mathbf{u}_{d,k}). \tag{4.17}$$

This suggests successive substitution, because of $\epsilon > 0$ being small. For ϵ and h not extremely small, one might wish to accelerate the process by a Newton-type device. However, the concomitant explicit computation of the Jacobian matrix of the right-hand side of (4.16b), (4.17) seems costly, and should therefore be avoided. Numerical differentiation for determination of the Jacobian matrix seems more appropriate.

Finally, we give a carrier independent formulation of the starting method. We use H, M, J_1 and D_k as in Section 1, cf. (1.23), (1.24), (1.27). Moreover, we use the notation $\tilde{\mathbf{u}}_{d,k} = (I - M)\mathbf{u}_{d,k}$. Then the equivalent form is (use (4.17))

$$\tilde{\mathbf{u}}_{d,k} = D_k \pi_k \Pi_{d,m} G(\mathbf{u}_{d,k}), \tag{4.18a}$$

$$u_0^h = \xi - \tilde{\mathbf{u}}_{d,k}(0, 0) + J_1 \tilde{\pi}_{k-1} M_m G(\mathbf{u}_{d,k}). \tag{4.18b}$$

(Compare (1.28a)–(1.28b).) Indeed, our starting method is a discretization of the type indicated in Section 1.

The Multistep Method. The above algorithm is obviously quite expensive, since on each subinterval, for instance on $(0, h)$, $(2d + 1) \times (k + 1)$ unknown 2-vectors are to be determined from the nonlinear equations (4.16a)–(4.16c). Using a multistep method as a discretization of (4.14a)–(4.14b), requires the determination of $2d + 1$ unknown 2-vectors in each step. Therefore, we propose the use of the multistep method (2.5). However, the starting values should be obtained by employing the self-starting method.

We return to (4.14a)–(4.14b). This is a stiff system of ordinary differential equations. For $\epsilon > 0$ small, the Jacobian matrix of this system is to a reasonable approximation given by the linear terms in (4.14a)–(4.14b), i.e., the factor ip/ϵ multiplying u_p. Hence, the Jacobian matrix has many eigenvalues on, or close to the imaginary axis, and having large modulus as well. Of course, (4.14a)–(4.14b) is a discretized hyperbolic partial differential equation (discretized in the τ-direction only) and the eigenvalues of the Jacobian matrix correspond to modes for the hyperbolic problem. Usually in the numerical treatment of hyperbolic problems, these modes must be represented reasonably well (energy conservation, or at least no strong damping). Here the smooth solution concept enables us to avoid that requirement. Therefore, we propose the use of the backward differentiation formulae (2.5) to solve (4.14a)–(4.14b) approximately, despite the strong damping properties of these multistep methods for large modes. Indeed, the strong damping property is advantageous. See the simple, but

illustrative discussion in Section 2 for properties of the multistep methods (2.5) in approximating a smooth solution.

The multistep method needs starting values. These values should be supplied by the self-starting method.

A variable order, variable stepsize implementation of the multistep methods (2.5) might be used. However, one should be aware of the instability of the higher order methods ($r \geqslant 3$) for eigenvalues of moderate size along, or close to, the imaginary axis.

Approximation of $\tilde{x}$. Both the self-starting method and the multistep method yield approximations for the u_p, $|p| \leqslant d$. Thus, $\tilde{x}(t)$ is directly approximated by

$$\Phi(t/\epsilon) \sum_{|p| \leqslant d} e^{ipt/\epsilon} u_p^h(t),$$

in which expression t and τ are coupled by $\tau = t/\epsilon$. For the multistep method we might need some additional interpolation, since only discrete values of u_p^h are obtained.

Special Cases. The above self-starting algorithm is defined for $k + 1$ arbitrary abscissae on $[0, h]$. In the actual algorithm these abscissae must be fixed, and Lobatto points, Radau points or Gaussian points seem a reasonable choice. In the case of Gaussian points, the algorithm might be changed slightly: instead of using a polynomial of degree $\leqslant k$ for u_0^h, one may use a polynomial of degree $\leqslant k + 1$ for u_0^h, but not for the other u_p^h. The same possibility applies for Radau points, but not for Lobatto points.

Also, in the multistep method one might use an Adams-Moulton multistep method for u_0^h, but not for the other u_p^h, $p \neq 0$.

5. Error Estimates. In this section we give a rigorous analysis of the self-starting method. We proceed by introducing a set of equations, equivalent to the ones which define the self-starting method. This new set of equations allows the use of the Banach fixed point theorem without the restriction that ϵ/h is small. The result is existence and uniqueness of a solution of the equations defined by the self-starting method. Also, error estimates are obtained, see Theorem 5.14 below. Then we use this basic result in analyzing some properties of the self-starting method: superconvergence; stability with respect to perturbations of the initial vector; global error estimates. These results are obtained without the restriction that ϵ/h is small. We also obtain stability with respect to perturbations of the right-hand side of the equations (rounding errors). However, this particular result requires a restriction of the form $\epsilon/h \leqslant c/k^2$, c a constant. Cf. Consequence 5.22 below.

An analysis based on the assumption ϵ/h small might very well be obtained. Cf. (4.17) and the successive substitution process suggested there. Nevertheless, we adopt an approach without the condition that ϵ/h is small. This permits a sharp distinction between the role of ϵ and the role of h. Such a distinction is important, because ϵ and h do play differing parts. The parameter ϵ is specified by the problem, i.e. the equation (1.1), while h is chosen in applying the self-starting method. Thus, from the point of view of numerical analysis, ϵ is not a free parameter, while h is a free parameter.

Readers not interested in technical details may proceed to Theorem 5.14 below; the technical results preceding Theorem 5.14 may be skipped as they are not required for understanding the statements of the results. However, one should take notice of the norms $\|\cdot\|$, $|||\cdot|||$ (see (5.5), (5.6) respectively), and the symbol $\eta_{d,m}$ (see the discussion immediately preceding Theorem 5.14).

To begin with we reproduce the relations which determine $[\mathbf{u}]_k$, cf. (1.28a)–(1.28b) and $\mathbf{u}_{d,k}$, cf. (4.18a)–(4.18b). These are

$$[\tilde{\mathbf{u}}]_k = D_{k-1}G([\mathbf{u}]_{k-1}), \tag{5.1a}$$

$$u_0 = \xi - [\tilde{\mathbf{u}}]_k(0, 0) + J_1 g_0([\mathbf{u}]_k), \tag{5.1b}$$

and

$$\tilde{\mathbf{u}}_{d,k} = D_k \pi_k \Pi_{d,m} G(\mathbf{u}_{d,k}), \tag{5.2a}$$

$$u_0^h = \xi - \tilde{\mathbf{u}}_{d,k}(0, 0) + J_1 \tilde{\pi}_{k-1} \tilde{g}_0(\mathbf{u}_{d,k}), \tag{5.2b}$$

respectively. Here, and henceforth, we use the notation of Section 1 and Section 4 respectively. In particular, recall $g_0(\mathbf{u}) = MG(\mathbf{u})$, $\tilde{g}_0(\mathbf{u}) = M_m G(\mathbf{u})$. We will make use of the following two function spaces.

$$C = \{\mathbf{u} \mid \mathbf{u}\colon [0, h] \times \mathbf{R} \longrightarrow \mathbf{R}^2, \text{ continuous, } 2\pi\text{-periodic in the second variable for all } t \in [0, h]\}. \tag{5.3}$$

$$C_{d,k} = \left\{\mathbf{u} \mid \mathbf{u}(t, \tau) = \sum_{|p| \leqslant d} e^{ip\tau} u_p^h(t),\ u_p^h \text{ a polynomial of degree} \leqslant k\right\}. \tag{5.4}$$

In particular, $\mathbf{u}_{d,k} \in C_{d,k} \subset C$. The Euclidean norm in $\mathbf{R}^2$ is (again) denoted by $|\cdot|$. Observe that $|\Phi(\tau)| = |\Phi^{-1}(\tau)| = 1$ in the corresponding matrix norm. In C we have the supremum norm $\|\cdot\|$,

$$\|\mathbf{u}\| = \sup_{(t,\tau) \in [0,h] \times \mathbf{R}} |\mathbf{u}(t, \tau)|. \tag{5.5}$$

An important part is played by the norm $|||\cdot|||$, which is defined on a subset of C by

$$|||\mathbf{u}||| = \max_{j=0,1,\ldots,k} \left\| \frac{\partial^j}{\partial t^j} \mathbf{u} \right\|. \tag{5.6}$$

Use of the norm $|||\mathbf{u}|||$ always implicitly presumes that the partial derivatives $\partial^j \mathbf{u}/\partial t^j$, $j \leqslant k$, are continuous, i.e. are elements of C.

If $\mathbf{v} = \mathbf{v}(t)$ rather than $\mathbf{v} = \mathbf{v}(t, \tau)$, we just interpret $\mathbf{v}$ as a function of (t, τ) (but one which is constant in the τ-direction for each t), and we use $\|\mathbf{v}\|$ and $|||\mathbf{v}|||$ with the obvious meaning.

The map G, cf. (1.7), is defined by

$$G(\mathbf{u})(t, \tau) = \Phi^{-1}(\tau) g(t, \Phi(\tau)\mathbf{u}(t, \tau)). \tag{5.7}$$

Thus, for smooth g, G is differentiable with respect to t and τ, and Fréchet differentiable with respect to $\mathbf{u}$. In particular, we have

$$\left(\frac{\partial}{\partial t}G\right)(\mathbf{u})(t,\tau) = \Phi^{-1}(\tau)g_t(t, \Phi(\tau)\mathbf{u}(t,\tau)) \tag{5.8}$$

with $g_t(t, x) = \partial g(t, x)/\partial t$. Also,

$$\left(\frac{\partial}{\partial u}G\right)(\mathbf{u})(t,\tau) = \Phi^{-1}(\tau)g_x(t, \Phi(\tau)\mathbf{u}(t,\tau))\Phi(\tau) \tag{5.9}$$

with $g_x(t, x) = \partial g(t, x)/\partial x$. Note that $(\partial G/\partial u)(\mathbf{u})(t, \tau)$ is a linear map from $\mathbf{R}^2 \longrightarrow \mathbf{R}^2$, $(\partial G/\partial u)(\mathbf{u})$ is a linear map from C into C, and $\partial G/\partial u$ is a map, not linear in general, from a suitable subset in C into the set of linear maps from C in C. Formulae similar to (5.8) and (5.9) are valid for the higher derivatives.

Henceforth it is assumed that Assumption 1.3 *holds true for some* $T \geqslant h_0 \geqslant h$.

We now begin the presentation of a sequence of technical results, which culminate in Theorem 5.14.

LEMMA 5.1. (i) $\partial/\partial t$ *is a bounded linear operator from* $C_{d,k}$ *into* $C_{d,k-1}$ *and* $\|\partial/\partial t\| = 2k^2/h$, $\forall d$. (*Cf.* (5.4).)

(ii) $\partial/\partial t$ *commutes with* H, M, $\Pi_{d,m}$ *and* M_m *on* $\{\mathbf{u} \mid \mathbf{u} \in C,\ \partial\mathbf{u}/\partial t \in C\}$. (*Cf.* (1.23), (1.24), (4.3) *and* (4.4).)

(iii) H *is a bounded linear operator from* C *into* C. *Moreover,* $\|H\| \leqslant 4\pi$.

(iv) M_m *is a bounded linear operator from* C *into* C. *Moreover,* $\|M_m\| = 1$.

(v) J_1 *is a bounded linear operator from* C *into* C. *Moreover,* $\|J_1\| = h$.

(vi) π_k, $\tilde{\pi}_{k-1}$ *are bounded linear operators from* C *into* C.

(vii) *All partial derivatives of* G *up to a certain order are bounded on* $\{\mathbf{u} \in C \mid \|\mathbf{u}\| \leqslant \|u_0^*\| + \delta\}$. (*Cf. Assumption* 1.3.)

Proof. In $C_{d,k}$, $\partial/\partial t$ operates on polynomials of degree $\leqslant k$. Thus $\partial/\partial t$ is bounded. The bound on its norm follows from the Markov-Bernstein inequality, which asserts that $|||f'||| \leqslant k^2|||f|||$, where f is a polynomial of degree $\leqslant k$. The norm $|||\cdot|||$ is the supremum norm on $[-1, 1]$. If f is the Chebyshev polynomial of the second kind of degree k, then the equality sign maintains. (Cf. Rivlin [17, p. 105].) This proves (i). The commutativity of $\partial/\partial t$ with $\Pi_{d,m}$ and with M_m is obvious, because $\Pi_{d,m}\mathbf{u}$, and $M_m\mathbf{u}$ are trigonometric polynomials with coefficients which are linear combinations of the $\mathbf{u}(\tau_j, t)$, cf. (4.2), (4.3), (4.4). The commutativity of $\partial/\partial t$ with H and with M follows, if $\partial/\partial t$ and J_2 commute. However, for all $\mathbf{u} \in C$ with $\|\partial\mathbf{u}/\partial t\| < \infty$, $\partial\mathbf{u}/\partial t \in C$, the definition of differentiability easily shows $\partial J_2/\partial t = J_2(\partial/\partial t)$. This proves (ii). (iii) and (iv) follow easily from the definition of H and of M_m, respectively, and (v) is an immediate consequence of the interval length h in the t-direction. (vi) is a consequence of the Lagrange interpolation formula. Finally, (vii) is just a transcription of Assumption 1.3(ii) into the G-terminology. □

Remark 5.2. $\|\pi_k\| \geqslant O(\log k)$, cf. Schönhage [19, p. 125]. Thus, $\|\pi_k\|$ is not uniformly bounded in k, $k \longrightarrow \infty$. This is probably also true for $\|\tilde{\pi}_{k-1}\|$.

LEMMA 5.3. $\|H\Pi_{d,m}\|$ *is bounded uniformly in* d *and* m, $m \geqslant 2d + 1$.

Proof. The arguments are very similar to the ones used in the proof of the Erdös-Turan Theorem, cf. Cheney [3]. We show that $\Pi_{d,m}$ is bounded uniformly in d, m from C into a larger space; then we use the fact that H is a bounded linear operator from this larger space back into C.

Let $\Gamma = \{p \mid p \in \mathbf{Z}, -((m-1)/2) \leqslant p \leqslant (m/2)\}$, and let

$$S = \left\{ \mathbf{v} \mid \mathbf{v}(\tau) = \sum_{p \in \Gamma} e^{ip\tau} a_p,\ a_p \in \mathbf{C} \right\}.$$

Let $\phi_p \in S$ be defined by $\phi_p(\tau_j) = \delta_{pj}$, for $0 \leqslant p, j \leqslant m-1$ and with $\tau_j = 2\pi j/m$. We introduce the operator Π_m as follows:

$$(\Pi_m \mathbf{u})(t, \tau) = \sum_{p=0}^{m-1} \mathbf{u}(t, \tau_p)\phi_p(\tau), \qquad t \in [0, h],\ \forall \tau.$$

Thus, $(\Pi_m \mathbf{u})(t, \cdot\,)$ is the trigonometric interpolation polynomial of $\mathbf{u}(t, \cdot\,)$ on the abscissae τ_j. Observe that

$$(\Pi_m \mathbf{u})(t, \tau) = \sum_{p \in \Gamma} e^{ip\tau} \tilde{u}_p(t).$$

Thus, since $\{p \mid p \in \mathbf{Z}, |p| \leqslant d\} \subset \Gamma$,

$$\int_0^{2\pi} |(\Pi_{d,m} \mathbf{u})(t, \tau)|^2 \, dt \leqslant \int_0^{2\pi} |(\Pi_m \mathbf{u})(t, \tau)|^2 \, dt.$$

Now we observe that

$$\int_0^{2\pi} \phi_p(\tau)\bar{\phi}_q(\tau) \, d\tau = \frac{1}{m}\delta_{pq}.$$

This is so, since the integral equals the discrete approximation

$$\frac{1}{m} \sum_{j=0}^{m-1} \phi_p(\tau_j)\bar{\phi}_q(\tau_j),$$

because $\phi_p, \phi_q \in S$. Consequently

$$\|(\Pi_{d,m} \mathbf{u})(t, \cdot\,)\|^2_{L^2(0,2\pi)}$$

$$\leqslant \|(\Pi_m \mathbf{u})(t, \cdot\,)\|^2_{L^2(0,2\pi)} \leqslant \frac{1}{m} \sum_{p=0}^{m-1} |\mathbf{u}(t, \tau_p)|^2 \leqslant \|\mathbf{u}\|^2.$$

Since H is bounded as a linear operator from $L^2(0, 2\pi)$ into $L^\infty(0, 2\pi)$, we finally obtain

$$\|H\Pi_{d,m} \mathbf{u}\| \leqslant \sup_{t \in (0,h)} \|(H\Pi_{d,m} \mathbf{u})(t, \cdot\,)\|$$

$$\leqslant \sup_{t \in (0,h)} \|H\|_{L^2 \to L^\infty} \|\mathbf{u}\| = \|H\|_{L^2 \to L^\infty} \|\mathbf{u}\|.$$

This concludes the proof of the lemma. □

Remark 5.4. It is known that $\|\Pi_m\| = O(\log m)$, cf. Schönhage [19, p. 127]. We expect a similar behavior for $\|\Pi_{d,m}\|$ if $2d + 1$ is close to m. Of course, the fact that $\|\Pi_m\| = O(\log m)$ is closely related to the bound $\|\pi_k\| \geqslant O(\log k)$, cf. Remark 5.2. □

In order to avoid the repetitious use of the same phrases, we will use c to denote a generic constant, not necessarily the same at each occurrence. The value of c may

depend on k, on the choice of the abscissae of π_k (on a unit interval), on δ (cf. Assumption 1.3(ii)) and on upper bounds (up to a certain order which depends on k) for the partial derivatives of G on the set $\{\mathbf{u} \mid \|\mathbf{u}\| \leqslant \|u_0^*\| + \delta\}$. Also, the value of c may depend on the upper bounds ϵ_0, h_0 for ϵ and h respectively. The value of c *never* depends on ϵ, h, d or m.

LEMMA 5.5. *Let* $\mathbf{u} \in C$, *and let* $G(\mathbf{u})$ *and* $G(\pi_k \Pi_m \mathbf{u})$ *be well defined. Then*

$$\pi_k \Pi_{d,m} G(\mathbf{u}) = \pi_k \Pi_{d,m} G(\pi_k \Pi_m \mathbf{u}),$$

for all k, d, m. (Π_m *is defined as in the proof of Lemma* 5.3 *above.*)

Proof. $\pi_k \Pi_{d,m} G(\mathbf{u})$ is determined by the values of $G(\mathbf{u})$ at the points (t_j, τ_p), with the t_j taken from among the abscissae $t_0, \ldots, t_k$ of π_k and with $\tau_p = 2\pi p/m$, $p = 0, 1, \ldots, m-1$. However, at these points we have $G(\mathbf{u}) = G(\pi_k \Pi_m \mathbf{u})$. □

LEMMA 5.6. *For all* $h > 0$,

(i) $\|\partial^j(\mathbf{u} - \pi_k \mathbf{u})/\partial t^j\| \leqslant ch^{m-j}\|\partial^m \mathbf{u}/\partial t^m\|$, $j = 0, 1, \ldots, m \leqslant k+1$,

(ii) $\|\partial^j(\mathbf{u} - \tilde{\pi}_{k-1} \mathbf{u})/\partial t^j\| \leqslant ch^{m-j}\|\partial^m \mathbf{u}/\partial t^m\|$, $j = 0, 1, \ldots, m \leqslant k$,

(iii) $\|\partial^j \pi_k \mathbf{u}/\partial t^j\| \leqslant c\|\partial^j \mathbf{u}/\partial t^j\|$, $j = 0, 1, \ldots, k+1$,

(iv) $\|\partial^j \tilde{\pi}_{k-1} \mathbf{u}/\partial t^j\| \leqslant c\|\partial^j \mathbf{u}/\partial t^j\|$, $j = 0, 1, \ldots, k$,

provided that the derivatives exist in C.

Proof. (i) and (ii) are an immediate consequence of a result of Ciarlet-Raviart (cf. [4], their Theorem 5, Section 4). Then

$$\left\| \frac{\partial^j}{\partial t^j} \pi_k \mathbf{u} \right\| \leqslant \left\| \frac{\partial^j}{\partial t^j} \mathbf{u} \right\| + \left\| \frac{\partial^j}{\partial t^j} (\pi_k \mathbf{u} - \mathbf{u}) \right\|.$$

By setting $m = j$ in (i) we obtain (iii). The proof of (iv) is similar. □

LEMMA 5.7. *Let* $\mathbf{u}, \mathbf{u} + \mathbf{w} \in \{\mathbf{u} \mid |||\mathbf{u}||| \leqslant |||u_0^*||| + \delta\}$. *Then for* $j = 0, 1, \ldots, k$ *we have*

$$\left\| \frac{\partial^j}{\partial t^j} \pi_k \{G(\mathbf{u} + \mathbf{w}) - G(\mathbf{u})\} \right\| \leqslant c \left\{ \|\mathbf{w}\| + \cdots + \left\| \frac{\partial^j}{\partial t^j} \mathbf{w} \right\| \right\}.$$

For $j = 0, 1, \ldots, k-1$, *we have*

$$\left\| \frac{\partial^j}{\partial t^j} \tilde{\pi}_{k-1} \{G(\mathbf{u} + \mathbf{w}) - G(\mathbf{u})\} \right\| \leqslant c \left\{ \|\mathbf{w}\| + \cdots + \left\| \frac{\partial^j}{\partial t^j} \mathbf{w} \right\| \right\}.$$

Proof. By Lemma 5.6(iii),

$$\left\| \frac{\partial^j}{\partial t^j} \pi_k \{G(\mathbf{u} + \mathbf{w}) - G(\mathbf{u})\} \right\| \leqslant c \left\| \frac{\partial^j}{\partial t^j} \{G(\mathbf{u} + \mathbf{w}) - G(\mathbf{u})\} \right\|.$$

Also note that

$$G(\mathbf{u} + \mathbf{w}) - G(\mathbf{u}) = \int_0^1 \left(\frac{\partial}{\partial u} G \right)(\mathbf{u} + s\mathbf{w})\mathbf{w}\, ds.$$

Because of the smoothness of G, we may interchange $\partial/\partial t$ and integration with respect to s. The set $\{\mathbf{u} \mid |||\mathbf{u}||| \leqslant |||u_0^*||| + \delta\}$ is convex; thus $\mathbf{u} + s\mathbf{w}$ belongs to this set for all $s \in [0, 1]$. Consequently because of Lemma 5.1(vii), we may write $(\partial^j/\partial t^j)(\partial G/\partial u) \cdot (\mathbf{u} + s\mathbf{w})\mathbf{w}$ as a linear combination of derivatives of $\mathbf{w}$, with coefficients which are derivatives of $\partial G/\partial u$, evaluated at $\mathbf{u} + s\mathbf{w}$ and at derivatives of $\mathbf{u} + s\mathbf{w}$. The highest deriva-

tive which occurs is of order $\leqslant j$. Thus, because of Lemma 5.1(vii), these coefficients are bounded uniformly in s. In fact, the bounds depend only on $|||u_0^*||| + \delta$ and on the bounds for certain derivatives of G. This proves the first inequality. The second inequality follows similarly from Lemma 5.6(iv). □

LEMMA 5.8. *Let $\tilde{\mathbf{u}} \in (I - M)C$, and let $|||\tilde{\mathbf{u}}||| \leqslant \delta/4$. Let η be a polynomial of degree $\leqslant k$ (in the variable t), and let $|||\eta||| \leqslant \delta/4$. Then there exists a constant h_0 such that for all $h \in (0, h_0]$, and all $m \geqslant 1$:*

(i) *the equation*

$$x = \xi + \eta + J_1 M_m G(\tilde{\mathbf{u}} + x)$$

has a unique solution x, $\|x\| \leqslant \|u_0^*\| + \delta/2$.

(ii) *the equation*

$$x^h = \xi + \eta + J_1 \tilde{\pi}_{k-1} M_m G(\tilde{\mathbf{u}} + x^h)$$

has a unique solution x^h, x^h a polynomial of degree $\leqslant k$. Also $\|x^h\| \leqslant \|u_0^*\| + \delta/2$.

(iii) *if $\eta = 0$, then $x - x^h$ obeys the following inequalities:*

$$\left\| \frac{\partial^j}{\partial t^j}(x - x^h) \right\| \leqslant ch^{k-j+1}, \qquad j = 0, 1, \ldots, k.$$

Proof. We use Picard iteration. G is Lipschitzian by Lemma 5.1(vii), $\|M_m\|$ is bounded, cf. Lemma 5.1(iv), and so is $\|\tilde{\pi}_{k-1}\|$, cf. Lemma 5.1(vi). It then follows easily that the Banach fixed point theorem applies, which proves (i) and (ii). Since the second equation is just a collocation type of approximation for the first one (both interpreted as differential equations), the results of Russell [18] apply. This leads to the estimates in (iii), but with constants which are m-dependent. Since $\|M_m\| \leqslant 1$, $\forall m$, the dependence of these constants on m is readily removed. This proves the third assertion. An independent proof, based on Lemma 5.6, is also possible. □

LEMMA 5.9. *Let $\tilde{\mathbf{v}}, \tilde{\mathbf{u}} \in (I - M)C$, and such that $|||\tilde{\mathbf{v}}|||, |||\tilde{\mathbf{u}}||| \leqslant \delta/4$. Let η_1, η_2 be polynomials of degree $\leqslant k$, and such that $|||\eta_1|||, |||\eta_2||| \leqslant \delta/4$. Let $h \in [0, h_0]$, with h_0 sufficiently small. Let $x^h, y^h \in \{x \mid \|x\| \leqslant \|u_0^*\| + \delta\}$ and let*

$$x^h = \xi + \eta_1 + J_1 \tilde{\pi}_{k-1} M_m G(\tilde{\mathbf{u}} + x^h),$$
$$y^h = \xi + \eta_2 + J_1 \tilde{\pi}_{k-1} M_m G(\tilde{\mathbf{v}} + y^h).$$

Then

$$\|x^h - y^h\| \leqslant (1 + ch)\|\eta_1 - \eta_2\| + ch\|\tilde{\mathbf{u}} - \tilde{\mathbf{v}}\|.$$

Moreover, for $j = 1, 2, \ldots, k$

$$\left\| \frac{\partial^j}{\partial t^j}(x^h - y^h) \right\| \leqslant c \sum_{p=0}^{j} \left\| \frac{\partial^p}{\partial t^p}(\eta_1 - \eta_2) \right\| + c \sum_{p=0}^{j-1} \left\| \frac{\partial^p}{\partial t^p}(\tilde{\mathbf{u}} - \tilde{\mathbf{v}}) \right\|.$$

Proof. Since Lemma 5.1(vii) applies, the results follow directly from the formula for $x^h - y^h$ and the fact that h is sufficiently small. □

Now let us introduce the map F from an open neighborhood of $0 \in C$ as follows: $F(u_0, \tilde{\mathbf{u}}) = x^h$ iff $x^h = \xi + u_0 + J_1 \tilde{\pi}_{k-1} M_m G(\tilde{\mathbf{u}} + x^h)$. Thus, the range of F consists of the polynomials of degree $\leqslant k$. Lemma 5.8 shows that F is well defined for all $\mathbf{u} \in C$ with $|||u_0||| \leqslant \delta/4$, $|||\tilde{\mathbf{u}}||| \leqslant \delta/4$. Also, Lemma 5.9 describes the Lipschitz character of F on this closed set.

By making use of F, the discretized equations (5.2a)–(5.2b) may be written equivalently as

$$\tilde{\mathbf{u}}_{d,k} = D_k \pi_k \Pi_{d,m} G(\mathbf{u}_{d,k}), \tag{5.10a}$$

$$u_0^h = F(-\tilde{\mathbf{u}}_{d,k}(0, 0), \tilde{\mathbf{u}}_{d,k}). \tag{5.10b}$$

In general, the right-hand side of these equations is *not a contraction* on a suitable domain. Therefore, we introduce the map F_0, defined by

$$F_0(\mathbf{u}) = F(-D_k \pi_k \Pi_{d,m} G(\mathbf{u})(0, 0), D_k \pi_k \Pi_{d,m} G(\mathbf{u})). \tag{5.11}$$

This map F_0 does have contraction properties, for $\epsilon > 0$ sufficiently small. This is the content of the next lemma.

LEMMA 5.10. *Let* $\mathbf{u}, \mathbf{u} + \mathbf{w} \in \{\mathbf{u} \mid |||\mathbf{u}||| \leqslant |||u_0^*||| + \delta, |||\tilde{\mathbf{u}}||| \leqslant \delta/4\}$. *Let* $\epsilon \in (0, \epsilon_0]$, $h \in (0, h_0]$, *with* ϵ_0, h_0 *sufficiently small. Then*

(i) F_0 *is well defined on* $\{\mathbf{u} \mid |||\mathbf{u}||| \leqslant |||u_0^*||| + \delta, |||\tilde{\mathbf{u}}||| \leqslant \delta/4\}$ *and* F_0 *maps this set into* $\{\mathbf{u} \mid |||\mathbf{u}||| \leqslant |||u_0^*||| + \delta/2, \tilde{\mathbf{u}} = 0\}$.

(ii) *For* $j = 0, 1, 2, \ldots, k$,

$$\left\| \frac{\partial^j}{\partial t^j} D_k \pi_k \Pi_{d,m} \{G(\mathbf{u} + \mathbf{w}) - G(\mathbf{u})\} \right\|$$

$$\leqslant c \left\{ \epsilon \|\mathbf{w}\| + \cdots + \epsilon \left\| \frac{\partial^j}{\partial t^j} \mathbf{w} \right\| + \epsilon^2 \left\| \frac{\partial^{j+1}}{\partial t^{j+1}} \mathbf{w} \right\| + \cdots + \epsilon^{k-j+1} \left\| \frac{\partial^k}{\partial t^k} \mathbf{w} \right\| \right\}.$$

For $j = 1, 2, \ldots, k$, *we have*

$$\left\| \frac{\partial^j}{\partial t^j} \{F_0(\mathbf{u} + \mathbf{w}) - F_0(\mathbf{u})\} \right\|$$

$$\leqslant c \left\{ \epsilon \|\mathbf{w}\| + \cdots + \epsilon \left\| \frac{\partial^{j-1}}{\partial t^{j-1}} \mathbf{w} \right\| + \epsilon^2 \left\| \frac{\partial^j}{\partial t^j} \mathbf{w} \right\| + \cdots + \epsilon^{k-j+2} \left\| \frac{\partial^k}{\partial t^k} \mathbf{w} \right\| \right\}.$$

For $j = 0$ *we have*

$$\|F_0(\mathbf{u} + \mathbf{w}) - F_0(\mathbf{u})\|$$

$$\leqslant c \left\{ \epsilon \|\mathbf{w}\| + \cdots + \epsilon^{p+1} \left\| \frac{\partial^p}{\partial t^p} \mathbf{w} \right\| + \cdots + \epsilon^{k+1} \left\| \frac{\partial^k}{\partial t^k} \mathbf{w} \right\| \right\}.$$

Proof. Because of Lemma 4.1, we have $\pi_k \Pi_{d,m} = \Pi_{d,m} \pi_k$. Because of Lemma 5.1(ii) and the definition of D_k (cf. (1.27)), we then obtain

$$\frac{\partial^j}{\partial t^j} D_k \pi_k \Pi_{d,m} = \epsilon \left\{ H\Pi_{d,m} \frac{\partial^j}{\partial t^j} + \cdots + (-\epsilon H)^{k-j} H\Pi_{d,m} \frac{\partial^k}{\partial t^k} \right\} \pi_k.$$

Now apply Lemma 5.3 (uniform boundedness of $H\Pi_{d,m}$), Lemma 5.1(iii) (boundedness of H), and the first estimate of (ii) follows immediately from Lemma 5.7.

This first estimate of (ii) is now used to prove (i). It follows from Lemma 5.8 that F_0 is well defined on the indicated set if $|D_k \pi_k \Pi_{d,m} G(\mathbf{u})(0, 0)| \leqslant \delta/4$ for all $\mathbf{u}$ in that set. Using the first estimate of (ii), we obtain

$$\begin{aligned} |D_k \pi_k \Pi_{d,m} G(\mathbf{u})(0, 0)| &\leqslant \|D_k \pi_k \Pi_{d,m} G(\mathbf{u})\| \\ &\leqslant \|D_k \pi_k \Pi_{d,m} \{G(\mathbf{u}) - G(0)\}\| + \|D_k \pi_k \Pi_{d,m} G(0)\| \\ &\leqslant c\epsilon + \|D_k \pi_k \Pi_{d,m} G(0)\|. \end{aligned}$$

Because of the smoothness of g, and thus G, and in particular by the smoothness of $G(0)$, it follows that $\|D_k \pi_k \Pi_{d,m} G(0)\| \leqslant c\epsilon$. Thus Lemma 5.8 applies, and thus F_0 is well defined on $\{\mathbf{u} \mid |||\mathbf{u}||| \leqslant |||u_0^*||| + \delta, |||\tilde{\mathbf{u}}||| \leqslant \delta/4\}$. The second set of estimates of (ii) follow directly from Lemma 5.9. The second assertion of (i) follows easily if we observe that $|F_0(\mathbf{u}) - \xi| \leqslant c + ch$. □

If $x \in \mathbf{R}^n$, $y \in \mathbf{R}^n$, we say that $x \leqslant y$ if $x_i \leqslant y_i$ for *all* coordinates x_i, y_i of x and y, respectively. If B and C are matrices with the same number of rows and columns, then we say $B \leqslant C$ if $B_{ij} \leqslant C_{ij}$ for *all* matrix coefficients B_{ij}, C_{ij} of B and C, respectively.

LEMMA 5.11. *Let $x, b \in \mathbf{R}^n$, and let L be an $n \times n$ matrix with nonnegative coefficients. Let $(I - L)^{-1}$ have nonnegative coefficients as well. Then $x \leqslant Lx + b$ implies $x \leqslant (I - L)^{-1} b$.*

Proof. Since the coefficients of $(I - L)^{-1}$ are nonnegative, applying $(I - L)^{-1}$ to both sides of the inequality $x \leqslant Lx + b$ yields the result. □

We apply this lemma in situations where

$$x = \left(\|\tilde{\mathbf{w}}\|, \ldots, \left\| \frac{\partial^k}{\partial t^k} \tilde{\mathbf{w}} \right\|, \|w_0\|, \ldots, \left\| \frac{\partial^{k'}}{\partial t^{k'}} w_0 \right\| \right)^T, \tag{5.12}$$

with $\mathbf{w}$ as in Lemma 5.10. Hence, the matrix L we require is given by

$$L = c \begin{pmatrix} B & B \\ D & D \end{pmatrix}, \tag{5.13}$$

$$B = \begin{pmatrix} \epsilon & \epsilon^2 & \cdots & \epsilon^{k+1} \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \epsilon^2 \\ \epsilon & \cdots & \cdots & \epsilon \end{pmatrix}, \quad D = \begin{pmatrix} \epsilon & \epsilon^2 & \cdots & \cdots & \epsilon^{k+1} \\ \epsilon & \epsilon^2 & \cdots & \cdots & \epsilon^{k+1} \\ \epsilon & \epsilon & \epsilon^2 & \cdots & \epsilon^k \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \epsilon & \epsilon & \cdots & \epsilon & \epsilon^2 \end{pmatrix}. \tag{5.14}$$

LEMMA 5.12. *Let L be given by* (5.13) *and* (5.14). *Then, for all $\epsilon \in (0, \epsilon_0]$, ϵ_0 sufficiently small, we have*

$$0 \leqslant (I - L)^{-1} \leqslant I + c\begin{pmatrix} B & B \\ B & B \end{pmatrix}.$$

Proof. The existence of $(I - L)^{-1}$ and the convergence of the series $I + L + L^2 + \cdots$ are obvious, since ϵ is small. This also shows $(I - L)^{-1} \geqslant 0$. Now

$$L \leqslant c\begin{pmatrix} B & B \\ B & B \end{pmatrix}.$$

It is easily verified that $B^2 \leqslant k\epsilon B$. Thus

$$L^2 \leqslant 2c^2\begin{pmatrix} B^2 & B^2 \\ B^2 & B^2 \end{pmatrix} \leqslant 2k\epsilon c^2\begin{pmatrix} B & B \\ B & B \end{pmatrix}.$$

Thus, for ϵ_0 sufficiently small, and all $\epsilon \in (0, \epsilon_0]$

$$L + L^2 + \cdots \leqslant \frac{1}{1 - 2kc^2\epsilon_0}\begin{pmatrix} B & B \\ B & B \end{pmatrix},$$

which proves the assertion. □

The results of Lemma 5.9–Lemma 5.12 are summarized in the following lemma:

LEMMA 5.13. *Assume*:

(i) $\mathbf{w}_{d,k}$ *is the solution of*

$$\tilde{\mathbf{w}}_{d,k} = D_k \pi_k \Pi_{d,m} G(\mathbf{w}_{d,k}) + \tilde{\eta},$$
$$w_0^h = \xi - \tilde{\mathbf{w}}_{d,k}(0, 0) + J_1 \tilde{\pi}_{k-1} M_m G(\mathbf{w}_{d,k}) + \eta_0,$$

with $\eta_0 \in MC_{d,k}$, $\tilde{\eta} \in (I - M)C_{d,k}$.

(ii) $\mathbf{u}_{d,k}$ *is the solution of the unperturbed equations* (5.2a)–(5.2b).

(iii) $\mathbf{u}_{d,k}, \mathbf{w}_{d,k} \in \{\mathbf{u} \mid |||\mathbf{u}||| \leqslant |||u_0^*||| + \delta, |||\tilde{\mathbf{u}}||| \leqslant \delta/4\}$.

(iv) ϵ_0, h_0, $|||\tilde{\eta}|||$ *and* $|||\eta_0|||$ *are sufficiently small.*

Let

$$x = \left(\|\tilde{\mathbf{w}}_{d,k} - \tilde{\mathbf{u}}_{d,k}\|, \ldots, \left\|\frac{\partial^k}{\partial t^k}(\tilde{\mathbf{w}}_{d,k} - \tilde{\mathbf{u}}_{d,k})\right\|\right)^T,$$

$$y = \left(\|w_0^h - u_0^h\|, \ldots, \left\|\frac{\partial^k}{\partial t^k}(w_0^h - u_0^h)\right\|\right)^T,$$

$$\chi = \left(\|\tilde{\eta}\|, \ldots, \left\|\frac{\partial^k}{\partial t^k}\tilde{\eta}\right\|\right)^T,$$

$$\psi = \left(\|\eta_0\|, \ldots, \left\|\frac{\partial^k}{\partial t^k}\eta_0\right\|\right)^T.$$

Also, let U and V be the $(k + 1) \times (k + 1)$ matrices,

$$U = c\begin{pmatrix} 1 & & & 0 \\ \cdot & \cdot & & \\ \cdot & & \cdot & \\ 1 & \cdot & \cdot & 1 \end{pmatrix}, \qquad V = c\begin{pmatrix} 1 & & & & \\ 1 & 0 & & 0 & \\ \cdot & \cdot & \cdot & & \\ \cdot & & \cdot & \cdot & \\ 1 & \cdot & \cdot & 1 & 0 \end{pmatrix},$$

respectively. Then, for all $\epsilon \in (0, \epsilon_0]$, $h \in (0, h_0]$ *and all m, d,*

$$\begin{pmatrix} x \\ y \end{pmatrix} \leqslant \begin{pmatrix} I + cB & cB \\ cB & I + cB \end{pmatrix} \begin{pmatrix} I & 0 \\ V & U \end{pmatrix} \begin{pmatrix} \chi \\ \psi \end{pmatrix},$$

where B is given by (5.14).

Proof. The proof of this lemma closely resembles the proof of Lemma 5.10. So we omit this proof. □

The statement of Theorem 5.14 below contains asymptotic bounds in ϵ and h, valid on the rectangle $(0, \epsilon_0] \times (0, h_0]$ in ϵ, h-space. We distinguish between $O(\epsilon^p h^q)$ and $O(\epsilon^s h^r)$ iff $(p - s)(q - r) < 0$. The term $O(\epsilon^p h^q)$ is incorporated into $O(\epsilon^s h^r)$ if $(p \geqslant s) \wedge (q \geqslant r)$. We make an exception for the terms $O(\epsilon^{k+1})$ and $O(\epsilon^{k+1}h)$, because their source is so different.

In the statement of Theorem 5.14 below, we encounter error-terms of the type

$$\left\| (I - \Pi_{d,m}) \frac{\partial^j}{\partial t^j} G([\mathbf{u}]_k) \right\|, \qquad j \leqslant k.$$

It is not so useful to give asymptotic bounds for these errors, especially if the functions, on which $I - \Pi_{d,m}$ operates, are of class C^∞. Therefore, we use the symbol $\eta_{d,m}$ to denote a common upper bound for those errors. Thus $\eta_{d,m} \longrightarrow 0$ for $d, m \longrightarrow \infty$. The rate of convergence of $\eta_{d,m}$ to zero depends on the smoothness of G, $[\mathbf{u}]_k$ etc. In particular, it depends on the smoothness, as a periodic function in τ, of $[\mathbf{u}]_k$, $G([\mathbf{u}]_k)$ etc. This may depend on ϵ, since the problems we are considering here have the tendency "to change the frequency" for increasing ϵ. However, since $\epsilon \in (0, \epsilon_0]$, we are always able to use an upper bound. Thus, $\eta_{d,m}$ is independent of ϵ. Also, $\eta_{d,m}$ is independent of h. It turns out that the dependence on h in an error term like $\|(I - \Pi_{d,m})\pi_k G([\mathbf{u}]_k)\|$ may be bounded from above for all $h \in (0, h_0]$. (This is done by Lemma 5.6.) Thus $\eta_{d,m}$ does not depend on ϵ and h. However it may, and will, depend on k.

The technical results just obtained will now be used to prove the following theorem (5.14) which characterizes properties of the smooth solutions exploited by the algorithm of Section 4.

THEOREM 5.14. *Assume*:

(i) *Assumption* 1.3 *holds.*

(ii) ϵ_0, h_0 *are sufficiently small in relation to k, to the choice of the abscissae of* π_k (*shifted to a unit interval*), *to* δ *as in Assumption* 1.3(ii), *and to the upper bounds for the partial derivatives of G up to a certain order* (*which depends on k*).

Then:

(a) *The equations* (5.2a)–(5.2b) *have a unique solution in the complete metric*

space $(S(\delta), \rho)$. *Here*

$$S(\delta) = \{\mathbf{u}_{d,k} \in C_{d,k} \mid |||\mathbf{u}_{d,k}||| \leqslant |||u_0^*||| + \delta,\ |||\tilde{\mathbf{u}}_{d,k}||| \leqslant \delta/4\},$$

and the metric ρ *is induced by* $|||\cdot|||$. *(Cf.* (5.6) *for* $|||\cdot|||$.)

(b) *If* $\mathbf{u}_{d,k} = \tilde{\mathbf{u}}_{d,k} + u_0^h$ *denotes the unique solution of* (5.2a)–(5.2b) *in* $S(\delta)$, *then*

$$\text{(b1)} \qquad \|u_0^h - [u_0]_k\| \leqslant c\{h^{k+1} + \epsilon^2 h^k + \cdots + \epsilon^{k+1}h\} + c\epsilon^{k+1} + \eta_{d,m}.$$

For $j = 1, 2, \ldots, k$,

$$\text{(b2)} \qquad \left\|\frac{\partial^j}{\partial t^j}(u_0^h - [u_0]_k)\right\| \leqslant c\{h^{k-j+1} + \epsilon^3 h^{k-j} + \cdots + \epsilon^{k-j+2}h\} + c\epsilon^{k+1} + \eta_{d,m}.$$

For $j = 0, 1, 2, \ldots, k$,

$$\text{(b3)} \qquad \left\|\frac{\partial^j}{\partial t^j}(\tilde{\mathbf{u}}_{d,k} - [\tilde{\mathbf{u}}]_k)\right\| \leqslant c\{\epsilon h^{k-j+1} + \cdots + \epsilon^{k-j+1}h\} + c\epsilon^{k+1} + \epsilon\eta_{d,m}.$$

Here c *is a generic constant, independent of* ϵ, h, d *and* m. $\eta_{d,m} \longrightarrow 0$ *for* $d, m \longrightarrow \infty$ *uniformly in* ϵ *and* h. *The rate of convergence of* $\eta_{d,m} \longrightarrow 0$ *is determined by the smoothness of* $[\mathbf{u}]_k$, $[\mathbf{u}]_{k+1}$ *as periodic functions in* τ.

Proof. Instead of (5.2a)–(5.2b) we consider the equivalent equations

$$\text{(5.15a)} \qquad \tilde{\mathbf{u}}_{d,k} = D_k \pi_k \Pi_{d,m} G(\tilde{\mathbf{u}}_{d,k} + u_0^h),$$

$$\text{(5.15b)} \qquad u_0^h = F_0(\tilde{\mathbf{u}}_{d,k} + u_0^h),$$

cf. (5.10), (5.11). The right-hand side of the above equations defines a map, which is well defined on $S(\delta)$. It follows from Lemma 5.10 that the image of $S(\delta)$ under this map is in $S(\delta)$, and it also follows from Lemma 5.10 that this map is Lipschitzian on $S(\delta)$ in the metric induced by $|||\cdot|||$. The Lipschitz constant is $O(\epsilon)$, uniformly in h, m, d, but not uniformly in k. Thus, with ϵ_0 sufficiently small, as in (ii) above, the Lipschitz constant is < 1. Thus, the Banach fixed point theorem applies. This proves assertion (a).

The error estimates, i.e. assertion (b), are obtained by comparing the solution $\mathbf{u}_{d,k}$ with a suitable approximation to $[\mathbf{u}]_k$. We choose $\mathbf{v}_{d,k} = \pi_k \Pi_{d,m}[\mathbf{u}]_k$ as this approximation. Insert this approximation into the equations (5.2a)–(5.2b). This yields a residual (local discretization error). Then we obtain bounds for the difference between $\mathbf{v}_{d,k}$ and the solution $\mathbf{u}_{d,k}$ by means of Lemma 5.13. Since bounds for the difference between $\mathbf{v}_{d,k}$ and $[u]_k$ are easily obtained, we thus get bounds for the difference between $\mathbf{u}_{d,k}$ (the discrete approximation) and $[\mathbf{u}]_k$.

So we must compute the residual η and we must estimate the difference between $\mathbf{v}_{d,k}$ and $[\mathbf{u}]_k$. Let us consider the difference $\mathbf{v}_{d,k} - [\mathbf{u}]_k$ first. We have

$$\begin{aligned} v_0^h - [u_0]_k &= \pi_k M_m[\mathbf{u}]_k - M[\mathbf{u}]_k \\ &= \pi_k(M_m - M)[\mathbf{u}]_k + (\pi_k - 1)[u_0]_k. \end{aligned}$$

The term $(\pi_k - 1)[u_0]_k$ is easily estimated through the use of Lemma 5.6. The term $\pi_k(M_m - M)[\mathbf{u}]_k$ may be estimated through the use of Lemma 5.6(iii). Also observe that $\partial/\partial t$ commutes with $M - M_m$ for $[\mathbf{u}]_k$ smooth (and $[\mathbf{u}]_k$ is smooth). Thus, for $j = 0, 1, \ldots, k$,

$$\left\|\frac{\partial^j}{\partial t^j}(v_0^h - [u_0]_k)\right\| \leqslant c\left\|(M_m - M)\frac{\partial^j}{\partial t^j}[\mathbf{u}]_k\right\| + ch^{k-j+1}.$$

In a similar manner we obtain for $j = 0, 1, \ldots, k$, that

$$\left\|\frac{\partial^j}{\partial t^j}(\tilde{\mathbf{v}}_{d,k} - [\tilde{\mathbf{u}}]_k)\right\|$$

$$\leqslant ch^{k-j+1}\left\|\Pi_{d,m}\frac{\partial^j}{\partial t^j}[\tilde{\mathbf{u}}]_k\right\| + c\left\|(I - \Pi_{d,m})\frac{\partial^j}{\partial t^j}[\tilde{\mathbf{u}}]_k\right\| + c\left\|(M - M_m)\frac{\partial^j}{\partial t^j}[\mathbf{u}]_k\right\|. \tag{5.16}$$

Since $[\tilde{\mathbf{u}}]_k$ is smooth, we may estimate $\|\Pi_{d,m}(\partial^j/\partial t^j)[\tilde{\mathbf{u}}]_k\|$ by a bound independent of d and m (this is not in conflict with Remark 5.4). $[\tilde{\mathbf{u}}]_k = O(\epsilon)$, (cf. (1.28a)) and so are its derivatives. Thus, the first two terms in the right-hand side of (5.16) are $h^{k-j+1}O(\epsilon) + \eta_{d,m}O(\epsilon)$. The last term in the right-hand side of (5.16) contains $[\mathbf{u}]_k$, and $[\mathbf{u}]_k = O(1)$. However, for $[\mathbf{u}]_k$ smooth, we know that $(M - M_m)(\partial^j/\partial t^j)[\mathbf{u}]_k$ is given by a series, the coefficients of which are taken from the Fourier expansion of $(\partial^j/\partial t^j)[\tilde{\mathbf{u}}]_k$. Thus,

$$\epsilon^{-1}\left\|(M - M_m)\frac{\partial^j}{\partial t^j}[\mathbf{u}]_k\right\| \longrightarrow 0 \quad \text{for } m \longrightarrow \infty.$$

Thus,

$$\epsilon^{-1}\left\|\frac{\partial^j}{\partial t^j}(\tilde{\mathbf{v}}_{d,k} - [\tilde{\mathbf{u}}]_k)\right\| = O(h^{k-j+1}) + \eta_{d,m} \quad \text{uniformly in } \epsilon \in (0, \epsilon_0].$$

This allows for estimates which are (probably) sharp with respect to their order in ϵ and in h.

The residual η has to be computed in much the same way. Use should be made of the above estimates, Theorem 1.4, Corollary 1.5, the expressions (5.1a)–(5.1b) for $[\mathbf{u}]_k$, Lemma 5.5 and Lemma 5.6. The details of this (tedious) computation are not especially illuminating. The result is: for $j = 0, 1, \ldots, k$,

$$\left\|\frac{\partial^j}{\partial t^j}\eta_0\right\| \leqslant ch^{k-j+1} + \eta_{d,m},$$

$$\left\|\frac{\partial^j}{\partial t^j}\tilde{\eta}\right\| \leqslant c\{\epsilon h^{k-j+1} + \cdots + \epsilon^{k-j+1}h\} + c\epsilon^{k+1} + \epsilon\eta_{d,m}.$$

Now apply Lemma 5.13, to conclude the proof of the theorem. □

Remark 5.15. The $O(\epsilon^{k+1})$-terms in statement (b) of Theorem 5.14 are due to the difference between D_k and D_{k-1} and to the occurrence of $G([\mathbf{u}]_{k-1})$ in (5.1a). The appearance of these $O(\epsilon^{k+1})$-terms is a consequence of comparing $\mathbf{u}_{d,k}$ to $[\mathbf{u}]_k$. One might ask whether the difference between $\Phi(t/\epsilon)\mathbf{u}_{d,k}(t, t/\epsilon)$ and $\tilde{x}(t)$, $\tilde{x}$ the exact solution of (1.1), contains an $O(\epsilon^{k+1})$-term. Numerical evidence suggests that the term

$O(\epsilon^{k+1})$ is also present in this difference. This would imply that there is no convergence, either to $[\mathbf{u}]_k$ or to $\tilde{u}$, for $h \to 0$, ϵ fixed, $m, d \to \infty$. □

It is well known that the error of numerical processes may be different (and better) in behavior, for $h \to 0$, at certain special points in the (t, τ)-space (*superconvergence*). Our approximation displays a similar effect. Let us assume that

$$\int_0^h f(t)\,dt - \int_0^h (\tilde{\pi}_{k-1} f)(t)\,dt = O(h^{s+1}), \qquad h \to 0,$$

for all sufficiently smooth f. Since

$$\int_0^h (\tilde{\pi}_{k-1} f)(t)\,dt = \int_0^h (\pi_k f)(t)\,dt$$

and since $\|f - \pi_k f\| = O(h^{k+1})$, $h \to 0$, it follows that $s \geqslant k+1$. We will characterize this situation by saying that π_k *defines a quadrature rule of order s.* The following theorem, giving a superconvergence result, is based on the results of Russell [18].

THEOREM 5.16. *If π_k defines a quadrature rule of order s, then we have*

$$|\mathbf{u}_{d,k}(h, 0) - [\mathbf{u}]_k(h, 0)|$$
$$\leqslant c\left\{h^{s+1} + h\|\tilde{\mathbf{u}}_{d,k} - [\tilde{\mathbf{u}}]_k\| + h\left\|\frac{\partial}{\partial t}\{\tilde{\mathbf{u}}_{d,k} - [\tilde{\mathbf{u}}]_k\}\right\| + h\eta_{d,m}\right\}.$$

Here c is a constant independent of ϵ, h, m and d.

Proof. Let x^h be the solution (cf. Lemma 5.8) of

$$x^h = \xi - [\tilde{\mathbf{u}}]_k(0, 0) + J_1 \tilde{\pi}_{k-1} g_0([\tilde{\mathbf{u}}]_k + x^h).$$

Then, $x^h(h) - [u_0]_k(h) = O(h^{s+1})$, see Russell [18]. We have

$$u_0^h = \xi - [\tilde{\mathbf{u}}]_k(0, 0) + J_1 \tilde{\pi}_{k-1} g_0([\tilde{\mathbf{u}}]_k + u_0^h) + \eta,$$

where η is a residual. We introduce the abbreviation $\tilde{\mathbf{z}} \equiv \tilde{\mathbf{u}}_{d,k} - [\tilde{\mathbf{u}}]_k$. Then it is easily seen that

$$\|\eta + \tilde{\mathbf{z}}(0, 0)\| \leqslant ch\|\tilde{\mathbf{z}}\| + h\eta_{d,m}.$$

Using Lemma 5.9 we get a bound for $\|u_0^h - x^h\|$. Here we employ the somewhat sharper bound

$$\|u_0^h - x^h - \eta\| \leqslant ch\|\eta\| \leqslant ch\|\tilde{\mathbf{z}}\| + h^2\eta_{d,m}.$$

Hence, upon collecting all these estimates, we get

$$\mathbf{u}_{d,k}(h, 0) - [\mathbf{u}]_k(h, 0)$$
$$= u_0^h(h) - [u_0]_k(h) + \tilde{\mathbf{z}}(h, 0)$$
$$= \{x^h(h) - [u_0]_k(h)\} + \{\tilde{\mathbf{z}}(h, 0) - \tilde{\mathbf{z}}(0, 0)\} + O(h\|\tilde{\mathbf{z}}\|) + O(h\eta_{d,m}).$$

This completes the proof of the theorem. □

Remark 5.17. If h is an integer multiple of $2\pi\epsilon$, then the periodicity in τ shows that there is superconvergence for $\mathbf{u}_{d,k}(h, h/\epsilon) - [\mathbf{u}]_k(h, h/\epsilon)$. This is how the result of Theorem 5.16 above will be made useful. □

We turn now to the question of *stability of the method.* To start with we consider the influence of a small perturbation ζ of the initial vector ξ. Thus, consider the equations

$$\widetilde{\mathbf{w}}_{d,k} = D_k \pi_k \Pi_{d,m} G(\mathbf{w}_{d,k}), \tag{5.17a}$$

$$w_0^h = \xi + \zeta - \widetilde{\mathbf{w}}_{d,k}(0, 0) + J_1 \widetilde{\pi}_{k-1} M_m G(\mathbf{w}_{d,k}). \tag{5.17b}$$

The result of the following lemma will be used to derive a stability statement; it will also be used to derive global error estimates.

LEMMA 5.18. *Let the assumptions of Theorem* 5.14 *hold. If* $|\zeta|$ *is sufficiently small, then the equations* (5.17a)–(5.17b) *have a unique solution* $\mathbf{w}_{d,k} \in S(\delta)$. *Moreover the estimates,*

$$\left\| \frac{\partial^j}{\partial t^j} (\widetilde{\mathbf{w}}_{d,k} - \widetilde{\mathbf{u}}_{d,k}) \right\| \leqslant \epsilon c |\zeta|,$$

$$\left\| \frac{\partial^j}{\partial t^j} (w_0^h - u_0^h) \right\| \leqslant c |\zeta|,$$

$$\|\mathbf{w}_{d,k}(0, h) - \mathbf{u}_{d,k}(0, h) - \zeta\| \leqslant ch|\zeta|,$$

$j = 0, 1, \ldots, k$, *with* c *a generic constant, independent of* ϵ, h, m, d, *are valid.*

Proof. Existence of a unique solution follows as in the proof of Theorem 5.14. The first two estimates are given by Lemma 5.13. The last estimate comes from

$$\mathbf{w}_{d,k}(h, 0) - \mathbf{u}_{d,k}(h, 0) = \int_0^h \frac{\partial}{\partial t} \{\mathbf{w}_{d,k}(t, 0) - \mathbf{u}_{d,k}(t, 0)\}\, dt + \zeta. \quad \square$$

The obvious conclusion of this lemma is the following Consequence 5.19.

Consequence 5.19. The self-starting method is stable with respect to small perturbations in the initial vector ξ, for all $\epsilon \in (0, \epsilon_0]$, $h \in (0, h_0]$, $\forall m, d$. $\square$

We are also able to obtain *global error* estimates. I.e., for the self-starting method applied on consecutive subintervals $(0, h)$, $(h, 2h)$, $\ldots \subset [0, T]$. E.g., on the subinterval $(h, 2h)$ the initial vector ξ is chosen as $\mathbf{u}_{d,k}(h, h/\epsilon)$, $\mathbf{u}_{d,k}$ being the approximation to $[\mathbf{u}]_k$ on $(0, h)$. Thus, if $h = 2\pi p\epsilon$, p integer, we may use Lemma 5.18 to describe the build-up of the global error at the nodes jh (in the uncoupled time t). This is a consequence of the periodicity in the τ-direction. Hence, with suitable assumptions on g we obtain the following Consequence 5.20.

Consequence 5.20. Let the self-starting method be applied on consecutive subintervals $(0, h)$, $(h, 2h)$, $(2h, 3h)$, $\ldots \subset [0, T]$. Let π_k define a quadrature rule of order s, $s \geqslant k + 1$. Then the global approximation error $e(t)$,

$$e(t) \equiv \widetilde{x}(t) - \Phi(t/\epsilon)\mathbf{u}_{d,k}(t, t/\epsilon)$$

is majorized by $(jh \leqslant T)$

$$|e(jh)| \leqslant ch^s + c\{\epsilon h^k + \cdots + \epsilon^k h\} + c\epsilon^{k+1} + \eta_{d,m},$$

$$\sup_{t \in (0, nh) \subset [0, T]} |e(t)| \leqslant ch^{k+1} + c\{\epsilon h^k + \cdots + \epsilon^k h + c\epsilon^{k+1}\} + \eta_{d,m}.$$

The generic constants c do not depend on ϵ, h, m, d, but they do depend on T and k. □

The proof is rather easy to obtain in view of previous results, and it will be omitted. However, we prefer to give a nonrigorous description of the error mechanism of the self-starting method.

Consider the subinterval $(0, h)$. Let ζ be a small perturbation in the initial vector ξ. The method determines envelopes on $(0, h)$, and those envelopes approximate corresponding envelopes of $[\mathbf{u}]_k$. At $t = 0$, each envelope, as determined by the method, approximates the corresponding envelope of $[\mathbf{u}]_k$, the error being dependent on ζ, but also on the parameters used in the method. This is the situation at $t = 0$ for each of the envelopes. These "initial" errors, at $t = 0$, in the envelopes are propagated (along the envelopes). During this propagation process a relative change of at most $O(h)$ may take place. So, the "initial" errors are encountered at $t = h$ multiplied by $1 + O(h)$. Also, in going from $t = 0$ to $t = h$, the error in each of the components picks up a local error. Hence, at $t = h$, the error in each envelope is $[1 + O(h)] \times$ {"initial" error}, plus a local error. The global error at $t = h$ is obtained by recombination of the envelope errors at $t = h$. If $h = 2\pi p\epsilon$, p integer, then the $[1 + O(h)] \times$ {"initial" error}'s recombine into the global error at $t = 0$, plus a perturbation of relative size $O(h)$. This is exactly the assertion of Lemma 5.18.

It is not so clear what happens if $h = 2\pi p\epsilon$, p not an integer. Rigorous, but straightforward, global error estimates are certainly not very sharp. This is quite fortunate, for such estimates contain terms like $\eta_{d,m}/h$. One expects that a certain amount of cancellation takes place, which reduces the linear build-up of local errors. But the extent to which this occurs is an open question.

In summary, these arguments show that the self-starting method goes through three stages: *decomposition, propagation along envelopes,* and *recombination.* The condition $h = 2\pi p\epsilon$, p integer, causes the decomposition and the recombination to take place in phase.

It is not so easy to give an analysis of the self-starting method, which is based directly on these ideas. In the first place, the situation is more complicated than sketched, because of the interaction between the envelopes during the propagation phase. Second, envelopes and Fourier series engender convergence questions. That is why we used a somewhat more abstract approach. However, the basic idea, outlined above, is preserved to some extent in treating the $MC_{d,k}$-component and the $(I - M)C_{d,k}$-component separately. This seems to be the important dichotomy.

We do not have a rigorous analysis for the algorithm based on the backward differentiation formulae (2.8). In fact a rigorous treatment of these methods is much more difficult, because these methods are not as directly related to $[\mathbf{u}]_k$ as is the self-starting method. However, if the above nonrigorous discussion has some sense in the multistep method case, then we see that in the multistep method situation only one stage occurs: propagation along the envelopes. No decomposition, nor recombination in the very strict sense of the self-starting method takes place. If this idea is correct, it follows that no condition of the type $h = 2\pi p\epsilon$, p integer, is necessary for a 'good'

global error behavior. For very simple linear problems and with the simplest backward differentiation formula (backward Euler method), it seems to be true that $h = 2\pi p\epsilon$, p integer, is not required. However, for those problems the number of envelopes is very small, and those simple examples are certainly not representative for the class of problems we want to deal with.

Finally, we consider the stability with respect to perturbations of the right-hand side. Usually, this type of stability is closely related to the stability with respect to a perturbation of the initial vector, cf. Consequence 5.19. Here the situation is completely different. Consider the perturbed equations (perturbation $\eta \in C_{d,k}$)

$$\widetilde{\mathbf{w}}_{d,k} = D_m \pi_k \Pi_{d,m} G(\mathbf{w}_{d,k}) + \widetilde{\eta}, \tag{5.18a}$$

$$w_0^h = \xi - \widetilde{\mathbf{w}}_{d,k}(0, 0) + J_1 \widetilde{\pi}_{k-1} M_m G(\mathbf{w}_{d,k}) + \eta_0 \tag{5.18b}$$

Estimates of the effect of these perturbations are given in the following lemma.

LEMMA 5.21. *Under the conditions of Theorem* 5.14, *let* $|||\widetilde{\eta}|||$, $|||\widetilde{\eta}_0|||$ *be sufficiently small. Then the equations* (5.18a)–(5.18b) *have a unique solution* $\mathbf{w}_{d,k} \in S(\delta)$, *and*

$$\|\widetilde{w}_{d,k} - \widetilde{\mathbf{u}}_{d,k}\| \leqslant c \left\{ \|\eta\| + \sum_{p=1}^{k} \epsilon^{p+1} \left\| \frac{\partial^p}{\partial t^p} \widetilde{\eta} \right\| + \sum_{p=0}^{k} \epsilon^{p+1} \left\| \frac{\partial^p}{\partial t^p} \eta_0 \right\| \right\},$$

$$\|w_0^h - u_0^h\| \leqslant c \left\{ \|\eta_0\| + \sum_{p=1}^{k} \epsilon^{p+1} \left\| \frac{\partial^p}{\partial t^p} \eta_0 \right\| + \sum_{p=0}^{k} \epsilon^{p+1} \left\| \frac{\partial^p}{\partial t^p} \widetilde{\eta} \right\| \right\}.$$

The (generic) constant c *does not depend on* ϵ, h, m, d.

Proof. Just apply Lemma 5.13. □

It is instructive to interpret this result. This is done by considering two different situations. First, we assume that $\widetilde{\eta}$, η_0 are smooth, i.e., their derivatives with respect to t are bounded uniformly in h, $h \in (0, h_0]$. Such perturbations may arise if, in actual computations, an approximation for g is used, the difference between g and its approximation being smooth. The obvious conclusion of Lemma 5.21 is that the method is stable with respect to smooth perturbations of the right-hand side, uniformly in ϵ and h for $\epsilon \in (0, \epsilon_0]$, $h \in (0, h_0]$. Second, we consider rough $\widetilde{\eta}$, η_0. These perturbations are characterized by the fact that $\|\partial\widetilde{\eta}/\partial t\|$, $\|\partial\eta_0/\partial t\|$ are realistically estimated by $2k^2\|\widetilde{\eta}\|/h$, $2k^2\|\eta_0\|/h$ respectively, cf. Lemma 5.1(i). Such rough perturbations may be caused by rounding errors due to the finite precision of the arithmetic used. Now the obvious conclusion of Lemma 5.21 is that the method is stable with respect to rough perturbations only if $2k^2\epsilon/h \leqslant c$, c a constant. These conclusions are summarized in the following Consequence 5.22.

Consequence 5.22. The self-starting method is stable with respect to smooth perturbations of the right-hand side, uniformly in ϵ, h for $\epsilon \in (0, \epsilon_0]$, $h \in (0, h_0]$ and uniformly in m, d. The method is stable with respect to rough perturbations of the right-hand side if $2\epsilon k^2/h \leqslant c$, for some constant c, and all m, d. □

We conclude this section with a remark concerning the initial value problem, cf. (1.1),

$$\frac{d}{dt}x = \frac{A}{\epsilon}x + g(t, x) + \frac{f(t)}{\epsilon}, \qquad x(0) = \xi. \tag{5.19}$$

The numerical example to be treated in Section 6 below involves a problem of this type.

If f is smooth, the substitution $x = y - A^{-1}f$ in (5.19) yields a problem of the type (1.1) in y. This shows the relation between (5.19) and (1.1).

We do not require a substitution of the above type if we seek to apply the self-starting method or the multistep method to (5.19) directly. However, theoretical results may be derived by making the substitution $x = y - A^{-1}\pi_k f$, and by applying the self-starting method to the problem in y (with $y = x + A^{-1}\pi_k f$). It is also possible to give a theoretical treatment without the explicit use of this transformation. It should be clear from the preceding sections and the analysis of this section how to proceed.

We just mention some results for the self-starting method applied to (5.19) directly. In general, the results of this section hold true for this problem as well, but there are a few small modifications in the formulation. E.g., the set $S(\delta)$ in Theorem 5.14 has to be changed to a closed sphere with center $A^{-1}f$ and radius greater than $|\xi + A^{-1}f(0)|$. In the error estimates of assertion (b) of Theorem 5.14, ϵ^p, $p > 0$, should be changed to ϵ^{p-1}. However, no term which is $O(1)$, $\epsilon \longrightarrow 0$, has to be changed to $O(1/\epsilon)$, $\epsilon \longrightarrow 0$! The global error estimates (cf. Consequence 5.20) now take the form

$$|e(jh)| \leqslant ch^{k+1} + c\{\epsilon h^k + \cdots + \epsilon^k h\} + c\epsilon^k + \eta_{d,m}, \tag{5.20}$$

$$\sup_{t\in(0,nh)} |e(t)| \leqslant ch^{k+1} + c\{\epsilon h^k + \cdots + \epsilon^k h\} + c\epsilon^k + \eta_{d,m}. \tag{5.21}$$

However, if $t = 0$ and $t = h$ are abscissae of π_k, then

$$|e(jh)| \leqslant ch^s + c\{\epsilon h^k + \cdots + \epsilon^k h\} + c\epsilon^k + \eta_{d,m}. \tag{5.22}$$

This follows because the $(\pi_k f - f)$-term now vanishes at the nodes jh. The estimate (5.21) does not change.

6. A Numerical Example. We apply the methods described in Section 4 to a simple problem. In order to have a problem for which the exact solution is known we construct an ordinary differential equation of the type (5.19) as follows:

The solution of the second order equation

$$\frac{d^2z}{dt^2} + \frac{z}{\epsilon^2} = \frac{e^{-t}}{\epsilon^2}, \qquad \epsilon > 0, \tag{6.1}$$

subject to the initial conditions

$$z(0) = 1 + \frac{1}{1+\epsilon^2}, \qquad z'(0) = -\frac{1}{1+\epsilon^2} \tag{6.2}$$

is given by

$$z(t) = \cos(t/\epsilon) + \frac{e^{-t}}{1+\epsilon^2}. \tag{6.3}$$

A nonlinear problem is obtained by introducing the new variable x,

$$z = x + \mu x^2, \qquad \mu > 0. \tag{6.4}$$

It follows that x and $y = \epsilon(d/dt)x$ satisfy the ordinary differential equation

$$\frac{d}{dt}\begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{\epsilon}\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \frac{\mu}{\epsilon}\begin{pmatrix} 0 \\ \dfrac{x^2 - 2y^2 - 2xe^{-t}}{1 + 2\mu x} \end{pmatrix} + \begin{pmatrix} 0 \\ \dfrac{e^{-t}}{\epsilon} \end{pmatrix}. \tag{6.5}$$

The initial conditions for x and y at $t = 0$ are easily computed by making use of (6.4). In particular, we have

$$x(t) = \frac{2z(t)}{1 + \sqrt{1 + 4\mu z(t)}}, \tag{6.6}$$

with z given by (6.3). We require that $x(t) \in \mathbf{R}$. Thus if we are interested in the interval $[0, T]$, we must restrict μ through

$$\mu \leqslant \left[4 \times \max_{t \in [0,T]} \{-z(t)\}\right]^{-1} \lesssim \frac{1}{4(1 - e^{-T})}. \tag{6.7}$$

We choose $T = 32\pi/100 \sim 1.005$. Thus we have $\mu \lesssim 0.4$.

For fixed ϵ, increasing μ increases the influence of the nonlinear term in (6.5). The influence of the nonlinear term determines the number of envelopes needed in order to achieve a specified accuracy. Thus, we expect that d (the number of envelopes is $2d + 1$) increases if μ increases, for fixed ϵ, and for the same accuracy.

As already observed in Section 5, the nonlinear term in (6.5) tends to "change the frequency" of the problem. It is well known that the odd terms in a Taylor expansion for $g(t, x)$ (expansion in x, about $x = 0$ for (1.1) and about $x = -A^{-1}f$ for (5.19)) are responsible for this frequency change. In this sense, the nonlinearity in (6.5) contains the small odd term $2\mu x$ in the denominator. The term $2xe^{-t}$ cancels in a Taylor series about $-A^{-1}f$. Thus the order of the odd terms in the nonlinearity in (6.5) is $O(\mu^2 x/\epsilon)$. If this argument is correct, then μ^2 is the parameter which determines the change in frequency. The methods of Section 4 deteriorate if a substantial change in frequency takes place. The number of envelopes required for even moderate accuracy becomes too large for all practical purposes.

The largest value of μ used in the examples below is $\mu = 0.3$. This corresponds to $\mu^2 \sim 0.1$, and this seems reasonable.

We employ three different methods in the examples to follow:

First Method (With Lobatto Points, $k = 1$). This method is the self-starting method used repeatedly on consecutive subintervals $(0, h)$, $(h, 2h)$, ... We choose $k = 1$. The projection π_k assigns to f its linear interpolation polynomial on the abscissae $t = 0$, $t = h$ (Lobatto abscissae for $k = 1$).

Second Method (With Lobatto Points, $k = 2$). This is the self-starting method used repeatedly on consecutive subintervals $(0, h)$, $(h, 2h)$, ... We choose $k = 2$. The projection π_k assigns to a function f its quadratic interpolation polynomial on the abscissae $t = 0$, $t = h/2$, $t = h$ (Lobatto abscissae for $k = 2$).

Third Method (*Multistep Method*). This is the method based on the backward differentiation formula

$$11x_n - 18x_{n-1} + 9x_{n-2} - 2x_{n-3} = 6hf_n \tag{6.8}$$

(standard multistep notation). We need starting values at $t = 0$, $t = h$ and $t = 2h$. These starting values are supplied by one step, stepsize $2h$, of the second method (with Lobatto points, $k = 2$). Then we obtain approximations at $t = 3h$, $t = 4h$, etc. by using (6.8).

In all examples we use the following conventions:

–The integration interval is $[0, 32\pi/100]$ $(32\pi/100 \sim 1.005)$.

–The values of m and d are kept constant over the integration interval.

–A fixed subinterval of length h is used for the methods with Lobatto points ($k = 1$, $k = 2$). For the multistep method, the starting values are obtained with subinterval length $2h$ (see above), and the multistep method itself is used with the constant stepsize h.

–We use a formulation of the methods involving $\Phi(\tau)$ and powers of $\Phi(\tau)$, cf. Section 3. Thus, we avoid complex arithmetic. The nonlinear equations are solved by a simple successive substitution process. This process corresponds directly to the one suggested in relation to (4.17). A similar process is used for the multistep method.

We perform only a finite number of successive substitution steps; consequently, we obtain approximations for the solutions of the discrete equations. Numerical experiments convince us of the accuracy of the results given in the tables and graphs. The error due to performing only a finite number of successive substitution steps is estimated to be typically less than 1% of the results given.

We consider global errors at the nodes as well as envelope errors. As in Consequence 5.20, let $e(t)$ denote the global error at the point t (t and τ coupled by $\tau = t/\epsilon$). Let $|\cdot|_1$ be the Hölder 1-norm. For the *methods with Lobatto points* we define the maximum error at the nodes, called E_{nod}, by

$$E_{\text{nod}} \equiv \max_{0 \leqslant j \leqslant n} |e(jh)|_1, \qquad nh = 32\pi/100. \tag{6.9}$$

For the multistep method this quantity is defined by

$$E_{\text{nod}} \equiv \max_{3 \leqslant j \leqslant n} |e(jh)|_1, \qquad nh = 32\pi/100. \tag{6.10}$$

Observe that the starting values are disregarded in E_{nod} for the multistep method.

The envelope errors are defined for the first coordinate of the solution of (6.5) only. This first coordinate of the solution is given by (6.6) and (6.3). Obviously, x may be written as $x = x(t, \tau)$, by replacing t/ϵ by τ in (6.6). Then, with convergent Fourier series,

$$x(t, \tau) = \sum_{p=0}^{\infty} a_p(t) \cos p\tau + \sum_{p=1}^{\infty} b_p(t) \sin p\tau. \tag{6.11}$$

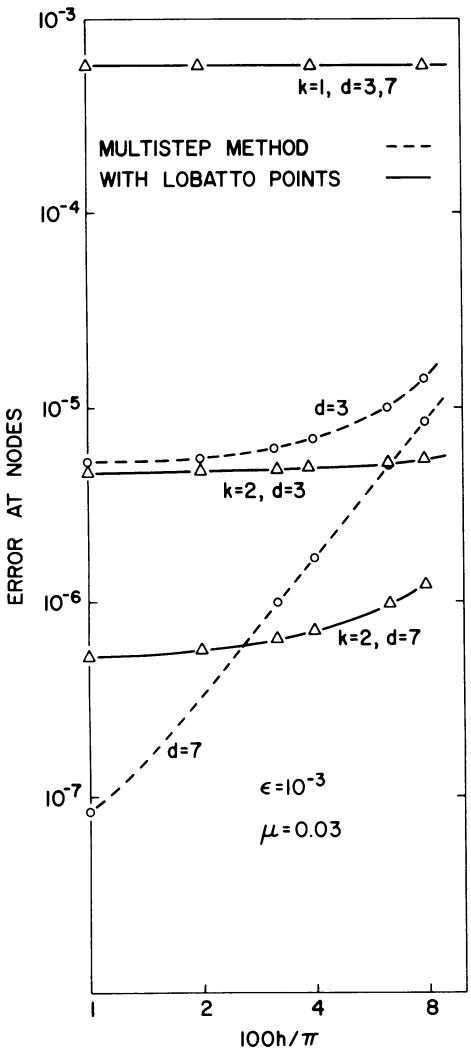

FIGURE 6-1. *Maximum error at the nodes for* (6.5)

We even have $b_p = 0$, $\forall p$. The discretization methods yield approximations $\tilde{a}_p$ for a_p, $0 \leqslant p \leqslant d$, and approximations $\tilde{b}_p$ for b_p, $1 \leqslant p \leqslant d$. For *the methods with Lobatto points the envelope error for* a_p is defined by

$$\|a_p - \tilde{a}_p\|_{L^\infty(0,32\pi/100)}. \tag{6.12}$$

For the multistep method the envelope error for a_p is defined by

$$\max_{3 \leqslant j \leqslant n} |a_p(jh) - \tilde{a}_p(jh)|, \qquad nh = 32\pi/100. \tag{6.13}$$

The envelope errors for b_p are defined similarly. Observe that the starting values are disregarded in the envelope errors for the multistep method.

The value of $a_p(t)$ is approximated, making use of a discrete Fourier transform on 64 points. Since $d \leqslant 15$ in all examples, this seems accurate enough. The $L^\infty(0, 32\pi/100)$-norm is estimated, using at least seven points per subinterval. Numerical

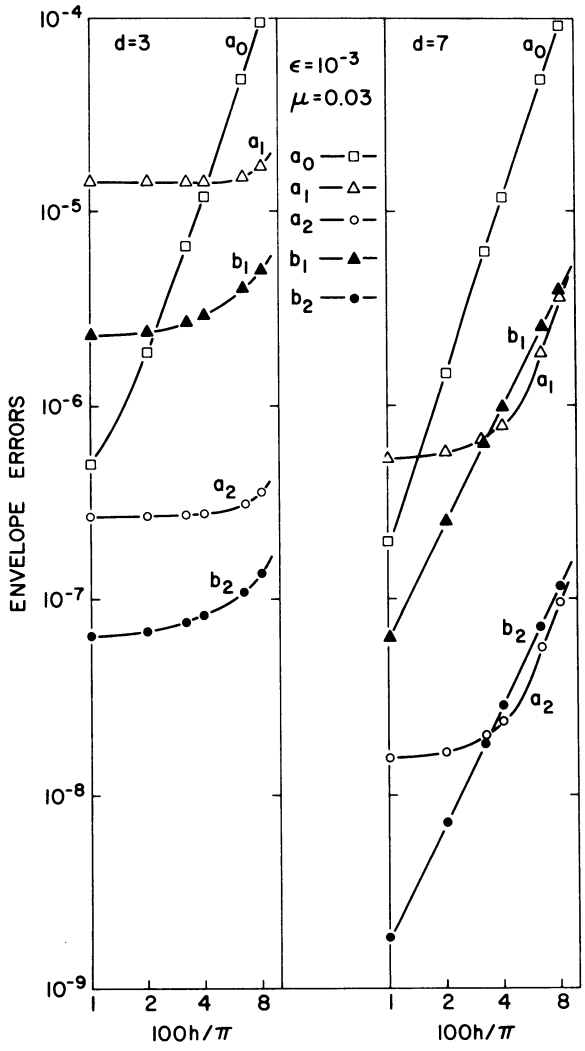

FIGURE 6-2. *Envelope errors for* (6.5)

experiments indicate that the error in estimating the $L^{\infty}(0, 32\pi/100)$-norm in this way is at most 4%.

In Figure 6-1 we display the maximum error at the nodes E_{nod} for the three methods in graphical form. The methods are applied to the problem (6.5) with $\epsilon = 0.001$, $\mu = 0.03$.

For the methods with Lobatto points, the length h of a subinterval ranges from $\pi/100 \sim 0.03$ to $8\pi/100 \sim 0.25$. For the multistep method we display the results as a function of the *equivalent stepsize.* The equivalent stepsize is defined as twice the actual stepsize. This is done because the method with Lobatto points ($k = 2$, subinterval length h) and the multistep method (stepsize $h/2$) use exactly the same set of points (abscissae) in t, τ-space (for equal d, m). Thus, the method with Lobatto points ($k = 2$, subinterval length h) and the multistep method (equivalent stepsize h) use the same amount of information about the differential equation.

The Lobatto points correspond to the abscissae of the trapezoidal rule ($k = 1$) and of Simpson's rule ($k = 2$). Thus, $s = 2$ for $k = 1$ and $s = 4$ for $k = 2$. (Cf. Theorem 5.16 for the definition of s.) Hence, we have the upperbounds, (cf. (6.9) and (5.22))

$$E_{\text{nod}} \leqslant \begin{cases} c(h^2 + \epsilon h) + c\epsilon + \eta_{d,m}, & k = 1, \\ c(h^4 + \epsilon h^2 + \epsilon^2 h) + c\epsilon^2 + \eta_{d,m}, & k = 2. \end{cases} \tag{6.14}$$

The graphs in Figure 6-1 illustrate this theoretical result.

For $k = 1$ we have $E_{\text{nod}} \sim 5.7E-4$ (with a 5% margin), for all h and d, m under consideration. Since $\epsilon = 10^{-3}$, we conclude that the term $O(\epsilon)$ is very much dominant in E_{nod} for $h \leqslant 1/4$ and all d, m, $d \geqslant 3$. For $k = 2$, the situation is more complicated. For h small, $h \lesssim 1/10$, the term $\eta_{d,m}$ (aliasing error) plays some part for $d = 3$, $m = 8$, while for $d = 7$, $m = 1$ the term $O(\epsilon^2)$ is dominant. The influence of the length h of a subinterval is (weakly) felt for $h \gtrsim 1/10$.

The error E_{nod} for the multistep method is roughly the same as E_{nod} for the method with Lobatto points ($k = 2$), for $d = 3$, $m = 8$. This is not so surprising: the aliasing error $\eta_{d,m}$ does not depend strongly on the discretization in the t-direction, provided that the resulting algorithm is stable. We have already seen that the aliasing error $\eta_{d,m}$ gives an important contribution to the error E_{nod} for the method with Lobatto points ($k = 2$) for $d = 3$, $m = 8$. So we expect the same, and even the same contribution, for the multistep method. This is confirmed by the graphs of Figure 6-1. Of course, this argument is made possible by the use of the equivalent stepsize.

For $d = 7$, $m = 16$ the behavior of E_{nod} for the multistep method and the behavior of E_{nod} for the method with Lobatto points ($k = 2$) differs considerably. A possible explanation is that the term $O(\epsilon^2)$ is not important (perhaps even absent) in E_{nod} for the multistep method. In this case the approximation properties which depend on h are not obscured by the term $O(\epsilon^2)$. The fact that the multistep method is not so strongly related to the asymptotic series as is the method with Lobatto points supports this explanation.

The error E_{nod} for the multistep method, $d = 7$, $m = 16$, in Figure 6-1 is approximately a straight line (there is some leveling off at $h \sim \pi/100$). Curiously enough, the slope of this line is 2.4 (in the figure the logarithmic scale in the horizontal direction is twice the one in the vertical direction). This implies that $E_{\text{nod}} \sim O(h^{2.4})$. We have no explanation for this behavior.

In Figure 6-2 we display, in graphical form, the envelope errors for the method with Lobatto points, $k = 2$. Cf. (6.12) for the envelope errors. We consider the envelope errors for a_0, a_1, a_2 and b_1, b_2. We observe that the envelope error admits the same bound as the supremum norm of the error in the total approximation. The supremum norm error for the total approximation is given by, cf. (5.21),

$$\sup_{t \in (0, 32\pi/100)} |e(t)| \leqslant ch^3 + c\{\epsilon h^2 + \epsilon^2 h\} + c\epsilon^2 + \eta_{d,m}. \tag{6.15}$$

		MULTISTEP METHOD			LOBATTO POINTS (k = 2)		
		h = 2π/100			h = 4π/100		
p	$\|\|a_p\|\|$	d = 3	d = 7	d = 15	d = 3	d = 7	d = 15
0	7.6(-1)	2.6(-3)	3.1(-6)	2.1(-6)	2.6(-3)	1.0(-5)	9.8(-6)
1	9.1(-1)	6.8(-3)	8.2(-6)	4.5(-6)	6.8(-3)	3.0(-5)	2.5(-5)
2	1.2(-1)	1.4(-2)	2.7(-5)	1.4(-6)	1.4(-2)	1.8(-5)	7.6(-6)
3	3.1(-2)	1.9(-2)	3.8(-5)	5.6(-7)	1.9(-2)	3.5(-5)	3.6(-6)
4	1.0(-2)		5.7(-5)	2.6(-7)		5.6(-5)	2.0(-6)
5	3.8(-3)		9.1(-5)	1.2(-7)		9.5(-5)	1.2(-6)
6	1.5(-3)		1.5(-4)	6.3(-8)		1.5(-4)	7.0(-7)
7	6.4(-4)		2.5(-4)	3.4(-8)		2.5(-4)	4.2(-7)
8	2.8(-4)			2.3(-8)			2.6(-7)
9	1.2(-4)			2.2(-8)			1.5(-7)
10	5.7(-5)			2.9(-8)			8.8(-8)
11	2.6(-5)			4.7(-8)			8.0(-8)
12	1.2(-5)			8.1(-8)			9.2(-8)
13	5.9(-6)			1.4(-7)			1.4(-7)
14	2.8(-6)			2.5(-7)			2.5(-7)
15	1.4(-6)			4.6(-7)			4.6(-7)
16	6.6(-7)						
17	3.2(-7)						

TABLE 6-1. *Envelope errors for* $\epsilon = 0.01$, $\mu = 0.3$

		MULTISTEP METHOD			LOBATTO POINTS (k = 2)		
		h = 2π/100			h = 4π/100		
p	$\|\|b_p\|\|$	d = 3	4 = 7	d = 15	d = 3	d = 7	d = 15
1	0	6.8(-2)	6.6(-5)	4.5(-6)	6.8(-2)	6.4(-5)	3.7(-6)
2	0	1.7(-2)	1.9(-5)	1.1(-6)	1.7(-2)	1.8(-5)	1.1(-6)
3	0	5.0(-3)	8.3(-6)	4.1(-7)	5.0(-3)	7.9(-6)	4.4(-7)
4	0		4.8(-6)	1.6(-7)		4.4(-6)	2.7(-7)
5	0		3.7(-6)	6.5(-8)		3.4(-6)	1.6(-7)
6	0		4.0(-6)	2.5(-8)		3.7(-6)	1.1(-7)
7	0		5.3(-6)	8.0(-9)		5.2(-6)	7.8(-8)
8	0			1.8(-9)			4.5(-8)
9	0			1.3(-10)			2.7(-8)
10	0			3.5(-10)			1.6(-8)
11	0			1.7(-10)			8.4(-9)
12	0			1.1(-9)			4.1(-9)
13	0			2.6(-9)			3.4(-9)
14	0			5.0(-9)			4.3(-9)
15	0			9.0(-9)			8.0(-9)
MAXIMUM ERROR AT THE NODES		6.5(-2)	4.0(-4)	6.3(-6)	6.4(-2)	3.8(-4)	1.4(-5)

TABLE 6-2. *Envelope errors, maximum error at the nodes,* $\epsilon = 0.01$, $\mu = 0.3$

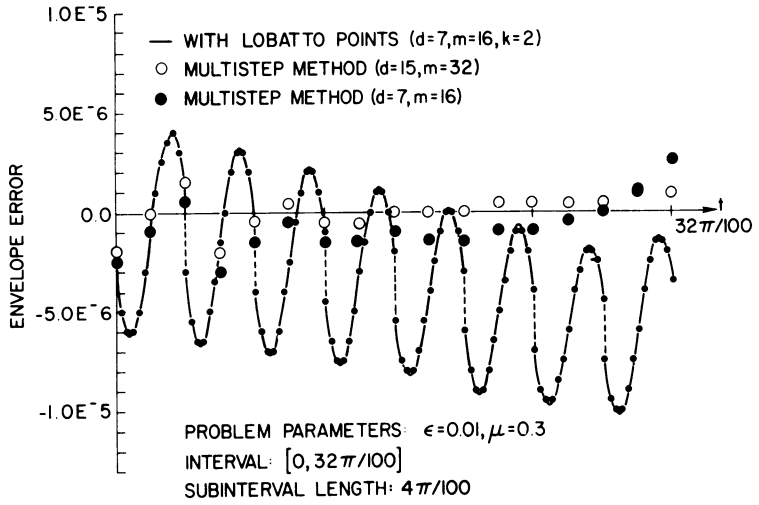

FIGURE 6-3. *Envelope error* $a_0 - \tilde{a}_0$ *for* (6.5)

Thus, we see that the aliasing error $\eta_{d,m}$ is the dominant error in the envelope errors for a_0, a_1, a_2, b_1, b_2 for $d = 3$, $m = 8$. In the other case, i.e., $d = 7$, $m = 16$, the envelope error for a_0 consists mainly of the term $O(h^3)$ (slope of the straight line for $\|a_0 - \tilde{a}_0\| \sim 3$ in Figure 6-2). The envelope errors for a_1 and a_2 are given by the term $O(\epsilon^2)$ (or perhaps even $O(\epsilon^3)$), for small h. But for $h \sim 1/4$ the term $O(h^3)$ makes itself felt. For $d = 7$, $m = 16$ the envelope errors $\|b_1 - \tilde{b}_1\|$ and $\|b_2 - \tilde{b}_2\|$ behave like $O(h^2)$ (slope of the corresponding lines in Figure 6-2 is approximately 2). We are inclined to believe that these errors represent the term $O(\epsilon h^2)$ in (6.15).

For the sake of completeness we note that the envelope errors for the multistep method (these errors are defined by (6.13)) for the problem parameters $\epsilon = 0.001$, $\mu = 0.03$ and for $d = 7$, $m = 16$ are: envelope error in a_0 is $O(h)$, envelope errors in a_1, a_2 are $O(h^3)$, envelope errors in b_1, b_2 are $O(h^2)$ (the corresponding curves are straight lines). This behavior is not so easily explained. It might be that these errors are mainly due to the errors in the starting values, even if the errors in the starting values themselves are not taken into account in the envelope error, cf. (6.13). In view of the results of Section 2, particularly Lemma 2.5, and in view of the very weak nonlinearity of the differential equation (6.5) (with $\epsilon = 0.001$, $\mu = 0.03$), it is implausible that the envelope errors for a_0, a_2, b_2 (these envelopes are obtained by the smooth solution concept) are due to the multistep method (6.8) itself. Indeed, the multistep method itself would yield an $O(h^4)$ behavior for these errors. Thus, the explanation in terms of errors in the starting values seems the more likely one. This strengthens our impression that *the multistep method is far superior to the self-starting method as a global method*; *the self-starting method should only be used to generate starting values.*

A more severe test for the methods is offered by the differential equation (6.5) with $\epsilon = 0.01$ and $\mu = 0.3$. In Table 6-1 we give the envelope errors for the multistep method (6.8) and for the method with Lobatto points ($k = 2$). The stepsize h (for the multistep method) as well as the subinterval length h (for the method with Lobatto

h	ERROR in $\|\cdot\|_1$-norm at $t = \frac{32\pi}{100}$		
$2\pi/1600$	1.01		
	3.2(-2)		
$2\pi/3200$	1.97(-1)	4.0(-3)	
	5.7(-3)		5.0(-5)
$2\pi/6400$	4.50(-2)	1.2(-5)	
	3.7(-4)		
$2\pi/12800$	1.10(-2)		

TABLE 6-3. *Richardson extrapolation table for trapezoidal rule,* $\epsilon = 0.001$, $\mu = 0.3$

points) are given in Table 6-1. The table also lists the values of d used. We have $m = 2d + 2$ in all cases. The $L^\infty(0, 32\pi/100)$-norm of the envelopes a_p is also given. Table 6-2 corresponds to Table 6-1 but yields the envelope errors for the b_p instead of the a_p. It also gives the errors E_{nod}.

As observed earlier, the multistep method and the method with Lobatto points ($k = 2$) have the same behavior if the aliasing error is dominant (cf. the case $d = 3$). Then from Table 6-1 and Table 6-2 it is seen that the aliasing error is dominant for $d = 3$, less dominant for $d = 7$, and the aliasing error is felt in the higher order envelopes only for $d = 15$. This is a strong experimental argument for the stability of these methods.

In Figure 6.3 we display the envelope error $a_0 - \tilde{a}_0$ itself, as a function of t, $t \in (0, 32\pi/100)$. The points are obtained through the lineprinter (20 lines above and below the horizontal axis). For the method with Lobatto points ($k = 2$) the error $a_0 - \tilde{a}_0$ is clearly discontinuous with jump discontinuities at the joins of the subintervals. Also, on each subinterval the error is well approximated by a third degree polynomial, as is to be expected. For $t \geqslant 3/4$, the envelope error $a_0 - \tilde{a}_0$ for the method with Lobatto points ($k = 2$) shows some weak signs of instability. This instability is due to the problem, more so than to the method. The exact solution of (6.5) with $\epsilon = 0.01$ and $\mu = 0.3$ does not exist for all $t \geqslant 0$; the solution exists on $[0, T)$ only, and in view of (6.6), (6.7) we have $T \sim 1.9$. Thus, the singularity in the solution itself makes itself felt for $t \geqslant 3/4$.

For $\epsilon = 0.01$, $\mu = 0.3$, we also solve the problem (6.5) by direct application of the trapezoidal rule. The trapezoidal rule is chosen because of its favorable stability properties for oscillatory problems. We use a constant stepsize h, and at each step we solve the nonlinear equations with a successive substitution process. We stop the successive substitution process if two consecutive iterates differ less than 10^{-6} (in the Hölder 1-norm). The result for $t = 32\pi/100$ is improved by Richardson extrapolation. The Hölder 1-norm of the approximation error at $t = 32\pi/100$ is listed in Table 6-3.

It would be desirable to obtain some insight into the efficiency of the methods of Section 4. It is clear that the behavior of the methods of Section 4 is almost independent of ϵ, $\epsilon \longrightarrow 0$. Thus, the methods of Section 4 are more efficient than any (classical) method provided that $\epsilon > 0$ is sufficiently small.

This is a relative statement about the efficiency of the methods of Section 4. At present, it is not possible to give more absolute information about their efficiency. However, it is clear that the methods of Section 4 solve a forced oscillation problem at each step, cf. Urabe [22]. Thus, the efficiency of these methods depends strongly on the efficiency of the process used in solving the forced oscillation problem. It is conceivable that a piecewise polynomial discretization, or a finite difference method, in the τ-direction leads to a more efficient algorithm.

Acknowledgement. The second author is grateful to IBM Corporation for giving him the opportunity to work for one year at the IBM T. J. Watson Research Center.

Mathematical Sciences Department
IBM Thomas J. Watson Research Center
Yorktown Heights, New York 10598

Wiskundig Seminarium
Vrije Universiteit
De Boelelaan 1081
Amsterdam, The Netherlands

1. V. AMDURSKY & A. ZIV, "On the numerical treatment of stiff highly-oscillatory systems," *SIAM J. Appl. Math.* (To appear.)

2. O. AXELSSON, "A class of A-stable methods," *BIT*, v. 9, 1969, pp. 185–199.

3. E. W. CHENEY, *Introduction to Approximation Theory*, McGraw-Hill, New York, 1966.

4. P. G. CIARLET & P. A. RAVIART, "General Lagrange and Hermite interpolation in R^n with applications to finite element methods," *Arch. Rational Mech. Anal.*, v. 46, 1972, pp. 177–199.

5. G. DAHLQUIST, "Numerical integration of ordinary differential equations," *Math. Scand.*, v. 4, 1956, pp. 33–50.

6. W. GAUTSCHI, "Numerical integration of ordinary differential equations based on trigonometric polynomials," *Numer. Math.*, v. 3, 1961, pp. 381–397.

7. C. W. GEAR, "The automatic integration of stiff ordinary differential equations," *Information Processing*, 68 (A. J. H. Morrel, Editor), North-Holland, Amsterdam, 1969, pp. 187–193.

8. C. W. GEAR, *Numerical Initial Value Problems in Ordinary Differential Equations*, Prentice-Hall, Englewood Cliffs, N. J., 1971.

9. P. HENRICI, *Discrete Variable Methods in Ordinary Differential Equations*, Wiley, New York, 1962.

10. F. HOPPENSTEADT & W. L. MIRANKER, *Numerical Methods for Stiff Systems of Differential Equations Related with Transistors, Tunnel Diodes, etc.*, Lecture Notes in Comput. Sci., vol. 10, Springer-Verlag, Berlin and New York, 1973, pp. 413–432.

11. F. HOPPENSTEADT & W. L. MIRANKER, "Differential equations having rapidly changing solutions," *J. Differential Equations*, v. 22, 1976, pp. 383–399.

12. I. KARASALO, "Minimum norm solutions of single stiff linear analytic differential equations," *J. Math. Anal. Appl.*, v. 51, 1975, pp. 516–538.

13. H. B. KELLER, "Numerical solution of boundary value problems for ordinary differential equations: Survey and some recent results on difference methods," in *Numerical Solutions of Boundary Value Problems for Ordinary Differential Equations* (A. K. Aziz, Editor), Academic Press, New York, 1975.

14. J. C. LAMBERT, *Computational Methods in Ordinary Differential Equations*, Wiley, London, 1973.

15. W. L. MIRANKER & G. WAHBA, "An averaging method for the stiff highly oscillatory problem," *Math. Comp.*, v. 30, 1976, pp. 383–399.

16. W. L. MIRANKER, M. van VELDHUIZEN & G. WAHBA, "Two methods for the stiff highly oscillatory problem," *Topics in Numerical Analysis,* III Conf., Dublin, 1976 (J. Miller, Editor), Academic Press, New York. (In preparation.)

17. T. J. RIVLIN, *The Chebyshev Polynomials,* Wiley, New York, 1974.

18. R. D. RUSSELL, "Collocation for systems of boundary value problems," *Numer. Math.,* v. 23, 1974, pp. 119–133.

19. A. SCHÖNHAGE, *Approximationstheorie,* de Gruyter, Berlin, 1971.

20. A. D. SNIDER & G. C. FLEMING, "Approximation by aliasing with applications to "certaine" stiff differential equations," *Math. Comp.,* v. 28, 1974, pp. 465–473.

21. E. L. STIEFEL & G. SCHEIFELE, *Linear and Regular Celestial Mechanics,* Springer-Verlag, Berlin, Heidelberg, New York, 1971.

22. M. URABE, "Galerkin's procedure for nonlinear periodic systems," *Arch. Rational Mech. Anal.,* v. 20, 1965, pp. 120–152.

23. R. WEISS, "The application of implicit Runge-Kutta methods to boundary value problems," *Math. Comp.,* v. 28, 1974, pp. 449–464.