

**STIFF SYSTEMS OF ORDINARY DIFFERENTIAL EQUATIONS.
II. BOUNDARY-VALUE PROBLEMS FOR
COMPLETELY STIFF SYSTEMS**

J. J. MAHONY and J. J. SHEPHERD

(Received 8 November 1979)

(Revised 2 March 1981)

Abstract

Solutions of the stiff system of linear differential equations

$$\epsilon \dot{\mathbf{x}} = A\mathbf{x} + \mathbf{r}$$

are obtained in a form yielding tight estimates of their properties, and conditions are obtained under which the operator norm of the map from \mathbf{r} to the solution \mathbf{x} does not become exponentially large for small values of ϵ . When these conditions are satisfied, the solutions are shown to be close to those of $A\mathbf{x} + \mathbf{r} = 0$, save at any singular points of A , and in boundary layers. The behaviour of solutions near admissible singular points is also obtained.

The results are used to characterize those boundary-value problems for the above system in which the solution defines maps from the data that are of “moderate” operator norm. This leads to a constructive existence theory for a limited class of boundary-value problems for the nonlinear system

$$\epsilon \dot{\mathbf{x}} = \mathbf{f}(t, \mathbf{x}, \epsilon).$$

It is suggested that the treatment of more general *classes* of boundary-value problems may be simplified using these results. By the use of simple examples, the problems involving large operator norms are shown to be related to the stability properties of the possible branches of the outer solutions close to those of

$$\mathbf{f}(t, \mathbf{x}, \epsilon) = 0.$$

1. Introduction

In a previous paper, [6], subsequently to be referred to as Part I, we have constructed a basis for the solution space of completely stiff homogeneous linear

systems of ordinary differential equations of the form

$$\epsilon \dot{x} = Ax, \quad (1.1)$$

under reasonably general assumptions about the matrix A and for suitably small, positive values of the parameter ϵ . In principle, once such a basis has been constructed, the method of variation of parameters may be used to construct all solutions of any inhomogeneous system

$$\epsilon \dot{x} = Ax + r, \quad (1.2)$$

for any given vector-valued function r .

The problem of solving any boundary-value problem associated with the system (1.2) is thus, in principle, reduced to that of obtaining answers to certain algebraic questions. However, as we shall see in the following discussion, a considerable number of the solutions obtained in this manner have properties that may be unacceptable within the context of many applications.

To illustrate the nature of the difficulty, we consider the very simple linear system

$$\begin{aligned} \epsilon \dot{x} &= -x, \\ \epsilon \dot{y} &= y, \end{aligned} \quad (1.3)$$

subject to the boundary conditions

$$\begin{aligned} y(0, \epsilon) &= 0, \\ bx(1, \epsilon) + y(1, \epsilon) &= c, \end{aligned} \quad (1.4)$$

when b and c are given real numbers independent of ϵ , with b non-zero. The unique solution to this boundary value problem is

$$\begin{aligned} x(t, \epsilon) &= (c/b) \exp((1-t)/\epsilon), \\ y(x) &= 0, \end{aligned} \quad (1.5)$$

and it is clear that, for any $\epsilon > 0$, the problem is well set in the sense of Hadamard, if natural norms (that is, uniform and scalar norms) are used. If, however, the objective was to evaluate $x(0, \epsilon)$ in terms of the parameters defining the problem then the operator norm of the map so defined becomes excessively large for small values of ϵ owing to the presence of the factor $\exp(1/\epsilon)$. In the context of many applications such a sensitive dependence on data may be unacceptable. It is interesting to note that, as illustrated by an example given by Eckhaus [3], such sensitivity may be associated with the failure of heuristic arguments. For stiff systems where errors of numerical integration will be of the order of a moderate power of the step length, an exponentially large factor in ϵ would have overwhelming effects. When problems exhibit such extreme sensitivity to the data defining them there is the possibility of obtaining "neighbouring" problems which have solutions with moderate operator norms.

Consider the system

$$\begin{aligned}\epsilon \dot{x} &= -x, \\ \epsilon \dot{y} &= y - k \exp(-1/\epsilon),\end{aligned}\tag{1.6}$$

where k is a parameter independent of ϵ . The unique solution of this system satisfying the boundary conditions (1.4) is

$$\begin{aligned}x &= b^{-1}\{c + k - k \exp(-1/\epsilon)\} \exp((1 - t)/\epsilon), \\ y &= -k \exp(-(1 - t)/\epsilon) + k \exp(-1/\epsilon).\end{aligned}$$

In general, the solution of this problem is subject to the same difficulties as the earlier example but there is a significant difference—it is possible to choose a value of k where the problem has a solution with moderate bounds. In the above example the choice $k = -c$ yields the solution

$$\begin{aligned}x &= (c/b) \exp(-t/\epsilon), \\ y &= c \exp(-(1 - t)/\epsilon) - c \exp(-1/\epsilon).\end{aligned}\tag{1.7}$$

Note that a choice of any $k + c$ of the form $O(1) \exp(-\epsilon^{-1})$ has a similar effect.

The significance of such a choice becomes apparent if we consider the nonlinear problem

$$\begin{aligned}\epsilon \dot{x} &= -x + \epsilon^2 f(x, y), \\ \epsilon \dot{y} &= y + \epsilon^2 g(x, y),\end{aligned}\tag{1.8}$$

with the same boundary conditions as previously. Suppose we were trying to validate some solution constructed by heuristic means for the above, where f and g are unbounded for x or y large. To choose the linear system (1.3) as a dominant for (1.8), while it might appear intuitively obvious, would lead to considerable difficulties, since (1.3) is not useful in the sense that it produces a solution $x(t, \epsilon)$ having supremum norm exponentially large as $\epsilon \rightarrow 0$. On the other hand, to choose the system (1.6) as such a dominant would remove this difficulty, and, provided $-c \exp(-\epsilon^{-1})$ lies in the range of g , would appear to offer the possibility of a contraction mapping proof of the existence of a solution of (1.8) close to (1.7).

The above examples are undoubtedly contrived but they serve to illustrate an important point. If one is attempting to validate an “approximation”, whether it be obtained by formal heuristics or numerical integration, for a boundary value problem involving a stiff nonlinear system, the choice of the linear dominant system can be of vital importance. Moreover, a satisfactory choice of a linear dominant cannot be considered in isolation from the boundary conditions. It is apparent therefore that one would like to distinguish, for an arbitrary linear boundary value problem, those problems, if any, which do not have over-sensitive dependence on the data. This paper has as its primary objective an examination of the properties of linear systems which are significant in any

attempt to characterize the 'nice' problems. As we are interested in the generation of useful approximations to solutions of nonlinear systems we are also concerned with the implications when no suitable linear dominant is available.

The first step is to develop suitable criteria as to what constitutes "over-sensitive dependence on data". The Hadamard concept of well-posed problems is the standard means of attempting to distinguish problems which do not have this undesirable feature. The examples already cited show that this is not adequate for our present purposes. A multiplier, exponentially large in ϵ , applied to the data is unsatisfactory for our purposes. We therefore introduce the idea of a *reasonably posed* problem, in which the solution is not only continuously dependent on the data, but the norm of the associated operator from data to solution is "not too large". Correspondingly, some problems will be *unreasonably posed*.

Clearly such a classification can have meaning only once norms have been defined and a meaning given to "not too large". We would argue strongly that the question, as to what is an appropriate definition and an appropriate meaning, should be resolved within the context of any specific application under consideration. To give specificity to our discussions we have settled on supremum norms and consider algebraically large terms like $\epsilon^{-\mu}$ for some positive μ as acceptable but not exponentially large terms such as $\exp(\epsilon^{-1})$. This choice is suggested by the results which have already been established in the literature, and it is also likely that it will be useful in the context of numerical solutions. We would emphasize that the thrust of our results would not be vitally affected by a more pragmatic choice.

These simple examples serve to explain the structure of this paper. We begin in Section 2 by constructing a particular integral of (1.2), in a form which gives us maximal control over the behaviour of all solutions, when only limited information is available on the function \mathbf{r} . This enables us to characterize, for a given homogeneous linear system (1.1), those functions \mathbf{r} for which the corresponding solutions of (1.2) depend *reasonably* on \mathbf{r} . In Section 3, we examine the structure of these reasonable solutions, and relate them to the results for the outer solution for (1.2), obtained by standard heuristic methods. Thus we are able to establish a set of sufficient conditions under which the heuristic arguments used to construct an outer solution lead to a reliable answer.

In Section 4 we consider the question of obtaining solutions of boundary-value problems for the system (1.2) when the solutions of such a system depend reasonably on \mathbf{r} . This leads us to a characterization of those boundary-value problems which depend reasonably on the boundary data. The results seem to emphasize that there is a very strong connection between the reasonable boundary conditions and the structural properties of the homogeneous system (1.1). Boundary conditions that lead to reasonable solutions constitute a small

subset of all boundary conditions which might be imposed. Such conclusions have significance in any attempt to construct a robust numerical scheme for the solution of boundary-value problems for stiff systems.

In Section 5 we turn our attention to methods for generating reasonable linear problems, and show how this leads to questions about the solvability of certain algebraic equations. In Section 6 we look briefly at the implications of this in the context of nonlinear singular perturbation theory. There, through simple examples of first-order systems, we see that the results are natural and are related to the selection of branches of possible outer solutions for stiff nonlinear systems. It would appear that the results described here may provide stability criteria for possible branches of outer solutions and serve to locate “interior layers”. It is anticipated that the results will permit the discussion of each boundary or interior layer in isolation of any other. They may also be of advantage in helping to generate useful numerical methods.

There is one aspect of the method of presentation, for the remaining two parts of this sequence of papers, which differs markedly from Part I. In that paper, because we recognized that numerical solutions would be available only for isolated values of ϵ , the theorems were stated in a manner free of the usual order notation of asymptotic analysis. The complications, which would be involved in maintaining such a presentation, become even more marked in the following work. Thus we revert to the standard terminology of asymptotic analysis which means that the stated assumptions cannot be satisfied by numerically derived approximations. However it would be possible to restate the assumptions and results, using the approach in Part I, so that the material is relevant where information is known only for a finite number of values of ϵ .

2. Construction of particular integrals

It is the purpose of this section to construct particular integrals for the completely stiff inhomogeneous linear system

$$\epsilon \dot{\mathbf{x}} = A(t, \epsilon)\mathbf{x} + \mathbf{r}(t, \epsilon), \quad 0 < t < T, \quad (2.1)$$

for a given $n \times n$ matrix function A and given n -vector function \mathbf{r} . Such systems are an intrinsic part of proof techniques that use the Contraction Mapping Theorem to validate trial solutions to nonlinear problems, obtained either analytically or numerically. Because the information supplied for such proofs is usually limited to bounds on the relevant functions, we will seek those forms for the particular integrals of (2.1) that will give us the maximum control over the behaviour of the general solutions of (2.1), while using only this limited data.

We note in passing that our methods could be applied to systems where supremum norms were available for functions of the form $\mathbf{r}(t, \epsilon) \exp(\epsilon^{-1}\alpha(t, \epsilon))$, with somewhat different results. An example to be considered in Section 6 will demonstrate that there may also be a general class of problems where this alternative approach may be significant.

We will apply the method of variation of parameters to construct a family of such particular integrals, and then will make an optimal choice within this family. The construction depends upon the properties of a fundamental matrix for the homogeneous system associated with (2.1) and to deal with this we will make considerable use of the basis for the solution space of this system that was constructed in Part I. The assumptions made about A in that reference will be adopted here also, and the reader is referred to that paper for precise details of these. Briefly, however, the thrust of these assumptions is that $A(t, \epsilon)$ may be diagonalized for all (t, ϵ) , with respect to a continuously differentiable basis of eigenvectors, for which estimates hold that are independent of ϵ in a neighbourhood of $\epsilon = 0$.

By applying these assumptions, we showed in Part I that, on any interval (which could be a subinterval of $[0, T]$, or the full interval $[0, T]$) on which an unchanging ordering may be defined on the real parts of the eigenvalue functions of $\lambda_i(t, \epsilon)$ of A , there exists a basis $\{x_i(t, \epsilon)\}_{i=1}^n$ for the solution space of the homogeneous system associated with (2.1) that takes the particularly simple form

$$x_i(t, \epsilon) = X_i(t, \epsilon) \exp(\zeta_i(t, \epsilon)), \quad i = 1, \dots, n, \quad (2.2)$$

where

$$\zeta_i(t, \epsilon) = \epsilon^{-1} \int_0^t \lambda_i(s, \epsilon) ds, \quad (2.3)$$

and

$$X_i(t, \epsilon) = s_i(t, \epsilon) + v_i(t, \epsilon). \quad (2.4)$$

Here, $s_i(t, \epsilon)$ is a suitably chosen eigenvector function for A , corresponding to the eigenvalue function $\lambda_i(t, \epsilon)$, while $v_i(t, \epsilon)$ is a vector function that tends to zero with ϵ , uniformly in t , on the relevant interval. In Part I we showed that X_i was the solution of an appropriately chosen integral equation (whose form changes for each value of i [6, Section 5]), and that as a consequence of the structure of this equation, $v_i(t, \epsilon)$ is continuously differentiable with respect to t for each $\epsilon > 0$. Further, successively closer approximations to v_i could be constructed by the use of a suitably defined iteration scheme. We also showed that, when the ordering described above changed at a finite number of interior points of $[0, T]$, a basis analogous to the above could be constructed, on each

subinterval where this ordering was unchanged, for which the analogues of the \mathbf{X}_i had the property (2.4), and we derived a transformation rule to continue from subinterval to subinterval.

If we define the matrix X by

$$X = [\mathbf{X}_1 \mathbf{X}_2 \dots \mathbf{X}_n], \quad (2.5)$$

and Z by

$$Z = \text{diag}(\exp \zeta_1, \exp \zeta_2, \dots, \exp \zeta_n), \quad (2.6)$$

we may apply the method of variation of parameters to the system (2.1), using XZ as a fundamental matrix, to obtain the representation

$$\mathbf{x}(t, \epsilon) = X \left\{ Z\mathbf{c} + \epsilon^{-1} Z \int' Z^{-1} X^{-1} \mathbf{r} \right\}, \quad (2.7)$$

for any solution $\mathbf{x}(t, \epsilon)$ of (2.1) on $[0, T]$, where the form (2.2) holds there. Here, \mathbf{c} is a constant n -vector (that may vary with ϵ). When changes in the ordering of eigenvalues occur at interior points of $[0, T]$, a form analogous to (2.7) holds in each of the respective subintervals. The lower terminal of the integral in (2.7) may be chosen arbitrarily so that, in this sense, the second term of (2.7) constitutes a family of particular integrals of (2.1).

We may note that the i th component of the vector in braces in (2.7) takes the form

$$c_i \exp \zeta_i + \epsilon^{-1} \exp \zeta_i \int' \exp(-\zeta_i) R_i, \quad (2.8)$$

where the vector \mathbf{R} is defined by

$$\mathbf{R} = X^{-1} \mathbf{r}. \quad (2.9)$$

We may exercise our freedom to choose the lower terminal of integration in (2.8), not only for various vectors \mathbf{c} , but also for each component c_i of such a vector. The exploitation of this freedom amounts, in fact, to a suitable redefinition of the constant c_i for each component of (2.8). In what follows, we aim to use this freedom to achieve the best estimates for the particular integral defined by the second term of (2.7), given limited estimates for the function \mathbf{r} .

At this point we digress to consider the types of integrals that may arise in such a particular integral. Thus we will concern ourselves with the behaviour, for small real positive values of the parameter ϵ , of integrals $I(\epsilon)$ of the form

$$I(\epsilon) = \int_a^b f(t, \epsilon) \exp(\epsilon^{-1} \{ \mu(t) + i\nu(t) \}) dt, \quad (2.10)$$

where f is uniformly bounded on $[a, b]$, independently of ϵ , and μ and ν are real-valued functions. It is apparent that the results of this section may be extended readily to the case where μ and ν depend on ϵ in a regular manner. If μ is not constant on $[a, b]$, and takes its maximum value at $t_\alpha \in [a, b]$, we may

apply extended forms of the Laplace estimation techniques described in Erdélyi [4, Section 2.2] to obtain the estimate

$$|I| \leq K_1 \varepsilon^\alpha \exp\{\varepsilon^{-1}\mu(t_\alpha)\} \|f\|, \tag{2.11}$$

where K_1 and α are positive numbers independent of ε , with α dependent on the nature of the maximum of $\mu(t)$, while $\|\cdot\|$ is the usual supremum norm for continuous functions, taken with respect to t on $[a, b]$. The form of the result is unchanged if the maximum value is taken at a finite number of points.

While the above estimate has considerable use, it is possible that it is an extravagant overestimate in cases where $\nu'(t)$ has no zero in a neighbourhood of each t_α . For, if f and $\mu + i\nu$ were analytic, the method of steepest descents [4, Section 2.5] would yield an estimate of smaller exponential order $\exp(\varepsilon^{-1}\{\mu(t_\beta) + i\nu(t_\beta)\})$, associated with either a singular point t_β of f or a saddle point t_β of $\mu + i\nu$. While this estimate is clearly an improvement, it is also apparent that the benefits obtained from such a procedure may be removed merely by replacing f by $f + \delta \exp\{-\varepsilon^{-1}i\nu(t)\}$ for any arbitrarily small δ . For, then, the dominating estimate of an integral of the type of (2.10) could be as large as

$$K_1 \varepsilon^\alpha \exp\{\varepsilon^{-1}\mu(t_\alpha)\} |\delta|. \tag{2.12}$$

When we consider using such a procedure as the above as part of an estimation technique to establish a contraction mapping proof in the space of continuous functions where only $\|f\|$ will be available, we are forced to the conclusion that the estimate (2.11) is the best available. This view is further reinforced when we consider the situation where f is a function defined by real-world data, or by numerical means, so that the uncertainty arising in its values could easily be of the form $\delta \exp\{-\varepsilon^{-1}i\nu(t)\}$ as given above, with the result that the tighter estimation procedure above would be rendered inappropriate. In fact, the only workable method of improving this estimate would be when holomorphic functions are involved, and the contour of integration may be deformed. This method has been utilized by Meyer [7] and Chapman and Mahony [2] for related problems in a simple case.

Similar considerations apply when $\mu(t)$ is constant on the interval so that while, if $\nu(t)$ is not constant, stationary phase techniques may show that the integral is smaller, the “best general” estimate is

$$|I| \leq K_2 \exp\{(\varepsilon^{-1}\mu)\} (b - a) \|f\|. \tag{2.13}$$

We now apply the considerations above to an estimation of the functional J_i , defined by

$$(J_i f)(t, \varepsilon) = \exp(\zeta_i(t, \varepsilon)) \int_{t_0}^t (-\zeta_i(s, \varepsilon)) f(s, \varepsilon) ds, \tag{2.14}$$

for some t_* lying in an interval $[a, b]$. It is clear that, if $\text{Re}\{\zeta_i(t, \epsilon) - \zeta_i(s, \epsilon)\}$ is positive for any s in $[t_*, t]$, the best estimate, as described by (2.11), will be exponentially large. Smaller estimates will be obtained if t_* is chosen so that $\text{Re}\{\zeta_i(t, \epsilon) - \zeta_i(s, \epsilon)\} \leq 0$ on this interval. Clearly, to make this so, we must choose t_* so that

$$\text{Re } \zeta_i(t_*, \epsilon) = \max_{[a,b]} \text{Re } \zeta_i(t, \epsilon). \tag{2.15}$$

When a number of such values exist, any one will suffice for this choice of t_* . This choice will ensure that the operator norm of J_i is not exponentially large, unless $\text{Re } \zeta_i$ has a local minimum, at t_m say, anywhere in (a, b) . For, when t and t_* are on opposite sides of t_m (and there must be such a t),

$$(J_i f)(t, \epsilon) \sim \exp\{\zeta_i(t, \epsilon) - \zeta_i(t_m, \epsilon)\}, \tag{2.16}$$

which is exponentially large as ϵ vanishes. We may summarize these results in the following lemma.

LEMMA 1. *Let any eigenvalue function $\lambda_i(t, \epsilon)$ be such that either $\text{Re } \lambda_i$ is one-signed or vanishes identically on $[a, b]$, or $\text{Re } \zeta_i$ has at most one local maximum at $t = t_\alpha$, interior to (a, b) , for all $0 < \epsilon \leq \epsilon_0$. Then, for the choice*

$$t_{*i} = \begin{cases} a & \text{if } \text{Re } \lambda_i < 0 \text{ or } \text{Re } \lambda_i \equiv 0, \\ b & \text{if } \text{Re } \lambda_i > 0, \\ t_\alpha & \text{otherwise,} \end{cases} \tag{2.17}$$

the functional J_i is a bounded linear map from the Banach space C_ϵ of continuous functions on $[a, b]$ into itself, where this space is normed, for each $\epsilon > 0$, by the supremum norm taken with respect to t . Further, we have

$$\|J_i\| \leq K_1 \epsilon^{\alpha_i} \quad \text{if } \text{Re } \lambda_i \neq 0, \tag{2.18}$$

and

$$\|J_i\| \leq K_2 \quad \text{if } \text{Re } \lambda_i \equiv 0, \tag{2.19}$$

on $[a, b]$, where K_1, K_2 and α_i are positive constants independent of ϵ , with $0 < \alpha_i \leq 1$.

Clearly, the constant α_i takes the value 1 when the first option in the hypotheses of the lemma holds while, when its value is less than unity, it is determined by the nature of the local maximum of $\text{Re } \zeta_i$ at t_α .

It is apparent from (2.14) that

$$\|J_i\| \leq \sup_{[a,b]} \exp(\text{Re } \zeta_i(t, \epsilon)) \left| \int_{t_{*i}}^t \exp(-\text{Re } \zeta_i(s, \epsilon)) ds \right|. \tag{2.20}$$

Now consider the image under J_i of the particularly simple function f given by

$$f(t, \epsilon) = \exp(i \operatorname{Im} \zeta_i(t, \epsilon)). \quad (2.21)$$

This function has unit norm and, moreover, we obtain from (2.14) that

$$\sup_{[a,b]} |(J_i f)(t, \epsilon)| = \|f\| \sup_{[a,b]} \left| \exp(\operatorname{Re} \zeta_i(t, \epsilon)) \int_{t_*}^t \exp(-\operatorname{Re} \zeta_i(s, \epsilon)) ds \right|. \quad (2.22)$$

Combining (2.20) and (2.22) and noting the definition of the operator norm for bounded linear operators [6, page 94] we arrive at

$$\|J_i\| = \sup_{[a,b]} \left| \exp(\operatorname{Re} \zeta_i(t, \epsilon)) \int_{t_*}^t \exp(-\operatorname{Re} \zeta_i(s, \epsilon)) ds \right|. \quad (2.23)$$

When the conditions of Lemma 1 are realized, this expression is bounded independently of ϵ , or vanishes with ϵ , a result entirely consistent with the findings of that lemma. However, if $\operatorname{Re} \zeta_i(t, \epsilon)$ has a local minimum, t_m , in (a, b) , the situation is somewhat different, as we have already noted. For, then, we get

$$\|J_i\| = C\epsilon^p \exp\left(\operatorname{Re}\left\{\zeta_i(t_*, \epsilon) - \zeta_i(t_m, \epsilon)\right\}\right) \quad (2.24)$$

for constants C and p independent of ϵ . Then, of course, $\|J_i\|$ is exponentially large as ϵ tends to zero. This merely serves to make more convincing our earlier observation regarding this particular case. What is more interesting is the fact that there is no choice of t_* that will remove this undesirable property. Quite apart from the effect of such a change on the results of Lemma 1, we can see from (2.23) that *any* other choice of t_* will lead to $\|J_i\|$ being exponentially large. Then results may be stated in the form of a lemma.

LEMMA 2. *Let $\operatorname{Re} \zeta_i(t, \epsilon)$ have a local minimum at $t = t_m$ in (a, b) for all $0 < \epsilon \leq \epsilon_0$. Then, for the choice of t_* as given by Lemma 1, the functional J_i satisfies the condition (2.24). Moreover, this exponentially large behaviour cannot be removed by any redefinition of t_* .*

The significance of Lemma 2 lies in the fact that there *do* exist functionals J_i such that no suitably bounded estimate of $\|J_i\|$ as given by (2.18) and (2.19) may be obtained, at least on the full set of continuous functions $f(t, \epsilon)$. In a sense, then, the choice of t_* given in Lemma 1 is optimal, in that any change in this choice can only jeopardize the results of that lemma, while doing nothing to alleviate the difficulties arising in Lemma 2. In subsequent sections, we will take care to restrict our class of functionals J_i so that the unbounded estimates of Lemma 2 do not arise.

The results obtained above may be applied to the second term in (2.7) to obtain our first general result of this section.

LEMMA 3. Let all the eigenvalue functions $\lambda_i(t, \epsilon)$ of A satisfy the conditions of Lemma 1 on $[0, T]$, together with the condition that there be no change in the ordering of these functions on this interval. Let $\mathbf{r}(t, \epsilon)$ be a continuous vector-valued function, normed for each $0 < \epsilon \leq \epsilon_0$, by the supremum norm taken with respect to t on $[0, T]$, and let t_{*i} be selected, for each $\lambda_i(t, \epsilon)$, as in Lemma 1. Define $\mathbf{x}_p(t, \epsilon)$ by

$$\mathbf{x}_p(t, \epsilon) = \epsilon^{-1}X(JX^{-1}\mathbf{r})(t, \epsilon), \quad (2.25)$$

where the vector-valued functional J is defined by

$$(J\phi)_k = J_k\phi_k, \quad (2.26)$$

with J_k as defined in Lemma 1.

Then $\mathbf{x}_p(t, \epsilon)$ so defined is a particular integral of the system (1.2) on $[0, T]$ that satisfies the bound

$$\|\mathbf{x}_p\| \leq Ce^{\alpha-1}\|\mathbf{r}\|, \quad (2.27)$$

where the norms are the usual supremum norms, while C and α are constants, independent of ϵ , with $C > 0$ and $0 \leq \alpha < 1$.

PROOF. If we define, for any vector $\mathbf{v}(t, \epsilon)$, the norm

$$\|\mathbf{v}\|^* = \sup_{[0, T]} |X^{-1}\mathbf{v}|, \quad (2.28)$$

it follows from the properties of the basis X that

$$\|\mathbf{v}\|^* = O(1)\|\mathbf{v}\|, \quad (2.29)$$

where the $O(1)$ factor does not vanish with ϵ , and $\|\cdot\|$ is the usual supremum norm for a vector-valued continuous function. Applying this, the definition (2.26) and the results of Lemma 1 to the expression in (2.25), we obtain the result (2.27).

REMARK. Note that, if all the eigenvalue functions have real parts that are not identically zero, we have $0 < \alpha \leq 1$ while, if these real parts are all *strictly* one-signed on $[0, T]$, we have $\alpha = 1$.

When we turn to the case where there are points interior to $[0, T]$ at which the ordering of the eigenvalue functions changes, the procedure is much the same, but with added complications. On each subinterval for which the ordering is constant, there exists a basis of the form of (2.7) for differing vectors X_k , and, using these we may construct a particular integral of the form shown in (2.25). Further, if \mathbf{x}_p and $\bar{\mathbf{x}}_p$ are any two particular integrals, defined on abutting intervals $[t_0, t_1]$ and $[t_1, t_2]$ respectively, where t_1 is a point of change of order, the difference $\{\bar{\mathbf{x}}_p(t_1, \epsilon) - \mathbf{x}_p(t_1, \epsilon)\}$ is well-defined and is, by the results of

Lemma 3, at worst $O(\varepsilon^{\alpha-1})$ for some $0 < \alpha < 1$. This presupposes, of course, that the hypotheses of Lemma 1 are realized on all of $[t_0, t_2]$. We might hope to remove this discontinuity, and hence to construct a particular integral of (1.2) over all of $[t_0, t_2]$ by appending solutions to the homogeneous system (1.1) and this is, in fact, what we do.

To carry out such a construction, we recall the methods applied in Section 6 of Part I. In Lemma 5 of that reference, a contraction mapping procedure was described by which constants c_k , $k = 1, \dots, n$, may be constructed such that, for any vector \mathbf{u} , the representation

$$\mathbf{u} = \sum_{\sigma} c_k \mathbf{X}_k(t_1, \varepsilon) + \sum_{\bar{\sigma}} c_k \bar{\mathbf{X}}_k(t_1, \varepsilon) \quad (2.30)$$

holds for all $0 < \varepsilon \leq \varepsilon_0$. Here, $\mathbf{X}_k(t, \varepsilon)$ and $\bar{\mathbf{X}}_k(t, \varepsilon)$ are the basis vectors arising in (2.2) and corresponding to $[t_0, t_1]$ and $[t_1, t_2]$, respectively, while $\sigma, \bar{\sigma}$ is any suitable partitioning of the integers $1, \dots, n$. The vector \mathbf{c} with components c_k was shown to satisfy

$$|\mathbf{c}| \leq O(1)|\mathbf{u}|, \quad (2.31)$$

where the $O(1)$ factor does not vanish with ε .

We now apply these results to obtain the following lemma.

LEMMA 4. *Let the eigenvalue functions $\lambda_i(t, \varepsilon)$ satisfy the conditions of Lemma 1 on $[0, T]$ and let there be a finite number of changes in the ordering of these functions on this interval. For each subinterval on which the ordering is constant, let t_* be defined by (2.17), and let $\mathbf{X}_k \exp(\zeta_k)$ be the local basis on such a subinterval analogous to (2.2). Then, there exists a particular integral of (1.2), continuously differentiable throughout all of $[0, T]$, that takes the form*

$$\mathbf{x}_p(t, \varepsilon) = X(t, \varepsilon)Z(t, \varepsilon)Z(t_*, \varepsilon)^{-1}\mathbf{c} + \varepsilon^{-1}X(t, \varepsilon)Z(t, \varepsilon) \int_{t_*}^t Z^{-1}X^{-1}\mathbf{r} \, ds, \quad (2.32)$$

on each such subinterval, where \mathbf{c} is a constant n -vector satisfying

$$\mathbf{c} = O(\varepsilon^{\alpha-1}), \quad (2.33)$$

for some $0 < \alpha \leq 1$, and X is the matrix defined by (2.5).

PROOF. We prove the result for a pair of abutting subintervals $[t_0, t_1]$ and $[t_1, t_2]$, where t_1 is a point of order change. The extension to the whole of $[0, T]$ is straightforward.

We first suppose that $x_p(t, \epsilon)$ and $\bar{x}_p(t, \epsilon)$ are the particular integrals constructed as in Lemma 3 and corresponding to $[t_0, t_1]$ and $[t_1, t_2]$, respectively. We can then define u by

$$u = \bar{x}_p(t_1, \epsilon) - x_p(t, \epsilon), \tag{2.34}$$

and this, by the results of Lemma 3 is, at worst $O(\epsilon^{\alpha-1})$, for some $0 < \alpha \leq 1$. In this way, we may construct the constants c_k defined in (2.30) for the bases $\bar{X}_k \exp(\zeta_k)$ and $\bar{\bar{X}}_k \exp(\zeta_k)$ on each subinterval, and it only remains to choose the partitioning σ and $\bar{\sigma}$. We choose σ to invoke only those X_i such that $\text{Re } \lambda_i \geq 0$ or $\text{Re } \lambda_i \equiv 0$ on $[t_0, t_1]$, and $\bar{\sigma}$ to invoke those X_i corresponding to $\text{Re } \lambda_i < 0$, somewhere on this first interval. This choice covers all possibilities, unless there is no reasonable particular integral on the first interval. Further, if for any i in $\bar{\sigma}$, $\text{Re } \lambda_i$ becomes positive on $[t_1, t_2]$, the corresponding ζ_i has $\text{Re } \zeta_i$ with a minimum on the joint interval, and is thus to be excluded from our construction, by the assumptions of the Lemma 3. Then, with the corresponding vector c determined from u as above, the first term in (2.29) may be continued into $[t_0, t_1]$ in the form

$$x_1(t, \epsilon) = \sum_{\sigma} c_k X_k(t, \epsilon) \exp(\zeta_k(t, \epsilon) - \zeta_k(t_1, \epsilon)), \tag{2.35}$$

and the second into $[t_1, t_2]$ in the form

$$x_2(t, \epsilon) = \sum_{\bar{\sigma}} c_k \bar{X}_k(t, \epsilon) \exp(\zeta_k(t, \epsilon) - \zeta_k(t_1, \epsilon)). \tag{2.36}$$

By the assumptions above, no exponential term is large, so we have the stated result, with the particular integral over all of $[t_0, t_2]$ given by

$$\bar{\bar{x}}_p(t, \epsilon) = \begin{cases} x_p(t, \epsilon) + x_1(t, \epsilon) & \text{for } t \in [t_0, t_1], \\ \bar{x}_p(t, \epsilon) + x_2(t, \epsilon) & \text{for } t \in [t_1, t_2]. \end{cases} \tag{2.37}$$

The process is obviously continuable to any finite number of subintervals for which the assumptions of Lemma 3 hold.

3. Relation to the outer solution

There are well-established heuristic methods for generating a formal approximation, generally known as the outer solution, for the differential system

$$\epsilon \dot{x} = Ax + r \tag{3.1}$$

when the matrix function $A(t, \epsilon)$ and the vector function $r(t, \epsilon)$ are suitably smooth. Then formal expansions may be obtained by rearranging the equation (3.1) in the form

$$x = -A^{-1}(r - \epsilon \dot{x}) \tag{3.2}$$

and then solving iteratively. Provided the derivatives involved exist and are suitably continuous, such an iteration procedure leads to the expansion

$$\mathbf{x} = -A^{-1}\mathbf{r} - \sum_{n=1}^N \epsilon^n A^{-1}(DA^{-1})^n \mathbf{r} + \epsilon^{N+1} A^{-1}(DA^{-1})^N \dot{\mathbf{x}}, \quad (3.3)$$

where, as in the above, D denotes the operation of differentiation with respect to t , and N is a selected positive integer.

We are here concerned with the status of the formal approximation obtained by deleting the last term above, that involves the derivative $\dot{\mathbf{x}}$. It is obvious that for some special systems, $(DA^{-1})^N \mathbf{r}$ may be identically zero for all (t, ϵ) and some value of N , in which case the above approach will lead to an exact solution in finite terms. In what follows, we will confine our attention to the cases for which the formal series does not terminate.

What is striking about the formal series is that, for any value of t , it is only the properties of A and \mathbf{r} at a given point that serve to fix the expansion here. Thus the formal approximation is defined at any t for smooth A and \mathbf{r} , provided that A is nonsingular there, regardless of whether A is singular elsewhere in the interval. Therefore, it is pertinent to ask questions about the status of the formal series at points where A is invertible, and the effects on this status of the points where A is singular. Further, if the above series is to be of use, it is desirable to obtain the corresponding form for the case when A is singular at a point. We will return to this aspect later in this section.

It is of some interest to compare the formal expansion (3.3) with the particular integral of (3.1) obtained in the previous section that has the form

$$\mathbf{x} = \epsilon^{-1} XZ \int' Z^{-1} X^{-1} \mathbf{r}, \quad (3.4)$$

where, for the moment, the lower terminals of integration are left arbitrary. In Section 2, we used the fact that, for certain choices of these terminals, the contribution from the integral above was dominated by that from a neighbourhood of the point t . We will need to make use of this here, but will leave these terminals arbitrary at the moment. For the present, we note that (3.4) may be expanded by the following procedure. We first note the identity

$$D\{Z^{-1} X^{-1} XZ\} = 0 \quad (3.5)$$

and, further, since XZ is a fundamental matrix on each subinterval where the eigenvalue functions of A have no change of order, we have

$$\epsilon D(XZ) = AXZ. \quad (3.6)$$

Thus we have from the above, provided that the range of integration does not include a singular point of A ,

$$\mathbf{x} = -XZ \int' D(Z^{-1} X^{-1}) A^{-1} \mathbf{r}, \quad (3.7)$$

which becomes, on integration by parts,

$$\mathbf{x} = -XZ[Z^{-1}X^{-1}A^{-1}\mathbf{r}]' + XZ \int^t Z^{-1}X^{-1}D(A^{-1}\mathbf{r}). \tag{3.8}$$

Obviously, this process may be continued as long as the regularity properties of A and \mathbf{r} hold, and will thus lead to an expansion closely related to the formal series obtained in (3.3). Moreover, for the choice t_* as the lower limit of integration, where t_* is that value chosen in Section 2, the last term in (3.8) is clearly

$$XJ(X^{-1}DA^{-1}\mathbf{r})(t, \epsilon), \tag{3.9}$$

where J is the operator of that section, and we may use the norm estimates of that section to discuss the conditions under which we may confirm the validity of the formal expansion (3.3) for general \mathbf{r} . We emphasize again that there will be special functions \mathbf{r} for which improved result will hold, but we are not concerned here with such questions.

First, we note that after any given number of steps of the iteration of the above process (3.8), the contribution from the lower terminal to the terms removed from the integral is a solution of the homogeneous equation for (3.1), and can be considered separately. It will become more significant in the context of a later discussion to be given in Section 4. The exact solution to be described by the formal series then must be of the form

$$\mathbf{x} = - \sum_{n=0}^{N-1} \epsilon^n (A^{-1}D)^n A^{-1}\mathbf{r} + XZ\mathbf{c} + \epsilon^{N-1}X(JX^{-1}(A^{-1}D)^N A^{-1}\mathbf{r}), \tag{3.10}$$

for some (not necessarily unique) determination of the constant vector \mathbf{c} . We can see from the above that the formal process (3.3) will define an asymptotic approximation to the solution if and only if there is a choice of constant \mathbf{c} such that

$$\left\| XZ \left\{ \mathbf{c} + \epsilon^N \int^t Z^{-1}X^{-1}(A^{-1}D)^N A^{-1}\mathbf{r} \right\} \right\| = o(\epsilon^N \|(A^{-1}D)^N A^{-1}\mathbf{r}\|), \tag{3.11}$$

for all (t, ϵ) and appropriate vector function $\mathbf{r}(t, \epsilon)$.

Recalling that we are still free to choose the lower terminal in the integral above, we can see that we are here faced with the same problem that arose in Section 2 and, here, the resolution is much the same as in that section, in that for a suitable choice of this terminal (and hence \mathbf{c}) the above relation is realized. To obtain this, we make the assumptions of Lemma 2 of Section 2, and thus restrict the above considerations to a subinterval on which the ordering of the real parts of the eigenvalue functions is constant. Further, the order relation (3.11) may also be obtained if we have suitably exponentially small estimates on $\|\mathbf{r}\|$, although we feel that, in general, this is an unlikely prospect. Thus, we conclude

that the lower terminal of (3.4) must be chosen to be the $t_*(\equiv \{t_{*i}\})$ of Lemma 2, namely such that for each component of the form

$$\exp(\zeta_i(t, \epsilon)) \int_{t_{*i}}^t (-\zeta_i(s, \epsilon)) v_i(s, \epsilon) ds, \tag{3.12}$$

Re ζ_i decreases monotonically away from Re $\zeta_i(t_{*i}, \epsilon)$, or else Re ζ_i is identically zero on the subinterval concerned. Further, we can use the same arguments as applied in the proof of Lemma 3 to establish the following theorem, the detailed proof of which is omitted, as it is essentially repetitious.

THEOREM 1. *Let N be a positive integer, and let \mathbf{r} and A be N times differentiable with respect to t on any interval (T_1, T_2) for which the ordering of the real parts of the eigenvalue function is preserved. Let $A(t, \epsilon)$ be invertible on this interval for given $\epsilon > 0$, and let $\|D^n \mathbf{r}\|$ and $\|D^n A^{-1}\|$ be bounded uniformly on (T_1, T_2) , with bounds independent of ϵ and n , for all $0 \leq n \leq N$. Let no eigenvalue function $\lambda_i(t, \epsilon)$ generate a function $\zeta_i(t, \epsilon)$ as defined by (2.3) for which Re ζ_i has a local minimum in (T_1, T_2) .*

Then there exists a solution $\mathbf{x}(t, \epsilon)$ of the differential system (3.1) on this interval such that

$$\lim_{\epsilon \rightarrow 0} \left\| \epsilon^{-N+\alpha} \left\{ \mathbf{x} + \sum_{n=0}^{N-1} \epsilon^n (A^{-1}D)^n A^{-1} \mathbf{r} \right\} \right\| = 0, \tag{3.13}$$

for some α satisfying $0 \leq \alpha < 1$, where $\|\cdot\|$ is the usual vector supremum norm on (T_1, T_2) . For any i such that Re ζ_i has a local minimum in this interval, the result will hold only for the restricted class of functions $\mathbf{r}(t, \epsilon)$ for which

$$\|X(JX^{-1}(A^{-1}D)^N A^{-1} \mathbf{r})\| \rightarrow 0 \tag{3.14}$$

as $\epsilon \rightarrow 0$.

REMARKS. Note that the arguments of Section 2 show that this last condition is very strong. It is also surprising to observe that the heuristic expansion (3.3) becomes invalid when Re ζ_i has a minimum, even if A is nonsingular at such a point.

When such a particular integral as (3.4) is used in the context of nonlinear problems, for the purpose of producing a contraction mapping proof, \mathbf{r} involves an error term, and a nonlinear term that involves the unknown \mathbf{x} itself. A bootstrap argument then shows us that the results of Theorem 1 will hold provided \mathbf{r} is continuously differentiable once with derivative bounded uniformly in t and ϵ . This implies certain differentiability properties of the nonlinearity occurring in the equation.

What is even more striking is the fact that the above result may be extended to the case where A is singular at an interior point of (T_1, T_2) , provided that at such a point a zero of a real eigenvalue $\lambda_i(t, \epsilon)$ corresponds to a maximum of the relevant $\zeta_i(t, \epsilon)^*$. When such a condition holds, the expansion (3.13) is valid throughout all of (T_1, T_2) except within a small neighbourhood of such a singularity of A . For we may write the i th component relative to the basis X of the particular integral (3.4) as

$$\epsilon^{-1} \exp(\zeta_i(t, \epsilon)) \int_{t_{*i}}^{t_{ii}} \exp(-\zeta_i)(X^{-1}\mathbf{r})_i - \exp(\zeta_i(t, \epsilon)) \int_{t_{ii}}^t (D(Z^{-1}X^{-1})A^{-1}\mathbf{r})_i, \quad (3.15)$$

for $t \geq t_{ii}$, where t_{ii} is suitably chosen and, in this case, t_{*i} coincides with the singularity of A . Clearly, when $t \leq t_{ii}$, the second integral does not occur, so that we are left to consider the first. We may now apply the iteration procedure described by (3.8) to (3.15) to obtain the second term in the required form (3.13). Moreover, the first term, for t_{ii} appropriately bounded away from t_{*i} is exponentially small so that this term may be absorbed into the remainder. This gives us the validity of the expansion (3.13) except in these small neighbourhoods of the relevant zeros of the appropriate eigenvalues λ_i .

To estimate this solution in the neighbourhood of such a singular point, the procedure is straightforward. At this point, the reader should note that the assumptions of Part I do not allow for a complex eigenvalue vanishing at an interior point of such an interval (T_1, T_2) . The ensuing discussion should be viewed in the light of this fact. The relevant component of this solution is representable in the form

$$\exp(\zeta_i(t, \epsilon)) \int_{t_{*i}}^t \exp(-\zeta_i(s, \epsilon))(X^{-1}\mathbf{r})_i(s, \epsilon) ds \quad (3.16)$$

relative to the basis X , where

$$\zeta_i'(t_{*i}, \epsilon) = 0. \quad (3.17)$$

All that is necessary to obtain asymptotic estimates as $\epsilon \rightarrow 0$ for such an integral is to use a local approximation to ζ_i in the neighbourhood of t_{*i} and the results given in Erdélyi [4, page 37] which show the integral to be dominated by the contribution from the integrand in a suitable neighbourhood of t_{*i} , the extent of which tends to zero with ϵ . If λ_i has a simple zero, the unbounded component of $A^{-1}\mathbf{r}$ is replaced by a term that matches smoothly, and which is such that the local contribution to the particular solution is no larger than $O(\epsilon^{-1/2})$. There are obvious generalizations to the case of higher-order zeros.

4. Reasonable linear problems

As we have noted in Section 2, the general solution of the linear equation (2.1) may be written in the form

$$\mathbf{x}(t, \varepsilon) = XZ\mathbf{c} + \varepsilon^{-1}X(JX^{-1}\mathbf{r})(t, \varepsilon), \quad (4.1)$$

where $(Jf)(t, \varepsilon)$ is the functional defined there, while $\mathbf{c} = \mathbf{c}(\varepsilon)$ is an n -vector of constants, and the choice of the lower limit in the integral has been discussed in that section. For the moment, we will assume that $[0, T]$ is an interval on which there are no changes in the order of eigenvalues, so that the basis represented by the square $n \times n$ matrix $X(t, \varepsilon)$ has the same form throughout this interval, and is thus of the elementary form given by equations (2.4).

Clearly, the first term in (4.1) defines a one-to-one correspondence between such vectors \mathbf{c} in R^n and basis vectors $\mathbf{X}_i(t, \varepsilon)$ so that, in the discussion that follows, we will use the same symbol to denote subspaces spanned by selected subsets of the \mathbf{X}_i (or \mathbf{x}_i) and the corresponding subspaces in R^n of vectors \mathbf{c} . The meaning should be clear from the context. We will also adopt the convention that, while XM represents a matrix for any (possibly nonsquare) compatible matrix M , this symbol may also be used to denote a subspace of R^n consisting of the appropriate linear combinations of the vectors \mathbf{X}_i . Again, no confusion should result, since the context will make the meaning clear.

When we come to consider boundary conditions relevant to the system (4.1), we can see that such conditions can take a wide variety of forms, both linear and nonlinear. Also, in many situations, these may not consist of the same number of linearly independent conditions as the order n of the original system. When such situations *do* arise, we envisage that there are parameters in the problem where values are to be determined in such a way that such a problem with otherwise apparently too many boundary conditions has a consistent solution. While nonlinear boundary conditions *do* occur, and are of interest in applications, we will restrict our attention here to linear ones, though the extension to nonlinear conditions may be made by analogous arguments. Thus, in this section, we will assume that there are essentially m_1 linearly independent boundary conditions applied at $t = 0$ and m_2 such conditions at $t = T$. We will suppose that these take the form

$$L\mathbf{x}(0, \varepsilon) = \mathbf{f}, \quad (4.2)$$

where L is an $m_1 \times n$ matrix of constants and \mathbf{f} is a constant m_1 -vector, and

$$R\mathbf{x}(T, \varepsilon) = \mathbf{g}, \quad (4.3)$$

where R and \mathbf{g} are $m_2 \times n$ matrix and m_2 -vector respectively, with analogous properties. We also anticipate that the entries of L and R may vary with ε and,

in such a case, we suppose that these are $O(1)$ for all ϵ in a neighbourhood of zero, and vanish at most algebraically with ϵ .

In principle, the boundary conditions (4.2) and (4.3) may be applied directly to the solution (4.1) to yield an equation determining the vector \mathbf{c} , and whether or not such a vector exists as a solution to such an equation is an algebraic question that can be resolved for each $\epsilon > 0$. However, to determine whether or not the solutions constructed in this way are reasonable in the sense described in Section 1, we must examine the structure of this problem in more detail.

Equations (4.1) to (4.3) show us that the map from the data \mathbf{f} , \mathbf{g} and \mathbf{r} of this boundary-value problem to the solution $\mathbf{x}(t, \epsilon)$ comprises two components—that defined by (4.2) and (4.3) mapping \mathbf{f} , \mathbf{g} and \mathbf{r} into \mathbf{c} and that defined by (4.1) and mapping \mathbf{c} and \mathbf{r} into \mathbf{x} . Thus, to establish whether or not this boundary-value problem is reasonably posed, we only need estimate the norms of these (linear) maps. Under the criteria laid down in Section 1, we will regard this problem as being *reasonably posed* if and only if the estimates of these norms are at worst algebraically large in ϵ for ϵ in a neighbourhood of zero. It should be recognised that in doing this, we are seeking sufficient conditions *only* that such a property may be held by this problem; there *may* be reasonably posed problems that lie outside this definition. In estimating these norms, we will adopt the usual Cartesian generalizations of the uniform and scalar norms, although these norms will depend on ϵ in this instance.

The discussion of the earlier sections of this paper, and also those of Part I, lead us to adopt the classification below for the modes of behaviour of eigenvalues $\lambda_i(t, \epsilon)$ of A and the corresponding vectors $\mathbf{X}_i(t, \epsilon)$. Although it would seem more natural to relate this classification to the eigenvectors $\mathbf{s}_i(t, \epsilon)$, the property (2.2) of the vectors $\mathbf{X}_i(t, \epsilon)$ shows that they are in one-to-one correspondence with these eigenvectors, so that the classification may be made relative to these, without any fear of ambiguity. Since the vectors \mathbf{X}_i also span R^n , the nett result is a small perturbation of the spaces that would have resulted from a classification on the basis of the eigenvectors.

Thus we make the decomposition

$$R^n = X^{(-)} \oplus X^{(+)} \oplus X^{(0)} \oplus X^{(R)}, \tag{4.4}$$

where the component spaces $X^{(-)}$, $X^{(+)}$, $X^{(0)}$, and $X^{(R)}$ are as defined below:

- (i) $\mathbf{X}_i \in X^{(-)}$ if and only if $\text{Re} \int_0^t \lambda_i$ has its maximum value at $t = 0$ and possibly a finite number of interior points;
- (ii) $\mathbf{X}_i \in X^{(+)}$ if and only if $\text{Re} \int_0^t \lambda_i$ has its maximum value at $t = T$, and possibly a finite number of interior points;
- (iii) $\mathbf{X}_i \in X^{(0)}$ if and only if $\text{Re} \int_0^T \lambda_i = 0$ and $\text{Re} \int_0^t \lambda_i$ is nowhere positive;
- (iv) $\mathbf{X}_i \in X^{(R)}$ if and only if $\text{Re} \int_0^t \lambda_i$ takes its maximum values at interior points only.

Note that, when viewed as *vector spaces*, these do not vary with (t, ϵ) in the sense that they consist of the same array of solutions X_i but that, when viewed as *matrices*, they display a pointwise variation.

Solutions $x(t, \epsilon)$ corresponding to elements of $X^{(-)}$ and $X^{(+)}$ display characteristic boundary-layer behaviour at $t = 0$ and $t = T$, respectively, and thus correspond to those arising in the more usual cases of eigenvalues having strictly negative and positive real parts, so common in the literature. Those corresponding to elements of $X^{(0)}$ may display this behaviour at either end-point of $[0, T]$, while those corresponding to elements of $X^{(R)}$ are distinctive in that, when normalized to be of order unity at either $t = 0$ or $t = T$, they display values in the interior of that interval that are exponentially large. Such solutions appear to correspond, in the present context, to the “resonance” solutions of Ackerberg and O’Malley [1], already mentioned in Section 1. We will also find it convenient to define one other subspace of R^n that may have a non-null intersection with all of the spaces defined above. As we noted in Section 2, the occurrence of an interior minimum of $\text{Re } \zeta_i(t, \epsilon)$ led to the second term of (4.1) being exponentially large, both at $t = 0$ and $t = T$. Such behaviour is of some interest, so we will deal with it by introducing the space X_{\min} , spanned by all those solutions $X_i(t, \epsilon)$ for which the corresponding $\text{Re } \zeta_i(t, \epsilon)$ has just such an interior minimum. Thus, in terms of this definition, our condition that exponentially large terms do *not* arise in the particular integral term of (4.1) is that the vector r have no component in X_{\min} ; that is, if we write $r = X R$, we require that $R_i \equiv 0$ for all $X_i \in X_{\min}$.

Thus, for r satisfying such a condition, we see on an application of Lemma 1 that the second term of (4.1) is bounded above by a term which is, at worst,

$$O(\epsilon^{-1})\|r\|, \tag{4.5}$$

where $\|\cdot\|$ is the usual norm (3.4). The first term is bounded above by a term of the form

$$\text{diag}\left(\exp\left\{\sup_t \text{Re } \zeta_i(t, \epsilon)\right\}\right)|c|. \tag{4.6}$$

For $X_i \in X^{(-)} \oplus X^{(0)}$ the first term in (4.6) is $O(1)$ as $\epsilon \rightarrow 0$. However, for X_i in the other spaces, it will become exponentially large. The only way we can make this map bounded for given ϵ in a neighbourhood of zero is to rescale the c ; that is, essentially, choose another norm for c . Thus, in what follows, we replace c by

$$Dc \equiv \text{diag}\left(\exp\left\{-\sup_t \text{Re } \zeta_i(t, \epsilon)\right\}\right)c \tag{4.7}$$

but use the same symbol for c ; only now we seek c 's that are $O(1)$. With this choice, we see that the map $(c, r) \rightarrow x$ defined by (4.1) is bounded with norm

bounded by $O(\epsilon^{-1})$ at worst. This bound will be improved to $O(\epsilon^{\alpha-1})$ for some $0 < \alpha \leq 1$ if there are no eigenvalue functions $\lambda_i(t, \epsilon)$ for which $\text{Re } \lambda_i \equiv 0$ on $[0, T]$. Note that, for $X_i \in X^{(-)} \oplus X^{(0)}$, the suprema in (4.7) are both zero, so that the corresponding components of \mathbf{c} are unchanged.

We now apply (4.1) to the boundary conditions (4.2) and (4.3) to obtain

$$LX(0, \epsilon)Z(0, \epsilon)D\mathbf{c} + \epsilon^{-1}LX(0, \epsilon)(JX^{-1}\mathbf{r})(0, \epsilon) = \mathbf{f}, \tag{4.8}$$

and

$$RX(T, \epsilon)Z(T, \epsilon)D\mathbf{c} + \epsilon^{-1}RX(T, \epsilon)(JX^{-1}\mathbf{r})(T, \epsilon) = \mathbf{g}, \tag{4.9}$$

respectively.

Under usual circumstances, we might expect that, at best, components of \mathbf{f} , \mathbf{g} and \mathbf{r} are algebraically small in ϵ for small $\epsilon > 0$ (or variations in these quantities display this behaviour). If, then, one of the data terms in (4.8) has a component in $LX^{(+)}(0, \epsilon)$ or $LX^{(R)}(0, \epsilon)$ that is not approximately exponentially small, the corresponding component of \mathbf{c} is premultiplied by an exponentially small term so that the map $(\mathbf{f}, \mathbf{g}) \rightarrow \mathbf{c}$ represented by these equations must have a norm that becomes exponentially large like $\exp(\epsilon^{-1})$ as ϵ vanishes. A similar situation arises for components of \mathbf{g} lying in $RX^{(-)}(T, \epsilon)$ or $RX^{(R)}(T, \epsilon)$.

We may combine the arguments of this section in the following result.

THEOREM 2. *Let the assumptions of Section 2 regarding the matrix A apply, and suppose that there are no changes in the ordering of the real parts of the eigenvalue functions $\lambda_i(t, \epsilon)$ of A on $[0, T]$. Let the operators L and R and the functions \mathbf{f} and \mathbf{g} be as given in (4.2) and (4.3), and suppose that $\mathbf{r}(t, \epsilon)$ is defined and continuous on $[0, T]$.*

Let α be an arbitrary n -vector, such that $|\alpha| = O(1)$ as $\epsilon \rightarrow 0$ and define the vectors $\mu(\epsilon)$ and $\nu(\epsilon)$ by

$$\mu \equiv Z(0, \epsilon)D\alpha \tag{4.10}$$

and

$$\nu = Z(T, \epsilon)D\alpha, \tag{4.11}$$

respectively.

Let the functions \mathbf{f} , \mathbf{g} and \mathbf{r} satisfy the conditions

(i) \mathbf{r} has no component in X_{\min} for all $(t, \epsilon) \in [0, T] \times (0, \epsilon_0]$,

(ii) $\mathbf{p}(\epsilon) \equiv \mathbf{f} - \epsilon^{-1}LX(0, \epsilon)(JX^{-1}\mathbf{r})(0, \epsilon) = O(\epsilon^a LX\mu)$ on $L(X^{(+)} \oplus X^{(R)})$, and

(iii) $\mathbf{q}(\epsilon) \equiv \mathbf{g} - \epsilon^{-1}RX(T, \epsilon)(JX^{-1}\mathbf{r})(T, \epsilon) = O(\epsilon^b RX\nu)$ on $R(X^{(-)} \oplus X^{(R)})$,

where a and b are suitably chosen constants, that may be negative.

Then the boundary-value problem (4.1) to (4.3) is reasonably posed if and only if

$$LX^{(-)}\mathbf{c} + LX^{(0)}\mathbf{c} = \mathbf{p}, \tag{4.12}$$

$$RX^{(+)}\mathbf{c} + RX^{(0)}\mathbf{c} = \mathbf{q} \tag{4.13}$$

is well-posed in the sense of Hadamard.

PROOF. The well-posed property of (4.12) and (4.13) ensures that it may tolerate small perturbations to the solution \mathbf{c} , provided the conditions indicated hold. Thus the equations (4.8) and (4.9) may be solved iteratively and the conditions (ii) and (iii) ensure that the norm of the operator $(\mathbf{f}, \mathbf{g}) \rightarrow \mathbf{c}$ is at most algebraically large in ε as $\varepsilon \rightarrow 0$.

It should be noted that the condition (i) of Theorem 1 is excessive, and the results obtained in Section 2 show us that, provided \mathbf{r} is suitably exponentially small on X_{\min} , it will not give rise to the unboundedness experienced in that section. However, we regard this degree of smallness to be so restrictive as to be unlikely to be establishable in most situations, so we will not pursue this line here, save to note its implications. This criticism could also be levelled at the degree of accuracy required in the constraints (ii) and (iii) on \mathbf{f} , \mathbf{g} and \mathbf{r} for certain subspaces of R^n . These conditions in fact imply that these vectors, and hence the solutions $\mathbf{x}(t, \varepsilon)$ corresponding to them, lie on a manifold in R^n defined by them. This is clearly an extension of the basic idea demonstrated by reference to the elementary example of Section 1. Moreover, the extreme sharpness of these conditions only serves to demonstrate that, for linear problems of the form (4.1) to (4.3), many boundary-value problems are unreasonable.

For the particular case of the homogeneous problem ($\mathbf{r} = \mathbf{0}$), the manifold mentioned above centres on

$$\mathbf{f} = \mathbf{0} \text{ on } L(X^{(+)} \oplus X^{(R)}), \quad \mathbf{g} = \mathbf{0} \text{ on } R(X^{(-)} \oplus X^{(R)}), \quad (4.14)$$

and has a diameter that is exponentially small as $\varepsilon \rightarrow 0$, so that the corresponding reasonable solutions $\mathbf{x}(t, \varepsilon)$ centre on solutions vanishing on these spaces. This is, of course, as would be expected.

We may also compare the results of this section with those commonly occurring in the literature. When $\text{Re } \lambda_i$ do not change sign on $[0, T]$, the $\text{Re } \xi_i$ are monotone there, and thus the spaces $X^{(0)}$, $X^{(R)}$ and X_{\min} are empty. Under these circumstances, the conditions of Theorem 1 reduce to the condition that $\mathbf{r}(t, \varepsilon)$ be bounded uniformly in (t, ε) and that \mathbf{f} and \mathbf{g} be exponentially small on $LX^{(+)}$ and $RX^{(-)}$, respectively. When we compare these with those conditions usually assumed in the literature, namely that \mathbf{f} and \mathbf{g} vanish on these spaces, we can see that the thrust of much of the work to date has been to construct solutions that are *reasonable* in the sense of this paper, without viewing them as solutions imbedded in a much wider class.

If we assume that $A(t, \varepsilon)$ is invertible, as well as some other differentiability properties about A and \mathbf{r} , we may write the conditions of Theorem 1 in terms of $-A^{-1}\mathbf{r}$, the outer solution of the equation (3.1).

COROLLARY 1. *Let $A(t, \varepsilon)$ be invertible for all $(t, \varepsilon) \in [0, T] \times (0, \varepsilon_0]$, and let $A^{-1}\mathbf{r}$ have a derivative with respect to t that is continuous in t and uniformly bounded with respect to (t, ε) . Further, suppose that X_{\min} is empty.*

Define the function $\Phi(t, \epsilon)$ by

$$\Phi(t, \epsilon) = -A^{-1}\mathbf{r}(t, \epsilon) + X(t, \epsilon)JX^{-1} \frac{d}{dt}(A^{-1}\mathbf{r})(t, \epsilon). \quad (4.15)$$

Then the conditions (i), (ii) and (iii) of Theorem 1 hold if and only if

- (i) $\bar{\mathbf{p}}(\epsilon) \equiv \mathbf{f} - L\Phi(0, \epsilon) = O(\epsilon^a LX\boldsymbol{\mu})$ on $L(X^{(+)} \oplus X^{(R)})$, and
 (ii) $\bar{\mathbf{q}}(\epsilon) \equiv \mathbf{g} - R\Phi(T, \epsilon) = O(\epsilon^b RX\boldsymbol{\nu})$ on $R(X^{(-)} \oplus X^{(R)})$, respectively, where $a, b, \boldsymbol{\mu}$ and $\boldsymbol{\nu}$ are as before.

PROOF. Clearly (i) is removed by assuming that X_{\min} is absent. We will thus be able to prove the desired result if we can show that $L\Phi(0, \epsilon)$ and $\epsilon^{-1}LX(0, \epsilon)(JX^{-1}\mathbf{r})(0, \epsilon)$ differ by a term of the form of the right-hand side of (i) on the relevant domain.

Using the analysis of Section 3, we obtain, on integrating by parts,

$$\epsilon^{-1}X(JX^{-1}\mathbf{r})(t, \epsilon) = \Phi(t, \epsilon) + X(t, \epsilon)Z(t, \epsilon)Z^{-1}(t_*, \epsilon)X^{-1}(t_*, \epsilon)A^{-1}\mathbf{r}(t_*, \epsilon), \quad (4.16)$$

where Φ is as defined above.

Now it is clear from the discussion of Section 2 that $D = Z^{-1}(t_*, \epsilon)$ and thus we have, on noting the assumptions regarding A and \mathbf{r} ,

$$Z(0, \epsilon)Z^{-1}(t_*, \epsilon) = O(\boldsymbol{\mu}) \quad (4.17)$$

and

$$Z(T, \epsilon)Z^{-1}(t_*, \epsilon) = O(\boldsymbol{\nu}), \quad (4.18)$$

and thus the result is proved by applying these to the conditions (ii) and (iii) of Theorem 1.

The results of this corollary merit a closer examination. First, it is clear from (4.15) that Φ consists of the formal "outer solution" $-A^{-1}\mathbf{r}$ together with a correction factor. Moreover, under the assumption made, this factor is $O(\epsilon^\beta)$ for some $0 < \beta < 1$, so that it represents a small "outer correction". The results (i) and (ii) of the corollary thus give clear-cut conditions on the end-values $L\Phi(0, \epsilon)$ and $R\Phi(T, \epsilon)$ in appropriate domains, such that the resulting solution to the boundary-value problem be reasonably posed. It should be noted that these conditions show that it is the values of this corrected outer solution that determine *which* problems are reasonably posed and not the end-values of the full solution, as one might have thought. What this result demonstrates is that both Φ and the full solution have end-values lying in the manifold of reasonable solutions mentioned earlier, under the conditions of the corollary.

The results of Corollary 1 may also be interpreted in another manner that has some significance for nonlinear problems. They show that, given an outer

solution, in this case $-A^{-1}\mathbf{r}$, we may construct a correction so that the resulting outer solution provides the basis for a reasonable solution of the boundary-value problem. In this case, the rule for the construction of this correction is clear and explicit, and no difficulties would be anticipated. Thus, this gives a means of “regularizing” an unreasonable problem, namely that of just using $-A^{-1}\mathbf{r}$ in the boundary conditions. We will return to this in the next section, where we apply it to nonlinear problems.

Where there are changes in the ordering defined on the real parts of the eigenvalue functions λ_i , the results of Theorem 1 require amendment. As we have shown in Part I, there is a well-defined algebraic process for continuing basis solutions across such points of order change. The rules for the selection of the continuation may be quite complex, depending on the number and manner of order changes. Once they have been implemented, however, they serve merely to modify the allocation of any basis solution to the spaces $X^{(-)}$, $X^{(+)}$, $X^{(0)}$, $X^{(R)}$ and X_{\min} . Thereafter, the analysis is essentially the same, with these redefinitions of the partitioning into subspaces. The reader is referred to Part I for some examples illustrating how the order changes can affect the partitioning. These also show that the effective dimension of the solution space may be changed after such a change of order, which affects the possible choices of boundary values that give rise to reasonably posed problems. They may also vary the number and location of reasonable boundary conditions.

5. Possible regularizations

As we have pointed out in the Introduction, it may be possible to make a minor modification to a linear system that transforms an unreasonable problem into a reasonable one. This could be useful in generating a contraction mapping proof for the existence of the solution of a nonlinear system close to a given candidate approximation. For such an iterative proof, the modification referred to above will change at each step, so that the resultant modification for the dominant linear system must be seen as part of the solution process. Thus the overall contractive process must also apply to this “regularization correction” as well as to the solution of the problem itself.

We thus start with the assumption that we are provided with a candidate approximation, \mathbf{w} , to the solution \mathbf{y} of the nonlinear system

$$\varepsilon \dot{\mathbf{y}} = \mathbf{f}(t, \mathbf{y}, \varepsilon), \quad 0 < t < T, \quad (5.1)$$

together with the linear boundary conditions (4.2) and (4.3), and seek to show that there is an exact solution $\mathbf{w} + \mathbf{z}$ of the full nonlinear boundary-value

problem close to it. The standard approach to such a problem is to compute the errors by which the candidate w fails to satisfy the required conditions.

Thus, let

$$\epsilon \dot{w} - f(t, w, \epsilon) = r, \tag{5.2}$$

$$Lw(0, \epsilon) = f + h \tag{5.3}$$

and

$$Rw(T, \epsilon) = g + k, \tag{5.4}$$

where L and R are the boundary operators introduced in Section 4, while f and g are known constant vectors.

For $w + z$ to be an exact solution, then,

$$\epsilon(\dot{w} + \dot{z}) - f(t, w + z, \epsilon) = 0, \tag{5.5}$$

$$Lz(0, \epsilon) = h \tag{5.6}$$

and

$$Rz(T, \epsilon) = k, \tag{5.7}$$

where h and k are known constant vectors, usually vanishingly small in ϵ . If $f(t, y, \epsilon)$ admits of a representation

$$f(t, w + z, \epsilon) = f(t, w, \epsilon) + f_y(t, w, \epsilon)z + Q(t, w, z, \epsilon), \tag{5.8}$$

where Q is a nonlinear term satisfying

$$\|Q\| \leq K \|z\|^{1+\mu} \tag{5.9}$$

for some $\mu > 0$, where the norms are the uniform norms, and K is a positive constant independent of ϵ , it follows that the nonlinear problem has a solution close to w if and only if there exists a solution z to

$$\epsilon \dot{z} = f_y(t, w, \epsilon)z + Q - r \tag{5.10}$$

for which $\|z\| = o(1)$ as $\epsilon \rightarrow 0$ and for which z satisfies the boundary conditions (5.6) and (5.7).

It is a straightforward process to show that if the linear part of (5.10), and hence the linear dominant of the nonlinear system, is a reasonable boundary-value problem (both with respect to the boundary conditions (5.3) and (5.4), and the inhomogeneity occurring in (5.10)), then there is such a z . The following theorem is one such result of the many possible, that may be established by the methods noted above.

THEOREM 3. *Let $w(t, \epsilon)$ be an approximate solution to the nonlinear problem (5.1), in the sense of (5.2) to (5.4), and suppose that*

$$h = \epsilon^\beta h_0, \quad k = \epsilon^\beta k_0 \tag{5.11}$$

and

$$r(t, \epsilon) = \epsilon^\gamma r_0(t, \epsilon), \tag{5.12}$$

for positive constants β and γ , and constant $O(1)$ vectors h_0 and k_0 , together with a continuous uniformly bounded (independently of ϵ) vector function r_0 . Let the assumptions of Theorem 2 regarding the matrix $A(t, \epsilon) \equiv f_y(t, w, \epsilon)$ and its eigenvalue functions apply, and assume that no eigenvalue function $\lambda_i(t, \epsilon)$ is such that $\text{Re } \zeta_i(t, \epsilon)$ has a local minimum interior to $[0, T]$.

Define, for all continuous function $\xi(t, \epsilon)$ such that $\|\xi\| < M$, where the norm is the uniform norm, and M is independent of ϵ , the vectors $p(\epsilon)$ and $q(\epsilon)$ by

$$p(\epsilon) \equiv h - \epsilon^{-1}LX(0, \epsilon)(JX^{-1}(Q(t, w, \xi, \epsilon) - r))(0, \epsilon) \tag{5.13}$$

and

$$q(\epsilon) \equiv k - \epsilon^{-1}RX(T, \epsilon)(JX^{-1}(Q(t, w, \xi, \epsilon) - r))(T, \epsilon), \tag{5.14}$$

respectively. Assume that, for M sufficiently small, $p(\epsilon)$ and $q(\epsilon)$ obey the conditions (ii) and (iii) of Theorem 2 while the equations (4.12) and (4.13) have a unique solution for each such pair p and q , that satisfies

$$|c| = O(\max\{|p|, |q|\}), \tag{5.15}$$

for small values of ϵ .

Let the constants β and γ satisfy the condition

$$\beta > \mu^{-1} \text{ and } \gamma > 1 + \mu^{-1}. \tag{5.16}$$

Then the nonlinear problem (5.1) has a unique solution $y(t, \epsilon)$ which satisfies the boundary conditions (4.2) and (4.3) and which is such that

$$y(t, \epsilon) = w(t, \epsilon) + O(\epsilon^\beta), \tag{5.17}$$

uniformly in t and ϵ .

PROOF. Using the results of Section 4, we may invert the boundary value problem for $z(t, \epsilon)$, defined above, to form the nonlinear integral equation

$$z(t, \epsilon) = XZDc + \epsilon^{-1}X\{JX^{-1}\{Q(t, w, z, \epsilon) - r\}\}(t, \epsilon), \tag{5.18}$$

where c is constructed from the boundary conditions (5.6) and (5.7). Under the assumptions of the theorem, the right-hand side of (5.18) has norm bounded by a term of the form

$$O(\epsilon^\beta) + O(\epsilon^{\alpha-1+\gamma}) + O(\epsilon^{\alpha-1})\|z\|^{1+\mu}, \tag{5.19}$$

for some α satisfying $0 < \alpha < 1$. The conditions (5.16) are sufficient to ensure that (5.19) is bounded above by a term of the form $O(1)\|z\|$ for any z lying in a ball $\|z\| = O(\epsilon^\beta)$. Thus the right-hand side of (5.19) maps this ball of functions z into itself. Moreover, this map is a contraction on this ball, as the specific form

of Q ensures. An application of the contraction mapping theorem now shows the existence of the required solution $\mathbf{z}(t, \epsilon)$ of (5.18), satisfying $\|\mathbf{z}\| = O(\epsilon^\beta)$.

As we noted above, the hypotheses of Theorem 3 are restrictive, and are sufficient only to ensure that the iterative procedure defined in the contraction mapping process should converge. There are obviously many relaxations of these conditions that could be made with a corresponding existence result being obtained. Those adopted are only significant in that, apart from giving a useful existence-uniqueness result, they ensure that in the iterative process $\mathbf{z}_n \rightarrow \mathbf{z}_{n+1}$, $n = 0, 1, 2, \dots$, defined by

$$\epsilon \dot{\mathbf{z}}_{n+1} = \mathbf{f}_y(t, \mathbf{w}, \epsilon) \mathbf{z}_{n+1} - \mathbf{r} + \mathbf{Q}(t, \mathbf{w}, \mathbf{z}_n, \epsilon), \quad (5.20)$$

together with boundary conditions (5.6) and (5.7) for \mathbf{z}_{n+1} , the (linear) boundary value problems determining each \mathbf{z}_n are reasonable in the sense of Section 4.

This constructive method of determining a solution of the nonlinear boundary value problem can also be utilized when the linear boundary value problems are *not* reasonable. However, in such circumstances, much tighter conditions are needed on the accuracy to which the trial approximation satisfies the differential equation and boundary conditions. As an illustration we consider the case where one eigenvalue function λ_i is such that the (necessarily) real ζ_i has a minimum at t_* . Using the best estimates of solutions of the linear differential equation, we have been able to establish we can show that the contraction mapping proof will be possible in a ball

$$\|\mathbf{z} - \mathbf{w}\| = O(\epsilon^\tau \exp(-\beta/\mu))$$

where, assuming $\zeta(T, \epsilon)$ is negative, $\beta = \zeta_i(T, \epsilon) - \zeta_i(t_*, \epsilon)$, and τ is any positive constant such that $\tau > \mu^{-1}(1 - \alpha)$. In order that the errors lead to a \mathbf{z} lying in this ball the errors must be such that

$$\|\mathbf{h}\|, \|\mathbf{k}\| = O(\epsilon^\tau \exp(-\beta/\mu)) \quad \text{and} \quad \|\mathbf{r}\| = O(\epsilon^\tau \exp((1 + \mu^{-1})\beta)).$$

Thus any trial approximation, which generates an unreasonable linear problem, can be justified by this means only if it can be constructed with suitable exponentially small errors. We know of no general means of generating such trial approximations as the standard heuristic arguments normally only reduce the errors to successively smaller algebraic orders. We believe that constructive methods are unlikely to be effective in such cases and we suspect that other fixed point methods will be needed. We will return briefly to this question after we have considered some simple examples in the next section. These serve to illustrate the fact that unreasonable linear problems can be associated with well-posed nonlinear boundary value problems.

6. Relevance to heuristic or numerical methods

We now consider briefly the implications of these results for practical methods of solution of the completely stiff nonlinear system

$$\epsilon \dot{\mathbf{y}} = \mathbf{f}(t, \mathbf{y}, \epsilon). \quad (6.1)$$

If any solution exhibits significant sustained oscillatory behaviour, any useful candidate approximation $\bar{\mathbf{Y}}$ must do the same and hence, in general, so will the entries of $\mathbf{f}_y(t, \bar{\mathbf{Y}}, \epsilon)$, so that none of our theory (recall the basic assumptions of Part I) is likely to be applicable. Any other solution type would have $\dot{\mathbf{y}}$ mainly small, so that candidate approximations will therefore be largely associated with the solutions of

$$\mathbf{f}(t, \mathbf{y}, \epsilon) = \mathbf{0}. \quad (6.2)$$

We would expect this association to be tempered by local failure in a number of boundary or interior layers, in which $\dot{\mathbf{y}}$ is large. At each t , equation (6.2) may be expected to define a number of possible values of the approximate solution \mathbf{Y}_k , and, by the Implicit Function Theorem [5, page 316], at any (t, \mathbf{Y}_k) at which the Jacobian derivative $\partial \mathbf{f} / \partial \mathbf{y}$ is nonsingular, we may expect, under relatively weak conditions, to be able to define continuous functions associated locally with each \mathbf{Y}_k . The possibility of obtaining one or more global solutions with the desired properties is easy to establish if this Jacobian matrix is nonsingular and the real parts of eigenvalue functions do not change sign. For, then, we may generate improved approximations to this \mathbf{Y}_k , by applying the linearization methods of Section 5, seeking small solutions for the difference $\mathbf{Y}_k - \mathbf{y}$, and thus constructing a solution to the nonlinear system (6.1). Further, if the actual boundary conditions to be imposed were of a form which could be generated within the limits of Theorems 2 or 3, we would then have established the existence of a solution or solutions of the full boundary-value problem for (6.1). It is worth noting that these assumptions regarding eigenvalues and boundary-values form the basis of most of the hypotheses involved in the current literature on this topic.

We anticipate, however, that it would be rare that any of the functions \mathbf{Y}_k so constructed would satisfy the boundary conditions imposed to the precision demanded by Theorem 2 and thus the total solution to the problem must be sought in a broader context in which genuinely nonlinear boundary or interior layers (that is, layers in which the solution changes rapidly so that our assumptions regarding $A = \mathbf{f}_y(t, \mathbf{y}_k, \epsilon)$ must almost certainly fail) may occur. What then is the role or value of the theory we have developed to this stage? We regard the theory as presented as providing the necessary basic stability theory for the outer solutions \mathbf{Y}_k as constructed above. Where the eigenvalue functions of A all

have one-signed real parts, our Theorem 3 serves to establish the forms of small linear boundary perturbations to such outer solutions which will lead to reasonable solutions, as defined earlier, over the major portion of the interval. Providing then that such approximations lie within the manifold of reasonable solutions described in Section 4, it will be possible to generate an exact solution of the system by the simple expedient of appending suitable boundary layers. Each boundary layer at t_β may thus be considered separately by assuming a local solution of the form

$$\mathbf{y}(t, \varepsilon) = \mathbf{Y}_k(t_\beta, \varepsilon) + \mathbf{u}(\tau), \quad (6.3)$$

where \mathbf{u} depends on an appropriate local scaling of t , so that we have an approximate boundary layer equation of the form

$$\varepsilon \dot{\mathbf{u}} = \mathbf{f}(0, \bar{\mathbf{Y}} + \mathbf{u}). \quad (6.4)$$

If we choose the boundary condition with the time scaled appropriately by means of $\tau = t/\varepsilon^\alpha$ for some $\alpha > 0$, we know that, if, for $\tau \rightarrow \pm\infty$, \mathbf{u} lies much outside the range of boundary conditions appropriate for reasonably posed problems, as described in Section 4, the outer solution already constructed will be unacceptable. For the differences between boundary values would be magnified exponentially within the zone where the outer solution was expected to hold, and any contraction mapping construction based on a linearization about this outer solution might be expected to fail.

Thus our results serve to provide boundary conditions for the boundary layers at each end that are of a precise and reliable nature, provided the real parts of the eigenvalue functions are one-signed or whenever the only change of sign of an eigenvalue λ_i is associated with a maximum of the function ζ_i . Then the boundary layers may be considered locally and independently. Moreover, if one *could* establish suitable trials for these boundary layer solutions, and could establish contraction proofs for the local boundary layer problems as described above, then the composite approximations given by (6.3) could probably be shown to be a genuine approximation to some solution $\mathbf{y}(t, \varepsilon)$. In such circumstances, our method provides a reliable technique for dealing with very small terms, which might play an extremely significant role, while working to only moderate accuracy.

Where, however, there are eigenvalue functions λ_i such that $\text{Re } \zeta_i$ has an interior minimum, the unreasonable nature of many linear boundary value problems does not necessarily carry over to nonlinear problems. For many unreasonable problems the difficulty is associated with a zero of some eigenvalue function of the linear operator A which we have seen is associated with the Fréchet derivative of \mathbf{f} . Thus, at points where this is singular, we can expect that different branches of the outer solutions come together or terminate. Certainly

there is no *a priori* reason to believe that, in such circumstances, the linear dominant based on the Fréchet derivative on an arbitrarily selected branch is necessarily appropriate. We believe that it is instructive to consider the properties of solutions of a few simple first-order systems as a guide to what might be expected in higher order systems. For these systems, simple geometric and continuity arguments may be used to obtain the qualitative properties of any solution. We will not give the details of such arguments, but will present our results diagrammatically. Note also that the results of these geometric demonstrations are entirely consistent with the stability-type arguments we have contemplated above.

The principal components of the arguments used relate to the sign of the slopes of the outer curves, the fact that these slopes must be large if the solution curve is not close to an "outer branch", and whether or not the solution curve will cross "outer branches". Close to the outer branch, the behaviour of the actual solutions may be deduced from our earlier results for linear systems, with the observation that this may not hold near singular points. Note that in the figures we have not used very large slopes to indicate boundary or interior layers.

EXAMPLE 1.

$$\epsilon y' = y^2 - t^2 \quad \text{on } (-a, b),$$

for positive a and b , with $a > b$. Here, possible outer solutions are $y = \pm t$ and, moreover, there is an open question as to the selection of one of these branches near $t = 0$. For the present, we will take the smooth continuations of these, so that our linearized forms about these selected outer solutions are given by

$$\epsilon \dot{x} = \pm 2tx, \tag{6.5}$$

together with the correction

$$\epsilon \dot{z} = \pm 2tz + z^2 \mp \epsilon. \tag{6.6}$$

There are two functions λ_{\pm} , given by $\pm 2t$ respectively, so that the two functions ζ_{\pm} are given by

$$\zeta_{\pm} = \pm e^{-1}(t^2 - a^2), \tag{6.7}$$

respectively. Here, the zeros of the ζ 's have been chosen so that both ζ_+ and ζ_- are non-positive on the interval.

The "positive branch" leads directly to an unreasonable linear problem and the qualitative arguments on slope show that there is no solution close to the positive branch throughout the whole interval. This demonstrates that the existence of a formal solution with a small error does not guarantee that there is a neighbouring outer solution. However, as the Jacobian matrix is singular at

$t = 0, y = 0$, it is possible to generate alternative possible outer solutions $y = |t|$ and $y = -|t|$. The coefficients in the dominating linear equation are only piecewise continuously differentiable so that minor extensions of our earlier results would be required. We may note that, for $t < 0$, the outer branch t is stable to the imposition of boundary conditions at $t = -a$ and error term integration from $t = -a$, while the branch $-t$ is stable to boundary conditions at $t = 0$, and error term integration from $t = 0$. The opposite behaviour holds for both on $t > 0$. Thus boundary values at $t = -a$ which can reach t near $t = -a$ lead to solutions following this outer solution and differing from it by a small amount, until $t = 0$ is reached. Thereafter, qualitative arguments applied to the original equation imply a switch to the branch $-t$. The end-values of these solutions are exponentially close together, as can be seen by using the values at $t = 0$ and the appropriate integral equations on $(-a, 0)$. However, whereabouts this narrow band of end-values lies can only be fixed to algebraic order in ϵ . This is why the constructive contraction mapping proof is so elusive. Elementary qualitative analysis of the full nonlinear system then reveals that the boundary layers that may be appended to yield solutions of this form must have initial values at $t = -a$ lying in the domain $y(-a) < a - \delta_1$ for some δ_1 positive and suitably chosen. In a like manner, we may construct solutions having end-values at $t = b$, and following t to $t = 0$ and there switching to $-t$, with terminating values lying in an exponentially small neighbourhood of this function. The schematic representation of these results is given in Figure 1.

By continuity arguments we predict that there are very small ranges of boundary values close to those of the reasonable solutions described above, from which solutions emanate being close to $\pm t$ on one or the other of the subintervals $(-a, 0)$ or $(0, b)$ and which fill the triangular regions shown in the figure. For clarity, these regions have been left clear in the present diagram. We note that the only admissible boundary values at $t = -a$, apart from an exponentially small set near $y = a$, are those corresponding to solutions which move rapidly to an exponentially small neighbourhood of $y = t$, which is the branch of the outer solution which is stable to small errors introduced at the left. The very small set of solutions close to $y = -t$ everywhere, for which Theorem 3 applies, delineates the domains of boundary values on left and right for which there is vastly different qualitative behaviour. We expect manifolds of such solutions to be extremely helpful in higher-dimensional problems.

EXAMPLE 2.

$$\epsilon y' = y(y^2 - 1) \quad \text{on } (-a, b).$$

We consider an example where branches terminate, so that we are faced with bifurcation phenomena.

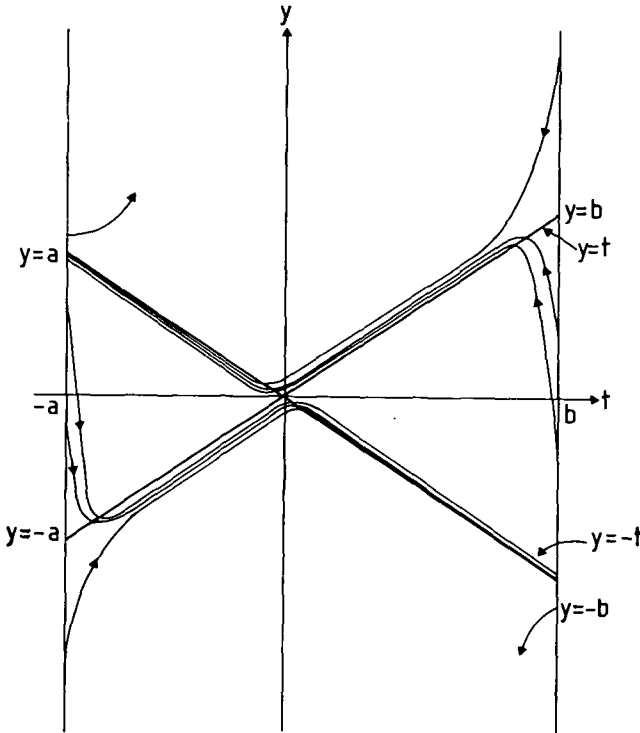


Figure 1. Solutions for $\epsilon y' = y^2 - t^2$.

On $t < 0$, there is only one possible outer solution, namely $y \equiv 0$, and the linear dominant equation is

$$\epsilon \dot{x} = -tx, \tag{6.8}$$

so that the sense of integration must be from $t = 0$, when we consider remainder terms associated with nonhomogeneous equations.

On $t > 0$, there are three possible outer branches, 0 and $\pm \sqrt{t}$. The first, $y \equiv 0$, leads to a linear system which has reasonable solutions only for integration from $t = 0$. Both the other branches, $\pm \sqrt{t}$, are stable for integration from the right, so that boundary conditions may be prescribed there. Moreover, apart from exponentially small neighbourhoods of $t = 0$, only boundary values applied at this end lead to bounded solutions. The situation is shown in Figure 2, for the upper half of the plane, the lower half being a mirror image. Almost all boundary conditions applied to the left lead to ill-posed problems.

Note that there is a marked solution $y = 0$ close to the branches which lead directly to unreasonable boundary value problems, but that there is a solution to the nonlinear system close to it. It has boundary values which form a manifold separating significantly different solutions.

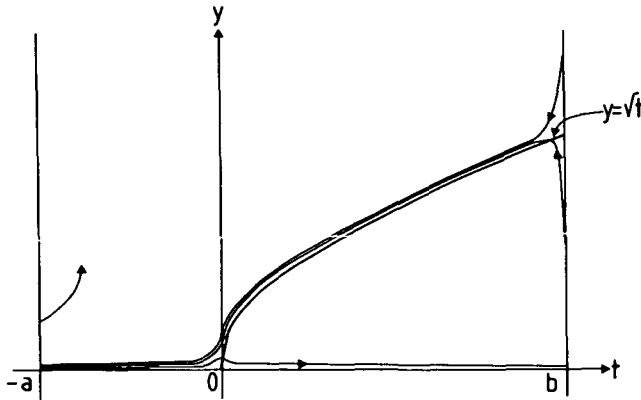


Figure 2. Solutions for $\epsilon y' = y(y^2 - 1)$.

EXAMPLE 3.

$$\epsilon y' = -t(y^2 - g^2), \text{ on } (-a, b),$$

where g is a positive function with $g' > 0$.

The possible outer solutions are

$$y \sim \pm g \tag{6.9}$$

and the linear dominant equation is

$$\epsilon \dot{x} = \mp 2tg(t)x. \tag{6.10}$$

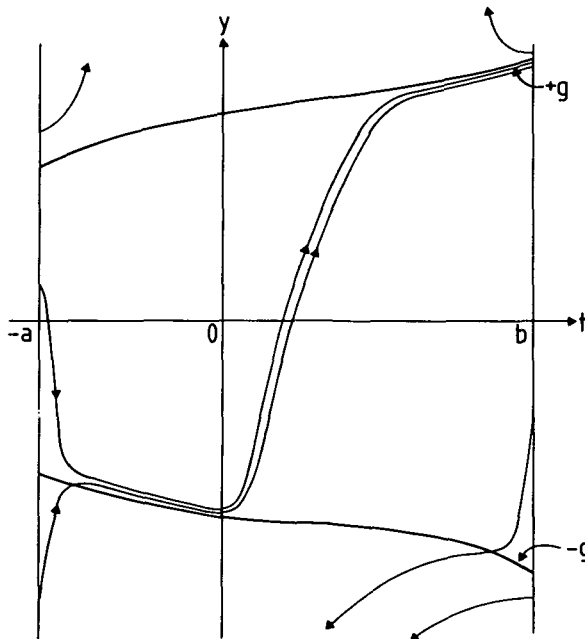


Figure 3. Solutions for $\epsilon y' = -t(y^2 - g^2)$.

The branch $y = g$ has $\text{Re } \zeta$ with a maximum, so that there is a marked solution close to $y = g$. The branch $y = -g$, for $t < 0$, is an attractor for boundary conditions at $t = -a$ such that $y(-a) < y_m(-a)$, where y_m is the marked solution. There is a small range of values exponentially close to $y_m(-a)$ that induce solutions that do not get close to $y = -g$. The solutions that have moved close to $y = -g$ in $t < 0$ transfer to a neighbourhood of $y = g$ in $t > 0$.

Careful analysis of the type used in Example 1 yields the diagram as shown in Figure 3 and from this it is easily seen that the natural place for the boundary conditions to be applied is at $t = -a$. As before, there are exponentially small ranges of boundary values, close to the reasonable solutions constructed, for which the resulting solutions fill the region $t < 0, |y| < g$.

Note that the interior layer is almost always located near $t = 0$ and it is only for the exponentially small ranges of special boundary values where this is not so.

It is striking to contrast this behaviour with that of the slightly modified problem with $g \equiv 1$. The problem can then be solved exactly and the position is then as illustrated in Figure 4.

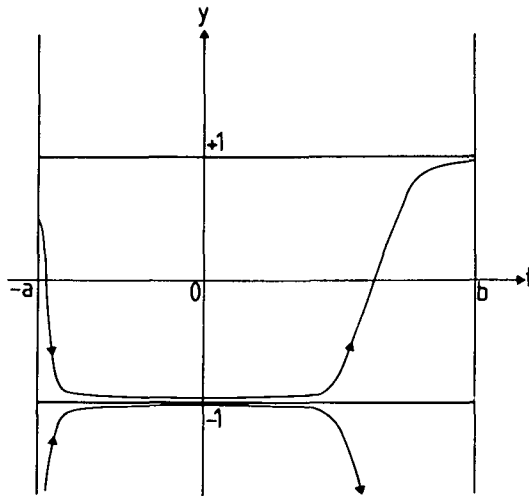


Figure 4. Solutions for $ey' = -t(y^2 - 1)$.

Here, the interior layer is not associated with the change of sign of the eigenvalue, but with the change of sign of ζ . The reason for this is that, for each branch of the outer solution (± 1 in this case), the correction equation is $\epsilon z' = \pm 2tz \mp 2tz^2$ in which there is no algebraically small error term, and so it is possible to establish the fact that z is not merely small, but is exponentially small like $\exp(-\epsilon^{-1} \int_{-a}^t 2s ds)$.

This serves to emphasize that there are marked differences of behaviour between branches in which Y , the outer solution, is constant, and branches in which the outer solution Y varies with t . Then, Y satisfies the differential equation exactly and adjustments are necessary only to cope with boundary conditions. Colloquially speaking, the system is then able to remember exponentially small terms, and these control the position of the interior layers. Where the outer solution varies, so that

$$f(t, Y_n) = -\epsilon \dot{Y}_{n-1}$$

continues to lead the algebraic order terms, the interior layers would appear to be expected to occur at the singular points of f_y .

Of course, the above arguments are merely plausible indications as to the results to be expected in general, together with some prescriptions as to methods of attack to be applied. However, we believe that they provide a more natural framework within which to work than has been available previously. They serve to indicate which branches of $f(t, Y) = 0$ should be selected, where to make the transfers, and what are the appropriate boundary conditions to apply to any boundary or interior layer. On occasions, this guide may prove to be fallible, in that it will lead to a boundary layer problem which cannot be solved. However, if it can be solved with reasonable solutions, then a contraction mapping argument ought to be usable. In general nonlinear systems, of course, we may not infer that this solution is the only solution or that it is the desired solution in any given application.

These examples serve to illustrate the fact that branches that are such that they will not allow us to construct an exact solution to the problem close to a given trial approximation with an error of algebraic order in ϵ may or may not be close to an actual solution. It is significant that, if there *is* a solution, solutions corresponding to quite closely neighbouring boundary values have markedly different behaviour. Note that this is not a serious disadvantage for the use of shooting methods in numerical integration, provided there is only one solution in this neighbourhood. The qualitatively different behaviour of neighbouring initial-value problems offers the potential for establishing a dichotomy from which the existence of at least one solution might be inferred. However, where more than one solution might be involved, as in the example

$$\epsilon^2 y'' = -y \{ g(x)^2 - y^2 \}, \quad 0 < x < 1,$$

with

$$y(0, \epsilon) = y(1, \epsilon) = 0 \quad \text{and} \quad g(x) > 0,$$

there may be significantly different solutions with neighbouring initial values. There is a positive solution which is close to $g(x)$ almost everywhere, and other

solutions with a small number of zeros lie close to $y = \pm g(x)$ almost everywhere. Since the values of $y'(0, \epsilon)$ distinguishing these solutions will be very close together, shooting methods would not seem appropriate.

There is no general reason to rule out of practical interest those boundary-value problems in which very special values of boundary data are required for the existence of a solution of a given type. The results we have obtained here, and the above examples suggest that constructive methods of the usual type are not likely to be satisfactory for existence proofs of these "special solutions". It would appear that other forms of fixed-point theorems may be more useful. One approach that may be useful is to demonstrate that there exist marked solutions of the system obtainable in such a form that

$$Q(t, w, z, \epsilon) - r = 0$$

on such appropriate subspaces as will allow the solution of the equation (5.18) together with (5.13) and (5.14) to be constructed by the usual method. As we can see, this condition will define a manifold of solutions. If we can establish the existence of a dichotomy of solutions centred on this, we may be in a position to infer an existence theorem for solutions in a suitable neighbourhood.

Such arguments cannot fix the number of solutions. However, if these solutions are close over the whole range, this will not matter, unless a count on the number of solutions is required.

Acknowledgement

This research was supported in part by the Australian Research Grants Committee.

References

- [1] R. C. Ackerberg and R. E. O'Malley, "Boundary layer problems exhibiting resonance", *Studies in Appl. Math.* 49 (1970), 277–295.
- [2] P. B. Chapman and J. J. Mahony, "Reflection of waves in a slowly varying medium", *SIAM J. Appl. Math.* 34 (1978), 303–319.
- [3] W. Eckhaus, "Formal approximations and singular perturbation", *SIAM Review* 19 (1977), 593–633.
- [4] A. E. Erdélyi, *Asymptotic expansions* (Dover, New York, 1956).
- [5] L. A. Liusternik and V. J. Sobolev, *Elements of functional analysis* (Hindustan Publishing Corp., India, 1974).

- [6] J. J. Mahony and J. J. Shepherd, "Stiff systems of ordinary differential equations. I. Completely stiff, homogeneous systems", *J. Austral. Math. Soc. B* 23 (1981), 17–51.
- [7] R. E. Meyer, "Gradual reflection of short waves", *SIAM J. Appl. Math.* 29 (1975), 481–492.

Department of Mathematics
University of Western Australia
Nedlands
Western Australia 6009