Comput. Math. Applic. Vol. 13, No. 12, pp. 925–936, 1987
Printed in Great Britain. All rights reserved

0097-4943/87 $3.00 + 0.00

UNIFORM AND OPTIMAL SCHEMES FOR STIFF INITIAL-VALUE PROBLEMS

P. A. Farrell
Department of Mathematical Sciences, Kent State University, Kent, OH 44242, U.S.A.
and Numerical Analysis Group, Trinity College, Dublin 2, Ireland

(*Received* 13 *June* 1986)

Communicated by E. Y. Rodin

Abstract—We formulate a class of difference schemes for stiff initial-value problems, with a small parameter ϵ multiplying the first derivative. We derive necessary conditions for uniform convergence with respect to the small parameter ϵ, that is the solution of the difference scheme u_i^h satisfies $|u_i^h - u(x_i)| \leqslant Ch$, where C is independent of h and ϵ. We also derive sufficient conditions for uniform convergence and show that a subclass of schemes is also optimal in the sense that $|u_i^h - u(x_i)| \leqslant C \min(h, \epsilon)$. Finally, we show that this class contains higher-order schemes.

Subject classification: primary 65L05; secondary 34E15.

1. INTRODUCTION

We consider the initial-value problem on the interval $\Omega = (0, \infty)$

$$Lu(x) \equiv \epsilon u'(x) + a(x)u(x) = f(x), \quad x \in \Omega, \tag{1a}$$

$$u(0) = A, \tag{1b}$$

where $a(x)$ and $f(x)$ are sufficiently smooth and the perturbation parameter ϵ is small and positive. In addition we assume

$$a(x) \geqslant \mathbf{a} \geqslant 0,$$

which is sufficient to guarantee that the operator L has a maximum principle and that the solution $u(x)$ of expressions (1a, b) is unique and bounded. We shall consider the following class of difference schemes:

$$L^h u_i^h \equiv \epsilon_i^h D_+ u_i^h + a_i^h u_{i+1}^h = f_i^h, \tag{2a}$$

$$u_0^h = A, \tag{2b}$$

where

$$D_+ u_i^h = (u_{i+1}^h - u_i^h)/h_i,$$

$$h_i = x_{i+1} - x_i,$$

$$\epsilon_i^h \geqslant 0 \tag{2c}$$

and

$$a_i^h \geqslant \mathbf{a}^h > 0, \tag{2d}$$

and in addition we may, for convenience, write

$$\epsilon_i^h = \epsilon \sigma_i^h \tag{3a}$$

and

$$a_i^h = \alpha_i a(x_{i+1}), \tag{3b}$$

where σ_i^h and σ_i are bounded and hence ϵ_i^h and a_i^h are bounded.

We wish to derive conditions on expressions (2a–d) necessary for uniform convergence, i.e. such that

$$|u_i^h - u(x_i)| \leqslant C \max_{0 \leqslant j \leqslant i} h_j^p,$$

where C and p are independent of i, h_i and ϵ and $p > 0$. We will also derive conditions sufficient for a class of difference schemes to be uniformly convergent.

For boundary-value problems, without turning points, uniform convergence is a strict criteria which is satisfied by only a small number of schemes. These schemes are the (exponentially) fitted schemes, which model the boundary-layer behaviour accurately. Other schemes, on a uniform mesh, either suffer instability or, due to artificial viscosity, introduce diffusion of the boundary layers. In the case of initial-value problems the same effects may be noted. The traditional approach is to refine the mesh in the area of the initial layer and, as we shall mention later, for a general solver, this would still be necessary when using a fitted scheme. However, one can derive a better intuitive understanding of the criteria necessary for a scheme by considering the concept of uniform convergence for initial-value problems also. The conditions for uniform convergence, which we derive, specify essentially that the transient behaviour must be modeled accurately. It is for this reason that the fitting technique is not easily applicable to a general solver for non-linear problems. The initial layers, in this case, may differ significantly in behaviour, as is noted in O'Reilly [1, 2]. The schemes proposed there are not of the same essential form as those proposed here. Nevertheless, one may expect that the closer a general scheme corresponds to the correctly fitted scheme for a problem the more accurate it will be on an arbitrary grid.

Another problem with applying the concept of uniform convergence to initial-value problems is that, although it requires that one model the transient behaviour well, it is not necessarily strict enough a criterion for behaviour outside the initial layer. Thus, as a fitted scheme may be more accurate initially but less accurate for large x. To impose a stricter criteria the concept of optimality was introduced [3]. A scheme is optimal and of order p if

$$|u_i^h - u(x_i)| \leqslant C \min(\max_{0 \leqslant j \leqslant i} h_j^p, \epsilon).$$

We shall consider a subclass of schemes deriving additional necessary and sufficient conditions for optimality. We shall also show that higher-order optimal schemes exist. Intuitively these conditions require additionally that the scheme should model the solution of the reduced equation sufficiently well.

Throughout the paper $\rho_i = h_i/\epsilon$ and C will denote a generic constant independent of i, h_i and ϵ.

Uniformly convergent schemes for this problem have been proposed by Doolan *et al.* [3], Carroll [4–6] and Miller [7]. Non-linear initial-value problems have been considered in Carroll [5] and O'Reilly [1, 2]. We are not interested in proposing further schemes in this paper, rather we wish to show what properties all these schemes have in common. It will be obvious later that a large class of schemes can be proposed which will satisfy the sufficient conditions. Since we propose the weakest possible conditions the error bound will not be the best possible for many schemes. However, it will be the best obtainable for a general scheme of the form (2a–d), that is, there is at least one scheme of that form which is only $O(h)$ uniformly convergent.

2. ANALYTIC RESULTS

In this section we collect some results concerning the solution of problem (1a, b). The first of these show that the solution satisfies a maximum principle and hence is uniformly stable. The second is a technical lemma and the last gives a first-order asymptotic expansion for the solution.

Lemma 2.1

(a) If $v_0(x)$ satisfies $v(0) \geqslant 0$ and $Lv_0(x) \geqslant 0$ for $x \in \Omega$ then $v_0(x) \geqslant 0$, $\forall\ x \in \Omega$.

(b) If $v_0(x)$ is the solution of problem (1) then

$$|u(x)| \leqslant \frac{1}{\alpha} \max_{0 \leqslant y \leqslant x} |f(y)|, \quad x \in \Omega.$$

Proof: Doolan *et al.*, Lemma 2.1 and 2.2 [3].

Lemma 2.2

Let $w(x)$ be a smooth function such that,

$$|(Lw(x))^{(i)}| \leqslant C[1+\epsilon^{-i}\exp(-\mathbf{a}x/\epsilon)], \quad \text{for } i \geqslant 0,$$

then

$$|(w^{(i)}(x)| \leqslant C[1+\epsilon^{-i}\exp(-\mathbf{a}x/\epsilon)], \quad \text{for } i \geqslant 0.$$

Proof: Doolan *et al.*, Lemma 7.1 [3].

The problem (1a, b) is a singlularly perturbed equation with an initial layer at $x = 0$. We may define the corresponding reduced solution $u_0(x)$ by

$$L_0u_0(x) \equiv a(x)u_0(x) = f(x), \quad x \in \Omega.$$

It is clear that the solution is

$$u_0(x) = \frac{f(x)}{a(x)}. \tag{4}$$

We may derive an asymptotic expansion for the solution $u(x)$ and a bound on the remainder and its derivatives as follows:

Lemma 2.3

The solution of problem (1a, b) may be written as

$$u(x) = u_0(x) + v_0(x) + \epsilon z(x), \tag{5}$$

where

$$v_0(x) = \gamma \exp[-a(0)x/\epsilon], \quad \gamma = \phi - u_0(0), \tag{6}$$

is the boundary layer function, $u_0(x)$ is given by equation (4), and

$$|z^{(i)}(x)| \leqslant C[1+\epsilon^{-i}\exp(-\mathbf{a}x/\epsilon)]. \tag{7}$$

Proof: Using, $Lu(x) = f(x)$, $a(x)u_0(x) = L_0u_0(x) = f(x)$, and the explicit form of $v_0(x)$, we get

$$\begin{aligned}\epsilon Lz(x) &= Lu(x) - Lu_0(x) - Lv_0(x)\\ &= f(x) - \epsilon u_0'(x) - a(x)u_0(x) - \epsilon v_0'(x) - a(x)v_0(x)\\ &= -\epsilon u_0'(x) - [a(x) - a(0)]v_0(x).\end{aligned}$$

Thus, using $a(x) - a(0) = xa'(\zeta)$ for $0 \leqslant \zeta \leqslant x$,

$$Lz(x) = -u_0'(x) - a'(\zeta)x/\epsilon v_0(x)$$

and thus, from the expression for $v_0(x)$,

$$|(Lz)^{(i)}| \leqslant C[1+\epsilon^{-i}\exp(-\mathbf{a}x/2\epsilon)].$$

Hence, by Lemma 2.2,

$$|z^{(i)}(x)| \leqslant C[1+\epsilon^{-i}\exp(-\mathbf{a}x/2\epsilon)].$$

3. NECESSARY CONDITIONS

We wish now to derive necessary conditions for the solution of the scheme (2a–d) to converge uniformly to the solution of problem (1a, b).

Assume that the solution of the difference schemes (2a–2d) exists $\forall\, h \leqslant H$ and $\forall\, \epsilon \leqslant \epsilon_0$ and that equations (3a, b) hold, $\lim \sigma_i^h$, $\lim a_i^h$ and $\lim f_i^h$ exist and

$$\lim f_i^h = \lim f(x_i)$$

and

$$\lim a_i^h = \lim a(x_i)$$

where, throughout this section, lim denotes the limit as $h_i \to 0$, $\rho_i = h_i/\epsilon \neq 0$ fixed and i fixed. With this definition,

$$\lim f(x_i) = f(0) \quad \text{and} \quad \lim a(x_i) = a(0)$$

and hence taking the limit of expressions (2a–d),

$$\lim \left[\frac{\sigma_i^h}{\rho_i}(u_{i+1}^h - u_i^h) + a_i^h u_{i+1}^h \right] = \lim f_i^h$$

becomes

$$\lim \left[\frac{\sigma_i^h}{\rho_i}(u_{i+1}^h - u_i^h) + a(0) u_{i+1}^h \right] = f(0). \tag{8}$$

Now, assuming uniform convergence, i.e. $\lim u_i^h = \lim u(x_i)$, and using Lemma 2.2,

$$\lim u_i^h = \lim u(x_i) = \lim [u_0(x_i) + v_0(x_i) + \epsilon z(x_i)]$$

$$= u_0(0) + \gamma \exp[-a(0)x_i/\epsilon] = \frac{f(0)}{a(0)} + \gamma \exp[-a(0)x_i/\epsilon]$$

and hence (8) becomes

$$\lim \left[\frac{\sigma_i^h}{\rho_i}(\exp[-a(0)\rho_i] - 1) + a(0)\exp[-a(0)\rho_i] \right] \gamma \exp[-a(0)x_i/\epsilon] = 0$$

or

$$\lim \sigma_i^h = \frac{a(0)\rho_i}{\exp[-a(0)\rho_i] - 1} = \sigma(a(0)\rho_i), \tag{9}$$

where

$$\sigma(x) = \frac{x}{e^x - 1} \tag{10}$$

is the generating function for the Bernoulli numbers.

It should be noted that, if we choose σ_i^h to satisfy condition (9) exactly, the difference scheme will be exact for a homogeneous constant coefficient problem. The necessary condition gives a minimum requirement on the scheme to model the transient behaviour of the problem accurately. It is in this sense that one refers to the scheme as *fitted*. Condition (9) is only an asymptotic (limiting) condition for h approaching 0. It is interesting to determine how closely this must be satisfied for finitely large values of h in order to obtain uniform convergence. We shall consider this question in the next section.

Notes:

(i) A minor modification of the above proof would yield

$$\lim \sigma_i^h = \alpha_i \sigma(a(0)\rho_i). \tag{11}$$

(ii) It is possible to show that $\sigma(x)$ has the following properties. These can be useful in verifying that schemes satisfy the necessary conditions.

(1) It is clear that

$$\sigma(-x) = \sigma(x) + x. \tag{12}$$

This can be useful in rewriting schemes in the form (2a–d).

(2) By continuation, we define

$$\sigma(0) = \lim_{x \to 0} \sigma(x) = 1. \tag{13}$$

(3) In addition if we write $\tau(x) = x \coth x$ then

$$\tau(x) = \tfrac{1}{2}[\sigma(2x) + \sigma(-2x)] = \tfrac{1}{2}[2\sigma(2x) + 2x] = \sigma(2x) + x. \tag{14}$$

(4) $$|\sigma(x_1)-\sigma(x_2)|\leqslant C|x_1-x_2| \tag{15}$$

w.l.o.g. assume $x_2>x_1$ then

$$\sigma(x_1)-\sigma(x_2)=(x_2-x_1)\frac{\mathrm{d}\sigma(z)}{\mathrm{d}x},\quad x_1\leqslant z\leqslant x_2.$$

It suffices to show

$$\left|\frac{\mathrm{d}\sigma(x)}{\mathrm{d}x}\right|=\left|\frac{(1-x)\mathrm{e}^x-1}{(\mathrm{e}^x-1)^2}\right|\leqslant C. \tag{16}$$

Since the only singularities of this function are at $-\infty$, 0 and ∞ we need only show it is bounded there. It is easy to verify

$$\lim_{x\to-\infty}\frac{\mathrm{d}\sigma(x)}{\mathrm{d}x}=-1,\quad \lim_{x\to 0}\frac{\mathrm{d}\sigma(x)}{\mathrm{d}x}=-\tfrac{1}{2},\quad \lim_{x\to\infty}\frac{\mathrm{d}\sigma(x)}{\mathrm{d}x}=0.$$

(5) If $x_2>x_1\geqslant 0$, then there exists some z, $x_1\leqslant z\leqslant x_2$, such that

$$|\sigma(x_1)-\sigma(x_2)|\leqslant C|x_2-x_1|(1+z)\mathrm{e}^{-z}.$$

This follows in the same manner as expression (15), except that one must show that

$$\left|\frac{\mathrm{d}\sigma(z)}{\mathrm{d}z}\frac{1}{\mathrm{e}^{-z}(1+z)}\right|\leqslant C,\quad \forall z\leqslant 0.$$

4. SUFFICIENT CONDITIONS

In this section we derive conditions sufficient for uniform convergence. First we shall consider the analogue of Lemma 2.1 for the difference scheme, which will establish a discrete maximum principle for L^h and also uniform stability for schemes (2a–d).

Lemma 4.1

(a) If v_i satisfies $v_0\geqslant 0$ and $L^h v_i\geqslant 0$ for $i>0$, then $v_i\geqslant 0$, for $i>0$.

(b) If u_i^h is the solution of schemes (2a–d) then

$$|u_i^h|\leqslant|u_0^h|+C\max_{0\leqslant j\leqslant i}|f_i^h|,\quad \text{for}\quad i>0. \tag{17}$$

Proof:

(a) Suppose that $v_k<0$ some k, hence there exists a

$$j\quad\text{such that}\quad v_{j+1}<0,\ v_i\geqslant 0,\quad \text{for}\quad i\leqslant j. \tag{18}$$

Then

$$L^h v_j=\left(-\frac{\epsilon_j^h}{h_j}\right)v_j+\left(-\frac{\epsilon_j^h}{h_j}+a_j^h\right)v_{j+1}<0,$$

using expressions (2a–d), (17) and (18). This is a contradiction and hence $v_i\geqslant 0\forall i$.

(b) Let

$$w_i=|u_0^h|+\frac{1}{\mathbf{a}^h}\max_{0\leqslant j\leqslant i}|L^h u_j^h|\pm u_i^h.$$

Then

$$w_0=|u_0^h|+\frac{1}{\mathbf{a}^h}\max_{0\leqslant j\leqslant i}|L^h u_j^h|\pm u_0^h\geqslant 0$$

and, using expression (2d),

$$L^h w_i = a_i^h\left(|u_0^h| + \frac{1}{\mathbf{a}^h}\max_{0\leqslant j\leqslant i}|L^h u_j^h|\right) \pm L^h u_i^h \geqslant \mathbf{a}^h\left(|u_0^h| + \frac{1}{\mathbf{a}^h}\max_{0\leqslant j\leqslant i}|L^h u_j^h|\right) \pm L^h u_i^h \geqslant 0.$$

The result now follows using proof (a) and $L^h u_i^h = f_i^h$.

We are now in a position to state the main theorem of this paper giving sufficient conditions for uniform convergence. We will defer the proof until later.

Theorem 4.1

Let u_i^h be the solution of schemes (2a–d), $u(x)$ be the solution of problem (1a, b). If, for given $\bar{x}$, some $0 \leqslant \eta_i \leqslant x_i + Ch_i$,

$$|\sigma_i^h - \sigma(a(\eta_i)\rho_i)| \leqslant C\rho_i p(\rho_i) x_i \exp[-a(\zeta)\rho_i], \quad \text{for} \quad 0 \leqslant x_i \leqslant \bar{x}, \tag{Ia}$$

$$|\epsilon_i^h - \epsilon| \leqslant Ch_i, \quad \text{for} \quad \bar{x} \leqslant x_i, \tag{Ib}$$

and

$$|a_i^h - a(x_{i+1})| \leqslant Ch_i, \tag{II}$$

$$|f_i^h - f(x_{i+1})| \leqslant Ch_i, \tag{III}$$

where $p(\rho_i)$ is a polynomial in ρ_i, $0 \leqslant \zeta \leqslant x_i + Ch_i$ and $h_i \leqslant \bar{x}$, then

$$|u_i^h - u(x_i)| \leqslant C \max_{0\leqslant j\leqslant i} h_j.$$

Notes on conditions (Ia, b)–(III). Conditions (II) and (III) are simply (uniform) consistency conditions as is condition (Ib). Condition (Ia) is a stronger condition which ensures that in the initial layer the solution of the difference scheme represents the rapidly varying component of the solution of expressions (Ia, b) adequately. It allows us to vary significantly from the asymptotic value of σ_i^h specified in expression (9).

In order to prove Theorem 4.1 we first rewrite the truncation error using Lemma 2.3,

$$\begin{aligned} L^h(u(x_i) - u_i^h) &= L^h u(x_i) - f_i^h = L^h u(x_i) - Lu(x_{i+1}) - f(x_{i+1}) - f_i^h \\ &= L^h u_0(x_i) - Lu_0(x_{i+1}) + L^h v(x_i) - Lv(x_{i+1}) + \epsilon(L^h z(x_i) - Lz(x_{i+1})) + f(x_{i+1}) - f_i^h. \end{aligned} \tag{19}$$

Rewriting the truncation error in this form will simplify the later proofs. It is also the natural choice if we consider the form of the difference scheme. We shall now proceed to bound the former sets of terms separately.

Lemma 4.2

Under the conditions of Theorem 4.1

$$|L^h v(x_i) - Lv(x_{i+1})| \leqslant C \min(h_i, \epsilon).$$

Proof: By adding terms which are each zero we may rewrite this as follows:

$$\begin{aligned} L^h v(x_i) - Lv(x_{i+1}) &= \epsilon_i^h D_+ v(x_i) + a_i^h v(x_{i+1}) - \epsilon\sigma(a(0)\rho_i) D_+ v(x_i) + a(0) v(x_{i+1}) + \epsilon v'(x_{i+1}) \\ &\quad + a(0) v(x_{i+1}) - \epsilon v'(x_{i+1}) + a(x_{i+1}) v(x_{i+1}) \\ &= [\epsilon_i^h - \epsilon\sigma(a(0)\rho_i)] D_+ v(x_i) + [a_i^h - a(x_{i+1})] v(x_{i+1}). \end{aligned} \tag{20}$$

Let us consider the first group of terms. In the region where condition (Ia) holds

$$\sigma_i^h = \sigma(a(\eta_i)\rho_i) + R_i,$$

where

$$|R_i| \leqslant C\rho_i p(\rho_i) x_i \exp[-a(\zeta)\rho_i].$$

Thus,

$$[\epsilon_i^h - \epsilon\sigma(a(0)\rho_i)] D_+ v(x_i) = \epsilon[\sigma(a(\eta_i)\rho_i) - \sigma(a(0)\rho_i)] D_+ v(x_i) + \epsilon R_i.$$

Now, using expression (15),

$$|\epsilon[\sigma(a(\eta_i)\rho_i)-\sigma(a(0)\rho_i)]D_+v(x_i)|\leqslant C\epsilon\eta_i\rho_i p_i(\rho_i)\exp[-a(\eta)\rho_i]\frac{\min(1,\rho_i)}{h_i}v(x_i)$$
$$\leqslant Cx_i\min(1,\rho_i)\exp[-a(\zeta)\rho_i]v(x_i)\leqslant C\min(h_i,\epsilon).$$

Finally,

$$|\epsilon R_i|\leqslant C\epsilon\rho_i p(\rho_i)x_i\exp[-a(\zeta)\rho_i]\leqslant C\epsilon\rho_i\exp[-a(\zeta)\rho_i]\leqslant C\min(h_i,\epsilon).$$

This is the required bound in the region where condition (Ia) holds.

In the region where condition (Ib) holds, since $x_i\geqslant\bar{x}$,

$$|\epsilon_i^h-\epsilon\sigma(a(0)\rho_i)|\,|D_+v(x_i)|=[|\epsilon_i^h-\epsilon|+|\epsilon\{\sigma(a(0)\rho_i)-1\}|]|D_+v(x_i)|$$
$$\leqslant Ch_i+\epsilon\rho_i\frac{\min(1,\rho_i)}{h_i}v(x_i)\leqslant C\min(1,\rho_i)\exp[-a(0)\bar{x}/\epsilon]\leqslant C\min(h_i,\epsilon).$$

The latter follows since $x_i\geqslant\bar{x}>0$ and if $h_i/\epsilon=\rho_i\leqslant 1$, then

$$\min(1,\rho_i)\exp[-a(0)x_i/\epsilon]\leqslant\rho_i\exp[-a(0)\bar{x}/\epsilon]\leqslant\frac{h_i}{a(0)\bar{x}}\frac{a(0)\bar{x}}{\epsilon}\exp[-a(0)\bar{x}/\epsilon]\leqslant Ch_i$$

and, if $\rho_i>1$, then

$$\min(1,\rho_i)\exp[-a(0)x_i/\epsilon]\leqslant\exp[-a(0)\bar{x}/\epsilon]\leqslant\frac{\epsilon}{a(0)\bar{x}}\frac{a(0)\bar{x}}{\epsilon}\exp[-a(0)\bar{x}/\epsilon]\leqslant C\epsilon.$$

Now consider the second set of terms in expression (20)

$$|\{a_i^h-a(x_{i+1})\}v(x_{i+1})|\leqslant Ch_i\exp[-a(0)x_i/\epsilon]\leqslant C\epsilon\frac{h_i}{\epsilon}\exp[-a(0)\bar{x}/\epsilon]\leqslant C\min(h_i,\epsilon),$$

since $h_i/\epsilon\leqslant x_i/\epsilon$ and $x_i/\epsilon\exp[-a(0)x_i/\epsilon]\leqslant C$. This concludes the proof of Lemma 4.2. □

Let us now consider the first term in expression (19). In order to conveniently bound these we first prove the following technical lemma.

Lemma 4.3

If conditions (Ia) and (II) are true then

$$|\epsilon_i^h-\epsilon|\leqslant C\min(h_i,\epsilon).$$

Proof:

$$|\epsilon_i^h-\epsilon|=|\epsilon(\sigma_i^h-1)|\leqslant\epsilon|\sigma_i^h-\sigma(a(\eta_i)\rho_i)|+\epsilon|\sigma(a(\eta_i)\rho_i)-1|$$
$$\leqslant C\epsilon\rho_i p(\rho_i)x_i\exp[-a(\zeta)\rho_i]+C\epsilon\min(1,\rho_i)\leqslant C\min(h_i,\epsilon).$$

Using this lemma we can now prove Lemma 4.4.

Lemma 4.4

Let

$$y(x)=u_0(x)+\epsilon z(x)$$

then, under the conditions of Theorem 4.2,

$$|L^hy(x_i)-Ly(x_{i+1})|\leqslant Ch_i.$$

Proof: Since, $|y^{(i)}(x)|=|u_0^{(i)}(x)|+\epsilon|z^{(i)}(x)|\leqslant C[1+\epsilon^{-i+1}\exp(-\mathbf{a}x/\epsilon)]$,

$$|L^hy(x_i)-Ly(x_{i+1})|\leqslant|\epsilon_i^hD_+y(x_i)-a_i^hy(x_{i+1})-\epsilon y'(x_{i+1})-a(x_{i+1})y(x_{i+1})|$$
$$\leqslant|\epsilon_i^hD_+y(x_i)-\epsilon y'(x_{i+1})|+|a_i^h-a(x_{i+1})|\,|y(x_{i+1})|$$
$$\leqslant|(\epsilon_i^h-\epsilon)y'(x_{i+1})|+|\epsilon_i^h[D_+y(x_i)-y'(x_{i+1})]|+|a_i^h-a(x_{i+1})|\,|y(x_{i+1})|$$
$$\leqslant C|\epsilon_i^h-\epsilon|+C\epsilon h_i|y''(\zeta)|+C|a_i^h-a(x_{i+1})|, \tag{21}$$

where $x_i \leqslant \zeta \leqslant x_{i+1}$,

$$\leqslant Ch_i + C\epsilon h_i\left(1 + \frac{1}{\epsilon}\exp(-\mathbf{a}\zeta/\epsilon)\right) \leqslant Ch_i + C\min(h_i, \epsilon) \leqslant Ch_i.$$

Proof of Theorem 4.1: Substituting (III) and the results of Lemma 4.2 and 4.4 in expression (19) gives

$$|L^h(u(x_i) - u_i^h)| \leqslant Ch_i$$

and, using Lemma 4.1,

$$|u(x_i) - u_i^h| \leqslant C \max_{0 \leqslant j \leqslant i} h_i.$$

We have thus shown that Theorem 4.1 gives sufficient conditions for uniform convergence. As we remarked earlier these conditions are quite general and are satisfied by a large number of schemes which have been proposed in the literature.

5. OPTIMAL CONVERGENCE

In order to show the stronger condition of optimality holds, we require stronger constraints on the coefficients. To guide us in the formulation of these constraints we derive the following necessary conditions for optimal convergence.

Let us assume that σ_i^h, a_i^h and f_i^h are bounded and consider the limit as $\epsilon \to 0$, for h_i and $i > 1$ fixed, which we denote by lim. Then

$$\lim u_i^h = u_\epsilon(x_i) = u_0(x_i) + v_0(x_i) + \epsilon z(x_i) = u_0(x_i).$$

Now, substituting this in the scheme,

$$\lim\left[\epsilon\sigma_i^h \frac{u_{i+1}^h - u_i^h}{h_i} + a_i^h u_{i+1}^h\right] = \lim f_i^h$$

we obtain

$$\lim a_i^h u_0(x_{i+1}) = \lim f_i^h$$

or, using expression (4),

$$\lim\left[f_i^h - a_i^h \frac{f(x_{i+1})}{a(x_{i+1})}\right] = 0. \tag{22}$$

This additional condition implies we should solve the reduced equation exactly and suggests that we should impose a condition which in the limit gives expression (22). Thus we get the following sufficient conditions for optimal convergence.

Theorem 5.1

Let u_i^h be the solution of

$$L^h u_i^h \equiv \epsilon_i^h D_+ u_i^h + a_i^h u_{i+1}^h = f_i^h \tag{23}$$

and $u(x)$ be the solution of problem (1a, b) then if, for given $\bar{x}$, some $0 \leqslant \eta_i \leqslant x_i + Ch_i$,

$$|\sigma_i^h - \sigma(a(\eta_i)\rho_i)| \leqslant C\rho_i p(\rho_i) x_i \exp[-a(\zeta)\rho_i], \quad \text{for} \quad 0 \leqslant x_i \leqslant \bar{x}, \tag{Ia}$$

$$|\epsilon_i^h - \epsilon| \leqslant C\min(h_i, \epsilon), \text{for} \quad \bar{x} \leqslant x_i \tag{Ib}$$

and

$$|a_i^h - a(x_{i+1})| \leqslant Ch_i, \tag{II}$$

$$\left|f_i^h - a_i^h \frac{f(x_{i+1})}{a(x_{i+1})}\right| \leqslant C\min(h_i, \epsilon). \tag{III}$$

where $p(\rho_i)$ is a polynomial in ρ_i, $0 \leqslant \zeta \leqslant x_i + Ch_i$ and $h_i \leqslant \bar{x}$, then

$$|u_i^h - u(x_i)| \leqslant C \min(\max_{0 \leqslant j \leqslant i} h_j^p, \epsilon).$$

It is clear that Lemma 4.2 suffices for the truncation error in $v_0(x)$. However to prove the theorem, we must refine the argument in Lemma 4.4. To accomplish this we consider $u_0(x)$ and $z(x)$ separately.

Lemma 5.1

Under the conditions of Theorem 5.1,

$$\epsilon\,|L^h z(x_i) - Lz(x_{i+1})| \leqslant C \min(h_i, \epsilon).$$

Proof: Using condition (Ia), Lemma 4.3, conditions (Ib), (II), (7) and $(x_{i+1}) \geqslant (x_i) \geqslant h_i$,

$$\begin{aligned}\epsilon\,|L^h z(x_i) - Lz(x_{i+1})| &\leqslant \epsilon\,|\epsilon_i^h D_+ z(x_i) - a_i^h z(x_{i+1}) - \epsilon z'(x_{i+1}) - a(x_{i+1})z(x_{i+1})|\\ &\leqslant \epsilon\,[|\epsilon_i^h D_+ z(x_i) - \epsilon z'(x_{i+1})| + |a_i^h - a(x_{i+1})|\,|z(x_{i+1})|]\\ &\leqslant \epsilon\,\{|(\epsilon_i^h - \epsilon)z'(x_{i+1})| + |\epsilon_i^h[D_+ z(x_i) - z'(x_{i+1})]|\} + C\epsilon h_i\\ &\leqslant C\epsilon\,|\epsilon_i^h - \epsilon\,|(1 + \epsilon^{-1}\exp[-\mathbf{a}(x_{i+1})/\epsilon]) + C\epsilon^2 h_i|z''(\zeta)| + C\epsilon h_i,\end{aligned}$$

where $x_i \leqslant \zeta \leqslant x_{i+1}$,

$$\begin{aligned}&\leqslant C\epsilon h_i\{1 + \epsilon^{-1}\exp[-\mathbf{a}(x_{i+1})/\epsilon]\} + C\epsilon^2 h_i\{1 + \epsilon^{-2}\exp[-\mathbf{a}(x_{i+1})/\epsilon]\}\\ &\leqslant C\epsilon h_i + C\epsilon\left(\frac{h_i}{\epsilon}\exp(-\mathbf{a}h_i/\epsilon)\right) \leqslant C \min(h_i, \epsilon).\end{aligned}$$

Of the terms in expression (19) there remains only

$$L^h u_0(x_i) - Lu_0(x_{i+1}) + f(x_{i+1}) - f_i^h.$$

We bound these in the following lemma.

Lemma 5.2

Under the assumptions of Theorem 5.1,

$$|L^h u_0(x_i) - Lu_0(x_{i+1}) + f(x_{i+1}) - f_i^h| \leqslant C \min(h_i, \epsilon).$$

Proof: Using, $u_0(x_{i+1}) = f(x_{i+1})/a(x_{i+1})$,

$$\begin{aligned}&|L^h u_0(x_i) - Lu_0(x_{i+1}) + f(x_{i+1}) - f_i^h|\\ &= |\epsilon_i^h D_+ u_0(x_i) - \epsilon u_0'(x_i) + a_i^h u_0(x_{i+1}) - a(x_{i+1})u_0(x_{i+1}) - f(x_{i+1}) - f_i^h|\\ &= |\epsilon_i^h D_+ u_0(x_i) - \epsilon u_0'(x_i)| + \left|a_i^h \frac{f(x_{i+1})}{a(x_{i+1})} - f_i^h\right|\\ &= |\epsilon_i^h - \epsilon\,|\,|D_+ u_0(x_i)| + \epsilon\,|D_+ u_0(x_i) - u_0'(x_i)| + C \min(h_i, \epsilon)\\ &= C \min(h_i, \epsilon) + C\epsilon h_i|u_0''(\zeta)| + C \min(h_i, \epsilon)\\ &= C \min(h_i, \epsilon).\end{aligned}$$

Combining Lemma's 4.2, 5.1 and 5.2 and, using Lemma 4.1, gives the result in Theorem 5.1.

6. SCHEMES WHICH SATISFY THESE CONDITIONS

A number of schemes have been proposed in the literature for initial-value problems. Doolan *et al.* [3], for example, proposed the following schemes:

$$\epsilon\sigma(-a(x_i)\rho_i)D_+ u_i^h + a(x_i)u_i^h = f(x_i), \tag{24}$$

$$\begin{aligned}\epsilon[\theta\sigma(-a(x_i)\rho_i) + (1-\theta)\sigma(a(x_i))\rho_i]D_+ u_i^h + a(x_i)[\theta u_i^h + (1-\theta)u_{i+1}^h]&\\ = \theta f(x_i) + (1-\theta)f(x_{i+1})&\end{aligned} \tag{25}$$

$$\epsilon\left[\frac{a(x_i)\rho_i}{2}\coth\frac{a(x_i)\rho_i}{2}\right]D_+u_i^h+\tfrac{1}{2}a(x_i)[u_i^h+u_{i+1}^h]=f(x_i), \tag{26}$$

$$\epsilon\sigma(a(0)\rho_i)D_+u_i^h+a(x_i)u_{i+1}^h=f(x_i), \tag{27}$$

$$\epsilon\sigma(a(0)\rho_i)D_+u_i^h+a(x_{i+1})u_{i+1}^h=f(x_{i+1}), \tag{28}$$

$$\epsilon\sigma(-a(x_{i+1})\rho_i)D_+u_i^h+a(x_{i+1})u_i^h=f(x_{i+1}). \tag{29}$$

It is immediately obvious that expressions (27) and (28) satisfy the conditions of Theorem 4.1. To see that expressions (24), (29) and (25) do also, we use equations (12) and (14) to rewrite them in the form (2a–d). using equation (12), we can rewrite equation (24) as

$$\epsilon[\sigma(a(x_i)\rho_i)+a(x_i)\rho_i]D_+u_i^h+a(x_i)u_i^h=f(x_i).$$

This is equivalent to

$$\epsilon\sigma(a(x_i)\rho_i)D_+u_i^h+a(x_i)u_{i+1}^h=f(x_i) \tag{30}$$

which clearly satisfies the conditions. The same argument holds also for expression (29).

Similarly, using expression (12), the scheme (25) can be transformed to

$$\epsilon\{\theta\sigma(a(x_i)\rho_i)+(1-\theta)\sigma(a(x_i))\rho_i\}D_+u_i^h+\theta\frac{a(x_i)h_i}{\epsilon}\frac{\epsilon}{h_i}(u_{i+1}^h-u_i^h)$$
$$+a(x_i)[\theta u_i^h+(1-\theta)u_{i+1}^h]=\theta f(x_i)+(1-\theta)f(x_{i+1})$$

which reduces to

$$\epsilon\sigma(a(x_i)\rho_i)D_+u_i^h+a(x_i)u_{i+1}^h=\theta f(x_i)+(1-\theta)f(x_{i+1}).$$

To transform (26), we use expression (14), giving

$$\epsilon\sigma(a(x_i)\rho_i)D_+u_i^h+\frac{\epsilon a(x_i)h_i}{2\epsilon}\frac{u_{i+1}^h-u_i^h}{h_i}+\frac{a(x_i)}{2}[u_i^h+u_{i+1}^h]=f(x_i)$$

which also reduces to expression (30). Thus all of these schemes satisfy the sufficient conditions in Theorem 4.1.

Having written the schemes in the form (2a–d), it is clear that schemes (24)–(27) do not satisfy the extra condition required for optimality by Theorem 5.1. However, for schemes (28) and (29) we have

$$a_i^h=a(x_{i+1}),\quad f_i^h=f(x_{i+1}).$$

Hence

$$f_i^h-a_i^h\frac{f(x_{i+1})}{a(x_{i+1})}=0,$$

which shows that these do satisfy this condition and hence are optimal.

A more interesting result is that a scheme (Example 9.1 of Doolan *et al.* [3]), which was shown to be $O(h^2)$ uniformly convergent there, can be shown to be optimal also. It thus answers the speculation as to whether there exist higher-order optimal schemes. This scheme is given by

$$\epsilon\sigma(-\tilde{a}_i\rho_i)D_+u_i^h+\tilde{a}_iu_i^h=f_i^h, \tag{31a}$$

where

$$\tilde{a}_i=[a(x_i)+a(x_{i+1})]/2 \tag{31b}$$

and

$$f_i^h=\frac{f(x_i)}{\rho_ia(x_i)}[1-\sigma(\tilde{a}_i\rho_i)]+\frac{f(x_{i+1})}{\rho_ia(x_{i+1})}[\sigma(-\tilde{a}_i\rho_i)-1]. \tag{31c}$$

We may rewrite expression (31a) in the form (2a–d), using the same method as for scheme (24).

Conditions (I) and (II) of Theorem 5.1 are thus satisfied. It remains to show that condition (III) is also satisfied. To show this we proceed as follows:

$$\begin{aligned} f_i^h - a_i^h\frac{f(x_{i+1})}{a(x_{i+1})} &= \frac{f(x_i)}{\rho_i a(x_i)}[1-\sigma(\tilde{a}_i\rho_i)] + \frac{f(x_{i+1})}{\rho_i a(x_{i+1})}[\sigma(-\tilde{a}_i\rho_i)-1] - \tilde{a}_i\frac{f(x_{i+1})}{a(x_{i+1})} \\ &= \frac{f(x_i)}{\rho_i a(x_i)}[1-\sigma(\tilde{a}_i\rho_i)] + \frac{f(x_{i+1})}{\rho_i a(x_{i+1})}[\sigma(-\tilde{a}_i\rho_i)-\tilde{a}_i\rho_i-1] \\ &= \frac{\epsilon}{h_i}[1-\sigma(\tilde{a}_i\rho_i)]\left[\frac{f(x_i)}{a(x_i)}-\frac{f(x_{i+1})}{a(x_{i+1})}\right], \end{aligned}$$

the latter, using expression (12) and regrouping. Thus, again using $|\sigma(y)-1| \leqslant \min(y, 1)$,

$$\left|f_i^h - a_i^h\frac{f(x_{i+1})}{a(x_{i+1})}\right| = \left|\epsilon[1-\sigma(\tilde{a}_i\rho_i)]D_+\left[\frac{f(x_i)}{a(x_i)}\right]\right| \leqslant C\epsilon\min(\rho_i, 1) = C\min(\mathrm{h}_i, \epsilon),$$

since $f(x)$ and $a(x)$ are sufficiently continuous and $a(x) \geqslant \mathbf{a} > 0$. This concludes the proof that scheme (31) satisfies the conditions of Theorem 5.1 and hence is an optimal $O(h^2)$ scheme.

We remark that all these schemes satisfy condition (Ia) on the whole interval of solution of the problem. This is unnecessarily restrictive. One need only satisfy condition (Ia) near the initial layer and switch to any standard scheme, which may be cast in the form (2a–d) and satisfies condition (Ib), once this region has been traversed.

Finally, to confirm that these bounds are attained in practice we consider the problem

$$u'(x) = -\lambda(u(x)-g(x)) + g'(x), \quad 0 \leqslant x \leqslant 10,$$

where

$$u(0) = 10, \quad g(x) = 10-(10+x)\exp(-x).$$

Tables 1–3 show the maximum absolute error at the nodes, for the backward Euler method, Trapezoidal method and for a number of fitted methods, for values of $\lambda = 20$, 200 and 10000 (corresponding to values of ϵ of 0.05, 0.005 and 0.00001).

Table 1. $\lambda = 20$ ($\epsilon = 0.05$)

Scheme	$h = 1/8$	$h = 1/16$	$h = 1/32$	$h = 1/64$
Backward Euler	2.02	1.57	9.18E − 1	5.05E − 1
Trapezoidal	1.93	5.57E − 1	1.22E − 1	3.06E − 2
Scheme (27)	6.52E − 1	2.82E − 1	1.29E − 1	6.13E − 2
Scheme (28)	2.82E − 1	1.83E − 1	1.04E − 1	5.49E − 2
Scheme [25 ($\theta = 0.5$)]	1.93E − 1	6.92E − 2	6.10E − 2	7.50E − 2
Scheme (31)	7.91E − 3	2.15E − 3	7.70E − 4	2.07E − 4

Table 2. $\lambda = 200$ ($\epsilon = 0.005$)

Scheme	$h = 1/8$	$h = 1/16$	$h = 1/32$	$h = 1/64$
Backward Euler	3.83E − 1	7.40E − 1	1.36	1.98
Trapezoidal	8.52	7.24	5.17	2.63
Scheme (27)	1.02	5.02E − 1	2.32E − 1	9.95E − 2
Scheme (28)	4.02E − 2	4.25E − 2	4.31E − 2	3.73E − 2
Scheme [25 ($\theta = 0.5$)]	4.90E − 1	2.30E − 1	9.48E − 2	3.46E − 2
Scheme (31)	2.13E − 3	1.01E − 3	4.18E − 4	1.38E − 4

Table 3. $\lambda = 10000$ ($\epsilon = 0.00001$)

Scheme	$h = 1/8$	$h = 1/16$	$h = 1/32$	$h = 1/64$
Backward Euler	7.95E − 4	1.59E − 3	3.19E − 3	6.39E − 3
Trapezoidal	9.99	19.7	27.7	28.4
Scheme (27)	1.06	5.47E − 1	2.77E − 1	1.39E − 1
Scheme (28)	8.05E − 5	8.50E − 5	8.75E − 5	8.87E − 5
Scheme [25 ($\theta = 0.5$)]	5.32E − 1	2.73E − 1	1.39E − 1	6.97E − 2
Scheme (31)	4.63E − 6	2.39E − 6	1.20E − 6	9.53E − 7

Schemes (24) and (26) do not differ from expression (27), nor does expression (29) from expression (28), since $a(x) = 1$ is a constant. If $a(x)$ is not a constant the solutions still do not differ significantly. The results show that, as predicted:

(1) schemes (24), (26) and (27) which satisfy the conditions of Theorem 4.1 but not of 5.1 are $O(h)$ uniformly convergent;

(2) schemes (28) and (29) are $O(\min(h, \epsilon))$ uniformly convergent since for $\epsilon = 0.00001$ the error is approx. $O(10^{-5})$ even for $h = 1/8$;

(3) scheme (25) for $\theta = 0.5$ shows peculiar characteristics since for this problem it is a fitted variation of a trapezoidal scheme;

(4) scheme (31) is uniformly $O(\min(h^2, \epsilon))$ since again for $\epsilon = 0.00001$ the error is approx. $O(10^{-5})$, whereas for $\epsilon = 0.05$ the error is $O(h^2)$.

It can also be noted that all the fitted schemes perform better than backward Euler or the Trapezoidal Rule.

REFERENCES

1. M. J. O'Reilly, On uniformly convergent finite difference methods for non-linear singular perturbation problems. Ph.D. Thesis, Trinity College, Dublin (1983).
2. M. J. O'Reilly, A uniformly convergent finite difference scheme for the singularly perturbed Riccati equation. In *Proc. BAIL III Conf Computational and Asymptotic Methods for Boundary and Initial Layers.* Boole Press, Dublin (1984).
3. E. P. Doolan, J. J. H. Miller and W. H. A. Schilders, *Uniform Numerical Methods for Problems with Initial and Boundary Layers.* Boole Press, Dublin (1980).
4. J. Carroll, On the implementation of exponentially fitted one-step methods for the numerical integration of stiff linear initial value problems. In *Proc. BAIL II Conf. on Computational and Asymptotic Methods for Boundary and Initial Layers.* Boole Press, Dublin (1982).
5. J. Carroll, Exponentially fitted one-step methods for the numerical integration of some stiff initial value problems. Ph.D. Thesis, Trinity College, Dublin (1983).
6. J. Carroll, A uniformly convergent exponentially fitted DIRK scheme. In *Proc. BAIL III Conference, on Computational and Asymptotic Methods for Boundary and Initial Layers.* Boole Press, Dublin (1984).
7. J. J. H. Miller, Optimal uniform difference schemes for linear initial value problems, *Comput. Math. Applic.* (in press).