

## Generalized OCI Schemes for Boundary Layer Problems\*

By Alan E. Berger, Jay M. Solomon, Melvyn Ciment,  
Stephen H. Leventhal and Bernard C. Weinberg

**Abstract.** A family of tridiagonal formally fourth-order difference schemes is developed for a class of singular perturbation problems. These schemes have no cell Reynolds number limitation and satisfy a discrete maximum principle. Error estimates and numerical results for this family of methods are given, and are compared with those for several other schemes.

**I. Introduction.** Mathematical models of diffusion-convection phenomena generally involve one or more spatial operators of the form

$$(1.1) \quad Lu = \epsilon u_{xx} + bu_x - du.$$

In many applications, the diffusion coefficient  $\epsilon$  is much smaller than the convection coefficient  $b$ ; for example, steady and unsteady viscous flow problems with large Reynolds numbers, and convective heat transport problems with large Peclet numbers. In this paper we consider difference approximations for  $Lu$  that are suitable when  $\epsilon$  is a parameter in  $(0, 1]$ , and  $b$  and  $d$  are smooth functions of  $x$  with  $b$  positive and  $d$  nonnegative in  $[0, 1]$ . Some of the difficulties encountered in these applications are typified by the singular perturbation problem

$$(1.2) \quad Lu = f(x) \quad \text{for } x \in (0, 1), \quad u(0) = \alpha_0, \quad u(1) = \alpha_1,$$

where  $f$  is smooth and  $\alpha_0$  and  $\alpha_1$  are given constants. With the assumptions made on  $b$  and  $d$ , it is well known that as  $\epsilon \rightarrow 0$  (1.2) has no turning points, and the solution of (1.2) exhibits a boundary layer adjacent to  $x = 0$ , e.g. [26]. Furthermore (1.2) possesses a maximum principle [23, p. 6]. Although we will concentrate on the treatment of (1.2), we are interested in schemes that can be applied as well to initial boundary value problems for  $au_t = Lu + f$ , and to nonlinear problems.

It has long been recognized that difficulties can arise when certain "centered" finite-difference and finite-element methods are applied to (1.2) when  $\epsilon$  is small. In particular, as pointed out in, e.g. [5], [6], [12], [13], [19], [25], such schemes when applied to (1.2) on a uniform grid have an inherent *formal cell Reynolds number limitation*. Namely, with a uniform mesh length  $h$ ,  $d = f = 0$ , and  $b$  constant, one

---

Received May 25, 1979.

1980 *Mathematics Subject Classification*. Primary 65L10, 65M05, 65M10; Secondary 34E15.

\* This work was supported jointly by the NSWC Independent Research Fund and NAVSEA and ONR and NBS.

finds that the *cell Reynolds number*  $bh/\epsilon$  must be bounded by some constant depending on the scheme in order to avoid spurious oscillations or gross inaccuracies. For small  $\epsilon$  this requires a prohibitive number of grid points, and so alternative approaches have been developed. One approach is to use a nonuniform mesh (which must be appropriately chosen) which is very fine “in the boundary layer” and coarser elsewhere, e.g. [4], [7], [18], [22]. Another approach has been to devise schemes which have no formal cell Reynolds number limitation. Schemes of this type have been constructed by using noncentered (“upwind”) differencing for the first derivative term, or, more generally, by adding an “artificial viscosity” to the diffusion coefficient  $\epsilon$ , e.g., [1], [5], [7], [10], [11], [13], [28], [29].

In this paper we confine our attention to the case of a uniform mesh, and develop and analyze a family of formally fourth-order accurate finite-difference representations for  $Lu$  having no formal cell Reynolds number limitation. We consider approximations to (1.1) on a uniform mesh  $x_j = jh$  ( $h = 1/J$ ,  $J$  a positive integer,  $j = 0, 1, \dots, J$ ) having the tridiagonal form

$$(1.3) \quad \frac{\epsilon}{h^2} (r_j^- u_{j-1} + r_j^c u_j + r_j^+ u_{j+1}) = q_j^-(Lu)_{j-1} + q_j^c(Lu)_j + q_j^+(Lu)_{j+1},$$

where  $u_j$  and  $(Lu)_j$  are approximations to  $u(x_j)$  and  $Lu(x_j)$ , respectively, (e.g., in the corresponding difference equation for (1.2),  $Lu_j$  is  $f_j$ ). This representation for  $Lu$  is said to be *explicit* when  $q_j^- = q_j^+ = 0$ , and *implicit* otherwise. Following the terminology of [6], a scheme of the form (1.3) will be referred to as an *operator compact implicit* (OCI) scheme if it is a formally fourth-order accurate representation of  $Lu$  (i.e., if its appropriately scaled truncation error is  $O(h^4)$  for fixed  $\epsilon$ ). Note that formal fourth-order accuracy is the highest that can be obtained by a scheme of the form (1.3) (cf. Remark 2.1 below). Such an OCI scheme, which we will refer to as the *standard OCI scheme*, was originally developed in the context (1.3) by Swartz [27]; this scheme is given in (2.6) below. For a recent review of some of the literature on higher-order (i.e., fourth-order) three-point finite-difference methods, we refer to [6]. There, it is shown that this standard OCI scheme, and various other higher-order schemes all have formal cell Reynolds number limitations.

Here, a new family of OCI schemes which generalizes the standard OCI scheme is obtained using the Taylor development of the truncation error for (1.3). These schemes are *polynomial schemes*, meaning that  $r_j^{-,c,+}$  (i.e.,  $r_j^-$ ,  $r_j^c$ ,  $r_j^+$ ) and  $q_j^{-,c,+}$  in (1.3) are polynomials in  $z \equiv h/\epsilon$ , the coefficients of the powers of  $z$  depending only on  $b_{j-1}$ ,  $b_j$ ,  $b_{j+1}$ ,  $d_{j-1}$ ,  $d_j$ ,  $d_{j+1}$ , and  $h$  (e.g., the well-known centered and upwind schemes for (1.2) are all polynomial schemes). The form of polynomial schemes is such that they can be conveniently implemented with standard iterative procedures when, for example,  $b$ ,  $d$ , and  $f$  depend on  $u$  (e.g., [6]). Our family of OCI schemes is chosen so that when applied to (1.2), they have no formal cell Reynolds number limitation; in particular, the resulting tridiagonal system of difference equations is diagonally dominant and satisfies a maximum principle corresponding to that satisfied by (1.2). In addition, these schemes are selected so that, when applied to initial-boundary value problems for  $au_t = Lu + f$  (by setting  $Lu = au_t - f$  in (1.3)) using, e.g., Crank-Nicolson time

discretization, the resulting method is unconditionally stable in the sense of von Neumann [24] for all  $\epsilon$  in  $(0, 1]$ . These schemes also admit an ADI factorization under appropriate conditions; see [6] and Section II below. This family is defined in Theorem 3.7 (the reader may well wish to read Theorem 3.7 and the paragraph following it before going through the preceding lengthy algebraic arguments and motivation).

These OCI schemes, while formally fourth order, approach a formally second-order accurate "upwind" scheme for the *reduced operator*  $\tilde{L}u \equiv bu_x - du$  as  $\epsilon \rightarrow 0$  with  $h$  fixed (i.e., as  $z = h/\epsilon$  becomes large). This phenomenon of reduction of formal order of accuracy in the limit  $\epsilon \rightarrow 0$  has been observed in [16], where it is demonstrated that explicit tridiagonal schemes satisfying certain (desirable) properties have at most formal *first*-order accuracy as  $\epsilon$  goes to 0. In Theorem 4.11 we show that any scheme of the form (1.3) satisfying certain properties has at most formal *second*-order accuracy as  $\epsilon$  goes to 0, and so the above-mentioned behavior of these OCI schemes is not unexpected. At the other limit, as  $z = h/\epsilon \rightarrow 0$ , these schemes approach the formally fourth-order accurate Störmer-Numerov scheme for  $eu_{xx} = f$  (i.e.,  $r_j^- = r_j^+ = 1$ ,  $r_j^c = -2$ ,  $q_j^- = q_j^+ = 1/12$ ,  $q_j^c = 10/12$ ), and so these schemes may indeed be viewed as *automatically* shifting their form from the higher-order accurate Störmer-Numerov scheme to a formally second-order accurate upwind scheme as  $z$  becomes large.

A complete error analysis for the full range of values of  $\epsilon$  will be given for our family of OCI schemes applied to (1.2); see Section IV. These error estimates are obtained using the comparison function approach of Kellogg and Tsan [16]. The error analysis demonstrates a fourth-order rate of convergence when  $\epsilon$  is fixed (i.e., the error in the approximate solution of (1.2) at all the grid points is bounded by  $C(\epsilon)h^4$  for some constant  $C(\epsilon)$  depending on  $\epsilon$  but not on  $h$ ). Furthermore, "away from the boundary layer," these schemes achieve  $O(h^2)$  convergence (uniformly for all  $\epsilon$  in  $(0, 1]$ ), and  $O(h^4/\epsilon^2)$  convergence when  $h \leq \epsilon$ . These schemes are  $O(h^2)$  at *all* the grid points when  $\epsilon$  is sufficiently small relative to  $h$  (i.e.  $\epsilon = O(h^2)$  for all these OCI schemes, and  $\epsilon = O(h^{3/2})$  for one particular OCI scheme). However, when  $\epsilon$  and  $h$  are of the same order of magnitude (i.e., when  $z \approx 1$ ), the error in the approximate solution of (1.2) "near the edge of the boundary layer" is  $O(1)$ .

The latter  $O(1)$  behavior is the case for all the standard polynomial schemes (e.g., the centered and upwind schemes, and the usual finite-element methods). This can be seen by directly comparing the exact solution with the finite-difference solution for the special case of (1.2) with  $u(0) = 1$ ,  $u(1) = 0$ ,  $f = d = 0$ , and  $b$  constant; cf. (2.12). The rough idea is that when  $h/\epsilon$  is a constant, this  $O(1)$  behavior will occur unless  $r_j^-/r_j^+$  becomes  $\exp(-bh/\epsilon)$  as  $h \rightarrow 0$ . Indeed, Miller [20] has shown that uniform convergence for any positive order (i.e., for some  $\delta > 0$  the error at all the mesh points is bounded by  $ch^\delta$  with  $c$  and  $\delta$  independent of  $h$  and  $\epsilon$ ) can be obtained only by schemes that incorporate an appropriate exponential character into their coefficients  $r_j^{-,c,+}$ .

An explicit scheme of this type was given some time ago in [2]. The proof of uniform first-order convergence for this scheme applied to (1.2) is of more recent vintage [14], [16], [21]. An implicit scheme of this type has been obtained in [9]

using a modification of the approach of [2], cf. also the references in [9]. Recently, this implicit scheme has been shown to be uniformly second-order convergent when applied to (1.2) when  $d = 0$  [3] (a statement of this result is given here in Theorem 4.10). As pointed out earlier, second order is the highest uniform order of convergence that one would expect to be able to obtain with a scheme of the form (1.3) satisfying certain properties (see Theorem 4.11), and thus the scheme [9] is optimal in this sense. It remains an open question if a (formally fourth-order) OCI scheme exists which is uniformly second-order convergent for (1.2). In some circumstances it may thus be advantageous to use a formally fourth-order scheme such as those developed here, while under other conditions it may be preferable to use a uniformly convergent scheme such as given in [9]; cf. the numerical results in Section V.

In the next section, basic notation is given and criteria for choosing desirable difference approximations to  $L$  having the form (1.3) are discussed. In Section III, polynomial OCI schemes are examined, and then a specific family of such schemes is given, and is shown to satisfy the criteria of Section II; see Theorem 3.7 and the paragraph following it. Using the approach of Kellogg and Tsan [16], a complete error analysis is given in Section IV for these schemes when applied to (1.2). The error behavior for the generalized OCI schemes is also compared with that of other schemes for (1.2). In the final section, numerical results for the OCI and several other schemes are given.

**II. Implicit Tridiagonal Finite-Difference Methods.**

2.1. *Notation and Preliminaries.* Consider a uniform mesh  $x_j = jh$  ( $j = 0, 1, 2, \dots, J$ ;  $J$  a positive integer) where  $h = 1/J$  is the mesh length. By definition, a tridiagonal finite-difference scheme for  $Lu$  given in (1.1) has the general form

$$(2.1) \quad \frac{\epsilon}{h^2} R(U_j) = Q(LU_j) \quad (j = 1, 2, \dots, J - 1),$$

where  $R$  and  $Q$  are tridiagonal operators defined by

$$(2.2a) \quad R(U_j) \equiv r_j^- U_{j-1} + r_j^c U_j + r_j^+ U_{j+1},$$

$$(2.2b) \quad Q(U_j) \equiv q_j^- U_{j-1} + q_j^c U_j + q_j^+ U_{j+1}.$$

Here and throughout the paper,  $U_j$  and  $LU_j$  are approximations of  $u(x_j)$  and  $Lu(x_j)$ , respectively. The coefficients  $r_j^{-,c,+}$ ,  $q_j^{-,c,+}$  are to be functions of  $b_{j-1}$ ,  $b_j$ ,  $b_{j+1}$ ,  $d_{j-1}$ ,  $d_j$ ,  $d_{j+1}$ ,  $h$ , and  $\epsilon$  ( $b_j = b(x_j)$ , etc.). Note that of the six coefficients in (2.2) one must be regarded as a multiplicative normalizing factor which has no effect on (2.1). Without loss of generality, throughout this paper (2.1) is assumed normalized so that for all  $j = 1, 2, \dots, J - 1$ ,

$$(2.3) \quad q_j^c \rightarrow \text{a positive constant as } h \rightarrow 0 \quad (\epsilon \text{ fixed}).$$

This normalization corresponds to that of standard explicit schemes for  $Lu$ ; e.g., the representation of  $Lu$  obtained using central difference approximations to  $u_{xx}$  and  $u_x$  which is given by (2.1) with  $r_j^+ = 1 + zb_j/2$ ,  $r_j^c = -(2 + hd_jz)$ ,  $r_j^- = 1 - zb_j/2$ ,  $q_j^- = q_j^+ = 0$ ,  $q_j^c = 1$  where  $z = h/\epsilon$ . Throughout the paper,  $\bar{R}$  will denote the  $(J - 1)$  by  $(J - 1)$  tridiagonal matrix  $\bar{R}_{ij} = \delta_{j+1}^i r_{j+1}^- + \delta_j^i r_j^c + \delta_{j-1}^i r_{j-1}^+$  ( $\delta_j^j = 1$ ;  $\delta_j^i = 0$ ,  $i \neq j$ ) and

$\bar{Q}$  is similarly defined in terms of  $Q$ . For simplicity of notation, the  $j$  index dependence of the coefficients  $r_j^-, q_j^-$ , etc. will be omitted when convenient.

The local truncation error  $\tau_j$  of (2.1) is defined in the usual way; i.e.,

$$(2.4) \quad \tau_j \equiv \frac{\epsilon}{h^2} R(u(x_j)) - Q(Lu(x_j)), \quad j = 1, \dots, J - 1.$$

For  $u(x)$  sufficiently smooth, the standard Taylor development of  $\tau_j$  for  $\epsilon$  fixed is given by

$$(2.5a) \quad \tau_j = T_j^0 u(x_j) + T_j^1 u^{(1)}(x_j) + \dots + T_j^6 u^{(6)}(x_j) + O(h^5),$$

where

$$(2.5b) \quad T_j^0 = \frac{\epsilon}{h^2} [r_j^+ + r_j^- + r_j^c + hz(q_j^+ d_{j+1} + q_j^c d_j + q_j^- d_{j-1})],$$

$$(2.5c) \quad T_j^1 = \frac{\epsilon}{h} \{r_j^+ - r_j^- - z[q_j^+ b_{j+1} + q_j^c b_j + q_j^- b_{j-1} - h(q_j^+ d_{j+1} - q_j^- d_{j-1})]\},$$

$$(2.5d) \quad T_j^2 = \frac{\epsilon}{2} \{(r_j^+ + r_j^-) - 2(q_j^+ + q_j^c + q_j^-) \\ - z[2(q_j^+ b_{j+1} - q_j^- b_{j-1}) - h(q_j^+ d_{j+1} + q_j^- d_{j-1})]\},$$

$$(2.5e) \quad T_j^\nu = \frac{\epsilon h^{\nu-2}}{\nu!} \{r_j^+ + (-1)^\nu r_j^- - \nu(\nu - 1)[q_j^+ + (-1)^\nu q_j^-] \\ - z[\nu(q_j^+ b_{j+1} + (-1)^{\nu-1} q_j^- b_{j-1}) \\ - h(q_j^+ d_{j+1} + (-1)^\nu q_j^- d_{j-1})]\} \quad (\nu = 3, 4, 5, 6).$$

The truncation error of (2.1) is said to be *formally of order p* if  $\tau_j = O(h^p)$  as  $h \rightarrow 0$  ( $\epsilon$  fixed) for  $j = 1, 2, \dots, J - 1$ . The normalization (2.3) makes the formal order of  $\tau_j$  well defined and directly comparable with that of well-known explicit methods; for example, the explicit centered difference method given earlier has, for fixed  $\epsilon$ ,  $\tau_j = O(h^2)$  as can be seen by direct substitution into (2.5). The following result will be needed in the next section.

*Remark 2.1.* Let (2.1) be normalized according to (2.3). Then  $\tau_j$  has formal order no larger than four. Furthermore, for  $\epsilon$  fixed, a necessary and sufficient condition for  $\tau_j = O(h^4)$  is that  $T_j^\nu = O(h^4)$  for  $\nu = 0, 1, 2, 3, 4$ .

*Proof.* The result follows directly from (2.5) with  $\epsilon$  fixed. For the first part, note that if  $T_j^2 = O(h^4)$  and  $T_j^4 = O(h^4)$  it follows from (2.5d) and (2.5e) that  $T_j^6 = \epsilon h^4 q_j^c / 200 + O(h^5)$ . Because of the normalization (2.3),  $T_j^6$  and hence  $\tau_j$  cannot be of order greater than 4. Turning to the second part of the remark, necessity is obvious. To prove sufficiency one need only show that  $T_j^5 = O(h^4)$ . Note that  $T_j^1 = O(h^4)$  and  $T_j^3 = O(h^4)$  imply that  $r_j^+ - r_j^- = O(h)$  and  $q_j^+ - q_j^- = O(h)$  which implies that  $T_j^5 = O(h^4)$ .

The standard OCI scheme [27] is defined by choosing  $R$  and  $Q$  such that  $T_j^\nu = 0$ ,  $\nu = 0, 1, 2, 3, 4$ . These five equations uniquely define this scheme (modulo a multiplicative factor that has no effect) which is given by

$$\begin{aligned}
 q_j^- &= 6 + (2b_{j+1} - 5b_j)z - b_j b_{j+1} z^2, \\
 (2.6a) \quad q_j^c &= 60 + 16(b_{j+1} - b_{j-1})z - 4b_{j-1} b_{j+1} z^2, \\
 q_j^+ &= 6 + (5b_j - 2b_{j-1})z - b_{j-1} b_j z^2; \\
 r_j^c &= -r_j^- - r_j^+ - hz(q_j^+ d_{j+1} + q_j^c d_j + q_j^- d_{j-1}), \\
 (2.6b) \quad 2r_j^- &= q_j^-(2 - 3zb_{j-1}) + q_j^c(2 - zb_j) + q_j^+(2 + zb_{j+1}) - 2hzq_j^- d_{j-1}, \\
 2r_j^+ &= q_j^-(2 - zb_{j-1}) + q_j^c(2 + zb_j) + q_j^+(2 + 3zb_{j+1}) - 2hzq_j^+ d_{j+1}.
 \end{aligned}$$

Note that (2.6b), which defines  $R$  in terms of  $Q$ , follows solely from  $T_j^0 = T_j^1 = T_j^2 = 0$ . The above OCI scheme was derived from approximation theory considerations by Swartz [27]. Later the corresponding formally third-order scheme for nonuniform meshes was obtained in [6] using a straightforward Taylor series development. In Section III, we will show that a multi-parameter family of OCI schemes can be produced by allowing  $T_j^3$  and  $T_j^4$  to be  $O(h^4)$  rather than zero. First we develop a set of criteria for selecting appropriate  $R$  and  $Q$ .

**2.2. Criteria for  $R$  and  $Q$ .** We consider two typical applications for schemes of the type (2.1). Our objective will be to formulate conditions on the operators  $R$  and  $Q$  which will insure that the resulting numerical methods for these applications will be well behaved.

The first application is the two-point boundary value problem

$$(2.7) \quad Lu = f(x), \quad x \in (0, 1); \quad u(0) = \alpha_0, \quad u(1) = \alpha_1,$$

where (as in Section I)  $\alpha_0$  and  $\alpha_1$  are given constants,  $b(x)$ ,  $d(x)$ , and  $f(x)$  are smooth on  $[0, 1]$ ,  $\epsilon$  is in  $(0, 1]$ , and  $d \geq 0$  and  $b > 0$  on  $[0, 1]$ . The application of (2.1) to the problem (2.7) results in

$$(2.8) \quad \frac{\epsilon}{h^2} R(U_j) = Q(f_j), \quad j = 1, \dots, J-1; \quad U_0 = \alpha_0, \quad U_J = \alpha_1.$$

Obviously one requirement is that the tridiagonal matrix  $\bar{R}$  associated with  $R$  must be invertible. Also, since a fundamental property of (2.7) is its maximum principle, it is natural and desirable to require that (2.8) possesses an analogous discrete maximum principle (i.e., if  $\alpha_0 \leq 0$ ,  $\alpha_1 \leq 0$ , and each  $f_j \geq 0$  then each  $U_j \leq 0$ ). The following conditions are sufficient to insure that these requirements are met:

$$(2.9) \quad r_j^- \geq 0, \quad r_j^+ > 0, \quad -r_j^c \geq r_j^- + r_j^+ \quad (j = 1, \dots, J-1);$$

$$(2.10) \quad q_j^- \geq 0, \quad q_j^+ \geq 0, \quad q_j^c > 0 \quad (j = 1, \dots, J-1).$$

Note that (2.9) implies that  $\bar{R}$  can be inverted by simple tridiagonal Gaussian decomposition; cf. [15, p.56]. Furthermore, the discrete maximum principle follows from (2.10) and

*Remark 2.2.* If (2.9) holds,  $U_0 \leq 0$ ,  $U_J \leq 0$ , and  $R(U_j) \geq 0$  for  $j = 1, 2, \dots, J-1$  then each  $U_j \leq 0$ .

*Proof.* If not, then there must be a  $k$  such that  $U_k = \max_j(U_j) > 0$ . It then follows from (2.9) that  $U_k = U_{k+1} = \dots = U_J \leq 0$ , a contradiction.

As suggested in [6], some indication of the correctness of the solution to (2.8) when  $\epsilon$  is small can be obtained by considering the special case when  $b(x) = b$ ,  $d(x) = d$  ( $b$  and  $d$  constants), and  $f(x) \equiv 0$ . For this case, the solution of (2.7) at  $x = x_j = jh$  is

$$(2.11) \quad u(x_j) = c_1 e^{x_j \alpha_-} + c_2 e^{x_j \alpha_+} = c_1 (e^{h \alpha_-})^j + c_2 (e^{h \alpha_+})^j,$$

where  $c_1$  and  $c_2$  are constants and

$$\alpha_{\pm} = -\frac{b}{2\epsilon} \left[ 1 \pm \left( 1 + \frac{4\epsilon d}{b^2} \right)^{1/2} \right].$$

Since  $b$  and  $d$  are constants,  $r_j^-$ ,  $r_j^+$ , and  $r_j^c$  are independent of  $j$ . Hence, the solution of (2.8) with  $f = 0$  is easily obtained as

$$(2.12) \quad U_j = c_3 (\mu_-)^j + c_4 (\mu_+)^j; \quad \mu_{\pm} = \frac{-r^c}{2r^+} \left[ 1 \pm \left( 1 - \frac{4r^+ r^-}{(r^c)^2} \right)^{1/2} \right],$$

where  $c_3$  and  $c_4$  are constants. Note that since  $\alpha_+ \geq 0$  and  $\alpha_- < 0$ ,  $U_j$  properly approximates the analytic solution (2.11) if and only if

$$(2.13) \quad 0 \leq \mu_- < 1 \quad \text{and} \quad \mu_+ \geq 1.$$

These conditions, if not always satisfied by a given scheme, are the source of the so-called formal cell Reynolds number limitation on  $h$ . For example, in [6] it was shown that for  $d = 0$  the standard OCI scheme satisfies (2.13) only when  $bh/\epsilon < 2\sqrt{3}$ . Although this severely limits  $h$  when  $\epsilon$  is small, it is an improvement over that of the explicit centered scheme which satisfies (2.13) only when  $bh/\epsilon \leq 2$ . Note that for (2.13) to hold, it is sufficient to require (2.9) and (since  $\mu_- \mu_+ = r^-/r^+$ )

$$(2.14) \quad r^+ > r^-.$$

In the following, (2.14) will be invoked locally at each mesh point  $j = 1, \dots, J - 1$ , when  $b$  and  $d$  are not constants.

We turn now to the application of schemes (2.1) to parabolic equations of the form

$$(2.15) \quad a(x, t) u_t - Lu = f(x, t).$$

Here  $a(x, t) \geq A_0 > 0$  and the coefficients  $b$  and  $d$  in  $Lu$  are functions of  $(x, t)$  which for fixed  $t$  satisfy the conditions imposed earlier in this section. For equations of this type, Ciment, Leventhal, and Weinberg [6] developed techniques which use the standard OCI scheme, to approximate the spatial operator, with a Crank-Nicolson or a three-level Lees type time integration method. Although we consider here only one space dimension, certain equations with two space dimensions can be considered using a splitting method which results in an ADI (alternating direction implicit) factorization; see [6] for details.

Following [6],  $LU_j$  in (2.1) is formally replaced by  $(au_t - f)_j$  and an appropriate

approximation for  $u_t$  is employed. We illustrate by applying a Crank-Nicolson approximation; i.e., for each  $j$

$$\frac{\epsilon}{2h^2} R^{n+\frac{1}{2}}(U_j^{n+1} + U_j^n) = Q^{n+\frac{1}{2}} \left[ \left( \frac{a_j^{n+\frac{1}{2}}}{\Delta t} \right) (U_j^{n+1} - U_j^n) - f_j^{n+\frac{1}{2}} \right],$$

where the  $n$  index indicates the  $t$  dependence ( $t^{n+1} = \Delta t + t^n$ ,  $t^{n+\frac{1}{2}} = t^n + \Delta t/2$ ,  $f_j^n = f(x_j, t^n)$ , etc. ( $R^{n+\frac{1}{2}}$  can be taken to be  $(R^n + R^{n+1})/2$ , etc.) The above can be rearranged into the following algorithmic form

$$(2.16) \quad \begin{aligned} & [(QA)^{n+\frac{1}{2}} - \lambda R^{n+\frac{1}{2}}](U_j^{n+1} - U_j^n) \\ & = 2\lambda R^{n+\frac{1}{2}}(U_j^n) + \Delta t Q^{n+\frac{1}{2}}(f_j^{n+\frac{1}{2}}), \quad j = 1, \dots, J-1, \end{aligned}$$

where  $\lambda = \epsilon \Delta t / (2h^2)$  and  $(QA)$  is the tridiagonal operator defined by  $(QA)U_j = \bar{q}_j^- a_{j-1} U_{j-1} + a_j^c a_j U_j + q_j^+ a_{j+1} U_{j+1}$ . The method (2.16) is for fixed  $\epsilon$  formally second-order accurate in  $\Delta t$  and fourth-order in  $h$  when (2.1) is an OCI scheme. Algorithms with the same general structure as (2.16) will result when either a Lees type or a fully implicit (Euler) time integration method is employed [6]. Moreover, in certain two-dimensional problems the individual ADI sweeps also have the above form [6]. Observe that the invertibility of  $\bar{Q}$  is not required in (2.16), however, it is used, at least formally, in many applications; for example, in formulating ADI methods for two-dimensional problems [6]. It therefore seems prudent to require that  $\bar{Q}^{-1}$  exists. Obviously, for (2.16) to be well defined, the tridiagonal matrix  $\bar{QA} - \lambda \bar{R}$  ( $\bar{A}$  is the diagonal matrix  $\bar{A}_{ij} = \delta_{ij}^1 a_j$ ) must be invertible; further, it is desirable that it be diagonally dominant.

To examine the stability of (2.16), we consider (2.15) with constant coefficients  $a$ ,  $b$ , and  $d$  and  $f = 0$  and perform the standard Fourier stability analysis [24]. For constant coefficients  $(QA)$  and  $R$  are independent of  $n$ ; thus, for  $f = 0$  (2.16) can be written as

$$(2.17) \quad (aQ - \lambda R)U_j^{n+1} = (aQ + \lambda R)U_j^n, \quad j = 1, \dots, J-1.$$

Since for constant coefficients  $r^{-,c,+}$  and  $q^{-,c,-}$  are independent of  $j$ , the substitution of  $U_j^n = \chi^n (e^{i\theta})^j$  into (2.17) yields

$$(2.18) \quad \chi = \frac{a + \lambda l(\theta)}{a - \lambda l(\theta)} \quad \text{where } l(\theta) = \frac{r^- e^{-i\theta} + r^+ e^{i\theta} + r^c}{q^- e^{-i\theta} + q^+ e^{i\theta} + q^c}.$$

For stability it is sufficient that  $|\chi| \leq 1$ . It follows from (2.18) that if  $a > 0$  and  $\lambda > 0$ , a necessary and sufficient condition for  $|\chi| \leq 1$  is  $\text{Re } l(\theta) \leq 0$ . Direct computation of  $\text{Re } l(\theta)$  yields

$$\begin{aligned} \text{Re } l(\theta) &= [r^c + (r^+ + r^-) \cos \theta][q^c + (q^+ + q^-) \cos \theta] + (r^+ - r^-)(q^+ - q^-) \sin^2 \theta \\ &= -\{(r^+ + r^-)(q^c - q^+ - q^-)(1 - \cos \theta) \\ &\quad - (r^c + r^+ + r^-)[q^c + \cos \theta(q^+ + q^-)] + 2 \sin^2 \theta(r^- q^+ + r^+ q^-)\}. \end{aligned}$$

Observe that if (2.9) and (2.10) hold and in addition

$$(2.19) \quad q^c \geq q^+ + q^-$$

holds, then  $\text{Re } l(\theta) \leq 0$  and (2.17) is unconditionally stable. It can also be shown that (2.9), (2.10), (2.19) insure unconditional stability for the methods using fully implicit or Lees type time discretization (Lees type is stable whenever  $|\chi| \leq 1$  for Crank-Nicolson; cf. [6]). Note also that if (2.9), (2.10) and (2.19) hold and, say,  $q^+ > 0$ , it follows that for constant coefficients the tridiagonal matrices  $\bar{Q}$  and  $(a\bar{Q} - \lambda\bar{R})$  are diagonally dominant and invertible by simple tridiagonal Gaussian decomposition [15]. We note that when a fully explicit (Euler) time integration method is used, a similar analysis suggests that stability could be a problem for all values of  $\Delta t$  if  $q^c = q^+ + q^-$ .

Observe that the condition (2.19) arises from consideration of the constant coefficient case. It will be seen that for the schemes presented in the next section the condition (2.19) cannot, in general, be satisfied when  $b(x)$  is not constant (cf. Remark 3.5). In the next section, we will replace (2.19) by the *heuristic* condition

$$(2.20) \quad b_j q_j^c \geq b_{j+1} q_j^+ + b_{j-1} q_j^-.$$

Recall that  $b(x)$  is assumed to be positive and bounded away from zero; hence, for constant coefficients (2.19) and (2.20) are equivalent. Also, when  $b(x)$  is not constant, (2.20) along with (2.10) and, say,  $q^+ > 0$  insures that  $\bar{Q}$  is invertible. Indeed, these conditions are sufficient to insure that  $\overline{QB}$  is invertible where  $\bar{B}$  is the diagonal matrix  $\bar{B}_{ij} = \delta_{ij} b_j$ . Diagonal dominance of the matrix  $\overline{QA} - \lambda\bar{R}$  will be examined at the end of Section III.

**III. Generalized OCI Schemes.** In this section, we present a broad family of formally fourth-order OCI approximations for  $Lu$  which generalize the standard OCI scheme given in (2.6). Our primary interest will be to exhibit a subfamily of schemes that satisfy the conditions formulated in Section II (see Theorem 3.7 and the paragraph following it) and, thus, are well behaved for  $0 < \epsilon \leq 1$  without requiring that  $h/\epsilon$  is bounded (i.e., there is no formal cell Reynolds number limitation). As will be seen later in this section, certain restrictions will be placed on  $h$ . These restrictions, however, are independent of  $\epsilon$  and involve only the coefficients  $b(x)$ ,  $d(x)$ , and  $a(x)$  (when (2.15) is considered). Throughout this section, the  $j$  index notation on  $r_j^-$ ,  $q_j^-$ , etc. will be dropped in order to simplify the notation.

3.1. *A Family of OCI Schemes.* As we have seen in the previous section, the standard OCI scheme is uniquely defined (within an unessential multiplicative constant) by the conditions  $T_j^\nu = 0$ ,  $\nu = 0, 1, 2, 3, 4$ ; cf. (2.5). These conditions represent the maximum number of lower-order terms in the Taylor development of the truncation error which can be “zeroed” by a scheme of the form (2.1). For the family of OCI schemes to be considered here, these conditions on (2.5) are weakened to

$$(3.1a) \quad T_j^0 = T_j^1 = T_j^2 = 0;$$

$$(3.1b) \quad T_j^3 = O(h^4), \quad T_j^4 = O(h^4) \quad (\epsilon \text{ fixed}), j = 1, \dots, J - 1.$$

As indicated earlier, the conditions (3.1a) imply that  $R$  is uniquely defined in terms of  $Q$  by (2.6b). The conditions (3.1b), therefore, serve as constraints on  $Q$  to insure formal fourth-order accuracy; cf. Remark 2.1. Note that it follows from (2.6b) and

(2.5e) that  $T_j^3$  and  $T_j^4$  can be written in terms of  $Q$  as

$$(3.2a) \quad T_j^3 = \epsilon h \left[ q^- - q^+ - \frac{1}{3} z (q^+ b_{j+1} + q^- b_{j-1} - \frac{1}{2} q^c b_j) \right],$$

$$(3.2b) \quad T_j^4 = \frac{\epsilon h^2}{12} [q^c - 5(q^+ + q^-) - z(q^+ b_{j+1} - q^- b_{j-1})], \quad z = h/\epsilon.$$

The conditions (3.1a) are motivated by the desire to find OCI schemes (2.1) that "zero" the maximum number of lower-order derivative terms in the Taylor development (2.5) and still allow the possibility of satisfying the conditions of Section 2.2 for  $z \in (0, \infty)$ . In this regard, we have the following

*Remark 3.1.* Consider a nontrivial scheme of the form (2.2) (i.e.,  $r^{-,c,+}$  and  $q^{-,c,+}$  do not vanish simultaneously) and assume that  $T_j^0 = T_j^1 = T_j^2 = T_j^3 = 0$ . Then  $q^+ \geq 0$ ,  $q^- \geq 0$ ,  $q^c \geq 0$ , and  $r^- \geq 0$  cannot all hold for  $z \in (0, \infty)$ .

*Proof.* To show this, use the expression for  $r^-$  given in (2.6b) and (3.2) (with  $T_j^3 = 0$ ) to write

$$2r^- = -q^+ - (\frac{1}{2} z b_j - 2)q^c - [2z(2b_{j-1} + h d_{j-1}) - 5]q^-.$$

For  $b_j, b_{j-1}$  positive,  $d_{j-1}$  nonnegative, and  $q^+ \geq 0$ ,  $q^- \geq 0$ ,  $q^c \geq 0$  (and not simultaneously zero), it follows that  $r^- < 0$  for  $z$  greater than some finite value.

We will consider the family of OCI schemes (2.1) which satisfy (3.1) and have  $q^-, q^+, q^c$ , defined as polynomials in  $z$  of order  $M$  at each mesh point  $j = 1, \dots, J-1$ , i.e.,

$$(3.3) \quad q^{-,c,+} = \sum_{\nu=0}^M q_{\nu}^{-,c,+} z^{\nu},$$

where the coefficients  $q_{\nu}^{-,c,+}$  are independent of  $\epsilon$ . It is understood that  $M$  is well defined in the sense that  $q_M^{-,c,+}$  are not all identically zero as functions of  $h$ . Note that the standard OCI scheme (2.6) belongs to this family. The implication of (3.1a) is that  $r^{-,c,+}$  defined by (2.6b) are also polynomials in  $z$  but of order  $M+1$ ; viz.,

$$(3.4a) \quad r^{-,c,+} = \sum_{\nu=0}^{M+1} r_{\nu}^{-,c,+} z^{\nu},$$

where, for example, the coefficients for  $r^-$  are given by

$$(3.4b) \quad r_0^- = q_0^- + q_0^c + q_0^+;$$

$$(3.4c) \quad r_{\nu}^- = q_{\nu}^- + q_{\nu}^+ + q_{\nu}^c - \frac{1}{2}(3b_{j-1} q_{\nu-1}^- + b_j q_{\nu-1}^c - b_{j+1} q_{\nu-1}^+) - h q_{\nu-1}^- d_{j-1},$$

for  $\nu = 1, \dots, M$ ;

$$(3.4d) \quad r_{M+1}^- = -\frac{1}{2}(3b_{j-1} q_M^- + b_j q_M^c - b_{j+1} q_M^+) - h q_M^- d_{j-1}.$$

The coefficients for  $r^c$  and  $r^+$  are of a similar form. To examine the implications of (3.1b), substitute (3.3) into (3.2) and impose (3.1b). The result is the following asymptotic relations as  $h \rightarrow 0$  ( $\epsilon$  fixed):

$$T_j^3 = \epsilon h [\tau_0^3 + \tau_1^3 z + \tau_2^3 z^2 + O(z^3)] = O(h^4),$$

$$T_j^4 = \frac{\epsilon h^2}{12} [\tau_0^4 + \tau_1^4 z + O(z^2)] = O(h^4),$$

where

$$(3.5a) \quad \tau_0^3 = q_0^- - q_0^+ = O(h^3), \quad \tau_0^4 = q_0^c - 5(q_0^+ + q_0^-) = O(h^2);$$

$$(3.5b) \quad \tau_1^3 = q_1^- - q_1^+ - \frac{1}{3} \left( q_0^+ b_{j+1} + q_0^- b_{j-1} - \frac{1}{2} q_0^c b_j \right) = O(h^2),$$

$$\tau_1^4 = q_1^c - 5(q_1^+ + q_1^-) - (q_0^+ b_{j+1} - q_0^- b_{j-1}) = O(h);$$

$$(3.5c) \quad \tau_2^3 = q_2^- - q_2^+ - \frac{1}{3} \left( q_1^+ b_{j+1} + q_1^- b_{j-1} - \frac{1}{2} q_1^c b_j \right) = O(h).$$

Observe that the condition (3.1b) yields five linear asymptotic relations and there are  $3(M + 1)$  coefficients  $q_{\nu}^{-,c,+}$  ( $\nu = 0, \dots, M$ ) to be specified. Because of the normalization (2.3),  $q_0^c$  must be positive; and it is convenient to take  $q_0^c = 60$ . For the case  $M = 0$ ,  $q_1^{-,c,+} = 0$  and  $q_2^{-,c,+} = 0$  and it is easily seen from (3.5a) and (3.5b) that  $\tau_1^3$  cannot be  $O(h^2)$ . Thus, there are no schemes of the form (3.3) with  $M = 0$  that satisfy (3.1). We, therefore, consider only  $M \geq 1$ . Observe that, accounting for (3.5) and the normalization of  $q_0^c$ , there are  $3(M - 1)$  coefficients which are unconstrained by the condition (3.1b). We point out that there are a variety of ways to satisfy the asymptotic relations (3.5). One possibility is to put  $\tau_0^3 = \tau_0^4 = \tau_1^3 = \tau_1^4 = \tau_2^3 = 0$  (with  $q_0^c = 60$ ). This approach, however, leads to relatively complex expressions for the coefficients of  $q^{-,c,+}$ . A different approach, which results in simpler expressions for  $q^{-,c,+}$ , will be used to define the schemes in Section 3.2.

We now examine some of the consequences of requiring that the schemes defined by (3.3) and (3.1) satisfy the conditions (2.9), (2.10), (2.14), and (2.20) for  $z \in (0, \infty)$ . Recall that these conditions were formulated assuming  $b(x) \geq B_1$  ( $B_1$  a positive constant) and  $d(x) \geq 0$ . As indicated in the following lemma not all of the above conditions need be considered.

LEMMA 3.2. Consider a scheme (2.1) which satisfies (3.1a). If the following conditions hold

$$(3.6) \quad q^- \geq 0, \quad q^+ > 0, \quad b_j q^c - b_{j+1} q^+ - b_{j-1} q^- \geq 0, \quad r^- \geq 0,$$

then  $q^c > 0$  and  $-r^c \geq r^+ + r^-$ ; furthermore, if in addition to (3.6),  $hd_{j+1}/b_{j+1} < 2$  then  $r^+ > r^-$  for  $z > 0$ .

Proof. The first inequality follows directly from the hypothesis. The second follows directly from (2.6b). To obtain the last inequality, use (2.6b) to write

$$(3.7) \quad \begin{aligned} r^+ - r^- &= z[b_{j-1} q^- + b_j q^c + b_{j+1} q^+ - h(q^+ d_{j+1} - q^- d_{j-1})] \\ &\geq z[(2b_{j-1} + hd_{j-1})q^- + (2b_{j+1} - hd_{j+1})q^+]. \end{aligned}$$

Suppose that a scheme defined by (3.3), (2.6b) satisfies (3.6) for  $z \in (0, \infty)$ . Since each of the quantities appearing in (3.6) are polynomials in  $z$ , it follows that the coefficient of the highest-order term in each of the respective polynomials must be nonnegative. In particular,

$$q_M^+ \geq 0, \quad q_M^- \geq 0, \quad b_j q_M^c - b_{j+1} q_M^+ - b_{j-1} q_M^- \geq 0, \quad \text{and}$$

$$2\kappa_{M+1}^- = -(4b_{j-1} + 2hd_{j-1})q_M^- - (b_j q_M^c - b_{j+1} q_M^+ - b_{j-1} q_M^-) \geq 0.$$

Note that the last three of the above can hold only if there is equality in each. We summarize this result in

**LEMMA 3.3.** *Consider a scheme defined by (3.3), (2.6b). If (3.6) holds for  $z \in (0, \infty)$ , then  $q_M^+ > 0$ ,  $q_M^- = 0$ ,  $b_j q_M^c = b_{j+1} q_M^+$ , and  $\kappa_{M+1}^- = 0$ ; further  $q_{M-1}^- \geq 0$ ,  $b_j q_{M-1}^c - b_{j+1} q_{M-1}^+ - b_{j-1} q_{M-1}^- \geq 0$ , and  $\kappa_M^- \geq 0$ .*

A consequence of Lemma 3.3 is

**Remark 3.4.** A necessary condition for a scheme (3.3), (2.6b) to satisfy  $r^+ > 0$  and (3.6) for  $z \in (0, \infty)$  is that  $hd_{j+1}/b_{j+1} \leq 2$ .

*Proof.* Substitute (3.3) into (2.6b) and use Lemma 3.3 to obtain

$$(3.8) \quad \begin{aligned} 2\kappa_{M+1}^+ &= b_j q_M^c + 3b_{j+1} q_M^+ - b_{j-1} q_M^- - 2hd_{j+1} q_M^+ \\ &= 2(2b_{j+1} - hd_{j+1})q_M^+. \end{aligned}$$

Since  $r^+ > 0$  for  $z \in (0, \infty)$  implies that  $\kappa_{M+1}^+ \geq 0$ , the result follows.

With the aid of Lemma 3.2, we can now prove a result alluded to in Section 2.2; viz.,

**Remark 3.5.** A scheme defined by (3.3), (2.6b) that satisfies  $q^+ > 0$ ,  $q^- \geq 0$ ,  $r^- \geq 0$  for  $z \in (0, \infty)$  cannot in general also satisfy (2.19) for  $z \in (0, \infty)$ .

*Proof.* Suppose that (2.19) holds for  $z \in (0, \infty)$ . This implies that  $q_M^c \geq q_M^+ + q_M^-$ . Also, if  $b_j > b_{j+1}$  and  $b_j \geq b_{j-1}$ , then (2.19) implies (2.20) for  $z \in (0, \infty)$  and by Lemma 3.3  $q_M^c = (b_{j+1}/b_j)q_M^+ < q_M^+$  which is a contradiction.

Formally as  $\epsilon \rightarrow 0$ ,  $Lu \rightarrow \tilde{L}u \equiv bu_x - du$  (the ‘‘reduced’’ operator). Consider now the limiting form as  $\epsilon \rightarrow 0$  ( $h$  fixed) of the schemes (3.3), (2.6b) that satisfy (3.6) for  $z \in (0, \infty)$ . Lemma 3.3 implies that as  $z \rightarrow \infty$ :  $q^- \sim q_{M-1}^- z^{M-1}$ ,  $q^+ \sim q_M^+ z^M$ ,  $q^c \sim (b_{j+1}/b_j)q_M^+ z^M$ ; and it follows from (2.6b) that  $r^- \sim \kappa_M^- z^M$ ,  $r^+ \sim q_M^+(2b_{j+1} - hd_{j+1})z^{M+1}$ ,  $r^c \sim -q_M^+ b_{j+1}[2 + (d_j/b_j)h]z^{M+1}$ . Substituting these results into (2.2) and dividing out a common factor on both sides, we find that

**Remark 3.6.** As  $\epsilon \rightarrow 0$  ( $h$  fixed) a scheme (3.3), (2.6b) satisfying (3.6) for  $z \in (0, \infty)$  takes the form

$$\frac{1}{h}(U_{j+1} - U_j) - \frac{1}{2} \left[ \left( \frac{d_{j+1}}{b_{j+1}} \right) U_{j+1} + \left( \frac{d_j}{b_j} \right) U_j \right] = \frac{1}{2} \left[ \left( \frac{1}{b_{j+1}} \right) \tilde{L}U_{j+1} + \left( \frac{1}{b_j} \right) \tilde{L}U_j \right].$$

The above limiting form is a formally second-order accurate one-sided (‘‘upwind’’) difference scheme for the reduced operator; cf. [1].

Obviously, to obtain schemes defined by (3.3), (3.1) that satisfy (3.6) for  $z \in (0, \infty)$ , both (3.5) and the conditions of Lemma 3.3 must be satisfied. For  $M = 1$  in (3.3), both sets of conditions cannot be satisfied. Indeed for  $M = 1$ ,  $q_2^{-,c,+} = 0$  and

Lemma 3.3 implies  $q_1^- = 0$  and  $b_j q_1^c = b_{j+1} q_1^+$ . It follows from (3.5c) that  $q_1^+ = O(h)$  and the condition  $\tau_1^3 = O(h^2)$  in (3.5b) cannot be satisfied. For  $M = 2$  in (3.3) the situation is more complex, but for the case of  $b(x) = b$  (a constant) and  $d(x) = 0$  it can be shown that the conditions of Lemma 3.3 and (3.5) are incompatible. We sketch here this demonstration leaving the details to the reader. Put  $q_0^c = 60$  and use (3.5a) and (3.5b) to obtain asymptotic relations for  $q_1^+$  and  $q_1^-$  in terms of  $q_1^c$ . Note that  $10q_1^- = q_1^c - 30b + O(h)$ . By Lemma 3.3,  $q_2^- = 0$  and  $q_2^+ = q_2^-$ . Now use (3.4c) and (3.5c) to obtain  $5\kappa_2^- = -(2q_1^c - 30b) + O(h)$ . But Lemma 3.3 also requires that  $q_1^- \geq 0$  and  $\kappa_2^- \geq 0$ . Hence, for sufficiently small  $h$  there is a contradiction. These observations have led us to the consideration of  $M = 3$  in (3.3) which will occupy the remainder of this section.

3.2. *A Subfamily (M = 3).* Consider now the schemes (3.3) with  $M = 3$  and

$$(3.9a) \quad q_0^c = 60, \quad q_0^+ = 6, \quad q_0^- = 6;$$

$$(3.9b) \quad q_1^+ = \frac{1}{10} q_1^c + 3b_j, \quad q_1^- = \frac{1}{10} q_1^c - 3b_j;$$

$$(3.9c) \quad q_2^+ = q_2^- + \frac{b_j}{10} q_1^c.$$

It can be verified by direct substitution that, if  $b(x) \in C^2[0, 1]$ , (3.9) satisfies (3.5). Thus, for this case the coefficients of (3.3) not defined by (3.9) or Lemma 3.3 are  $q_1^c, q_2^-, q_2^c,$  and  $q_3^+$ . It is convenient to introduce the following scaling for these coefficients

$$(3.10) \quad p_1 = \frac{1}{10} q_1^c/b_j, \quad p_2 = q_2^-/b_j^2, \quad p_3 = q_2^c/b_j^2, \quad p_4 = q_3^+/b_j^3.$$

In general,  $p_1, p_2, p_3, p_4$  may depend on  $j$ ; however, for simplicity of notation, we are omitting the  $j$  indices on these quantities. Incorporating (3.9), (3.10), and Lemma 3.3 into (3.3), we obtain

$$(3.11) \quad \begin{aligned} q_j^- &= 6 + (p_1 - 3)\rho + p_2\rho^2, \\ q_j^c &= 60 + 10p_1\rho + p_3\rho^2 + (b_{j+1}/b_j)p_4\rho^3, \\ q_j^+ &= 6 + (p_1 + 3)\rho + (p_1 + p_2)\rho^2 + p_4\rho^3, \end{aligned}$$

where  $\rho = b_j z = b_j h/\epsilon$  is the *local cell Reynolds number*. In the following, we shall require that  $h$  be sufficiently small so that  $10b_j - b_{j-1} - b_{j+1} > 0$  and  $hd_{j+1} < 2b_{j+1}$  for  $j = 1, \dots, J-1$ . The latter condition was used in Lemma 3.2 (and to insure that  $r^+ > 0$ , see Remark 3.4); the former condition implies that  $b_j q_j^c > b_{j+1} q_j^+ + b_{j-1} q_j^-$  for  $z = 0$ ; see (3.11).

The parameters  $p_1, p_2, p_3,$  and  $p_4$  appearing in (3.11) are now determined for each  $j = 1, \dots, J-1$ , so that (3.6) holds for  $\rho \in (0, \infty)$ . It will be seen that this places no restriction on  $p_1$  and, further,  $p_2, p_3, p_4$  are bounded from below. To establish this, we use a recursive approach which is briefly as follows: Consider any fixed value for  $p_1$  and determine the range of values for  $p_2$  such that  $q^- \geq 0$  and  $q^+ - q_3^+ z^3 = q^+ - p_4 \rho^3 \geq 0$  for  $\rho \in (0, \infty)$ . Having chosen such a  $p_2$ , next determine the range of values for  $p_3$  such that  $b_j q^c - b_{j+1} q^+ - b_{j-1} q^- \geq 0$  and  $r^- - \kappa_3^- z^3 \geq 0$  for

$\rho \in (0, \infty)$ . Finally, after choosing such a  $p_3, p_4$  is chosen so that  $r^- \geq 0$  and  $q^+ > 0$  for  $\rho \in (0, \infty)$ . Note from (3.11) that both  $q^-$  and  $q^+ - p_4\rho^3$  are of the form  $a_0 + a_1\rho + a_2\rho^2$  where the coefficients depend only on  $p_1$  and  $p_2$ , and  $a_0 > 0$ . A necessary and sufficient condition for such a quadratic to be nonnegative on  $(0, \infty)$  is that the discriminant be nonpositive when  $a_1 < 0$ , and  $a_2 \geq 0$  when  $a_1 \geq 0$ . Using this, we find that  $q^- \geq 0$  and  $q^+ - p_4\rho^3 \geq 0$  for  $\rho \in (0, \infty)$  if and only if

$$(3.12) \quad p_2 \geq \pi_0 \quad \text{where } 24\pi_0 = \left. \begin{cases} 0, & \text{if } 3 \leq p_1 \\ (p_1 - 3)^2, & \text{if } -3(3 - 2\sqrt{2}) \leq p_1 \leq 3 \\ -24p_1, & \text{if } -3 \leq p_1 \leq -3(3 - 2\sqrt{2}) \\ (p_1 - 3)^2 - 12p_1, & \text{if } p_1 \leq -3 \end{cases} \right\}.$$

The quantities  $b_jq^c - b_{j+1}q^+ - b_{j-1}q^-$  and  $r^- - \kappa_3 z^3$  are also quadratics with coefficients depending only on  $p_1, p_2$ , and  $p_3$  and with the zero-order coefficient positive. For the first quantity this follows directly from (3.11); for the second quantity use (3.11), (3.10) and (3.4) (note that  $\kappa_4^- = 0$ ). In order that  $b_jq^c - b_{j+1}q^+ - b_{j-1}q^- \geq 0$  for  $\rho \in (0, \infty)$ , we take  $p_3 \geq \pi_1(b_{j+1}/b_j, b_{j-1}/b_j)$  where

$$(3.13a) \quad \pi_1(\xi, \eta) = (\xi + \eta)p_2 + \xi p_1 + \left. \begin{cases} 0, & \text{if } \sigma \geq 0 \\ 3\sigma^2(10 - \xi - \eta)/8, & \text{if } \sigma \leq 0 \end{cases} \right\},$$

with  $\sigma \equiv p_1/3 + (\eta - \xi)/(10 - \xi - \eta)$ ,

and in order that  $r^- - \kappa_3 z^3 \geq 0$  for  $\rho \in (0, \infty)$ , we take  $p_3 \geq \pi_2(b_{j+1}/b_j, b_{j-1}/b_j)$  where

$$(3.13b) \quad \pi_2(\xi, \eta) = 15 - 2p_2 + (\tilde{S} - 1)p_1 - 3(\xi + \tilde{S}) + \left. \begin{cases} 0, & \text{if } 2p_1 - \tilde{S} \geq 0 \\ (2p_1 - \tilde{S})^2/8, & \text{if } 2p_1 - \tilde{S} \leq 0 \end{cases} \right\},$$

with  $2\tilde{S} \equiv 3\eta - \xi + 10 + 2h\eta(d_{j-1}/b_{j-1})$ .

Hence, to satisfy both conditions, we take

$$(3.13c) \quad p_3 \geq \max\{\pi_1(b_{j+1}/b_j, b_{j-1}/b_j), \pi_2(b_{j+1}/b_j, b_{j-1}/b_j)\}.$$

To find the range of values for  $p_4$ , consider the condition  $\kappa_3^- \geq 0$ . Note that from (3.4c), (3.11), and (3.10) we have

$$(3.14a) \quad \kappa_3^-/b_j^3 = (1 + b_{j+1}/b_j)p_4 - \frac{1}{2}\pi_3(b_{j+1}/b_j, b_{j-1}/b_j),$$

where

$$(3.14b) \quad \begin{aligned} \pi_3(\xi, \eta) &= p_3 + [3\eta - \xi + 2h(d_{j-1}/b_{j-1})\eta]p_2 - \xi p_1 \\ &= p_3 - \pi_1(\xi, \eta) + \bar{\pi}_1 + 2\eta[2 + h(d_{j-1}/b_{j-1})]p_2. \end{aligned}$$

In the above,  $\bar{\pi}_1$  denotes the third term in (3.13a). Note that  $\bar{\pi}_1$  is nonnegative. Since  $p_2$  is also nonnegative, it, therefore, follows from (3.14b) and (3.13c) that

$\pi_3(b_{j+1}/b_j, b_{j-1}/b_j)$  is nonnegative. Hence, in order to satisfy  $\kappa_3^- \geq 0$  and  $p_4 \geq 0$ , we take

$$(3.15) \quad p_4 \geq \frac{1}{2}[1 + b_{j+1}/b_j]^{-1} \pi_3(b_{j+1}/b_j, b_{j-1}/b_j).$$

It is obvious from the above that  $r^- \geq 0$  and  $q^+ \geq 0$  for  $\rho \in (0, \infty)$  if (3.12), (3.13c), (3.15) hold. We now show that under these conditions,  $q^+ > 0$  for  $\rho \in (0, \infty)$ . Note that  $q^+ = 0$  is possible only if  $p_4 = 0$  which in turn is possible only when equality is taken in (3.15) and  $\pi_3 = 0$ ; further, by (3.14b),  $\pi_3 = 0$  only if  $p_2 = 0$  which by (3.12) is possible only if  $p_1 \geq 3$ . But for  $p_1 \geq 3, p_2 = 0, p_4 = 0$  it follows from (3.11) that  $q^+ > 0$  for  $\rho \in (0, \infty)$ . We summarize the above results in

**THEOREM 3.7.** *Assume that  $h$  is sufficiently small so that  $10b_j - b_{j+1} - b_{j-1} > 0$  and  $hd_{j+1}/b_{j+1} < 2$  for  $j = 1, \dots, J - 1$ . Then, the OCI schemes defined by (2.1), (3.11), and (2.6b) with  $p_1$  arbitrary and  $p_2, p_3$ , and  $p_4$  satisfying (3.12), (3.13c), and (3.15), respectively, will satisfy the following for each  $j, j = 1, \dots, J - 1$ , and  $0 < h/\epsilon < \infty$ :*

$$q^- \geq 0, \quad q^c > 0, \quad q^+ > 0, \quad b_j q^c \geq b_{j+1} q^+ + b_{j-1} q^-, \quad \text{and}$$

$$r^+ > r^- \geq 0, \quad -r^c \geq r^+ + r^-.$$

Note that the parameters  $p_1$  and  $p_2$  do not necessarily depend on  $j$  and, thus, in practical calculations can be chosen, subject to (3.12), once and for all. The lower bounds for  $p_3, p_4$ , however, indicate that these parameters will need to be determined for each  $j, j = 1, \dots, J - 1$ . Concerning the choice of the parameters, the error analysis in Section IV indicates that better convergence behavior is achieved when  $r^-/r^+$  is small, particularly, in the limit as  $z \rightarrow \infty$ . This suggests choosing  $p_1, p_2, p_3, p_4$  so that the coefficient of the highest degree term in  $r^-$  is made as small as possible. In this regard, if  $p_4$  is defined by taking equality in (3.15), then  $\kappa_3^- = 0$ , cf. (3.14a). If one also takes  $p_1 \geq 3, p_2 = 0$ , and  $p_3$  defined by equality in (3.13c), it follows (as we shall see below) that  $\kappa_2^- = O(h)$ . Furthermore,  $\kappa_1^- = O(h)$  when  $p_1 = 3$ . For definiteness, and motivated by the above observations, for the error analysis in Section IV and the numerical experiments in Section V, we will consider the following sub-family of the OCI schemes defined in Theorem 3.7:

(3.16a) schemes defined by (2.1), (3.11), (2.6b) with

(3.16b)  $p_1$  any fixed constant (i.e., the same for  $j = 1, \dots, J - 1$ ), and  $p_2 = \pi_0$  (see (3.12)), and

(3.16c) for each  $j = 1, \dots, J - 1, p_3$  and  $p_4$  are defined by equality in (3.13c) and (3.15), respectively,

under the assumption that  $h$  is restricted so that

$$(3.16d) \quad 10b_j - b_{j-1} - b_{j+1} \geq B_0 > 0, \quad 2 - hd_{j+1}/b_{j+1} \geq B_0 > 0$$

$$(j = 1, \dots, J - 1),$$

where  $B_0$  is any positive constant smaller than  $8 \cdot \min\{b(x)\}$  and 2. The error analysis, and the numerical results in Section V tend to suggest the choice  $p_1 = 3$  in (3.16)

when solving (2.7). For the case  $d = 0$  and  $b$  a constant, and with  $\rho \equiv bh/\epsilon$ , this scheme is given by

$$(3.17a) \quad q^- = 6, \quad q^c = 60 + 30\rho + 9\rho^2 + 1.5\rho^3, \quad q^+ = 6 + 6\rho + 3\rho^2 + 1.5\rho^3,$$

$$(3.17b) \quad r^- = 72, \quad r^+ = 72 + 72\rho + 36\rho^2 + 12\rho^3 + 3\rho^4, \quad r^c = -r^- - r^+.$$

One can observe that  $r^-/r^+$  from (3.17) is the (4,0) Padé approximation to  $e^{-\rho}$  (e.g., [8], [30, p. 269]).

We remark that even if the schemes given in (3.16) appear somewhat complicated, it is actually quite simple to program the evaluation of  $r_j^{-,c,+}$ ,  $q_j^{-,c,+}$  (using (3.11) and (2.6b)). The relative efficiency of (3.16) will depend on (among other things) the balance between its possibly better accuracy (in principle requiring fewer mesh points to obtain a satisfactory solution) and the somewhat larger number of operations required to evaluate each coefficient of  $R$  and  $Q$ . The use of fewer mesh points is particularly significant in applications where many operations are required to evaluate  $b$  and  $d$  at each grid point.

3.3. *Some Properties of R for Schemes (3.16).* The following technical results concerning (3.16) will be used in the error analysis in Section IV. Let  $j$  be given and suppose  $b_{j+1} = b_{j-1} = b_j$  and  $d = 0$ . Then for a scheme (3.16) the resulting  $R$  coefficients (denoted by  $\tilde{r}_j^{-,c,+}$ ) are given by

$$(3.18a) \quad \tilde{r}_j^- = 72 + (12p_1 - 36)\rho + (6 - 5p_1 + 2p_2 + p_3)\rho^2,$$

$$(3.18b) \quad \tilde{r}_j^+ = 72 + (12p_1 + 36)\rho + (6 + 7p_1 + 2p_2 + p_3)\rho^2 \\ + (1.5p_1 + p_2 + .5p_3 + 2p_4)\rho^3 + 2p_4\rho^4,$$

$$(3.18c) \quad \tilde{r}_j^c = -\tilde{r}_j^- - \tilde{r}_j^+ \quad \text{where } \rho = b_j h/\epsilon.$$

Observe that, in the above,  $p_3$  and  $p_4$  are constants defined by equality in (3.13c) and (3.15) with  $b_{j+1} = b_{j-1} = b_j$  and  $d_{j-1} = 0$ . The quantities  $\tilde{r}_j^\pm$  will be used to approximate  $r_j^\pm$ . In what follows, it will be convenient to consider  $r_j^\pm$  as functions of  $\rho = b_j h/\epsilon$ ; viz.,

$$(3.19) \quad r_j^+ = \sum_{\nu=0}^4 \kappa_\nu^+ \rho^\nu, \quad r_j^- = \sum_{\nu=0}^2 \kappa_\nu^- \rho^\nu,$$

where it is understood that  $\kappa_\nu^\pm$  in general depend on  $j$ . (Note the change in notation from (3.4)). The corresponding coefficients in (3.18) will be denoted by  $\tilde{\kappa}_\nu^\pm$ . From the definitions (3.16), (3.18), (3.19), it can be seen that

*Remark 3.8.* Given a scheme (3.16), for each  $j$  ( $j = 1, \dots, J-1$ ) it is true that

$$(3.20a) \quad |\kappa_1^- - \tilde{\kappa}_1^-| + |\kappa_2^- - \tilde{\kappa}_2^-| \leq Ch \quad \text{and} \quad \kappa_0^- = 72,$$

$$(3.20b) \quad |\kappa_1^+ - \tilde{\kappa}_1^+| + \dots + |\kappa_4^+ - \tilde{\kappa}_4^+| \leq Ch \quad \text{and} \quad \kappa_0^+ = 72,$$

where  $C$  is a constant depending only on  $S_1 \equiv \{p_1, B_0, \min b, \max(b, d, |b'(x)|)\}$ .

With this result one can show that

LEMMA 3.9. Consider an OCI scheme (3.16). Then there exists a constant  $c > 0$  such that for  $h \leq c$  and for each  $j = 1, \dots, J - 1$ ,

$$(3.21) \quad \kappa_1^+ - \kappa_1^- \geq B_2 > 0, \quad \kappa_3^+ > 0, \quad \kappa_4^+ \geq B_3 > 0,$$

where the positive constants  $c, B_2$ , and  $B_3$  depend only on  $S_1$ . Also,  $\tilde{r}_j^{-,c,+}$  and  $q_j^{-,c,+}$  considered as polynomials in  $z$  (or  $\rho$ ) have coefficients that are bounded in magnitude by some constant  $B_4$  depending only on  $S_1$ .

*Proof.* The second part of the lemma is obvious from (3.16). For the first part, Remark 3.8 indicates that we need only consider  $\tilde{r}_j^{-,+}$  given by (3.18) (i.e., the case  $b_{j+1} = b_{j-1} = b_j, d = 0$ ). Now from the way in which the schemes (3.16) are chosen, the coefficients of  $\rho^2$  in  $q^-, q^+$ , and  $q^c - q^- - q^+$  are nonnegative; and hence,  $p_2, p_1 + p_2$ , and  $p_3 - p_1 - 2p_2$  are nonnegative. Further, from (3.12),  $p_2 > 0$  when  $p_1 < 3$ . From (3.15),  $2p_4 = p_3/2 + p_2 - p_1/2$  and so  $\tilde{\kappa}_3^+ = p_1 + 2p_2 + p_3 \geq 2(p_1 + p_2) + 2p_2 > 0$ . Now  $\tilde{\kappa}_1^+ - \tilde{\kappa}_1^- = 72$ , so it remains to deal with  $\tilde{\kappa}_4^+ = 2p_4$ . But  $2p_4 \geq (p_1/2 + p_2) + p_2 - p_1/2 = 2p_2$  which is positive when  $p_1 < 3$ . Suppose  $p_1 \geq 3$ . Then  $\pi_1 = p_1$  while  $\pi_2 = 5p_1 - 6$  so  $p_3 = 5p_1 - 6$  and  $2p_4 = 2p_1 - 3$  and the result follows.

3.4. *Diagonal Dominance for Parabolic Problems.* We now consider the tri-diagonal matrix  $\overline{QA} - (\Delta t/2zh)\overline{R}$  that arises in applications of (2.2) to parabolic equations; cf. Section 2.2. We confine our attention to the class of schemes defined in Theorem 3.7 and determine sufficient conditions for the diagonal dominance of this matrix for  $z \in (0, \infty)$ . Observe that if the matrix  $\overline{QA}$  is itself diagonally dominant for  $z \in (0, \infty)$  then the diagonal dominance of  $\overline{QA} - (\Delta t/2zh)\overline{R}$  would follow immediately from (2.9). However,  $\overline{QA}$  cannot in general be diagonally dominant for  $z \in (0, \infty)$  because of the particular form of the highest-order terms in  $q^+$  and  $q^c$  required by Lemma 3.3. As a preliminary result consider the condition

$$(3.22) \quad a_j(q^c - q_3^c z^3) - a_{j+1}(q^+ - q_3^+ z^3) - a_{j-1}q^- \geq 0.$$

Note that (3.22) is the condition for diagonal dominance of the product of  $\overline{A}$  with the matrix  $\overline{Q}$  in which the highest-order terms in  $q^c, q^+$  are removed. As can be readily seen from (3.11) and (3.10), this condition is analogous to the condition (2.20) with the function  $b(x)$  replaced by  $a(x, t)$  ( $t$  fixed). Recall that it is assumed that  $a(x, t) \geq A_0 > 0$ . Hence, (3.22) is satisfied for  $z \in (0, \infty)$  if  $10a_j - a_{j+1} - a_{j-1} > 0$  and  $p_3 \geq \pi_1(a_{j+1}/a_j, a_{j-1}/a_j)$ . This result is now used to prove the following

THEOREM 3.10. Assume that  $h$  is sufficiently small so that  $hd_{j+1}/b_{j+1} < 2, 10b_j - b_{j+1} - b_{j-1} > 0$ , and  $10a_j - a_{j+1} - a_{j-1} > 0, j = 1, \dots, J - 1$ . Consider an OCI scheme defined in Theorem 3.7 and such that for  $j = 1, \dots, J - 1$ ,

$$p_3 \geq \max\{\pi_1(b_{j+1}/b_j, b_{j-1}/b_j), \pi_2(b_{j+1}/b_j, b_{j-1}/b_j), \pi_1(a_{j+1}/a_j, a_{j-1}/a_j)\}.$$

Then for  $k$  a positive constant satisfying

$$k > \max\left\{0, \frac{a_{j+1}b_j - a_j b_{j+1}}{2b_j b_{j+1}}\right\} \text{ for } j = 1, \dots, J - 1,$$

the tridiagonal matrix  $\overline{QA} - (k/z)\overline{R}$  is strictly diagonally dominant for  $z \in (0, \infty)$ .

Note that for the applications in Section 2.2,  $k = \Delta t/2h$ . Hence, the above condition on  $k$  is satisfied if  $Ch^2 \leq \Delta t$  where  $C$  is independent of  $\epsilon$  and  $h$ .

*Proof.* The hypothesis implies that the results of Theorem 3.7 and (3.22) hold for  $z \in (0, \infty)$ . Note that  $\phi^c = q^c a_j - (k/z)r^c > 0$ . Let  $\phi^\pm = q^\pm a_{j\pm 1} - (k/z)r^\pm$ . We will establish that  $\Phi = \phi^c - |\phi^+| - |\phi^-| > 0$ . First assume that  $\phi^+ \geq 0$ , then

$$\begin{aligned} \Phi &= q^c a_j - q^+ a_{j+1} \mp q^- a_{j-1} + (k/z)(-r^c + r^+ \pm r^-) \\ &\geq q^c a_j - q^+ a_{j+1} - q^- a_{j-1} + (k/z)(-r^c + r^+ - r^-). \end{aligned}$$

It follows from (2.6b) that

$$\begin{aligned} -r^c + r^+ - r^- &> z[q^+(3b_{j+1} - hd_{j+1}) + q^c b_j - q^- b_{j-1}] > 2zq^+ b_{j+1} \\ &\geq 2z^4 q_3^+ b_{j+1}. \end{aligned}$$

The last inequality in the above is obtained using the fact that  $q^+ - z^3 q_3^+ \geq 0$ . It then follows using (3.22) that

$$\Phi > z^3 [a_j q_3^c - a_{j+1} q_3^+ + 2kb_{j+1} q_3^+] = z^3 [a_j b_{j+1}/b_j - a_{j+1} + 2kb_{j+1}] q_3^+.$$

Since  $q_3^+ = p_4 b_j^3 \geq 0$ , we conclude that for  $k$  restricted as indicated,  $\Phi > 0$  if  $\phi^+ \geq 0$ . Assume now that  $\phi^+ \leq 0$ , then

$$\Phi = q^c a_j + q^+ a_{j+1} \mp q^- a_{j-1} + (k/z)(-r^c - r^+ \pm r^-) \geq q^c a_j + q^+ a_{j+1} - q^- a_{j-1}.$$

Since  $q^+ > 0$ ,  $q^+ - q_3^+ z^3 \geq 0$ ,  $q_3^c \geq 0$ , it follows using (3.22) that  $\Phi > 0$  for  $\phi^+ \leq 0$  which completes the proof.

#### IV. Error Estimates.

4.1. *Statement of Results.* In this section we will establish an error estimate for the family of OCI schemes (3.16) applied to the problem (2.7). Also for comparison, the convergence behavior for several other schemes applied to (2.7) will be given. In this section (unless otherwise stated) it will be assumed that

$$(4.1) \quad 0 < B_1 \leq b(x) \quad \text{for some constant } B_1,$$

$$(4.2) \quad b(x) \text{ and } d(x) \quad \text{are in } C^m[0, 1],$$

with  $B_5 \equiv \|b\|$  and  $B_6 \equiv \|d\|$  in  $C^m[0, 1]$  (i.e.,  $B_5$  bounds the magnitude of the  $i$ th derivative of  $b$  for  $i = 0, 1, \dots, m$ , etc.). The function  $f$  in (2.7) may depend on  $\epsilon$  such that  $f(x, \epsilon)$  has  $m$  continuous  $x$ -derivatives satisfying

$$(4.3) \quad |f^{(i)}(x, \epsilon)| \leq B_7 + B_7 \epsilon^{-i} \exp(-\delta_1 x/\epsilon)$$

for  $i = 0, 1, \dots, m$ ,  $x$  in  $[0, 1]$ , and  $\epsilon$  in  $(0, 1]$ , where  $B_7$  and  $\delta_1$  are positive constants (independent of  $x$  and  $\epsilon$ ). The choice  $m = 5$  is used in the analysis of (3.16), and is indeed sufficient for the other schemes to be considered here. Let  $B_8$  denote  $|\alpha_0| + |\alpha_1|$ .

To state (and prove) the error estimate, we need to describe a particular function associated with each scheme (3.16). Let a specific value for  $p_1$  be given and consider for the moment the situation where  $b(x)$  is some positive constant  $\beta$  and  $d = 0$ . Then  $p_2$  and  $p_3$  defined by (3.16) are constants depending only on  $p_1$ . Define  $\bar{p}_3$  by

$$(4.4) \quad \bar{p}_3 = \begin{cases} p_3 + 1 & \text{if } p_1 < 3 \\ p_3 + h & \text{if } p_1 \geq 3 \end{cases},$$

and  $\bar{p}_4$  by equality in (3.15) with  $p_3 = \bar{p}_3$ , i.e.,  $\bar{p}_4 = (\bar{p}_3 + 2p_2 - p_1)/4$ . Now define  $\mu^- = \mu^-(\beta)$  and  $\mu^+ = \mu^+(\beta)$  by (cf. (3.18))

$$(4.5a) \quad \mu^- = 72 + (12p_1 - 36)\rho + (6 - 5p_1 + 2p_2 + \bar{p}_3)\rho^2,$$

$$(4.5b) \quad \begin{aligned} \mu^+ = 72 + (12p_1 + 36)\rho + (6 + 7p_1 + 2p_2 + \bar{p}_3)\rho^2 \\ + (1.5p_1 + p_2 + .5\bar{p}_3 + 2\bar{p}_4)\rho^3 + 2\bar{p}_4\rho^4, \end{aligned}$$

where  $\rho = \beta z = \beta h/\epsilon$ , and define

$$(4.5c) \quad \mu = \mu(\beta) = \mu^-(\beta)/\mu^+(\beta).$$

Throughout the rest of this paper  $c, C, c_i, C_i$  ( $i = 1, 2, \dots$ ) will be used to denote generic positive constants which may depend on elements of the set  $S = S_2 \cup \{B_0, p_1\}$  where  $S_2 = \{B_1, B_5, B_6, B_7, \delta_1, B_8\}$ , but which are independent of  $h$  and  $\epsilon$ . Also,  $\{U_j\}$  ( $j = 0, 1, \dots, J$ ) will be used to denote the approximation to the solution of (2.7),  $\{u(x_j)\}$ , obtained by (2.8) and whichever scheme is being discussed. Our error estimate is given by

**THEOREM 4.1.** *Let  $\{U_j\}$  be obtained by an OCI scheme (3.16). Then there are constants  $\sigma$  and  $C$  depending only on  $S$  such that for  $j = 1, \dots, J - 1$ ,*

$$(4.6a) \quad |U_j - u(x_j)| \leq Ch^4\epsilon^{-2} + Ch^4\epsilon^{-4}\mu(\sigma) \quad \text{when } h \leq \epsilon,$$

$$(4.6b) \quad |U_j - u(x_j)| \leq Ch^2 + C\mu(\sigma)^j \quad \text{when } \epsilon \leq h,$$

with  $\mu(\sigma)$  defined by (4.5).

For  $\epsilon$  fixed, (4.6a) clearly implies  $O(h^4)$  convergence. Also, when  $\epsilon$  is sufficiently small relative to  $h$ , (4.6b) shows that the error is  $O(h^2)$  at every mesh point. In particular, this is true when  $p_1 < 3$  and  $\epsilon \leq Ch^2$ ; when  $p_1 > 3$  and  $\epsilon \leq Ch^{5/3}$ ; and when  $p_1 = 3$  and  $\epsilon \leq Ch^{3/2}$ . To verify this, one need only check that  $\mu(\sigma)$  has the required behavior as  $oh/\epsilon \rightarrow \infty$ . The numerical results given in the next section indicate that the estimates (4.6) are "sharp." Later in this section we will also prove

**COROLLARY 4.2.** *Let  $\{U_j\}$  be obtained by an OCI scheme (3.16), and let  $x$  be a given number in  $(0, 1)$ . Then there is a constant  $C_x$  depending only on  $x$  and  $S$  such that*

$$(4.7a) \quad |U_j - u(x_j)| \leq C_x h^4 \epsilon^{-2} \quad \text{for } h \leq \epsilon, x_j \text{ in } [x, 1],$$

$$(4.7b) \quad |U_j - u(x_j)| \leq C_x h^2 \quad \text{for } \epsilon \leq h, x_j \text{ in } [x, 1].$$

In particular, Corollary 4.2 shows that the error is always  $O(h^2)$  "away from the boundary layer." The above results make rigorous the comments in the introduction concerning the error behavior of the OCI schemes. The proof of Theorem 4.1 utilizes the comparison function techniques employed by Kellogg and Tsan in [16]. The analysis of the OCI schemes, however, is more complicated because of the more complex

nature of the schemes, and in particular (as will be seen) because  $q^-$  is not zero. Before giving the proof, we present, for comparison, convergence results for several other schemes for (2.7).

4.2. *Convergence Results for Other Schemes for (2.7).* Using the same procedures as in the proof of Theorem 4.1, one can prove the following result for the standard OCI scheme:

**THEOREM 4.3.** *Let  $\{U_j\}$  be obtained by the standard OCI scheme (2.6). There are constants  $\sigma, c,$  and  $C$  depending only on  $S_2$  such that if  $h/\epsilon \leq c,$  then one has*

$$(4.8) \quad |U_j - u(x_j)| \leq Ch^4 \epsilon^{-2} + Ch^4 \epsilon^{-4} (1 + oh/\epsilon)^{-j} \quad \text{for } j = 1, \dots, J - 1.$$

We have briefly considered the scheme resulting from the use of interpolating cubic finite elements to solve (2.7) with  $b$  constant and  $d = 0.$  Using nodal condensation (see for example [5], [29]) followed by a local quadratic approximation of  $f$  to obtain  $q^{-,c,+}$  (i.e.,  $f$  at  $x_j \pm h/3$  and  $x_j \pm 2h/3$  is evaluated using the quadratic passing through  $f(x_{j-1}), f(x_j), f(x_{j+1})$ ), we obtain the following formally fourth-order scheme for (2.7) ( $b$  constant,  $d = 0$ ),

$$(4.9a) \quad \begin{aligned} q^- &= 6 - 3\rho + .2\rho^2 & \text{where } \rho &= bh/\epsilon, \\ q^c &= 60 + .8\rho^2, & q^+ &= 6 + 3\rho + .2\rho^2. \end{aligned}$$

$$(4.9b) \quad \begin{aligned} r^- &= 72 - 36\rho + 7.2\rho^2 - .6\rho^3, \\ r^+ &= 72 + 36\rho + 7.2\rho^2 + .6\rho^3, & r^c &= -r^- - r^+. \end{aligned}$$

Note that (2.6b) is valid for (4.9), and  $r^-/r^+$  is the (3,3) Padé approximation to  $e^{-\rho}.$  By the same type of analysis as used for Theorems 4.1 and 4.3 one has

*Remark 4.4.* Assume  $b$  is constant and  $d = 0.$  Let  $\{U_j\}$  be obtained by (4.9). There are constants  $\sigma, c,$  and  $C$  depending only on  $S_2$  such that if  $h/\epsilon \leq c,$  then

$$(4.10) \quad |U_j - u(x_j)| \leq Ch^4 \epsilon^{-2} + Ch^4 \epsilon^{-4} (1 + oh/\epsilon)^{-j} \quad \text{for } j = 1, \dots, J - 1.$$

Since (4.9) suffers a formal cell Reynolds number limitation, we have not pursued the derivation and analysis of the corresponding difference scheme for variable  $b$  and  $d.$

We now present results for the classical centered and upwind schemes. The centered scheme for (2.7) is given by

$$(4.11) \quad \begin{aligned} r_j^- &= 1 - b_j h/(2\epsilon), & r_j^c &= -2 - h^2 d_j/\epsilon, \\ r_j^+ &= 1 + b_j h/(2\epsilon), & q_j^c &= 1, & q_j^- &= q_j^+ = 0. \end{aligned}$$

Abrahamsson, Keller, and Kreiss [1] have obtained an error estimate for (4.11) applied to (2.7) by expanding the true and approximate solutions in powers of  $\epsilon, h,$  and  $h/\epsilon.$  This approach permits treatment of certain systems generalizing (2.7). For the scalar equation (2.7), the following result is also obtainable by comparison function techniques.

**THEOREM 4.5 (ABRAHAMSSON, KELLER AND KREISS [1]).** *Let  $\{U_j\}$  be obtained by the centered scheme (4.11). There are constants  $\sigma, c,$  and  $C$  depending only on*

$S_2$  such that if  $h/\epsilon \leq c$ , then

$$(4.12) \quad |U_j - u(x_j)| \leq Ch^2 + Ch^2\epsilon^{-2}(1 + oh/\epsilon)^{-j} \quad \text{for } j = 1, \dots, J - 1.$$

The classical first-order upwind scheme for (2.7) is given by

$$(4.13) \quad \begin{aligned} r_j^- &= 1, & r_j^c &= -2 - b_j h/\epsilon - h^2 d_j/\epsilon, \\ r_j^+ &= 1 + b_j h/\epsilon, & q_j^c &= 1, & q_j^- &= q_j^+ = 0. \end{aligned}$$

For this scheme one has

**THEOREM 4.6 (KELLOGG AND TSAN [16]).** *Let  $\{U_j\}$  be obtained by the upwind scheme (4.13). Then there are constants  $\sigma$  and  $C$  depending only on  $S_2$  such that for  $j = 1, \dots, J - 1$ ,*

$$(4.14a) \quad |U_j - u(x_j)| \leq Ch + Ch\epsilon^{-1}(1 + oh/\epsilon)^{-j} \quad \text{when } h \leq \epsilon,$$

$$(4.14b) \quad |U_j - u(x_j)| \leq Ch + C(1 + oh/\epsilon)^{-j} \quad \text{when } \epsilon \leq h.$$

Note that (4.14b) implies that the error at all the grid points is  $O(h)$  when  $\epsilon \leq Ch^2$ .

An upwind scheme which is second order for "small"  $\epsilon$  was presented and analyzed in [1] (see (1.12) therein). A slight variant of this scheme is

$$(4.15) \quad \begin{aligned} r_j^- &= 1, & r_j^c &= -2 - .5(\dot{b}_j + b_{j+1})h/\epsilon - d_j h^2/(2\epsilon), \\ r_j^+ &= 1 + .5(b_j + b_{j+1})h/\epsilon - d_{j+1} h^2/(2\epsilon), & q_j^c &= q_j^+ = 1/2, & q_j^- &= 0. \end{aligned}$$

The scheme (4.15) satisfies  $T_j^0 = T_j^1 = 0$ . The following error estimate of [1] can be obtained by the comparison function approach as well.

**THEOREM 4.7 (ABRAHAMSSON, KELLER AND KREISS [1]).** *Let  $\{U_j\}$  be obtained by the upwind scheme (4.15). There are constants  $\sigma$ ,  $c$ , and  $C$  depending only on  $S_2$  such that if  $h \leq c$  then for  $j = 1, \dots, J - 1$ , one has*

$$(4.16a) \quad |U_j - u(x_j)| \leq Ch^2 + Ch\epsilon + Ch\epsilon^{-1}(1 + oh/\epsilon)^{-j} \quad \text{when } h \leq \epsilon,$$

$$(4.16b) \quad |U_j - u(x_j)| \leq Ch^2 + C(1 + oh/\epsilon)^{-j} \quad \text{when } \epsilon \leq h.$$

This scheme is  $O(h)$  for fixed  $\epsilon$ , while achieving  $O(h^2)$  at all the grid points when  $\epsilon \leq Ch^3$ . We note that (4.15) can also be obtained using "upwind" linear finite elements [13] with a certain treatment of the integral containing  $d(x)$ .

It can be observed that of the schemes presented thus far, (3.16) displays the best theoretical convergence behavior. We now turn our attention to uniformly accurate schemes. The first such scheme, originally given by Allen and Southwell in [2], is

$$(4.17) \quad \begin{aligned} r_j^- &= \rho \cdot \exp(-\rho)/[1 - \exp(-\rho)], & r_j^+ &= r_j^- e^\rho, \\ r_j^c &= -r_j^- - r_j^+ - d_j h^2/\epsilon, & q_j^c &= 1, & q_j^- &= q_j^+ = 0, \quad \text{where } \rho = b_j h/\epsilon. \end{aligned}$$

The scheme (4.17) has been shown in [14], [16], [21] to be uniformly  $O(h)$  for (2.7). The sharpest estimate is obtained in [16]; viz.,

**THEOREM 4.8 (KELLOGG AND TSAN [16]).** *Let  $\{U_j\}$  be obtained by the scheme (4.17). Then there are constants  $\sigma$  and  $C$  depending only on  $S_2$  such that for  $j = 1, \dots, J-1$ ,*

$$(4.18) \quad |U_j - u(x_j)| \leq Ch^2/(h + \epsilon).$$

Note that the estimate (4.18) implies uniform  $O(h)$  convergence, as well as  $O(h^2)$  behavior for fixed  $\epsilon$ . The following scheme was formulated by El-Mistikawy and Werle in [9]; cf. also the references in [9]. When  $d = 0$ , it has the form given below (note (4.19) can also be obtained by using "upwind" linear finite elements with "optimal" one-point quadrature; see [13]).

$$(4.19) \quad \begin{aligned} r_j^- &= \rho^- \exp(-\rho^-)/[1 - \exp(-\rho^-)], & r_j^+ &= \rho^+/[1 - \exp(-\rho^+)], \\ r_j^c &= -r_j^- - r_j^+, & q_j^- &= (1 - r_j^-)/(2\rho^-), & q_j^+ &= (r_j^+ - 1)/(2\rho^+), & q_j^c &= q_j^- + q_j^+, \\ & & & \text{where } \rho^- &= (b_{j-1} + b_j)h/(2\epsilon), & \rho^+ &= (b_j + b_{j+1})h/(2\epsilon). \end{aligned}$$

Using comparison function techniques, the following preliminary error estimate can be obtained:

**THEOREM 4.9 (BERGER, SOLOMON AND CIMENT [3]).** *Assume  $d = 0$ , and let  $\{U_j\}$  be obtained by (4.19). Then there are constants  $\sigma$  and  $C$  depending only on  $S_2$  such that for  $j = 1, \dots, J-1$ ,*

$$(4.20a) \quad |U_j - u(x_j)| \leq Ch^2 + Ch^2 \epsilon^{-1} \exp(-\alpha x_j/\epsilon) \quad \text{when } h \leq \epsilon,$$

$$(4.20b) \quad |U_j - u(x_j)| \leq Ch^2 + C\epsilon \cdot \exp(-\alpha x_{j-1}/\epsilon) \quad \text{when } \epsilon \leq h.$$

Using this result in a manner to be indicated in Section 4.3, one can obtain the following estimate:

**THEOREM 4.10 (BERGER, SOLOMON AND CIMENT [3]).** *Assume (4.1),  $d = 0$ , and  $b(x)$  and  $f(x)$  are in  $C^5[0, 1]$  (here  $f$  does not depend on  $\epsilon$ ). Let  $\{U_j\}$  denote the approximate solution of (2.7) obtained by (4.19). Then there is a constant  $C$  depending only on  $b, f, \alpha_0, \alpha_1$  such that*

$$(4.21) \quad |U_j - u(x_j)| \leq Ch^2 \quad \text{for } j = 1, \dots, J-1, \epsilon \text{ in } (0, 1].$$

The scheme given in [9], in the case  $d \neq 0$ , can be written in the following form. Let  $\bar{n}_1$  denote the negative root of  $\epsilon w^2 + (b_{j-1} + b_j)w/2 - (d_{j-1} + d_j)/2 = 0$ , and let  $\bar{k}_1$  denote the nonnegative root. Define  $n_1 = h\bar{n}_1$  and  $k_1 = h\bar{k}_1$ . Similarly define  $n_2$  and  $k_2$  using the quadratic  $\epsilon w^2 + (b_j + b_{j+1})w/2 - (d_j + d_{j+1})/2$ . Define the functions  $e(w) \equiv \exp(w)$ ,  $g(w) \equiv (e(w) - 1)/w$  with  $g(0) \equiv 1$ ; and let  $2v_1 \equiv [1 - e(n_1 - k_1)]^{-1}$  and  $2v_2 \equiv [1 - e(n_2 - k_2)]^{-1}$ . The scheme [9] then has the form

$$(4.22) \quad \begin{aligned} r_j^- &= e(n_1)/g(n_1 - k_1), & r_j^+ &= e(-k_2)/g(n_2 - k_2), \\ r_j^c &= -n_1 - 1/g(n_1 - k_1) + k_2 - 1/g(n_2 - k_2), \\ q_j^- &= g(n_1)v_1 - e(n_1)g(-k_1)v_1, \\ q_j^+ &= g(-k_2)v_2 - e(-k_2)g(n_2)v_2, & q_j^c &= q_j^- + q_j^+. \end{aligned}$$

One can verify that when  $d = 0$ , (4.22) becomes (4.19). The results of a numerical experiment (to be reported in [3]) are consistent with the assertion that (4.22) is uniformly  $O(h^2)$  for (2.7) when  $b, d$ , and  $f$  are “smooth.” It appears to us that the extension of Theorems 4.9 and 4.10 to the case  $d \neq 0$  by comparison function techniques would involve a formidable amount of algebra, if not more serious difficulties.

The following result suggests that  $O(h^2)$  is the best uniform order of accuracy obtainable by the schemes we are considering:

**THEOREM 4.11.** *Consider an arbitrary scheme of the form (2.1) (here  $\tau_j^-, \dots, q_j^+$  can be arbitrary functions of  $b_{j-1}, b_j, b_{j+1}, d_{j-1}, d_j, d_{j+1}, h$ , and  $\epsilon$ ), and let  $\tau_j$  be defined by (2.4)–(2.5). Assume  $\tau_j^- \geq 0$ , (2.10) is true, and (for proper scaling) assume  $q_j^c \equiv 1$ . Then it cannot be true that  $|T_j^k| \leq Ch^{2+\delta}$  for  $\epsilon$  in  $(0, 1]$  and  $k = 1, 2, 3$  (here  $C$  and  $\delta$  denote positive constants independent of  $h$  and  $\epsilon$ ).*

The proof (given later) is similar to that of Theorem 4.5 of [16] which states that if also  $q^- \equiv q^+ \equiv 0$ , then  $|T_j^1| + |T_j^2|$  cannot be uniformly  $O(h^{1+\delta})$ . It is an open question whether there is a scheme of the form (2.1) which has an error behavior combining the best aspects of the generalized OCI schemes and the uniformly  $O(h^2)$  scheme, perhaps behaving like  $O(h^4/(h^2 + \epsilon^2))$ .

**4.2. Outline of the Proof of Theorem 4.1.** We now outline the proof of Theorem 4.1 in a sequence of lemmas (lengthy proofs are deferred to the end of the section). The first two lemmas bound the behavior of the solution  $u$  of (2.7), and are used in estimating the truncation error  $\tau_j$ .

**LEMMA 4.12 (KELLOGG AND TSAN [16]).** *Equation (2.7) has a unique solution, and there are positive constants  $\delta$  and  $C$  depending only on  $S_2$  such that*

$$(4.23) \quad |u^{(i)}(x)| \leq C[1 + \epsilon^{-i}\exp(-2\delta x/\epsilon)] \quad \text{for } i = 0, 1, \dots, m + 1$$

*( $m$  is that of (4.2)).*

**LEMMA 4.13** (slight extension of Lemma 2.4 of [16]). *The solution of (2.7) can be written in the form*

$$(4.24a) \quad u(x) = (-\epsilon u_x(0)/b(0))\exp(-b(0)x/\epsilon) + w(x),$$

where, for some positive constants  $\delta$  and  $C$  depending only on  $S_2$

$$(4.24b) \quad |w^{(i)}(x)| \leq C[1 + \epsilon^{-i+1}\exp(-2\delta x/\epsilon)] \quad \text{for } i = 0, 1, \dots, m + 1.$$

Note that by Lemma 4.12,  $|\epsilon u_x(0)/b(0)| \leq C_1$ .

The next two remarks, along with Remark 2.2, give the basis of the comparison function approach used here. As in [16], two comparison functions,  $\phi_j \equiv -2 + x_j$  and  $\psi_j = \psi_j(\beta) \equiv -[\mu(\beta)]^j$  (for some  $\beta > 0$  to be chosen), will be utilized. The function  $\phi_j$  is used to estimate the error where  $u$  is “well behaved,” while  $\psi_j$  estimates the error “near” the boundary layer at  $x = 0$ . The error estimate is obtained using  $\phi_j, \psi_j$ , and bounds on the truncation error, as indicated in the next two results.

**Remark 4.14.** Let  $e_j = u(x_j) - U_j$ , then the truncation error defined in (2.4) is  $\tau_j = \epsilon h^{-2} Re_j$  for  $j = 1, \dots, J - 1$ .

*Remark 4.15.* Let  $k_1(h, \epsilon) \geq 0$  and  $k_2(h, \epsilon) \geq 0$  such that

$$\epsilon h^{-2} R(\pm e_j + k_1(h, \epsilon)\phi_j + k_2(h, \epsilon)\psi_j) \geq 0 \quad \text{for } j = 1, \dots, J - 1.$$

Then (by Remark 2.2)  $|e_j| \leq k_1(h, \epsilon)|\phi_j| + k_2(h, \epsilon)|\psi_j|$ .

It, thus, remains to find lower bounds for  $\epsilon h^{-2} R\phi_j$  and  $\epsilon h^{-2} R\psi_j$ , and to bound  $|\tau_j|$ , thus determining  $k_1(h, \epsilon)$  and  $k_2(h, \epsilon)$  and, hence, giving the estimate for  $|e_j|$ . For the rest of the proof, it is convenient to take  $h$  bounded above by some ‘‘small’’ constant (independent of  $\epsilon$ ). The following observation shows that this is permissible.

*Remark 4.16.* Consider any scheme of the form (2.1) with  $T_j^0 \equiv T_j^1 \equiv 0$ . Assume that for some given  $h$  and  $\epsilon$  (2.9) and (2.10) are valid. Then each component of the solution  $\{U_j\}$  of (2.8) (for the given  $h$  and  $\epsilon$ ) is bounded by some constant  $C_1$  which depends only on  $\min b, \max|f|$ , and  $|\alpha_0| + |\alpha_1|$ .

This follows from Remark 2.2 and the fact that if  $C$  is a constant, then  $\epsilon h^{-2} R(C\phi_j \pm U_j) = Q(CL\phi_j \pm f_j)$  which is nonnegative for  $C$  sufficiently large. The lower bound for  $\epsilon h^{-2} R\phi_j$  is given in

LEMMA 4.17. *There are positive constants  $c$  and  $c_1$  depending only on  $S$  such that when  $h \leq c_1$  then for  $j = 1, \dots, J - 1$ ,*

$$(4.25a) \quad \epsilon h^{-2} R\phi_j \geq c \quad \text{when } h \leq \epsilon,$$

$$(4.25b) \quad \epsilon h^{-2} R(\epsilon^3 h^{-3} \phi_j) \geq c \quad \text{when } \epsilon \leq h.$$

*Proof.* By Lemma 3.9, for  $h$  sufficiently small,  $p_4 \geq c_2$  ( $c_2$  a positive constant). Since  $T_j^0 \equiv T_j^1 \equiv 0$ ,  $\epsilon h^{-2} R\phi_j = QL\phi_j \geq (q_j^- + q_j^c + q_j^+)B_1$ . By Theorem 3.7, (2.10) is valid. Hence, since  $q_j^+$  as a polynomial in  $\rho$  (see (3.11)) is positive for  $\rho \geq 0$ , (4.25a) holds. Now by the definition of  $p_2, q_j^+ - p_4\rho^3 \geq 0$  and (4.25b) follows.

We now turn our attention to  $\psi_j$ . The following technical result is used to bound  $\exp(-\delta h/\epsilon)$  (with  $\delta$  as in Lemmas 4.12 and 4.13) by  $\mu$ .

LEMMA 4.18. *Let  $\delta$  be a given positive number. Then there is a constant  $c$  depending on  $\delta$  and  $p_1$  such that if  $0 < \beta \leq c$ , then*

$$(4.26) \quad \exp(-\delta z) \leq \mu(\beta) \quad \text{for } 0 < z < \infty.$$

*Proof.* For  $z \leq c_1$  ( $c_1$  some positive constant),  $\exp(-\delta z) \leq 1 - \delta z/2$  while  $\mu(\beta) \geq 1 - 2\beta z$  which resolves the case of  $z$  near zero. Now  $\mu^-(\beta) > 0$  for all  $\rho = \beta z$ , so  $\mu^-(\beta) \geq c_2 > 0$  where  $c_2$  depends only on  $p_1$ . Thus,  $\mu \geq c_2/\mu^+(\beta)$ ; and hence, it suffices to find  $\beta$  such that  $c_2 \geq \mu^+(\beta)/\exp(\delta z)$ . But for any  $\beta \leq 1$  this will be true for  $z$  larger than some constant  $C_3$ . Now for  $c_1 \leq z \leq C_3$ ,  $\mu(\beta) \rightarrow 1$  as  $\beta \rightarrow 0$ , and so the result follows.

The desired lower bound for  $\epsilon h^{-2} R\psi_j$  is given by

LEMMA 4.19. *Let  $\delta$  be as in Lemmas 4.12 and 4.13. There exist constants  $c_1$  and  $c_2$  depending only on  $S$  such that when  $h \leq c_1$  and  $\beta \leq c_2$  then for  $j = 1, \dots, J - 1$ ,*

$$(4.27a) \quad \epsilon h^{-2} R(\epsilon\psi_j(\beta)) \geq c_\beta \mu(\beta)^j \geq c_\beta \exp(-\delta x_j/\epsilon) \quad \text{for } h \leq \epsilon,$$

$$(4.27b) \quad \epsilon h^{-2} R(\epsilon^3 h^{-2} \psi_j(\beta)) \geq c_\beta \mu(\beta)^j \geq c_\beta \exp(-\delta x_j/\epsilon) \quad \text{for } \epsilon \leq h,$$

$$(4.27c) \quad \epsilon h^{-2} R(\epsilon^3 h^{-2} \psi_j(\beta)/\mu(\beta)) \geq c_\beta \mu(\beta)^{j-1} \geq c_\beta \exp(-\delta x_{j-1}/\epsilon) \quad \text{for } \epsilon \leq h,$$

where the constant  $c_\beta$  depends only on  $S$  and on  $\beta$ .

The proof is rather involved and so will be given later.

It remains to estimate  $\tau_j$ . As in [16], essential use will be made of the integral form of the remainder in Taylor's theorem (which can be derived by repeated integration by parts). For a sufficiently smooth function  $g(x)$ , and numbers  $a$  and  $p$

$$(4.28) \quad \begin{aligned} R_n(a, p, g) &\equiv g(p) - \sum_{i=0}^n g^{(i)}(a) \frac{(p-a)^i}{i!} = g^{(n+1)}(\xi) \frac{(p-a)^{n+1}}{(n+1)!} \\ &= \frac{1}{n!} \int_a^p (p-s)^n g^{(n+1)}(s) ds. \end{aligned}$$

Here  $\xi$  is a point between the points  $a$  and  $p$ . The derivative form of  $R_n$  will be used when  $h \leq \epsilon$ . However, when  $\epsilon \leq h$ , the integral form will be used, since when  $p = a \pm h$  and  $g = \exp(-\beta x/\epsilon)$ , this "gains" an  $\epsilon$  while only "giving up" an  $h$  (in comparison to the derivative form of  $R_n$ ). The proof of Theorem 4.1 is completed by estimating  $|\tau_j|$  and, thus, finding the  $k_1$  and  $k_2$  of Remark 4.15.

4.3. *Proofs and Additional Remarks.* We first treat Theorem 4.1 in the case  $h \leq \epsilon$ . Recalling that  $T_j^0 = T_j^1 = T_j^2 = 0$  for the OCI schemes (3.16),

$$(4.29) \quad \begin{aligned} \tau_j &= T_j^3 u_j^{(3)} + T_j^4 u_j^{(4)} + [\epsilon h^3 (r_j^+ - r_j^-)/120 + \epsilon h^3 (q_j^- - q_j^+)/6] \mu_j^{(5)} \\ &\quad - b_{j-1} R_3(x_j, x_{j-1}, u^{(1)}) q_j^- - b_{j+1} R_3(x_j, x_{j+1}, u^{(1)}) q_j^+ \\ &\quad + d_{j-1} R_4(x_j, x_{j-1}, u) q_j^- + d_{j+1} R_4(x_j, x_{j+1}, u) q_j^+ \\ &\quad + \epsilon h^{-2} R_5(x_j, x_{j-1}, u) r_j^- + \epsilon h^{-2} R_5(x_j, x_{j+1}, u) r_j^+ \\ &\quad - \epsilon R_3(x_j, x_{j-1}, u^{(2)}) q_j^- - \epsilon R_3(x_j, x_{j+1}, u^{(2)}) q_j^+. \end{aligned}$$

The expansion out to terms involving  $u^{(6)}$  when  $h \leq \epsilon$  is necessary in order to demonstrate the formal fourth-order accuracy of the OCI schemes. Analyzing a formally second-order scheme generally only requires Taylor expansion out to  $u^{(4)}$  terms. To bound the terms in (4.29), we recall the fact that the schemes (3.16) satisfy the conditions (3.5), and so  $|T_j^3| \leq C \epsilon h z^3$  and  $|T_j^4| \leq C \epsilon h^2 z^2$  for  $h \leq \epsilon$ . Using Lemma 4.12 to bound the derivatives in (4.29) and choosing a  $\beta$  sufficiently small so that Lemmas 4.18 and 4.19 are valid, we obtain, for  $h \leq \epsilon$ ,

$$(4.30) \quad \begin{aligned} |\tau_j| &\leq C \{ \epsilon h z^3 (1 + \epsilon^{-3} \mu^j) + \epsilon h^2 z^2 (1 + \epsilon^{-4} \mu^j) + \epsilon h^3 z (1 + \epsilon^{-5} \mu^j) \\ &\quad + h^4 (1 + \epsilon^{-5} \mu^{j-1}) + \epsilon h^{-2} h^6 (1 + \epsilon^{-6} \mu^{j-1}) + \epsilon h^4 (1 + \epsilon^{-6} \mu^{j-1}) \}. \end{aligned}$$

Since  $h \leq \epsilon$ , there is a constant  $c(\beta)$  independent of  $h$  and  $\epsilon$  such that  $0 < c(\beta) \leq \mu(\beta)$ ; and so, each  $\mu^{j-1}$  in (4.30) may be replaced by  $\mu^j$ . Then Remarks 4.14. and 4.15 and Lemmas 4.17 and 4.19 yield (4.6a).

We now obtain the estimate (4.6b) for the case  $z \geq 1$ . When  $j \geq 2$  we write the truncation error in the form

$$\begin{aligned}
\tau_j &= \epsilon h^{-2} R_2(x_j, x_{j-1}, u) r_j^- + \epsilon h^{-2} R_2(x_j, x_{j+1}, u) r_j^+ \\
(4.31) \quad &+ \{-\epsilon R_0(x_j, x_{j-1}, u^{(2)}) - b_{j-1} R_1(x_j, x_{j-1}, u^{(1)}) + d_{j-1} R_2(x_j, x_{j-1}, u)\} q_j^- \\
&+ \{-\epsilon R_0(x_j, x_{j+1}, u^{(2)}) - b_{j+1} R_1(x_j, x_{j+1}, u^{(1)}) + d_{j+1} R_2(x_j, x_{j+1}, u)\} q_j^+,
\end{aligned}$$

while when  $j = 1$ , we replace the term  $W_1 \equiv -b_{j-1} R_1(x_j, x_{j-1}, u^{(1)}) q_j^-$  by

$$(4.32) \quad W_1 = -b_{j-1} R_0(x_j, x_{j-1}, u^{(1)}) q_j^- - b_{j-1} h u_j^{(2)} q_j^-.$$

We now appeal to (4.24) and observe by linearity, the error in approximating  $u$  is the sum of the errors in separately using the difference scheme for  $w(x)$  and for the first term in (4.24a). To treat  $w(x)$ , we utilize the integral form of  $R_n$  given in (4.28) with Lemma 4.13, and also the fact that

$$(4.33) \quad \int_{x_{j-1}}^{x_j} \exp(-2\delta s/\epsilon) ds \leq \epsilon \delta^{-1} \exp(-2\delta x_{j-1}/\epsilon).$$

We, thus, find that the truncation error  $\tau_j$  in (4.31) for the function  $w(x)$  is bounded by

$$\begin{aligned}
|\tau_j(w)| &\leq C \left\{ [\epsilon r_j^- + \epsilon q_j^- + h q_j^- + h^2 q_j^-] \int_{x_{j-1}}^{x_j} (1 + \epsilon^{-2} e^{-2\delta s/\epsilon}) ds \right. \\
(4.34) \quad &+ [\epsilon r_j^+ + \epsilon q_j^+ + h q_j^+] \int_{x_j}^{x_{j+1}} (1 + \epsilon^{-2} e^{-2\delta s/\epsilon}) ds \left. \right\}.
\end{aligned}$$

Then using (4.33) and Lemma 4.18, we obtain

$$\begin{aligned}
|\tau_j(w)| &\leq C \{ \epsilon h r_j^- + \epsilon h q_j^- + h^2 q_j^- + \epsilon h r_j^+ + \epsilon h q_j^+ + h^2 q_j^+ \} \\
(4.35) \quad &+ C \{ r_j^- + q_j^- + z q_j^- + h z q_j^- \} e^{-\delta x_{j-1}/\epsilon} \mu(\beta)^{j-1} \\
&+ C \{ r_j^+ + q_j^+ + z q_j^+ \} e^{-\delta x_j/\epsilon} \mu(\beta)^j.
\end{aligned}$$

The term which produced the factor  $z q_j^-$  (i.e.,  $W_1$  with  $u$  replaced by  $w$ ) will be handled separately when  $j = 1$ . Except for this, by noting that  $z^k \exp(-\delta z)$  is bounded ( $k$  any given positive integer,  $z \geq 0$ ) and by using Lemmas 4.17 and 4.19, we would have the following estimate for the error  $e_j(w)$  when  $\epsilon \leq h$ ;

$$(4.36) \quad |e_j(w)| \leq C \left\{ h^2 + \epsilon^3 h^{-2} \max_j (r_j^- + q_j^-) \mu^{j-1} + \mu^j \right\}.$$

We also have

*Remark 4.20.* The second summand on the right side of (4.36) is bounded by  $C\mu^j + Ch^2$  when  $\epsilon \leq h$ .

*Proof.* This follows by recalling (3.11), (3.18a), (3.20a), and then verifying that for  $\epsilon \leq h$ ;  $\epsilon^3 h^{-2} (r_j^- + q_j^-) \mu^+ \leq C\mu^-$  when  $h^2 \leq \epsilon$ , and  $\epsilon^3 h^{-2} (r_j^- + q_j^-) \leq Ch^2$  when  $\epsilon \leq h^2$ . The cases  $p_1 < 3$ ,  $p_1 = 3$ , and  $p_1 > 3$  are done individually.

To complete the treatment of  $w$ , at  $j = 1$  we use the expression (4.32) for  $W_1$  (with  $u$  replaced by  $w$ ). Now

$$(4.37a) \quad |W_1| \leq C q_1^- (h + \mu^{j-1}) + C h q_1^- (1 + \epsilon^{-1} e^{-2\delta x_j/\epsilon})$$

and so at  $j = 1$  it is true that

$$(4.37b) \quad |W_1| \leq Cq_j^- \mu^{j-1},$$

which completes the proof of (4.36).

To complete the proof of (4.6b) it remains to bound the error resulting from approximating  $v(x) \equiv \exp(-b(0)x/\epsilon)$  (cf. (4.24a)) by the difference scheme. Let  $V_j$  be the approximate solution; cf. (4.38b) below. Then  $e_j(v) \equiv |V_j - v_j| \leq |V_j| + |v_j| \leq |V_j| + \mu^j$  (the latter step by Lemma 4.18), so it remains only to bound  $|V_j|$ . The differential equation satisfied by  $v(x)$  is  $\epsilon v_{xx} + bv_x - dv = g$  where

$$(4.38a) \quad g(x) = [(b(0) - b(x))b(0)/\epsilon - d(x)]v(x);$$

and so

$$(4.38b) \quad \epsilon h^{-2} R V_j = G_j \equiv q_j^- g_{j-1} + q_j^c g_j + q_j^+ g_{j+1}.$$

But

$$(4.39) \quad |G_j| \leq C\{q_j^-(x_{j-1}/\epsilon + 1)v_{j-1} + q_j^c(x_j/\epsilon + 1)v_j + q_j^+(x_{j+1}/\epsilon + 1)v_{j+1}\}.$$

Since  $x_{j-1} = 0$  at  $j = 1$ , and  $y^k \exp(-b(0)y/2)$  is bounded for  $y \geq 0$  ( $k$  any fixed positive integer), and recalling Lemma 4.18, we have

$$(4.40) \quad |G_j| \leq Cq_j^- \mu^{j-1} + C\mu^j.$$

At  $j = 0$  and  $J$ ,  $V_j = v_j \leq \mu^j$ , so the discrete maximum principle (Remark 2.2) implies that

$$(4.41) \quad |V_j| \leq C\epsilon^3 h^{-2} \mu^{j-1} \max_j(q_j^-) + C\epsilon^3 h^{-2} \mu^j + \mu^j \quad \text{for } \epsilon \leq h.$$

This result, together with Remark 4.20 completes the proof of Theorem 4.1.

*Proof of Corollary 4.2.* Let  $\bar{x}$  be given in  $(0, 1)$ . We demonstrate that  $h^4 \epsilon^{-4} \mu(\sigma)^j \leq C(\bar{x})h^4 \epsilon^{-2}$  for  $h \leq \epsilon$  and  $x_j$  in  $(\bar{x}, 1)$ , and that  $\mu(\sigma)^j \leq C(\bar{x})h^2$  for  $\epsilon \leq h$  and  $x_j$  in  $(\bar{x}, 1)$ . First, we verify that

*Remark 4.21.* Given  $\sigma > 0$ , there is a  $\beta > 0$  such that  $\mu(\sigma) \leq (1 + \beta h/\epsilon)^{-1}$ .

*Proof.* Recalling (4.5) and Lemma 3.9, this is easy to obtain for  $h/\epsilon$  near 0, and for  $h/\epsilon$  large. For  $h/\epsilon$  in between;  $\mu(\sigma) \leq c < 1$  for some constant  $c$ , and so choosing  $\beta$  small enough so that  $(1 + \beta h/\epsilon)^{-1} > c$  gives the result.

Now if  $x_j \in (\bar{x}, 1)$ , then  $j \geq \bar{x}/h$ , so for the result when  $h \leq \epsilon$ , it suffices to find  $C(\bar{x})$  such that

$$(4.42) \quad (1 + \beta h/\epsilon)^{-\bar{x}/h} \leq C(\bar{x})\epsilon^2 \quad \text{for } h \leq \epsilon,$$

which is true if and only if

$$(4.43) \quad -\bar{x}h^{-1} \ln(1 + \beta h/\epsilon) \leq \ln C(\bar{x}) + 2 \ln(\epsilon) \quad \text{for } h \leq \epsilon.$$

Observe that  $\ln(1 + \beta h/\epsilon) \geq \beta h\epsilon^{-1} - .5\beta^2 h^2 \epsilon^{-2}$ . Thus, the term on the left side of (4.43) is bounded above by  $-\beta\bar{x}\epsilon^{-1} + .5\bar{x}\beta^2 h\epsilon^{-2}$  which, since  $h \leq \epsilon$ , is bounded above by  $(-\beta\bar{x} + .5\bar{x}\beta^2)\epsilon^{-1}$ . Since we may assume that  $\beta < 1$ , the left side of (4.43) is bounded above by  $-.5\beta\bar{x}/\epsilon$ . Hence, (4.43) is valid if  $C(\bar{x})$  can be found such that  $-.5\beta\bar{x}/\epsilon \leq \ln C(\bar{x}) + 2 \ln(\epsilon)$ , or if  $\epsilon^{-2} \exp(-.5\beta\bar{x}/\epsilon) \leq C(\bar{x})$ . Such a  $C(\bar{x})$  indeed exists.

For the case  $\epsilon \leq h$ , it suffices to find a  $C(\bar{x})$  for which

$$(4.44) \quad (1 + \beta h/\epsilon)^{-\bar{x}/h} \leq C(\bar{x})h^2 \quad \text{for } \epsilon \leq h.$$

But since  $\epsilon \leq h$ ,  $(1 + \beta h/\epsilon) \geq (1 + \beta)$ ; and so, the left side of (4.44) is bounded above by  $(1 + \beta)^{-\bar{x}/h} = \exp(-\bar{x}h^{-1}\ln(1 + \beta))$ ; and the result follows. We observe that

*Remark 4.22.* By using (4.42) and (4.44), "interior estimates" can be obtained as well for the other polynomial schemes described in this section.

We next comment briefly on the proofs of Theorems 4.9 and 4.10 which will be given elsewhere [3]. The demonstration of Theorem 4.9 follows the general pattern of the proof of Theorem 4.8 done by Kellogg and Tsan [16]. The analysis of the first term on the right of (4.24a), however, involves algebra even more involved than that encountered in [16]. With Theorem 4.9 in hand, Theorem 4.10 is obtained by a "bootstrapping" technique, using the decomposition of the solution  $u$  suggested by its uniformly valid asymptotic expansion as  $\epsilon \rightarrow 0$  (e.g., see Smith [26, p. 257]); viz.,

$$(4.45) \quad u(x) = A_0(x) + \frac{c}{b(x)} \exp \left[ -\epsilon^{-1} \int_0^x b(\xi) d\xi \right] + \epsilon R_0(x, \epsilon),$$

where  $c$  is a constant,  $A_0$  is smooth, and  $R_0(x, \epsilon)$  satisfies an equation of the same general form as (2.7). Since  $A_0$  is smooth and (4.19) is formally  $O(h^2)$  for all  $\epsilon$  in  $(0, 1]$ , it follows that the contribution to the error from approximating  $A_0$  using (4.19) is  $O(h^2)$ . The error from approximating  $R_0$  is demonstrated to be given by (4.20), so the error from  $\epsilon R_0$  is given by  $\epsilon$  times (4.20) which is  $O(h^2)$ . The final step in obtaining (4.21) is to directly show that (4.19) is uniformly  $O(h^2)$  for the middle term on the right side of (4.45). This involves lengthy but rather methodical Taylor series expansions along with the comparison function methods.

*Proof of Theorem 4.11.* Suppose the contrary, and consider the situation with  $b$  constant and  $d = 0$ , and set  $\rho = bh/\epsilon$ . Use the expressions in (2.5) for  $T^1$  and  $T^2$  to find that

$$(4.46a) \quad 2r^- = (2 - 3\rho)q^- + (2 - \rho)q^c + 2q^+ + \rho q^+ - h\epsilon^{-1}T^1 + 2\epsilon^{-1}T^2,$$

$$(4.46b) \quad r^+ - r^- = \rho q^- + \rho q^c + \rho q^+ + h\epsilon^{-1}T^1,$$

and from the expression for  $T^3$ ,

$$(4.46c) \quad (r^+ - r^-) + (6 - 3\rho)q^- + (-6 - 3\rho)q^+ = 6\epsilon^{-1}h^{-1}T^3.$$

Using the expression for  $r^+ - r^-$  in (4.46b) in (4.46c) yields

$$(4.47) \quad (6 - 2\rho)q^- + \rho q^c - 6q^+ + h\epsilon^{-1}T^1 - 6\epsilon^{-1}h^{-1}T^3 = 2\rho q^+.$$

Using the expression for  $\rho q^+$  given in (4.47) in (4.46a) produces

$$(4.48) \quad 2r^- = (5 - 4\rho)q^- + (2 - .5\rho)q^c - q^+ - .5h\epsilon^{-1}T^1 + 2\epsilon^{-1}T^2 - 3\epsilon^{-1}h^{-1}T^3.$$

Now since it is assumed that  $q^c = 1$ ,  $q^- \geq 0$ ,  $q^+ \geq 0$ , it follows that

$$(4.49) \quad 2r^- \leq (5 - 4\rho)q^- + 2 - .5\rho - .5h\epsilon^{-1}T^1 + 2\epsilon^{-1}T^2 - 3\epsilon^{-1}h^{-1}T^3.$$

Multiplying (4.49) through by  $\epsilon h^{-1}$ , letting  $\epsilon = ch$  for some constant  $c$ , and supposing  $T^1, T^2, T^3$  are  $O(h^{2+\delta})$  for some  $\delta > 0$  then implies that for  $h$  and  $c$  sufficiently small  $r^- < 0$ , a contradiction. Note also that Theorem 4.5 of [16] can be demonstrated in a similar fashion using (4.46a).

The last result which needs to be proven is Lemma 4.19.

*Proof of Lemma 4.19.* Since  $-r_j^- \geq r_j^- + r_j^+$ , we have for any given  $\beta > 0$ ,

$$(4.50a) \quad \epsilon h^{-2} R \psi_j(\beta) \geq \epsilon h^{-2} (-r_j^- \mu^{j-1} + (r_j^- + r_j^+) \mu^j - r_j^+ \mu^{j+1})$$

$$(4.50b) \quad = \epsilon h^{-2} \mu^{j-1} r_j^+ (1 - \mu)(\mu - r_j^-/r_j^+).$$

Recall (4.5), set  $\rho = \beta h/\epsilon$ , and write  $\mu$  in the form

$$(4.51) \quad \mu(\beta) = \mu^-/\mu^+ = (72 + \mu_1^- \rho + \mu_2^- \rho^2)/(72 + \mu_1^+ \rho + \dots + \mu_4^+ \rho^4).$$

We estimate the last three factors in (4.50b) in order to obtain (4.27). Now for  $h \leq c$  and  $z \leq c_1$  it is true that  $r_j^+ \geq c_0$  (since  $r_j^+$  at  $z = 0$  is 72 and Lemma 3.9 obtains). Also for  $h \leq c$  and  $z \geq c_3$ , we have  $r_j^+ \geq c_2 z^4$ ; cf. (3.21). For  $c_1 \leq z \leq c_3$ ,  $\tilde{r}_j^+$  (given by (3.18)) is  $\geq c_5$  (by Theorem 3.7). Then Remark 3.8 implies that for  $h \leq c$  and  $c_1 \leq z \leq c_3$ ,  $r_j^+ \geq c_6$ . Combining these results we have

$$(4.52) \quad r_j^+ \geq c_7 \text{ for } h \leq c, z \leq 1; \text{ and } r_j^+ \geq c_8 z^4 \text{ for } h \leq c, z \geq 1.$$

We next examine  $(1 - \mu)$ . We may suppose throughout that  $\beta < 1$ . Now since  $(1 + y)^{-1} = 1 - y + y^2 - \dots$  for  $|y| < 1$ , we have for  $\rho$  near 0

$$(4.53a) \quad \mu = 1 - (\mu_1^+ - \mu_1^-)\rho/72 + O(\rho^2).$$

Recalling (4.5), we then have

$$(4.53b) \quad \mu = 1 - \rho + O(\rho^2),$$

$$(4.53c) \quad 2\rho \geq 1 - \mu \geq \rho/2 \text{ for } \rho \leq c.$$

In particular, the latter holds for  $\beta \leq c$  and  $z \leq 1$ . Now consider any given  $\beta < 1$ . For  $z \geq c_1(\beta)$  (i.e.,  $c_1$  can depend on  $\beta$  as well as the set  $S$ ) and  $h \leq c$  one has  $\mu \leq 1/2$  and so  $1 - \mu \geq 1/2$ . For  $1 \leq z \leq c_1(\beta)$  and  $h \leq c$ , one finds by Theorem 3.7 that  $\mu^+ > \mu^-$  and, hence,  $1 - \mu \geq c_2(\beta) > 0$ , and so,

$$(4.54) \quad 1 - \mu \geq c_3(\beta) > 0 \text{ for } h \leq c, z \geq 1.$$

We now have remaining the task of finding an adequate lower bound for the last factor in (4.50b). We first consider the situation for  $z \leq 1$ . Following the same type of reasoning used in obtaining (4.53) (recall Remark 3.8 and (3.18)), there results

$$(4.55) \quad r_j^-/r_j^+ = \tilde{r}_j^-/\tilde{r}_j^+ + Cz^2 \leq 1 - \min(b)z + Cz^2 \text{ for } z \leq c;$$

and so, using (4.53b), there are positive constants  $c_1$  and  $c$  such that

$$(4.56) \quad \mu - r_j^-/r_j^+ \geq \min(b)z - 1 + 1 - 2\beta z + O(z^2) \geq c(\beta)z \text{ for } z \leq c_1, \beta \leq c.$$

Now for  $h \leq c$  and  $c_1 \leq z \leq 1$ , we have  $r_j^-/r_j^+ \leq c_2 < 1$  while for  $\beta$  sufficiently small and  $z \leq 1$  we have  $\mu(\beta) \geq (1 + c_2)/2$ . Hence, we conclude that

$$(4.57) \quad \mu - r_j^-/r_j^+ \geq c_1(\beta)z \text{ for } h \leq c, z \leq 1, \beta \leq c.$$

Combining (4.50), (4.52), (4.53c), (4.57), and Lemma 4.18 yields (4.27a).

We now show that  $\mu - r_j^-/r_j^+ \geq c(\beta)\mu$  when  $z \geq 1$ ,  $h \leq c_1$ ,  $\beta \leq c_2$  which with (4.50), (4.52), (4.54), and Lemma 4.18 proves (4.27b, c). For any given  $\beta$ , there are positive constants  $c$  and  $c^1$  such that

$$(4.58a) \quad \mu \geq c^1 \text{ when } 0 \leq \beta z \leq 1 \text{ and } h \leq c.$$

Also, there is an  $\eta > 1$  and a  $c > 0$  such that

$$(4.58b) \quad r_j^-/r_j^+ \leq c^{1/2} \quad \text{for } z \geq \eta, h \leq c,$$

and so, since  $\mu \leq 1$ ,

$$(4.58c) \quad \mu - r_j^-/r_j^+ \geq c^{1/2} \geq c^1 \mu/2 \quad \text{when } 0 \leq \beta z \leq 1, z \geq \eta, \text{ and } h \leq c.$$

Since  $0 < B_1 \leq b(x) \leq B_5$ , there is a constant  $M < 1$  and a constant  $c > 0$  such that

$$(4.58d) \quad r_j^-/r_j^+ \leq M < 1 \quad \text{when } 1 \leq z \leq \eta \text{ and } h \leq c.$$

Since  $\mu(\beta) \rightarrow 1$  as  $\beta \rightarrow 0$ , there are positive constants  $c_1, c_2, c_3$  such that

$$(4.58e) \quad \mu - r_j^-/r_j^+ \geq \mu - M \geq c_3 \geq c_3 \mu \quad \text{when } \beta \leq c_2, 1 \leq z \leq \eta, \text{ and } h \leq c_1.$$

By reviewing (4.58) we find that it only remains to deal with the situation where  $z \geq \eta$  and  $\beta z > 1$ . Henceforth, the latter conditions are assumed to obtain. Note that for any given  $\beta, \mu^- \leq C\beta^2 z^2$  while, analogously to (4.52),  $\mu^+ \geq \bar{c}_8 \beta^4 z^4$  when  $h \leq c$ , and so,

$$(4.59a) \quad \mu \leq c\beta^{-2} z^{-2} \quad \text{when } h \leq c_1.$$

Also, there are constants  $c$  and  $C$  such that  $\mu^+ \leq C\beta^4 z^4$  for  $h \leq c$ . Consider the case  $p_1 < 3$ . Then from the definition (4.4) of  $\bar{p}_3, \mu^- \geq \beta^2 z^2$ ; and so,  $\mu \geq \beta^2 z^2 / (C\beta^4 z^4) \geq (c_7/\beta^2) z^{-2}$ . However, recalling (4.52),  $r_j^-/r_j^+ \leq Cz^2 / (c_8 z^4) = c_9 z^{-2}$ . Hence, if  $\beta$  is chosen small enough so that  $(c_7/\beta^2) \geq 2c_9$ , then

$$(4.59b) \quad \mu - r_j^-/r_j^+ \geq c_9 z^{-2} \quad \text{for } \beta \leq c_2, h \leq c_1;$$

and so, (4.59) gives  $\mu - r_j^-/r_j^+ \geq c(\beta)\mu$ . Taking into account the definition of  $\bar{p}_3$ , the analysis of the cases  $p_1 = 3$  and  $p_1 > 3$  is similar.

**V. Numerical Experiments.** In this section we present some numerical results for the schemes that have been previously discussed. These results indicate that the convergence rates described in Section IV are "sharp," and give, for one particular problem, a comparison of the accuracy of the various schemes. The problem on which the numerical experiments were conducted was

$$(5.1) \quad \epsilon u_{xx} + b(x)u_x = f(x, \epsilon), \quad x \text{ in } (0, 1); \quad u(0) = \alpha_0, \quad u(1) = \alpha_1(\epsilon),$$

where  $b(x) = (x + 1)^3$  and the solution  $u(x)$  is given by

$$(5.2) \quad u(x) = \frac{1}{b(x)} \exp \left[ -\epsilon^{-1} \int_0^x b(\xi) d\xi \right] + \exp(-x/2).$$

The function  $u(x)$  determines  $f(x, \epsilon), \alpha_0$ , and  $\alpha_1(\epsilon)$ . Note that the resulting  $f(x, \epsilon)$  satisfies (4.3) and further  $|\alpha_1(\epsilon)|$  is bounded for  $\epsilon$  in  $(0, 1]$ . The motivation for this choice of the form of the exact solution  $u$  comes from the expression (4.45). The first term on the right side of (5.2) has the form of the "most singular part" of a solution of (5.1), while the second term is smooth. Note that as  $\epsilon$  becomes much smaller than the mesh size  $h$ , the first term becomes virtually zero at all grid points except  $x = 0$ . Thus for  $\epsilon \ll h, u(x)$  and, in general, its numerical approximation are dominated by the second term on the right side of (5.2).

In order to conveniently generate a wide variation of  $h$  and  $\epsilon$  for each scheme considered, problem (5.1) (with solution (5.2)) was run with  $\epsilon \equiv h^p$  for various values of  $p$ . For each scheme and each value of  $p$ , the mesh length  $h$  was successively halved starting with  $h = 1/32$  and ending with  $h = 1/2048$ . The numerical results are presented in Tables 1–9; each table summarizes the results from using a particular scheme. The first column in each table gives the number of subintervals  $J$ ; by definition the mesh size  $h$  is  $1/J$ . In each of the other columns numerical results are given for a particular value of  $p$ . For example, in Table 1, the column under  $\epsilon = h^5$  gives the maximum error at all the mesh points,  $E_\infty \equiv \text{maximum}_{j=1, \dots, J-1} |u_j - U_j|$ , for  $J = 32, 64, \dots, 2048$ , where for each given  $J$  problem (5.1) was solved with  $\epsilon = h^5 = (1/J)^5$ . The numerical rate of convergence is determined from the  $E_\infty$  values for two successive values of  $J$  (e.g.  $E_\infty^1$  and  $E_\infty^2$  corresponding to  $h = 1/J$  and  $h = 1/(2J)$ , respectively) by

$$(5.3) \quad \text{rate} \equiv (\ln E_\infty^1 - \ln E_\infty^2) / \ln(2).$$

Thus, for example, the errors in Table 1 for  $\epsilon = h^5$  for  $J = 32$  and  $J = 64$  are  $2.3\text{E}-5$  and  $3.9\text{E}-6$ , respectively; and the corresponding experimental rate of convergence defined by (5.3) is 2.58. The last row in each table gives the rate predicted by the error estimates presented in Section IV for the given value of  $p$ .

TABLE 1. Numerical results for the generalized OCI scheme (3.16) with  $p_1 = 3$ .

	$\epsilon=1$	$\epsilon=h^5$	$\epsilon=h^{.75}$	$\epsilon=h$	$\epsilon=h^{1.5}$	$\epsilon=h^2$	$\epsilon=h^3$
J	$E_\infty$ Rate	$E_\infty$ Rate	$E_\infty$ Rate	$E_\infty$ Rate	$E_\infty$ Rate	$E_\infty$ Rate	$E_\infty$ Rate
32	4.4E-7 3.89	2.3E-5 2.58	1.9E-4 1.43	2.0E-3 .28	6.3E-3 1.15	3.0E-5 3.03	7.7E-6 1.97
64	3.0E-8 3.94	3.9E-6 2.53	7.0E-5 1.31	1.7E-3 .14	2.8E-3 1.57	3.7E-6 2.58	2.0E-6 1.99
128	2.0E-9 3.90	6.7E-7 2.46	2.8E-5 1.17	1.5E-3 .07	9.6E-4 1.78	6.2E-7 2.22	5.0E-7 1.99
256	1.3E-10	1.2E-7 2.37	1.3E-5 1.10	1.4E-3 .04	2.8E-4 1.87	1.3E-7 2.06	1.2E-7 2.00
512	round off error	2.4E-8 2.28	5.9E-6 1.05	1.4E-3 .02	7.6E-5 1.92	3.2E-8 2.02	3.1E-8 2.00
1024		4.9E-9 2.22	2.8E-6 1.02	1.4E-3 .01	2.0E-5 1.94	7.8E-9 2.01	7.8E-9 2.00
2048		1.1E-9	1.4E-6	1.4E-3	5.3E-6	2.0E-9	2.0E-9
pred	4.00	2.00	1.00	.00	2.00	2.00	2.00

TABLE 2. Numerical results for the generalized OCI scheme (3.16) with  $p_1 = 0$ .

	$\epsilon=1$	$\epsilon=h^{.5}$	$\epsilon=h^{.75}$	$\epsilon=h$	$\epsilon=h^{1.5}$	$\epsilon=h^2$	$\epsilon=h^3$
J	$E_{\infty}$ Rate	$E_{\infty}$ Rate	$E_{\infty}$ Rate	$E_{\infty}$ Rate	$E_{\infty}$ Rate	$E_{\infty}$ Rate	$E_{\infty}$ Rate
32	3.5E-7 3.97	1.9E-5 2.75	1.7E-4 1.60	2.6E-3 .36	6.4E-2 .25	6.7E-3 1.73	1.6E-5 2.65
64	2.2E-8 3.98	2.8E-6 2.64	5.6E-5 1.42	2.0E-3 .19	5.4E-2 .49	2.0E-3 1.86	2.5E-6 2.24
128	1.4E-9 3.91	4.5E-7 2.53	2.1E-5 1.24	1.8E-3 .09	3.8E-2 .66	5.5E-4 1.93	5.3E-7 2.07
256	9.5E-11	7.8E-8 2.41	8.8E-6 1.15	1.7E-3 .05	2.4E-2 .78	1.5E-4 1.97	1.3E-7 2.02
512	round off error	1.5E-8 2.31	4.0E-6 1.09	1.6E-3 .02	1.4E-2 .85	3.7E-5 1.98	3.1E-8 2.00
1024		3.0E-9 2.23	1.9E-6 1.05	1.6E-3 .01	7.8E-3 .90	9.4E-6 1.99	7.8E-9 2.00
2048		6.3E-10	9.0E-7	1.6E-3	4.2E-3	2.4E-6	2.0E-9
pred	4.00	2.00	1.00	.00	1.00	2.00	2.00

TABLE 3. Numerical results for the generalized OCI scheme (3.16) with  $p_1 = 6$ .

	$\epsilon=1$	$\epsilon=h^{.5}$	$\epsilon=h^{.75}$	$\epsilon=h$	$\epsilon=h^{1.5}$	$\epsilon=h^{5/3}$	$\epsilon=h^2$
J	$E_{\infty}$ Rate	$E_{\infty}$ Rate	$E_{\infty}$ Rate	$E_{\infty}$ Rate	$E_{\infty}$ Rate	$E_{\infty}$ Rate	$E_{\infty}$ Rate
32	9.6E-7 3.82	5.0E-5 2.41	4.1E-4 1.24	4.1E-3 .16	1.0E-2 .96	2.7E-3 1.74	1.0E-4 2.73
64	6.8E-8 3.91	9.4E-6 2.41	1.7E-4 1.18	3.7E-3 .08	5.4E-3 1.25	8.0E-4 1.87	1.5E-5 2.75
128	4.5E-9 3.93	1.8E-6 2.36	7.7E-5 1.07	3.5E-3 .04	2.3E-3 1.38	2.2E-4 1.93	2.3E-6 2.68
256	3.0E-10	3.4E-7 2.29	3.7E-5 1.04	3.4E-3 .02	8.6E-4 1.43	5.7E-5 1.96	3.6E-7 2.55
512	round off error	7.0E-8 2.22	1.8E-5 1.01	3.3E-3 .01	3.2E-4 1.46	1.5E-5 1.98	6.1E-8 2.40
1024		1.5E-8 2.16	8.9E-6 .99	3.3E-3 .01	1.2E-4 1.48	3.8E-6 1.99	1.2E-8 2.25
2048		3.4E-9	4.5E-6	3.3E-3	4.2E-5	9.4E-7	2.4E-9
pred	4.00	2.00	1.00	.00	1.50	2.00	2.00

TABLE 4. Numerical results for the Swartz OCI scheme (2.6).

	$\epsilon=1$	$\epsilon=h^{.5}$	$\epsilon=h^{.75}$	$\epsilon=h$
J	$E_{\infty}$ Rate	$E_{\infty}$ Rate	$E_{\infty}$ Rate	$E_{\infty}$ Rate
32	2.0E-7	7.7E-6	1.3E-4	4.0E-3
	3.99	2.49	1.22	-2.95
64	1.3E-8	1.4E-6	5.7E-5	3.1E-2
	4.00	2.33	1.12	-8.74
128	7.8E-10	2.7E-7	2.6E-5	1.3E1
	4.15	2.20	1.05	2.73
256	4.4E-11	6.0E-8	1.3E-5	2.0E0
		2.12	1.03	.18
512	round off error	1.4E-8	6.2E-6	1.8E0
		2.07	1.02	.26
1024		3.3E-9	3.1E-6	1.5E0
		2.06	1.01	.01
2048		7.8E-10	1.5E-6	1.5E0
pred	4.00	2.00	1.00	.00

TABLE 5. Numerical results for the centered scheme (4.11).

	$\epsilon=1$	$\epsilon=h^{.5}$	$\epsilon=h^{.75}$	$\epsilon=h$
J	$E_{\infty}$ Rate	$E_{\infty}$ Rate	$E_{\infty}$ Rate	$E_{\infty}$ Rate
32	1.6E-4	2.5E-3	9.5E-3	4.6E-2
	2.01	1.29	.80	.19
64	3.9E-5	1.0E-3	5.4E-3	4.0E-2
	2.00	1.25	.68	.10
128	9.8E-6	4.4E-4	3.4E-3	3.7E-2
	2.00	1.21	.62	.05
256	2.4E-6	1.9E-4	2.2E-3	3.6E-2
	2.00	1.17	.58	.03
512	6.1E-7	8.4E-5	1.5E-3	3.5E-2
	2.00	1.13	.55	.01
1024	1.5E-7	3.9E-5	1.0E-3	3.5E-2
	2.02	1.10	.53	.01
2048	3.8E-8	1.8E-5	6.7E-4	3.5E-2
pred	2.00	1.00	.50	.00

TABLE 6. Numerical results for the first-order one-sided scheme (4.13).

	$\epsilon=1$	$\epsilon=h^{.5}$	$\epsilon=h^{.75}$	$\epsilon=h$	$\epsilon=h^{1.5}$	$\epsilon=h^2$	$\epsilon=h^3$
J	$E_{\infty}$ Rate	$E_{\infty}$ Rate	$E_{\infty}$ Rate	$E_{\infty}$ Rate	$E_{\infty}$ Rate	$E_{\infty}$ Rate	$E_{\infty}$ Rate
32	1.2E-2 .98	4.0E-2 .57	7.5E-2 .29	1.4E-1 .04	1.4E-1 .36	3.1E-2 .92	3.9E-3 1.14
64	6.2E-3 .99	2.7E-2 .56	6.1E-2 .27	1.4E-1 .02	1.1E-1 .42	1.6E-2 .96	1.7E-3 1.09
128	3.1E-3 1.00	1.9E-2 .55	5.1E-2 .24	1.3E-1 .01	8.0E-2 .45	8.3E-3 .98	8.2E-4 1.05
256	1.6E-3 1.00	1.3E-2 .54	4.3E-2 .25	1.3E-1 .01	5.9E-2 .47	4.2E-3 .99	4.0E-4 1.02
512	7.9E-4 1.00	8.7E-3 .53	3.6E-2 .24	1.3E-1 .00	4.2E-2 .48	2.1E-3 1.00	2.0E-4 1.01
1024	3.9E-4 1.00	6.0E-3 .52	3.1E-2 .24	1.3E-1 .00	3.0E-2 .49	1.1E-3 1.00	9.7E-5 1.01
2048	2.0E-4	4.2E-3	2.6E-2	1.3E-1	2.2E-2	5.4E-4	4.8E-5
pred	1.00	.50	.25	.00	.50	1.00	1.00

TABLE 7. Numerical results for the second-order one-sided scheme (4.15).

	$\epsilon=1$	$\epsilon=h^{.5}$	$\epsilon=h^{.75}$	$\epsilon=h$	$\epsilon=h^{1.5}$	$\epsilon=h^2$	$\epsilon=h^3$
J	$E_{\infty}$ Rate	$E_{\infty}$ Rate	$E_{\infty}$ Rate	$E_{\infty}$ Rate	$E_{\infty}$ Rate	$E_{\infty}$ Rate	$E_{\infty}$ Rate
32	1.7E-2 .95	3.3E-2 .54	6.4E-2 .23	1.3E-1 -.02	1.3E-1 .32	2.6E-2 .88	7.6E-4 1.89
64	8.6E-3 .97	2.3E-2 .53	5.5E-2 .21	1.3E-1 -.01	1.0E-1 .40	1.4E-2 .94	2.1E-4 1.95
128	4.4E-3 .99	1.6E-2 .51	4.7E-2 .21	1.3E-1 -.01	7.9E-2 .44	7.5E-3 .97	5.3E-5 1.97
256	2.2E-3 .99	1.1E-2 .50	4.1E-2 .22	1.3E-1 .00	5.8E-2 .46	3.8E-3 .98	1.4E-5 1.99
512	1.1E-3 1.00	7.8E-3 .49	3.5E-2 .23	1.3E-1 .00	4.2E-2 .48	1.9E-3 .99	3.4E-6 1.99
1024	5.6E-4 1.00	5.6E-3 .49	3.0E-2 .23	1.3E-1 .00	3.0E-2 .48	9.7E-4 1.00	8.6E-7 2.00
2048	2.8E-4	4.0E-3	2.6E-2	1.3E-1	2.2E-2	4.9E-4	2.2E-7
pred	1.00	.50	.25	.00	.50	1.00	2.00



It seems fair to state that the agreement between the predicted and experimental rates is rather good for all the schemes. The tables also give (even if for only one particular problem) a comparison of the numerical behavior of the various methods. Of the generalized OCI schemes (3.16) run, the one with  $p_1 = 3$  seems overall the best for this problem. The generalized OCI schemes are the most accurate schemes of those tested when  $\epsilon$  is not "near"  $h$ , while the uniformly second-order accurate scheme (4.19) gives the best results for  $\epsilon$  "near"  $h$ .

**Acknowledgment.** The authors thank Professor Bruce Kellogg for many very helpful discussions and comments and for kindly providing us with an early draft of [16].

Code R44  
Naval Surface Weapons Center  
Silver Spring, Maryland 20910

Code R44  
Naval Surface Weapons Center  
Silver Spring, Maryland 20910

Mathematical Analysis Division  
National Bureau of Standards  
Washington, D. C. 20234

Gulf Science and Technology Company  
Simulation Research Section  
P. O. Drawer 2038  
Pittsburgh, Pennsylvania 15230

Scientific Research Associates  
P. O. Box 498  
Glastenbury, Connecticut 06033

1. L. R. ABRAHAMSSON, H. B. KELLER & H.-O. KREISS, "Difference approximations for singular perturbations of systems of ordinary differential equations," *Numer. Math.*, v. 22, 1974, pp. 367–391.
2. D. N. DE G. ALLEN & R. V. SOUTHWELL, "Relaxation methods applied to determine the motion, in two dimensions, of a viscous fluid past a fixed cylinder," *Quart. J. Mech. Appl. Math.*, v. 8, 1955, pp. 129–145.
3. A. E. BERGER, J. M. SOLOMON & M. CIMENT, "On a uniformly accurate difference method for a singular perturbation problem." (In preparation.)
4. T. H. CHONG, "A variable mesh finite difference method for solving a class of parabolic differential equations in one space variable," *SIAM J. Numer. Anal.*, v. 15, 1978, pp. 835–857.
5. I. CHRISTIE & A. R. MITCHELL, "Upwinding of high order Galerkin methods in conduction-convection problems," *Internat. J. Numer. Methods Engrg.*, v. 12, 1978, pp. 1764–1771.
6. M. CIMENT, S. H. LEVENTHAL & B. C. WEINBERG, "The operator compact implicit method for parabolic equations," *J. Computational Phys.*, v. 28, 1978, pp. 135–166.
7. F. W. DORR, "The numerical solution of singular perturbations of boundary value problems," *Siam J. Numer. Anal.*, v. 7, 1970, pp. 281–313.
8. B. L. EHLE, "A-stable methods and Pade approximations to the exponential," *SIAM J. Math. Anal.*, v. 4, 1973, pp. 671–680.
9. T. M. EL-MISTIKAWY & M. J. WERLE, "Numerical method for boundary layers with blowing—the exponential box scheme," *AIAA J.*, v. 16, 1978, pp. 749–751.
10. J. C. HEINRICH, P. S. HUYAKORN, O. C. ZIENKIEWICZ & A. R. MITCHELL, "An upwind finite element scheme for two-dimensional convective transport equation," *Internat. J. Numer. Methods Engrg.*, v. 11, 1977, pp. 131–143.
11. J. C. HEINRICH & O. C. ZIENKIEWICZ, "Quadratic finite element schemes for two-dimensional convective-transport problems," *Internat. J. Numer. Methods Engrg.*, v. 11, 1977, pp. 1831–1844.

12. R. S. HIRSH & D. H. RUDY, "The role of diagonal dominance and cell Reynolds number in implicit difference methods for fluid mechanics problems," *J. Computational Phys.*, v. 25, 1974, pp. 304–310.
13. T. J. R. HUGHES, W. K. LIU & A. BROOKS, "Finite element analysis of incompressible viscous flows by the penalty function formulation," *J. Computational Phys.*, v. 30, 1979, pp. 1–60.
14. A. M. IL'IN, "Differencing scheme for a differential equation with a small parameter affecting the highest derivative," *Mat. Zametki*, v. 6, 1969, pp. 237–248=*Math. Notes*, v. 6, 1969, pp. 596–602.
15. E. ISAACSON & H. B. KELLER, *Analysis of Numerical Methods*, Wiley, New York, 1966.
16. R. B. KELLOGG & A. TSAN, "Analysis of some difference approximations for a singular perturbation problem without turning points," *Math. Comp.*, v. 32, 1978, pp. 1025–1039.
17. H.-O. KREISS, "Difference approximations for singular perturbation problems," *Numerical Solutions of Boundary Value Problems for Ordinary Differential Equations* (A. K. Aziz, Ed.), Academic Press, New York, 1975, pp. 199–211.
18. H.-O. KREISS & N. NICHOLS, *Numerical Methods for Singular Perturbation Problems*, Lecture Notes in Phys., Vol. 58 (R. Glowinski and J. L. Lions, Eds.), Springer-Verlag, Berlin, 1976, pp. 544–558.
19. D. C. L. LAM & R. B. SIMPSON, "Centered differencing and the box scheme for diffusion convection problems," *J. Computational Phys.*, v. 22, 1976, pp. 486–500.
20. J. J. H. MILLER, "Some finite difference schemes for a singular perturbation problem," in *Constructive Function Theory*, Proc. Internat. Conf. on Constr. Fun. Theory, Blagoevgrad, 30 May–4 June 1977. (To appear.)
21. J. J. H. MILLER, "Sufficient conditions for the convergence, uniformly in epsilon, of a three point difference scheme for a singular perturbation problem," in *Praktische Behandlung von Differentialgleichungen in Anwendungsgebieten* (Proc. Conf., Oberwolfach, 1977), Lecture Notes in Math., Springer, Berlin. (To appear.)
22. C. E. PEARSON, "On a differential equation of boundary layer type," *J. Math. and Phys.*, v. 49, 1968, pp. 134–154.
23. M. H. PROTTER & H. F. WEINBERGER, *Maximum Principles in Differential Equations*, Prentice-Hall, Englewood Cliffs, N. J., 1967.
24. R. D. RICHTMYER & K. W. MORTON, *Difference Methods for Initial-Value Problems*, 2nd ed., Interscience, New York, 1967.
25. P. J. ROACHE, *Computational Fluid Dynamics*, Hermosa, Albuquerque, N. M., 1972.
26. D. R. SMITH, "The multivariable method in singular perturbation analysis," *SIAM Rev.*, v. 17, 1975, pp. 221–273.
27. B. K. SWARTZ, "The construction of finite difference analogs of some finite element schemes," in *Mathematical Aspects of Finite Elements in Partial Differential Equations* (C. de Boor, Ed.), Academic Press, New York, 1974, pp. 279–312.
28. M. VAN VELDHUIZEN, "Higher order methods for a singularly perturbed problem," *Numer. Math.*, v. 30, 1978, pp. 267–279.
29. M. VAN VELDHUIZEN, "Higher order schemes of positive type for singular perturbation problems," *Numerical Analysis of Singular Perturbation Problems* (P. W. Hemker and J. J. H. Miller, Eds.), Academic Press, New York, 1979, pp. 361–383.
30. R. S. VARGA, *Matrix Iterative Analysis*, Prentice-Hall, Englewood Cliffs, N. J., 1962.