SIAM J. SCI. COMPUT.
Vol. 18, No. 1, pp. 182–202, January 1997
© 1997 Society for Industrial and Applied Mathematics
010

IMPROVED QUASI-STEADY-STATE-APPROXIMATION METHODS FOR ATMOSPHERIC CHEMISTRY INTEGRATION*

L. O. JAY†, A. SANDU‡, F. A. POTRA§, AND G. R. CARMICHAEL¶

Abstract. In the last fifteen years the quasi-steady-state-approximation (QSSA) method has been a commonly used method for integrating stiff ordinary differential equations arising from atmospheric chemistry problems. In this paper a theoretical analysis of the QSSA method is developed, stressing its strengths and its weaknesses. This theory leads to practical improvements to the QSSA method. New algorithms, including symmetric and extrapolated QSSA are presented.

Key words. atmospheric chemistry, stiff ordinary differential equations, quasi-steady-state approximation, extrapolation, differential-algebraic equations

AMS subject classifications. 65L05, 65L06, 80A30

PII. S1064827595283033

1. Introduction.

As our scientific understanding of atmospheric chemistry and dynamics has expanded in recent years, so has our ability to construct comprehensive models which describe the relevant processes. (Carmichael, Peters, and Kitada [5], Jacob et al. [18], and Dentener and Crutzen [7] are examples of regional and global scale atmospheric chemistry models in use today.) However, these comprehensive atmospheric chemistry models are computationally intensive because the governing equations are nonlinear, highly coupled, and stiff. As with other computationally intensive problems, the ability to fully utilize these models remains severely limited by today's computer technology.

The large computational requirements in the study of chemically perturbed environments arise from the complexity of the chemistry of the atmosphere. Integration of the chemistry rate equations typically consumes as much as 90 percent of the total CPU time! Obviously, more efficient integration schemes for the chemistry solvers would result in immediate benefits through the reduction of CPU time necessary for each simulation. As more and more chemical species and reactions are added to the chemical scheme for valid scientific reasons the need for faster yet more accurate chemical integrators becomes even more critical.

Efficient chemistry integration algorithms for atmospheric chemistry have been obtained by carefully exploiting the particular properties of the model. One of the commonly used methods is the QSSA method of Hesstvedt, Hov, and Isaacsen [16]. The performance of the QSSA scheme can be further improved by using the lumping technique which leads to mass conservation of groups of species. Practical QSSA per-

*Received by the editors March 13, 1995; accepted for publication (in revised form) March 4, 1996. This research was supported in part by the DOE under grant DE-FG02-94ER 61855 and by the Center for Global and Regional Environmental Research.

http://www.siam.org/journals/sisc/18-1/28303.html

†Department of Computer Science, University of Minnesota, 4-192 EE/CS Bldg, 200 Union Street S.E., Minneapolis, MN 55455-0159 (na.ljay@na-net.ornl.gov). The work of this author was supported by the Fonds National Suisse de la Recherche Scientifique, Switzerland.

‡Program in Applied Mathematics and Computational Sciences, The University of Iowa, Iowa City, IA 52246 (sandu@cgrer.uiowa.edu).

§Departments of Mathematics and Computer Science, The University of Iowa, Iowa City, IA 52246 (potra@math.uiowa.edu).

¶Center for Global and Regional Environmental Research and the Department of Chemical and Biochemical Engineering, The University of Iowa, Iowa City, IA 52246 (gcarmich@icaen.uiowa.edu).

formance is discussed in the instructive paper [26], by Shieh, Chang, and Carmichael, where different integrators are compared on specific atmospheric chemistry problems. An evaluation of the local truncation error of the QSSA scheme can be found in [30].

There are many specially tailored methods in use in atmospheric chemistry models. One of the first proposed methods, and one which has been extensively used, is the hybrid predictor-corrector algorithm of Young and Boris [33]. Species are divided into stiff and nonstiff; the explicit Euler method (predictor) and an explicit trapezoidal method (corrector) are used for the nonstiff part, while the stiff part is integrated with a modified midpoint scheme.

Sillman in [27] developed an integration scheme based on the implicit Euler formula. Following a careful analysis of sources and sinks of odd hydrogen radicals in the troposphere, the author reorders the vector of species such that the resulting Jacobian is nearly lower block triangular; this enables an elegant "decoupling" between short-lived species (integrated implicitly) and long-lived species (integrated semi-implicitly). The scheme is efficient but difficult to generalize.

Hertel, Berkowicz, Christensen, and Hov [15] proposed an algorithm based on the implicit Euler method. Using only linear operators it preserves the total mass. The nonlinear system is solved using functional iterations. The main idea is to speed up these iterations using explicit solutions for several groups of species. The method seems to work fine for very large step-sizes.

A particularly clear approach was taken by Gong and Cho [11]. They divide the species into slow and fast, according to their lifetimes. The slow species are estimated using an explicit Euler scheme; the implicit ones are integrated with the implicit Euler scheme (and Newton–Raphson iterations for solving the nonlinear system). As a last step, the slow species are "corrected," reiterating the explicit Euler step.

A fancy projection/forward differencing method was proposed by Elliot, Turco, and Jacobson [9]. The species are grouped together in families. The distribution of the constituents inside a family is recalculated before each integration step using an implicit relation and solving the corresponding nonlinear system. (This "projection" can be viewed as a "predictor.") Then the integration is carried out for families using a significantly improved time step.

Dabdub and Seinfeld in [6] investigated an extrapolation algorithm whose underlying numerical scheme is based on a QSSA predictor and on a hybrid corrector (with a trapezoidal method for nonstiff components and a modified QSSA formula for the stiff components). The authors report good results; however, a theoretical analysis of the method is not presented.

Verwer [29] proposed an extension of QSSA to a second-order consistent scheme and also a "two-step method" which is the second-order backward differential formula (BDF) plus Gauss–Seidel iterations for solving the nonlinear system. (According to the author, these iterations perform similarly to the modified Newton method but with less overhead.) The two-step method enables very large step-sizes.

A different approach was taken in [19] by Jacobson and Turco. The 3-D calculations are vectorized around the grid-cell dimension (a very interesting idea) and advantage is taken of the sparse structure of Jacobians and a specific reordering of species (that makes Jacobians close to lower triangular form).

In this paper we look in detail at the widely used QSSA method and demonstrate that significant improvements in the efficiency of this type of method can be achieved. We consider two extrapolation algorithms based on QSSA. In particular we obtain an order-2 method that uses two function evaluations per step which we call the *extrapolated QSSA* method. We also construct a nontrivial modification of the well-known GBS (Gragg–Bulirsch–Stoer) extrapolation algorithm based on an ap-

propriate QSSA modification. In particular we obtain an order-2 method that uses three function evaluations per step which we call (for good reason) the *symmetric QSSA* method. In the stiff case the extrapolated methods no longer have a higher order than the *plain QSSA* method does. Nevertheless, we prove that the *extrapolated QSSA* method and the *symmetric QSSA* method have a smaller error constant, which explains their superior performance. We also prove that under certain conditions the *plain QSSA* method is convergent when applied to a particular singular perturbation problem. Numerical experiments on a test problem used in a regional scale model are also presented.

2. Plain, DAE, and iterated QSSA. If $y \in \mathbf{R}^n$ denotes the vector of concentrations, the differential equations arising from the chemical mass balance relation can be written in the form

$$\frac{dy_j}{dt} = P_j(y_1, \ldots, y_n) - D_j(y_1, \ldots, y_n) y_j \quad \text{for } j = 1, \ldots, n, \tag{1}$$

where $P_j(y)$ and $D_j(y)y_j$ are *production* and *destruction* terms, respectively. These equations have an exponential analytical solution provided that $P_j(y)$ and $D_j(y)$ are constant. For an initial value $y(t_0) = y_0$ and a step-size h the approximation

$$y_j(t_0 + h) \approx \frac{P_j(y_0)}{D_j(y_0)} - \left(\frac{P_j(y_0)}{D_j(y_0)} - y_{0,j} \right) \cdot e^{-hD_j(y_0)} =: \bar{y}_j(t_0 + h) \tag{2}$$

forms the basis of the QSSA method. For species with a very long *lifetime* $\tau_j = 1/D_j$, i.e., with very small D_j, this equation can be simplified by replacing the exponential term with $1 - hD_j(y_0)$, thus obtaining the explicit Euler formula

$$y_j(t_0 + h) \approx y_{0,j} + h \left(P_j(y_0) - D_j(y_0) y_{0,j} \right). \tag{3}$$

For species with a very short lifetime, i.e., with very large positive D_j, the following steady-state relation is obtained:

$$y_j(t_0 + h) \approx \frac{P_j(y(t_0 + h))}{D_j(y(t_0 + h))}. \tag{4}$$

For short-lived species these equalities form a system of nonlinear equations which is usually solved by a fixed-point iteration scheme. This is in fact equivalent to solving the system of differential-algebraic equations (DAE) obtained by replacing in (1) the differential equations corresponding to short-lived species by their corresponding steady-state equations

$$\begin{aligned} \frac{dy_j}{dt} &= P_j(y_1, \ldots, y_n) - D_j(y_1, \ldots, y_n) y_j, \ j \in \mathcal{J} \ , \\ 0 &= P_i(y_1, \ldots, y_n) - D_i(y_1, \ldots, y_n) y_i, \ i \in \mathcal{I} \ , \end{aligned} \tag{5}$$

where $\mathcal{I}$ is the set of indices corresponding to the short-lived species and the set $\mathcal{J}$ consists of the remaining indices. We call the scheme based on (2)–(3)–(4) the *DAE QSSA* method. This is clearly distinct from the method consisting of applying (2) to all species which will be called the *plain QSSA* method or simply the QSSA method. We note that the *DAE QSSA* method described in this paper is usually known in the literature as the QSSA method and has been extensively used in solving atmospheric chemistry equations.

Consider now the *plain QSSA* scheme. By construction we have

$$\bar{y}(t_0) = y_0 = y(t_0) \ , \qquad \bar{y}'(t_0) = P(y_0) - D(y_0) y_0 = y'(t_0).$$

A simple analysis for the second derivatives at t_0 gives

$$\bar{y}''(t_0) = -D(y_0)\bar{y}'(t_0) = -\,D(y_0)(P(y_0) - D(y_0)y_0)\ ,$$

$$\begin{aligned} y''(t_0) &= P_y(y_0)y'(t_0) - D_y(y_0)(y'(t_0), y(t_0)) - D(y_0)y'(t_0) \\ &= (P_y(y_0) - D(y_0))\,(P(y_0) - D(y_0)y_0) - D_y(y_0)(P(y_0) - D(y_0)y_0, y_0)\ , \end{aligned}$$

showing that $\bar{y}''(t_0) \neq y''(t_0)$ in general. Thus the order of *plain QSSA* is equal to one.

In an attempt to improve *plain QSSA*, the chemists working on atmospheric models have developed the *iterated QSSA* method. The formula (2) is reapplied with P_j and D_j recomputed at the point $y_1 := \bar{y}(t_0 + h)$, giving,

$$\tilde{y}_j(t_0 + h) := \frac{P_j(y_1)}{D_j(y_1)} - \left(\frac{P_j(y_1)}{D_j(y_1)} - y_{0,j}\right) \cdot e^{-hD_j(y_1)}. \tag{6}$$

The work per step is approximately doubled, as compared to *plain QSSA*. Numerical experiments have shown that *iterated QSSA* performs better than *plain QSSA* (in terms of precision/work ratio) only for large tolerances.

3. Extrapolation algorithms based on QSSA. A natural way to build new methods based on QSSA in the hope of better efficiency is to consider extrapolation algorithms. Some extrapolation methods have proved to be successful for very stiff problems arising in chemistry, e.g., extrapolation based on the linearly implicit Euler method or on the linearly implicit midpoint rule, see [2, 8] and [14, Section IV.9]. Therefore, extrapolation cannot be a priori discarded as a viable technique for solving the stiff systems arising in atmospheric chemistry. In general for high accuracy requirements extrapolation to high order is used, but here we are mainly interested in low-order extrapolation since the accuracy requirements in atmospheric chemistry are low. In this paper we will consider two extrapolation algorithms based on QSSA. Extrapolation is based on the existence of an asymptotic expansion in h-powers for the global error. In the presence of stiffness such an expansion does not hold in general, however. Nevertheless, extrapolation may already lead to a certain improvement just by reducing the error constants.

From the nonstiff situation the extrapolation algorithm based on QSSA is defined as follows. Considering a step-size H and a sequence of positive integers $n_1 < n_2 < n_3 < \ldots$, we perform n_j times the QSSA formula (2) with step-size $h_j = H/n_j$, and denote the result by T_{j1}. We then extrapolate these values via the recursion

$$T_{j,k+1} = T_{jk} + \frac{T_{jk} - T_{j-1,k}}{(n_j/n_{j-k}) - 1}. \tag{7}$$

The extrapolated values T_{jk} are approximations of order k to the exact solution $y(t + H)$ in the nonstiff situation.

Another type of extrapolation algorithm makes use of asymptotic expansions in even powers of h. The following algorithm is similar to the well-known GBS algorithm [13, Formula II.9.13] but it is based on QSSA. We compute

$$y_1 = e^{-D(y_0)h}\left(y_0 - D(y_0)^{-1}P(y_0)\right) + D(y_0)^{-1}P(y_0), \tag{8a}$$

$$y_{i+1} = e^{-D(y_i)2h}\left(y_{i-1} - D(y_i)^{-1}P(y_i)\right) + D(y_i)^{-1}P(y_i) \tag{8b}$$

$$\text{for } i = 1, \ldots, 2n - 1$$

and then perform the following step:

$$S_h(t_n) = e^{-D(y_{2n})h}\left(y_{2n-1} - D(y_{2n})^{-1}P(y_{2n})\right) + D(y_{2n})^{-1}P(y_{2n}), \tag{8c}$$

where $t_n = t_0 + 2nh$. The extrapolation algorithm is slightly different. Here, considering a step-size H and a sequence of positive integers $n_1 < n_2 < n_3 < \ldots$, we perform the algorithm (8a)–(8c) with step-size $h_j = H/(2n_j)$ and denote the result by $T_{j1} := S_{h_j}(t_n)$. We then extrapolate these values with the recursion

$$T_{j,k+1} = T_{jk} + \frac{T_{jk} - T_{j-1,k}}{(n_j/n_{j-k})^2 - 1}. \tag{9}$$

The extrapolated values T_{jk} are approximations of order $2k$ to the exact solution in the nonstiff situation.

In the next two sections we analyze what may happen with stiffness by considering a singular perturbation problem and its related reduced system.

4. The reduced system of a singular perturbation problem. Since the differential equations (1) modeling chemical reactions are generally stiff, the well-known phenomenon of order reduction may occur for the integration method [22]. As a simplified model problem for the forthcoming analysis we consider the following *singular perturbation problem*:

$$y' = -\,D_1(y,z)y + P_1(y,z)\ , \tag{10a}$$

$$z' = -\left(\frac{1}{\varepsilon}D_2(y,z) + D_3(y,z)\right)z + \left(\frac{1}{\varepsilon}P_2(y,z) + P_3(y,z)\right) \tag{10b}$$

with $0 < \varepsilon \ll 1$ and $D_2(y,z)$ supposed to be a diagonal matrix strictly positive definite in a neighborhood of the solution. The above division into two classes of species is rather restrictive, but it will give certain insights into the behavior of the different algorithms based on QSSA in the presence of stiffness.

The equations (10a)–(10b) can be rewritten as

$$\begin{aligned} y' &= D_1(y,z)(-y + C_1(y,z))\ , \\ z' &= D_4(y,z)(-z + C_4(y,z))\ , \end{aligned} \tag{11}$$

where

$$\begin{aligned} C_1(y,z) &= D_1(y,z)^{-1}P_1(y,z)\ , \qquad & D_4(y,z) &= \frac{1}{\varepsilon}D_2(y,z) + D_3(y,z)\ , \\ P_4(y,z) &= \frac{1}{\varepsilon}P_2(y,z) + P_3(y,z)\ , \qquad & C_4(y,z) &= D_4(y,z)^{-1}P_4(y,z)\ . \end{aligned}$$

Multiplying the equation (10b) by ε and letting $\varepsilon \to 0$ we obtain the *reduced system*

$$y' = -\,D_1(y,z)y + P_1(y,z) = D_1(y,z)(-y + C_1(y,z)) =: f(y,z)\ , \tag{12a}$$

$$0 = -\,D_2(y,z)z + P_2(y,z) = D_2(y,z)(-z + C_2(y,z)) =: g(y,z)\ . \tag{12b}$$

We assume that

$$g_z(y,z) \text{ is invertible} \tag{13}$$

in a neighborhood of the solution which implies that the differential-algebraic system (12a)–(12b) has *index one* (cf. [14]). This assumption is actually quite natural for species with very short life-times (see (5)). In order to prove the convergence of the QSSA algorithms we will need the stability assumption

(14) $C_2(y,z) = D_2(y,z)^{-1}P_2(y,z)$ is a contraction in z for the norm $\|\cdot\|$

to be satisfied in a neighborhood of the solution. We denote the related contractivity constant by ρ. We will see in Theorem 5.1 that (14) implies (13).

Let us apply the QSSA method to the stiff equations (10a)–(10b). Since $D_2(y,z)$ is a diagonal matrix with strictly positive coefficients we can take the limit $\varepsilon \to 0$ and we obtain

$$y_1 = e^{-D_1(y_0,z_0)h}\,(y_0 - C_1(y_0,z_0)) + C_1(y_0,z_0), \tag{15a}$$

$$z_1 = C_2(y_0,z_0). \tag{15b}$$

This is the definition of the *direct approach* of the QSSA method applied to the reduced problem (12a)–(12b). It will help us later on in Section 5 for the convergence analysis of the QSSA method applied to the singular perturbation problem (10a)–(10b).

Now we restrict our analysis to the differential-algebraic system (12a)–(12b) of index one and the method (15a)–(15b). Differentiating the algebraic equation (12b) with respect to t and omitting the function arguments we obtain

$$z' = (I - C_{2z})^{-1}C_{2y}D_1(-y + C_1).$$

By expanding into Taylor series the exact and the numerical solutions, it can be seen that the local error $\delta y_h(t_0) := y_1 - y(t_0+h)$ and $\delta z_h(t_0) := z_1 - z(t_0+h)$ of the QSSA method (15a)–(15b) is given by

$$\begin{aligned}\delta y_h(t_0) = -\frac{h^2}{2}(&D_{1y_0}\,(D_{10}(-y_0+C_{10}),-y_0+C_{10})\\ &+ D_{10}C_{1y_0}D_{10}(-y_0+C_{10})\\ &+ D_{1z_0}(-y_0+C_{10},z_0') + D_{10}C_{1z_0}z_0') + O(h^3),\end{aligned} \tag{16a}$$

$$\delta z_h(t_0) = -hz_0' + O(h^2), \tag{16b}$$

where the subscript 0 indicates that the function arguments are the initial values (y_0,z_0). We have given the complete expression of the first term of the error because we will make a comparison with some other methods later on. It must be noticed that even if $C_{2z}(y,z) = 0$ the local error remains $\delta y_h(t_0) = O(h^2)$ and $\delta z_h(t_0) = O(h)$. In the following theorem we give a perturbed asymptotic expansion of the global error for a constant step-size application of the method (15a)–(15b).

THEOREM 4.1. *Consider the index one system* (12a)–(12b) *with consistent initial values* (y_0,z_0) *and suppose that* (14) *is satisfied in a neighborhood of the solution. Then the global error of the QSSA method* (15a)–(15b) *at* $t_i = t_0 + ih$ *satisfies for* $ih \le H$,

$$\begin{aligned}y_i - y(t_i) &= ha_1(t_i) + h^2(a_2(t_i) + \alpha_i^2) + O(h^3),\\ z_i - z(t_i) &= h(b_1(t_i) + \beta_i^1) + O(h^2)\ .\end{aligned}$$

The error terms are uniformly bounded for H sufficiently small. The functions $a_1(t)$, $a_2(t)$, and $b_1(t)$ are smooth. The perturbations α_i^2, β_i^1 are independent of h and they do not vanish in general. At t_0 we have $a_1(t_0) = 0$, $a_2(t_0) + \alpha_0^2 = 0$ and $b_1(t_0) + \beta_0^1 = 0$.

Proof. To start the proof, we first show convergence of order-1 for the QSSA method (direct approach). It is worth noting that this part of the proof remains valid for variable step-sizes with $h = \max_i |h_i|$. We use standard techniques (see, e.g., [12, Theorem 4.4] and [14, Theorem VI.7.5]). We denote two neighboring QSSA solutions by $\{\widetilde{y}_n, \widetilde{z}_n\}$, $\{\widehat{y}_n, \widehat{z}_n\}$ and their difference by $\Delta y_n = \widetilde{y}_n - \widehat{y}_n$, $\Delta z_n = \widetilde{z}_n - \widehat{z}_n$. We suppose for the moment that

$$\|\widehat{y}_n - y(t_n)\| \le C_0 h, \quad \|\widehat{z}_n - z(t_n)\| \le C_1 h, \quad \|\Delta y_n\| \le C_2 h^2, \quad \|\Delta z_n\| \le C_3 h. \tag{17}$$

This will be justified by induction below. For the QSSA method (15a)–(15b) we have the inequalities

$$\|\Delta y_{n+1}\| \le \|\Delta y_n\| + O(h\|\Delta y_n\| + h\|\Delta z_n\|), \tag{18a}$$

$$\|\Delta z_{n+1}\| \le \rho \cdot \|\Delta z_n\| + O(\|\Delta y_n\| + h\|\Delta z_n\|) \tag{18b}$$

with $0 \le \rho < 1$. Applying [14, Lemma VI.2.9] we get

$$\|\Delta y_n\| \le C_4 \left(\|\Delta y_0\| + h\|\Delta z_0\|\right),$$

$$\|\Delta z_n\| \le C_5 \left(\|\Delta y_0\| + (h + \rho^n) \cdot \|\Delta z_0\|\right).$$

If (y_n^k, z_n^k) with $k \le n$ denotes the QSSA solution starting on the exact solution at t_k, then the previous formula and (16a)–(16b) imply

$$\|y_n^k - y_n^{k+1}\| \le C_4 \left(\|\delta y_h(t_k)\| + h\|\delta z_h(t_k)\|\right) \le C_6 h^2,$$

$$\|z_n^k - z_n^{k+1}\| \le C_5 \left(\|\delta y_h(t_k)\| + (h + \rho^{n-k-1}) \cdot \|\delta z_h(t_k)\|\right) \le C_7 h^2 + C_8 \rho^{n-k-1} h.$$

Summing up we obtain

$$\sum_{k=0}^{n-1} \|y_n^k - y_n^{k+1}\| \le C_9 h, \qquad \sum_{k=0}^{n-1} \|z_n^k - z_n^{k+1}\| \le C_{10} h + \frac{C_{11}}{1-\rho} h \le C_{12} h.$$

Since the constants C_6, C_7, C_8, C_9, and C_{12} do not depend on the constants C_0, C_1, C_2, and C_3, the assumption (17) is justified by induction on n provided the constants C_0, C_1, C_2, and C_3 are chosen sufficiently large and h sufficiently small.

In the second part of our proof we assume that the step-size h is constant. As in [14, Theorem VI.4.3] we are looking for a perturbed asymptotic expansion of the global error of the form

$$y_i - y(t_i) = \sum_{j=1}^{N} h^j (a_j(t_i) + \alpha_i^j) + O(h^{N+1}),$$

$$z_i - z(t_i) = \sum_{j=1}^{N} h^j (b_j(t_i) + \beta_i^j) + O(h^{N+1})$$

with smooth functions $a_j(t)$, $b_j(t)$, and perturbations α_i^j, β_i^j satisfying $a_j(t_0) + \alpha_0^j = 0$, $b_j(t_0) + \beta_0^j = 0$, and

$$\alpha_i^1 \to 0 \quad \text{for} \quad i \to \infty. \tag{19}$$

For this purpose we construct recursively truncated expansions

$$\widehat{y}_i = y(t_i) + \sum_{j=1}^{M} h^j (a_j(t_i) + \alpha_i^j) + h^{M+1}\alpha_i^{M+1},$$
$$\widehat{z}_i = z(t_i) + \sum_{j=1}^{M} h^j (b_j(t_i) + \beta_i^j),$$

such that when inserted into (15a)–(15b) we have

$$\widehat{y}_{i+1} = e^{-D_1(\widehat{y}_i,\widehat{z}_i)h}\left(\widehat{y}_i - C_1(\widehat{y}_i,\widehat{z}_i)\right) + C_1(\widehat{y}_i,\widehat{z}_i) + O\left(h^{M+2}\right),$$
$$\widehat{z}_{i+1} = C_2(\widehat{y}_i,\widehat{z}_i) + O\left(h^{M+1}\right).$$

We first develop the above expressions into Taylor series at t_i to obtain conditions for the smooth functions. We then develop the terms involved with the perturbations at t_0 to obtain conditions for the perturbations independently of h. Each power of h leads to two types of conditions, one for the smooth functions $a_j(t)$, $b_j(t)$ and the other for the perturbations α_i^j, β_i^j. After some tedious computations we have the following results. For $M = 0$ we simply obtain the condition $\alpha_{i+1}^1 = \alpha_i^1$ for the perturbations. Therefore by the hypothesis (19) we must necessarily have $\alpha_i^1 = 0$ for all $i \geq 0$. For $i = 0$ it implies that $a_1(t_0) = 0$. For $M = 1$ the smooth functions $a_1(t)$ and $b_1(t)$ must satisfy

$$\begin{aligned} 0 = {} & D_{2z}(t)\left(z(t), b_1(t)\right) + D_2(t)\left(z'(t) + b_1(t)\right) \\ & - P_{2z}(t)b_1(t) + D_{2y}(t)\left(z(t), a_1(t)\right) - P_{2y}(t)a_1(t), \end{aligned} \tag{20a}$$

$$\begin{aligned} a_1'(t) = {} & -\frac{1}{2}y''(t) - D_1(t)a_1(t) - D_{1y}(t)\left(y(t), a_1(t)\right) \\ & - D_{1z}(t)\left(y(t), b_1(t)\right) + P_{1y}(t)a_1(t) + P_{1z}(t)b_1(t) \\ & + \frac{1}{2}D_1(t)\left(-D_1(t)y(t) + P_1(t)\right). \end{aligned} \tag{20b}$$

We have used the notation $D_1(t) := D_1\left(y(t), z(t)\right)$, etc. We can compute $b_1(t)$ from (20a) because of the invertibility of the matrix $g_z(t) = -D_{2z}(t)(z(t), \cdot) - D_2(t) + P_{2z}(t)$. We then insert its expression into (20b), leading to a linear differential equation for $a_1(t)$ with initial condition $a_1(t_0) = 0$. Therefore $a_1(t)$ and $b_1(t)$ are determined uniquely from the two above equations. Putting $t = t_0$ in (20a), we have $b_1(t_0) \neq 0$ in general, implying that $\beta_0^1 \neq 0$. For the perturbations β_i^1 and α_i^2 we get the recurrences

$$\beta_{i+1}^1 = D_2(t_0)^{-1}\left(P_{2z}(t_0)\beta_i^1 - D_{2z}(t_0)\left(z(t_0), \beta_i^1\right)\right),$$
$$\alpha_{i+1}^2 = \alpha_i^2 + P_{1z}(t_0)\beta_i^1 - D_{1z}(t_0)\left(y(t_0), \beta_i^1\right).$$

Therefore, in general $\beta_i^1 \neq 0$ and $\alpha_i^2 \neq 0$ for all i. The remainder can be estimated as in part d of the proof of [14, Theorem VI.4.3]. We obtain recurrence relations similar to (18a)–(18b). □

The process of determining the perturbed asymptotic expansion may be continued if the perturbations are exponentially decaying to zero. For $j \geq 2$, $a_j(t)$ and $b_j(t)$ are

computed similarly to $a_1(t)$ and $b_1(t)$, and we obtain other very intricate recurrence relations for α_i^{j+1} and β_i^j. In fact it is not worthwhile to continue this process, because here the aim of computing a perturbed asymptotic expansion is to see if the extrapolated values could be of higher order than one. Unfortunately, this cannot happen since only the smooth function terms $a_1(t)$, $a_2(t)$, and $b_1(t)$ are eliminated by extrapolation, not the perturbation terms $\beta_i^1 \neq 0$ and $\alpha_i^2 \neq 0$. Thus Theorem 4.1 shows that the order of the standard extrapolation (7) of QSSA remains equal to one for all extrapolated values and this is a negative result. We can therefore expect that in the stiff situation the standard extrapolation of the QSSA values will generally not improve the order of the *plain QSSA*. This result was confirmed numerically. Although the order remains equal to one when doing extrapolation, the error constants are actually smaller and this can imply a certain improvement in efficiency for the first values of the extrapolation tableau.

We call the element T_{22} of the extrapolation tableau (7) with $n_1 = 1$ and $n_2 = 2$ the *extrapolated QSSA* method. Applied with a step-size $H = 2h$ this method can be expressed as a multistage method as follows:

$$
\begin{aligned}
Y_1 &= y_0 + (e^{-D(y_0)2h} - 1)(y_0 - C(y_0)),\\
Y_2 &= y_0 + (e^{-D(y_0)h} - 1)(y_0 - C(y_0)),\\
Y_3 &= Y_2 + (e^{-D(Y_2)h} - 1)(Y_2 - C(Y_2)),\\
y_1 &= 2Y_3 - Y_1,
\end{aligned}
\tag{21}
$$

and it necessitates only two function evaluations. It is an order-2 method in the nonstiff case. We analyze what happens to this method when applied to the reduced system (12a)–(12b). We get for the direct approach

$$
\begin{aligned}
Y_1 &= y_0 + (e^{-D_1(y_0,z_0)2h} - 1)(y_0 - C_1(y_0,z_0)), \qquad Z_1 \;=\; C_2(y_0,z_0),\\
Y_2 &= y_0 + (e^{-D_1(y_0,z_0)h} - 1)(y_0 - C_1(y_0,z_0)), \qquad Z_2 \;=\; C_2(y_0,z_0),\\
Y_3 &= Y_1 + (e^{-D_1(Y_2,Z_2)h} - 1)(Y_2 - C_1(Y_2,Z_2)), \qquad Z_3 \;=\; C_2(Y_2,Z_2),\\
y_1 &= 2Y_3 - Y_1, \qquad z_1 \;=\; 2Z_3 - Z_1.
\end{aligned}
$$

Using Taylor series to compute the local error of this method, we arrive at

$$
\begin{aligned}
\delta y_H(t_0) = -\frac{H^2}{2}&(D_{1z0}\left(-y_0 + C_{10}, (I - C_{2z0})^{-1}C_{2y_0}D_{10}(-y_0 + C_{10})\right)\\
&+ D_{10}C_{1z0}(I - C_{2z0})^{-1}C_{2y_0}D_{10}(-y_0 + C_{10})) + O(H^3),\\
\delta z_H(t_0) = H(I - (I - C_{2z0})^{-1})&C_{2y_0}D_{10}(-y_0 + C_{10}) + O(H^2).
\end{aligned}
$$

We clearly see that the local error of this method contains fewer terms than the error (16a)–(16b) of *plain QSSA*. We observe that if $C_{2_z}(y,z) = 0$ the local error of the *extrapolated QSSA* method is $\delta y_H(t_0) = O(H^2)$ and $\delta z_H(t_0) = O(H^2)$. For this method, using a convergence proof similar to that given in Theorem 4.1 for *plain QSSA*, we obtain convergence of order-1 for the y- and z-components but with a smaller error constant.

A similar analysis for the GBS-type algorithm (8a)–(8c) would be very intricate, but it has been observed numerically that there is no significant improvement when the extrapolation algorithm is used. Nevertheless, the first element of the extrapolation tableau with $n_1 = 2$ gives good results. Applied with a step-size $H = 2h$ this multistage method reads

$$\begin{aligned} Y_1 &= y_0 + (e^{-D(y_0)h} - 1)(y_0 - C(y_0)), \\ Y_2 &= y_0 + (e^{-D(Y_1)2h} - 1)(y_0 - C(Y_1)), \\ Y_3 &= Y_1 + (e^{-D(Y_2)h} - 1)(Y_1 - C(Y_2)), \\ y_1 &= Y_3, \end{aligned} \tag{22}$$

and it necessitates three function evaluations. We call this method the *symmetric QSSA* method. It is an order-2 method in the nonstiff case. We analyze what happens to this method when applied to the reduced system (12a)–(12b). We get for the direct approach

$$\begin{aligned} Y_1 &= y_0 + (e^{-D_1(y_0,z_0)h} - 1)(y_0 - C_1(y_0, z_0)), \qquad Z_1 = C_2(y_0, z_0), \\ Y_2 &= y_0 + (e^{-D_1(Y_1,Z_1)2h} - 1)(y_0 - C_1(Y_1, Z_1)), \qquad Z_2 = C_2(Y_1, Z_1), \\ Y_3 &= Y_1 + (e^{-D_1(Y_2,Z_2)h} - 1)(Y_1 - C_1(Y_2, Z_2)), \qquad Z_3 = C_2(Y_2, Z_2), \\ y_1 &= Y_3, \qquad z_1 = Z_3. \end{aligned}$$

Using Taylor series to compute the local error of this method, we arrive at

$$\begin{aligned} \delta y_H(t_0) = \frac{H^2}{2} &\left(D_{1z0}\left(-y_0 + C_{10}, \left(\frac{1}{2}I - (I - C_{2z0})^{-1}\right) C_{2y_0} D_{10}(-y_0 + C_{10})\right)\right. \\ &\left. + D_{10} C_{1z0}\left(\frac{1}{2}I - (I - C_{2z0})^{-1}\right) C_{2y_0} D_{10}(-y_0 + C_{10})\right) \\ &+ O(H^3), \\ \delta z_H(t_0) = H &\left(I + \frac{1}{2}C_{2z0} - (I - C_{2z0})^{-1}\right) C_{2y_0} D_{10}(-y_0 + C_{10}) + O(H^2). \end{aligned}$$

We clearly see that the local error of this method contains fewer terms than the error (16a)–(16b) of *plain QSSA*. It is also clear here that the extrapolation algorithm (9) cannot increase the order because the error of the first element of the extrapolation tableau does not have an asymptotic expansion in even powers of H. In fact any explicit method of QSSA type cannot be of order greater than one for the reduced system (12a)–(12b) because of the presence of the expression $(I - C_{2z})^{-1}$ in the first derivative of the exact solution for the z-component. We observe that if $C_{2z}(y, z) = 0$ the local error of the *symmetric QSSA* method is $\delta y_H(t_0) = O(H^2)$ and $\delta z_H(t_0) = O(H^2)$. For this method, using a convergence proof similar to that given in Theorem 4.1 for *plain QSSA*, we obtain convergence of order-1 for the y- and z-components but with a smaller error constant.

5. Convergence of QSSA for the singular perturbation problem. In this section we give a proof of convergence under certain conditions of the *plain QSSA* method when applied to the singularly perturbation problem (10a)–(10b).

Because we are mainly interested in smooth solutions to (10a)–(10b) (see [14, Section VI.2]) we require as a stability assumption that the logarithmic norm of $g_z(y,z)$ satisfies [14, Formula VI.2.11]

$$\mu(g_z(y,z)) < 0 \tag{23}$$

in an ε-independent neighborhood of the solution. By definition, the logarithmic norm of a matrix A is given by

$$\mu(A) = \lim_{h\to 0, h>0} \frac{\|I + hA\| - 1}{h}, \tag{24}$$

where I is the identity matrix.

In the following theorem we show that the stability assumptions (14) and (23) may be related for the matrix norm induced by the max-norm $\|z\|_\infty = \max_{i=1}^n |z_i|$. For other norms some counterexamples below demonstrate that the two assumptions are unrelated.

THEOREM 5.1. *In a neighborhood of the solution*

1. *if $C_2(y,z)$ is a contraction in z for the norm $\|\cdot\|$, then $g_z(y,z)$ is invertible;*
2. *if in addition $D_2(y,z)$ is diagonal and strictly positive definite, and the induced matrix norm of any diagonal matrix $D = diag(d_{11},\ldots,d_{nn})$ satisfies $\|D\| = \max_{i=1}^n |d_{ii}|$, then the real parts of the eigenvalues of $g_z(y,z)$ are strictly negative;*
3. *moreover, if $C_2(y,z)$ is a contraction in z for the max-norm, then for the induced logarithmic norm we have*

$$\mu_\infty(g_z(y,z)) < 0 \ .$$

Conversely

4. *if $\mu(g_z(y,z)) < 0$, then the real parts of the eigenvalues of $g_z(y,z)$ are strictly negative and $g_z(y,z)$ is therefore invertible;*
5. *if $\mu_\infty(g_z(y,z)) < 0$, $D_2(y,z)$ is diagonal and strictly positive definite, and the diagonal elements of $C_{2z}(y,z)$ are nonnegative, then $C_2(y,z)$ is a contraction in z for the max-norm.*

Proof. For part 1 we rewrite

$$g(y,z) = D_2(y,z)\,(-z + C_2(y,z))\,.$$

Differentiating this expression with respect to z leads to

$$g_z(y,z) = D_{2z}(y,z)\,(-z + C_2(y,z)) + D_2(y,z)\,(-I + C_{2z}(y,z))\,.$$

Since $g(y_0,z_0) = 0$ and $D_2(y_0,z_0)$ is invertible we have $-z_0 + C_2(y_0,z_0) = 0$. Hence we get

$$g_z(y_0,z_0) = D_2(y_0,z_0)\,(-I + C_{2z}(y_0,z_0))\,.$$

Because $C_2(y,z)$ is a contraction in z with constant ρ for the given norm $\|\cdot\|$ we have equivalently for the induced matrix norm

$$\|C_{2z}(y,z)\| \le \rho < 1 \ . \tag{25}$$

Since $D_2(y,z)$ is invertible, this completes the proof of the first part of the theorem.

For part 2 we suppose by contradiction that there exists an eigenvalue λ of $g_z(y,z)$ with a nonnegative real part. We denote by $v \neq 0$ a corresponding eigenvector. We will show that $v = 0$, giving the desired contradiction. We use the notation $D := D_2(y_0,z_0)$ and $C := C_{2z}(y_0,z_0)$. We have $D(C-I)v = \lambda v$ which implies that $(I + \lambda D^{-1} - C)v = 0$. We thus obtain

$$(I + \lambda D^{-1})(I - \left(I + \lambda D^{-1}\right)^{-1} C)v = 0 \ .$$

The matrix $I + \lambda D^{-1}$ is clearly invertible. The matrix $I - (I + \lambda D^{-1})^{-1} C$ is invertible, too, because of the estimate

$$\| \left(I + \lambda D^{-1}\right)^{-1} C\| \leq \| \left(I + \lambda D^{-1}\right)^{-1} \| \cdot \|C\| \leq \frac{1}{|1 + \lambda / \max_{i=1}^{n} d_{ii}|} \cdot \rho \leq \rho < 1.$$

We thus arrive at the contradiction $v = 0$.

For part 3, the logarithmic norm associated with the max-norm of a matrix A is given by [13, Formula I.10.20']

$$\mu_\infty(A) = \max_{i=1}^{n} \left(a_{ii} + \sum_{j \neq i} |a_{ij}| \right) .$$

For the matrix $C - I$ we get

$$\mu_\infty(C-I) = \max_{i=1}^{n} \left(c_{ii} - 1 + \sum_{j \neq i} |c_{ij}| \right) \leq \max_{i=1}^{n} \left(\sum_{j=1}^{n} |c_{ij}| \right) - 1 = \|C\|_\infty - 1 < 0.$$

For the matrix $D(C-I)$ we thus have the estimate

$$\mu_\infty \left(D(C-I)\right) = \max_{i=1}^{n} \left(d_{ii} \left(c_{ii} - 1 + \sum_{j \neq i} |c_{ij}| \right) \right) \leq \min_{i=1}^{n} d_{ii} \cdot \mu_\infty(C-I) < 0.$$

Conversely, for part 4, if a matrix A satisfies $\mu(A) < \alpha$ then the real part of the eigenvalues of A are strictly smaller than α. This result is a simple consequence of the definition of the logarithmic norm (24). We suppose by contradiction that there exists an eigenvalue λ of A satisfying $\mathrm{Re}(\lambda) \geq \alpha$ with a corresponding eigenvector v of unit norm. We have for $h > 0$ sufficiently small

$$\frac{\|(I + hA)v\| - 1}{h} = \frac{|1 + h\lambda| - 1}{h} \geq \frac{1 + h\mathrm{Re}(\lambda) - 1}{h} = \mathrm{Re}(\lambda) \geq \alpha \ ,$$

implying that $\mu(A) \geq \alpha$ and giving the desired contradiction.

Finally for the last part, we have by hypothesis that

$$\mu_\infty \left(D(C-I)\right) = \max_{i=1}^{n} \left(d_{ii} \left(c_{ii} - 1 + \sum_{j \neq i} |c_{ij}| \right) \right) < 0.$$

Since d_{ii} and c_{ii} are supposed to be positive we obtain

$$\sum_{j=1}^{n} |c_{ij}| - 1 < 0 \text{ for all } i.$$

Thus we get

$$\|C\|_\infty = \max_{i=1}^{n} \left(\sum_{j=1}^{n} |c_{ij}| \right) < 1. \qquad \square$$

Here is a counterexample which shows that (14) does not imply (23) in general. We take

$$C_2(y,z) = \begin{pmatrix} 0.9z_2 \\ 0.9z_1 \end{pmatrix}, \qquad D_2(y,z) = \begin{pmatrix} 0.1 & 0 \\ 0 & 10 \end{pmatrix}.$$

Although $C_2(y,z)$ is a contraction for the 1-norm $\|z\|_1 = \sum_{i=1}^{n} |z_i|$ and the Euclidean norm $\|z\|_2 = (\sum_{i=1}^{n} |z_i|^2)^{1/2}$, for the corresponding induced logarithmic norms (see [13, Theorem I.10.5]) we have $\mu_1(g_z(y,z)) = 8.9$ and $\mu_2(g_z(y,z)) \approx 1.67$. However, we can notice that $C_2(y,z)$ is a contraction for the max-norm and $\mu_\infty(g_z(y,z)) = -0.01$. Most common matrix norms satisfy the condition enounced in the part 2 of Theorem 5.1, e.g., for all norms induced by the p-norms $\|z\|_p = (\sum_{i=1}^{n} |z_i|^p)^{1/p}$ with $p \geq 1$. Here is a counterexample for a norm which cannot satisfy this condition. We take

$$C_2(y,z) = \begin{pmatrix} 5.9z_1 - 5z_2 \\ 5z_1 - 4.1z_2 \end{pmatrix}, \qquad D_2(y,z) = \begin{pmatrix} 200 & 0 \\ 0 & 2 \end{pmatrix}.$$

The spectral radius of $C_{2z}(y,z)$ is equal to 0.9, hence there exists a norm $\|\cdot\|$ for which $\|C_{2z}(y,z)\| < 0.95$ say, but an eigenvalue of $g_z(y,z)$ is approximately equal to 978.99. A concrete example in $\mathbf{R}^2$ of a norm whose induced matrix norm does not satisfy the hypothesis in part 2 of Theorem 5.1 is given by $\|z\| = |z_2| + |z_2 - z_1|$. There are also counterexamples for the converse part of Theorem 5.1. We choose

$$C_2(y,z) = \begin{pmatrix} 2z_2 \\ 0 \end{pmatrix}, \qquad D_2(y,z) = \begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix}.$$

For the 1-norm and the Euclidean norm we have $\mu_1(g_z(y,z)) = -1$ and $\mu_2(g_z(y,z)) \approx -0.69$, but $\|C_{2z}(y,z)\|_1 = 2$ and $\|C_{2z}(y,z)\|_2 = 2$; i.e., $C_2(y,z)$ is not a contraction for these norms.

We now analyze the behavior of the QSSA method when applied to the singular perturbation problem (10a)–(10b). We will do an analysis similar to that in [14, Section VI.2]. We are mainly interested in smooth solutions of the form

$$y(t) = y^0(t) + \varepsilon y^1(t) + \varepsilon^2 y^2(t) + \ldots, \tag{26a}$$

$$z(t) = z^0(t) + \varepsilon z^1(t) + \varepsilon^2 z^2(t) + \ldots. \tag{26b}$$

Inserting these expansions into (10a)–(10b), multiplying (10b) by ε, and comparing equal powers of ε we get for ε^0,

$$\begin{aligned} y^{0\prime} &= -\, D_1(y^0, z^0) y^0 + P_1(y^0, z^0), \\ 0 &= -\, D_2(y^0, z^0) z^0 + P_2(y^0, z^0), \end{aligned}$$

for ε^1,

$$\begin{aligned} y^{1\prime} =& - D_{1y}(y^0,z^0)(y^0,y^1) - D_1(y^0,z^0)y^1 + P_{1y}(y^0,z^0)y^1 - D_{1z}(y^0,z^0)(y^0,z^1) \\ & + P_{1z}(y^0,z^0)z^1, \\ z^{0\prime} =& - D_{2y}(y^0,z^0)(z^0,y^1) + P_{2y}(y^0,z^0)y^1 - D_{2z}(y^0,z^0)(z^0,z^1) - D_2(y^0,z^0)z^1 \\ & + P_{2z}(y^0,z^0)z^1 - D_3(y^0,z^0)z^0 + P_3(y^0,z^0), \end{aligned}$$

etc. For the QSSA method

$$y_{n+1} = y_n + k_n, \qquad k_n = (e^{-D_{1,n}h} - 1)(y_n - C_{1,n}), \tag{27a}$$

$$z_{n+1} = z_n + \ell_n, \qquad \ell_n = (e^{-D_{4,n}h} - 1)(z_n - C_{4,n}), \tag{27b}$$

we consider similar expansions

$$y_n = y_n^0 + \varepsilon y_n^1 + \varepsilon^2 y_n^2 + \dots, \qquad k_n = k_n^0 + \varepsilon k_n^1 + \varepsilon^2 k_n^2 + \dots, \tag{28a}$$

$$z_n = z_n^0 + \varepsilon z_n^1 + \varepsilon^2 z_n^2 + \dots, \qquad \ell_n = \ell_n^0 + \varepsilon \ell_n^1 + \varepsilon^2 \ell_n^2 + \dots. \tag{28b}$$

We use the notation $D_{1,n}$ for $D_1(y_n, z_n)$, $D_{1,n}^0$ for $D_1(y_n^0, z_n^0)$, etc. $C_4(y,z)$ can be developed in powers of ε as follows:

$$C_4 = D_2^{-1}(1 + \varepsilon D_2^{-1} D_3)^{-1}(P_2 + \varepsilon P_3) = C_2 + \varepsilon D_2^{-1}(-D_2^{-1} D_3 P_2 + P_3) + O(\varepsilon^2).$$

THEOREM 5.2. *Consider the singular perturbation problem* (10a)–(10b) *with* $D_2(y,z)$ *diagonal and strictly positive definite, satisfying the assumptions* (14) *and* (23) *for the max-norm, and admitting a smooth solution of the form* (26a)–(26b) *with initial values* (y_0, z_0). *Then for any fixed constant* $c > 0$ *the global error of the QSSA method* (27a)–(27b) *satisfies for* $\varepsilon \le ch$

$$y_n - y(t_n) = O(h)\ , \qquad z_n - z(t_n) = O(h)$$

uniformly for $h \le h_0$ *and* $nh \le const.$

Before giving the proof of this theorem we first need a perturbation lemma.

LEMMA 5.3. *Consider the singular perturbation problem* (10a)–(10b) *with* $D_2(y,z)$ *diagonal and strictly positive definite, satisfying the assumptions* (14) *and* (23) *for the max-norm and the QSSA method* (27a)–(27b). *Assume that* $\|\widehat{z}_n - C_2(\widehat{y}_n, \widehat{z}_n)\|_\infty \le Ah$, $\|\widehat{y}_n - y_n\|_\infty \le Bh$, $\|\widehat{z}_n - z_n\|_\infty \le Ch$, $\|\delta_n\|_\infty \le Dh$, *and* $\|\theta_n\|_\infty \le Eh$. *Then for any fixed constant* $c > 0$, *the perturbed values*

$$\widehat{y}_{n+1} = e^{-D_1(\widehat{y}_n, \widehat{z}_n)h}(\widehat{y}_n - C_1(\widehat{y}_n, \widehat{z}_n)) + C_1(\widehat{y}_n, \widehat{z}_n) + \delta_n, \tag{29a}$$

$$\widehat{z}_{n+1} = e^{-D_4(\widehat{y}_n, \widehat{z}_n)h}(\widehat{z}_n - C_4(\widehat{y}_n, \widehat{z}_n)) + C_4(\widehat{y}_n, \widehat{z}_n) + \theta_n \tag{29b}$$

satisfy

$$\|\widehat{y}_{n+1} - y_{n+1}\|_\infty \le (1 + Fh)\|\widehat{y}_n - y_n\|_\infty + Gh\|\widehat{z}_n - z_n\|_\infty + \|\delta_n\|_\infty, \tag{30a}$$

$$\|\widehat{z}_{n+1} - z_{n+1}\|_\infty \le K\|\widehat{y}_n - y_n\|_\infty + (\alpha + Lh)\|\widehat{z}_n - z_n\|_\infty + \|\theta_n\|_\infty \tag{30b}$$

for $\varepsilon \leq ch$, $h \leq h_0$, where $\alpha < 1$. The constants F, G, and K do not depend on the constants A, B, C, D, and E. The constant L depends on the constants A, B, and C.

Proof. For the y-component the result is obtained by direct estimation. For the z-component this result is proved by applying the mean value theorem. We consider the vector-valued function

$$F(y,z) = e^{-D_4(y,z)h}(z - C_4(y,z)) + C_4(y,z)$$

for (y,z) in a $O(h)$-neighborhood of (y_n, z_n). Its partial derivatives are

$$F_y(y,z) = e^{-D_4(y,z)h}\left(-hD_{4y}(y,z)(z - C_4(y,z)) + I - C_{4y}(y,z)\right) + C_{4y}(y,z),$$

$$F_z(y,z) = e^{-D_4(y,z)h}\left(-hD_{4z}(y,z)(z - C_4(y,z)) + I - C_{4z}(y,z)\right) + C_{4z}(y,z).$$

We have the estimate $\|F_y(y,z)\|_\infty \leq k$ independently of the constants A, B, C, D, and E. In this lemma we consider by hypothesis values satisfying $\|z - C_2(y,z)\|_\infty \leq \ell h$ where the constant ℓ depends on the constants A, B, and C. We get (omitting the function arguments)

$$\|F_z\|_\infty \leq m\ell h + \|e^{-D_4 h}(I - C_{2z}) + C_{2z}\|_\infty,$$

where m is independent of the constants A, B, C, D, and E. For $h > 0$ we have

$$\begin{aligned}
\|e^{-D_4 h}(I - C_{2z}) + C_{2z}\|_\infty &= \max_{i=1}^{n}\left(e^{-D_{4ii}h} + (1 - e^{-D_{4ii}h})\sum_{j=1}^{n}|C_{2zij}|\right) \\
&\leq \max_{i=1}^{n}\left(e^{-D_{4ii}h} + (1 - e^{-D_{4ii}h})\|C_{2z}\|_\infty\right) \\
&\leq \alpha < 1
\end{aligned}$$

as a consequence of $e^{-D_{4ii}h} + (1 - e^{-D_{4ii}h})\rho \leq \alpha < 1$ for all i. ∎

We are now in position to give a proof of Theorem 5.2.

Proof of Theorem 5.2. We insert y_n^0 and z_n^0 into the QSSA method (27a)–(27b). According to Theorem 4.1 the reduced system is convergent of order-1 so that $\|z_n^0 - C_2(y_n^0, z_n^0)\|_\infty \leq Ah$. The defects satisfy

$$\delta_n = y_{n+1}^0 - y_n^0 - \left(e^{D_1(y_n^0, z_n^0)h} - 1\right)\left(y_n^0 - C_1(y_n^0, z_n^0)\right) = 0$$

$$\begin{aligned}
\theta_n &= z_{n+1}^0 - C_4(y_n^0, z_n^0) - e^{D_4(y_n^0, z_n^0)h}\left(z_n^0 - C_4(y_n^0, z_n^0)\right) \\
&= z_{n+1}^0 - C_2(y_n^0, z_n^0) - e^{D_4(y_n^0, z_n^0)h}\left(z_n^0 - C_2(y_n^0, z_n^0)\right) + O(\varepsilon) = O(h)\ ,
\end{aligned}$$

i.e., $\|\theta_n\|_\infty \leq Eh$. Denoting $\Delta y_n = y_n^0 - y_n$ and $\Delta z_n = z_n^0 - z_n$, we assume that (y_n, z_n) and (y_n^0, z_n^0) satisfy

$$\|\Delta y_n\|_\infty \leq Bh\ , \qquad \|\Delta z_n\|_\infty \leq Ch; \tag{31}$$

this will be justified by induction below. We apply Lemma 5.3 to obtain

$$\|\Delta y_{n+1}\|_\infty \leq (1 + Fh)\|\Delta y_n\|_\infty + Gh\|\Delta z_n\|_\infty,$$

$$\|\Delta z_{n+1}\|_\infty \leq K\|\Delta y_n\|_\infty + (\alpha + Lh)\|\Delta z_n\|_\infty + Eh,$$

where the constants F, G, and K do not depend on the constants A, B, C, and E. The constant L depends on the constants A, B, and C but does not vary with n. We can apply [14, Lemma VI.2.9] to get the desired result. The hypotheses (31) are satisfied by induction on n provided the constants A, B, and C are chosen sufficiently large and h is sufficiently small, but independently of ε. □

6. Description of the test problem. To test the properties of different numerical methods we have chosen the Carbon Bond Mechanism IV (CBM-IV) (Gery et al., [10]), consisting of 32 chemical species involved in 70 thermal and 11 photolytic reactions. The concentration of H_2O is held fixed throughout the simulation. This mechanism is designed for the numerical simulation of chemical processes in urban and in regional scale models. For the numerical experiments the chemical mechanism is run for a simulation time of 5 days. The rate constants and initial conditions follow the IPCC[1] Chemistry Intercomparison study (see [21]) scenario 3 ("Bio"). An operator-splitting environment[2] is simulated with a time step of 20 [*minutes*] for the transport scheme. Emission levels of 0.01 [*ppb/hour*] of NO, 0.01 [*ppb/hour*] of NO_2 and 0.1 [*ppb/hour*] of isoprene are considered. These emissions are injected in the system in equal quantities at the beginning of each 20 [*minutes*] interval.

To describe the stiffness of the problem we have computed both the eigenvalues λ_i of the Jacobian and the destruction rates D_i. The relation $-D_i \approx \lambda_i$ allows us to associate the eigenvalues with the largest negative real parts to certain short-lived species. For the real part of the spectrum we have found the following values: $-8.11 \cdot 10^8$ $[O(^1D)]$, $-8.26 \cdot 10^4$ $[O(^3P)]$ $-2.47 \cdot 10^3$ $[ROR]$, -3.5 $[OH]$, -4.2 $[TO_2]$, all others being in the interval $[-0.14, \ -10^{-8}]$. The problem is very stiff since time steps of 1 [*nanosecond*] are prohibitively small considering an integration interval of 20 [*minutes*] and the low accuracy required. For this problem the fact that the eigenvalues with the largest negative real parts are isolated and can be associated with certain species indicates that the singular perturbation model (10a)–(10b) makes sense.

7. Numerical results. In this section the results for the test problem are compared with the solution computed by the code RADAU5 of Hairer and Wanner [14] with very tight tolerances rtol $= 10^{-12}$ and atol $= 10^{-10}$ $[molecules/cm^3]$.

As a measure of the accuracy we have employed the *number of accurate digits (NAD)* computed as follows

$$NAD = \frac{1}{N} \sum_{i=1}^{N} NAD_i, \qquad NAD_i = -\log_{10}(ERR_i),$$

where N is the number of species, ERR_i a measure of the relative error in the numerical solution of species i, and NAD_i the corresponding number of accurate digits. With the "exact" solution $y(t)$ (computed by RADAU5) and the numerical solution $\hat{y}(t)$ at hand at discrete times $\{t_j = t_0 + j \cdot \Delta t, \ 0 \le j \le M\}$ the measure of the relative error is computed as follows:

$$ERR_i = \sqrt{\frac{1}{|\mathcal{J}_i|} \cdot \sum_{j \in \mathcal{J}_i} \left| \frac{y_i(t_j) - \hat{y}_i(t_j)}{y_i(t_j)} \right|^2}, \qquad \mathcal{J}_i = \{0 \le j \le M : \ |y_i(t_j)| \ge a\}.$$

[1]Intergovernmental Panel on Climate Change.

[2]The atmospheric convection–diffusion–reaction equation is solved with the method of fractional steps [32]; chemistry and transport are considered separately and integrated with different step-sizes.

The threshold factor used here is $a = 1$ $[molecules/cm^3]$. If the set $\mathcal{J}_i$ is empty, the value of ERR_i is neglected. The purpose of considering the above defined error measure instead of the root mean square norm ($a = 0$ $[molecules/cm^3]$) is to suppress from the error calculation the values at times when the absolute value of the concentration falls below $a = 1$ $[molecules/cm^3]$; these values are very likely corrupted and the corresponding large relative errors say nothing about the general computational accuracy. From a physical standpoint, for atmospheric chemistry applications, values of $a = 1$ $[molecules/cm^3]$ or less can be assimilated to a complete extinction of the species.

In what follows we denote the current step-size by h. The integrators used are the following:

1. *Plain QSSA* integrates all the species with formula (2).
2. *DAE QSSA* is used with a dynamic partition of the species into slow, fast, and normal. At each time step we have:
   - If $\tau_i > 100 \cdot h$ the species is slow and is integrated with (3).
   - If $\tau_i < 0.1 \cdot h$ the species is fast and is integrated with (4).
   - Otherwise, formula (2) is applied.
3. *Iterated QSSA* is similar to DAE QSSA but has one extra iteration (6).
4. *CHEMEQ* (see [33], implemented in CALGRID) is used as specified in [25]:
   - If $\tau_i < 0.2 \cdot h$ the species is fast and is integrated with (4).
   - If $\tau_i > 5 \cdot h$ the species is slow and is integrated with the nonstiff CHEMEQ formula.
   - For all other species the CHEMEQ stiff formula is used.
5. *Extrapolated QSSA* (21) uses the difference $y_1 - Y_3 = Y_3 - Y_1$ in the error estimator for the step-size control.
6. *Symmetric QSSA* (22) uses for the step-size control the difference $y_1 - Y_4$ where Y_4 is just one cheap extra QSSA step using the function evaluation at y_0 needed for Y_1,

$$Y_4 = y_0 + (e^{-D(y_0)2h} - 1)(y_0 - C(y_0)).$$

7. *TWOSTEP* is based on the variable step size, two-step BDF2 [28, 29, 31]. Instead of a modified Newton process, Gauss–Seidel iterations are used for solving the nonlinear system of equations. This technique carefully exploits the production-loss form of the differential equation (see [28] for details). We have used the original implementation obtained directly from the authors. To accelerate the convergence of the Gauss–Seidel iterations, the species have been sorted in decreasing order relative to the size of their destruction rates.
8. *VODE* (a BDF code, see [3, 4]) is similar to LSODE (Livermore Solver for ODE, see [17]), widely used by atmospheric modelers. VODE is considered to have several advantages over LSODE when used to integrate systems of ODE arising from chemical kinetics (see [4]). In order to take full advantage of the sparsity pattern of the Jacobian, VODE has been modified as described in [23] by replacing the general factorization and substitution routines `dgefa` and `dgesl` with specialized sparse routines. Results for both the standard and the modified VODE are presented.

All integrators have been used with a lower bound of 0.01 [*seconds*] imposed on the step-size.

The emissions of NO, NO_2, and $ISOP$ introduce transient regimes at the beginning of each hourly interval. At these moments, a complete restart is carried

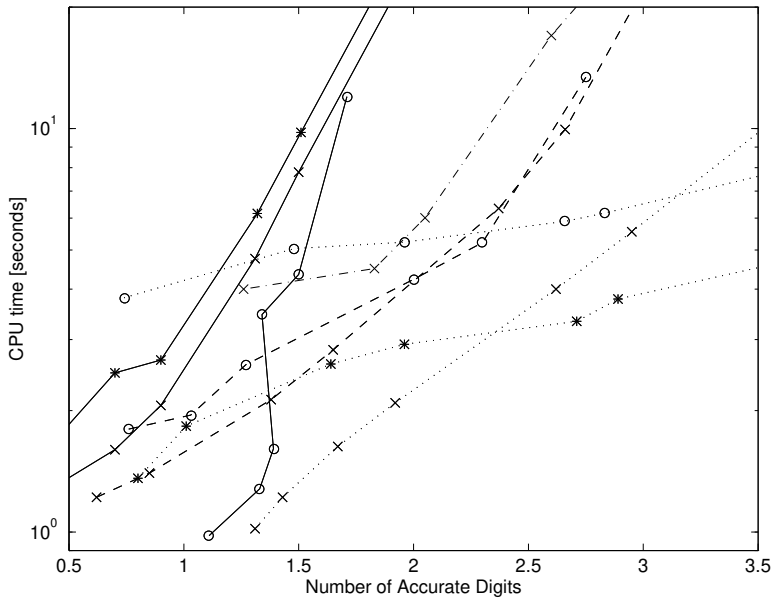

FIG. 1. *Work-precision diagram for* CBM-IV. *Plain QSSA (solid with "∗"), DAE QSSA (solid with "x"), iterated QSSA (solid with "o"), CHEMEQ (dashed-dotted with "x"), extrapolated QSSA (dashed with "x"), symmetric QSSA (dashed with "o"), TWOSTEP (dotted with "x"), VODE (dotted with "o"), and sparse VODE (dotted with "∗").*

out for all integrators. More exactly, an exit from the integration subroutine is performed, the step-size is reset to its default value of 1 [*second*], and the subroutine is called again.[3] In a 3-D operator-splitting model each two consecutive calls to the chemical kinetics integrator are separated by a step of the transport scheme, which may change significantly the concentration values. As a consequence, for comprehensive atmospheric models, a periodic restart of the chemical integrator is a necessity.

Figure 1 reports the CPU time versus the NDA for the different integrators.

The efficiency of *plain QSSA* is improved by treating with *DAE QSSA* separately the steady-state species on one hand and the slow species on the other hand. This conclusion is in agreement with the practical experience of QSSA users.

The extra function evaluation used in *iterated QSSA* pays back for large values of rtol; if more accuracy is needed then this strategy is not better than the classic *DAE QSSA* approach. Several numerical tests have shown that employing more than one iteration decreases the efficiency of *iterated QSSA*.

VODE uses the highest order formulas among the tested algorithms. This fact can be observed from the smaller slope in the work-precision diagram of Figure 1. A high order method pays back when an accurate solution is needed; sparse VODE is the most efficient for 2.5 or more NAD. In the low accuracy range required by atmospheric chemistry simulation the off-the-shelf code VODE is not competitive, since its performance is affected by frequent restarts. This is one of the reasons why atmospheric scientists have chosen to develop their own integrators rather than using general solvers. However, if the linear algebra is done such that full advantage of

[3]Each call to VODE has been done with `istate` = 1.

the structure of the problem is taken (see [23]) the computational time of VODE is greatly reduced[4] and the code becomes competitive.

Extrapolated QSSA and *symmetric QSSA* perform well compared with *DAE QSSA*, *iterated QSSA* or CHEMEQ (especially when a NAD higher than 1 is required) but not better than sparse VODE or TWOSTEP. In 3-D atmospheric models, two accurate digits in the solution of chemical kinetics equations is an acceptable value. More precision is thought to be redundant due to inaccuracies in the transport scheme; less precision can have an unpredictable effect on the overall accuracy through the transport scheme with an operator-splitting algorithm. For this level of accuracy (two significant digits) TWOSTEP performs the best among the tested numerical integrators.

A componentwise analysis of the numerical error shows that smooth components like O_3 are integrated correctly by all methods. However, the species involved in fast photochemistry are integrated less precisely. Peaks of error appear exactly during sunset and sunrise periods (although in the measure reported here this is not apparent). The two new methods are more accurate and efficient than the classic QSSA ones or CHEMEQ, but they are not as fast as the BDF codes TWOSTEP and sparse VODE.

The experimental conclusions presented here are restricted to the model used and to the set of algorithms employed. More numerical tests are necessary before drawing a general conclusion. The authors are currently involved in a comprehensive benchmark work that will test most of the old and new algorithms (see [24]).

8. Concluding remarks. QSSA-based algorithms are explicit methods and yet they enjoy a remarkable stability. They behave like implicit methods although their evaluation formula is explicit. Although their relative error can be large, we must mention that their absolute error is small and that the QSSA solutions are close to the exact solution even for rapidly varying components like NO; QSSA-based methods preserve quite well the overall behavior of the solution. This explains why these methods have been successfully employed for many years for problems where relatively large errors are accepted and small computing times are desired.

In [30] the local truncation error for the plain QSSA scheme is shown to be only $O(h)$ for the components with small lifetimes $\tau_i \ll h$. However, numerical experiments have shown that the QSSA solutions still converge to the exact solution. The fact that the local order reduction is not felt globally is in line with the theoretical convergence analysis presented here.

The analysis and experiments show that the most attractive features of QSSA-type methods are their small computational time and their easy coding, while their main weaknesses are their low order and their relatively low accuracy. In an attempt to overcome these weaknesses, the analysis of QSSA has led us to two new methods, the *extrapolated* and the *symmetric QSSA*. They clearly perform better than the classic QSSA versions and the hybrid algorithm CHEMEQ. However, they are not as fast as the BDF codes sparse VODE and TWOSTEP for the test problem presented here. When considering a complete 3-D model involving transport of chemical species, QSSA-type methods allow for lumping of species that results in increased efficiency. Whether the BDF codes will benefit as much from lumping remains to be seen. Preliminary work with TWOSTEP [31] shows promise in this direction.

[4]The use of sparse linear algebra routines with VODE reduces the total computational time for our test problem by a factor between two to three.

All the tested methods show small computational times for low accuracy. To see why computational speed is so important, let us mention that on a HP-935A workstation 1 day of chemistry simulation with *DAE QSSA* (at a single grid-point, with our comprehensive model) takes roughly 1 second. A 3-D model may have $50 \times 50 \times 20$ grid-points (a realistic value for a regional model) and the chemistry must be evaluated at each grid-point. A simple calculation shows that, with a serial code, 1 day of simulation will need at least 15 CPU hours (without counting the transport part and the overhead associated with reading and writing megabytes of data). The net result is that the simulation time is of the same order as the wall clock time. One possible solution is to move the codes on more powerful machines (e.g., a version of STEM-II, see [5], is currently running on a CRAY-C90). Another direction would be to take advantage of the inner parallelism of the problem and to write parallel versions of the simulation codes (some work has also been done in this direction).

Work still needs to be done to develop special integration methods that will improve more dramatically both the speed and the accuracy of the atmospheric chemistry modeling codes.

Acknowledgments. We would like to thank Ernst Hairer for his comments on a preliminary version of sections 3 to 5. We also wish to thank Valeriu Damian-Iordache for his help in coding the test problem.

REFERENCES

[1] R. D. Atkinson, D. L. Baulch, R. A. Cox, R. F. Hampson, Jr., J. A. Kerr, and J. Troe, *Evaluated kinetic and photochemical data for atmospheric chemistry*, J. Chem. Kinetics, 21 (1989), pp. 115–190.

[2] G. Bader and P. Deuflhard, *A semi-implicit mid-point rule for stiff systems of ordinary differential equations*, Numer. Math., 41 (1983), pp. 373–398.

[3] P. N. Brown, G. D. Byrne, and A. C. Hindmarsh, *VODE: A Variable Step ODE Solver*, SIAM J. Sci. Statist. Comput., 10 (1989), pp. 1038–1051.

[4] G. D. Byrne and A. M. Dean, *The numerical solution of some chemical kinetics models with VODE and CHEMKIN* II, Comput, Chem., 17 (1993), pp. 297–302.

[5] G. R. Carmichael, L. K. Peters, and T. Kitada, *A second generation model for regional-scale transport/chemistry/deposition*, Atmospheric Environ., 20 (1986), pp. 173–188.

[6] D. Dabdub and J. H. Seinfeld, *Extrapolation techniques used in the solution of stiff ODEs associated with chemical kinetics of air quality models*, Atmospheric Environ., 29 (1995), pp. 403–410.

[7] R. Dentener and P. Crutzen, *Reaction of N_2O_5 on tropospheric aerosols: Impact of the global distributions of NO_x, O_3 and OH*, J. Geophys. Res., 98 (1993), pp. 7149–7163.

[8] P. Deuflhard, *Recent progress in extrapolation methods for ordinary differential equations*, SIAM Rev., 27 (1985), pp. 505–535.

[9] S. Elliot, R. P. Turco, and M. Z. Jacobson, *Tests on combined projection/forward differencing integration for stiff photochemical family systems at long time step*, Comput. Chem., 17 (1993), pp. 91–102.

[10] M. W. Gery, G. Z. Whitten, J. P. Killus, and M. C. Dodge, *A photochemical kinetics mechanism for urban and regional scale computer modeling*, J. Geophys. Res., 94 (1989), pp. 12925–12956.

[11] W. Gong and H. R. Cho, *A numerical scheme for the integration of the gas phase chemical rate equations in* 3D *atmospheric models*, Atmospheric Environ., 27A (1993), pp. 2147–2160.

[12] E. Hairer, Ch. Lubich, and M. Roche, *The Numerical Solution of Differential-Algebraic Systems by Runge-Kutta Methods*, Springer-Verlag, Berlin, New York, 1989.

[13] E. Hairer, S. P. Norsett, and G. Wanner, *Solving Ordinary Differential Equations* I. *Nonstiff Problems*, Springer-Verlag, Berlin, 1993.

[14] E. Hairer and G. Wanner, *Solving Ordinary Differential Equations* II. *Stiff and Differential-Algebraic Problems*, Springer-Verlag, Berlin, 1991.

[15] O. Hertel, R. Berkowicz, J. Christensen, and O. Hov, *Test of two numerical schemes for use in atmospheric transport-chemistry models*, Atmospheric Environ., 27A (1993), pp. 2591–2611.

[16] E. Hesstvedt, O. Hov, and I. Isaacsen, *Quasi-steady-state-approximation in air pollution modelling: comparison of two numerical schemes for oxidant prediction*, Internat. J. Chem. Kinetics, 10 (1978), pp. 971–994.

[17] A. Hindmarsch, *ODEPACK: A Systematized Collection of ODE Solvers*, North–Holland, Amsterdam, 1983.

[18] D. J. Jacob, J. A. Logan, G. M. Gardner, C. M. Spivakovsky, R. M. Yevich, S. C. Wofsy, S. Sillman, and M. J. Prather, *Factors regulating ozone over the United States and its export to the global atmosphere*, J. Geophys. Res., 98 (1993), pp. 14817–14826.

[19] M. Z. Jacobson and R. P. Turco, *SMVGEAR: A sparse-matrix, vectorized Gear code for atmospheric models*, Atmospheric Environ., 17 (1994), pp. 273–284.

[20] F. W. Lurmann, A. C. Loyd, and R. Atkinson, *A chemical mechanism for use in long-range transport/acid deposition computer modeling*, J. Geophys. Res., 91 (1986), pp. 10905–10936.

[21] J. Olson, M. Prather, T. Berntseu, G. R. Carmichael, R. Chatfield, P. Connell, R. Derwent, L. Horowitz, S. Jin, M. KanaKidou, P. Kasithatla, R. Kotomartini, M. Kuhn, K. Lau, S. Sillman, J. Penner, L. Perliski, F. Stordal, A. Thompson, and O. Wild, *Results from the IPCC Photochemical Model Intercomparison (Photo Comp): Some Insights into Tropospheric Chemistry*, J. Geophys. Res., submitted.

[22] A. Prothero and A. Robinson, *On the stability and accuracy of one-step methods for solving stiff systems of ordinary differential equations*, Math. Comput., 28 (1974), pp. 145–162.

[23] A. Sandu, F. A. Potra, V. Damian, and G. R. Carmichael, *Efficient Implementation of Fully Implicit Methods for Atmospheric Chemistry*, Technical report 79, Department of Mathematics, University of Iowa, Iowa City, IA, 1995.

[24] A. Sandu, J. G. Verwer, M. van Loon, G. R. Carmichael, F. A. Potra, D. Dabdub, and J. H. Seinfeld, *Benchmarking Stiff ODE Solvers for Atmospheric Chemistry Problems*, Technical report 85, Department of Mathematics, University of Iowa, Iowa City, IA, 1996.

[25] R. D. Saylor and G. D. Ford, *On the comparison of numerical methods for the integration of kinetic equations in atmospheric chemistry and transport models*, Atmospheric Environ., 29 (1995), pp. 2585–2593.

[26] D. Shyan-Shu Shieh, Y. Chang, and G. R. Carmichael, *The evaluation of numerical techniques for solution of stiff ODE arising from chemical kinetic problems*, Environ. Software, 3 (1988), pp. 28–38.

[27] S. Sillman, *A numerical solution for the equations of tropospheric chemistry based on an analysis of sources and sinks of odd hydrogen*, J. Geophys. Res., 96 (1991), pp. 20735–20744.

[28] J. G. Verwer, *Gauss-Seidel iterations for stiff ODEs from chemical kinetics*, SIAM J. Sci. Comput., 15 (1994), pp. 1243–1250.

[29] J. Verwer, *Explicit Methods for Stiff ODEs from Atmospheric Chemistry*, Preprint NM-R94, EMEP MSC-W Norwegian Meteorological Institute, P.O. Box 43 Blindern, N-0313 Oslo 3, Norway, 1994.

[30] J. Verwer and M. van Loon, *An evaluation of explicit pseudo-steady-state approximation schemes for stiff ODE systems from chemical kinetics*, J. Comput. Physics, 113 (1994), pp. 347–352.

[31] J. Verwer, J. G. Blom, M. van Loon, and E. J. Spee, *A comparison of stiff ODE solvers for atmospheric chemistry problems*, Atmospheric Environ., 30 (1996), pp. 49–58.

[32] N. N. Yanenko, *The Method of Fractional Steps*, Springer-Verlag, New York, Heidelberg, Berlin, 1971.

[33] T. R. Young and J. P. Boris, *A numerical technique for solving stiff ODE associated with the chemical kinetics of reactive flow problems*, J. Phys. Chem., 81 (1977), pp. 2424–2427.