Nonlinear Analysis: Modelling and Control, 2006, Vol. 11, No. 3, 247–261

Modelling Complex Chemical Processes in Homogeneous Solutions: Automatic Numerical Simulation[*]

O.V. Klymenko, I. B. Svir

Mathematical and Computer Modelling Laboratory
Kharkov National University of RadioElectronics
14 Lenin Avenue, Kharkov, 61166, Ukraine
svir@kture.kharkov.ua

Received: 21.04.2006 **Revised:** 17.07.2006 **Published online:** 01.09.2006

Abstract. Two algorithms for the determination of the necessary limit of local error for the numerical solution of ordinary differential equation (ODE) systems describing homogeneous chemical and biochemical processes, and for the evaluation of their stiffness are developed. The approach for finding the necessary limit of local error of a numerical ODE solver is justified by the proof of the corresponding theorems. The application of the new algorithms implemented in version 2.1 of KinFitSim software to the simulation of real chemical systems is considered on the example of Belousov-Zhabotinsky reaction.

Keywords: stiff ordinary differential equations, Gear's method, homogeneous chemical process, Belousov-Zhabotinsky reaction.

1 Introduction

Although mathematical models of real-world problems are becoming more and more complex, many of them can still be formulated in terms of ordinary differential equations (ODEs). The range of processes which are described by ODEs spans over mechanics, biology, medicine, chemistry and other areas that are of great interest in modern science.

At present, particularly interesting and important problems are found in biology where the study of biochemical reactions continually taking place in living

---

[*] The authors would like to acknowledge NATO for the Reintegration Grant NUKR.RIG.981488.

organisms is crucial to understanding their role in regulating biological processes. If all species participating in a reaction mechanism are uniformly distributed in space the corresponding mathematical model will consist only of rate laws, i.e. of ODEs. Biochemical processes are usually extremely complex and thus their models are particularly demanding to the quality of numerical methods for their solution. However, even simple chemical reactions may require the application of advanced numerical methods for their simulation due to corresponding systems of ODEs being stiff [1]. This happens when rates of chemical reactions are very different which means that some components of the solution change much faster than others. In this case standard numerical methods such as Runge-Kutta methods and Adams methods [1] fail and stiff-stable methods must be used. In addition to their enhanced stability these methods are typically more accurate owing to their implicit nature. However, even the use of appropriate methods cannot guarantee that a numerical solution is adequate as will be shown below.

Consider a general homogeneous chemical process involving n species ($n \geq 2$) which consists of m elementary reaction steps ($m \geq 1$). Another assumption that we make here is that variations of temperature and pressure during the observation period are insignificant and therefore rate constants of individual reactions do not vary with time. Formally, such process may be represented by a matrix stoichiometric equation [1, 2]:

$$\boldsymbol{Y}\boldsymbol{\alpha} = \mathbf{0}, \tag{1}$$

where $\boldsymbol{\alpha} \in \mathbb{Z}^{n\times m}$ is the stoichiometric matrix whose columns correspond to stoichiometric vectors of individual reactions and $\boldsymbol{Y} \equiv \{Y_1, Y_2, \dots, Y_n\}$ is the vector of symbolic species names. The matrix $\boldsymbol{\alpha}$ gives proportions of reacting species in elementary reaction steps and can be represented as the difference of two matrices $\boldsymbol{\alpha} = \boldsymbol{\pi} - \boldsymbol{\rho}$ where $\boldsymbol{\pi}, \boldsymbol{\rho} \in \mathbb{Z}_+^{n\times m}$ are the product and reagent stoichiometric matrices respectively.

The generalised mathematical model of a homogeneous chemical process has been previously described [2] and using the notations introduced here may be presented in the following form:

$$\frac{dy_k}{dt} = \sum_{j=1}^{m} (\pi_{kj} - \rho_{kj}) \Big(k_j^+ \prod_{i=1}^{n} y_i^{\rho_{ij}} - k_j^- \prod_{i=1}^{n} y_i^{\pi_{ij}} \Big), \quad k = \overline{1, n}, \tag{2}$$

where y_k is the concentration of species Y_k, π_{ij} and ρ_{ij} are the elements of the product and reagent stoichiometric matrices respectively, $k_j^+, k_j^- \in \mathbb{R}_+$ are the rate constants of the forward and reverse reactions in jth elementary reaction step, $t \in [0, T]$ where T is the observation period length which is assumed finite.

The system of ODEs (2) is subject to the following initial conditions

$$\boldsymbol{y}(0) = \boldsymbol{y}^0, \tag{3}$$

where $\boldsymbol{y}(t) \equiv \{y_1(t), y_2(t), \ldots, y_n(t)\}^T$ is the vector of concentrations and $\boldsymbol{y}^0 \equiv \{y_1(0), y_2(0), \ldots, y_n(0)\}^T$ is the vector of initial concentrations.

The ODE system (2) is non-linear and therefore cannot be resolved analytically in the general case. Hence there is a need to exploit approximate numerical methods for its solution. There exist a number of general and specialised numerical methods for the solution of systems of ODEs [1, 3–5]. The most popular modern numerical methods for solving ODEs comprise linear multistep methods [5] including different types of predictor-corrector schemes, explicit and implicit Runge-Kutta methods [1, 5]. The backward differentiation formulae (BDF) of different orders [4, 7] (also known as the Gear's method) and implicit Runge-Kutta methods [1, 8] are widely used to treat stiff ODE systems. In computational practice these methods usually incorporate an algorithm for adaptive time step size control, which provides a more or less uniform error distribution over the course of simulation. These algorithms determine the size of the next time step based on the estimated value of local error at the current time step and a predefined limit of local integration error. A typical algorithm of this sort utilises an expression of the form

$$h_{k+1} = \omega h_k \left(\frac{\varepsilon}{r_{k+1}}\right)^{\frac{1}{s+1}}, \tag{4}$$

where h_k and h_{k+1} are the successive time steps, ε is the limit of local error, r_{k+1} is the estimate of local error at $(k+1)$th step, $\omega \leq 1$ is a safety factor against overestimation of the step size and s is the order of a method.

Yet, such an algorithm must also take into account the stability properties of a numerical integration method when choosing the size of the next time step. Stiff problems pose especially harsh restrictions on the size of the integration time step and consequently special methods have been developed for such problems.

However, for some of these methods such as Gear's method [4] or implicit Runge-Kutta methods [1] the stability regions have not been determined or have been established only approximately. Hence, if during the process of numerical solution the adaptive step selection algorithm that is unaware of the stability properties of the pertinent numerical method selects a step size, which is outside of the actual (but unknown) stability region, the simulation is likely to fail or lead to a physically non-realistic solution. In particular, during numerical simulation of a homogeneous chemical reaction mechanism some concentrations may become negative while nevertheless satisfying the prescribed limit of local error. This can lead to obtaining wrong results even though the resulting concentration distributions may look not unreasonable (see below).

Consider, for example, the Belousov-Zhabotinsky reaction mechanism [9] written in symbolic form as

$$\begin{aligned} &A+Y \xrightarrow{k_1} X, && k_1 = 4.72\,\mathrm{l\,mol^{-1}s^{-1}},\\ &X+Y \xrightarrow{k_2} P, && k_2 = 3\times 10^9\,\mathrm{l\,mol^{-1}s^{-1}},\\ &B+X \xrightarrow{k_3} 2X+Z, && k_3 = 1.5\times 10^4\,\mathrm{l\,mol^{-1}s^{-1}},\\ &2X \xrightarrow{k_4} Q, && k_4 = 4\times 10^7\,\mathrm{l\,mol^{-1}s^{-1}},\\ &Z \xrightarrow{k_5} Y, && k_5 = 1\ \mathrm{s^{-1}}, \end{aligned} \tag{5}$$

with the following initial concentrations of the species: $[A]_0 = [B]_0 = 0.066\,\mathrm{M}$, $[Z]_0 = 0.002\,\mathrm{M}$, $[P]_0{=}[Q]_0{=}[X]_0{=}[Y]_0{=}0\,\mathrm{M}$ ($1\,\mathrm{M}{=}1\mathrm{mol\,l^{-1}}$). The mathematical model of the reaction scheme (5) according to the generalised model (2) is

$$\begin{aligned} da/dt &= -k_1 ay,\\ db/dt &= -k_3 bx,\\ dp/dt &= k_2 xy,\\ dq/dt &= k_4 x^2,\\ dx/dt &= k_1 ay - k_2 xy + k_3 bx - 2k_4 x^2,\\ dy/dt &= -k_1 ay - k_2 xy + k_5 z,\\ dz/dt &= k_3 bx - k_5 z, \end{aligned} \tag{6}$$

where concentrations of species are designated with corresponding lower-case letters and initial conditions are as above.

Fig. 1(a) shows the simulation results obtained using the Gear's method with the limit of local error $\varepsilon = 10^{-5}$. Clearly the numerical solution in this case does not exhibit rapid "jumps" that appear in the converged concentrations with a period of approximately $16\,s$ (the converged solution was obtained also using the Gear's method with the limit of local error $\varepsilon = 10^{-12}$; see Fig. 1(b). Instead, non-physical oscillations with small amplitude emerge in the numerical solution around the time of appearance of the first "jump" in the converged solution (Fig. 2). The concentrations of species Y and Z become negative in this region

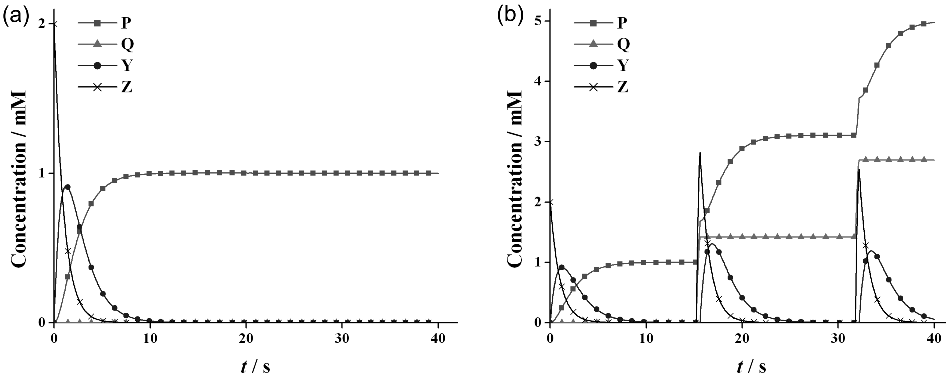

Fig. 1. Computed concentration distributions for the Belousov-Zhabotinsky reaction: (a) numerical solution with limit of local error $\varepsilon = 10^{-5}$; (b) converged solution.

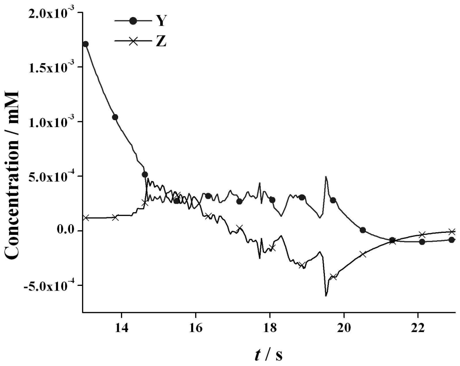

Fig. 2. Oscillations in concentrations of Y and Z computed numerically with limit of local error $\varepsilon = 10^{-5}$.

which causes the numerical solution to converge to some steady solution instead of that for the “oscillating” Belousov-Zhabotinsky reaction (Fig. 1(b)). This however is not obvious from Fig. 1(a) where all shown concentration distributions do not look implausible.

This example shows that systems of ODEs of the form (2) (and especially stiff systems) cannot be integrated properly with an arbitrarily set limit of local error and there is a need in an algorithm for the determination of a threshold of local error that would guarantee obtaining adequate numerical solutions.

In this paper, we formulate and justify two algorithms for the determination of stiffness of the general model of a homogeneous reaction mechanism and for the determination of the threshold of local error which guarantees obtaining physically meaningful results.

2 *A priori* estimation of the stiffness of ODE systems

An ODE system is stiff if the stiffness coefficient defined by

$$S(t) = \max_{1\le i\le n} \mathrm{Re}(-\lambda_i) \Big/ \min_{1\le i\le n} \mathrm{Re}(-\lambda_i), \tag{7}$$

where $\lambda_i,\ i=\overline{1,n}$ are eigen-values of the Jacobian of the right-hand side (r.h.s.) of (2), is much greater than unity [1]. In general it is impossible to determine stiffness of an ODE system prior to its solution because the stiffness coefficient is a function of time. Therefore, even if its value evaluated at $t=0$ (this can be done without solving the system) is close to unity an ODE system cannot be guaranteed to be non-stiff since for $t>0$ the value of $S(t)$ can exceed 1 by several orders of magnitude.

Nevertheless, a robust criterion for the determination of stiffness of an ODE system of the form (2) can be formulated on the basis of *a priori* knowledge of the rate constants k_j^+ and k_j^-, $j=\overline{1,m}$. Generally these rate constants have different dimensions due to different reaction orders, the fact which prevents their direct comparison. To avoid this difficulty we introduce *equivalent* rate constants in the following way. First, we note that for each step j in the mechanism its orders in the forward (+) and reverse (–) directions are expressed respectively as

$$p_j^+ = \sum_{i=1}^{n} \rho_{ij}, \quad p_j^- = \sum_{i=1}^{n} \pi_{ij}. \tag{8}$$

Equivalent forward and reverse rate constants are then defined as

$$k^+_{j,equiv} = k^+_j \; y_{0,max}^{p^+_j - 1}, \quad k^-_{j,equiv} = k^-_j \; y_{0,max}^{p^-_j - 1}, \tag{9}$$

respectively where $y_{0,max}$ is the maximum initial concentration among reacting species. According to the definitions (9), equivalent rate constants have units of s^{-1} and thus the introduction of these definitions may be interpreted as the replacement of all reaction steps in the mechanism with corresponding pseudo-first order reactions. Obviously, not all multi-component reactions may be treated as pseudo-first order ones but this approach allows one to (approximately) compare the relative rates of different reaction steps in the mechanism.

Since the dimensions of all equivalent rate constants (9) are equal (s^{-1}) they can be directly compared with each other. Thus the following *initial* stiffness coefficient can be introduced

$$S_0 = \frac{\max\bigl(\max\limits_{1\leq j\leq m} k^+_{j,equiv}, \max\limits_{1\leq j\leq m} k^-_{j,equiv}\bigr)}{\min\left(\min\limits_{\substack{1\leq j\leq m \\ k^+_{j,equiv}>0}} k^+_{j,equiv}, \min\limits_{\substack{1\leq j\leq m \\ k^-_{j,equiv}>0}} k^-_{j,equiv}\right)}, \tag{10}$$

where the numerator equals maximum among all forward and reverse equivalent rate constants and the denominator equals minimum non-zero equivalent rate constant. We assume that an ODE system under consideration is stiff if $S_0 \geq 100$ and non-stiff otherwise.

In practice the value of S_0 calculated from (10) allows one to choose between the non-stiff (e.g. Adams-Moulton [1, 5]) and stiff (e.g. Gear [4]) solvers. Although the suggested bound (10) usually overestimates the value of the stiffness coefficient (7) it promptly detects problem stiffness which allows avoiding program breakdowns due to incorrect choice of the solution method.

Considering the above example of the Belousov-Zhabotinsky reaction one can evaluate S_0 to be 6.36×10^8 which clearly indicates that the ODE system (6) describing this reaction scheme is stiff.

3 Choosing the limit of local error

Let us rewrite the Cauchy problem (2), (3) in the following (vector) form to simplify notations:

$$\boldsymbol{y}' = \boldsymbol{f}(\boldsymbol{y}), \tag{11}$$

$$\boldsymbol{y}(0) = \boldsymbol{y}^0. \tag{12}$$

The autonomous ODE system (11) satisfies the conditions of the theorem of existence and uniqueness of the solution [10] since the function $\boldsymbol{f}(\boldsymbol{y})$ is continuously differentiable in the area

$$\overline{G} = \left\{ \boldsymbol{y}\colon 0 \le y_i(t) \le b,\ 0 \le t \le T,\ i = \overline{1,n} \right\}, \tag{13}$$

i.e. $f_i \in C^1(\overline{G}),\ i = \overline{1,n}$. It also follows from continuous differentiability of $f_i,\ i = \overline{1,n}$ in $\overline{G}$ that these functions and their derivatives are bounded:

$$\left|f_i(\boldsymbol{y})\right| \le M_0, \quad \left|\frac{df_i}{dy_j}\right| \le M_1, \quad t \in [0,T], \quad i,j = \overline{1,n}. \tag{14}$$

The constant b in (13) is an appropriately chosen upper boundary of the variation of all the functions $y_i(t),\ i = \overline{1,n}$, which may be estimated as

$$b = M_0 T. \tag{15}$$

The exact solution to (11), (12) (or (2), (3)) thus satisfies the following two-sided inequalities

$$0 \le y_i(t) \le b, \quad t \in [0,T], \quad i = \overline{1,n}, \tag{16}$$

since concentrations must remain non-negative and be bounded above, which follows from continuous differentiability of functions $y_i(t),\ i = \overline{1,n}$ on a closed set [11].

We can now formulate the following theorem about the properties of the exact solution to the Cauchy problem (2), (3):

Theorem 1. *For every τ such that $0 < \tau < T$ the exact solution to* (2), (3) *on the interval $[\tau, T]$ satisfies the two-sided inequalities*

$$0 < y_i(t) < b, \quad i = \overline{1,n}, \tag{17}$$

where b is defined in (15).

Proof. Functions $f_i(\boldsymbol{y}),\ i = \overline{1,n}$ are not identical zeros by construction. From this fact and inequalities (16) it follows that for the components $y_k(t)$ of the concentration vector with zero initial conditions ($y_k^0 = 0$) the corresponding r.h.s. functions $f_k(\boldsymbol{y})$ must be strictly positive in the vicinity of $t = 0$. Therefore for sufficiently small $\tau > 0$ all components of $\boldsymbol{y}$ will be strictly positive at $t = \tau$, i.e. $y_i(\tau) > 0,\ i = \overline{1,n}$.

Let us now consider functions $y_i(t),\ i = \overline{1,n}$ on the interval $[\tau, T]$ and prove by contradiction that they do not vanish there. Suppose that one component of the concentration vector, $y_k(t)$, vanishes at $t = t_0$, $t_0 \in [\tau, T]$. $y_k(t)$ cannot vanish at more than a finite number of points of $[\tau, T]$ since otherwise the identity $y_k(t) \equiv 0$ would hold true, which contradicts equation (2). Additionally, $y_k(t) \geq 0$ according to inequality (16). Therefore $t = t_0$ is a minimum point of $y_k(t)$. Then the derivative, dy_k/dt, at this point must also equal zero and change its sign from negative to positive when passing $t = t_0$ from left to right together with the r.h.s. function $f_k(\boldsymbol{y})$.

Consider now the function

$$f_k(\boldsymbol{y}) = \sum_{j=1}^{m} (\pi_{kj} - \rho_{kj}) \Big[k_j^+ \prod_{i=1}^{n} y_i^{\rho_{ij}} - k_j^- \prod_{i=1}^{n} y_i^{\pi_{ij}} \Big] \tag{18}$$

at $t = t_0$. The number of terms in the sum on the r.h.s. of (18) can be reduced since the terms describing elementary reaction steps in which species Y_k does not take part are identical zeros. Denote m_k^ρ the number of elementary reactions in which Y_k is a reactant and $\big\{j_s(k)\big\}_{s=1}^{m_k^\rho}$ is a subset of indexes $j = \overline{1,n}$ corresponding to such reactions. Similarly, m_k^π is the number of elementary reactions in which species Y_k is a product and $\big\{l_s(k)\big\}_{s=1}^{m_k^\pi}$ is a subset of reaction indexes corresponding to such reactions.

Next we note that if Y_k is a reactant in jth reaction then at $t = t_0$ the first term in square brackets in (18) is equal to zero due to a zero multiplier $y_k^{\rho_{kj}}(t_0)$. Likewise, if Y_k is a product of jth reaction then at $t = t_0$ the second term in square brackets in (18) equals zero. Thus the expression for $f_k(\boldsymbol{y})$ at $t = t_0$ can be rewritten as

$$\begin{aligned} f_k\big(\boldsymbol{y}(t_0)\big) &= \sum_{j=1}^{m_k^\rho} \rho_{kj_s(k)} k^-_{j_s(k)} \prod_{\substack{i=\overline{1,n}\\ i\neq k}} y_i^{\pi_{ij_s(i)}}(t_0) \\ &+ \sum_{j=1}^{m_k^\pi} \pi_{kl_s(k)} k^+_{l_s(k)} \prod_{\substack{i=\overline{1,n}\\ i\neq k}} y_i^{\rho_{il_s(i)}}(t_0). \end{aligned} \tag{19}$$

Clearly this expression is strictly positive at $t = t_0$ because all concentrations but $y_k(t_0)$ are non-zero, constants $\rho_{kl_s(k)}$ and $\pi_{kj_s(k)}$ are positive and rate constants $k^-_{j_s(k)}$ and $k^+_{l_s(k)}$ are non-negative. Moreover, there exists at least one non-zero rate constant since otherwise species Y_k does not participate in any reaction step and should be mapped out.

Whence the r.h.s. of (19) is strictly positive which contradicts our supposition that $y_k(t)$ vanishes at $t = t_0$ and has a minimum at this point, i.e. $f_k(\boldsymbol{y}(t_0)) = 0$. On the other hand, $y_k(t)$ cannot reach b according to the definition of the latter in equation (15) which completes the proof. ☐

Let us now apply the Euler's method [3,5,6] to obtain an approximate solution of the Cauchy problem (11), (12). The main iterative formula of the method is

$$\boldsymbol{y}^{i+1} = \boldsymbol{y}^i + h_i \boldsymbol{f}(\boldsymbol{y}^i) + \boldsymbol{\eta}^i, \tag{20}$$

where h_i is the ith integration step length and $\boldsymbol{\eta}^i$ is the vector of discretisation error at ith step whose components can be estimated as

$$\max_{1\leq j\leq n} |\eta^i_j| \leq c_0 h_i^2, \tag{21}$$

where c_0 is a constant independent of t and $\boldsymbol{y}$. Using these notations we formulate the following theorem:

Theorem 2. *If* $f_j \in C^1(\overline{G})$, $j = \overline{1,n}$ *and there exists a solution* $\boldsymbol{y}(t)$ *to the Cauchy problem* (11), (12) *in the closed interval* $[0, T]$ *satisfying the assumptions of Theorem* 1 *then there exists such a limit of local error* ε_0 *of the Euler's method that for any limit of local error* $\varepsilon < \varepsilon_0$ *the inequalities*

$$\bar{c}h < \min_i y^i_j < \max_i y^i_j < b - \bar{c}h, \tag{22}$$

where $h = \max_i h_i$ *and*

$$\bar{c} = \frac{c_0 + 0.5nM_0M_1}{\sqrt{n}M_1}(e^{nTM_1} - 1), \quad j = \overline{1,n} \tag{23}$$

will hold true for an approximate solution computed using the Euler's method (20) *with adaptive step-size control, eq.* (4)*, with the local error limit set at* ε*.*

Proof. Consider the vector of local errors in numerical solution of (11), (12) obtained according to equation (20) on ith integration step,

$$\boldsymbol{\epsilon}^i = \boldsymbol{y}(t_i) - \boldsymbol{y}^i. \tag{24}$$

Its value can be estimated by expanding the exact solution to the Cauchy problem (11), (12) into a Taylor series around t_i and eliminating the terms of the order higher than one:

$$\begin{aligned}\boldsymbol{y}(t_{i+1}) &= \boldsymbol{y}(t_i) + h_i \frac{d\boldsymbol{y}}{dt}\Big|_{t=t_i} + \frac{h_i^2}{2}\frac{d^2\boldsymbol{y}}{dt^2}\Big|_{t=\xi} \\ &= \boldsymbol{y}(t_i) + h_i \boldsymbol{f}\big(\boldsymbol{y}(t_i)\big) + \frac{h_i^2}{2}\frac{d\boldsymbol{f}}{dt}\Big|_{t=\xi} \\ &= \boldsymbol{y}(t_i) + h_i \boldsymbol{f}\big(\boldsymbol{y}(t_i)\big) + \frac{h_i^2}{2}\Big(\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{y}}\frac{d\boldsymbol{y}}{dt}\Big)\Big|_{t=\xi},\end{aligned}$$

where the remainder is written in Lagrange's form with $\xi \in [t_i, t_{i+1}]$ and $\partial\boldsymbol{f}/\partial\boldsymbol{y}$ is the Jacobian of the r.h.s. of (11). The local error on the time step $(i+1)$ can then be written as

$$\begin{aligned}\boldsymbol{\epsilon}^{i+1} &= \boldsymbol{\epsilon}^i + h_i\Big(\boldsymbol{f}\big(\boldsymbol{y}(t_i)\big) - \boldsymbol{f}(\boldsymbol{y}^i)\Big) + \frac{h_i^2}{2}\Big(\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{y}}\frac{d\boldsymbol{y}}{dt}\Big)\Big|_{t=\xi} - \boldsymbol{\eta}^i \\ &= \boldsymbol{\epsilon}^i + h_i\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{y}}\Big|_{\boldsymbol{y}=\boldsymbol{\theta}}\boldsymbol{\epsilon}^i + \frac{h_i^2}{2}\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{y}}\Big|_{t=\xi}\boldsymbol{f}\big(\boldsymbol{y}(\xi)\big) - \boldsymbol{\eta}^i,\end{aligned}$$

where we applied the mean-value theorem of the differential calculus and vector $\boldsymbol{\theta}$ lies between $\boldsymbol{y}(t_i)$ and $\boldsymbol{y}^i$. The norm of the vector of local errors can be estimated as

$$\begin{aligned}\|\boldsymbol{\epsilon}^{i+1}\| &= \|\boldsymbol{\epsilon}^i\| + h_i\Big\|\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{y}}\Big|_{\boldsymbol{y}=\boldsymbol{\theta}}\Big\|\|\boldsymbol{\epsilon}^i\| + \frac{h_i^2}{2}\Big\|\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{y}}\Big|_{t=\xi}\Big\|\big\|\boldsymbol{f}\big(\boldsymbol{y}(\xi)\big)\big\| + \|\boldsymbol{\eta}^i\| \\ &\le (1 + nhM_1)\|\boldsymbol{\epsilon}^i\| + h^2M,\end{aligned}$$

where $M = c_0\sqrt{n} + n^{3/2}M_0M_1/2$. The latter can be rewritten in the form

$$\|\boldsymbol{\epsilon}^{i+1}\| \le h\frac{M}{nM_1}(e^{(i+1)nhM_1} - 1).$$

Hence, the maximum error due to discretisation according to the Euler's method allows the following upper estimate to be devised:

$$\|\boldsymbol{\epsilon}^{i}\| \le \epsilon = h\frac{M}{nM_1}(e^{nTM_1} - 1), \quad \forall i. \tag{25}$$

The same estimate holds also for the components of the vector $\boldsymbol{\epsilon}^i$ since $|\epsilon^i_j| \le \|\boldsymbol{\epsilon}^i\|$, $j = \overline{1,n}$.

Using the estimate (25) and the definition of local error (24) we can write

$$\left|y_j(t_i) - y^i_j\right| \le \epsilon, \quad j = \overline{1,n}.$$

Rearranging the latter we obtain the following two-sided inequality for the numerical solution

$$y_j(t_i) - \epsilon \le y^i_j \le y_j(t_i) + \epsilon,\ j = \overline{1,n}. \tag{26}$$

Since $y_j(t)$ is continuous on the closed interval $[0, T]$ it attains its minimum and maximum values on this interval according to the extreme-value theorem for continuous functions [11]. We denote those values as $y_{j,\min}$ and $y_{j,\max}$ respectively. We know from Theorem 1 that the exact solution $\boldsymbol{y}(t)$ to the Cauchy problem (11), (12) satisfies (17) for all $t > 0$. Therefore, $y_{j,\min} > 0$. Then there exists such step $h_{1,j}$ that for all $h < h_{1,j}$ the value of local error ϵ estimated by (25) will be less than $\epsilon_{1,j} = y_{j,\min}/2$ and the inequality

$$y^i_j \ge y_j(t) - \epsilon > \epsilon \tag{27}$$

will be satisfied. Analogously, there exists such step $h_{2,j}$ that for all $h < h_{2,j}$ the value of local error ϵ will be less than $\epsilon_{2,j} = (b - y_{j,\max})/2$ and the following inequality will hold:

$$y^i_j \le y_j(t) + \epsilon > b - \epsilon. \tag{28}$$

Hence, selecting the limit of local error as

$$\varepsilon_0 = \min_{1\le j\le n} \min(\epsilon_{1,j}, \epsilon_{2,j}) \tag{29}$$

and combining inequalities (25)–(28) we obtain that for $h = h_0 = \varepsilon_0/\bar{c}$ where $\bar{c} = \frac{M}{nM_1}\left(e^{nTM_1} - 1\right)$ the numerical solution to the Cauchy problem (11), (12) will satisfy the two-sided inequality (22). The inequality (22) will also be satisfied for any $h < h_0$. Thus defining $\varepsilon = h\bar{c} < h_0\bar{c} = \varepsilon_0$ we obtain the desired result. ☐

It follows from Theorem 2 that there always exists a limit of local error for which the numerical solution obtained by the Euler's method will lie in $\overline{G}$ and therefore remain strictly positive for all $t > 0$. This result remains valid for other numerical methods for ODEs which differ from the Euler's method by a higher order of approximation, i.e. when $\max\limits_{1 \le j \le n} |\eta_j^i| \le \tilde{c}h^p$ where $\tilde{c}$ is a constant independent of t and $\boldsymbol{y}$ and p is the order of a method.

We can now devise a criterion for the determination of the threshold of local error necessary for the solution to lie in the physical domain thus eliminating non-physical oscillations: *numerically computed concentrations in the Cauchy problem* (2), (3) *must remain non-negative*

$$y_j^i \ge 0, \quad j = \overline{1, n} \tag{30}$$

for any step number $i = 1, 2, \ldots$.

In practice the above criterion is applied as follows. The numerical solution starts with an initial limit of local error (say, $\varepsilon = 10^{-5}$). Then if condition (30) is violated during calculations the numerical solution process restarts with the limit of local error $\varepsilon := \varepsilon/10$. If necessary, this algorithm is repeated several times until the numerical solution satisfies (30) at all integration steps. Theorem 2 ensures that such a threshold of local error can be found for any Cauchy problem of the form (2), (3). However, in computational practice it is reasonable to restrict the decrease of the local error limit ε by the machine precision of a computer, which is used for calculations.

Turning back to the Belousov-Zhabotinsky reaction [9] and applying to it the above algorithm implemented in software package KinFitSim [2, 12–14] for kinetic simulation and fitting experimental data we find that the maximum limit (threshold) of local error which allows obtaining an adequate numerical solution is $\varepsilon = 10^{-8}$. Setting the Gear's method tolerance to a value less than or equal to 10^{-8} results in a numerical solution which is free of non-physical oscillations and follows the fully converged solution (see Fig. 1(b)). This exemplifies the fact that even if the required accuracy of the numerical solution is lower than ε_0 a physically meaningful solution cannot be obtained with the limit of local error ε set at values greater than ε_0.

4 Conclusions

The algorithms devised and justified in this work allow one to determine the necessary accuracy of the numerical solution of ODE systems, which represent mathematical models of homogeneous chemical and biochemical processes, and to assess their stiffness. The application of these algorithms ensures that an appropriate method is applied for the numerical solution and that the result lies within the permissible region.

The methods developed here have been implemented in the latest version of KinFitSim package (version 2.1) [13] in the form of an automatic numerical integration procedure. Thus the user's intervention into the solution process is eliminated such that both the solution method and its parameters are selected by the program based on the analysis of the mathematical model to be solved. Hence, the user is only required to enter a reaction mechanism and corresponding initial parameter values (initial concentrations and rate constants) prior to simulation. This automatic simulation procedure has been tested of numerous kinetic reaction mechanisms including both stiff and non-stiff in the course of numerical simulation.

References

1. G. Hall, J. M. Watt (Eds.), *Modern Numerical Methods for Ordinary Differential Equations*, Clarendon Press, Oxford, UK, 1976.

2. A. V. Klymenko, I. B. Svir, Modelling of kinetic mechanisms for photochemical analysis, *Management information. Systems and devices. All-Ukr. Sci. Interdep. Mag.,* **121**, pp. 30–34, 2002.

3. N. S. Bakhvalov, *Numerical Methods,* Nauka, Moscow, 1973.

4. C. W. Gear, *Numerical Initial Value Problem in Ordinary Differential Equations,* Prentice-Hall, Englewood Cliffs, NJ, 1971.

5. E. Hairer, S. P. Nørsett, G. Vanner, *Solution of Ordinary Differential Equations. Nonstiff Problems,* Mir, Moscow, 1990.
6. E. A. Volkov, *Numerical Methods,* Nauka, Moscow, 1982.
7. K. R. Jackson, The numerical solution of large systems of stiff IVPs for ODEs, *Appl. Numer. Math.,* **20**(1–2), pp. 5–20, 1996.
8. M. E. Hosea, L. E. Shampine, Analysis and implementation of TR-BDF2, *Appl. Numer. Math.,* **20**(1–2), pp. 21–37, 1996.
9. A. M. Zhabotinsky, Periodic processes of the oxidation of malonic acid in solution (Study of the kinetics of Belousov's reaction), *Biofizika,* **9**, pp. 306–311, 1964.
10. L. S. Pontryagin, *Ordinary differential equations,* Nauka, Moscow, 1982.
11. T. M. Apostol, *Calculus,* Vol. 1, 2nd ed., John Wiley & Sons, Inc., 1967.
12. I. B. Svir, A. V. Klymenko, M. S. Platz, 'KinFitSim' – a software package to fit kinetic data to a user selected mechanism, *Comput. Chem.,* **26**(4), pp. 379–386, 2002.
13. I. B. Svir, O. V. Klymenko, A. I. Oleinick, M. S. Platz, Kinfitsim (version 2.1) – a powerful tool for kinetic simulation of any reaction mechanism and fitting of any number of pairs of theoretical and experimental data sets, *Radioelektronika i Informatika,* **4**, pp. 22–25, 2004.
14. E. M. Tippmann, M. S. Platz, I. B. Svir, O. V. Klymenko, Evidence for specific solvation of two halocarbene amides, *J. Am. Chem. Soc.,* **126**(18), pp. 5750–5762, 2004.