ADVANCES IN MATHEMATICAL SCIENCES
Vol. 01 (2019) 16-26
www.knowvel.com

Original Research Article

PHASE SPACE ERROR CONTROL WITH VARIABLE TIME-STEPPING ALGORITHMS APPLIED TO THE FORWARD EULER METHOD FOR AUTONOMOUS DYNAMICAL SYSTEMS

R. Vigneswaran[a] | S. Thilaganathan[b]

ABSTRACT

We consider a phase space stability error control for numerical simulation of dynamical systems. Standard adaptive algorithm used to solve the linear systems perform well during the finite time of integration with fixed initial condition and performs poorly in three areas. To overcome the difficulties faced the Phase Space Error control criterion was introduced. A new error control was introduced by R. Vigneswaran and Tony Humbries which is generalization of the error control first proposed by some other researchers. For linear systems with a stable hyperbolic fixed point, this error control gives a numerical solution which is forced to converge to the fixed point. In earlier, it was analyzed only for forward Euler method applied to the linear system whose coefficient matrix has real negative eigenvalues. In this paper we analyze forward Euler method applied to the linear system whose coefficient matrix has complex eigenvalues with negative large real parts. Some theoretical results are obtained and numerical results are given.

Keywords: Adaptivity, fixed point, linear systems, variable step-size

AUTHOR AFFILIATION

[a] *Department of Mathematics and Statistics, Faculty of Science, University of Jaffna, Sri Lanka*

[b]*Department of Physical Science, Vavuniya Campus, University of Jaffna, Sri Lanka*

CORRESPONDENCE

R. Vigneswaran, Department of Mathematics and Statistics, University of Jaffna, Sri Lanka

Email: rvicky58@gmail.com

PUBLICATION HISTORY

Received: May 31, 2019

Accepted: June 12, 2019

ARTICLE ID: AMS-72

1. INTRODUCTION

Variable time stepping methods are often used to solve the dynamical systems defined by autonomous initial value ordinary differential equations:

$$y_t = f(y), \qquad y(0) = y^0 \in \Re^m, \tag{1.1}$$

where $f : \Re^m \to \Re^m$ is assumed to be Lipchitz continuous. It is globally accepted that efficient algorithm must be adaptive; that is, the step-size must be varied according to some error measures. In contrast to the fixed step-size case, a dynamical system oriented theory for variable step-size algorithm is far from complete. Only the standard adaptive algorithm performs well during finite time integration with fixed initial conditions and it is observed that typical adaptive algorithm fails or shows poor behaviour in three areas. Many researchers analysed and found those areas and contributed as research articles.

The first area is in spurious fixed points, was identified in [1]; it was shown that most of adaptive explicit Runge-Kutta methods admit stable spurious fixed points for arbitrary small tolerances. The second area is around fixed point. It was analysed and proved in Hall [3], that the standard adaptive algorithm fails to provide the correct dynamical system in this very simple and important scenario. Clear illustration was given in [4]. The third is identified near saddle points. The standard adaptive algorithm performs poorly near saddle points. It is clearly illustrated in [4] that a chaotic attractor it is often the unstable manifolds of the fixed point lead the flow on the attractor. The numerical solution will thus only be given good approximation to the flow on the attractor if it directs the unstable manifolds well. To do this it must produce a good approximation to the local unstable manifolds. Those three poor behaviours of standard algorithm will lead to the introduction of a new phase space error control.

Next, now we describe the standard error control which performs poorly near fixed points as mentioned above. In order to state precise results we focus on the embedded Explicit Runge-Kutta (ERK) pairs. Further details of these methods can be found for example in [2, 7].

Let t_n denote a sequence of (unequally spaced) grid points in time and let y_n denote an approximation to $y(t_n)$. Given y_n and a step-size $h_n := t_{n+1} - t_n$, the ERK pair is defined by

$$Y_i = y_n + h_n \sum_{j=1}^{i-1} a_{ij} f(Y_i), 1 \le i \le s, \tag{1.2}$$

$$y_{n+1} = y_n + h_n \sum_{i=1}^{s} b_i f(Y_i), \tag{1.3}$$

$$\tilde{y}_{n+1} = y_n + h_n \sum_{i=1}^{s} \tilde{b}_i f(Y_i). \tag{1.4}$$

Here $\{a_{ij}, b_i, \tilde{b}_i\}$ for $1 \le i \le s$ and $1 \le j \le i-1$ are the coefficients of the formula pair and $\tilde{y}_{n+1}$ is a subsidiary approximation that is used for error control. If $\tilde{y}_{n+1}$ is a lower-order approximation than y_n, then the pair is said to be operating in extrapolation mode.

In the typical local error control, the difference $y_n - \tilde{y}_{n+1}$ yields an estimate of the local error control which can be used to alter the step-size during integration. An estimate of the local error control is bounded at each time–step by a user-defined tolerance τ which allows the step-size to either increase or decrease over the next step. Let

$$E(y_n, h_n) = \frac{1}{h_n^{\rho}} (y_n - \tilde{y}_{n+1}) \tag{1.5}$$

be an approximation to the local truncation error over a step with $\rho = 0$ (Error-per-step (EPS)) or with $\rho = 1$ (Error per unit step (EPUS)). The error estimate $\| E(y_n, h_n) \|$ is used for two purposes, error control and step- size selection. For both cases (EPS &EPUS), the step size h_n is chosen at each step such that

$$\| E(y_n, h_n) \| \le \tau, \tag{1.6}$$

where $0 < \tau << 1$. In this case an approximation y_{n+1} is regarded as acceptable, otherwise the step-size is rejected and re-computed with a smaller step-size until the constraint (1.6) becomes true. The standard formula for next step is

$$h_{n+1} = \left(\frac{\gamma^{\tau}}{\| E(y_n, h_n) \|} \right)^{1/\tilde{q}} h_n, \tag{1.7}$$

where $\tilde{q}$ is the largest integer such that $\| E(y_n, h_n) \| = O(h_n^{\tilde{q}})$.

So, $\tilde{q} = \min(p, q) + 1 - \rho$. The constant safety factor $\gamma \in (0,1)$ is included to avoid rejecting too many steps.

2. Phase Space Error Controls

Higham et al. [4] proposed the Phase Space (PS) error control given by

$$\left\| y_{n+1} - y_n - \frac{1}{2} h_n (f(y_{n+1}) + f(y_n)) \right\| \le \frac{1}{2} \varphi h_n \| f(y_{n+1}) + f(y_n) \|, \tag{2.1}$$

where $\varphi \in (0,1)$ is a constant.

The new error control, the Phase Space θ, (PS_θ) error control was introduced in [5]. In this error control, the numerical solution $\{y_n\}_{n=0}^{\infty}$ satisfies the error constraint

$$\| y_{n+1} - y_n - h_n [(1-\theta) f(y_n) + \theta f(y_{n+1})] \| \le \varphi h_n \| (1-\theta) f(y_n) + \theta f(y_{n+1}) \|, \tag{2.2}$$

at every step, where $\varphi \in (0,1)$ is a user defined parameter akin to tolerance, and $\theta \in (0,1]$ is also a parameter to be chosen. This is a generalization of the PS error control introduced in [4], which is corresponded to (2.2) with $\theta = \frac{1}{2}$. It was seen in [5] that this error control automatically controls the step- size relative to the stability limit.

In [6], the behaviour of the forward Euler method under PS_θ error control (2.2) when applied to the linear system

$$y_t = Ay, \qquad y(0) = y^0 \in \Re^m, \tag{2.3}$$

with the real $m \times m$ matrix A was discussed when the eigenvalues of A are real and negative. When the forward Euler method is applied to the above system (2.3), the numerical solution $\{y^n\}$ evolves according to

$$y^{n+1} = R(h_n A) y^n, \tag{2.4}$$

Where $R(h_n A)$ is the stability polynomial matrix given by

$$R(h_n A) = I + h_n A. \tag{2.5}$$

With (2.4), the PS_θ constraint (2.2) becomes

$$\left\| R(h_n A) - 1 - h_n [(1-\theta)A + \theta A R(h_n A)] y^n \right\| \le \varphi h_n \left\| [(1-\theta)A + \theta A R(h_n A)] y^n \right\|, \tag{2.6}$$

for any vector norm ||.||.

Tony Humphries and Vigneswaran [6] established the following theorems and confirmed these by numerical experiments.

Theorem 2.1

Consider the forward Euler method under PS_θ error control (2.2) in $\|.\|_\infty$ with $\varphi < \frac{\theta}{1-\theta}$ applied to the system

$$y_t = \wedge y, \qquad \wedge = Diag\,[\lambda_1, \lambda_2, \ldots, \lambda_m], \qquad \lambda_i < 0, \forall i, \quad y(0) = y^0 \in \Re^m \tag{2.7}$$

with $\lambda_1 < \lambda_2 < \ldots < \lambda_{m-1} < \lambda_m < 0,$ and the initial conditions satisfy

$y(0) = [y_1^0, \cdots y_m^0] \in \Re^m$ with $y_m^0 \neq 0$. Then $\| y^n \|_\infty \to 0$ as $n \to \infty$ with the following:

1. $y_m^n \to 0$ Monotonically as $n \to \infty$;
2. If $\lambda_i \ge \frac{\theta(1+\varphi)}{\varphi} \lambda_m$ then $y_i^n \to 0$ and $\frac{y_i^n}{y_m^n} \to 0$ both monotonically as $n \to \infty$;
3. If $\frac{\theta(1+\varphi)}{\varphi} \lambda_m > \lambda_i \ge \left[\frac{2\theta(1+\varphi)}{\varphi} - 1 \right] \lambda_m$,

then $y_i^n \to 0$ and $\left| \frac{y_i^n}{y_m^n} \right| \to 0$ as $n \to \infty$;

4. For all remaining components of y^n we have $y_i^n \to 0$

as $n \to \infty$ with $\limsup_{n \to \infty} \left| \frac{y_i^n}{y_m^n} \right| < \frac{\varphi}{\theta - \frac{\varphi}{1+\varphi}}$;

5. Let θ_n be the angle between y^n and $[0,0,\ldots 0,1] \in \Re^m$. Then

$$\liminf_{n \to \infty} \cos\theta_n \ge \frac{1}{\sqrt{1 + (m-1)\left(\frac{\varphi}{\theta - \frac{\varphi}{1+\varphi}} \right)^2}} \ge 1 - \frac{1}{2}(m-1)\frac{\varphi^2}{\theta^2} + O(\varphi^3).$$

These results were extended to arbitrary norms and to non-diagonal linear systems in the following theorem.

Theorem 2.2
Consider the forward Euler method under PS_θ Error control (2.2) with sufficiently small φ applied to the linear system (2.3) where the matrix A is diagonalizable with negative real eigenvalues $\lambda_i, i = 1, 2, ..., m$ satisfying $\lambda_1 < \lambda_2 < \ldots < \lambda_{m-1} < \lambda_m < 0$. Then $\| y^n \| \to 0$ as $n \to \infty$.

The step- size selection strategies used in [4, 5] were not entirely satisfactory.

Tony Humphries and Vigneswaran [6] introduced a new step-size selection strategy based on the step-sizes derived from the standard error control and PS_θ error control respectively. It was also shown in [6] that the step-size tends to a constant value when PS_θ constraint applied to the initial value problem (1.1).

Now we focus on the forward Euler method under PS_θ error control (2.2) when applied to the linear system

$$y_t = Ay, \qquad y(0) = y^0 \in C^m, \tag{2.8}$$

with $m \times m$ matrix A having complex eigenvalues with negative real parts. We confirmed the results in [6] for matrix A having complex eigenvalues with negative real parts. The same step-size selection strategy is applied which was introduced in [6].

3. Linear Systems with Complex Eigenvalues

In [8], the results obtained for the linear systems with real and negative eigenvalues in [6] were confirmed by numerical experiments for the linear systems whose eigenvalues are complex with negative real parts.

The following lemma establishes algebraic inequalities which will be essential in the proof of our main result.

Lemma 3.1

Suppose $\theta \in (0,1]$, $\varphi \in (0,1), k > 0 > Re(\lambda_m)$ and

$$-\frac{1}{Re(\lambda_m)} < \frac{\theta(1+\varphi)}{\varphi}, \tag{3.1}$$

then

$$(1 + 2k\theta \,\mathrm{Re}(\lambda_m) + k^2\theta^2 |\lambda_m|^2)\varphi^2 < k^2\theta^2 |\lambda_m|^2 \tag{3.2}$$

And

$$\frac{\varphi}{\theta(1-\varphi^2)|\lambda_m|^2}[\varphi + Re(\lambda_m) + \sqrt{Re(\lambda_m)^2 + (1-\varphi^2)Im(\lambda_m)^2}] < k. \tag{3.3}$$

Proof:

Let $\lambda_m = \alpha_m + i\beta_m$. If $1 + 2k\theta\alpha_m \le 0$, then (3.2) holds trivially for any, $\varphi \in (0,1)$. Moreover in this case

$$-\frac{1}{Re(\lambda_m)} < 2\theta < \frac{\theta(1+\varphi)}{\varphi},$$

so (3.1) holds. Suppose now that $1 + 2k\,\theta\,\alpha_m > 0$. Then $-(1+1/k\theta\alpha_m)) > 1$ and hence (3.1) implies that

$$\varphi < -\frac{1}{1+\frac{1}{k\,\theta\alpha_m}} < 1 \quad \text{and}$$
$$\varphi^2 < \frac{k^2\,\theta^2{\alpha_m}^2}{1+2k\,\theta\,\alpha_m + k^2\,\theta^2{\alpha_m}^2}. \tag{3.4}$$

Also, $1 + 2k\,\theta\,\alpha_m > 0$ implies that

$$(1 + 2k\,\theta\,Re(\lambda_m) + k^2\,\theta^2|\lambda_m|^2){\alpha_m}^2 \le (1 + 2k\,\theta\,\alpha_m + k^2\,\theta^2{\alpha_m}^2)|\lambda_m|^2$$

and combining this with (3.4) implies (3.2). Thus we have established (3.2) in all cases. Now (3.2) implies that

$1 < -2k\,\theta\,\alpha_m + \frac{k^2\,(1-\varphi^2)\theta^2|\lambda_m|^2}{\varphi^2}$,
and multiplying by $(1-\varphi^2)|\,\lambda_m|$ and rearranging,

$$\alpha_m{}^2 + (1-\varphi^2)\beta_m{}^2 < \varphi^2\alpha_m{}^2 - 2k\theta\alpha_m(1-\varphi^2)\,|\,\lambda_m{}^2\,| + \frac{k^2(1-\varphi^2)\theta^2\,|\,\lambda_m\,|^4}{\varphi^2}.$$

Taking square roots and rearranging implies (3.3).

Theorem 3.2

Consider the forward Euler method under PS_θ error control (2.2) in $\|.\|_\infty$, where $\varphi \in (0,1)$ satisfies

$$\varphi < \frac{\theta}{1-\theta} \tag{3.5}$$

applied to the system

$$y_t = \wedge y, \quad \wedge = Diag\,[\lambda_1, \lambda_2, \ldots, \lambda_m],\ y(0) = y^0 \in C^m, \tag{3.6}$$

where $Re(\lambda_i) < 0$ and $|Im(\lambda_i)| < |Re(\lambda_i)|$ for all $i = 1,2,\ldots,m$ and the initial condition satisfies $y_m{}^0 \neq 0$.

Let H_i: $\frac{\varphi}{\theta(1-\varphi^2)|\lambda_i|^2}\,[\varphi + Re(\lambda_i) + \sqrt{(Re(\lambda_i))^2 + (1-\varphi^2)(Im(\lambda_i))^2}\,]$

and order the eigenvalues so that $H_m \geq H_{m-1} \geq \ldots \geq H_1 \geq 0$. Then $||y^n||_\infty \to 0$ *as* $n \to \infty$ with:

1. $\left|y_m{}^n\right| \to 0$ monotonically as $n \to \infty$;
2. If $\lambda_i \neq \lambda_m$ and $0 > \mathrm{Re}(\lambda_i) > \frac{\theta(1+\varphi)}{\varphi}\mathrm{Re}(\lambda_m)$, then $|\,y_i{}^n\,| \to 0$ and $\left|\frac{y_i{}^n}{y_m{}^n}\right| \to 0$ both monotonically as $n \to \infty$
3. If $0 > \frac{|\lambda_i|^2}{2Re(\lambda_i)} > \frac{\theta(1+\varphi)}{\varphi}Re(\lambda_m)$ then $\left|y_i{}^n\right| \to 0$ and $\left|\frac{y_i{}^n}{y_m{}^n}\right| \to 0$ both monotonically as $n \to \infty$;
4. For all remaining components of y^n , we have $|\,y_i{}^n\,| \to 0$ as $n \to \infty$.

Proof:

For the system (3.6) the stability polynomial matrix $R(h_n A) = I + h_n A$ is a diagonal matrix which is expressed as

$$R(h_n A) = Diag[1+h_n\lambda_1, 1+h_n\lambda_2, \ldots, 1+h_n\lambda_m]. \tag{3.7}$$

With the ∞-norm $\|.\|_\infty$, from (3.7) and (2.6) we have

$$\left\|\begin{bmatrix} -\theta h_n{}^2\lambda_1{}^2 y_1{}^n \\ -\theta h_n{}^2\lambda_2{}^2 y_2{}^n \\ . \\ . \\ . \\ \theta h_n{}^2\lambda_m{}^2 y_m{}^n \end{bmatrix}\right\|_\infty \leq \varphi h_n \left\|\begin{bmatrix} \lambda_1(1+\theta h_n\lambda_1) y_1{}^n \\ \lambda_2(1+\theta h_n\lambda_2) y_2{}^n \\ . \\ . \\ . \\ \lambda_m(1+\theta h_n\lambda_m) y_m{}^n \end{bmatrix}\right\|_\infty .$$

This implies $h_n\theta\,|\,\lambda_i{}^2 y_i{}^n\,| \leq \varphi\,|\,\lambda_i(1+\theta\lambda_i h_n)\,||\,y_i{}^n\,|$ (3.8)

for at least one $i \in \{1,2,\ldots,m\}$ and hence

$$h_n^2\theta^2 \mid \lambda_i^2 \mid \le \varphi^2 \mid 1+\theta\lambda_i h_n \mid^2 . \tag{3.9}$$

Let $\lambda_i = \alpha_i + i\beta_i$. Then (3.9) is equivalent to $h_n^2\theta^2(\alpha_i^2+\beta_i^2) \le \varphi^2[(1+\theta\alpha_i h_n)^2 + (h_n\theta\beta_i)^2]$. Rearranging as a quadratic in h_n we see that i^{th} condition holds if and only if.

$$h_n \le H_i := \frac{\varphi}{\theta(1-\varphi^2)(\alpha_i^2+\beta_i^2)}\left[\varphi\alpha_i + \sqrt{\alpha_i^2+(1-\varphi^2)\beta_i^2}\right] > 0 .$$

Since at each step $h_n \le H_i \le H_m$ for some $i \in \{1,2,...,m\}$, it follows that $h_n \le H_m$ for every step h_n satisfying PS_θ error constraint (2.2), and so (3.9) is satisfied with $i = m$ at every step.

Note that for the forward Euler method $R(h_n\lambda_i) = I + h_n\lambda_i$ and so $|R(h_n\lambda_i)|<1$ if and only if $h_n < -\frac{2\alpha_i}{(\alpha_i^2+\beta_i^2)}$ and $\mathrm{Re}(R(h_n\lambda_i)) > 0$ if and only if $h_n < -\frac{1}{\alpha_i}$. But the assumption $|\beta_i| \le |\alpha_i|$ implies that $-\frac{1}{\alpha_i} \le -\frac{2\alpha_i}{(\alpha_i^2+\beta_i^2)}$, hence if $h_n \le -\frac{1}{\alpha_i}$ then $|R(h_n\lambda_i)|<1$ and $\mathrm{Re}(R(h_n\lambda_i)) > 0$.

Now we establish (1)-(4)

1. The inequality (3.5) implies that $\frac{\theta(1+\varphi)}{\varphi} > 1$, thus we can apply Lemma (3.1) with $k = -\frac{1}{\alpha_m}$ and (3.3) implies that

$h_n \le H_m < -\frac{1}{\alpha_m} \le -\frac{2\alpha_m}{\alpha_m^2+\beta_m^2}$ for all $n \ge 0$ and the result follows.

2. Applying Lemma (3.1) with $k = -\frac{1}{\alpha_i}$ we conclude that $h_n \le H_m < -\frac{1}{\alpha_i}$ for all $n \ge 0$, and it follows that $\left|y_i^n\right| \to 0$ monotonically as $n \to \infty$. To show that $\left|\frac{y_i^n}{y_m^n}\right| \to 0$, consider

$$\frac{|\lambda_i|^2}{2\alpha_i} \ge \alpha_i \ge \frac{\theta(1+\varphi)}{\varphi}\alpha_m .$$

This implies

$$\frac{\alpha_m}{|\lambda_m|^2} \ge \frac{\theta(1+\varphi)}{\varphi}\frac{2\alpha_m^2}{|\lambda_m|^2}\frac{\alpha_i}{|\lambda_i|^2} > \frac{\alpha_i}{|\lambda_i|^2} \tag{3.10}$$

Since $\frac{\theta(1+\varphi)}{\varphi} > 1$ and $\frac{2\alpha_m^2}{|\lambda_m|^2} \ge 1$.

Thus

$$(\alpha_m-\alpha_i)|\lambda_i|^2 > -\alpha_i\left(|\lambda_i|^2 - |\lambda_m|^2\right). \tag{3.11}$$

Now there are two cases to consider.

(a). If $|\lambda_i|^2 > |\lambda_m|^2$ then (3.11) implies

$$\frac{2(\alpha_m-\alpha_i)}{\left(|\lambda_i|^2-|\lambda_m|^2\right)} > \frac{-2\alpha_i}{|\lambda_i|^2} \ge h_n .$$

This implies $$\frac{2(\alpha_m-\alpha_i)}{\left(|\lambda_i|^2-|\lambda_m|^2\right)} > h_n .$$

That is, $$2(\alpha_m-\alpha_i) > h_n\left(|\lambda_i|^2-|\lambda_m|^2\right).$$

This implies,

$2h_n\left(\alpha_m-\alpha_i\right)>h_n^2\left(\left|\lambda_i\right|^2-\left|\lambda_m\right|^2\right)$.

By rearranging the components and adding 1 on both sides, we get

$1+2h_n\alpha_i+h_n^2\left|\lambda_i\right|^2>1+2h_n\alpha_m+h_n^2\left|\lambda_m\right|^2$.

Thus

$|R(h_n\lambda_i)|^2=(1+h_n\alpha_i)^2+h_n^2\beta_i^2<(1+h_n\alpha_m)^2+h_n^2\beta_m^2=|R(h_n\lambda_m)|^2$

(b). If $\left|\lambda_i\right|^2<\left|\lambda_m\right|^2$ then (3.10) implies

$\frac{\alpha_m}{\left|\lambda_m\right|^2}>\frac{\alpha_i}{\left|\lambda_i\right|^2}\geq\frac{\alpha_i}{\left|\lambda_m\right|^2}$. This implies $\alpha_m>\alpha_i$.

Thus

$|R(h_n\lambda_i)|^2=1+2h_n\alpha_i+h_n^2|\lambda_i|^2<1+1+2h_n\alpha_m+h_n^2|\lambda_m|^2=|R(h_n\lambda_m)|^2.$

In both cases $\left|\frac{{y_i}^n}{{y_m}^n}\right|\to 0$ and hence the result follows.

3. Applying Lemma (3.1) with $k=-\frac{2\alpha_i}{|\lambda_i|^2}$, we conclude that $h_n\leq H_m<-\frac{2\alpha_i}{|\lambda_i|^2}$ for all $n\geq 0$ and it follows that $|{y_i}^n|\to 0$ monotonically as $n\to\infty$. The proof of

$\left|\frac{{y_i}^n}{{y_m}^n}\right|\to 0$ as $n\to\infty$ is the same as in part (2).

4. For all remaining components, $0>\frac{\theta(1+\varphi)}{\varphi}\mathrm{Re}(\lambda_m)>\frac{|\lambda_i|^2}{2\,\mathrm{Re}(\lambda_i)}>\mathrm{Re}(\lambda_i)$.

If the i^{th} index of the constraint (3.8) fails, then $H_i<h_n\leq H_j$. By the Lemma (3.1) with $m=j$ and $k=-\frac{1}{\alpha_i}$,

we obtain $H_i<h_n\leq H_j\leq-\frac{2\alpha_j}{{\alpha_j}^2+{\beta_j}^2}$.

For all remaining components, $|{y_i}^n|\to 0$ as $n\to\infty$ as in the proof of part (2).

Remark: These results are extended to arbitrary norms and to non-diagonal linear system in the following theorem.

Theorem (3.3)

Consider the system (3.6) where $\mathrm{Re}(\lambda_i)<0$ and there exists C ≥1 such that |Im (**λ**i)|² ≤C |Re (**λ**i)|² for all i=1,2,…,m and the initial conditions satisfies $y_m^0\neq 0.$ under the numerical approximation by the forward Euler method with PS_θ

Error control (2.2) in ||.||∞ where $\varphi\epsilon(0,1)$ satisfies

$\varphi<\frac{2\theta}{1+C-2\theta}$, (3.12)

Let $H_i:=\frac{\varphi}{\theta(1-\varphi^2)\left|\lambda_i\right|^2}\left[\varphi\,\mathrm{Re}(\lambda_i)+\sqrt{\mathrm{Re}(\lambda_i)^2+(1-\varphi^2)\,\mathrm{Im}(\lambda_i)^2}\right]$ and order the eigenvalues so that $H_m\geq H_{m-1}\geq\ldots\geq H_1>0.$

Then $\left\|y^n\right\|_\infty\to 0$ as $n\to\infty$ with:

1. $\left|y_m^{\,n}\right| \to 0$ monotonically as $n \to \infty$;
2. If $\lambda_i \neq \lambda_m$ and $0 > \mathrm{Re}(\lambda_i) > \frac{\theta(1+\varphi)}{\varphi}\mathrm{Re}(\lambda_m)$, then $|\, y_i^{\,n}| \to 0$ and $\left|\frac{y_i^{\,n}}{y_m^{\,n}}\right| \to 0$ both monotonically as $n \to \infty$
3. If $0 > \frac{|\lambda_i|^2}{2Re(\lambda_i)} > \frac{\theta(1+\varphi)}{\varphi} Re(\lambda_m)$ then $\left|y_i^{\,n}\right| \to 0$ and $\left|\frac{y_i^{\,n}}{y_m^{\,n}}\right| \to 0$ both monotonically as $n \to \infty$;
4. For all remaining components of y^n , we have $|\, y_i^{\,n}| \to 0$ as $n \to \infty$.

Proof:

Follow the theorem (3.3) to establish that $h_n \leq H_m$ for every step h_n satisfying PS_θ error constraint already defined. Now establish (1).

1. Equation (3.12) implies that

$\beta_m^2 \leq C\alpha_m^2 < \left\lfloor \frac{2\theta(1+\varphi)}{\varphi} - 1 \right\rfloor \alpha_m^2,$

Hence

$\frac{|\lambda_m|^2}{2\alpha_m^2} < \frac{\theta(1+\varphi)}{\varphi}$

Thus we can apply the lemma (3.1) with $k = \frac{2\alpha_m}{|\lambda_m|^2}$ to deduce that $h_n \leq H_m < -\frac{2\alpha_m}{|\lambda_m|^2}$ which implies that $|R(h_n, \lambda_m)| < 1$ for all $n \geq 0$, and the result follows.

2. Equation (3.12) implies that

$\beta_i^2 \leq C\alpha_i^2 < \left\lfloor \frac{2\theta(1+\varphi)}{\varphi} - 1 \right\rfloor \alpha_i^2,$

Hence

$\frac{|\lambda_i|^2}{2\alpha_i^2} < \frac{\theta(1+\varphi)}{\varphi}$

Thus we can apply the lemma (3.1) with $k = \frac{2\alpha_i}{|\lambda_i|^2}$ to deduce that $h_n \leq H_m < -\frac{2\alpha_i}{|\lambda_i|^2}$ which implies that $|R(h_n, \lambda_i)| < 1$ for all $n \geq 0$, and the result $|\, y_i^{\,n}| \to 0$ and the proof of $\left|\frac{y_i^{\,n}}{y_m^{\,n}}\right| \to 0$ is the same as in the part (2) of the theorem (3.2).

3. It is similar to the proof of the above part (2).
4. For all remaining components of y^n, if the i^{th} condition of (3.9) fails then j^{th} component of right hand side as follows $H_i < h_n \leq H_j$, That is

$\beta_j^2 \leq C\alpha_j^2 < \left\lfloor \frac{2\theta(1+\varphi)}{\varphi} - 1 \right\rfloor \alpha_j^2,$

$\Rightarrow \frac{|\lambda_j|^2}{2\alpha_j^2} < \frac{\theta(1+\varphi)}{\varphi}$

Hence the result follows as the above part(1).

Remark: These results are extended to arbitrary norms and to non-diagonal linear system in the theorem.

Theorem (3.4)

Consider the forward Euler method under PS_θ error control is defined with sufficiently small φ applied to the linear system

$y_t = Ay,$

where the matrix A is diagonalizable with complex eigenvalues λ_i, $i = 1, 2, \ldots, m$ satisfying $\mathrm{Re}(\lambda_i) < 0$ and $\left|\mathrm{Im}(\lambda_i)\right| < \left|\mathrm{Re}(\lambda_i)\right|$ for all $i = 1, 2, \ldots, m$ and the initial condition satisfies $y_m^0 \neq 0$.

Let

$H_i := \frac{\varphi}{\theta(1-\varphi^2)|\lambda_i|^2}\left[\varphi \mathrm{Re}(\lambda_i) + \sqrt{\mathrm{Re}(\lambda_i)^2 + (1-\varphi^2)\,\mathrm{Im}(\lambda_i)^2}\right]$ and order the eigenvalues so that $H_m \geq H_{m-1} \geq \ldots \geq H_1 > 0$. Then $\|y^n\|_\infty \to 0$ as $n \to \infty$.

Proof:

Since the matrix A is diagonalizable, there exists a non-singular matrix P such that $P^{-1}AP = D$ a diagonal matrix whose diagonal entries are $\lambda_1, \lambda_2, \ldots \lambda_m$. then the stability polynomial matrix $R(h_n, A) = I + h_n A$ satisfies

$$P^{-1}R(h_n, A)P = Diag[1 + h_n\lambda_1, 1 + h_n\lambda_2, \ldots, 1 + h_n\lambda_m] = \bar{R}(h_n D)(Say) \tag{3.13}$$

With $||R(h_nA) - I - h_n[(1-\theta)A + \theta AR(h_nA)]y^n|| \leq \varphi h_n||[(1-\theta)A + \theta AR(h_nA)]y^n||$

Becomes

$$||P\{\bar{R}(h_nD) - I - h_n[(1-\theta)D + \theta D\bar{R}(h_nD)]\}z^n|| \leq \varphi h_n||\{P[(1-\theta)D + \theta D\bar{R}(h_nD)]\}z^n||, \tag{3.14}$$

Where $z^n = P^{-1}y^n$. Now we define new norm $\|.\|_P$by

$$\|x\|_P = \|xP\|, \qquad \forall x \in \mathbb{R}^m \tag{3.15}$$

With this norm, the constraint (3.14) becomes

$$||\{\bar{R}(h_nD) - I - h_n[(1-\theta)D + \theta D\bar{R}(h_nD)]\}z^n||_P \leq \varphi h_n||\{[(1-\theta)D + \theta D\bar{R}(h_nD)]\}z^n||_P \tag{3.16}$$

Since the norms are equivalent on finite dimensional linear space, $\exists c_1, c_2 > 0$ such that

$$c_1\|x\|_\infty \leq \|x\|_P \leq c_2\|x\|_\infty, \quad \forall x \in \mathbb{R}^m \tag{3.17}$$

By combining (3.16) and (3.17), we obtain

$$||\{\bar{R}(h_nD) - I - h_n[(1-\theta)D + \theta D\bar{R}(h_nD)]\}z^n||_\infty \leq \varphi h_n||\{[(1-\theta)D + \theta D\bar{R}(h_nD)]\}z^n||_\infty \tag{3.18}$$

$$\left\|\begin{bmatrix} -\theta h_n^2\lambda_1^2 z_1^n \\ -\theta h_n^2\lambda_2^2 z_2^n \\ \vdots \\ -\theta h_n^2\lambda_m^2 z_m^n \end{bmatrix}\right\|_\infty \leq \varphi_1 h_n \left\|\begin{bmatrix} \lambda_1(1+\theta\lambda_1 h_n)z_1^n \\ \lambda_2(1+\theta\lambda_2 h_n)z_2^n \\ \vdots \\ \lambda_m(1+\theta\lambda_m h_n)z_m^n \end{bmatrix}\right\|_\infty$$

Where $\varphi_1 = \varphi\left(\frac{c_2}{c_1}\right)$ (< 1 for sufficiently smallφ). This implies that at least one of the following

$$h_n\theta\lambda_i^2|z_i^n| \leq -\varphi_1\lambda_i|1 + \theta\lambda_i h_n||z_i|^n, \quad i = 1,2,\ldots,m \tag{3.19}$$

Must hold. Since φ is sufficiently small, we can choose φ so that $\varphi < \frac{c_1}{c_2}\frac{\theta}{(1-\theta)}$. Hence by theorem (3.2), we obtain that $\|z^n\| \longrightarrow 0$ as $n \longrightarrow \infty$. This implies that $\|y^n\| \longrightarrow 0$ as $n \longrightarrow \infty$ for any norm $\|.\|$ since $y^n = Pz^n$ and P is non-singular.

4. Numerical Results

We consider the forward Euler (RK1(2))method applied to the system

$y_t = \begin{bmatrix} -3 & -1 \\ 1 & -3 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$. $y = [y_1, y_2]^T$ and

$y(0) = [0.9 \quad 10^{-4}]^T$. The above matrix has eigenvalues $-3 \pm i$. A typical standard algorithm as defined in introductory section with $\tau = 10^{-2}$ produces the dynamics is in Figure 1.

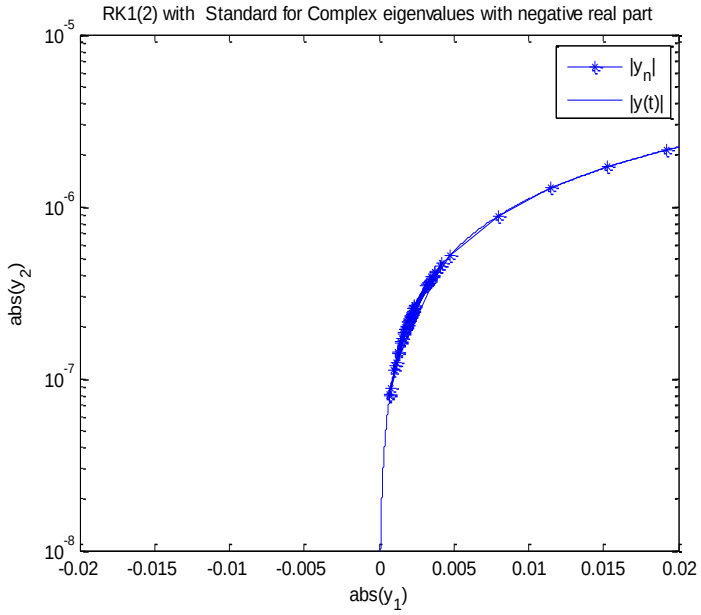

Figure 1. Numerical solutions of standard algorithm near stable fixed point for RK1(2) having complex eigenvalues.

In Figure 1 numerical solution is little deviated from the fixed point. It is not possible to converge the solution towards the fixed point. If we apply the RK1(2) method with combined PS_θ error control and standard error control and φ = 0.1, we obtain numerical solution in Figure 2, where we can see how numerical solution converges to fixed point.

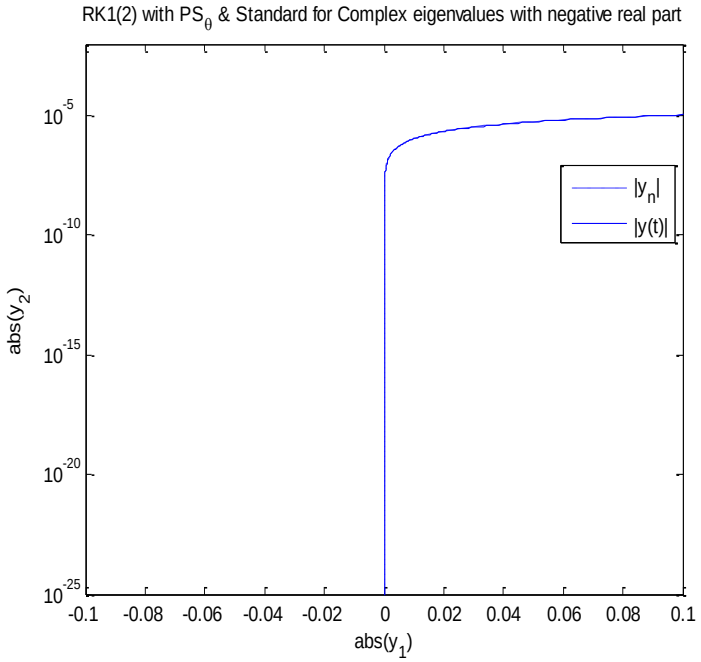

Figure 2. Numerical solution using RK1(2) combining with PS_θ augmented algorithm near a stable fixed point

Figure 3 shows the step-size sequences used by two algorithms. Some step-sizes rejected in standard algorithm whilst the PS_θ algorithm has no rejections and quickly converges to a constant value.

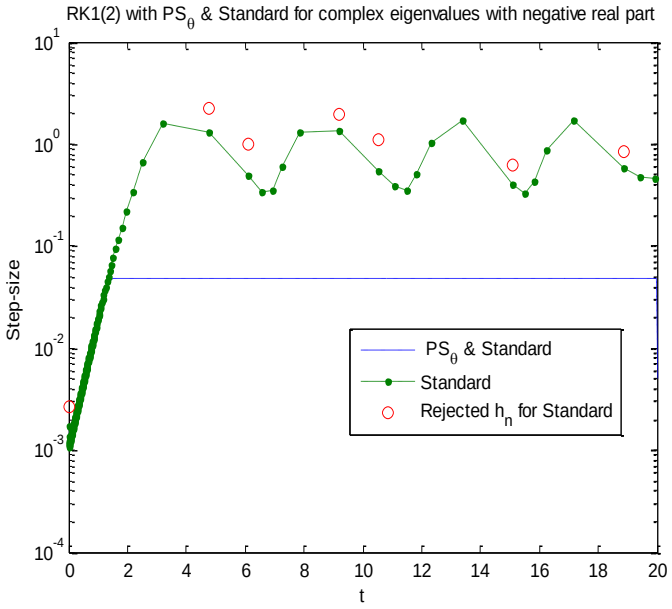

Figure 3. Step-sizes used by the standard and PS_θ augmented algorithms

5. Conclusion

In this research we analysed Phase space error control for forward Euler method applied to the linear systems whose coefficient matrix has complex eigenvalues with negative real parts and it is shown numerical solution forced to converge to the fixed point.

We will analyse the Phase space error control for s-stage general explicit Runge-Kutta methods applied to the linear systems whose coefficient matrix has real and negative eigenvalues in the forthcoming paper

Conflict of Interests

The authors declare that there is no conflict of interest related to the publication of this article.

References

[1] M. A. Aves, D. F. Griffiths, and D. J. Higham, "Does Error Control Suppress Spuriosity?," SIAM J. Numer. Anal., vol. 34, no. 2, pp. 756–778, 2003.

[2] E. Hairer, S. P. Nϕrsett, and G. Wanner, "Solving ordinary differential quations I–nonstiff problems", volume I, Springer– Verlag, Berlin Heidelberg, Second Edition, New York, 1993.

[3] G. Hall, "Equilibrium states of Runge Kutta schemes," ACM Trans. Math. Softw., vol. 11, no. 3, pp. 289–301, 1985.

[4] D. J. Higham, A. R. Humphries, and R. J. Wain, "Phase Space Error Control for Dynamical Systems," SIAM J. Sci. Comput., vol. 21, no. 6, pp. 2275–2294, 2003.

[5] A. R. Humphries, and N. Christodoulou, "Phase space error control for dynamical systems II", Research Report No: 2000-13:2275, University of Sussex, 2000.

[6] T. Humphries and R. Vigneswaran, "Phase Space Stability Error Control with Variable Time-stepping Runge-Kutta Methods for Dynamical Systems," Appl. Numer. Anal. Comput. Math., vol. 1, no. 2, pp. 469–488, 2004.

[7] L. F. Shampine, "Numerical solution of ordinary differential equations", Chapman and Hall, London, 1994.

[8] R. Vigneswaran, and S. Thilakanathan, "A combined error control with forward euler method for dynamical systems", International Journal of Mathematical and Computational Sciences, vol. 3, no. 5, 2016.