CONTRACTIVITY-PRESERVING EXPLICIT 2-STEP, 6-STAGE, 6-DERIVATIVE HERMITE-BIRKHOFF-OBRECHKOFF ODE SOLVER OF ORDER 13

By
Abdulrahman Alzahrani
July 2015

A Thesis
submitted to the Faculty of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements
for the degree of
Master in Science in Mathematics

Abstract

In this thesis, we construct a new optimal contractivity-preserving (CP) explicit, 2-step, 6-stage, 6-derivative, Hermite–Birkhoff–Obrechkoff method of order 13, denoted by HBO(13) with nonnegative coefficients, for solving nonstiff first-order initial value problems $y' = f(t, y)$, $y(t_0) = y_0$. This new method is the combination of a CP 2-step, 6-derivative, Hermite–Obrechkoff of order 9, denoted by HO(9), and a 6-stage Runge–Kutta method of order 5, denoted by RK(6,5). The new HBO(13) method has order 13. We compare this new method, programmed in MATLAB, to Adams–Bashforth–Moulton method of order 13 in PECE mode, denoted by ABM(13), by testing them on several frequently used test problems, and show that HBO(13) is more efficient with respect to the CPU time, the global error at the endpoint of integration and the relative energy error. We show that the new HBO(13) method has a larger scaled interval of absolute stability than ABM(13) in PECE mode. The Shu–Osher form of the new CP HBO(13) method is given in the Appendix B .

Key words: Contractivity preserving; explicit Hermite-Birkhoff-Obrechkoff method; Adams-Bashforth-Moulton method; comparing ODE solvers; N-body simulation.

Acknowledgements

First and foremost, praises and thanks to Allah (the God), the Almighty, for His showers of blessings throughout my research work to complete the research successfully.

I would like to express my deepest and sincere gratitude to my wonderful supervisors, **Rémi Vaillancourt** and **Thierry Giordano** for giving me the opportunity to do research and providing invaluable guidance throughout this research. Their encouragement, support, excellent advices, understanding, vision, sincerity and motivation have deeply inspired me. They have taught me the methodology to carry out the research and to present the research works as clearly as possible. It was a great privilege and honor to work and study under their guidance. I am extremely grateful for what they have offered to me. I am also eternally grateful to **Dr Truong Nguyen-Ba** for his help, continuous advices, valuable discussions and cooperation. I thank him so much for his time and great effort he has offered me during my M.Sc. studies. Also, I express my thanks to **Prof. Benoit Dionne**, Director of Graduate Programs at the Department of Mathematics and Statistics at the University of Ottawa, for giving me the useful information and valuable advices which I needed as a graduate student.

I would also like to express the profound gratitude from my deep heart to my beloved parents for their love, prayers, caring, continuous support and significant sacrifices for educating and preparing me for my future. I am very much grateful to my beloved wife for her love, patience, understanding, support and sharing my happiness and sadness. She was always there cheering me up and stood by me during my stay in Ottawa. Special thanks go to my beloved sons for the joy and happiness they have brought to my life. They gave me the hope and strength to keep going when I faced difficult times. I would also like to thank my brothers and sisters who were always supporting and encouraging me with their best wishes.

My very sincere thanks to my friends **Said Alqarni** and **Naif Alzahrani** for their precious friendship, valuable advices, encouragement and continuous support during my stay in Ottawa. I would also like to thank my colleague **Taqi Aldin Shatnawi** who was always willing to help and give his best suggestions.

Finally, I take this opportunity to extend my sincere appreciation to the government of Saudi Arabia, Saudi Arabia's Education Ministry and King Saud University for their financial support to complete this thesis successfully.

Dedication

I lovingly dedicate this thesis to:

My parents, **Musaad** *and* **Khadra**, *who have always loved me and whose good examples have taught me to work hard for the things that I aspire to achieve.*

My wife, **Samirah**, *for all of her love and support.*

My sons, **Saood** *and* **Musaad**, *who have been a source of motivation and strength during moments of despair and discouragement.*

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview of the new contractivity-preserving (CP) HBO(13) method

We cast a contractivity-preserving (CP) explicit 2-step Hermite–Obrechkoff method [21] and a 6-stage Runge–Kutta method of order 5 into an optimal CP explicit 2-step 6-stage Hermite–Birkhoff–Obrechkoff method of order 13 named HBO(13). The 2 steps consist in the current step and 1 backstep. The name of the method was chosen because it uses Hermite–Birkhoff interpolation polynomials and y' to $y^{(6)}$ at step points like Obrechkoff methods. The link between the two types of methods is that values at off-step points are obtained by means of predictors which use values at previous points.

Milne [18] was the first to have advocated the use of multiderivative, multistep Obrechkoff formulae for the numerical solution of differential equations. More recently, Huang and Innanen [11] introduced both a new form of the classical Adams–Cowell methods and new multiderivative, multistep methods. Some of them have larger stability interval and smaller truncation error than classical multistep methods.

HBO(13) is designed for solving nonstiff systems of first-order initial value problems of the form

$$y' = f(t, y), \qquad y(t_0) = y_0, \qquad \text{where} \quad ' = \frac{d}{dt} \quad \text{and} \quad y \in \mathbb{R}^n, \tag{1.1.1}$$

in the case where y' to $y^{(6)}$ can be calculated analytically or recursively, for instance, in dynamical systems [2, 3, 4, 10, 23, 29].

Similar to SSP Runge–Kutta methods (RK), which can be written as a convex combination of forward Euler method (FE) steps (see for example [6, 12, 27, 28]), CP HBO(13) methods can be written as a convex combination of steps of the special 6 derivatives extension of FE, which we denote by S(6):

$$y_{n+1} = y_n + \Delta t f(t_n, y_n) + \sum_{m=2}^{6} \eta_m (\Delta t)^m f^{(m-1)}(t_n, y_n) \tag{1.1.2}$$

where the coefficients η_m are smaller or equal to $\frac{1}{m!}$. If $\eta_m = \frac{1}{m!}$, for $m = 2, 3, \dots, 6$, then S(6) reduces to the usual Taylor method of order 6. The error in S(6) is of order $\ell \geq 2$ where ℓ is the smallest integer $2 \leq m \leq 6$ such that $\eta_m < \frac{1}{m!}$. If S(6) is contractive (in a given norm), then HBO(13) will be contractive (in the same norm) as a convex combination of S(6) with modified step sizes.

The region of absolute stability of HBO(13) is derived under the assumption that two solutions, y and $\tilde{y}$, of Problem (1.1.1) are *contractive*:

$$\|y(t + \Delta t) - \tilde{y}(t + \Delta t)\| \leq \|y(t) - \tilde{y}(t)\|, \quad \forall\, \Delta t \geq 0. \tag{1.1.3}$$

We assume that there exists a maximum stepsize $\Delta t_{\mathrm{S}(6)}$ such that f satisfies a discrete analog of (1.1.3) when S(6) is used with $\Delta t \leq \Delta t_{\mathrm{S}(6)}$:

$$\begin{aligned} \|y_{n+1} - \tilde{y}_{n+1}\| = \Bigg\| & y_n + \Delta t f(t_n, y_n) + \sum_{m=2}^{6} \eta_m (\Delta t)^m f^{(m-1)}(t_n, y_n) \\ & - \left(\tilde{y}_n + \Delta t f(t_n, \tilde{y}_n) + \sum_{m=2}^{6} \eta_m (\Delta t)^m f^{(m-1)}(t_n, \tilde{y}_n) \right) \Bigg\| \leq \|y_n - \tilde{y}_n\|, \end{aligned} \tag{1.1.4}$$

where y_n and $\tilde{y}_n$ are two numerical solutions generated by S(6) with different neighbouring initial values $y_0 = y(t_0)$ and $\tilde{y}_0 = \tilde{y}(t_0)$.

We are interested in a higher-order HBO(13) that maintains the *contractivity-preserving property*:

$$\|y_{n+1} - \tilde{y}_{n+1}\| \leq \max\{\|y_n - \tilde{y}_n\|, \|y_{n-1} - \tilde{y}_{n-1}\|\}, \tag{1.1.5}$$

for $0 \leq \Delta t \leq \Delta t_{\max} = c\Delta t_{\mathrm{S}(6)}$ whenever inequality (1.1.4) holds. Here c, called the CP coefficient, depends only on the numerical integration method but not on f. This definition of the CP coefficient of HBO(13) follows closely the definition of the SSP coefficient for RK [6].

In [14], similar types of methods have been constructed and tested on DETEST problems [13].

The objective of HBO(13) is to maintain the CP property (1.1.5) while achieving higher-order accuracy, perhaps with a modified time-step restriction, measured here with the CP coefficient $c(\mathrm{HBO}(13))$:

$$\Delta t \leq c(\mathrm{HBO}(13))\Delta t_{\mathrm{S}(6)}. \tag{1.1.6}$$

The CP coefficient describes the ratio of the maximal HBO(13) time step to the time step $\Delta t_{\mathrm{S}(6)}$, for which condition (1.1.5) holds.

1.2 Thesis objectives

The objectives of this thesis are:

- To construct a new explicit optimal CP Hermite–Birkhoff–Obrechkoff method of order 13 with nonnegative coefficients and with the largest CP coefficient obtained by `fmincon` in the MATLAB Optimization Toolbox.
- To fully derive the order conditions for the new HBO(13) method.
- To derive the canonical Shu-Osher form of HBO(13).
- To describe the region of absolute stability of HBO(13).
- To compare the performance on the basis of the CPU time and the global error at the endpoint of integration between HBO(13) and ABM(13), by solving the

nine nonstiff DETEST problems [13]. The Adams–Bashforth–Moulton solver of order 13, ABM(13), is a well known and widly used method to solve nonstiff problems.

- **B1** The growth of two conflicting populations.
- **B3** A nonlinear chemical reaction.
- **B5** Euler equations of motion for a rigid body without external forces.
- **D1** Kepler's two body problem with eccentricity $\epsilon = 0.1$.
- **D2** As in D1 except with eccentricity $\epsilon = 0.3$.
- **D3** As in D1 except with eccentricity $\epsilon = 0.5$.
- **D4** As in D1 except with eccentricity $\epsilon = 0.7$.
- **D5** As in D1 except with eccentricity $\epsilon = 0.9$.
- **E2** Derived from Van der Pol's equation with $\epsilon = 1$.

Also, we compare the performance on the basis of the CPU time and the global error at the endpoint of integration between HBO(13) and ABM(13), by solving the following two problems:

- Hénon–Heiles's problem [10].
- A problem in Galactic dynamics [3].

We present the eleven test problems considered in this thesis in the Appendix A. These problems are accepted in the numerical ODE community as a standard benchmark between high order ODE solvers and have been used by several authors (see for example [13, 14]).

- To compare the performance of HBO(13) and ABM(13) after a 350–orbit integration of Kepler's two body problem D2, D3, D4 and D5 on the basis of the CPU time and the relative energy error.

- To compare the growth of the relative energy error of HBO(13) and ABM(13) on 10000–orbit integration of Kepler's two body problem D3, D4 and D5.

1.3 Thesis organization

This thesis has five chapters. In Chapter 1, we define the new CP HBO(13) method, the goals and contributions of the thesis. The remainder of the thesis is organized as follows:

- In Chapter 2, we review some literature on first-order systems of ordinary differential equations, some numerical methods and the stability of some ODE's solvers.
- Chapter 3 is divided in three parts. In the first part (Sections 3.1, 3.2 and 3.3), we present the (CP) s-stage HBO methods, based on the combination of (CP) HO methods with RK(s,5) methods. We include the formulae and the order conditions of our new HBO(13) method. The second part (Sections 3.4, 3.5, 3.6 and 3.7) presents the transformation between the Butcher and Shu-Osher forms as well as the canonical Shu-Osher form and the formulation of the optimization problem to obtain the optimal HBO(13). The last part (Section 3.8) describes the region of absolute stability of HBO(13).
- The implementation of HBO(13) and the numerical results are found in Chapter 4. In Section 4.1, we introduce the test problems used for the comparisons. We compare the methods HBO(13) and ABM(13) using:
  - In Section 4.2, the global error at the endpoint of integration as a function of CPU time and CPU percentage efficiency gain.
  - In Section 4.3, the relative energy error, (EE(t)), as a function of CPU time.
  - In Section 4.4, the growth of relative energy error (EE(t)) as a function of time t.
- The conclusion of this thesis is presented in Chapter 5.

1.4 Thesis contributions

We believe that the results presented in this thesis are new and hope that they will contribute positively to the contractivity-preserving theory. Our contributions of this thesis include in particular:

- The construction of the new explicit optimal CP Hermite–Birkhoff–Obrechkoff method of order 13 with nonnegative coefficients and with the largest CP coefficient obtained by `fmincon` in the MATLAB Optimization Toolbox.
- The derivation of the order conditions for the new HBO(13) method.
- The implementation of HBO(13) in MATLAB, and the computation of its region of absolute stability.
- Our results show in particular that HBO(13) has a larger scaled interval of absolute stability than ABM(13) in PECE mode. We show also that the MATLAB version of HBO(13) requires fewer steps than ABM(13) on the problems introduced in Section 1.2 which often used to test higher order ODE solvers.
- Our results show that the new method HBO(13) has lower global error at the endpoint of integration and uses less CPU time than ABM(13) in solving the eleven problems considered in this thesis.
- Our results show that the new HBO(13) method performs better than ABM(13) in solving Kepler's two-body problem D2, D3, D4 and D5 after a 350–orbit integration.
- Our results show that the new HBO(13) method wins over ABM(13) on Kepler's two-body problem D3, D4 and D5 on the basis of the relative energy error as a function of 10000 periods integration time.

Chapter 2

Background Material

Most results presented in this chapter can be found in [16].

2.1 First-order systems of ordinary differential equations

In this thesis we study *first-order systems of ordinary differential equations* . Let I be an open interval of $\mathbb{R}$ and D be a domain of $\mathbb{R} \times \mathbb{R}^N$ (i.e., an open connected set). A first-order system of ordinary differential equation is written

$$y' = f(t, y(t)), \qquad \text{where} \qquad y' = \frac{dy}{dt}, \tag{2.1.1}$$

and $y : I \to \mathbb{R}^N$ with $(t, y(t)) \in D$ and f is a function from D to $\mathbb{R}^N$. Let $t_0 \in I$ and $(t_0, y_0) \in D$. To find a solution $y(t)$ of (2.1.1), satisfying $y(t_0) = y_0$ is known as an "initial value problem" that is written as

$$y' = f(t, y(t)), \qquad y(t_0) = y_0. \tag{2.1.2}$$

Before stating Theorem 2.1.1, let us recall the following definition

Definition 2.1.1. *Let D be a domain of $\mathbb{R} \times \mathbb{R}^N$. A function $f : D \to \mathbb{R}^N$ is said to satisfy a "Lipschitz condition in its second variable" if there exists a constant L, known as a Lipschitz constant, such that for every (t, y) and $(t, y^*) \in D$, then*

$$\| f(t, y) - f(t, y^*) \| \leq L \| y - y^* \| . \tag{2.1.3}$$

Theorem 2.1.1. *Let D be a domain of $\mathbb{R} \times \mathbb{R}^N$ and $f : D \to \mathbb{R}^N$ be a continuous function, satisfying a Lipschitz condition in its second variable, then there exists an open interval $I \subset \mathbb{R}$ and a unique solution $y : I \to \mathbb{R}^N$ with $(t, y(t)) \in D$ of the initial value problem* (2.1.2).

2.2 Solvers for nonstiff ordinary differential equations

2.2.1 Different types of methods

Several types of numerical methods to find solutions of ODE's have been developed. They can be divided in one-step and in multi-steps methods. One-step methods contain a number of stages (A s-stage method has $(s-1)$ predictors to compute f at intermediary points (called off-step points) and one integration formula to compute f at the end of the step). One-step methods use only the information from one previous point and off-step points to get the approximation solution at the next point. Euler and Runge-Kutta methods are examples of one-step methods. In contrast, multistep methods, for instance Adams-Bashforth methods and Adams-Moulton methods, use the information from several previous points to get the approximation solution at the next point.

2.2.2 Rung-Kutta methods

Runge-Kutta methods were primarily developed around 1900 by C. Runge and M. W. Kutta. Runge-Kutta methods use multiple evaluations of the function f at each step by extending Euler's method. They are one-step methods. They use only one initial point, denoted by (x_0, y_0), to compute the next point denoted by (x_1, y_1) or more generally, y_i is used to compute y_{i+1}. Usually, the step size h is variable. Using this method, we get the approximation y_{n+1} for $y(x_{n+1})$ from the previously given approximation y_n. The coefficients of this method are free parameters which satisfy a Taylor series expansion through some order in the time step h. The general s-stage

Runge-Kutta method is defined by

$$y_{n+1} = y_n + h \sum_{i=1}^{s} b_i f(x_n + c_i h, Y_i)$$
$$\text{where} \quad Y_1 = y_n \quad \text{and} \quad Y_i = y_n + h \sum_{j=1}^{s} a_{ij} f(x_n + c_j h, Y_j), \quad i = 2, 3, \ldots, s. \tag{2.2.1}$$

We always assume that the following condition holds:

$$c_i = \sum_{j=1}^{s} a_{ij}, \qquad i = 1, 2, ..., s. \tag{2.2.2}$$

It is convenient to display the coefficients occurring in (2.2.1) in the following form, known as a *Butcher array*:

$$\begin{array}{c|cccc} c_1 & a_{11} & a_{12} & \cdots & a_{1s} \\ c_2 & a_{21} & a_{22} & \cdots & a_{2s} \\ \vdots & \vdots & \vdots & & \vdots \\ c_s & a_{s1} & a_{s2} & \cdots & a_{s5} \\ \hline & b_1 & b_2 & \cdots & b_s \end{array} \tag{2.2.3}$$

We define the s-dimensional vectors c and b and the $(s \times s)$-matrix A by

$$c = [c_1, c_2, \cdots, c_s]^{\mathrm{T}}, \quad b = [b_1, b_2, \cdots, b_s]^{\mathrm{T}}, \quad A = [a_{ij}].$$

If in (2.2.1) we have that $a_{ij} = 0$ for $j \geq i$ and $1 \leq i \leq s$, then each of Y_i is given explicitly in terms of previously computed Y_j, for $1 \leq j \leq i-1$, and the corresponding method is called an *explicit* Runge–Kutta method. Since there is one evaluation of f per stage, there will be s evaluations of f per step.
The explicit s-stage Runge–Kutta methods also can be written as

$$y_{n+1} = y_n + h \sum_{i=1}^{s} b_i k_i, \tag{2.2.4}$$

$$\text{where} \quad k_i = f\left(x_n + c_i h, y_n + h \sum_{j=1}^{i-1} a_{ij} k_j\right) \quad \text{and} \quad c_i = \sum_{j=1}^{i-1} a_{ij}, \ \text{for } 1 \leq i \leq s.$$

The coefficients of this method are free parameters which satisfy a Taylor series expansion through some order in the time step h. To determine the coefficients of s-stage Runge–Kutta methods (2.2.4) in order that they have order p, we match the Taylor expansion of the solution generated by the Runge–Kutta method with the Taylor expansion of the exact solution. As an example, we consider the 3-stage Runge-Kutta method

$$y_{n+1} = y_n + h(b_1 k_1 + b_2 k_2 + b_3 k_3), \tag{2.2.5}$$

$$\begin{aligned} k_1 &= f(x_n, y_n), \\ k_2 &= f(x_n + c_2 h, y_n + h a_{21} k_1), \\ k_3 &= f(x_n + c_3 h, y_n + h a_{31} k_1 + h a_{32} k_2). \end{aligned}$$

We expand $y(x_{n+1})$ about x_n as a Taylor series, then we get

$$y(x_{n+1}) = y(x_n) + h y^{(1)}(x_n) + \frac{1}{2!} h^2 y^{(2)}(x_n) + \frac{1}{3!} h^3 y^{(3)}(x_n) + O(h^4), \tag{2.2.6}$$

where

$$\begin{aligned} y^{(1)}(x_n) &= f, \\ y^{(2)}(x_n) &= f_x + f f_y, \\ y^{(3)}(x_n) &= f_{xx} + 2 f f_{xy} + f^2 f_{yy} + f_y (f_x + f f_y), \end{aligned}$$

and $f = f(x, y)$, $f_x = \dfrac{\partial f(x,y)}{\partial x}$, $f_{xx} = \dfrac{\partial^2 f(x,y)}{\partial x^2}$, $f_{xy} = f_{yx} = \dfrac{\partial^2 f(x,y)}{\partial x \partial y}$.

To short the notation, we define

$$F = f_x + f f_y, \quad G = f_{xx} + 2 f f_{xy} + f^2 f_{yy}, \tag{2.2.7}$$

then (2.2.6) becomes

$$y(x_{n+1}) = y(x_n) + h f + \frac{1}{2!} h^2 F + \frac{1}{3!} h^3 (F f_y + G) + O(h^4). \tag{2.2.8}$$

We expand now the 3-stage Runge-Kutta method (2.2.5), as follows:

Firstly, we expand k_2, defined in (2.2.5), about the point (x_n, y_n) as a Taylor series, we obtain

$$k_2 = f + h c_2 (f_x + k_1 f_y) + \frac{1}{2!} h^2 c_2^2 (f_{xx} + 2 k_1 f_{xy} + k_1^2 f_{yy}) + O(h^3).$$

Using the notation in (2.2.7), then we write the expansion of k_2 as

$$k_2 = f + h\ c_2 F + \frac{1}{2!} h^2 c_2^2 G + O(h^3). \tag{2.2.9}$$

Secondly, we use a similar strategy for k_3, we get

$$k_3 = f + hc_3 F + h^2 (c_2 a_{32} F f_y + \frac{1}{2!} c_3^2 G) + O(h^3). \tag{2.2.10}$$

Finally, we substitute equations (2.2.9) and (2.2.10) into (2.2.5), we get

$$\begin{aligned} y_{n+1} = y(x_n) &+ h(b_1 + b_2 + b_3) f + h^2 (b_2 c_2 + b_3 c_3) F \\ &+ \frac{1}{2!} h^3 [2 b_3 c_2 a_{32} F f_y + (b_2 c_2^2 + b_3 c_3^2) G] + O(h^3). \end{aligned} \tag{2.2.11}$$

The coefficients must satisfy the following constraints in order to match the expansions (2.2.6) and (2.2.11):

$$\begin{aligned} b_1 + b_2 + b_3 &= 1, \\ b_2 c_2 + b_3 c_3 &= \frac{1}{2!}, \\ b_2 c_2^2 + b_3 c_3^2 &= \frac{1}{3}, \\ b_3 c_2 a_{32} &= \frac{1}{3!}, \end{aligned}$$

which are called order conditions for third-order explicit Rung-Kutta methods.

2.2.3 Adams–Bashforth–Moulton (ABM) methods

The methods of Adams are another improvement of Euler's method which were considered even earlier than Runge–Kutta methods. These were devised by John Couch Adams in order to solve a problem of F. Bashforth (see [9], pages 356-358). Adams methods are either explicit methods, known as *Adams–Bashforth methods* or implicit methods, known as *Adams–Moulton methods*. Adams–Bashforth methods have the following form

$$y_{n+1} = y_n + h(\beta_k f_n + \cdots + \beta_0 f_{n-k+1}), \tag{2.2.12}$$

where k is the number of steps, β_i, for $i = 0, 1, \ldots, k$, are constants and
$f_i = f(x_i, y_i)$, for $i = n-k+1,\ n-k+2, \ldots, n$.
The first examples of explicit Adams methods are

$$
\begin{aligned}
k = 1: &\quad y_{n+1} = y_n + hf_n. \qquad \text{(explicit Euler method)}\\
k = 2: &\quad y_{n+1} = y_n + h(\tfrac{3}{2}f_n - \tfrac{1}{2}f_{n-1}).\\
k = 3: &\quad y_{n+1} = y_n + h(\tfrac{23}{12}f_n - \tfrac{16}{12}f_{n-1} + \tfrac{5}{12}f_{n-2}).\\
k = 4: &\quad y_{n+1} = y_n + h(\tfrac{55}{24}f_n - \tfrac{59}{24}f_{n-1} + \tfrac{37}{24}f_{n-2} - \tfrac{9}{24}f_{n-3}).
\end{aligned}
$$

The Adams–Moulton methods have the following form

$$y_{n+1} = y_n + h(\beta_k f_{n+1} + \cdots + \beta_0 f_{n-k+1}). \tag{2.2.13}$$

The first examples of implicit Adams methods are

$$
\begin{aligned}
k = 0: &\quad y_{n+1} = y_n + hf_{n+1} = y_n + hf(x_{n+1}, y_{n+1}).\\
k = 1: &\quad y_{n+1} = y_n + h(\tfrac{1}{2}f_{n+1} + \tfrac{1}{2}f_n).\\
k = 2: &\quad y_{n+1} = y_n + h(\tfrac{5}{12}f_{n+1} + \tfrac{8}{12}f_n - \tfrac{1}{12}f_{n-1}).\\
k = 3: &\quad y_{n+1} = y_n + h(\tfrac{9}{24}f_{n+1} + \tfrac{19}{24}f_n - \tfrac{5}{24}f_{n-1} + \tfrac{1}{24}f_{n-2}).
\end{aligned}
$$

As shown in [9], the methods (2.2.13) give in general more accurate approximations to the exact solution than (2.2.12). Adams methods of order 1 to 13 or 14 are frequently used; beyond order 14, they lack stability. The region of absolute stability shrinks as the order increases. The Adams methods are usually implemented in a "predictor–corrector" form. That is, a preliminary calculation is carried out using the Bashforth form of the method. The approximate solution at a new step value is then used to evaluate an approximation of the derivative value at the new point. This derivative approximation is then used in the Moulton formula in place of the derivative at the new point. There are many alternatives for the next step, and we will describe just one of them. Let y_n^* denote the approximation to $y(x_n)$ found during the Bashforth part of the step calculation and y_n the improved approximation found in the Moulton part of the step. Temporarily denote by β_i^* the value of β_i in the Bashforth formula so that β_i will denote only the Moulton coefficient. The value of k corresponding to the Bashforth formula will be denoted here by k^*. Usually k and k^* are related by $k^* = k + 1$ so that both formulae have the same order $p = k + 1$.
In the Bashforth stage of the calculation we compute

$$y_n^* = y_{n-1} + h\sum_{i=1}^{k^*} \beta_i^* f(x_{n-i}, y_{n-i}), \tag{2.2.14}$$

and in the Moulton stage we compute

$$y_n = y_{n-1} + h\beta_0 f(x_n, y_n^*) + h\sum_{i=1}^{k} \beta_i f(x_{n-i}, y_{n-i}). \tag{2.2.15}$$

Methods of this type are referred to as "predictor–corrector" methods because the overall computation in a step consists of a preliminary prediction of the answer followed by a correction of this first predicted value. The use of (2.2.14) and (2.2.15) requires two calculations of the function f in each step of the computation. Such a scheme is referred to as being in "predict–evaluate–correct–evaluate" or "PECE" mode.

2.2.4 Contractivity preserving methods and the growth of error

When solving numerically an initial value problem (2.1.2), it is desirable to ensure that errors in the initial conditions or numerical errors at a given step do not grow excessively as they are propagated in subsequent steps. A sufficient condition for a k-step numerical method to have this property is the notion of contractivity preserving.

Definition 2.2.1. *A k-step numerical method is contractivity preserving if*

$$\|y_{n+1} - \widetilde{y}_{n+1}\| \leq \max\left\{\|y_n - \widetilde{y}_n\|, \|y_{n-1} - \widetilde{y}_{n-1}\|, \ldots, \|y_{n-k+1} - \widetilde{y}_{n-k+1}\|\right\}, \tag{2.2.16}$$

where for $n-k+1 \leq j \leq n+1$, y_j and $\widetilde{y}_j$ denote the approximate solutions at time t_j of the initial value problem (2.1.2) and y_{j+1} and $\widetilde{y}_{j+1}$ denote the corresponding numerical solutions at the next time step t_{j+1} of the initial value problem (2.1.2).

By interpreting $\widetilde{y}_n$ as a perturbation of y_n due to numerical errors, we see that contractivity implies that these errors do not grow as they are propagated [15].

2.2.5 Contractivity-preserving explicit Hermite-Obrechkoff ODE solver

Since HBO(13) will be based on the combination of contractivity-preserving (CP) Hermite–Obrechkoff of order 9, HO(9), and RK(6,5), we present briefly CP, explicit, 2-step, HO ODE solvers (see [21] for details). To construct HO methods, we replace the forward Euler (FE) method,

$$y_{n+1} = y_n + \Delta t f(t_n, y_n), \tag{2.2.17}$$

used by Gottlieb et al [6] and Huang [12] in establishing strong stability preserving (SSP) Runge–Kutta (RK) methods as convex combinations of FE methods. We rewrite HO as a convex combination of the special d derivative extension of FE, which we denote by S(d):

$$y_{n+1} = y_n + \Delta t f(t_n, y_n) + \sum_{m=2}^{d} \eta_m (\Delta t)^m f^{(m-1)}(t_n, y_n), \tag{2.2.18}$$

where the coefficients η_m satisfy the inequality $\eta_m \leq \frac{1}{m!}$. If the equality holds, then S(d) reduces to the Taylor method of order d. The error in S(6) is of order $\ell \geq 2$ where ℓ is the smallest integer $2 \leq m \leq d$ such that $\eta_m < \frac{1}{m!}$. If S(d) is contractive in a given norm, then HO will be contractive as a convex combination of S(d) with modified step sizes. HO methods require one formula to perform the integration step from t_n to t_{n+1}. To construct an HO method, we use a Hermite interpolation polynomial as a 2–step integration formula with d derivative of y to obtain y_{n+1} up to order $p \geq d$

$$y_{n+1} = \sum_{\ell=0}^{1} \left[\gamma_{\ell,0}\ y_{n-\ell} + \sum_{m=1}^{d} (\Delta t)^m\ \gamma_{\ell,m}\ y_{n-\ell}^{(m)} \right], \tag{2.2.19}$$

with step size Δt and consistency condition: $\gamma_{0,0} + \gamma_{1,0} = 1$. The 2–step HO methods seek to achieve a higher order HO that maintains the contractivity-preserving property:

$$\|y_{n+1} - \widetilde{y}_{n+1}\| \leq \max\left\{\|y_n - \widetilde{y}_n\|, \|y_{n-1} - \widetilde{y}_{n-1}\|\right\}. \tag{2.2.20}$$

The aim of CP HO, in general, is to maintain the CP property (2.2.20) while achieving a higher order accuracy than the special d derivative extension of FE, S(d) (2.2.18). An example of this kind of methods is HO(13) [21] which usually requires significantly fewer function evaluations and significantly less CPU time than the Taylor method of order 13 and the Runge-Kutta method DP(8,7)13M to achieve the same global error when solving standard N-body problems.

2.3 Contractivity-Preserving (CP) methods and introduction of Shu-Osher form

2.3.1 Contractivity-preserving Hermite-Obrechkoff of order 8, HO(8)

By using the same idea as in Section 2.2.4, we can construct HO(8). We use a Hermite interpolation polynomial as a 2–step integration formula with 6 derivatives of y to obtain y_{n+1} to order 8,

$$y_{n+1} = \sum_{\ell=0}^{1} \left[\gamma_{\ell,0}\; y_{n-\ell} + \sum_{m=1}^{6} (\Delta t)^m \; \gamma_{\ell,m}\; y_{n-\ell}^{(m)} \right], \tag{2.3.1}$$

with the consistency condition

$$\gamma_{0,0} + \gamma_{1,0} = 1. \tag{2.3.2}$$

We also have to satisfy the order conditions

$$\sum_{m=0}^{j} \left[\gamma_{0,m} \frac{(0)^{j-m}}{(j-m)!} + \gamma_{1,m} \frac{(-1)^{j-m}}{(j-m)!} \right] = \frac{1}{j!}, \qquad j = 1, 2, \ldots, 6, \tag{2.3.3}$$

$$\sum_{m=0}^{6} \left[\gamma_{0,m} \frac{(0)^{j-m}}{(j-m)!} + \gamma_{1,m} \frac{(-1)^{j-m}}{(j-m)!} \right] = \frac{1}{j!}, \qquad j = 7, 8, \tag{2.3.4}$$

where $0^0 = 1$ by convention.

2.3.2 Shu-Osher form of Runge-Kutta methods

Explicit strong stability preserving (SSP) or contractivity-preserving (CP) Runge–Kutta methods have been introduced by Shu and Osher in [7] to guarantee that spatial discretizations that are total variation diminishing (TVD) and total variation bounded (TVB) when coupled with forward Euler will still generate TVD and TVB solutions when coupled with these high order Runge-Kutta methods. The key observation in the development of (CP) methods was that specific Runge-Kutta methods can be

written as a convex combination of Euler's methods, so that any convex functional properties of Euler's method will carry over to these Runge-Kutta methods.
An explicit Runge-Kutta method is frequently written in *Butcher* form

$$\begin{aligned} Y_i &= y_n + \Delta t \sum_{j=1}^{s} a_{ij} F_j, \qquad 2 \leq i \leq s, \\ y_{n+1} &= y_n + \Delta t \sum_{j=1}^{s} b_j F_j, \end{aligned} \tag{2.3.5}$$

where $\quad Y_1 = y_n, \quad F_1 = f_n \quad$ and $\quad F_j := f(t_n + c_j \Delta t, Y_j), \quad j = 2, 3, \ldots, s.$
It can also be written in *Shu-Osher* form as follows

$$\begin{aligned} Y_i &= \sum_{j=1}^{i-1} \alpha_{ij} Y_j + \Delta t \beta_{ij} F_j, \qquad i = 2, 3, \ldots, s+1, \\ y_{n+1} &= Y_{s+1}, \end{aligned} \tag{2.3.6}$$

with consistency conditions $\sum_{j=1}^{i-1} \alpha_{ij} = 1$. Shu-Osher form is convenient because, if all the coefficents α_{ij} and β_{ij} are non-negative, each stage of RK method in Shu-Osher form can be rewritten as convex combinations of forward Euler steps, with a modified time step. The 6-stages Rung-Kutta method of order 5 in *Shu-Osher* form is written as follows

$$\begin{aligned} Y_i &= \sum_{j=1}^{i-1} \alpha_{ij} Y_j + \Delta t \beta_{ij} F_j, \qquad i = 2, 3, \ldots, 7, \\ y_{n+1} &= Y_7. \end{aligned} \tag{2.3.7}$$

It has been noticed in [24] that there is no s-stage CP Runge-Kutta method with $\alpha_{ij} \geq 0,\ \beta_{ij} \geq 0,$ for $j = 0, 1, \ldots, i-1$ and $i = 2, 3, \ldots, s+1$ with order $p > 4$.

2.4 Stability of the ODE solvers

We present briefly linear k–step methods before introducing the stability of ODE solvers. As in [16], linear k–step methods are of the form

$$\sum_{j=0}^{k} \alpha_j y_{n+j} = \Delta t \sum_{j=0}^{k} \beta_j f_{n+j}, \tag{2.4.1}$$

where α_j and β_j are constants subject to the conditions $\alpha_k = 1$ and $|\alpha_0| + |\beta_0| \neq 0$.

The difference equation (2.4.3) associated with a numerical method for ODEs is obtained by applying the linear k-step method (2.4.1) to the test equation

$$y' = \lambda y \qquad \text{where} \quad \lambda \in \mathbb{R}. \tag{2.4.2}$$

By (2.4.1) and (2.4.2), the difference equation takes the form

$$\sum_{j=0}^{k} (\alpha_j - \Delta t \beta_j \lambda) y_{n+j} = 0, \tag{2.4.3}$$

and the characteristic equation is

$$\sum_{j=0}^{k} (\alpha_j - \widehat{h}\beta_j) r^j = 0, \tag{2.4.4}$$

with $\widehat{h} = \lambda \Delta t$. The stability of the linear k-step method (2.4.1) requires that the roots of the characteristic equation (2.4.4) must satisfy the root condition defined as follows

Definition 2.4.1. *[16] The method* (2.4.1) *is said to satisfy* ***the root condition*** *if all of the roots of the characteristic equation* (2.4.4) *have modulus less or equal to one, and those of modulus one are simple.*

We then have

Definition 2.4.2. *[16] The linear k-step method* (2.4.1) *is said to be* ***absolutely stable*** *for a given $\widehat{h}$ if for that $\widehat{h}$ all the roots of the characteristic equation* (2.4.4) *satisfy the root condition and to be* ***absolutely unstable*** *for that $\widehat{h}$ otherwise .*

Definition 2.4.3. [*16*] *The linear k-step method* (2.4.1) *is said to have* ***region of absolute stability*** *$\mathcal{R}_{\mathcal{A}}$ where $\mathcal{R}_{\mathcal{A}}$ is a region of the complex $\widehat{h}$-plane, if it is absolutely stable for all $\widehat{h} \in \mathcal{R}_{\mathcal{A}}$. The intersection of $\mathcal{R}_{\mathcal{A}}$ with the real axis is called the* ***interval of absolute stability****.*

For a fair comparison of the performance of different ODE solvers, the interval of absolute stability (defined in Definition 2.4.3) is divided by the number of evaluations of f per step, thus giving the scaled interval of absolute stability.

2.5 Regions of absolute stability of RK(6,5), HO(8) and ABM(13)

2.5.1 Region of absolute stability of RK(6,5)

Recall (as in Subsection 2.2.2) that the general *explicit* 6-stage Runge-Kutta method of order 5, RK(6,5), is defined by

$$y_{n+1} = y_n + h\sum_{i=1}^{6} b_i f(x_n + c_i h, Y_i), \tag{2.5.1}$$

where $Y_1 = y_n$ and $Y_i = y_n + h\sum_{j=1}^{5} a_{ij} f(x_n + c_j h, Y_j)$, for $2 \le i \le 6$.

To obtain the (unscaled) region of absolute stability, $\mathcal{R}$, of RK(6,5), we apply RK(6,5) formula (2.5.1) with constant step $h = \Delta t$ to the test equation (2.4.2), that is $f(t, y) = \lambda y$. Thus, we obtain, with $\widehat{h} = \lambda\Delta t$;

$$\begin{aligned} y_{n+1} &= y_n + \widehat{h}\ \sum_{i=1}^{6} b_i Y_i, \\ \text{where}\quad Y_i &= y_n + \widehat{h}\ \sum_{j=1}^{5} a_{ij} Y_j, \qquad i = 2, 3, \ldots, 6, \end{aligned} \tag{2.5.2}$$

Since RK(6,5) is an explicit numerical method, the (6×6)-matrix $A = [a_{ij}]$ is strictly lower triangular. For convenience, we define Y, b and $e \in \mathbb{R}^6$ by $Y := [Y_1, Y_2, \ldots, Y_6]^T$, $b := [b_1, b_2, \ldots, b_6]^T$ and $e := [1, 1, \ldots, 1]^T$; then (2.5.2) can be written:

$$Y = y_n e + \widehat{h}AY, \qquad y_{n+1} = y_n + \widehat{h}b^T Y. \tag{2.5.3}$$

Using the first equation of (2.5.3), we get $(I - \widehat{h}A)Y = y_n e$. As A is nilpotent, $(I - \widehat{h}A)$ is invertible, we have

$$Y = (I - \widehat{h}A)^{-1} y_n e. \tag{2.5.4}$$

Using the second equation of (2.5.3) and (2.5.4), we get

$$y_{n+1} = [1 + \widehat{h}\ b^T (I - \widehat{h}A)^{-1} e]\ y_n, \tag{2.5.5}$$

where I is the (6×6) identity matrix.

As in Section 2.4, using the equation (2.5.5), we get the difference equation

$$C_0\ y_n + C_1\ y_{n+1} = 0, \tag{2.5.6}$$

and the associated linear characteristic equation:

$$C_0 + C_1\ r = 0, \tag{2.5.7}$$

where $C_1 = 1$, $C_0 = 1 + \widehat{h}\ b^T(I - \widehat{h}A)^{-1}e$. A complex number $\widehat{h}$ is in $\mathcal{R}$ if the root of the characteristic equation (2.5.7) satisfies the root condition which was defined in Defintion 2.4.1. This root depends on the order conditions of RK(6,5) which can be obtained as in Subsection 2.2.2 . RK(6,5) has 17 conditions (equations) in 21 parameters (unknowns) to be satisfied, so there exists an infinite family of explicit RK(6,5). We will consider one member of the family of explicit RK(6,5) to give an idea about the region of absolute stability of RK(6,5). DOPRI(5,4) [16, page 204] is one member of the family of explicit RK(6,5) which has the root $r = 1 + \widehat{h} + \widehat{h}^2/2 + \widehat{h}^3/6 + \widehat{h}^4/24 + \widehat{h}^5/120 + \widehat{h}^6/600$. By applying the scanning technique (see [16, pp. 70 and 204]) we can find the region of absolute stability of DOPRI(5,4) which is indicated in Fig. 1. The grey regions in Fig. 1 are symmetric about the real axis, and Fig. 1 shows only the grey regions in the half-plane $\mathrm{Im}(\widehat{h}) > 0$.

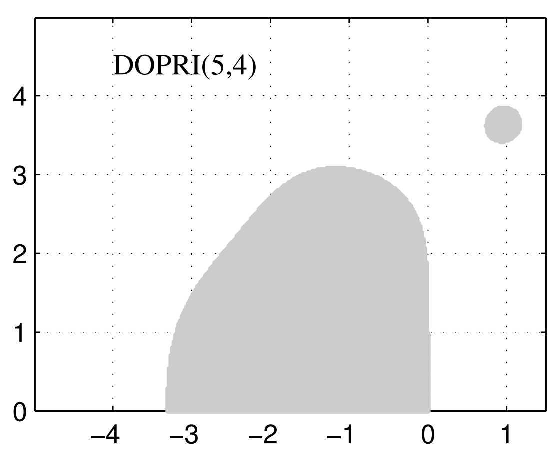


Figure 1: The grey regions depict the unscaled region of absolute stability of DOPRI(5,4) with the unscaled stability interval $(-3.3, 0)$.

2.5.2 Region of absolute stability of HO(8)

The integration formula of HO(8) is given by

$$y_{n+1} = \sum_{\ell=0}^{1} \left[\gamma_{\ell,0}\ y_{n-\ell} + \sum_{m=1}^{6} (\Delta t)^m\ \gamma_{\ell,m}\ y_{n-\ell}^{(m)} \right]. \tag{2.5.8}$$

To obtain the (unscaled) region of absolute stability, $\mathcal{R}$, of HO(8), we apply the integration formula (2.5.8) with constant step Δt to the test equation (2.4.2) as above Thus, we obtain

$$y_{n+1} = \sum_{\ell=0}^{1} \left[\gamma_{\ell,0}\ y_{n-\ell} + \sum_{m=1}^{6} (\lambda \Delta t)^m\ \gamma_{\ell,m}\ y_{n-\ell} \right]. \tag{2.5.9}$$

If we let $\widehat{h} = \lambda \Delta t$ and rewrite (2.5.8) as a function of y_{n+1}, y_n and y_{n-1} only, we then get

$$y_{n+1} = \left[\gamma_{0,0} + \sum_{m=1}^{6} (\widehat{h})^m\ \gamma_{0,m} \right] y_n + \left[\gamma_{1,0} + \sum_{m=1}^{6} (\widehat{h})^m\ \gamma_{1,m} \right] y_{n-1}. \tag{2.5.10}$$

As in Subsection 2.5.1, we get the difference equation

$$\sum_{j=0}^{2} C_j y_{n+(j-1)} = 0, \tag{2.5.11}$$

and the associated linear characteristic equation:

$$\sum_{j=0}^{2} C_j r^j = 0, \tag{2.5.12}$$

where $C_2 = 1$, $C_1 = -\left[\gamma_{0,0} + \sum_{m=1}^{6} (\widehat{h})^m\ \gamma_{0,m} \right]$ and $C_0 = -\left[\gamma_{1,0} + \sum_{m=1}^{6} (\widehat{h})^m\ \gamma_{1,m} \right]$.

A complex number $\widehat{h}$ is in $\mathcal{R}$ if the two roots of the characteristic equation (2.5.12) satisfy the root condition defined in Defintion 2.4.1. The region of absolute stability of HO(8) is indicated in Fig. 2 . As in Subsection 2.5.1, the grey region in Fig. 2 is symmetric about the real axis, and Fig. 2 shows only the grey region in the half-plane Im($\widehat{h}$)> 0.

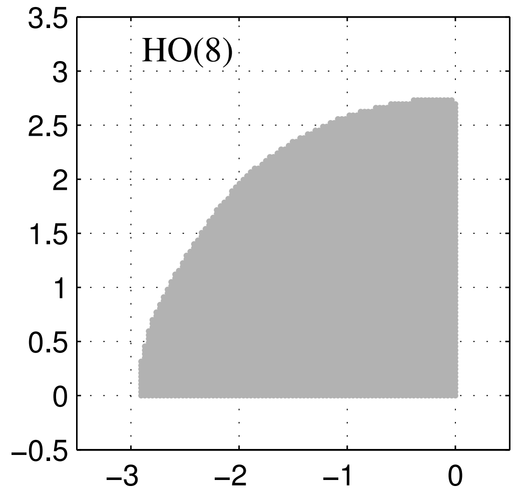


Figure 2: The grey region depicts the unscaled region of absolute stability of HO(8) with the unscaled stability interval $(-2.905, 0)$.

2.5.3 Region of absolute stability of ABM(13)

L. F. Shampine, M. K. Gordon, have plotted stability regions for several variants of the Adams methods for orders $k = 1, 2, \dots, 12$ [25, pp. 135–140]. They plotted the stability region of PECE method with Adams–Bashforth (predictor) formula of order 12 and Adams–Moulton (corrector) formula of order 13. This PECE method is actually ABM(13). It is found that the (unscaled) stability interval of ABM(13) is $(-0.062, 0)$.

Chapter 3

Contractivity-preserving HBO Methods

In 2002, Ruuth and Spiteri [24] showed that an s-stage CP RK method with nonnegative coefficients and order $p > 4$ does not exist. In this thesis, we construct an optimal CP HBO(13) method with nonnegative coefficients, which has order conditions analogous to the order conditions of RK(s,5).

3.1 CP s-stage HBO methods based on combining CP HO methods with RK(s,5) methods

3.1.1 General HBO formulation and notation

The following notation will be used in this chapter:

Notation 3.1.1.

We will denote by

- k *the number of steps of a given method,*
- s *the number of stages of a given method per time step,*
- d *the number of derivatives of* y *as in Subsection (2.2.4),*

- *HO(k,d,p) the k-step, d-derivative, (CP) Hermite–Obrechkoff method of order p,*
- *HBO(k,s,d,p) the k-step, s-stage, d-derivative, (CP) Hermite–Birkhoff–Obrechkoff method of order p.*

Hence, the method we study in this thesis is: (CP) HBO(2,6,6,13). To simpilfy notation, we will denote it by HBO(13). Moreover, as the HBO and HO methods are contractivity-preserving, we will omit (CP).
In the formula (3.1.1) and (3.1.2), we will use the following notation

- The abscissa vector $[c_1, c_2, c_3, \ldots, c_s]^T$, $0 \leq c_j \leq 1$, defines the off-step points $t_n + c_j\Delta t$, $j = 1, 2, \ldots, s$. In all cases $c_1 = 0$.
- For $1 \leq j \leq s$, F_j denotes the jth-stage derivative $f(t_n + c_j\Delta t, Y_j)$, where $Y_1 = y_n$ and Y_j is the jth-stage value.

A HBO(k,s,d,p) method to perform integration from t_n to t_{n+1} is defined by the following s formulae:

- Hermite-Birkhoff (HB) polynomials are used as predictors to obtain the stage values Y_i,

$$
\begin{aligned}
Y_i = v_{\mathrm{BU,i}} y_n + \Delta t\left[\sum_{j=1}^{i-1} a_{ij} F_j\right] + \sum_{m=2}^{d} (\Delta t)^m \gamma_{0,i,m} y_n^{(m)} \\
+ \sum_{j=1}^{k-1} A_{\mathrm{BU},ij} y_{n-j} + \Delta t\left[\sum_{j=1}^{k-1} B_{\mathrm{BU},ij} f_{n-j}\right] + \sum_{j=1}^{k-1}\left[\sum_{m=2}^{d} (\Delta t)^m \gamma_{j,i,m} y_{n-j}^{(m)}\right], \\
i = 2, 3, \ldots, s, \quad (3.1.1)
\end{aligned}
$$

- A HB polynomial is used as an integration formula to obtain y_{n+1} to order p,

$$
\begin{aligned}
y_{n+1} = v_{\mathrm{BU,s+1}} y_n + \Delta t\left[\sum_{j=1}^{s} b_j F_j\right] + \sum_{m=2}^{d} (\Delta t)^m \gamma_{0,s+1,m} y_n^{(m)} \\
+ \sum_{j=1}^{k-1} A_{\mathrm{BU},s+1,j} y_{n-j} + \Delta t\left[\sum_{j=1}^{k-1} B_{\mathrm{BU},s+1,j} f_{n-j}\right] + \sum_{j=1}^{k-1}\left[\sum_{m=2}^{d} (\Delta t)^m \gamma_{j,s+1,m} y_{n-j}^{(m)}\right], \\
(3.1.2)
\end{aligned}
$$

in the above formulae, the coefficients

1. $v_{\mathrm{BU,i}}, A_{\mathrm{BU},ij}, B_{\mathrm{BU},ij}, \gamma_{0,i,m}$ and $\gamma_{j,i,m}$, for $2 \le i \le s+1$ and $1 \le j \le k-1$,

2. a_{ij}, for $2 \le i \le s$ and $1 \le j \le i-1$,

3. b_j, for $1 \le j \le s$

are the coefficients that we compute to obtain a good approximation y_{n+1} of the solution $y(t_{n+1})$.
The subscript BU refers to the Butcher form, while the subscript SO will be used later for Shu-Osher form and canonical Shu-Osher form.

3.1.2 Construction of the order conditions

To derive the order conditions of HBO(k, s, d, p), we use the following expressions coming from the backsteps of the methods:

$$
\begin{aligned}
B_i(j) &= \sum_{\ell=1}^{k-1} A_{\mathrm{BU},ij} \frac{(-\ell)^j}{j!} + \sum_{\ell=1}^{k-1} B_{\mathrm{BU},ij} \frac{(-\ell)^{j-1}}{(j-1)!} + \gamma_{0,i,j} + \sum_{\ell=1}^{k-1} \left[\sum_{m=2}^{j} \gamma_{1,i,m} \frac{(-\ell)^{j-m}}{(j-m)!} \right], \\
B(j) &= \sum_{\ell=1}^{k-1} A_{\mathrm{BU},s+1,j} \frac{(-\ell)^j}{j!} + \sum_{\ell=1}^{k-1} B_{\mathrm{BU},s+1,j} \frac{(-\ell)^{j-1}}{(j-1)!} + \gamma_{0,s+1,j} + \sum_{\ell=1}^{k-1} \left[\sum_{m=2}^{j} \gamma_{1,s+1,m} \frac{(-\ell)^{j-m}}{(j-m)!} \right], \\
&\qquad \text{for} \quad 1 \le j \le d, \quad 2 \le i \le s \\
B_i(j) &= \sum_{\ell=1}^{k-1} A_{\mathrm{BU},ij} \frac{(-\ell)^j}{j!} + \sum_{\ell=1}^{k-1} B_{\mathrm{BU},ij} \frac{(-\ell)^{j-1}}{(j-1)!} + \sum_{\ell=1}^{k-1} \left[\sum_{m=2}^{d} \gamma_{1,i,m} \frac{(-\ell)^{j-m}}{(j-m)!} \right], \\
&\qquad j = d+1, d+2, \ldots, p, \quad i = 2, 3, \ldots, s, \\
B(j) &= \sum_{\ell=1}^{k-1} A_{\mathrm{BU},s+1,j} \frac{(-\ell)^j}{j!} + \sum_{\ell=1}^{k-1} B_{\mathrm{BU},s+1,j} \frac{(-\ell)^{j-1}}{(j-1)!} + \sum_{\ell=1}^{k-1} \left[\sum_{m=2}^{d} \gamma_{1,s+1,m} \frac{(-\ell)^{j-m}}{(j-m)!} \right], \\
&\qquad \text{for} \quad d+1 \le j \le p+1, \quad 2 \le i \le s.
\end{aligned}
\tag{3.1.3}
$$

Forcing an expansion of the numerical solution produced by formulae (3.1.1)–(3.1.2) to agree with the Taylor expansion of the true solution, we obtain multistep- and several RK-type order conditions that must be satisfied by HBO(k, s, d, p) methods.

First, we need to satisfy the set of consistency conditions:

$$v_{\mathrm{BU},i} + \sum_{j=1}^{k-1} A_{\mathrm{BU},ij} = 1, \quad i = 2, 3, \ldots, s+1. \tag{3.1.4}$$

Second, to reduce the large number of RK-type order conditions (see [19]), we impose the following simplifying assumptions:

$$\sum_{j=1}^{i-1} a_{ij} c_j^k + k! B_i(k+1) = \frac{1}{k+1}\, c_i^{k+1}, \qquad \begin{cases} i = 2, 3, \ldots, s, \\ k = 0, 1, \ldots, p-5 \end{cases} \tag{3.1.5}$$

where $c_1^0 = 1$ by convention. Thus, in the case of order $p > 5$, twelve remaining sets of equations have to be solved,

$$\sum_{i=1}^{s} b_i c_i^k + k! B(k+1) = \frac{1}{k+1}, \qquad k = 0, 1, \ldots, p-1, \tag{3.1.6}$$

$$\sum_{i=2}^{s} b_i \left[\sum_{j=1}^{i-1} a_{ij}\, \frac{c_j^{p-4}}{p-4!} + B_i(p-3) \right] + B(p-2) = \frac{1}{(p-2)!}, \tag{3.1.7}$$

$$\sum_{i=2}^{s} b_i \frac{c_i}{p-2} \left[\sum_{j=1}^{i-1} a_{ij}\, \frac{c_j^{p-4}}{(p-4)!} + B_i(p-3) \right] + B(p-1) = \frac{1}{(p-1)!}, \tag{3.1.8}$$

$$\sum_{i=2}^{s} b_i \left[\sum_{j=1}^{i-1} a_{ij}\, \frac{c_j^{p-3}}{(p-3)!} + B_i(p-2) \right] + B(p-1) = \frac{1}{(p-1)!}, \tag{3.1.9}$$

$$\sum_{i=2}^{s} b_i \left[\sum_{j=1}^{i-1} a_{ij} \left[\sum_{k=1}^{j-1} a_{jk}\, \frac{c_k^{p-4}}{(p-4)!} + B_j(p-3) \right] + B_i(p-2) \right] + B(p-1) = \frac{1}{(p-1)!}, \tag{3.1.10}$$

$$\sum_{i=2}^{s} b_i \frac{c_i^2}{(p-2)(p-1)} \left[\sum_{j=1}^{i-1} a_{ij}\, \frac{c_j^{p-4}}{(p-4)!} + B_i(p-3) \right] + B(p) = \frac{1}{p!}, \tag{3.1.11}$$

$$\sum_{i=2}^{s} b_i \frac{c_i}{p-1} \left[\sum_{j=1}^{i-1} a_{ij}\, \frac{c_j^{p-3}}{(p-3)!} + B_i(p-2) \right] + B(p) = \frac{1}{p!}, \tag{3.1.12}$$

$$\sum_{i=2}^{s} b_i \frac{c_i}{p-1} \left[\sum_{j=1}^{i-1} a_{ij} \left[\sum_{k=1}^{j-1} a_{jk}\, \frac{c_k^{p-4}}{(p-4)!} + B_j(p-3) \right] + B_i(p-2) \right] + B(p) = \frac{1}{p!}, \tag{3.1.13}$$

$$\sum_{i=2}^{s} b_i\left[\sum_{j=1}^{i-1} a_{ij}\,\frac{c_j^{p-2}}{(p-2)!} + B_i(p-1)\right] + B(p) = \frac{1}{p!}, \tag{3.1.14}$$

$$\sum_{i=2}^{s} b_i\left[\sum_{j=1}^{i-1} a_{ij}\frac{c_j}{p-2}\left[\sum_{k=1}^{j-1} a_{jk}\,\frac{c_k^{p-4}}{(p-4)!} + B_j(p-3)\right] + B_i(p-1)\right] + B(p) = \frac{1}{p!}, \tag{3.1.15}$$

$$\sum_{i=2}^{s} b_i\left[\sum_{j=1}^{i-1} a_{ij}\left[\sum_{k=1}^{j-1} a_{jk}\,\frac{c_k^{p-3}}{(p-3)!} + B_j(p-2)\right] + B_i(p-1)\right] + B(p) = \frac{1}{p!}, \tag{3.1.16}$$

$$\sum_{i=2}^{s} b_i\left\{\sum_{j=1}^{i-1} a_{ij}\left[\sum_{k=1}^{j-1} a_{jk}\left(\sum_{\ell=1}^{k-1} a_{k\ell}\,\frac{c_\ell^{p-4}}{(p-4)!} + B_k(p-3)\right) + B_j(p-2)\right] + B_i(p-1)\right\} + B(p) = \frac{1}{p!}. \tag{3.1.17}$$

These order conditions are simply RK order conditions with backstep parts $B_i(\cdot)$ and $B(\cdot)$.

3.2 Construction of the HBO(13) method

We construct 6-stage Hermite–Birkhoff–Obrechkoff methods, as a subclass of multi-derivative, multistep, multistage, methods, by the formulae (3.2.1) to (3.2.6).

Let Δt denote the step size. The abscissa vector $[c_1, c_2, c_3, \ldots, c_6]^T$ defines the six off-step points $t_n + c_j\Delta t$, for $1 \le j \le 6$. In order to perform integration step from t_n to t_{n+1}, the predictors P_j, for $2 \le j \le 6$ and the integration formulae IF are computed as follows:

(P_2) A Hermite-Birkhoff polynomial is used as predictor P_2 to obtain Y_2 to order 9,

$$Y_2 = v_{\mathrm{BU},2}y_n + \Delta t\left[\sum_{j=1}^{i-1} a_{2j}F_j\right] + \sum_{m=2}^{6}(\Delta t)^m\gamma_{0,2,m}y_n^{(m)} + A_{\mathrm{BU},2}y_{n-1} + \Delta t B_{\mathrm{BU},2}f_{n-1} + \sum_{m=2}^{6}(\Delta t)^m\gamma_{1,2,m}y_{n-1}^{(m)}, \tag{3.2.1}$$

where for $1 \le j \le 6$, $F_j := f(t_n + c_j\Delta t, Y_j)$, denotes the stage derivatives, $F_1 = f_n$ and $Y_1 = y_n$. In all cases $c_1 = 0$.

(P$_3$) A Hermite-Birkhoff polynomial is used as predictor P$_3$ to obtain Y_3 to order 9,

$$Y_3 = v_{\mathrm{BU},3} y_n + \Delta t \left[\sum_{j=1}^{i-1} a_{3j} F_j \right] + \sum_{m=2}^{6} (\Delta t)^m \gamma_{0,3,m} y_n^{(m)}$$
$$+ A_{\mathrm{BU},3} y_{n-1} + \Delta t B_{\mathrm{BU},3} f_{n-1} + \sum_{m=2}^{6} (\Delta t)^m \gamma_{1,3,m} y_{n-1}^{(m)}, \quad (3.2.2)$$

(P$_4$) A Hermite-Birkhoff polynomial is used as predictor P$_4$ to obtain Y_4 to order 9,

$$Y_4 = v_{\mathrm{BU},4} y_n + \Delta t \left[\sum_{j=1}^{i-1} a_{4j} F_j \right] + \sum_{m=2}^{6} (\Delta t)^m \gamma_{0,4,m} y_n^{(m)}$$
$$+ A_{\mathrm{BU},4} y_{n-1} + \Delta t B_{\mathrm{BU},4} f_{n-1} + \sum_{m=2}^{6} (\Delta t)^m \gamma_{1,4,m} y_{n-1}^{(m)}, \quad (3.2.3)$$

(P$_5$) A Hermite-Birkhoff polynomial is used as predictor P$_5$ to obtain Y_5 to order 9,

$$Y_5 = v_{\mathrm{BU},5} y_n + \Delta t \left[\sum_{j=1}^{i-1} a_{5j} F_j \right] + \sum_{m=2}^{6} (\Delta t)^m \gamma_{0,5,m} y_n^{(m)}$$
$$+ A_{\mathrm{BU},5} y_{n-1} + \Delta t B_{\mathrm{BU},5} f_{n-1} + \sum_{m=2}^{6} (\Delta t)^m \gamma_{1,5,m} y_{n-1}^{(m)}, \quad (3.2.4)$$

(P$_6$) A Hermite-Birkhoff polynomial is used as predictor P$_6$ to obtain Y_6 to order 9,

$$Y_6 = v_{\mathrm{BU},6} y_n + \Delta t \left[\sum_{j=1}^{i-1} a_{6j} F_j \right] + \sum_{m=2}^{6} (\Delta t)^m \gamma_{0,6,m} y_n^{(m)}$$
$$+ A_{\mathrm{BU},6} y_{n-1} + \Delta t B_{\mathrm{BU},6} f_{n-1} + \sum_{m=2}^{6} (\Delta t)^m \gamma_{1,6,m} y_{n-1}^{(m)}, \quad (3.2.5)$$

(IF) A Hermite-Birkhoff polynomial is used as integration formula IF to obtain y_{n+1} to order 13,

$$y_{n+1} = v_{\mathrm{BU},7} y_n + \Delta t \left[\sum_{j=1}^{6} b_j F_j \right] + \sum_{m=2}^{6} (\Delta t)^m \gamma_{0,7,m} y_n^{(m)}$$
$$+ A_{\mathrm{BU},7} y_{n-1} + \Delta t B_{\mathrm{BU},7} f_{n-1} + \sum_{m=2}^{6} (\Delta t)^m \gamma_{1,7,m} y_{n-1}^{(m)}. \quad (3.2.6)$$

Here the subscript BU refers to the Butcher form, while the subscript SO will be used later for the Shu–Osher form.

Formulae (3.2.1)–(3.2.6) are the Butcher form of HBO(13).

Note that the set of derivatives $y_n^{(m)}$, $m = 2, 3, \ldots, 6$, is computed only once per step at $t = t_n$. The defining formulae of 2-step HBO(13) involve the usual RK parameters c_i, $a_{i,j}$ and b_j and the Taylor expansion parameters $\gamma_{0,i,j}$ and $\gamma_{1,i,j}$. Then, we can represent a HBO(13) method by its *coefficient scheme* $(A, b, \boldsymbol{\gamma}_0, \boldsymbol{\gamma}_1)$ where A denotes the (6×6) matrix $A = (a_{i,j})$, b the 6-vector $b = (b_1, b_2, \ldots, b_6)^T$, $\boldsymbol{\gamma}_0$ and $\boldsymbol{\gamma}_1$ the (7×5) matrices $\boldsymbol{\gamma}_0 = (\gamma_{0,i,j})$ and $\boldsymbol{\gamma}_1 = (\gamma_{1,i,j})$ of Taylor expansion parameters $\gamma_{0,i,j}$ and $\gamma_{1,i,j}$, respectively. We can display the coefficient scheme $(A, b, \boldsymbol{\gamma}_0, \boldsymbol{\gamma}_1)$ and the c_i in the Butcher tableau and the (7×5) matrices $\boldsymbol{\gamma}_i$, $i = 0, 1$, as follows.

$$
\begin{array}{c|ccccc}
c_1 & & & & & \\
c_2 & a_{2,1} & & & & \\
c_3 & a_{3,1} & a_{3,2} & & & \\
\vdots & \vdots & \vdots & \ddots & & \\
c_6 & a_{6,1} & a_{6,2} & \cdots & a_{6,5} & \\
\hline
 & b_1 & b_2 & \cdots & b_5 & b_6
\end{array}
\qquad
\boldsymbol{\gamma}_i = \begin{bmatrix}
0 & 0 & \cdots & 0 \\
\gamma_{i,2,2} & \gamma_{i,2,3} & \cdots & \gamma_{i,2,6} \\
\gamma_{i,3,2} & \gamma_{i,3,3} & \cdots & \gamma_{i,3,6} \\
\vdots & & \cdots & \vdots \\
\gamma_{i,7,2} & \gamma_{i,7,3} & \cdots & \gamma_{i,7,6}
\end{bmatrix}. \tag{3.2.7}
$$

3.3 Order conditions for HBO(13) method

To derive the order conditions of 6-stage HBO(13), we use the following expressions coming from the backsteps of the method:

$$
\begin{aligned}
B_i(j) &= A_{\mathrm{BU},i}\,\frac{(-1)^j}{j!} + B_{\mathrm{BU},i}\,\frac{(-1)^{j-1}}{(j-1)!} + \gamma_{0,i,j} + \sum_{m=2}^{j}\gamma_{1,i,m}\,\frac{(-1)^{j-m}}{(j-m)!},\\
B(j) &= A_{\mathrm{BU},7}\,\frac{(-1)^j}{j!} + B_{\mathrm{BU},7}\,\frac{(-1)^{j-1}}{(j-1)!} + \gamma_{0,7,j} + \sum_{m=2}^{j}\gamma_{1,7,m}\,\frac{(-1)^{j-m}}{(j-m)!},\\
&\qquad\qquad\text{for}\quad 1\le j\le 6,\quad 2\le i\le 6,\\
B_i(j) &= A_{\mathrm{BU},i}\,\frac{(-1)^j}{j!} + B_{\mathrm{BU},i}\,\frac{(-1)^{j-1}}{(j-1)!} + \sum_{m=2}^{6}\gamma_{1,i,m}\,\frac{(-1)^{j-m}}{(j-m)!},\\
&\qquad\qquad\text{for}\quad 7\le j\le 13,\quad 2\le i\le 6,\\
B(j) &= A_{\mathrm{BU},7}\,\frac{(-1)^j}{j!} + B_{\mathrm{BU},7}\,\frac{(-1)^{j-1}}{(j-1)!} + \sum_{m=2}^{6}\gamma_{1,7,m}\,\frac{(-1)^{j-m}}{(j-m)!},\\
&\qquad\qquad j = 7, 8, \ldots, 14.
\end{aligned}
\tag{3.3.1}
$$

where $\gamma_{0,i,1} = 0, i = 2, 3, \ldots, 7$.

To obtain the order conditions for HBO(13) method, we do the following:
Firstly, we match Taylor expansions of the exact solution and numerical solution produced by formulae (3.2.1)–(3.2.6) up to order 13. This will produce a very large number of RK-type order conditions of order 13. These RK-type order conditions include the set of consistency conditions which are

$$v_{\mathrm{BU},i} + A_{\mathrm{BU},i} = 1, \quad i = 2, 3, \ldots, 7. \tag{3.3.2}$$

Secondly, to reduce the very large number of RK-type order conditions of order 13 (see [19]), we impose the following simplifying assumptions:

$$\sum_{j=1}^{i-1} a_{ij}c_j^k + k!B_i(k+1) = \frac{1}{k+1}\,c_i^{k+1}, \qquad \begin{cases} i = 2, 3, \ldots, 6, \\ k = 0, 1, \ldots, 8. \end{cases} \tag{3.3.3}$$

where $c_1^0 = 1$ by convention. The set of equations (3.3.3) is derived from the Taylor expansions of each stage Y_i up to order 9. We use the set of simplifying assumptions (3.3.3) to reduce the very large number of RK-type order conditions of order 13 to twelve remaining sets of equations of order 13 to be solved. These order conditions are comparable to the set of order conditions of RK of order 5:

$$\sum_{i=1}^{6} b_i c_i^k + k!B(k+1) = \frac{1}{k+1}, \qquad k = 0, 1, \ldots, 12, \tag{3.3.4}$$

$$\sum_{i=2}^{6} b_i \left[\sum_{j=1}^{i-1} a_{ij}\, \frac{c_j^9}{9!} + B_i(10)\right] + B(11) = \frac{1}{11!}, \tag{3.3.5}$$

$$\sum_{i=2}^{6} b_i \frac{c_i}{11} \left[\sum_{j=1}^{i-1} a_{ij}\, \frac{c_j^9}{9!} + B_i(10)\right] + B(12) = \frac{1}{12!}, \tag{3.3.6}$$

$$\sum_{i=2}^{6} b_i \left[\sum_{j=1}^{i-1} a_{ij}\, \frac{c_j^{10}}{10!} + B_i(11)\right] + B(12) = \frac{1}{12!}, \tag{3.3.7}$$

$$\sum_{i=2}^{6} b_i \left[\sum_{j=1}^{i-1} a_{ij} \left[\sum_{k=1}^{j-1} a_{jk}\, \frac{c_k^9}{9!} + B_j(10)\right] + B_i(11)\right] + B(12) = \frac{1}{12!}, \tag{3.3.8}$$

$$\sum_{i=2}^{6} b_i \frac{c_i^2}{11 \times 12} \left[\sum_{j=1}^{i-1} a_{ij}\, \frac{c_j^9}{9!} + B_i(10)\right] + B(13) = \frac{1}{13!}, \tag{3.3.9}$$

$$\sum_{i=2}^{6} b_i \frac{c_i}{12} \left[\sum_{j=1}^{i-1} a_{ij}\, \frac{c_j^{10}}{10!} + B_i(11)\right] + B(13) = \frac{1}{13!}, \tag{3.3.10}$$

$$\sum_{i=2}^{6} b_i \frac{c_i}{12} \left[\sum_{j=1}^{i-1} a_{ij} \left[\sum_{k=1}^{j-1} a_{jk}\, \frac{c_k^9}{9!} + B_j(10)\right] + B_i(11)\right] + B(13) = \frac{1}{13!}, \tag{3.3.11}$$

$$\sum_{i=2}^{6} b_i \left[\sum_{j=1}^{i-1} a_{ij}\, \frac{c_j^{11}}{11!} + B_i(12)\right] + B(13) = \frac{1}{13!}, \tag{3.3.12}$$

$$\sum_{i=2}^{6} b_i \left[\sum_{j=1}^{i-1} a_{ij} \frac{c_j}{11} \left[\sum_{k=1}^{j-1} a_{jk}\, \frac{c_k^9}{9!} + B_j(10)\right] + B_i(12)\right] + B(13) = \frac{1}{13!}, \tag{3.3.13}$$

$$\sum_{i=2}^{6} b_i \left[\sum_{j=1}^{i-1} a_{ij} \left[\sum_{k=1}^{j-1} a_{jk}\, \frac{c_k^{10}}{10!} + B_j(11)\right] + B_i(12)\right] + B(13) = \frac{1}{13!}, \tag{3.3.14}$$

$$\sum_{i=2}^{6} b_i \left\{\sum_{j=1}^{i-1} a_{ij} \left[\sum_{k=1}^{j-1} a_{jk} \left(\sum_{\ell=1}^{k-1} a_{k\ell}\, \frac{c_\ell^9}{9!} + B_k(10)\right) + B_j(11)\right] + B_i(12)\right\} + B(13) = \frac{1}{13!}. \tag{3.3.15}$$

These order conditions are the RK order conditions with backstep parts $B_i(\cdot)$ and $B(\cdot)$.

3.4 Shu-Osher and modified Shu-Osher forms of HBO(13) method for deriving the CP property

In this section, we determine the Shu-Osher and modified Shu-Osher forms of a subclass of Butcher form of HBO(13) method (see conditions (3.4.3) and (3.4.4)). And we use them to express the method HBO(13) as convex combinations of S(6), defined in (1.1.2). Therefore, it will perserve the *contractivity-perserving property* (1.1.5). For convenience, we rewrite HBO(13) formulae (3.2.1)–(3.2.6) as follows:

$$Y_i = v_{\mathrm{BU,i}} y_n + \Delta t \left[\sum_{j=1}^{i-1} a_{ij} F_j \right] + \sum_{m=2}^{6} (\Delta t)^m \gamma_{0,i,m} y_n^{(m)} + A_{\mathrm{BU},i} y_{n-1} + \Delta t B_{\mathrm{BU},i} f_{n-1} + \sum_{m=2}^{6} (\Delta t)^m \gamma_{1,i,m} y_{n-1}^{(m)}, \qquad i = 2, 3, \ldots, 6 \quad (3.4.1)$$

$$y_{n+1} = v_{\mathrm{BU,7}} y_n + \Delta t \left[\sum_{j=1}^{6} b_j F_j \right] + \sum_{m=2}^{6} (\Delta t)^m \gamma_{0,7,m} y_n^{(m)} + A_{\mathrm{BU,7}} y_{n-1} + \Delta t B_{\mathrm{BU,7}} f_{n-1} + \sum_{m=2}^{6} (\Delta t)^m \gamma_{1,7,m} y_{n-1}^{(m)}. \quad (3.4.2)$$

We let

$$v_{\mathrm{BU,i}} \neq 0, \text{for} \;\; 2 \leq i \leq 7, \quad (3.4.3)$$

and similar to the generalization to HB methods [20] of the Shu-Osher form of RK methods [27], we present how to write HBO(13) in its Shu-Osher form.

Firstly, let

$$\lambda_{i\ell} > 0, \qquad \sum_{\ell=1}^{i-1} \lambda_{i\ell} = 1, \;\text{for}\;\; 3 \leq i \leq 7. \quad (3.4.4)$$

Then, formulae (3.4.1) and (3.4.2) become

$$Y_i = \left[\sum_{\ell=1}^{i-1}\lambda_{i\ell}\right]v_{\mathrm{BU,i}}y_n + \Delta t\left[\sum_{j=1}^{i-1}a_{ij}F_j\right] + \sum_{m=2}^{6}(\Delta t)^m\gamma_{0,i,m}y_n^{(m)}$$
$$+ A_{\mathrm{BU},i}y_{n-1} + \Delta t B_{\mathrm{BU},i}f_{n-1} + \sum_{m=2}^{6}(\Delta t)^m\gamma_{1,i,m}y_{n-1}^{(m)}, \text{ for } 3 \le i \le 6, \quad (3.4.5)$$

$$y_{n+1} = \left[\sum_{\ell=1}^{6}\lambda_{7\ell}\right]v_{\mathrm{BU,7}}y_n + \Delta t\left[\sum_{j=1}^{6}b_jF_j\right] + \sum_{m=2}^{6}(\Delta t)^m\gamma_{0,7,m}y_n^{(m)}$$
$$+ A_{\mathrm{BU,7}}y_{n-1} + \Delta t B_{\mathrm{BU,7}}f_{n-1} + \sum_{m=2}^{6}(\Delta t)^m\gamma_{1,7,m}y_{n-1}^{(m)}. \quad (3.4.6)$$

Secondly, we express the term y_n in formulae (3.4.1) as a function of Y_i, for $2 \le i \le 6$

$$y_n = \frac{1}{v_{\mathrm{BU,i}}}\left\{Y_i - \Delta t\left[\sum_{j=1}^{i-1}a_{ij}F_j\right] - \sum_{m=2}^{6}(\Delta t)^m\gamma_{0,i,m}y_n^{(m)}\right.$$
$$\left. - A_{\mathrm{BU},i}y_{n-1} - \Delta t B_{\mathrm{BU},i}f_{n-1} - \sum_{m=2}^{6}(\Delta t)^m\gamma_{1,i,m}y_{n-1}^{(m)}\right\}. \quad (3.4.7)$$

We do not want to be confused with the index i so we replace the index i by ξ in formulae (3.4.7) to obtain

$$y_n = \frac{1}{v_{\mathrm{BU},\xi}}\left\{Y_\xi - \Delta t\left[\sum_{j=1}^{\xi-1}a_{\xi j}F_j\right] - \sum_{m=2}^{6}(\Delta t)^m\gamma_{0,\xi,m}y_n^{(m)}\right.$$
$$\left. - A_{\mathrm{BU},\xi}\ y_{n-1} - \Delta t B_{\mathrm{BU},\xi}\ f_{n-1} - \sum_{m=2}^{6}(\Delta t)^m\gamma_{1,\xi,m}y_{n-1}^{(m)}\right\},$$
$$\text{for } 2 \le \xi \le 6. \quad (3.4.8)$$

Thirdly, for $3 \le i \le 6$ and $\ell = 1, 2, \ldots, i-1$, we replace y_n in (3.4.1) by the right hand side of (3.4.8) with $\xi = \ell$. Similarly, we replace y_n in (3.4.2) by the right hand side of (3.4.8) with $\xi = \ell$. If we set

$$\alpha_{ij} = \frac{\lambda_{ij}v_{\mathrm{BU,i}}}{v_{\mathrm{BU,j}}}, \text{ for } \ j = 2, 3, \ldots, i-1, \text{and } \ 3 \le i \le 7, \quad (3.4.9)$$

hence, after some calculations, formulae (3.4.1) with $i = 2$, (3.4.5) and (3.4.6) become

$$\begin{aligned} Y_i &= \left[\sum_{j=1}^{i-1} \alpha_{ij} Y_j + \Delta t\, \beta_{ij} F_j\right] + \sum_{m=2}^{6} (\Delta t)^m \delta_{0,i,m} y_n^{(m)} \\ &\quad + \left[A_i y_{n-1} + \Delta t B_i f_{n-1}\right] + \sum_{m=2}^{6} (\Delta t)^m \delta_{1,i,m} y_{n-1}^{(m)}, \text{ for } \ 2 \le i \le 7, \\ y_{n+1} &= Y_7, \end{aligned} \tag{3.4.10}$$

where the coefficients are

$$\alpha_{i1} = v_{\mathrm{BU,i}} - \sum_{\ell=2}^{i-1} \alpha_{i\ell}\ v_{\mathrm{BU},\ell}, \quad \text{for } \ 3 \le i \le 7, \tag{3.4.11}$$

$$\alpha_{21} = v_{\mathrm{BU},2}, \tag{3.4.12}$$

$$\beta_{ij} = a_{ij} - \sum_{\ell=j+1}^{i-1} \alpha_{i\ell}\ a_{\ell j}, \quad \text{for } \ 3 \le i \le 6 \ \text{ and } \ j = 1, 2, 3, \ldots, i-1, \tag{3.4.13}$$

$$\beta_{21} = a_{21}, \tag{3.4.14}$$

$$\beta_{7j} = b_j - \sum_{\ell=j+1}^{6} \alpha_{7\ell}\ a_{\ell j}, \qquad \text{for } \ 1 \le j \le 6, \tag{3.4.15}$$

$$A_i = A_{\mathrm{BU},i} - \sum_{\ell=2}^{i-1} \alpha_{i\ell}\ A_{\mathrm{BU},\ell}, \qquad \text{for } \ 3 \le i \le 7, \tag{3.4.16}$$

$$A_2 = A_{\mathrm{BU},2}, \tag{3.4.17}$$

$$B_i = B_{\mathrm{BU},i} - \sum_{\ell=2}^{i-1} \alpha_{i\ell}\ B_{\mathrm{BU},\ell}, \qquad \text{for } \ 3 \le i \le 7, \tag{3.4.18}$$

$$B_2 = B_{\mathrm{BU},2}, \tag{3.4.19}$$

$$\delta_{0,i,m} = \gamma_{0,i,m} - \sum_{\ell=2}^{i-1} \alpha_{i\ell}\ \gamma_{0,\ell,m}, \qquad \text{for } \ 3 \le i \le 7 \ \text{ and } \ 2 \le m \le 6, \tag{3.4.20}$$

$$\delta_{0,2,m} = \gamma_{0,2,m}, \qquad \text{for } \ 2 \le m \le 6, \tag{3.4.21}$$

$$\delta_{1,i,m} = \gamma_{1,i,m} - \sum_{\ell=2}^{i-1} \alpha_{i\ell}\ \gamma_{1,\ell,m}, \qquad \text{for } \ 3 \le i \le 7 \ \text{ and } \ 2 \le m \le 6, \tag{3.4.22}$$

$$\delta_{1,2,m} = \gamma_{1,2,m}, \qquad \text{for } \ 2 \le m \le 6, \tag{3.4.23}$$

with consistency conditions:

$$\sum_{j=1}^{i-1} \alpha_{ij} + A_i = 1, \qquad i = 2, 3, \dots, 7. \tag{3.4.24}$$

In fact, for $i = 2, 3, \dots, 7$, we have:

$$\begin{aligned}
\sum_{j=1}^{i-1} \alpha_{ij} + A_i &= \alpha_{i1} + \sum_{j=2}^{i-1} \alpha_{ij} + A_i \\
&= v_{\mathrm{BU,i}} - \sum_{\ell=2}^{i-1} \alpha_{i\ell}\, v_{\mathrm{BU},\ell} + \sum_{j=2}^{i-1} \alpha_{ij} + A_i && \text{(by (3.4.11))} \\
&= v_{\mathrm{BU,i}} - \sum_{\ell=2}^{i-1} \alpha_{i\ell}\, v_{\mathrm{BU},\ell} + \sum_{j=2}^{i-1} \alpha_{ij} + A_{\mathrm{BU},i} - \sum_{\ell=2}^{i-1} \alpha_{i\ell}\, A_{\mathrm{BU},\ell} && \text{(by (3.4.16))} \\
&= v_{\mathrm{BU,i}} + A_{\mathrm{BU},i} - \sum_{\ell=2}^{i-1} \alpha_{i\ell}\, (v_{\mathrm{BU},\ell} + A_{\mathrm{BU},\ell}) + \sum_{j=2}^{i-1} \alpha_{ij} \\
&= v_{\mathrm{BU,i}} + A_{\mathrm{BU},i} - \sum_{\ell=2}^{i-1} \alpha_{i\ell} + \sum_{j=2}^{i-1} \alpha_{ij} && \text{(by (3.3.2))} \\
&= 1 - \sum_{\ell=2}^{i-1} \alpha_{i\ell} + \sum_{j=2}^{i-1} \alpha_{ij} = 1. && \text{(by (3.3.2))}
\end{aligned}$$

Form (3.4.10) is called the Shu–Osher form of HBO(13).

Now, we set

$$v_i = \alpha_{i1}, \text{ for } \ 2 \le i \le 7, \tag{3.4.25}$$

and

$$w_i = \beta_{i1}, \text{ for } \ 2 \le i \le 7, \tag{3.4.26}$$

in (3.4.10), then we have the *modified Shu–Osher form* of HBO(13):

$$\begin{aligned}
Y_i = &\left[v_i y_n + \Delta t\, w_i f_n + \sum_{m=2}^{6} (\Delta t)^m \delta_{0,i,m} y_n^{(m)}\right] + \Bigg[A_i y_{n-1} + \Delta t B_i f_{n-1} \\
&+ \sum_{m=2}^{6} (\Delta t)^m \delta_{1,i,m} y_{n-1}^{(m)}\Bigg] + \left[\sum_{j=2}^{i-1} \alpha_{ij} Y_j + \Delta t\, \beta_{ij} F_j\right], \quad i = 2, 3, \dots, 7,
\end{aligned} \tag{3.4.27}$$

$$y_{n+1} = Y_7.$$

This form generalizes the modified Shu–Osher form for RK methods (see [27, 7]). From (3.4.3), (3.4.4), (3.4.11) and (3.4.12), we note that

$$\alpha_{ij} \neq 0, \text{ for } \ 2 \leq i \leq 7, \text{ and } \ j = 1, 2, \ldots, i-1. \tag{3.4.28}$$

If we assume that

$$A_i \neq 0, \text{ for } \ 2 \leq i \leq 7, \tag{3.4.29}$$

then we can rearrange the stage values Y_i, for $2 \leq i \leq 7$, in (3.4.27) as follows:

$$\begin{aligned} Y_i = v_i \Bigg[y_n + \Delta t \, \frac{w_i}{v_i} f_n + \sum_{m=2}^{6} (\Delta t)^m \frac{\delta_{0,i,m}}{v_i} y_n^{(m)} \Bigg] + A_i \Bigg[y_{n-1} + \Delta t \frac{B_i}{A_i} f_{n-1} \\ + \sum_{m=2}^{6} (\Delta t)^m \frac{\delta_{1,i,m}}{A_i} y_{n-1}^{(m)} \Bigg] + \sum_{j=2}^{i-1} \alpha_{ij} \left(Y_j + \Delta t \, \frac{\beta_{ij}}{\alpha_{ij}} F_j \right), \end{aligned} \tag{3.4.30}$$

with consistency conditions:

$$v_i + A_i + \sum_{j=2}^{i-1} \alpha_{ij} = 1, \text{ for } \ 2 \leq i \leq 7. \tag{3.4.31}$$

Clearly, (3.4.30) is the convex combination of S(6) (1.1.2) with the step sizes $\Delta t \, \frac{w_i}{v_i}$, $\Delta t \frac{B_i}{A_i}$ and $\Delta t \, \frac{\beta_{ij}}{\alpha_{ij}}$ whenever v_i, A_i, α_{ij} are strictly positive. Note that (3.4.30) is a subclass of modified Shu-Osher form of HBO(13) (3.4.27) (see conditions (3.4.28) and (3.4.29)).

From (3.4.30), we obtain the difference $Y_i - \widetilde{Y}_i$, for $2 \leq i \leq 7$, as follows:

$$\begin{aligned} Y_i - \widetilde{Y}_i = v_i \Bigg[\left(y_n + \Delta t \, \frac{w_i}{v_i} \, f_n + \sum_{m=2}^{6} (\Delta t)^m \, \frac{\delta_{0,i,m}}{v_i} \, y_n^{(m)} \right) \\ - \left(\tilde{y}_n + \Delta t \, \frac{w_i}{v_i} \, \tilde{f}_n + \sum_{m=2}^{6} (\Delta t)^m \, \frac{\delta_{0,i,m}}{v_i} \, \tilde{y}_n^{(m)} \right) \Bigg] \\ + A_i \Bigg[\left(y_{n-1} + \Delta t \, \frac{B_i}{A_i} f_{n-1} + \sum_{m=2}^{6} (\Delta t)^m \, \frac{\delta_{1,i,m}}{A_i} \, y_{n-1}^{(m)} \right) \\ - \left(\tilde{y}_{n-1} + \Delta t \, \frac{B_i}{A_i} \, \tilde{f}_{n-1} + \sum_{m=2}^{6} (\Delta t)^m \, \frac{\delta_{1,i,m}}{A_i} \, \tilde{y}_{n-1}^{(m)} \right) \Bigg] \end{aligned}$$

$$+\sum_{j=2}^{i-1}\alpha_{ij}\left[\left(Y_j+\Delta t\,\frac{\beta_{ij}}{\alpha_{ij}}\,F_j\right)-\left(\widetilde{Y}_j+\Delta t\,\frac{\beta_{ij}}{\alpha_{ij}}\,\widetilde{F}_j\right)\right], \tag{3.4.32}$$

where $\widetilde{Y}_i$ defined in Section 2.6. Provided all the coefficients of (3.4.30) are nonnegative, the following straightforward extension of a result presented in [8, 12] holds.

Theorem 3.4.1. *If f satisfies condition* (1.1.4) *of the S(6) method, then the* 2*-step,* 6*-stage,* 6*-derivative, HBO(13) method* (3.4.30)*, under the assumption that all coefficients of* (3.4.30) *are nonnegative, satisfies the CP property*

$$\|y_{n+1}-\tilde{y}_{n+1}\|\le\max_{0\le j\le 1}\|y_{n-j}-\tilde{y}_{n-j}\|,$$

provided

$$\Delta t\le c_{\text{feasible}}\,\Delta t_{S(6)},$$

where

- *the feasible CP coefficient, c_{feasible} is the minimum of the following numbers:*

$$\begin{aligned} r_{i0}&=\frac{v_i}{w_i}, \quad i=2,3,\dots,7,\\ \min_{j=2,3,\dots,i-1}&\left\{\frac{\alpha_{ij}}{\beta_{ij}}\right\}, \quad i=3,4,\dots,7,\\ r_{i1}&=\frac{A_i}{B_i}, \quad i=2,3,\dots,7,\end{aligned} \tag{3.4.33}$$

- *For $\ell=0,1$, $i=2,3,\dots,7$ and $2\le m\le 7$, the following conditions are imposed on $\delta_{\ell,i,m}$,*

$$\frac{\delta_{0,i,m}}{v_i}\le\left[\frac{1}{r_{i0}}\right]^m\frac{1}{m!},\qquad \frac{\delta_{1,i,m}}{A_i}\le\left[\frac{1}{r_{i1}}\right]^m\frac{1}{m!}, \tag{3.4.34}$$

with the convention that $a/0=+\infty$.

Proof. The difference $Y_i-\widetilde{Y}_i$ of HBO(13) can be rewritten as a convex combination of the three terms on the right-hand side of (3.4.32). Thus, by convexity of the norm $\|\cdot\|$, we have, for $2\le i\le 7$,

$$\|Y_i-\widetilde{Y}_i\|\le v_i\left\|\left(y_n+\Delta t\,\frac{w_i}{v_i}\,f_n+\sum_{m=2}^{6}(\Delta t)^m\,\frac{\delta_{0,i,m}}{v_i}\,y_n^{(m)}\right)\right.$$

$$
\begin{aligned}
&\qquad - \left(\tilde{y}_n + \Delta t\, \frac{w_i}{v_i}\, \tilde{f}_n + \sum_{m=2}^{6} (\Delta t)^m\, \frac{\delta_{0,i,m}}{v_i}\, \tilde{y}_n^{(m)} \right) \Bigg\| \\
&+ A_i \Bigg\| \left(y_{n-1} + \Delta t\, \frac{B_i}{A_i}\, f_{n-1} + \sum_{m=2}^{6} (\Delta t)^m\, \frac{\delta_{1,i,m}}{A_i}\, y_{n-1}^{(m)} \right) \\
&\qquad - \left(\tilde{y}_{n-1} + \Delta t\, \frac{B_i}{A_i}\, \tilde{f}_{n-1} + \sum_{m=2}^{6} (\Delta t)^m\, \frac{\delta_{1,i,m}}{A_i}\, \tilde{y}_{n-1}^{(m)} \right) \Bigg\| \\
&+ \sum_{j=2}^{i-1} \alpha_{ij} \Bigg\| \left(Y_j + \Delta t\, \frac{\beta_{ij}}{\alpha_{ij}}\, F_j \right) - \left(\tilde{Y}_j + \Delta t\, \frac{\beta_{ij}}{\alpha_{ij}}\, \tilde{F}_j \right) \Bigg\|. \qquad (3.4.35)
\end{aligned}
$$

The first two terms on the right-hand side of (3.4.35) have the following upper bounds:

$$
\begin{aligned}
v_i &\Bigg\| \left(y_n + \Delta t\, \frac{w_i}{v_i}\, f_n + \sum_{m=2}^{6} (\Delta t)^m\, \frac{\delta_{0,i,m}}{v_i}\, y_n^{(m)} \right) \\
&\qquad - \left(\tilde{y}_n + \Delta t\, \frac{w_i}{v_i}\, \tilde{f}_n + \sum_{m=2}^{6} (\Delta t)^m\, \frac{\delta_{0,i,m}}{v_i}\, \tilde{y}_n^{(m)} \right) \Bigg\| \\
&+ A_i \Bigg\| \left(y_{n-1} + \Delta t\, \frac{B_i}{A_i}\, f_{n-1} + \sum_{m=2}^{6} (\Delta t)^m\, \frac{\delta_{1,i,m}}{A_i}\, y_{n-1}^{(m)} \right) \\
&\qquad - \left(\tilde{y}_{n-1} + \Delta t\, \frac{B_i}{A_i}\, \tilde{f}_{n-1} + \sum_{m=2}^{6} (\Delta t)^m\, \frac{\delta_{1,i,m}}{A_i}\, \tilde{y}_{n-1}^{(m)} \right) \Bigg\| \\
&\le v_i \big\| y_n - \tilde{y}_n \big\| + A_i \big\| y_{n-1} - \tilde{y}_{n-1} \big\| \quad \text{by (1.1.4) and (3.4.34)} \\
&\le (v_i + A_i) \max_{0 \le \ell \le 1} \| y_{n-\ell} - \tilde{y}_{n-\ell} \|, \quad i = 2, 3, \ldots, 7,
\end{aligned}
$$

since

$$
\frac{1}{r_{i0}} \Delta t \le \frac{\Delta t}{c_{\text{feasible}}} \le \Delta t_{\text{S}(6)} \quad \text{and} \quad \frac{1}{r_{i1}} \Delta t \le \frac{\Delta t}{c_{\text{feasible}}} \le \Delta t_{\text{S}(6)}.
$$

The third term on the right-hand side of (3.4.35) has the following upper bound:

$$\begin{aligned}
&\sum_{j=2}^{i-1} \alpha_{ij} \left\| \left(Y_j + \Delta t \, \frac{\beta_{ij}}{\alpha_{ij}} \, F_j \right) - \left(\widetilde{Y}_j + \Delta t \, \frac{\beta_{ij}}{\alpha_{ij}} \, \widetilde{F}_j \right) \right\| \\
&\qquad \le \sum_{j=2}^{i-1} \alpha_{ij} \left\| \left(Y_j + \frac{\Delta t}{c_{\text{feasible}}} \, F_j \right) - \left(\widetilde{Y}_j + \frac{\Delta t}{c_{\text{feasible}}} \, \widetilde{F}_j \right) \right\| \\
&\qquad \le \sum_{j=2}^{i-1} \alpha_{ij} \max_{0 \le \ell \le 1} \| y_{n-\ell} - \tilde{y}_{n-\ell} \| \,, \quad i = 2, 3, \ldots, 7, \quad \text{by (1.1.4)},
\end{aligned}$$

since $\dfrac{\beta_{ij}}{\alpha_{ij}} \Delta t \le \dfrac{\Delta t}{c_{\text{feasible}}} \le \Delta t_{\text{S(6)}}$.

Because v_i, A_i and α_{ij} are strictly positive and $v_i + A_i + \sum_{j=2}^{i-1} \alpha_{ij} = 1$, we have the inequality

$$\begin{aligned}
\left\| Y_i - \widetilde{Y}_i \right\| &\le (v_i + A_i) \max_{0 \le \ell \le 1} \| y_{n-\ell} - \tilde{y}_{n-\ell} \| + \sum_{j=2}^{i-1} \alpha_{ij} \max_{0 \le \ell \le 1} \| y_{n-\ell} - \tilde{y}_{n-\ell} \| \\
&\le \max_{0 \le j \le 1} \| y_{n-j} - \tilde{y}_{n-j} \| \,, \quad i = 2, 3, \ldots, 7.
\end{aligned}$$

We thus obtain $\left\| Y_i - \widetilde{Y}_i \right\| \le \max_{0 \le j \le 1} \| y_{n-j} - \tilde{y}_{n-j} \|$ for $i = 2, 3, \ldots, 7$. In particular, this yields $\| y_{n+1} - \tilde{y}_{n+1} \| \le \max_{0 \le j \le 1} \| y_{n-j} - \tilde{y}_{n-j} \|$ for $i = 7$. □

Note that the coefficients v_i, w_i, α_{ij}, β_{ij}, A_i, B_i, $\delta_{\ell,i,m}$ of HBO(13) (3.4.30) that satisfy conditions (3.4.34), with their corresponding Butcher form coefficients satisfying order conditions (3.3.2) to (3.3.15), will produce a feasible CP coefficient, c_{feasible}, defined in Theorem 3.4.1 and a *feasible* HBO(13) in Shu–Osher form (3.4.30). What we really want is not merely a feasible method HBO(13) in Shu–Osher form but the best HBO(13) that has the largest possible CP coefficient.

3.5 Butcher form and modified Shu-Osher form in compact vector notation

In this section, we recall that Gottlieb, Ketcheson and Shu presented canonical Shu–Osher forms in compact vector notation for RK methods (see [7, Sections 3.1–3.4] for details). Our construction of canonical Shu–Osher form of HBO(13) proceeds in this section and in Section 3.6.

A. *Vector notation*

Vector and matrix notations help to represent a HBO method in canonical Shu–Osher form. Let us denote:

- $\boldsymbol{v}_0, \boldsymbol{w}_0, \boldsymbol{A}_{\mathrm{BU}}$ and $\boldsymbol{B}_{\mathrm{BU}}$, the vectors of $\mathbb{R}^7$ given by

$$\boldsymbol{v}_0 = [0, v_{\mathrm{BU},2}, v_{\mathrm{BU},3}, \ldots, v_{\mathrm{BU},7}]^T, \quad \boldsymbol{w}_0 = [0, a_{21}, a_{31}, \ldots, a_{61}, b_1]^T,$$

$$\boldsymbol{A}_{\mathrm{BU}} = [0, A_{\mathrm{BU},2}, A_{\mathrm{BU},3}, \ldots, A_{\mathrm{BU},7}]^T \text{ and } \boldsymbol{B}_{\mathrm{BU}} = [0, B_{\mathrm{BU},2}, B_{\mathrm{BU},3}, \ldots, B_{\mathrm{BU},7}]^T.$$

- $\boldsymbol{\gamma}_0$ and $\boldsymbol{\gamma}_1$, the rectangular (7×5)-matrices, defined in (3.2.27).
- $\boldsymbol{\beta}_0 \in \mathbb{R}^{7\times 7}$, the strictly lower triangular matrix

$$\boldsymbol{\beta}_0 = \begin{bmatrix} 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ a_{21} & 0 & 0 & 0 & \cdots & 0 & 0 \\ a_{31} & a_{32} & 0 & 0 & \cdots & 0 & 0 \\ a_{41} & a_{42} & a_{43} & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \ddots & \ddots & & \vdots \\ a_{61} & a_{62} & a_{63} & \cdots & a_{65} & 0 & 0 \\ b_1 & b_2 & b_3 & \cdots & b_5 & b_6 & 0 \end{bmatrix}, \tag{3.5.1}$$

whose coefficients come from the Butcher form (3.4.1) and (3.4.2) of HBO(13). Moreover, set also the $(7 \times N)$-matrices

$$\boldsymbol{Y} = [0, Y_2, \ldots, Y_7]^T, \text{and } \boldsymbol{F} = [0, F_2, \ldots, F_7]^T, \tag{3.5.2}$$

and the $(5 \times N)$-matrices

$$\begin{aligned}\boldsymbol{\varphi}_n &= [(\Delta t)^2 y_n^{(2)}, (\Delta t)^3 y_n^{(3)}, \dots, (\Delta t)^6 y_n^{(6)}]^T, \text{and} \\ \boldsymbol{\varphi}_{n-1} &= [(\Delta t)^2 y_{n-1}^{(2)}, (\Delta t)^3 y_{n-1}^{(3)}, \dots, (\Delta t)^6 y_{n-1}^{(6)}]^T,\end{aligned} \tag{3.5.3}$$

defined by the N-vectors: Y_j, F_j for $j = 2, \dots, 7$, y_j, f_j for $j = n-1, n$, $Y_1 = y_n$, $F_1 = f_n$, $Y_7 = y_{n+1}$ and $F_7 = f_{n+1}$.

B. *Butcher form in compact vector notation*

With the above notation, the Butcher form (3.4.1) and (3.4.2) of HBO(13) can be written in vector notation as:

$$\begin{aligned}\boldsymbol{Y} &= \boldsymbol{v}_0 y_n^T + \Delta t \boldsymbol{w}_0 f_n^T + \boldsymbol{\gamma}_0 \boldsymbol{\varphi}_n \\ &\quad + \boldsymbol{A}_{\text{BU}} y_{n-1}^T + \Delta t \boldsymbol{B}_{\text{BU}} f_{n-1}^T + \boldsymbol{\gamma}_1 \boldsymbol{\varphi}_{n-1} + \Delta t \boldsymbol{\beta}_0 \boldsymbol{F}, \\ y_{n+1} &= Y_7,\end{aligned} \tag{3.5.4}$$

with the consistency conditions:

$$\boldsymbol{v}_0 + \boldsymbol{A}_{\text{BU}} = \boldsymbol{e}_7, \tag{3.5.5}$$

where $\boldsymbol{e}_7$ is the 7-vector

$$\boldsymbol{e}_7 = [0, 1, 1, \dots, 1]^T \in \mathbb{R}^7. \tag{3.5.6}$$

C. *Modified Shu-Osher form in compact vector notation*

Using the equations and notations (3.4.11) to (3.4.23), and (3.4.25) and (3.4.26), we set

- the vectors $\boldsymbol{v}, \boldsymbol{w}, \boldsymbol{A}_{\text{SO}}, \boldsymbol{B}_{\text{SO}} \in \mathbb{R}^7$

$$\boldsymbol{v} = [0, v_2, v_3, \dots, v_7]^T, \quad \boldsymbol{w} = [0, w_2, w_3, \dots, w_7]^T,$$

$$\boldsymbol{A}_{\text{SO}} = [0, A_2, A_3, \dots, A_7]^T \quad \text{and} \quad \boldsymbol{B}_{\text{SO}} = [0, B_2, B_3, \dots, B_7]^T.$$

- the strictly lower triangular (7×7)-matrices

$$\boldsymbol{\alpha} = (\alpha_{ij}) \quad \text{and} \quad \boldsymbol{\beta} = (\beta_{ij}).$$

- the rectangular (7×5)-matrices with zero first row,

$$\boldsymbol{\delta}_0 = (\delta_{0,i,m}) \quad \text{and} \quad \boldsymbol{\delta}_1 = (\delta_{1,i,m}).$$

where the components v_i, w_i, α_{ij}, β_{ij}, A_i, B_i, $\delta_{\ell,i,m}$ come from the modified Shu–Osher form (3.4.27) of HBO(13).
Using the equations (3.4.11) to (3.4.23), we have

$$\boldsymbol{v} = (\boldsymbol{I} - \boldsymbol{\alpha})\,\boldsymbol{v}_0, \quad \boldsymbol{w} = (\boldsymbol{I} - \boldsymbol{\alpha})\,\boldsymbol{w}_0, \quad \boldsymbol{\delta}_0 = (\boldsymbol{I} - \boldsymbol{\alpha})\,\boldsymbol{\gamma}_0, \tag{3.5.7}$$

$$\boldsymbol{A}_{\mathrm{SO}} = (\boldsymbol{I} - \boldsymbol{\alpha})\,\boldsymbol{A}_{\mathrm{BU}}, \quad \boldsymbol{B}_{\mathrm{SO}} = (\boldsymbol{I} - \boldsymbol{\alpha})\,\boldsymbol{B}_{\mathrm{BU}}, \tag{3.5.8}$$

$$\boldsymbol{\delta}_1 = (\boldsymbol{I} - \boldsymbol{\alpha})\,\boldsymbol{\gamma}_1, \quad \boldsymbol{\beta} = (\boldsymbol{I} - \boldsymbol{\alpha})\,\boldsymbol{\beta}_0. \tag{3.5.9}$$

Therefore, the modified Shu–Osher form (3.4.27) of HBO(13) can be written in vector notation as:

$$\begin{aligned} \boldsymbol{Y} &= \boldsymbol{v} y_n^T + \Delta t \boldsymbol{w} f_n^T + \boldsymbol{\delta}_0 \boldsymbol{\varphi}_n + \boldsymbol{A}_{\mathrm{SO}} y_{n-1}^T + \Delta t \boldsymbol{B}_{\mathrm{SO}} f_{n-1}^T + \boldsymbol{\delta}_1 \boldsymbol{\varphi}_{n-1} \\ &\quad + \boldsymbol{\alpha} \boldsymbol{Y} + \Delta t \boldsymbol{\beta} \boldsymbol{F}, \\ y_{n+1} &= Y_7, \end{aligned} \tag{3.5.10}$$

with the consistency conditions (3.4.24) in vector form,

$$\boldsymbol{v} + \boldsymbol{A}_{\mathrm{SO}} + \boldsymbol{\alpha} \boldsymbol{e}_7 = \boldsymbol{e}_7, \tag{3.5.11}$$

where $\boldsymbol{e}_7$, defined in (3.5.6). Note that, as the first row of matrices $\boldsymbol{Y}, \boldsymbol{F}$ are equal to zero, α_{i1} and β_{i1}, for $2 \le i \le 7$, are not used in formulae (3.5.10) and are replaced by v_i and w_i, for $2 \le i \le 7$, respectively.

3.6 Canonical Shu-Osher form of HBO(13) method written solely in terms of vectors and matrices of Butcher form for deriving the CP property

In this section, we consider a subclass of HBO(13) methods, for which we will optimize the CP coefficients in Section 3.7. This subclass is defined by the following extra

condition:
Condition (r): The matrices $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ of the modified Shu-Osher form of HBO(13) (see 3.5.10) are proportional, i.e. $\boldsymbol{\alpha} = r\boldsymbol{\beta}$, for $r > 0$.
Then by (3.5.9), we have: $\boldsymbol{\beta} = (\boldsymbol{I} - r\boldsymbol{\beta})\,\boldsymbol{\beta}_0$ and therefore

$$\boldsymbol{\beta}\,(\boldsymbol{I} + r\boldsymbol{\beta}_0) = \boldsymbol{\beta}_0. \tag{3.6.1}$$

As $\boldsymbol{\beta}_0$ is strictly lower triangular, $(\boldsymbol{I} + r\boldsymbol{\beta}_0)$ is invertible and

$$\boldsymbol{\beta} = \boldsymbol{\beta}_0\,(\boldsymbol{I} + r\boldsymbol{\beta}_0)^{-1} = (\boldsymbol{I} + r\boldsymbol{\beta}_0)^{-1}\,\boldsymbol{\beta}_0. \tag{3.6.2}$$

By *Condition* (r), we have:

$$\begin{aligned}(\boldsymbol{I} - \boldsymbol{\alpha})\,(\boldsymbol{I} + r\boldsymbol{\beta}_0) &= (\boldsymbol{I} - r\boldsymbol{\beta})\,(\boldsymbol{I} + r\boldsymbol{\beta}_0)\\ &= \boldsymbol{I} + r\boldsymbol{\beta}_0 - r\boldsymbol{\beta}\,(\boldsymbol{I} + r\boldsymbol{\beta}_0)\\ &= \boldsymbol{I} + r\boldsymbol{\beta}_0 - r\boldsymbol{\beta}_0\\ &= \boldsymbol{I}.\end{aligned}$$

Hence,

$$(\boldsymbol{I} - \boldsymbol{\alpha}) = (\boldsymbol{I} + r\boldsymbol{\beta}_0)^{-1}. \tag{3.6.3}$$

Using the equations (3.5.7) to (3.5.9) and (3.6.3), the modified Shu-Osher form of HBO(13) (3.5.10) becomes under *Condition* (r):

$$\begin{aligned}\boldsymbol{Y} = (\boldsymbol{I} + r\boldsymbol{\beta}_0)^{-1}\,\big[&\boldsymbol{v}_0 y_n^T + \Delta t\boldsymbol{w}_0 f_n^T + \boldsymbol{\gamma}_0\boldsymbol{\varphi}_n\\ &+ \boldsymbol{A}_{\text{BU}} y_{n-1}^T + \Delta t\boldsymbol{B}_{\text{BU}} f_{n-1}^T + \boldsymbol{\gamma}_1\boldsymbol{\varphi}_{n-1} + \boldsymbol{\beta}_0\,(r\boldsymbol{Y} + \Delta t\boldsymbol{F})\big],\end{aligned} \tag{3.6.4}$$

with the consistency relations

$$(\boldsymbol{I} + r\boldsymbol{\beta}_0)^{-1}\,\boldsymbol{v}_0 + (\boldsymbol{I} + r\boldsymbol{\beta}_0)^{-1}\,\boldsymbol{A}_{\text{BU}} + r\,(\boldsymbol{I} + r\boldsymbol{\beta}_0)^{-1}\,\boldsymbol{\beta}_0\boldsymbol{e}_7 = \boldsymbol{e}_7, \tag{3.6.5}$$

which is equivalent to the consistency conditions (3.5.5). The modified Shu-Osher form of HBO(13) (3.5.10) under *Condition* (r), shown in (3.6.4), is called *the Canonical Shu-Osher form of HBO(13) method written solely in terms of vectors and matrices of Butcher form.*

Let f be a function satisfying the assumptions of Theorem 2.1.1 and the condition (1.1.4) of the S(6) method. Let us denote the coefficients of the modified Shu-Osher form of HBO(13) (3.6.4) as

$$\boldsymbol{v}_r = (\boldsymbol{I} + r\boldsymbol{\beta}_0)^{-1}\,\boldsymbol{v}_0, \qquad \boldsymbol{w}_r = (\boldsymbol{I} + r\boldsymbol{\beta}_0)^{-1}\,\boldsymbol{w}_0, \tag{3.6.6}$$

$$\boldsymbol{\delta}_0, r = (\boldsymbol{I} + r\boldsymbol{\beta}_0)^{-1}\,\boldsymbol{\gamma}_0, \qquad \boldsymbol{\delta}_1, r = (\boldsymbol{I} + r\boldsymbol{\beta}_0)^{-1}\,\boldsymbol{\gamma}_1, \tag{3.6.7}$$

$$\boldsymbol{A}_{\mathrm{SO}}, r = (\boldsymbol{I} + r\boldsymbol{\beta}_0)^{-1}\,\boldsymbol{A}_{\mathrm{BU}}, \qquad \boldsymbol{B}_{\mathrm{SO}}, r = (\boldsymbol{I} + r\boldsymbol{\beta}_0)^{-1}\,\boldsymbol{B}_{\mathrm{BU}}, \tag{3.6.8}$$

$$\boldsymbol{\beta}_r = (\boldsymbol{I} + r\boldsymbol{\beta}_0)^{-1}\,\boldsymbol{\beta}_0. \tag{3.6.9}$$

Under *Condition* (r), Theorem 3.4.1 can be rewritten as follows:

Theorem 3.6.1. *Let f be as above. Then the 2-step, 6-stage, 6-derivatives HBO(13) method (3.6.4) satisfies the CP property*

$$\|y_{n+1} - \tilde{y}_{n+1}\| \leq \max_{0\leq j\leq 1} \|y_{n-j} - \tilde{y}_{n-j}\|, \ \textit{for} \ \ \Delta t \leq r\Delta t_{S(6)},$$

if the following conditions are satisfied:

- *the coefficients (3.6.6) to (3.6.9) are component-wise positive (see Remark 3.6.1),*
- $$r \leq r_{i0} = \frac{v_i}{w_i}, \quad r \leq r_{i1} = \frac{A_i}{B_i}, \ \textit{for} \ \ 2 \leq i \leq 7, \tag{3.6.10}$$
- $$\delta_{0,i,m} \leq \frac{v_i}{r_{i0}^m\ m!}, \quad \delta_{1,i,m} \leq \frac{A_i}{r_{i1}^m\ m!}, \ \textit{for} \ \ 2 \leq i \leq 7 \ \ \textit{and} \ \ 2 \leq m \leq 6, \tag{3.6.11}$$

with the convention that $a/0 = +\infty$.

Remark 3.6.1. *We noticed from the numerical computations that we can have a feasible HBO(13) method with v_i, w_i are both 0, for $2 \leq i \leq 7$ or A_i, B_i are both 0, for $2 \leq i \leq 7$.*

3.7 Formulation of the optimization problem to obtain the optimal HBO(13)

Theorem 3.6.1 indicates sufficient conditions on the parameter r of *Condition* (r) should satisfy for the 2-step, 6-stage, 6-derivatives HBO(13) method (3.6.4) to be contractivity-preserving. Hence each representation $(\boldsymbol{v}_r, \boldsymbol{w}_r, \boldsymbol{\delta}_{0,r}, \boldsymbol{A}_{\mathrm{SO},\boldsymbol{r}}, \boldsymbol{B}_{\mathrm{SO},\boldsymbol{r}}, \boldsymbol{\delta}_{1,r}, \boldsymbol{\alpha}_r, \boldsymbol{\beta}_r)$, of HBO(13) (3.6.4), produces a feasible CP coefficient r if the following conditions are satisfied:

$$\boldsymbol{v}_r = (\boldsymbol{I} + r\boldsymbol{\beta}_0)^{-1}\,\boldsymbol{v}_0 \geq 0 \tag{3.7.1}$$
$$\boldsymbol{w}_r = (\boldsymbol{I} + r\boldsymbol{\beta}_0)^{-1}\,\boldsymbol{w}_0 \geq 0 \tag{3.7.2}$$
$$\boldsymbol{\delta}_{0,r} = (\boldsymbol{I} + r\boldsymbol{\beta}_0)^{-1}\,\boldsymbol{\gamma}_0 \geq 0 \tag{3.7.3}$$
$$\boldsymbol{\delta}_{1,r} = (\boldsymbol{I} + r\boldsymbol{\beta}_0)^{-1}\,\boldsymbol{\gamma}_1 \geq 0 \tag{3.7.4}$$
$$\boldsymbol{A}_{\mathrm{SO},\boldsymbol{r}} = (\boldsymbol{I} + r\boldsymbol{\beta}_0)^{-1}\,\boldsymbol{A}_{\mathrm{BU}} \geq 0 \tag{3.7.5}$$
$$\boldsymbol{B}_{\mathrm{SO},\boldsymbol{r}} = (\boldsymbol{I} + r\boldsymbol{\beta}_0)^{-1}\,\boldsymbol{B}_{\mathrm{BU}} \geq 0 \tag{3.7.6}$$
$$\boldsymbol{\beta}_r = \boldsymbol{\beta}_0\,(\boldsymbol{I} + r\boldsymbol{\beta}_0)^{-1} \geq 0 \tag{3.7.7}$$

together with

- conditions (3.6.10), (3.6.11), and
- order conditions (3.3.2)–(3.3.15) for order 13.

Let c(HBO(13)) denotes the supremum of the CP coefficients r and consider the following definition

Definition 3.7.1. *The effective CP coefficient of CP HBO(13) method is denoted by*

$$c_{\mathit{eff}}(HBO(13)) = \frac{c(HBO(13))}{l}, \tag{3.7.8}$$

where l is the number of function evaluations of HBO(13) method per time step and $c(HBO(13))$, defined above.

Gottlieb et al [8] pointed out that the coefficients c_{eff} provide a fair comparison between methods of the same order. Since HBO(13) method contains many free parameters, the Matlab Optimization Toolbox was used to search for the methods with largest c(HBO(13)) under the tolerance 10^{-12} on the objective function c(HBO(13)) provided all the constraints are satisfied to the tolerance 8×10^{-14}.

We have not found any CP general linear 2-step multistage methods of order 13 with nonnegative coefficients in the literature. Our study indicates that HBO(13) exists and has fairly good CP coefficients. Our best method of order 13 has

$c(\text{HBO}(13)) = 0.540$.

The formula of the new HBO(13) is listed in Appendix B with its c(HBO(13)), c_{eff}(HBO(13)) and abscissa vector.

3.8 Region of absolute stability of HBO(13)

To obtain the region of absolute stability, $\mathcal{R}$, of HBO(13), we apply the predictor P_i, for $2 \le i \le 6$ (3.4.1) and the integration formula (3.4.2) with constant step $h = \Delta t$ to the linear test equation

$$y' = \lambda y, \qquad y_0 = 1.$$

Thus we obtain

$$Y_i = v_{\text{BU,i}} y_n + \widehat{h} \sum_{j=1}^{i-1} a_{ij} Y_j + \sum_{m=2}^{6} (\widehat{h})^m \gamma_{0,i,m} y_n + A_{\text{BU},i} y_{n-1} + \widehat{h} B_{\text{BU},i} y_{n-1} + \sum_{m=2}^{6} (\widehat{h})^m \gamma_{1,i,m} y_{n-1}, \qquad i = 2, 3, \dots, 6, \quad (3.8.1)$$

and

$$y_{n+1} = v_{\text{BU},7} y_n + \widehat{h} \left[\sum_{j=1}^{6} b_j Y_j \right] + \sum_{m=2}^{6} (\widehat{h})^m \gamma_{0,7,m} y_n + A_{\text{BU},7} y_{n-1} + \widehat{h} B_{\text{BU},7} y_{n-1} + \sum_{m=2}^{6} (\widehat{h})^m \gamma_{1,7,m} y_{n-1}, \quad (3.8.2)$$

where $\widehat{h} = \lambda\Delta t$. We can see that Y_i in (3.8.1) is expressed only in terms of $y_{n-\ell}$, $\ell = 0, 1$, since $c_1 = 0$. Then, y_{n+1} in (3.8.2) is expressed only in terms of $y_{n-\ell}$, $\ell = 0, 1$; thus we obtain the following second-order difference equation and associated linear characteristic equation:

$$\sum_{j=0}^{2} C_j y_{n+(j-1)} = 0, \qquad \sum_{j=0}^{2} C_j r^j = 0. \tag{3.8.3}$$

The coefficients C_j in the above characteristic equation can be found by symbolic computation. A complex number $\widehat{h}$ is in $\mathcal{R}$ if the two roots of the characteristic equation satisfy the root condition which is defined in Definition 2.4.1. The scanning method used to find $\mathcal{R}$ is similar to the one used for Runge–Kutta methods (see [16, pp. 70 and 204]). The grey region in Fig. 3 depicts $\mathcal{R}$ for HBO(13) with the unscaled interval of stability $(-2.79, 0)$. As in Subsections 2.5.1 and 2.5.2, the grey region in Fig. 3 is symmetric about the real axis, and Fig. 3 shows only the grey region in the half-plane Im$(\widehat{h})> 0$. It is seen that HBO(13) has a larger scaled interval of absolute stability $(-0.253, 0)$ than the Adams–Bashforth–Moulton method of order 13 $(-0.03, 0)$ in PECE mode [25, pp. 139–140].

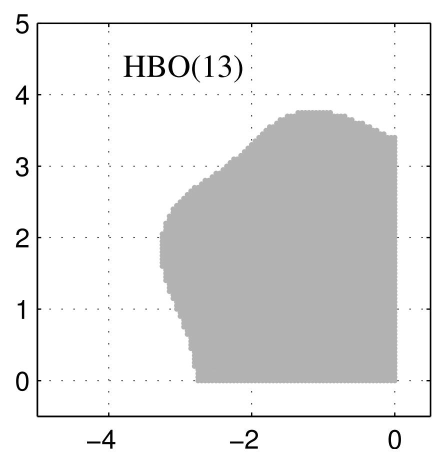


Figure 3: The grey region depicts the unscaled region of absolute stability of HBO(13) with the unscaled stability interval $(-2.79, 0)$.

Chapter 4

Numerical Results

4.1 Implementation and problems used for comparison

We compare the numerical performances of HBO(13) and ABM(13) on each of the test problems listed below. We give in Appendix A, a short description of this list of problems.

- **B1** The growth of two conflicting populations.
- **B3** A nonlinear chemical reaction.
- **B5** Euler equations of motion for a rigid body without external forces.
- **D1** Kepler's two body problem with eccentricity $\epsilon = 0.1$.
- **D2** As in D1 except with eccentricity $\epsilon = 0.3$.
- **D3** As in D1 except with eccentricity $\epsilon = 0.5$.
- **D4** As in D1 except with eccentricity $\epsilon = 0.7$.
- **D5** As in D1 except with eccentricity $\epsilon = 0.9$.
- **E2** Derived from Van der Pol's equation with $\epsilon = 1$.

- Hénon–Heiles's problem [10].
- A problem in Galactic dynamics [3].

The starting value for HBO(13) is calculated by MATLAB's `ode113` under the stringent tolerance 5×10^{-14}. The higher order derivatives y' to $y^{(6)}$ of Taylor series are calculated at each integration step by known recurrence formulae (see, for example, [9, pp. 46–49], [17, 1, 2]). Deprit and Zahar [5] showed that such recursive computation is very effective in achieving high accuracy, even with little computing time and large step sizes. Our comparison of the performances of HBO(13) and ABM(13) proceeds in four steps. **Firstly**, we implement both HBO(13) and ABM(13) in MATLAB. **Secondly**, we collect the CPU time, the global error at the endpoint of integration and the relative energy error. **Thirdly**, we compute the CPU percentage efficiency gain (here, the CPU time estimates are obtained by using MATLAB's `polyfit`). **Lastly**, we draw figures and tables according to the data collected through previous steps and compare the results.
Computations were performed in MATLAB Version 8.3.0.532 (R2014a) on a Mac Laptop with the following characteristics:

☐ Processor: 2.6GHz quad-core Intel Core i7.

☐ Memory: 8GB 1600MHz DDR3.

☐ Operating system: Mac OS X Version 10.9.5.

4.2 CPU time of HBO(13) and ABM(13)

Firstly, we compute the global error at the endpoint of integration interval as a function of the CPU time of HBO(13) and ABM(13). If $v = [v_1, v_2, \dots, v_N]^T \in \mathbb{R}^N$, recall that $||v||_\infty = \max\{|v_j|; 1 \leq j \leq N\}$. The global error at the endpoint t_f of integration interval, denoted by GE, is taken in the infinity norm,

$$\text{GE} = ||y_f - z_f||_\infty,$$

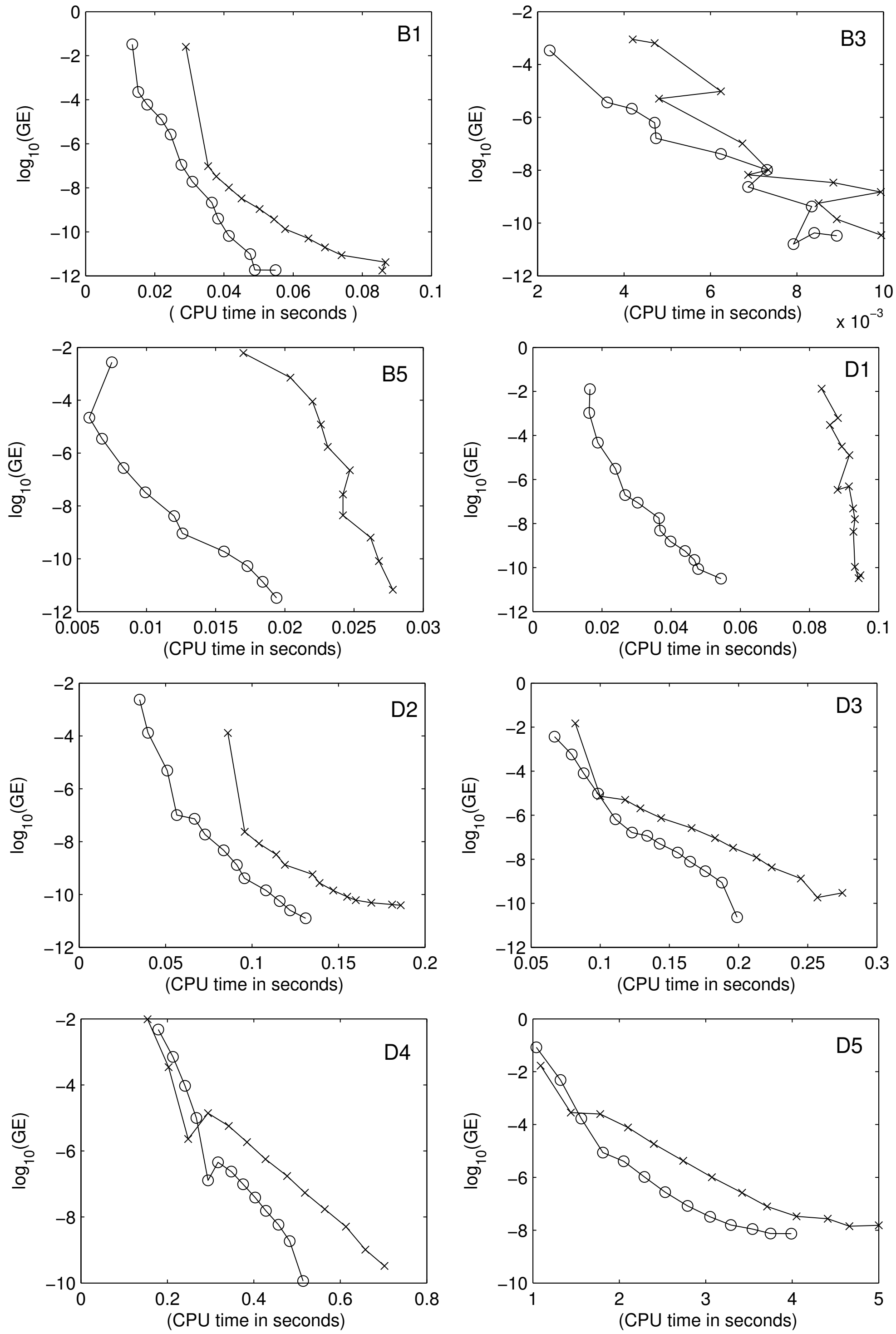
where y_f is the numerical value obtained by HBO(13) and z_f is the "exact solution" obtained by MATLAB's `ode113` with stringent tolerance 5×10^{-14}. The numerical tests have been repeated 20 times to obtain a better estimation of the CPU time data. In Fig. 4 and Fig. 5, the performance of HBO(13) and ABM(13) is compared on the eleven problems of Section 4.1, on the basis of $\log_{10}$(GE) (vertical axis) as a function of CPU time in seconds (horizontal axis). It is seen, from Fig. 4 and Fig. 5, that HBO(13) compares favorably with ABM(13) at stringent tolerance, on the basis of $\log_{10}$(GE) (vertical axis) as a function of CPU time (horizontal axis) for these eleven problems (listed in Section 4.1).

The *CPU percentage efficiency gain* (CPU PEG) is defined by the formula (cf. Sharp [26]),

$$(\text{CPU PEG})_i = 100 \left[\frac{\sum_j \text{CPU}_{2,ij}}{\sum_j \text{CPU}_{1,ij}} - 1 \right], \tag{4.2.1}$$

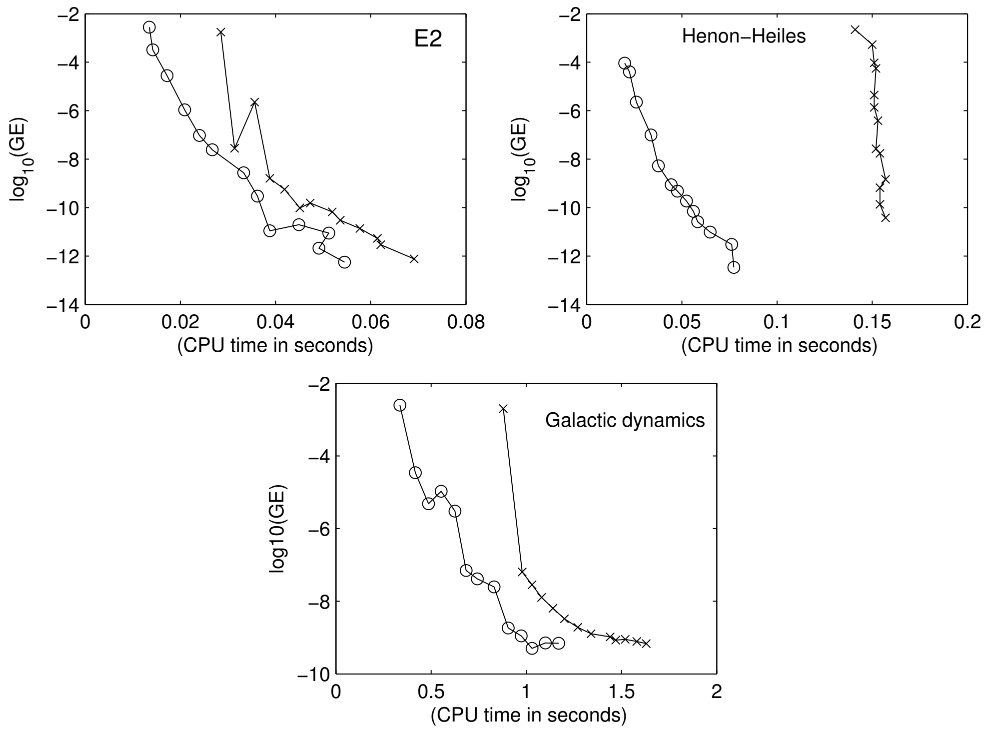
where $\text{CPU}_{1,ij}$ and $\text{CPU}_{2,ij}$ are the estimates of CPU time of methods 1 and 2, respectively, associated with problem i, and $j = -\log_{10}(\text{GE estimate})$. To compute $\text{CPU}_{2,j}$ and $\text{CPU}_{1,j}$ appearing in (4.2.1), we approximate the data $(\log_{10}(\text{GE}), \log_{10}(\text{CPU}))$ in a least-squares sense by MATLAB's `polyfit`. Then, for chosen integer values of the summation index j, we take $-\log_{10}(\text{GE estimate}) = j$ and obtain $\log_{10}$(CPU estimate) from the approximating curve, and finally the estimate of CPU time.

The CPU PEG listed in Table 1 shows that HBO(13) wins on the basis of $\log_{10}$(GE) as a function of CPU time for the listed problems on the time interval $[0, t_f]$.



HBO(13) ○ and ABM(13) ×

Figure 4: Log_{10} (GE) (vertical axis) as a function of CPU time (horizontal axis) for the problems on hand.



HBO(13) ∘ and ABM(13) ×

Figure 5: Log_{10} (GE) (vertical axis) as a function of CPU time (horizontal axis) for the problems on hand.

Problem	CPU PEG of HBO(13) over ABM(13)
B1 $(t_f = 20)$	50 %
B3 $(t_f = 20)$	27 %
B5 $(t_f = 20)$	118 %
D1 $(t_f = 16\pi)$	190 %
D2 $(t_f = 16\pi)$	52 %
D3 $(t_f = 16\pi)$	31 %
D4 $(t_f = 16\pi)$	27 %
D5 $(t_f = 16\pi)$	23 %
E2 $(t_f = 20)$	41 %
Hénon–Heiles $(t_f = 70)$	275 %
Galactic dynamics $(t_f = 500)$	62 %

Table 1: CPU PEG of HBO(13) over ABM(13) for the listed problems.

4.3 CPU time of HBO(13) and ABM(13) after a 350-orbit integration of Kepler's two-body problem

Secondly, we compute the relative energy error (EE(t)) as a function of CPU time of HBO(13) and ABM(13) after a 350-orbit integration of a Hamiltonian system as in [11]. The Hamiltonian for Kepler's two-body problem is

$$H_{\text{Kepler}} = \frac{1}{2}(y_3^2 + y_4^2) - (y_1^2 + y_2^2)^{-1/2}.$$

The relative energy error (EE(t)) at time t is defined as

$$\text{EE}(t) = \left| \frac{\text{E}(t) - \text{E}(0)}{\text{E}(0)} \right|,$$

where E(t) is the energy at time t. For this comparison, we used Kepler's two-body problem with eccentricities of 0.3, 0.5, 0.7 and 0.9 and an interval of integration of $[0, 700\pi]$. Fig. 6 is the graph of $\log_{10}$(EE) as a function of CPU time in seconds for HBO(13) and ABM(13) after a 350-orbit integration of Kepler's two-body problem with eccentricities of 0.3, 0.5, 0.7 and 0.9. It is seen, from Fig. 6, that HBO(13) compares favorably with ABM(13).

The CPU PEGs of HBO(13) over ABM(13) after a 350-orbit integration of Kepler's two-body problem with eccentricities of 0.3, 0.5, 0.7 and 0.9 respectively are listed in Table 2. From Table 2, we see that HBO(13) wins on the basis of the relative energy error (EE) and CPU time.

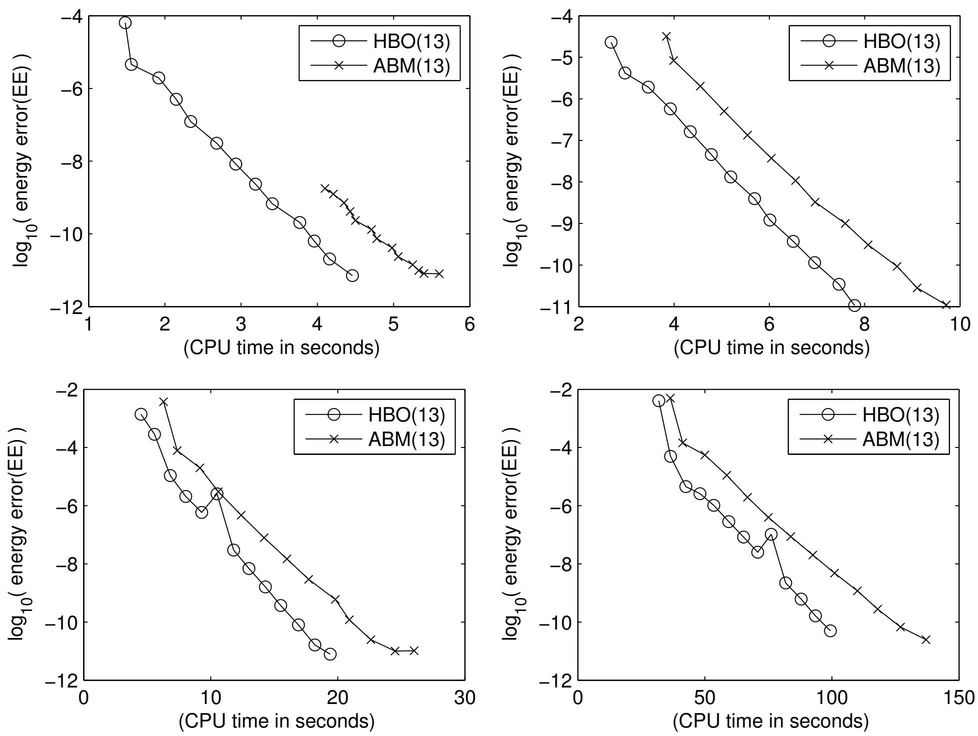


Figure 6: Log_{10} (EE) (vertical axis) as a function of CPU time (horizontal axis) on Kepler's two-body problem with $e = 0.3$ (*Top left*), $e = 0.5$ (*Top right*), $e = 0.7$ (*Bottom left*) and $e = 0.9$ (*Bottom right*) respectively. The interval of integration is $[0, 700\pi]$.

CPU PEG for Kepler's two-body problem with:	
$e = 0.3$	32%
$e = 0.5$	25 %
$e = 0.7$	28 %
$e = 0.9$	27 %

Table 2: CPU PEG of HBO(13) over ABM(13) on a 350-orbit integration of Kepler's two-body problem with $e = 0.3$, $e = 0.5$, $e = 0.7$ and $e = 0.9$ respectively.

4.4 Relative energy error of HBO(13) and ABM(13) on a 10000-orbit integration of Kepler's two-body problem

In our last comparison, we compare the relative energy error (EE(t)) on a 10000-orbit integration of Kepler's two-body problem with eccentricities of $e = 0.5$, $e = 0.7$ and $e = 0.9$. Fig. 7 gives the graph of relative energy error EE(t) as a function of time t for $e = 0.5$, $e = 0.7$ and $e = 0.9$ over an interval of 10000 periods. Fig. 7 shows that the *contractivity-preserving property* reduces the growth of error. Fig. 8 gives the smoothed graph of logarithmic scaled EE(t) as a function of logarithmic scaled t for $e = 0.5$, $e = 0.7$ and $e = 0.9$ over an interval of 10000 periods. The smoothing removed the small amplitude high frequency oscillations in the original data and was done by using the MATLAB's `filter` command with the window size of 20. The initial stepsize was chosen so that ABM(13) and HBO(13) used the same CPU time.

For $e = 0.5$, $e = 0.7$ and $e = 0.9$, the relative energy error of ABM(13) is about 25, 25 and 34 times the relative energy error of HBO(13), respectively, across the interval of integration. These results are consistent with the CPU PEGs listed in Table 2 for two-body problem with $e = 0.5$, $e = 0.7$ and $e = 0.9$.
Moreover, we used linear least-squares to fit the power law $C_1 t^{C_2}$ to the graphs of Fig 8. The values of C_2 are listed in Table 3 and are in good agreement with the expected asymptotic value of one for non-symplectic methods.

	C_2 of $C_1 t^{C_2}$ for Kepler's two-body problem with:		
Method	$e = 0.5$	$e = 0.7$	$e = 0.9$
HBO(13)	1.017	1.028	1.040
ABM(13)	0.995	0.994	0.993

Table 3: Exponent C_2 of power law $C_1 t^{C_2}$ fitted to the error graphs of Fig. 8 for Kepler's two-body problem with $e = 0.5$, $e = 0.7$ and $e = 0.9$ respectively.

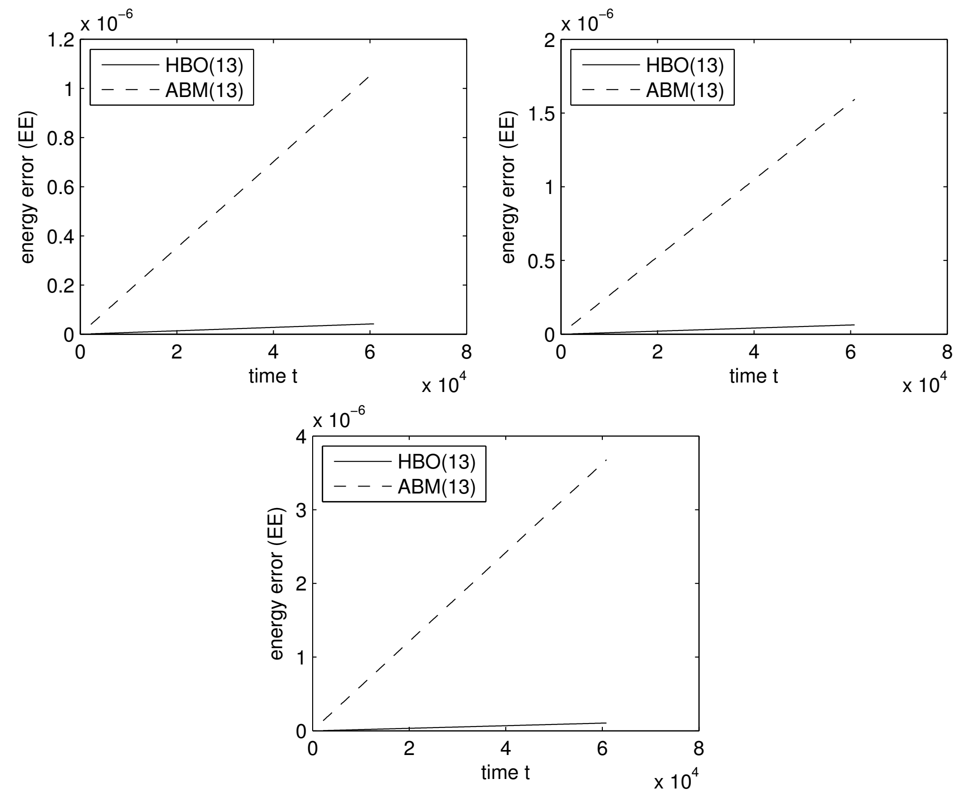


Figure 7: Growth of relative energy error (EE) (vertical axis) as a function of t (horizontal axis) on Kepler's two-body problem with $e = 0.5$ (*Top left*), $e = 0.7$ (*Top right*) and $e = 0.9$ (*Bottom*) respectively. The interval of integration is $[0, 20000\pi]$

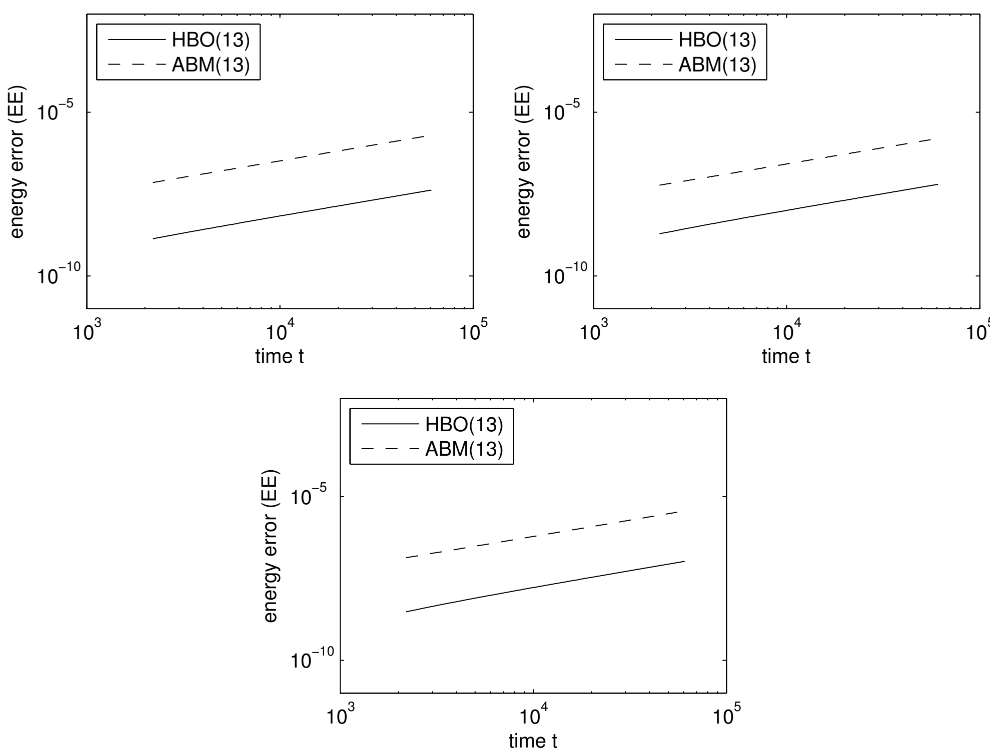


Figure 8: Growth of logarithmic scaled relative energy error (EE) (vertical axis) as a function of logarithmic scaled t (horizontal axis) on Kepler's two-body problem with $e = 0.5$ (*Top left*), $e = 0.7$ (*Top right*) and $e = 0.9$ (*Bottom*) respectively. The interval of integration is $[0, 20000\pi]$.

Chapter 5

Conclusion

A new contractivity-preserving (CP) explicit, 2-step, 6-stage, 6-derivative Hermite–Birkhoff–Obrechkoff method, HBO(13), of order 13 with nonnegative coefficients was constructed for solving nonstiff first-order initial value problems $y' = f(t, y)$, $y(t_0) = y_0$, where f satisfies some smoothness conditions (see Section 2.1). This new method is the combination of a CP, 2-step, 6-derivative, Hermite–Obrechkoff method of order 9 and a 6-stage Runge–Kutta method of order 5. The new HBO(13) method has order 13. This new method was tested on the following nonstiff DETEST problems [13]:

- **B1** The growth of two conflicting populations.
- **B3** A nonlinear chemical reaction.
- **B5** Euler equations of motion for a rigid body without external forces.
- **D1** Kepler's two body problem with eccentricity $\epsilon = 0.1$.
- **D2** As in D1 except with eccentricity $\epsilon = 0.3$.
- **D3** As in D1 except with eccentricity $\epsilon = 0.5$.
- **D4** As in D1 except with eccentricity $\epsilon = 0.7$.
- **D5** As in D1 except with eccentricity $\epsilon = 0.9$.

- **E2** Derived from Van der Pol's equation with $\epsilon = 1$.

and on the following two problems:

- Hénon–Heiles's problem [10].
- A problem in Galactic dynamics [3].

The new method HBO(13) has lower global error at the endpoint of integration interval and uses less CPU time than Adams–Bashforth–Moulton method of order 13 in PECE mode, denoted by ABM(13), in solving the eleven problems listed above. Also, HBO(13) was compared with ABM(13) after a 350-orbit integration of Kepler's two-body problem D2, D3, D4 and D5. We showed that HBO(13) performed better than ABM(13) on the basis of the CPU time and the relative energy error. Furthermore, we showed that HBO(13) is better than ABM(13) on Kepler's two-body problem D3, D4 and D5 on the basis of the relative energy error as a function of 10000 periods integration time.

Appendices

Appendix A

TEST problems

In 1979, T. E. Hull et al. [13] presented a bank of test problems called DETEST for comparing different numerical solvers of ODE's. Since then, DETEST problems have been used by several authors (see for example [13, 14, 20, 21]). In this thesis, we utilize 9 of the DETEST problems and we describe them briefly below [13]:

1. **B1:** The growth of two conflicting populations:

$$y_1' = 2(y_1 - y_1 y_2), \qquad y_1(0) = 1,$$

$$y_2' = -(y_2 - y_1 y_2), \qquad y_2(0) = 3.$$

2. **B3:** A nonlinear chemical reaction:

$$y_1' = -y_1, \qquad y_1(0) = 1,$$

$$y_2' = y_1 - y_2^2, \qquad y_2(0) = 0,$$

$$y_3' = y_2^2, \qquad y_3(0) = 0.$$

3. **B5:** Euler equations of motion for a rigid body without external forces:

$$y_1' = y_2\ y_3, \qquad y_1(0) = 0,$$

$$y_2' = -y_1\ y_3, \qquad y_2(0) = 1,$$

$$y_3' = 0.51\ y_1\ y_2, \qquad y_3(0) = 1.$$

4. **D1:** Kepler's two body problem:

$$y_1' = y_3, \qquad y_1(0) = 1 - \varepsilon,$$

$$y_2' = y_4, \qquad y_2(0) = 0,$$

$$y_3' = \frac{-y_1}{(y_1^2 + y_2^2)^{\frac{3}{2}}}, \qquad y_3(0) = 0,$$

$$y_4' = \frac{-y_2}{(y_1^2 + y_2^2)^{\frac{3}{2}}}, \qquad y_4(0) = \sqrt{\frac{1+\varepsilon}{1-\varepsilon}},$$

where $\varepsilon = 0.1$ is the eccentricity of the orbit.

5. **D2:** As in D1 except with eccentricity $\varepsilon = 0.3$.

6. **D3:** As in D1 except with eccentricity $\varepsilon = 0.5$.

7. **D4:** As in D1 except with eccentricity $\varepsilon = 0.7$.

8. **D5:** As in D1 except with eccentricity $\varepsilon = 0.9$.

All **D1**, **D2**, **D3**, **D4** and **D5** problems are derived from the orbit equations:

$$x'' = \frac{-x}{r^3}, \qquad x(0) = 1 - \varepsilon, \qquad x'(0) = 0,$$

$$y'' = \frac{-y}{r^3}, \qquad y(0) = 0, \qquad y'(0) = \sqrt{\frac{1+\varepsilon}{1-\varepsilon}},$$

$$r^2 = x^2 + y^2,$$

with solution

$$x = \cos u - \varepsilon, \qquad x' = \frac{-\sin u}{1 - \varepsilon \cos u},$$

$$y = \sqrt{1-\varepsilon^2} \sin u, \qquad y' = \frac{\sqrt{1-\varepsilon^2} \cos u}{1 - \varepsilon \cos u},$$

where $u - \varepsilon \sin u - t = 0$.

9. **E2:**

$$y_1' = y_2, \qquad y_1(0) = 2,$$

$$y_2' = (1 - y_1^2)y_2 - y_1, \qquad y_2(0) = 0.$$

This is an ODE system derived from Van der Pol's equation:

$$y'' - (1 - y^2)y' + y = 0.$$

Also, we campare HBO(13) to ABM(13) by testing them on the following two problems:

1. **Hénon–Heiles's problem:**

$$y_1' = y_3, \qquad y_1(0) = 0,$$

$$y_2' = y_4, \qquad y_2(0) = 0.20,$$

$$y_3' = -y_1 - 2\ y_1\ y_2, \qquad y_3(0) = 0.42,$$

$$y_4' = -y_2 - y_1^2 + y_2^2, \qquad y_4(0) = 0.20,$$

more details about this problem can be found in [10].

2. **A problem in Galactic dynamics:**

$$y_1' = y_4 + 0.25\, y_2, \qquad y_1(0) = 2.5,$$

$$y_2' = y_5 - 0.25\, y_1, \qquad y_2(0) = 0,$$

$$y_3' = y_6, \qquad y_3(0) = 0,$$

$$y_4' = 0.25\ y_5 - \frac{2\ y_1}{(1.25)^2 \left(\frac{y_1^2}{(1.25)^2} + y_2^2 + \frac{y_3^2}{(0.75)^2} \right)}, \qquad y_4(0) = 0,$$

$$y_5' = 0.25\, y_4 - \frac{2\ y_2}{\left(\frac{y_1^2}{(1.25)^2} + y_2^2 + \frac{y_3^2}{(0.75)^2} \right)}, \qquad y_5(0) = 1.68,$$

$$y_6' = -\frac{2\ y_3}{(0.75)^2 \left(\frac{y_1^2}{(1.25)^2} + y_2^2 + \frac{y_3^2}{(0.75)^2} \right)}, \qquad y_6(0) = 0.20,$$

more details about this problem can be found in [3].

Appendix B

The formula of the new HBO(13)

The appendix lists the canonical Shu–Osher form of the new HBO(13) method with its $c(\text{HBO}(13))$, $c_{\text{eff}}(\text{HBO}(13))$, abscissa vector σ and unscaled stability interval.

HBO(13) $c(\text{HBO}(13)) = 0.54051153051649181$, $c_{\text{eff}}(\text{HBO}(13)) = 0.0491$, $\sigma = [0, 0.71555502095573598, 0.74591974875879885, 0.82960696020132418, 0.93986892931739696, 0.68209546216033279]^T$ and unscaled stability interval $(-2.79, 0)$.

$$
\begin{aligned}
Y_2 &= 6.6070574215939470\,\text{e-}01 y_{n-1} + 3.3929425784060535\,\text{e-}01 y_n \\
&+ 7.4853269685835921\,\text{e-}01 \Delta t f_{n-1} + 6.2772806625677147\,\text{e-}01 \Delta t f_n \\
&+ 4.2401720048232616\,\text{e-}01 \Delta t^2 y_{n-1}^{(2)} + 2.5017211930381755\,\text{e-}01 \Delta t^2 y_n^{(2)} \\
&+ 1.4432681394236077\,\text{e-}01 \Delta t^3 y_{n-1}^{(3)} + 7.6604621420481925\,\text{e-}02 \Delta t^3 y_n^{(3)} \\
&+ 2.8735270366287514\,\text{e-}02 \Delta t^4 y_{n-1}^{(4)} + 1.1732463724989753\,\text{e-}02 \Delta t^4 y_n^{(4)} \\
&+ 3.1216853637924848\,\text{e-}03 \Delta t^5 y_{n-1}^{(5)} + 0.0 \Delta t^5 y_n^{(5)} \\
&+ 1.4125015828714865\,\text{e-}04 \Delta t^6 y_{n-1}^{(6)} + 5.0644552089026875\,\text{e-}04 \Delta t^6 y_n^{(6)} \\
Y_3 &= 5.0725452174769037\,\text{e-}01 y_{n-1} + 4.1725297510951104\,\text{e-}01 y_n \\
&+ 2.8752717032024577\,\text{e-}01 \Delta t f_{n-1} + 7.7195943389181787\,\text{e-}01 \Delta t f_n \\
&+ 5.2644020984180064\,\text{e-}02 \Delta t^2 y_{n-1}^{(2)} + 1.4018663967318618\,\text{e-}01 \Delta t^2 y_n^{(2)} \\
&+ 2.3488860620683224\,\text{e-}03 \Delta t^3 y_{n-1}^{(3)} + 1.9878841136844256\,\text{e-}02 \Delta t^3 y_n^{(3)}
\end{aligned}
$$

$$
\begin{aligned}
&+ 0.0\Delta t^4 y_{n-1}^{(4)} + 6.3582790269573856\,\text{e-}03\Delta t^4 y_n^{(4)}\\
&+ 1.2702937328522049\,\text{e-}04\Delta t^5 y_{n-1}^{(5)} + 0.0\Delta t^5 y_n^{(5)}\\
&+ 2.3367443959533051\,\text{e-}05\Delta t^6 y_{n-1}^{(6)} + 0.0\Delta t^6 y_n^{(6)}\\
&+ 7.5492503142798581\,\text{e-}02 Y_2 + 1.3966862662607935\,\text{e-}01\Delta t F_2.\\
Y_4 &= 6.5967243474242121\,\text{e-}01 y_{n-1} + 2.2656639463279032\,\text{e-}01 y_n\\
&+ 7.7478290355532309\,\text{e-}01\Delta t f_{n-1} + 4.1917032633196977\,\text{e-}01\Delta t f_n\\
&+ 3.1264654754690491\,\text{e-}01\Delta t^2 y_{n-1}^{(2)} + 2.8778251959722112\,\text{e-}01\Delta t^2 y_n^{(2)}\\
&+ 6.3941812403468079\,\text{e-}02\Delta t^3 y_{n-1}^{(3)} + 0.0\Delta t^3 y_n^{(3)}\\
&+ 7.1017470318959072\,\text{e-}03\Delta t^4 y_{n-1}^{(4)} + 5.8719690361103182\,\text{e-}03\Delta t^4 y_n^{(4)}\\
&+ 3.5996438430403006\,\text{e-}04\Delta t^5 y_{n-1}^{(5)} + 4.3428622677961292\,\text{e-}04\Delta t^5 y_n^{(5)}\\
&+ 0.0\Delta t^6 y_{n-1}^{(6)} + 0.0\Delta t^6 y_n^{(6)} + 0.0 Y_2 + 0.0\Delta t F_2\\
&+ 1.1376117062478852\,\text{e-}01 Y_3 + 2.1046946124550345\,\text{e-}01\Delta t F_3.\\
Y_5 &= 5.8102799690388296\,\text{e-}01 y_{n-1} + 2.7356614581491429\,\text{e-}01 y_n\\
&+ 6.2512740861375393\,\text{e-}01\Delta t f_{n-1} + 5.0612453272459346\,\text{e-}01\Delta t f_n\\
&+ 2.2321916144997420\,\text{e-}01\Delta t^2 y_{n-1}^{(2)} + 2.7985648342544800\,\text{e-}01\Delta t^2 y_n^{(2)}\\
&+ 3.9454464482277268\,\text{e-}02\Delta t^3 y_{n-1}^{(3)} + 0.0\Delta t^3 y_n^{(3)}\\
&+ 3.6738799318486535\,\text{e-}03\Delta t^4 y_{n-1}^{(4)} + 8.1924449985334817\,\text{e-}03\Delta t^4 y_n^{(4)}\\
&+ 1.4903338638699046\,\text{e-}04\Delta t^5 y_{n-1}^{(5)} + 1.2164623626872834\,\text{e-}04\Delta t^5 y_n^{(5)}\\
&+ 0.0\Delta t^6 y_{n-1}^{(6)} + 0.0\Delta t^6 y_n^{(6)} + 0.0 Y_2 + 0.0\Delta t F_2 + 0.0 Y_3 + 0.0\Delta t F_3\\
&+ 1.4540585728120278\,\text{e-}01 Y_4 + 2.6901527362840638\,\text{e-}01\Delta t F_4.\\
Y_6 &= 5.1340348144918235\,\text{e-}01 y_{n-1} + 4.6351217028272523\,\text{e-}01 y_n\\
&+ 2.7818804157848459\,\text{e-}01\Delta t f_{n-1} + 8.5754353813657036\,\text{e-}01\Delta t f_n\\
&+ 4.8744453549424173\,\text{e-}02\Delta t^2 y_{n-1}^{(2)} + 1.6750298964583885\,\text{e-}01\Delta t^2 y_n^{(2)}\\
&+ 2.9112290425571579\,\text{e-}03\Delta t^3 y_{n-1}^{(3)} + 3.1978703560973341\,\text{e-}02\Delta t^3 y_n^{(3)}\\
&+ 0.0\Delta t^4 y_{n-1}^{(4)} + 9.3686202649562369\,\text{e-}03\Delta t^4 y_n^{(4)}\\
&+ 1.1899229090206578\,\text{e-}05\Delta t^5 y_{n-1}^{(5)} + 0.0\Delta t^5 y_n^{(5)}\\
&+ 2.1808084717690859\,\text{e-}06\Delta t^6 y_{n-1}^{(6)} + 1.2504381244475798\,\text{e-}04\Delta t^6 y_n^{(6)}
\end{aligned}
$$

$$+ 5.2707126899719991\,\text{e-}03 Y_2 + 9.7513418167703304\,\text{e-}03 \Delta t F_2$$
$$+ 1.7813635578120598\,\text{e-}02 Y_3 + 3.2956994573452636\,\text{e-}02 \Delta t F_3$$
$$+ 0.0 Y_4 + 0.0 \Delta t F_4 + 0.0 Y_5 + 0.0 \Delta t F_5.$$
$$y_{n+1} = 4.4785731431519327\,\text{e-}01 y_{n-1} + 3.0620658750636298\,\text{e-}01 y_n$$
$$+ 2.3774953886124439\,\text{e-}01 \Delta t f_{n-1} + 5.6651259079295457\,\text{e-}01 \Delta t f_n$$
$$+ 5.5214853296717298\,\text{e-}02 \Delta t^2 y_{n-1}^{(2)} + 3.5588913205343700\,\text{e-}02 \Delta t^2 y_n^{(2)}$$
$$+ 7.0738306097583139\,\text{e-}03 \Delta t^3 y_{n-1}^{(3)} + 1.3550346563864980\,\text{e-}02 \Delta t^3 y_n^{(3)}$$
$$+ 5.0532250141140838\,\text{e-}04 \Delta t^4 y_{n-1}^{(4)} + 6.2682487139903084\,\text{e-}04 \Delta t^4 y_n^{(4)}$$
$$+ 1.6325730539221063\,\text{e-}05 \Delta t^5 y_{n-1}^{(5)} + 0.0 \Delta t^5 y_n^{(5)} + 0.0 \Delta t^6 y_{n-1}^{(6)} + 0.0 \Delta t^6 y_n^{(6)}$$
$$+ 3.0376207957774675\,\text{e-}04 Y_2 + 5.6199000840460053\,\text{e-}04 \Delta t F_2$$
$$+ 0.0 Y_3 + 0.0 \Delta t F_3 + 0.0 Y_4 + 0.0 \Delta t F_4$$
$$+ 8.0795449764181029\,\text{e-}02 Y_5 + 1.4947960441653491\,\text{e-}01 \Delta t F_5$$
$$+ 1.6483688633468502\,\text{e-}01 Y_6 + 3.0496460672573128\,\text{e-}01 \Delta t F_6.$$

Bibliography

[1] R. Barrio, Performance of the Taylor series method for ODE/DAEs, Appl. Math. Comput., 163 (2005), 525–545.

[2] R. Barrio, F. Blesa, M. Lara, VSVO formulation of the Taylor method for the numerical solution of ODEs, Comput. Math. Appl., 50 (2005), 93–111.

[3] J. Binney, S. Tremaine, Galactic Dynamics, Princeton Univ Press, 1987.

[4] R. A. Broucke, Numerical integration of periodic orbits in the main problem of artificial satellite theory, Celest. Mech. Dyn. Astr., 58 (1994), 99–123.

[5] A. Deprit and R. M. W. Zahar, Numerical integration of an orbit and its concomitant variations, Z. Angew. Math. Phys., 17 (1966), 425–430.

[6] S. Gottlieb, D. I. Ketcheson, C.-W. Shu, High order strong stability preserving time discretization, J. Sci. Comput., 38(3) (2009), 251–289. DOI: 10.1007/s10915-008-9239-z.

[7] S. Gottlieb, D. I. Ketcheson, C.-W. Shu, Strong Stability Preserving Runge–Kutta and Multistep Time Discretizations, World Scientific, Singapore, 2011.

[8] S. Gottlieb, C.-W. Shu, E. Tadmor, Strong stability-preserving high-order time discretization methods, SIAM Rev., 43 (2001), 89–112.

[9] E. Hairer, S. P. Nørsett, G. Wanner, Solving Ordinary Differential Equations I. Nonstiff Problems. Springer-Verlag, Berlin, 1993.

[10] H. Hénon, C. Heiles, The applicability of the third integral of motion. Some numerical examples, Astron. J. 69 (1964), 73–79.

[11] T. Y. Huang, K. Innanen, A survey of multiderivative multistep integrators, Astron. J., 112(3) (1996), 1254–1262.

[12] C. Huang, Strong stability preserving hybrid methods, Appl. Numer. Math., 59 (2009), 891–904.

[13] T. E. Hull, W. H. Enright, B. M. Fellen, A. E. Sedgwick, Comparing numerical methods for ordinary differential equations. SIAM J. Numer. Anal. 9 (1972), 603–637.

[14] C. A. Kennedy, M. K. Carpenter, R. M. Lewis, Low-storage, explicit Runge–Kutta schemes for the compressible Navier–Stokes equations, Appl. Numer. Math., 35 (2000), 177–219.

[15] D. I. Ketcheson, Highly efficient strong stability preserving Runge–Kutta methods with low-storage implementations, SIAM J. Sci. Comput., 30 (4)(2008), 2113–2136.

[16] J. D. Lambert, Numerical Methods for Ordinary Differential Systems: The Initial Value Problem. Wiley, London, 1991.

[17] M. Lara, A. Elipe, M. Palacios, Automatic programming of recurrent power series, Math. Comput. Simulation, 49 (1999), 351–362.

[18] W. E. Milne, A note on the numerical integration of differential equations, J. Res. Nat. Bur. Standards, 43 (1949), 537–542.

[19] T. Nguyen-Ba, E. Kengne, R. Vaillancourt, One-step 4-stage Hermite–Birkhoff–Taylor ODE solver of order 12, Can. Appl. Math. Q., 16 (1)(2008), 77–94.

[20] T. Nguyen-Ba, H. Nguyen-Thu, T. Giordano, and R. Vaillancourt, Strong-stability-preserving 7-stage Hermite–Birkhoff time-discretization methods, J. Sci. Comput., 50(1) (2012), 63–90, DOI 10.1007/s10915-011-9473-7.

[21] Nguyen-Ba, T., Desjardins, S. J., Sharp, P. W., Vaillancourt, R., 2013. Contractivity-preserving, explicit, Hermite–Obrechkoff ODE solver of order 13, Celest. Mech. Dyn. Astr., (2013), DOI 10.1007/s10569-013-9520-9.

[22] P. J. Prince, J. R. Dormand, High order embedded Runge–Kutta formulae, J. Comput. Appl. Math., 7 (1981), 67–75.

[23] E. Rabe, Determination and survey of periodic Trojan orbits in the restricted problem of three bodies, Astron. J., 66 (1961), 500–513.

[24] S. J. Ruuth and R. J. Spiteri, Two barriers on strong-stability-preserving time discretization methods, J. Sci. Comput. 17 (2002), 211-220.

[25] L. F. Shampine, M. K. Gordon, Computer Solution of Ordinary Differential Equations: The Initial Value Problem, Freeman, San Francisco, CA, 1975.

[26] P. W. Sharp, Numerical comparison of explicit Runge–Kutta pairs of orders four through eight, ACM Trans. Math. Software, 17 (1991), 387–409.

[27] C.-W. Shu, S. Osher, Efficient implementation of essentially non-oscillatory shock-capturing schemes, J. Comput. Phys., 77 (1988), 439–471.

[28] R. J. Spiteri, S. J. Ruuth, A new class of optimal high-order strong-stability-preserving time-stepping schemes, SIAM J. Numer. Anal., 40 (2002), 469–491.

[29] V. Szebehely, Theory of Orbits, The Restricted Problem of Three Bodies, Acad. Press, New York, 1967.