Variable Step-Size Implicit-Explicit Linear Multistep Methods for Time-Dependent PDEs

by

Dong Wang

B.Eng., Harbin Institute of Technology, 1992

M.Eng., Harbin Institute of Technology, 1996

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
IN THE DEPARTMENT
OF
MATHEMATICS

SIMON FRASER UNIVERSITY
Summer 2005

APPROVAL

Name: Dong Wang

Degree: Master of Science

Title of thesis: Variable Step-Size Implicit-Explicit Linear Multistep Methods for Time-Dependent PDEs

Examining Committee: Dr. Ralf Wittenberg
Chair

Dr. Steven Ruuth
Senior Supervisor

Dr. Mary Catherine Kropinski
Supervisor

Dr. Jim Verner
Supervisor

Dr. Robert Russell
Internal/External Examiner

Date Approved: July 7, 2005

SIMON FRASER UNIVERSITY

PARTIAL COPYRIGHT LICENCE

Abstract

The objectives of this thesis are to design, analyze and numerically investigate easily implementable Variable Step-Size Implicit-Explicit (VSIMEX) Linear Multistep Methods for time-dependent PDEs.

The thesis begins with a derivation of the family of second-order, two-step VSIMEX schemes with two free parameters. A zero-stability analysis of these VSIMEX schemes gives analytical results on the restriction of the step-size ratio for general second-order VSIMEX schemes. The family of third-order, three-step VSIMEX schemes with three free parameters is also derived. A zero-stability analysis of these VSIMEX schemes gives numerical values for the step-size restrictions. A fourth-order, four-step VSIMEX scheme and its stability properties are also studied.

Numerically, we apply our new VSIMEX schemes to the 1-D advection-diffusion and Burgers' equations. The expected orders of convergence are achieved, and accurate approximate solutions are obtained. Our results demonstrate the superiority of VSIMEX schemes over classical IMEX schemes in solving Burgers' equation.

Acknowledgments

I express my utmost gratitude to my senior supervisor, Dr. Steven Ruuth, for suggesting the topic, for his guidance, encouragement and support in the development of this thesis, for his efforts made in reading and correcting this thesis; without that this thesis would not have been executed.

I thank all members of examining committee, Dr. Robert Russell, Dr. Jim Verner, Dr. Mary Catherine Kropinski and Dr. Ralf Wittenberg, for their invaluable comments and corrections to this thesis.

Special thanks must go to Dr. Wentao Sun for his many helpful hints and supports in carrying out numerical experiments. Many thanks also go to Colin Macdonald and Benjamin Ong for their carefully proofreading of this thesis.

Financial support received from NSERC and the Department of Mathematics is much obliged and appreciated.

Finally, I thank my wife, Huihua, for her countless support and encouragement during my studies at Simon Fraser University, graduate program.

Dedication

To
my parents,
my wife, Huihua
and
my daughter, Firmiana

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 IMEX Schemes

Many problems in physics, engineering, chemistry, biology and other areas involve the numerical solution of time-dependent Partial Differential Equations (PDEs). Some types of PDEs can conveniently be transformed into large systems of Ordinary Differential Equations (ODEs) in time by doing spatial discretizations based on finite difference methods, finite volume methods, spectral methods or finite element methods.

For large systems of ODEs with both stiff and nonstiff parts, the classical Implicit-Explicit (IMEX) linear multistep methods treat the stiff part implicitly and the nonstiff part explicitly (see Ascher, Ruuth and Wetton [1]). This has proven to be a powerful technique. The main idea behind these IMEX methods developed in [1] is outlined below.

Large systems of ODEs typically have the form

$$\dot{u} = f(u) + g(u). \tag{1.1}$$

The term $f(u)$ in (1.1) is a nonstiff and possibly nonlinear term which we do not wish to integrate implicitly. This could be because the Jacobian of $f(u)$ is non-symmetric and non-definite and an iterative solution of the implicit equations is desired, or the Jacobian could be dense, requiring the inversion of a full matrix at each time step (see Ascher et al. [1]), or else an explicit scheme may be preferred for ease of implementation.

On the other hand, the $g(u)$ term in (1.1) is stiff. Consequently, an implicit time integrator should be used to prevent the unrealistically small time steps which arise from an explicit treatment.

Hence for solving ODEs (1.1), numerical schemes which integrate the $g(u)$ term implicitly and $f(u)$ term explicitly are highly desired. Such implicit-explicit methods are referred to as IMEX schemes in Ascher et al. [1].

1.2 Motivation for VSIMEX Methods

For solutions of ODEs (1.1) with different time scales, i.e. solutions rapidly varying in some regions of the time domain while slowly changing in other regions, variable step-size schemes are often essential to obtain computationally efficient, accurate results. For example, small time steps may be necessary to capture rapidly varying initial transients, while large time steps may be desirable to capture the subsequent slowly changing, long-term evolution of the system.

Standard IMEX linear multistep methods are designed for the case of constant step-sizes. Thus starting values must be computed every time the temporal step-size is varied for standard IMEX schemes.

A commonly used approach for handling variable step-sizes for linear multistep methods is the interpolation method (see Nordsieck [9]). At each time-step, a standard fixed step-size linear multistep formula is applied, and if necessary the required starting values are approximated by interpolating through the past saved values. Unfortunately, this process is sufficiently complicated that it is often avoided in practice.

The motivation behind this thesis is to provide the end-users of IMEX schemes with easily implementable variable step-size IMEX (VSIMEX) linear multistep schemes.

1.3 Overview

The overview of this thesis is as follows:

In Chapter 2, general VSIMEX linear multistep schemes are defined. The order conditions for order-s, s-step VSIMEX schemes are derived from the local truncation error. This is followed by the derivation of different VSIMEX schemes up to fourth-order. First-order, one-step IMEX schemes are also VSIMEX schemes. A general class of second-order, two-step VSIMEX schemes with two parameters is developed, and some popular second-order IMEX schemes and their corresponding variable step-size versions are presented. A particular parameterization of third-order, three-step

VSIMEX schemes with three parameters is given and a fourth-order VSIMEX scheme is also presented.

In Chapter 3, the linear stability analysis of IMEX schemes is discussed and reviewed, and stability contours for different order IMEX schemes are plotted. The second part of this chapter deals with the zero-stability analysis of VSIMEX schemes. Zero-stability imposes restrictions on the step-size ratios required to ensure that VSIMEX schemes are stable in the limit as the step-sizes approach zero.

In Chapter 4, numerical experiments for the linear advection-diffusion equation and Burgers' equation are carried out using various IMEX and VSIMEX schemes. Accurate approximate solutions are obtained, and the expected orders of convergence for our VSIMEX schemes are verified for a variety of time-step strategies.

In Chapter 5, a summary of this thesis and ideas for future work are presented.

Chapter 2

Lower and Higher Order VSIMEX Schemes

In this chapter, different VSIMEX linear multistep schemes up to fourth-order are derived. The first-, second- and third-order VSIMEX linear multistep schemes are families of methods which admit one, two and three free parameters respectively. For the fourth-order VSIMEX linear multistep scheme, we focus on the fourth-order variable step-size, semi-implicit, Backward Differentiation Formula (VSSBDF4).

2.1 General VSIMEX Linear Multistep Methods

We consider an arbitrary grid $\{t_n\}$ and denote the step-size $k_{n+j} = t_{n+j+1} - t_{n+j}$. Furthermore, assume that the previous s approximations U^{n+j} to $u(t_{n+j}), j = 0, 1, \ldots, s-1$, are known.

The general s-step VSIMEX linear multistep schemes for ODEs (1.1) take the form

$$\frac{1}{k_{n+s-1}} \sum_{j=0}^{s} \alpha_{j,n} U^{n+j} = \sum_{j=0}^{s-1} \beta_{j,n} f(U^{n+j}) + \sum_{j=0}^{s} C_{j,n} g(U^{n+j}), \tag{2.1}$$

where $\alpha_{s,n} \neq 0$, $C_{s,n} \neq 0$ and $s \geq 2$. The variable coefficients $\alpha_{j,n}$, $\beta_{j,n}$ and $C_{j,n}$ are functions of the mesh ratios $\omega_i = k_i/k_{i-1}$ for $i = n+1, ..., n+s-1$, $s \geq 2$ and must satisfy the order conditions (2.4) listed below.

Ascher et al. [1] proved for fixed step-sizes k that s-step IMEX schemes achieve at most order-s accuracy and that this is achieved by an s-parameter family of schemes.

In this thesis, we only consider s-step, $O(\overline{k}^s)$ VSIMEX linear multistep schemes, where $\overline{k}$ is the average temporal step-size.

Begin by assuming that the mesh ratios k_i/k_n and the variable coefficients $\alpha_{j,n}$, $\beta_{j,n}$ and $C_{j,n}$ are all bounded for $i = n+1, ..., n+s-1$, replace the approximate solutions U^{n+j}, $j = 0, 1, ..., s$ by the corresponding exact solutions $u(t_{n+j})$ in the variable coefficient difference equation (2.1) to obtain the Local Truncation Error (LTE)

$$\text{LTE} = \frac{1}{k_{n+s-1}} \sum_{j=0}^{s} \alpha_{j,n} u(t_{n+j}) - \sum_{j=0}^{s-1} \beta_{j,n} f(u(t_{n+j})) - \sum_{j=0}^{s} C_{j,n} g(u(t_{n+j})). \tag{2.2}$$

For a smooth function $u(t)$, expanding equation (2.2) in a Taylor series about t_n yields

$$\begin{aligned}
\text{LTE} = \frac{1}{k_{n+s-1}} \Bigg\{ & \alpha_{0,n} u(t_n) + \sum_{j=1}^{s} \alpha_{j,n} \Bigg[u(t_n) + u'(t_n) \sum_{i=0}^{j-1} k_{n+i} + \\
& + \frac{u''(t_n)}{2!} \left(\sum_{i=0}^{j-1} k_{n+i} \right)^2 + \cdots + \frac{u^{(p)}(t_n)}{p!} \left(\sum_{i=0}^{j-1} k_{n+i} \right)^p \Bigg] \Bigg\} \\
& - \beta_{0,n} f(u(t_n)) - \sum_{j=1}^{s-1} \beta_{j,n} \Bigg[f(u(t_n)) + \frac{df}{dt}(u(t_n)) \sum_{i=0}^{j-1} k_{n+i} + \cdots \\
& + \frac{1}{(p-1)!} \frac{d^{(p-1)} f}{dt^{(p-1)}}(u(t_n)) \left(\sum_{i=0}^{j-1} k_{n+i} \right)^{p-1} \Bigg] - C_{0,n} g(u(t_n)) \\
& - \sum_{j=1}^{s} C_{j,n} \Bigg[g(u(t_n)) + \frac{dg}{dt}(u(t_n)) \sum_{i=0}^{j-1} k_{n+i} + \cdots \\
& + \frac{1}{(p-1)!} \frac{d^{(p-1)} g}{dt^{(p-1)}}(u(t_n)) \left(\sum_{i=0}^{j-1} k_{n+i} \right)^{p-1} \Bigg] + O(k_n^p).
\end{aligned} \tag{2.3}$$

Applying equation (1.1) to the LTE (2.3), a pth-order VSIMEX scheme is obtained provided that the following constraints for $\alpha_{j,n}$, $\beta_{j,n}$ and $c_{j,n}$ hold:

$$\begin{aligned}
& \sum_{j=0}^{s} \alpha_{j,n} = 0, \\
& \sum_{j=1}^{s} \alpha_{j,n} \left(\sum_{i=0}^{j-1} k_{n+i} \right) = k_{n+s-1} \sum_{j=0}^{s-1} \beta_{j,n} = k_{n+s-1} \sum_{j=0}^{s} C_{j,n}, \\
& \vdots
\end{aligned} \tag{2.4}$$

$$\begin{aligned}\frac{1}{p!}\sum_{j=1}^{s}\alpha_{j,n}\left(\sum_{i=0}^{j-1}k_{n+i}\right)^{p} &= k_{n+s-1}\frac{1}{(p-1)!}\sum_{j=1}^{s-1}\beta_{j,n}\left(\sum_{i=0}^{j-1}k_{n+i}\right)^{p-1} \\ &= k_{n+s-1}\frac{1}{(p-1)!}\sum_{j=1}^{s}C_{j,n}\left(\sum_{i=0}^{j-1}k_{n+i}\right)^{p-1}.\end{aligned}$$

Taken together, these constraints are known as the order conditions.

2.2 First-Order VSIMEX Schemes

First-order, one-step IMEX schemes are actually VSIMEX schemes, since they allow for variable time-stepping. This one-parameter family of schemes for (1.1) can be expressed as (see Ascher et al. [1])

$$U^{n+1} - U^n = k_n f(U^n) + k_n[(1-\gamma)g(U^n) + \gamma g(U^{n+1})]. \tag{2.5}$$

The leading order term in the LTE in (2.5) is given by

$$\frac{k_n}{2}\ddot{u}(t_n) - k_n\gamma\frac{dg}{dt}(t_n), \tag{2.6}$$

which suggests the restriction $\gamma \in [0,1]$ to maintain a moderate LTE.

Some schemes in this one-parameter family are familiar. For example $\gamma = 0$ gives the Forward Euler scheme

$$U^{n+1} - U^n = k_n f(U^n) + k_n g(U^n). \tag{2.7}$$

Since this is a fully explicit scheme, rather than an IMEX scheme, we will not consider it further.

When $\gamma = \frac{1}{2}$, we have

$$U^{n+1} - U^n = k_n f(U^n) + \frac{1}{2}k_n[g(U^n) + g(U^{n+1})], \tag{2.8}$$

which applies the second-order, one-step Crank-Nicolson method to $g(u)$ and the Forward Euler method to $f(u)$.

Another choice, $\gamma = 1$, yields

$$U^{n+1} - U^n = k_n f(U^n) + k_n g(U^{n+1}), \tag{2.9}$$

which applies Backward Euler to $g(u)$ and Forward Euler to $f(u)$. As we know, the Backward Euler method is the first-order member of the class of Backward Differentiation Formulas (BDFs) (see Lambert [8]).

In later sections, we will also develop some order-p VSIMEX schemes ($p = 2, 3, 4$) similar to (2.9), which apply BDFs to g and extrapolate f to time step t_{n+p}. Those schemes will be referred to as order-p Variable Step-size Semi-implicit BDF (VSSBDFp) schemes.

In practice, at least a second-order time integrator is desirable since a second-order spatial discretization is often used. In the next section, we derive the general second-order, two-step VSIMEX schemes with two free parameters, and highlight some particular VSIMEX schemes whose corresponding IMEX schemes are quite familiar to us.

2.3 Second-Order VSIMEX Schemes

Second-order, two-step VSIMEX schemes admit two free parameters. If our VSIMEX schemes are centered in time about time-step $t_{n+1+\gamma}$ to second-order, we derive second-order VSIMEX schemes, viz., a family of schemes involving two parameters (γ, c) for which equation (2.1) is

$$\frac{1}{k_{n+1}} \sum_{j=0}^{2} \alpha_{j,n} U^{n+j} = \sum_{j=0}^{1} \beta_{j,n} f(U^{n+j}) + \sum_{j=0}^{2} C_{j,n} g(U^{n+j}) \tag{2.10}$$

where

$$\begin{aligned}
\alpha_{0,n} &= \frac{(2\gamma - 1)\omega_{n+1}^2}{1 + \omega_{n+1}}, \\
\alpha_{1,n} &= (1 - 2\gamma)\omega_{n+1} - 1, \\
\alpha_{2,n} &= \frac{1 + 2\gamma\omega_{n+1}}{1 + \omega_{n+1}}, \\
\beta_{0,n} &= -\gamma\omega_{n+1}, \\
\beta_{1,n} &= 1 + \gamma\omega_{n+1}, \\
C_{0,n} &= \frac{c}{2}, \\
C_{1,n} &= 1 - \gamma - (1 + \frac{1}{\omega_{n+1}})\frac{c}{2}, \\
C_{2,n} &= \gamma + \frac{c}{2\omega_{n+1}}.
\end{aligned} \tag{2.11}$$

In the constant step size case (i.e., if we set all consecutive step-size ratios $\omega_{n+1} = 1$ for all $n = 0, 1, \ldots, N-2$, where N is the number of total nodes in time interval $[0, T]$), the schemes (2.10) reduce to the family of IMEX schemes (see Ascher et al. [1])

$$\begin{aligned}\frac{1}{k}[(\gamma+\frac{1}{2})U^{n+2} - 2\gamma U^{n+1} + (\gamma-\frac{1}{2})U^n] = \\ (\gamma+1)f(U^{n+1}) - \gamma f(U^n) + \\ (\gamma+\frac{c}{2})g(U^{n+2}) + (1-\gamma-c)g(U^{n+1}) + \frac{c}{2}g(U^n),\end{aligned} \tag{2.12}$$

where k is the constant temporal step-size.

Some VSIMEX schemes corresponding to familiar IMEX schemes are as follows:

- $(\gamma, c) = (\frac{1}{2}, 0)$ gives

$$\begin{aligned}\frac{1}{k_{n+1}}\left[U^{n+2} - U^{n+1}\right] &= \left(1 + \frac{1}{2}\omega_{n+1}\right) f(U^{n+1}) - \frac{1}{2}\omega_{n+1} f(U^n) \\ &\quad + \frac{1}{2}\left[g(U^{n+2}) + g(U^{n+1})\right].\end{aligned} \tag{2.13}$$

  Since it applies Crank-Nicolson to the stiff term and the variable step-size second-order Adams-Bashforth scheme to the nonstiff term, this scheme will be referred to VSCNAB (Variable Step-size Crank-Nicolson, Adams-Bashforth).

  The corresponding constant step-size IMEX version (CNAB) is a popular scheme in computational fluid dynamics (see Ascher et al. [1]). It has the form

$$\frac{1}{k}\left[U^{n+2} - U^{n+1}\right] = \frac{3}{2}f(U^{n+1}) - \frac{1}{2}f(U^n) + \frac{1}{2}\left[g(U^{n+2}) + g(U^{n+1})\right]. \tag{2.14}$$

- $(\gamma, c) = (1, 0)$ gives

$$\begin{aligned}\frac{1}{k_{n+1}}\left[\frac{1+2\omega_{n+1}}{1+\omega_{n+1}}U^{n+2} - (1+\omega_{n+1})U^{n+1} + \frac{\omega_{n+1}^2}{1+\omega_{n+1}}U^n\right] = \\ (1+\omega_{n+1})f(U^{n+1}) - \omega_{n+1}f(U^n) + g(U^{n+2}).\end{aligned} \tag{2.15}$$

  As mentioned in Section 2.2, this scheme applies a variable step-size BDF2 scheme to the stiff part and extrapolates the nonstiff part to time step t_{n+2}. This scheme will be referred to as the variable step-size second-order semi-implicit BDF (VSS-BDF2).

  The corresponding constant step-size IMEX version (SBDF2) is given by

$$\frac{1}{k}\left[\frac{3}{2}U^{n+2} - 2U^{n+1} + \frac{1}{2}U^n\right] = 2f(U^{n+1}) - f(U^n) + g(U^{n+2}). \tag{2.16}$$

- $(\gamma, c) = (0, 1)$ gives

$$\frac{1}{k_{n+1}}\left[\frac{1}{1+\omega_{n+1}}U^{n+2} + (\omega_{n+1}-1)U^{n+1} - \frac{\omega_{n+1}^2}{1+\omega_{n+1}}U^n\right] = f(U^{n+1}) + \frac{1}{2}\left[\frac{1}{\omega_{n+1}}g(U^{n+2}) + (1-\frac{1}{\omega_{n+1}})g(U^{n+1}) + g(U^n)\right]. \quad (2.17)$$

The corresponding constant step-size IMEX scheme is

$$\frac{1}{2k}\left(U^{n+2} - U^n\right) = f(U^{n+1}) + \frac{1}{2}\left[g(U^{n+2}) + g(U^n)\right], \quad (2.18)$$

which applies a scheme somewhat like the Crank-Nicolson to the stiff part and a Leap-Frog scheme to nonstiff part. In Ascher et al. [1], scheme (2.18) is referred to as CNLF (Crank-Nicolson, Leap-Frog). Correspondingly, we call scheme (2.17) VSCNLF (Variable Step-size CNLF).

- $(\gamma, c) = (\frac{1}{2}, \frac{1}{8})$ gives

$$\frac{1}{k_{n+1}}\left(U^{n+2} - U^{n+1}\right) = \frac{1}{2}\left[(2+\omega_{n+1})f(U^{n+1}) - \omega_{n+1}f(U^n)\right] + \frac{1}{16\omega_{n+1}}\left[(8\omega_{n+1}+1)g(U^{n+2}) + (7\omega_{n+1}-1)g(U^{n+1}) + \omega_{n+1}g(U^n)\right]. \quad (2.19)$$

The corresponding constant step-size scheme, referred to as the Modified CNAB method in Ascher et al. [1] is

$$\frac{1}{k}\left(U^{n+2} - U^{n+1}\right) = \frac{3}{2}f(U^{n+1}) - \frac{1}{2}f(U^n) + \frac{9}{16}g(U^{n+2}) + \frac{3}{8}g(U^{n+1}) + \frac{1}{16}g(U^n). \quad (2.20)$$

Similar to CNAB, the modified CNAB scheme has a small truncation error. For problems with small mesh Reynolds numbers, this modified scheme will be preferred due to its superior damping of high frequency errors ([1]).

- $(\gamma, c) = (\frac{1}{2}, -\frac{\omega_{n+1}^2}{3(1+\omega_{n+1})})$ yields

$$\frac{1}{k_{n+1}}\left(U^{n+2} - U^{n+1}\right) = \left(1 + \frac{\omega_{n+1}}{2}\right)f(U^{n+1}) - \frac{\omega_{n+1}}{2}f(U^n) + \frac{1}{6(1+\omega_{n+1})}[(3+2\omega_{n+1})g(U^{n+2}) + (3+\omega_{n+1})(1+\omega_{n+1})g(U^{n+1}) - \omega_{n+1}^2 g(U^n)], \quad (2.21)$$

which treats the stiff g term using the two-step variable step-size implicit Adams-Moulton scheme, and the nonstiff f term using the two-step variable step-size explicit Adams-Bashforth scheme. We refer to (2.21) as VSAMAB.

The corresponding constant step-size AMAB scheme is given by

$$\frac{1}{k}\left(U^{n+2}-U^{n+1}\right)=\frac{3}{2}f(U^{n+1})-\frac{1}{2}f(U^{n}) + \frac{1}{12}\left[5g(U^{n+2})+8g(U^{n+1})-g(U^{n})\right]. \tag{2.22}$$

Schemes (2.21) and (2.22) are rarely useful in practical computation since they exhibit poor linear stability. See Figure 3.10 of the next chapter for further details.

- $(\gamma, c) = (\frac{1}{2}, -\omega_{n+1})$ yields the fully explicit second-order, two-step variable step-size Adams-Bashforth scheme

$$\begin{aligned}\frac{1}{k_{n+1}}\left(U^{n+2}-U^{n+1}\right) &= (1+\frac{\omega_{n+1}}{2})\left[f(U^{n+1})+g(U^{n+1})\right]. \\ &\quad -\frac{\omega_{n+1}}{2}[f(U^{n})+g(U^{n})].\end{aligned} \tag{2.23}$$

This is not a VSIMEX scheme, so we will not consider it further.

2.4 Third-Order VSIMEX Schemes

Third-order, three-step VSIMEX schemes admit three free parameters. One particular parametrization for equation (2.1) can be derived by introducing the three parameters (γ, θ, c). This leads to

$$\frac{1}{k_{n+2}}\sum_{j=0}^{3}\alpha_{j,n}U^{n+j}=\sum_{j=0}^{2}\beta_{j,n}f(U^{n+j})+\sum_{j=0}^{3}C_{j,n}g(U^{n+j}), \tag{2.24}$$

where

$$
\begin{aligned}
\alpha_{0,n} &= -\frac{\omega_{n+1}^3\omega_{n+2}^2[3\gamma^2\omega_{n+2}+2\gamma(1-\omega_{n+2})-1]}{(1+\omega_{n+1})[1+\omega_{n+1}(1+\omega_{n+2})]},\\
\alpha_{1,n} &= -\frac{\omega_{n+2}^2[\gamma\omega_{n+1}\omega_{n+2}(2-3\gamma)+(1-2\gamma)(1+\omega_{n+1})]}{1+\omega_{n+2}},\\
\alpha_{2,n} &= -\frac{\omega_{n+1}(1+\gamma\omega_{n+2})[1+\omega_{n+2}(3\gamma-2)]+\omega_{n+2}(2\gamma-1)+1}{1+\omega_{n+1}}-\theta,\\
\alpha_{3,n} &= \frac{1+2\gamma\omega_{n+2}+\omega_{n+1}(1+\gamma\omega_{n+2})(1+3\gamma\omega_{n+2})}{(1+\omega_{n+2})[1+\omega_{n+1}(1+\omega_{n+2})]}+\theta,
\end{aligned}
$$

$$
\begin{aligned}
\beta_{0,n} &= \frac{\omega_{n+1}^2\omega_{n+2}[6\gamma(1+\gamma\omega_{n+2})+\theta(3+2\omega_{n+2})]}{6(1+\omega_{n+1})},\\
\beta_{1,n} &= -\gamma\omega_{n+2}[1+\omega_{n+1}(1+\gamma\omega_{n+2})]-\frac{\theta\omega_{n+2}[3+\omega_{n+1}(3+2\omega_{n+2})]}{6},\\
\beta_{2,n} &= \frac{(1+\gamma\omega_{n+2})[1+\omega_{n+1}(1+\gamma\omega_{n+2})]}{(1+\omega_{n+1})}\\
&\quad+\theta\left[1+\frac{\omega_{n+2}}{2}+\frac{\omega_{n+1}\omega_{n+2}(3+2\omega_{n+2})}{6(1+\omega_{n+1})}\right],
\end{aligned}
\tag{2.25}
$$

$$
\begin{aligned}
C_{0,n} &= \frac{\theta\omega_{n+1}^2\omega_{n+2}(3+2\omega_{n+2})}{6(1+\omega_{n+1})}-c,\\
C_{1,n} &= \frac{c(1+\omega_{n+1})[1+\omega_{n+1}(1+\omega_{n+2})]-\omega_{n+1}^2\omega_{n+2}^2\gamma(1-\gamma)}{\omega_{n+1}^2(1+\omega_{n+2})}\\
&\quad-\frac{\theta\omega_{n+2}[3+\omega_{n+1}(3+2\omega_{n+2})]}{6},\\
C_{2,n} &= \frac{\omega_{n+1}^2\omega_{n+2}(1-\gamma)(1+\gamma\omega_{n+2})-c[1+\omega_{n+1}(1+\omega_{n+2})]}{\omega_{n+1}^2\omega_{n+2}}\\
&\quad+\theta\left[1+\frac{\omega_{n+2}}{2}+\frac{\omega_{n+1}\omega_{n+2}(3+2\omega_{n+2})}{6(1+\omega_{n+1})}\right],\\
C_{3,n} &= \frac{\omega_{n+1}^2\omega_{n+2}\gamma(1+\gamma\omega_{n+2})+c(1+\omega_{n+1})}{\omega_{n+1}^2\omega_{n+2}(1+\omega_{n+2})}.
\end{aligned}
$$

Remark:

- Restricting scheme (2.24) to constant step-sizes leads to the three-parameter family of third-order IMEX schemes appearing in Ascher et al. [1]

$$\begin{aligned}
&\frac{1}{k}\left[\left(\frac{1}{2}\gamma^2+\gamma+\frac{1}{3}+\theta\right)U^{n+3}+\left(-\frac{3}{2}\gamma^2-2\gamma+\frac{1}{2}-\theta\right)U^{n+2}\right.\\
&\left.+\left(\frac{3}{2}\gamma^2+\gamma-1\right)U^{n+1}+\left(-\frac{1}{2}\gamma^2+\frac{1}{6}\right)U^{n}\right]\\
=\;&\left(\frac{\gamma^2+3\gamma}{2}+1+\frac{23}{12}\theta\right)f(U^{n+2})-\left(\gamma^2+2\gamma+\frac{4}{3}\theta\right)f(U^{n+1})\\
&+\left(\frac{\gamma^2+\gamma}{2}+\frac{5}{12}\theta\right)f(U^{n})+\left(\frac{\gamma^2+\gamma}{2}+c\right)g(U^{n+3})\\
&+\left(1-\gamma^2-3c+\frac{23}{12}\theta\right)g(U^{n+2})\\
&+\left(\frac{\gamma^2-\gamma}{2}+3c-\frac{4}{3}\theta\right)g(U^{n+1})+\left(\frac{5}{12}\theta-c\right)g(U^{n}). \qquad (2.26)
\end{aligned}$$

- Scheme (2.24) is centered at time step $t_{n+2+\gamma}$ whenever $\theta = 0$.

- Letting $\theta \to \pm\infty$ in (2.25) yields the variable step-size third-order fully explicit Adams-Bashforth scheme.

Note that choosing $(\gamma, \theta, c) = (1, 0, 0)$ in scheme (2.24) yields the variable step-size third-order semi-implicit BDF (VSSBDF3)

$$\frac{1}{k_{n+2}} \sum_{j=0}^{3} \alpha_{j,n} U^{n+j} = \sum_{j=0}^{2} \beta_{j,n} f(U^{n+j}) + g(U^{n+3}), \tag{2.27}$$

where

$$\begin{aligned}
\alpha_{0,n} &= -\frac{\omega_{n+1}^3 \omega_{n+2}^2 (1+\omega_{n+2})}{(1+\omega_{n+1})(1+\omega_{n+1}+\omega_{n+1}\omega_{n+2})}, \\
\alpha_{1,n} &= \omega_{n+2}^2 (\omega_{n+1} + \frac{1}{1+\omega_{n+2}}), \\
\alpha_{2,n} &= -1 - \omega_{n+2} - \frac{\omega_{n+1}\omega_{n+2}(1+\omega_{n+2})}{1+\omega_{n+1}}, \\
\alpha_{3,n} &= 1 + \frac{\omega_{n+2}}{1+\omega_{n+2}} + \frac{\omega_{n+1}\omega_{n+2}}{1+\omega_{n+1}(1+\omega_{n+2})}, \\
\beta_{0,n} &= \frac{\omega_{n+1}^2 \omega_{n+2}(1+\omega_{n+2})}{1+\omega_{n+1}}, \\
\beta_{1,n} &= -\omega_{n+2}[1+\omega_{n+1}(1+\omega_{n+2})], \\
\beta_{2,n} &= \frac{(1+\omega_{n+2})[1+\omega_{n+1}(1+\omega_{n+2})]}{1+\omega_{n+1}}.
\end{aligned} \tag{2.28}$$

2.5 Fourth-Order VSIMEX Schemes

For the fourth-order, four-step VSIMEX schemes, we only develop the variable step-size fourth-order, four-step semi-implicit BDF (VSSBDF4) scheme. This particular fourth-order scheme for equation (2.1) has the form

$$\frac{1}{k_{n+3}}\sum_{j=0}^{4}\alpha_{j,n}U^{n+j}=\sum_{j=0}^{3}\beta_{j,n}f(U^{n+j})+g(U^{n+4}), \tag{2.29}$$

where

$$\begin{aligned}
\alpha_{0,n}&=\frac{1+\omega_{n+3}}{1+\omega_{n+1}}\frac{A_2}{A_1}\frac{\omega_{n+1}^4\omega_{n+2}^3\omega_{n+3}^2}{A_3},\\
\alpha_{1,n}&=-\omega_{n+2}^3\omega_{n+3}^2\frac{1+\omega_{n+3}}{1+\omega_{n+2}}\frac{A_3}{A_2},\\
\alpha_{2,n}&=\omega_{n+3}\left[\frac{\omega_{n+3}}{1+\omega_{n+3}}+\omega_{n+2}\omega_{n+3}\frac{A_3+\omega_{n+1}}{1+\omega_{n+1}}\right],\\
\alpha_{3,n}&=-1-\omega_{n+3}\left[1+\frac{\omega_{n+2}(1+\omega_{n+3})}{1+\omega_{n+2}}(1+\frac{\omega_{n+1}A_2}{A_1})\right],\\
\alpha_{4,n}&=1+\frac{\omega_{n+3}}{1+\omega_{n+3}}+\frac{\omega_{n+2}\omega_{n+3}}{A_2}+\frac{\omega_{n+1}\omega_{n+2}\omega_{n+3}}{A_3},\\
\beta_{0,n}&=-\omega_{n+1}^3\omega_{n+2}^2\omega_{n+3}\frac{1+\omega_{n+3}}{1+\omega_{n+1}}\frac{A_2}{A_1},\\
\beta_{1,n}&=\omega_{n+2}^2\omega_{n+3}\frac{1+\omega_{n+3}}{1+\omega_{n+2}}A_3,\\
\beta_{2,n}&=-A_2A_3\frac{\omega_{n+3}}{1+\omega_{n+1}},\\
\beta_{3,n}&=\frac{\omega_{n+2}(1+\omega_{n+3})}{1+\omega_{n+2}}\frac{(1+\omega_{n+3})(A_3+\omega_{n+1})+\frac{1+\omega_{n+1}}{\omega_{n+2}}}{A_1},\\
A_1&=1+\omega_{n+1}(1+\omega_{n+2}),\\
A_2&=1+\omega_{n+2}(1+\omega_{n+3}),\\
A_3&=1+\omega_{n+1}A_2.
\end{aligned} \tag{2.30}$$

In the constant temporal step-size case, equation (2.29) reduces to the fourth-order SBDF (see equation (33) in Ascher et al. [1]). This scheme is given by

$$\begin{aligned}
&\frac{1}{k}\left[\frac{25}{12}U^{n+4}-4U^{n+3}+3U^{n+2}-\frac{4}{3}U^{n+1}+\frac{1}{4}U^{n}\right]=\\
&\quad 4f(U^{n+3})-6f(U^{n+2})+4f(U^{n+1})-f(U^{n})+g(U^{n+4}).
\end{aligned} \tag{2.31}$$

Chapter 3

Stability Analyses of IMEX and VSIMEX Schemes

If the constraints (2.4) are satisfied, then the VSIMEX methods (2.1) are consistent. By the Lax-Richtmeyer Theorem (see Strikwerda [10]), IMEX methods are convergent when applied to a well-posed initial value problem, provided stability is ensured (see Ascher et al. [1]).

Hence, we are also interested in the stability properties of our VSIMEX schemes, namely, what restrictions on step-size ratios ω_i, $i = n+1, \ldots, n+s-1$ are required in order to ensure order-s, s-step VSIMEX schemes are stable.

It is difficult to develop a linear stability analysis of VSIMEX schemes due to the freedom introduced by the step-size ratios, ω_i. In this thesis, we focus on zero-stability. This analysis appears in Section 3.2.

When constant step-sizes are utilized, our VSIMEX schemes reduce to IMEX schemes. To better understand the stability properties associated with these IMEX schemes, a linear stability analysis is presented in Section 3.1.

3.1 Linear Stability Analysis of IMEX Schemes

In this thesis, our linear stability analysis of IMEX schemes mainly follows the methods developed in Ascher et al. [1].

The model problem for our linear stability analysis of IMEX schemes is the 1-D

advection-diffusion equation,

$$u_t = au_x + bu_{xx}, \tag{3.1}$$

subject to periodic boundary conditions on the interval [0,1]. Here a, b are constants and $b \geq 0$.

In the spatial discretization, central finite differences schemes are used to approximate u_x and u_{xx} and we denote by D_1 and D_2 the first- and second-order derivative approximations in x respectively. Then equation (3.1) is discretized into a system of semi-discrete equations in time

$$\begin{aligned} \dot{U}_j &= aD_1U_j + bD_2U_j \\ &= a\frac{U_{j+1} - U_{j-1}}{2h} + b\frac{U_{j+1} - 2U_j + U_{j-1}}{h^2} \end{aligned} \tag{3.2}$$

where $1 \leq j \leq M$ and M is the number of spatial mesh grids. Here we assume M is even, and the spatial mesh grid size is $h = \frac{1}{M}$.

By applying the discrete Fourier transform (DFT) and its Inverse DFT (see Trefethen [11]) to (3.2) we obtain for every j

$$\begin{aligned} \frac{1}{2\pi}\sum_{\ell=-\frac{M}{2}+1}^{M/2} e^{i\ell j 2\pi h}\hat{\dot{U}}_\ell &= \frac{a}{2h}\left[\frac{1}{2\pi}\sum_{\ell=-\frac{M}{2}+1}^{M/2}(e^{i\ell(j+1)2\pi h} - e^{i\ell(j-1)2\pi h})\hat{U}_\ell\right] + \\ &\quad \frac{b}{h^2}\left[\frac{1}{2\pi}\sum_{\ell=-\frac{M}{2}+1}^{M/2}(e^{i\ell(j-1)2\pi h} - 2e^{i\ell j 2\pi h} + e^{i\ell(j+1)2\pi h})\hat{U}_\ell\right], \end{aligned} \tag{3.3}$$

where

$\hat{U}_\ell$, the DFT of U_ℓ, is given by

$$\hat{U}_\ell = 2\pi h\sum_{j=1}^{M} e^{-i\ell 2\pi h j}U_j, \qquad \ell = -\frac{M}{2}+1, \cdots, \frac{M}{2}, \tag{3.4}$$

and its Inverse DFT is given by

$$U_j = \frac{1}{2\pi}\sum_{\ell=-\frac{M}{2}+1}^{M/2} e^{i\ell 2\pi h j}\hat{U}_\ell, \qquad j = 1, 2, \cdots, M. \tag{3.5}$$

By equating the coefficients in (3.3) and making use of the following identities,

$$e^{i\ell(j+1)2\pi h} - e^{i\ell(j-1)2\pi h} = i2e^{i\ell j 2\pi h}\sin(2\pi\ell h), \tag{3.6}$$

$$e^{i\ell(j-1)2\pi h} - 2e^{i\ell j2\pi h} + e^{i\ell(j+1)2\pi h} = 2e^{i\ell j2\pi h}\left[\cos(2\pi\ell h) - 1\right], \tag{3.7}$$

we find for every wavenumber ℓ

$$\dot{\hat{U}}_\ell = i\frac{a}{h}\sin(2\pi\ell h)\hat{U}_\ell + \frac{2b}{h^2}\left[\cos(2\pi\ell h) - 1\right]\hat{U}_\ell, \qquad \ell = -\frac{M}{2}+1,\cdots,\frac{M}{2}. \tag{3.8}$$

Letting

$$\begin{aligned} \alpha_\ell &= \frac{2b}{h^2}\left[\cos(2\pi\ell h) - 1\right], \\ \beta_\ell &= \frac{a}{h}\sin(2\pi\ell h), \end{aligned} \tag{3.9}$$

then we can rewrite (3.8) as

$$\dot{\hat{U}}_\ell = i\beta_\ell\hat{U}_\ell + \alpha_\ell\hat{U}_\ell. \tag{3.10}$$

From (3.9),

$$\frac{(\alpha_\ell + \frac{2b}{h^2})^2}{(\frac{2b}{h^2})^2} + \frac{\beta_\ell^2}{(\frac{a}{h})^2} = 1, \tag{3.11}$$

it follows that the coefficient pairs $(\alpha_\ell, \beta_\ell)$ lie on an ellipse. See Figure 3.1.

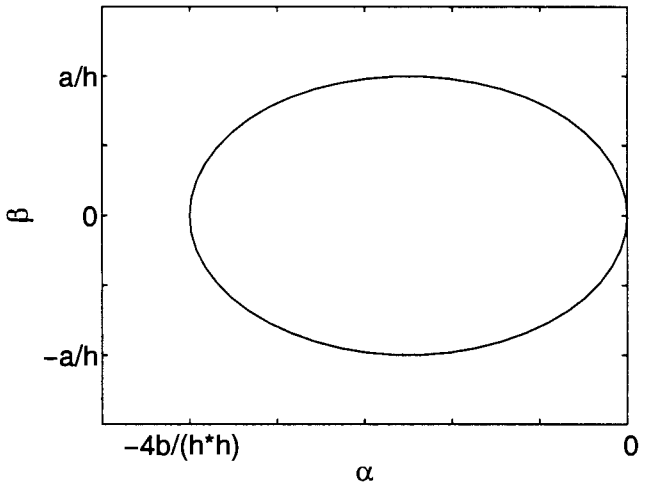


Figure 3.1: Ellipse of (α, β).

The general order-s, s-step multistep IMEX schemes for $\dot{u} = f(u) + g(u)$ have the form

$$\frac{1}{k}U^{n+s} + \frac{1}{k}\sum_{j=0}^{s-1} a_j U^{n+j} = \sum_{j=0}^{s-1} b_j f(U^{n+j}) + \sum_{j=0}^{s} c_j g(U^{n+j}), \tag{3.12}$$

where a_j, b_j and c_j are constant coefficients, and k is the temporal step-size.

Applying the IMEX scheme (3.12) to (3.10), yields

$$\frac{1}{k}\hat{U}_\ell^{n+s} + \frac{1}{k}\sum_{j=0}^{s-1} a_j \hat{U}_\ell^{n+j} = \sum_{j=0}^{s-1} b_j i\beta_\ell \hat{U}_\ell^{n+j} + \sum_{j=0}^{s} c_j \alpha_\ell \hat{U}_\ell^{n+j}, \tag{3.13}$$

which is a linear difference equation with constant coefficients for each fixed wavenumber ℓ.

The solution to (3.13) is given by (see Bender & Orszag [3])

$$\hat{U}_\ell^{n+s} = p_1 r_1^{n+s} + p_2 r_2^{n+s} + \cdots + p_s r_s^{n+s}, \tag{3.14}$$

where r_i is the ith root of the characteristic polynomial $\rho(r)$ defined by

$$\rho(r) = (1 - kc_s\alpha_\ell)r^s + \sum_{j=0}^{s-1}(a_j - ikb_j\beta_\ell - kc_j\alpha_\ell)r^j, \tag{3.15}$$

and p_i is constant when r_i is simple and a polynomial of degree $d-1$ in n for multiple roots r_i with multiple degree of d.

To ensure linear stability, the IMEX schemes (3.12) must satisfy the root conditions, i.e., all of the roots of the characteristic polynomial (3.15) have modulus less than or equal to unity, and those of modulus unity are simple (see Lambert [8]). In practice, determining the time-stepping restrictions is equivalent to finding the largest time-step for which the ellipse in Figure 3.1 lies in the absolute stability region of the IMEX schemes.

We can also obtain another important property for an IMEX scheme. From (3.10) we know the exact solution is given by

$$\hat{U}_\ell(t) = e^{(\alpha_\ell + i\beta_\ell)t}\hat{U}_\ell(0). \tag{3.16}$$

Over one time step

$$|\hat{U}_\ell(t_{n+1})| = e^{\alpha_\ell k}|\hat{U}_\ell(t_n)|, \tag{3.17}$$

which indicates the magnitude of $\hat{U}_\ell$ is amplified by a factor $e^{\alpha_\ell k}$ over a time interval of length k. We plot the corresponding amplification contours in the (α, β) plane in Figure 3.2.

Figure 3.2 suggests that the roots of the characteristic polynomial of (3.15) should be small for large and negative α_ℓ. This fact corresponds to the fast decay property of high frequency modes when b is large.

Next, the linear stability properties for different IMEX schemes are presented. For notational convenience, we simply rewrite (3.10) as

$$\dot{V} = i\beta V + \alpha V. \tag{3.18}$$

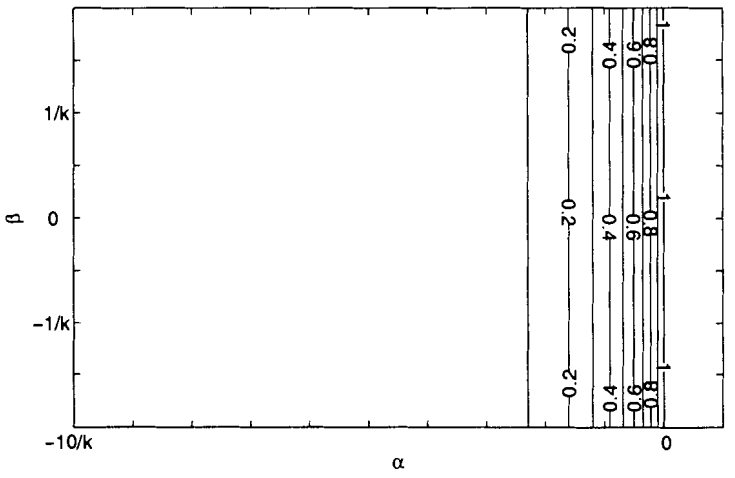


Figure 3.2: Amplification contours over a time interval of length k.

3.1.1 First-Order Methods

To test the linear stability of first-order IMEX schemes, we apply (2.5) to the model test problem $\dot{V} = i\beta V + \alpha V$ to obtain

$$V^{n+1} - V^n = i\beta k V^n + \alpha k \left[(1-\gamma)V^n + \gamma V^{n+1}\right]. \tag{3.19}$$

Solving for V^{n+1}, we get

$$V^{n+1} = \xi(\alpha, \beta) V^n, \tag{3.20}$$

where

$$\xi(\alpha, \beta) = \frac{1 + ik\beta + k(1-\gamma)\alpha}{1 - k\gamma\alpha}. \tag{3.21}$$

The stability region is given by

$$\{(\alpha, \beta) : \ |\xi(\alpha, \beta)| \leq 1\}. \tag{3.22}$$

We plot the stability region contours for the first order SBDF with $\gamma = 1$ in Figure 3.3 , and for two other first-order IMEX schemes, $\gamma = \frac{2}{3}$ and $\gamma = \frac{1}{2}$ in Figure 3.4 and Figure 3.5 respectively.

Among all first-order IMEX schemes, the SBDF1 scheme possesses the largest stability region and the strongest decay for α large and negative. Analytically, we can verify this decay property by letting $\alpha \to -\infty$ in (3.21). This leads to $\xi(-\infty, \beta) = 0$ for $\gamma = 1$ (i.e., SBDF1). For smaller γ values, the corresponding first-order IMEX scheme will not have as strong a decay property and may even be unstable for large b. Hence, among the first order IMEX schemes, the SBDF scheme is often preferred to the others.

Next, we examine the linear stability along the β-axis. For the origin $(\alpha, \beta) = (0, 0)$, we have

$$|\xi(0,0)| = 1. \tag{3.23}$$

This indicates that the origin of the $\alpha - \beta$ plane lies in the stability region. For other $(0, \beta)$ with $\beta \neq 0$,

$$|\xi(0,\beta)| = |1 + ik\beta| = \sqrt{1 + k^2\beta^2} > 1. \tag{3.24}$$

Thus, it follows that first-order IMEX schemes are unstable for $b = 0, a \neq 0$ in (3.1).

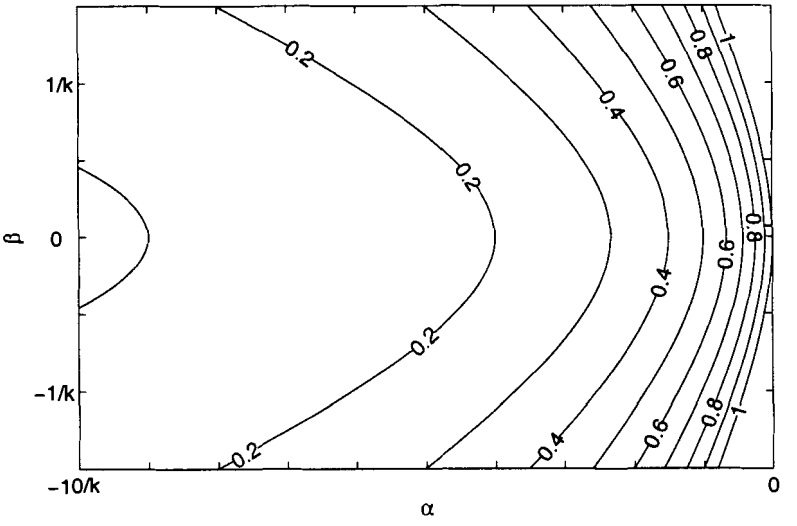


Figure 3.3: Linear stability contour plot for SBDF1 ($\gamma = 1$).

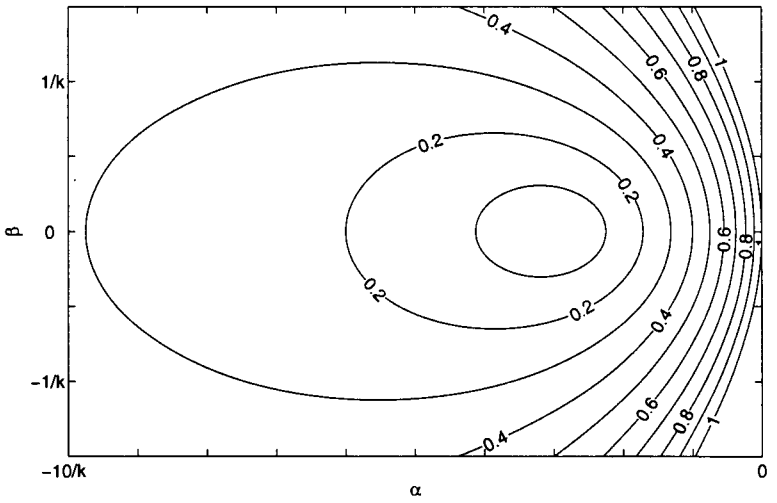


Figure 3.4: Linear stability contour plot for $\gamma = \frac{2}{3}$.

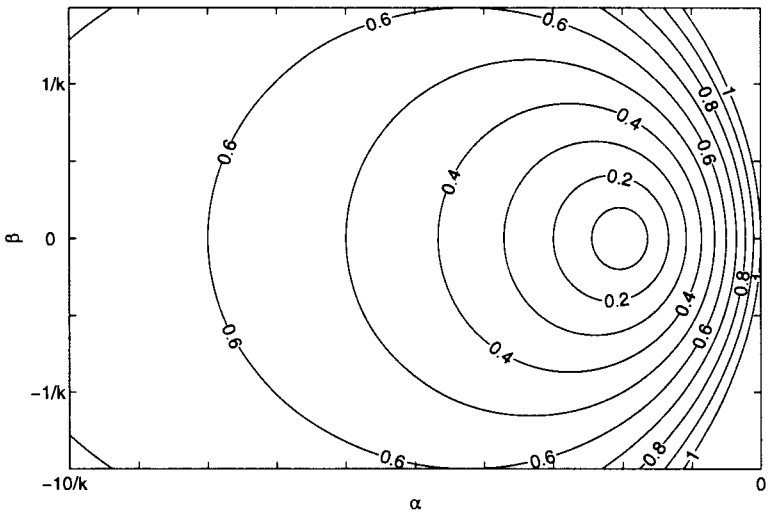


Figure 3.5: Linear stability contour plot for $\gamma = \frac{1}{2}$.

3.1.2 Second-Order Methods

To determine the linear stability contours for second-order methods, we apply equation (2.12) to the model test problem $\dot{V} = i\beta V + \alpha V$. This yields

$$\begin{aligned}\frac{1}{k}\left[(\gamma+\frac{1}{2})V^{n+2} - 2\gamma V^{n+1} + (\gamma-\frac{1}{2})V^{n}\right] = \\ i\beta\left[(\gamma+1)V^{n+1} - \gamma V^{n})\right] + \\ \alpha\left[(\gamma+\frac{c}{2})V^{n+2} + (1-\gamma-c)V^{n+1}) + \frac{c}{2}V^{n}\right],\end{aligned} \tag{3.25}$$

and the corresponding second-degree characteristic polynomial is given by (see Ascher et al. [1])

$$\begin{aligned}\rho(z) &= \left[\gamma+\frac{1}{2} - k\alpha(\gamma+\frac{c}{2})\right]z^2 - [2\gamma + ik\beta(\gamma+1) \\ &\quad + k\alpha(1-\gamma-c)]z + \gamma - \frac{1}{2} + ik\beta\gamma - k\alpha\frac{c}{2}.\end{aligned} \tag{3.26}$$

Letting ξ_1, ξ_2 denote the two roots of $\rho(z)$ in (3.26) for any given pairs (α, β) and (γ, c), we can solve explicitly for ξ_1 and ξ_2 by using the quadratic formula. The stability region for any pair (γ, c) is thus

$$\{(\alpha,\beta)\colon \max\{|\xi_1|, |\xi_2|\} \leq 1\}. \tag{3.27}$$

Some stability contours for particular IMEX schemes discussed in Section 2.3 are presented in Figures 3.6 to 3.10 below.

For the 1-D advection-diffusion equation (3.1), and over the family of second-order, two-step IMEX methods, Ascher et al. [1] demonstrate that the SBDF2 scheme allows the largest stable time step when the discrete diffusion part dominates (i.e. $\frac{b}{ah} > \frac{1}{2}$), and that the CNLF scheme allows the largest stable time step when the discrete convection part dominates (i.e. $\frac{b}{ah} < \frac{1}{2}$). Also, the popular CNAB is competitive only when $\frac{b}{ah} \approx \frac{1}{2}$.

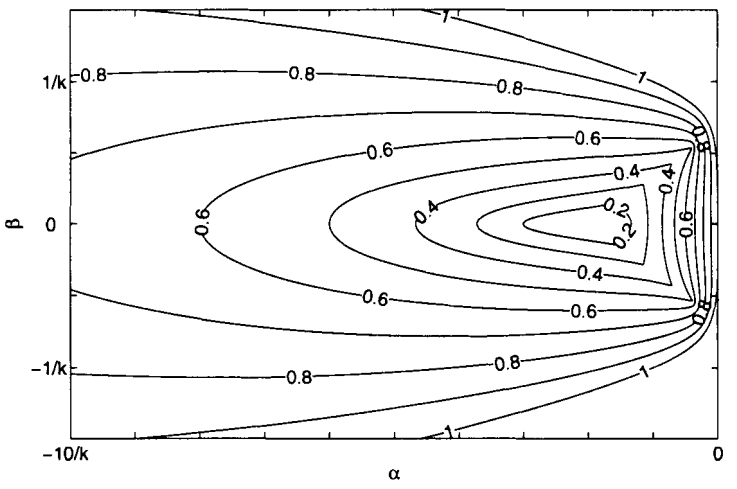


Figure 3.6: Linear stability contours for CNAB ($\gamma = \frac{1}{2}, c = 0$).

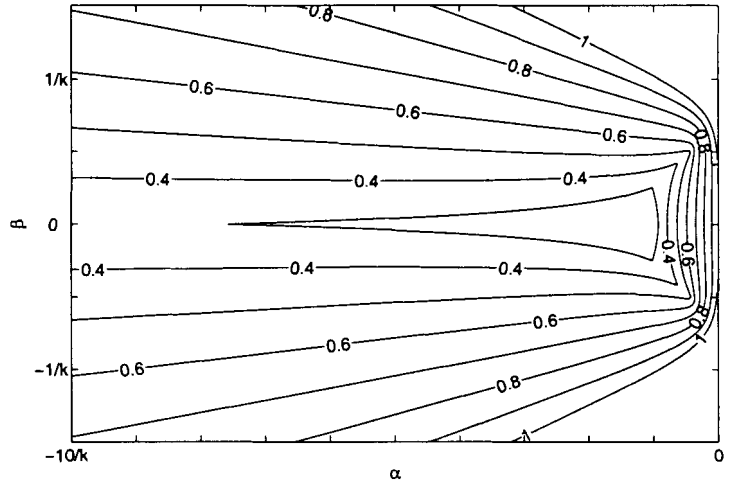


Figure 3.7: Linear stability contours for Modified CNAB ($\gamma = \frac{1}{2}, c = \frac{1}{8}$).

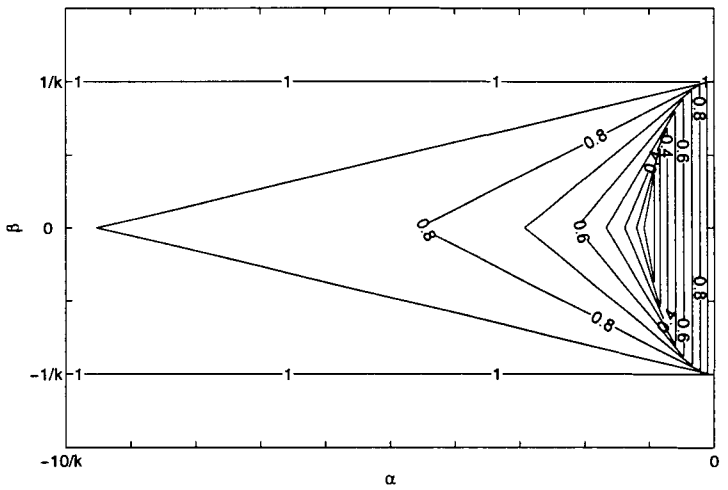


Figure 3.8: Linear stability contours for CNLF ($\gamma = 0, c = 1$).

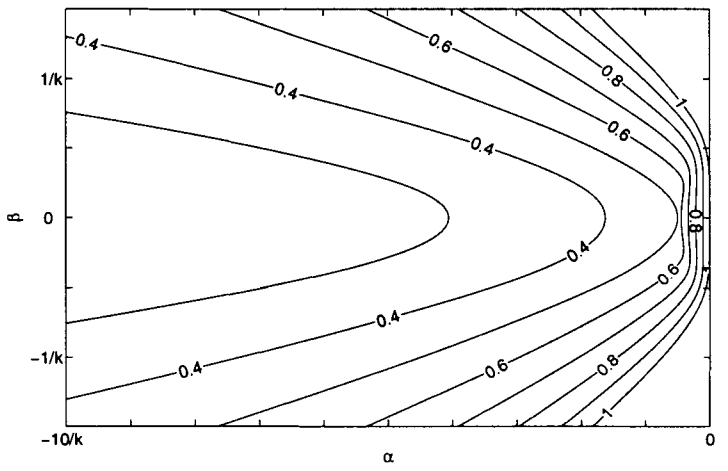


Figure 3.9: Linear stability contours for SBDF ($\gamma = 1, c = 0$).

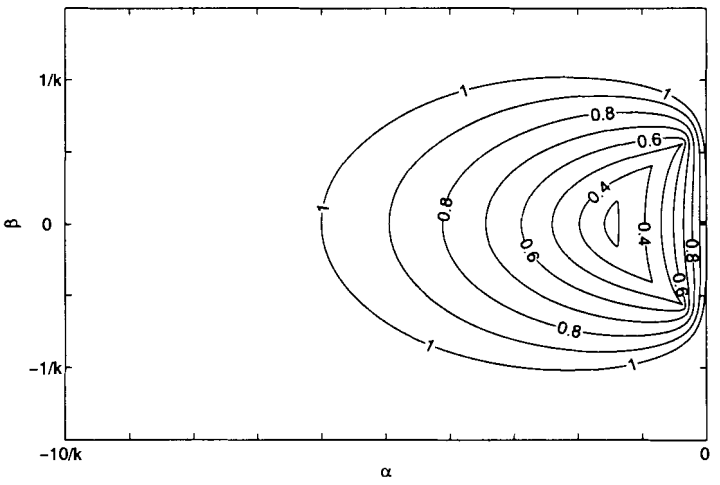


Figure 3.10: Linear stability contours for AMAB ($\gamma = \frac{1}{2}, c = -\frac{1}{6}$).

3.1.3 Third-Order Methods

Applying the general third-order, three-step IMEX scheme (2.26) to the model problem $\dot{V} = i\beta V + \alpha V$, yields

$$\begin{aligned}
&\frac{1}{k}\left[\left(\frac{1}{2}\gamma^2+\gamma+\frac{1}{3}+\theta\right)V^{n+3}+\left(-\frac{3}{2}\gamma^2-2\gamma+\frac{1}{2}-\theta\right)V^{n+2}\right.\\
&\left.+\left(\frac{3}{2}\gamma^2+\gamma-1\right)V^{n+1}+\left(-\frac{1}{2}\gamma^2+\frac{1}{6}\right)V^{n}\right]\\
=\;& i\beta\left[\left(\frac{\gamma^2+3\gamma}{2}+1+\frac{23}{12}\theta\right)V^{n+2}-\left(\gamma^2+2\gamma+\frac{4}{3}\theta\right)V^{n+1}\right.\\
&\left.+\left(\frac{\gamma^2+\gamma}{2}+\frac{5}{12}\theta\right)V^{n}\right]+\alpha\left[\left(\frac{\gamma^2+\gamma}{2}+c\right)V^{n+3}+\left(1-\gamma^2-3c+\frac{23}{12}\theta\right)V^{n+2}\right.\\
&\left.+\left(\frac{\gamma^2-\gamma}{2}+3c-\frac{4}{3}\theta\right)V^{n+1}+\left(\frac{5}{12}\theta-c\right)V^{n}\right]. \qquad (3.28)
\end{aligned}$$

The corresponding third degree characteristic polynomial $\rho(z)$ is given by

$$\begin{aligned}\rho(z) = &\left[\frac{1}{2}\gamma^2+\gamma+\frac{1}{3}+\theta-k\alpha\left(\frac{\gamma^2+\gamma}{2}+c\right)\right]z^3-\left[\frac{3}{2}\gamma^2+2\gamma\right.\\&\left.-\frac{1}{2}+\theta+ik\beta\left(\frac{\gamma^2+3\gamma}{2}+1+\frac{23}{12}\theta\right)+k\alpha\left(1-\gamma^2-3c+\frac{23}{12}\theta\right)\right]z^2\\&+\left[\frac{3}{2}\gamma^2+\gamma-1+ik\beta\left(\gamma^2+2\gamma+\frac{4}{3}\theta\right)-k\alpha\left(\frac{\gamma^2-\gamma}{2}+3c-\frac{4}{3}\theta\right)\right]z\\&-\frac{1}{2}\gamma^2+\frac{1}{6}-ik\beta\left(\frac{\gamma^2+\gamma}{2}+\frac{5}{12}\theta\right)-k\alpha\left(\frac{5}{12}\theta-c\right).\end{aligned} \tag{3.29}$$

Define the three roots of polynomial (3.29) by ξ_1, ξ_2 and ξ_3 for any given pairs (α, β) and (γ, θ, c). The stability region for any pair (γ, θ, c) is thus

$$\{(\alpha,\beta)\colon\ \max\{|\xi_1|,|\xi_2|,|\xi_3|\}\le 1\}\,. \tag{3.30}$$

Some stability contours are displayed in Figures 3.11 to 3.13.

Ascher et al. [1] prove that the third-order SBDF scheme has the strongest asymptotic decay among the third-order IMEX schemes and is stable on the β-axis for sufficiently small β.

The plots for two other interesting third-order methods appear in Figures 3.12 and 3.13. These plots indicate that the corresponding methods possess large stability regions as well as strong high frequency decay.

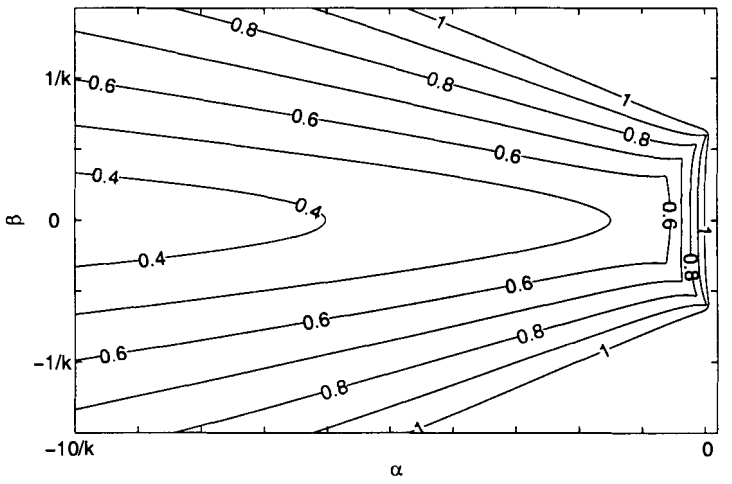


Figure 3.11: Linear stability contours for SBDF3 $(\gamma, \theta, c) = (1, 0, 0)$.

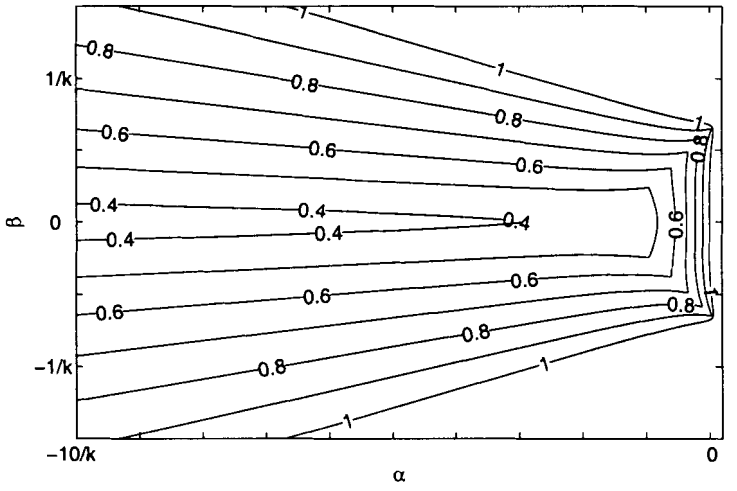


Figure 3.12: Linear stability contours for $(\gamma, \theta, c) = (0, -2.036, -0.876)$.

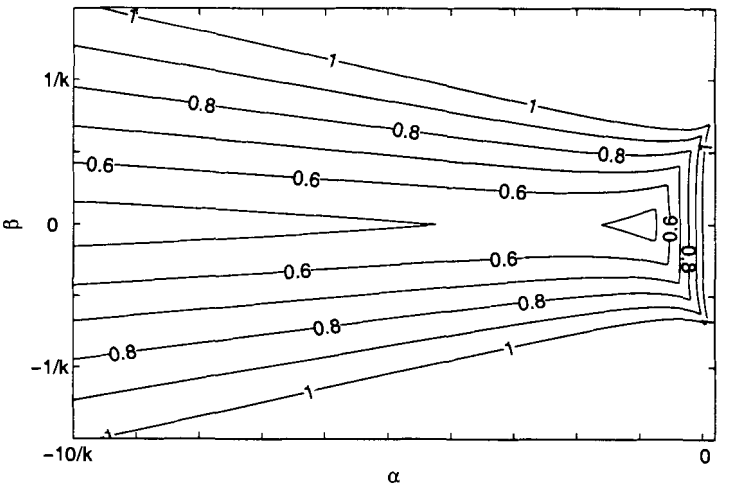


Figure 3.13: Linear stability contours for $(\gamma, \theta, c) = (0.5, -1.25, -0.52)$.

3.1.4 A Fourth-Order Method

As shown in Ascher et al. [1], the general fourth-order, four-step IMEX scheme is a four-parameter family of methods. In that paper, the authors mainly investigate the fourth-order SBDF method. This scheme has good stability properties for solving stiff problems.

Applying (2.31) to our test equation $\dot{V} = i\beta V + \alpha V$, yields

$$\begin{aligned} &\frac{25}{12}V^{n+4} - 4V^{n+3} + 3V^{n+2} - \frac{4}{3}V^{n+1} + \frac{1}{4}V^{n} = \\ &\quad ik\beta[4V^{n+3} - 6V^{n+2} + 4V^{n+1} - V^{n}] + k\alpha V^{n+4}. \end{aligned} \tag{3.31}$$

The corresponding fourth degree characteristic polynomial $\rho(z)$ is given by

$$\begin{aligned} \rho(z) = &\left(\frac{25}{12} - k\alpha\right) z^4 - (4 + i4k\beta)z^3 + (3 + i6k\beta)z^2 \\ &- \left(\frac{4}{3} + i4k\beta\right) z + \frac{1}{4} + ik\beta. \end{aligned} \tag{3.32}$$

Define the four roots of polynomial (3.32) by ξ_1, ξ_2, ξ_3 and ξ_4. The stability region for the SBDF4 scheme is thus

$$\{(\alpha, \beta)\colon \max\{|\xi_1|, |\xi_2|, |\xi_3|, |\xi_4|\} \leq 1\}, \tag{3.33}$$

the linear stability contours are plotted in Figure 3.14.

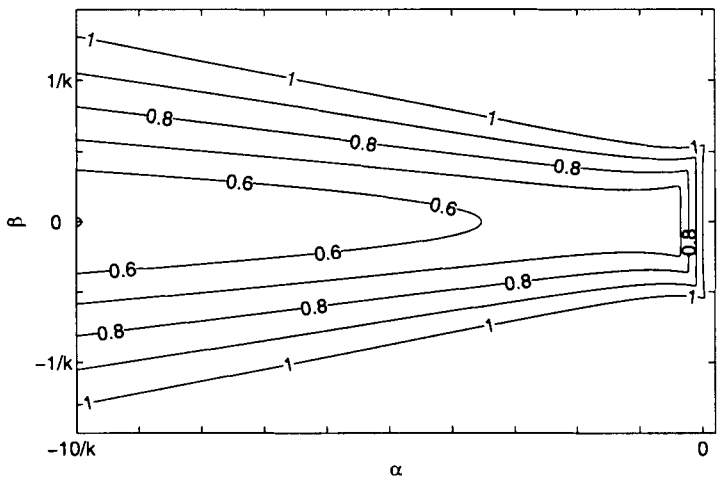


Figure 3.14: Linear stability contours for SBDF4.

We are interested in finding IMEX schemes with good linear stability properties. We also want the corresponding VSIMEX schemes to have good zero-stability properties. Thus we consider the zero-stability of VSIMEX schemes in the next section.

3.2 Zero-Stability Analysis of VSIMEX Schemes

Zero-stability is concerned with the stability of a numerical method in the limit as the step-size goes to zero; that is, zero-stability measures how computational errors, such as errors in the starting values, round-off errors, etc. propagate as the computation proceeds and as the temporal step-size approaches zero.

It is known that a zero-stable numerical scheme is insensitive to perturbations such as round-off errors (see Lambert [8]). Moreover, necessary and sufficient conditions for a linear multistep method to be convergent are that it is both consistent and zero-stable (see Lambert [8]). Thus, zero-stability is an essential property of any usable linear multistep method.

Zero-stability analysis deals with the behaviour of solutions as the step-size approaches zero. This can be interpreted in terms of the numerical solution of the linear

multistep method applied to the differential equation (see Hairer et al. [6])

$$\dot{u} = 0. \tag{3.34}$$

Although this test problem is special in the sense that the right hand side of (3.34) denoted by $F(t, u)$ equals zero, the stability characteristics for this equation determine the stability for the situation when $F(t, u)$ is not identically zero. This is because the solution to the homogeneous equation (3.34) is embedded in the solution to any equation (see Burden and Faires [4]).

The zero-stability analysis of consistent linear multistep methods is carried out using the root conditions. This is described in many reference books, such as Lambert [8], Hairer et al. [6] and Burden and Faires [4], etc.

In this thesis, our goal is to find restrictions on the step-size variations that are required to ensure our VSIMEX schemes are zero-stable. Our presentation mainly follows Hairer et al. [6].

Applying order-s, s-step VSIMEX method (2.1) to the scalar differential equation $\dot{u} = 0$ yields the variable coefficient difference equation (make the coefficient of U^{n+s} equal to 1)

$$U^{n+s} + \sum_{j=0}^{s-1} \alpha_{j,n} U^{n+j} = 0. \tag{3.35}$$

We define the polynomial $\rho_n(z)$ of degree s,

$$\rho_n(z) = z^s + \sum_{j=0}^{s-1} \alpha_{j,n} z^j \tag{3.36}$$

and remark that the consistency condition of our VSIMEX schemes implies

$$\rho_n(1) = 0, \qquad n = 0, 1, \ldots, N - s, \tag{3.37}$$

and N is the number of total nodes in time interval $[0, T]$.

Define the divided polynomials $\rho_n^*(z)$ of degree $s - 1$ (Grigorieff [5])

$$\begin{aligned} \rho_n^*(z) &= \frac{\rho_n(z)}{z - 1} \\ &= z^{s-1} + \sum_{j=0}^{s-2} \alpha_{j,n}^* z^j. \end{aligned} \tag{3.38}$$

From (3.36) and (3.38), we have

$$z^s + \alpha_{s-1,n} z^{s-1} + \ldots + \alpha_{1,n} z + \alpha_{0,n} = (z-1)\left(z^{s-1} + \alpha^*_{s-2,n} z^{s-2} + \ldots + \alpha^*_{1,n} z + \alpha^*_{0,n}\right). \tag{3.39}$$

Equating the coefficients of the corresponding terms yields

$$\begin{aligned} \alpha^*_{s-2,n} &= 1 + \alpha_{s-1,n}, \\ \alpha^*_{0,n} &= -\alpha_{0,n}, \\ \alpha^*_{s-j-1,n} - \alpha^*_{s-j,n} &= \alpha_{s-j,n}, \quad \text{for } j = 2, \cdots, s-1. \end{aligned} \tag{3.40}$$

If we introduce the vector $U_n = (U^{n+s-1}, U^{n+s-2}, \ldots, U^n)^T$, then equation (3.35) becomes

$$\begin{bmatrix} U^{n+s} \\ U^{n+s-1} \\ \vdots \\ U^{n+2} \\ U^{n+1} \end{bmatrix} = \begin{bmatrix} -\alpha_{s-1,n} & -\alpha_{s-2,n} & \cdots & . & -\alpha_{0,n} \\ 1 & 0 & \ldots & . & 0 \\ & 1 & & . & 0 \\ & & \ddots & \vdots & \vdots \\ & & & 1 & 0 \end{bmatrix} \begin{bmatrix} U^{n+s-1} \\ U^{n+s-2} \\ \vdots \\ U^{n+1} \\ U^n \end{bmatrix}. \tag{3.41}$$

For convenience, we write

$$U_{n+1} = A_n U_n, \tag{3.42}$$

where A_n, as in (3.41) is known as the companion matrix. It can be shown that the roots of the polynomial $\rho_n(z)$ are the eigenvalues of the companion matrix A_n.

Similar to A_n, the companion matrix A^*_n associated with the divided polynomial ρ^*_n in (3.38) is given by

$$A^*_n = \begin{bmatrix} -\alpha^*_{s-2,n} & -\alpha^*_{s-3,n} & \cdots & . & -\alpha^*_{0,n} \\ 1 & 0 & \ldots & . & 0 \\ & 1 & \ldots & . & 0 \\ & & \ddots & \vdots & \vdots \\ & & & 1 & 0 \end{bmatrix} \tag{3.43}$$

Definition 3.1. (Hairer et al. [6]) *VSIMEX schemes (2.1) are called zero-stable, if*

$$\|A_{n+m} A_{n+m-1} \cdots A_{n+1} A_n\| < \infty \tag{3.44}$$

for all n and $m \geq 0$ and $\|\cdot\|$ is an appropriate subordinate matrix norm[1].

[1] $\|A\| = \sup \frac{\|Av\|}{\|v\|}$, where vector $v \neq 0$, and $\|AV\|$, $\|v\|$ are vector norm values.

We are also interested in exploring the relationship between A_n and A_n^* defined above. This can be easily done by introducing matrix T and its inverse T^{-1} (Hairer et al. [6], Grigorieff [5])

$$T = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ & 1 & 1 & \dots & 1 \\ & & 1 & \dots & 1 \\ & 0 & & \ddots & \vdots \\ & & & & 1 \end{bmatrix}, \quad T^{-1} = \begin{bmatrix} 1 & -1 & & & 0 \\ & 1 & -1 & & \\ & & 1 & \ddots & \\ & 0 & & \ddots & -1 \\ & & & & 1 \end{bmatrix} \tag{3.45}$$

with dimensions $s \times s$.

A simple calculation leads to

$$T^{-1} A_n T = \begin{bmatrix} A_n^* & 0 \\ e_{s-1}^T & 1 \end{bmatrix}, \tag{3.46}$$

where block matrix A_n^* is defined in (3.43) with dimension $(s-1) \times (s-1)$; $e_{s-1}^T = [0, \cdots, 0, 1]$ with dimension $1 \times (s-1)$; the zero block matrix has dimension $(s-1) \times 1$ and 1 is just a scalar.

The next theorem is taken from Hairer et al. [6] and Grigorieff [5].

Theorem 3.2. *The order-s, s-step VSIMEX scheme (2.1) is zero-stable if and only if the following two conditions are satisfied for all n and $m \geq 0$:*

$$\begin{array}{ll} (a) & \left\| A_{n+m}^* \cdots A_{n+1}^* A_n^* \right\| < \infty \\ (b) & \left\| e_{s-1}^T \sum_{j=n}^{n+m} \left(\prod_{i=n}^{j-1} A_i^* \right) \right\| < \infty. \end{array} \tag{3.47}$$

Proof. From (3.46), we see

$$A_n = T \begin{bmatrix} A_n^* & 0 \\ e_{s-1}^T & 1 \end{bmatrix} T^{-1}, \tag{3.48}$$

so

$$A_{n+1} A_n = T \begin{bmatrix} A_{n+1}^* A_n^* & 0 \\ e_{s-1}^T (A_n^* + I) & 1 \end{bmatrix} T^{-1}. \tag{3.49}$$

By mathematical induction, we can deduce

$$A_{n+m} \cdots A_{n+1} A_n = T \begin{bmatrix} A_{n+m}^* \cdots A_{n+1}^* A_n^* & 0 \\ e_{s-1}^T \sum_{j=n}^{n+m} \left(\prod_{i=n}^{j-1} A_i^* \right) & 1 \end{bmatrix} T^{-1}, \tag{3.50}$$

and therefore, the assertion follows from Definition 3.1. □

An important attribute of Theorem 3.2 lies in that the dimension of the matrices under consideration is reduced by one (Hairer et al. [6]). This property is especially useful for the zero-stability analysis of second-order, two-step VSIMEX methods which is provided in Section 3.2.2.

From the above discussion, we have the following observations:

- The purpose of introducing the companion matrix A_n in (3.41) is to conveniently set up the framework for the zero-stability analysis. By the recurrence relation (3.42), we can easily get

$$U_{n+m+1} = A_{n+m}A_{n+m-1}\cdots A_{n+1}A_nU_n. \tag{3.51}$$

  To measure whether computational errors (or perturbations) in U_n are under control as the computation proceeds and the step-size approaches zero, we simply require that $\|A_{n+m}\cdots A_{n+1}A_n\|$ is bounded for some suitably chosen matrix norm. This naturally leads to the Definition 3.1.

- In general, the companion matrix A_n^* involves the step-size ratios $\omega_j, j = n+1,\ldots,n+s-1$. The variable coefficients $\alpha_{j,n}^*, j = 0,\ldots,s-2$ are functions of step-size ratios, hence the Theorem 3.2 will impose restrictions on these values ω_j in order to ensure zero-stability.

- An appropriate matrix norm for companion matrices A_n or A_n^* must be chosen in order to apply the Definition 3.1 and Theorem 3.2. As we shall show in Sections 3.2.3 and 3.2.4, theoretical restrictions on the step-size ratios for third- and fourth-order VSIMEX schemes vary according to which matrix norm is adopted. This implies that the allowable step-size variations should be computed with respect to some suitably chosen matrix norms.

In the following sections, we will use the Definition 3.1 and Theorem 3.2 to derive the zero-stability restrictions on step-size ratios for order-s, s-step VSIMEX schemes ($s = 1, 2, 3$ and 4).

3.2.1 First-Order VSIMEX Schemes

To test the zero-stability property, apply (2.5) to $\dot{u} = 0$. This yields

$$U^{n+1} - U^n = 0 \tag{3.52}$$

and the companion matrix A_n defined in Section 3.2 is simply

$$A_n = [1] \tag{3.53}$$

for all n.

Thus Definition 3.1 is satisfied for schemes (2.5) and first-order VSIMEX schemes (2.5) are zero-stable for any step-size sequence.

3.2.2 Second-Order VSIMEX Schemes

Applying second-order, two-step VSIMEX schemes (2.10) to $\dot{u} = 0$ and rearranging so that the coefficients of U^{n+2} equals 1 yields

$$U^{n+2} + \lambda_{1,n}U^{n+1} + \lambda_{0,n}U^n = 0, \tag{3.54}$$

where

$$\begin{aligned} \lambda_{1,n} &= \frac{(1+\omega_{n+1})[(1-2\gamma)\omega_{n+1} - 1]}{1 + 2\gamma\omega_{n+1}}, \\ \lambda_{0,n} &= \frac{(2\gamma - 1)\omega_{n+1}^2}{1 + 2\gamma\omega_{n+1}}. \end{aligned}$$

Rewriting (3.54) in matrix-vector form by introducing vector $U_{n+1} = (U^{n+2}, U^{n+1})^T$,

$$\begin{bmatrix} U^{n+2} \\ U^{n+1} \end{bmatrix} = \begin{bmatrix} -\lambda_{1,n} & -\lambda_{0,n} \\ 1 & 0 \end{bmatrix} \begin{bmatrix} U^{n+1} \\ U^n \end{bmatrix}. \tag{3.55}$$

Thus, we have

$$U_{n+1} = A_n U_n, \tag{3.56}$$

where the companion matrix is

$$A_n = \begin{bmatrix} -\lambda_{1,n} & -\lambda_{0,n} \\ 1 & 0 \end{bmatrix}. \tag{3.57}$$

Also, the companion matrix A_n^* for the divided polynomial ρ_n^* of degree one is just

$$\begin{aligned} A_n^* &= [-\lambda_{0,n}^*] \\ &= [\lambda_{0,n}]. \end{aligned} \tag{3.58}$$

In this case Theorem 3.2 takes a simple form, which is presented in Theorem 3.3 (Grigorieff [5]).

Theorem 3.3. *The second-order, two-step VSIMEX scheme (2.10) is zero-stable if and only if the following two conditions are satisfied for all n and $m \geq 0$,*

$$\begin{aligned} &(a) \quad \sup\left(|\lambda_{0,n+m} \cdots \lambda_{0,n+1}\lambda_{0,n}|\right) < \infty \\ &(b) \quad \sup\left(|1 + \lambda_{0,n} + \lambda_{0,n}\lambda_{0,n+1} + \cdots + \lambda_{0,n}\lambda_{0,n+1} \cdots \lambda_{0,n+m}|\right) < \infty. \end{aligned}$$

We observe that if $|\lambda_{0,i}| < 1$ for all $i = n, n+1, \ldots, n+m$ then conditions (a) and (b) in Theorem 3.3 will be satisfied and imply zero-stability.

Recall

$$\lambda_{0,i} = \frac{(2\gamma - 1)\omega_{i+1}^2}{1 + 2\gamma\omega_{i+1}}.$$

Solving for $|\lambda_{0,i}| < 1$, we have the following results:

1. For $0 \leq \gamma < \frac{1}{2}$, if $0 < \omega_{i+1} \leq \frac{1}{1-2\gamma}$, then $|\lambda_{0,i}| \leq 1$, and conditions (a) and (b) will be satisfied.

   In this case, if all $\omega_{i+1}, i = n, n+1, \ldots, n+m$ take the upper bound value $\frac{1}{1-2\gamma}$, then $\lambda_{0,i} = -1$, and condition (b) in Theorem 3.3 will still be satisfied. This can be easily verified from (3.50) and (3.58), since

   $$\begin{bmatrix} \lambda_{0,n+m} \cdots \lambda_{0,n+1}\lambda_{0,n} & 0 \\ \sum_{j=n}^{n+m} \prod_{i=n}^{j-1} \lambda_{0,i} & 1 \end{bmatrix} = \begin{cases} \begin{bmatrix} -1 & 0 \\ 1 & 1 \end{bmatrix} & \text{if m is even} \\ I_{2\times 2} & \text{if m is odd} \end{cases} \tag{3.59}$$

   where $I_{2\times 2}$ is an identity matrix.

2. For $\frac{1}{2} < \gamma \leq 1$, if $0 < \omega_{i+1} < \frac{\gamma + \sqrt{\gamma^2 + 2\gamma - 1}}{2\gamma - 1}$, then $|\lambda_{0,i}| < 1$, and conditions (a) and (b) will be satisfied.

   In this case, to ensure zero-stability, ω_{i+1} can not take the upper bound value for all i, otherwise $\lambda_{0,i} = 1$ and condition (b) in Theorem 3.2 will be violated.

3. For $\gamma = \frac{1}{2}$, VSIMEX is zero-stable for any step-size sequence.

   In this case, since $\lambda_{0,i} = 0$ for all i, there is no restriction on ω_{i+1}.

In summary, we present the following Corollary 3.4,

Corollary 3.4. *Consider the family of second-order, two-step VSIMEX schemes (2.10) with two parameters (γ, c). Suppose*

$$\begin{cases} 0 < \omega_{i+1} \leq \frac{1}{1-2\gamma} & \texttt{if} \quad 0 \leq \gamma < \frac{1}{2} \\ 0 < \omega_{i+1} < \frac{\gamma+\sqrt{\gamma^2+2\gamma-1}}{2\gamma-1} & \texttt{if} \quad \frac{1}{2} < \gamma \leq 1 \end{cases}$$

for all $i > 0$. Then the underlying VSIMEX scheme is zero-stable. In particular, VSIMEX schemes with $\gamma = \frac{1}{2}$ are zero-stable for any step-size sequence.

We plot the ratios of step-size ω versus the parameter γ in Figure 3.15 based on Corollary 3.4. Figure 3.16 shows magnified views of the maximum ratios for small and large γ.

In summary, we make two comments on the step-size ratio restrictions for ensuring zero-stability for second-order, two-step VSIMEX schemes. First, these are analytical results. Second, the results are independent of the choice of matrix norms for the companion matrix A_n in Definition 3.1. As shown in the Sections 3.2.3 and 3.2.4, zero-stability results for third- and fourth-order VSIMEX schemes are quite different. Those results will be numerical and will depend on the choice of matrix norm.

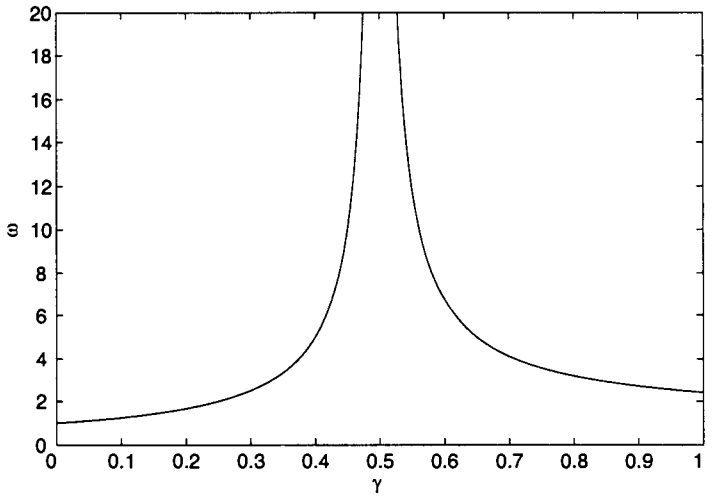
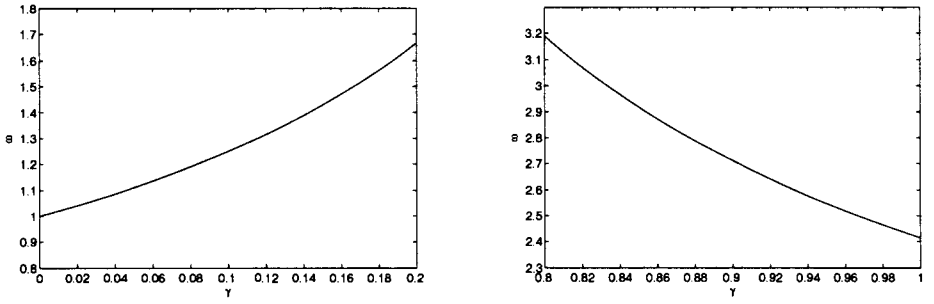


Figure 3.15: Maximum step-size ratio ω vs parameter γ for second-order, two-step VSIMEX schemes (2.10).



(a) Small γ case

(b) Large γ case

Figure 3.16: Maximum step-size ratio ω vs small and large parameter γ for second-order, two-step VSIMEX schemes (2.10).

3.2.3 Third-Order VSIMEX Schemes

Similar to the zero-stability analysis in the second-order, two-step VSIMEX case, we apply our third-order, three-step VSIMEX schemes (2.24) to $\dot{u} = 0$, and rearrange so that the coefficient of U^{n+3} equals 1, which results in

$$U^{n+3} + \sum_{j=0}^{2} \lambda_{j,n} U^{n+j} = 0, \qquad n = 0, 1, 2, \ldots, N-3, \tag{3.60}$$

where

$$\lambda_{j,n} = \frac{\alpha_{j,n}}{\alpha_{3,n}}, \qquad j = 0, 1, 2, \quad n = 0, 1, 2, \ldots, N-3, \tag{3.61}$$

N is the number of total nodes in time interval $[0, T]$ and $\alpha_{j,n}$, $\alpha_{3,n}$ are defined in (2.25).

From (3.43), the companion matrix A_n^* corresponding to the reduced characteristic polynomial $\rho_n^*(z)$ for the nth-step of (3.60) is

$$A_n^* = \begin{bmatrix} -\lambda_{1,n}^* & -\lambda_{0,n}^* \\ 1 & 0 \end{bmatrix}, \tag{3.62}$$

where

$$\begin{aligned} \lambda_{1,n}^* &= 1 + \lambda_{2,n}, \\ \lambda_{0,n}^* &= -\lambda_{0,n}. \end{aligned} \tag{3.63}$$

To apply Theorem 3.2, a suitable matrix norm for A_n^* has to be chosen. It is clear that if

$$\left\| A_j^* \right\| \leq 1, \quad \text{for } j = n, n+1, \cdots \tag{3.64}$$

then Theorem 3.2 implies that the underlying third-order VSIMEX scheme is zero-stable. Thus, the allowable zero-stable step-size ratios will be dependent on the chosen matrix norm. Also, because there are two step-size ratios ω_{n+1}, ω_{n+2} involved in the computation of U^{n+3} in (2.24), it is difficult to derive analytically a sufficient condition for the restrictions on the step-size ratios. Thus, we numerically derive restrictions on the step-size ratios.

In the following, we choose some matrix norms for A_n^* and calculate the corresponding zero-stable step-size ratios based on $\left\| A_j^* \right\| \leq 1$ for all j.

- ∞-norm

  If the ∞-norm is used, then $\|A_n^*\|_\infty \leq 1$ implies

  $$|\lambda_{1,j}^*| + |\lambda_{0,j}^*| < 1 \quad \text{for all } j. \tag{3.65}$$

  By solving this inequality, we can find permissible step-size ratios for the different third-order, three-step VSIMEX schemes. These are shown in Table 3.1. We also compute the largest values R such that for constant step-size ratios $\omega \in [0, R)$ the magnitudes of all eigenvalues of matrix A_n^* (which is independent of n in this case) are less than one. See Table 3.1. The significances of R are as follows:

  - Third-order, three-step VSIMEX schemes satisfy the root conditions (described in Section 3.1) whenever step-size ratios $\omega \in [0, R)$.
  - R is an upper bound for the zero-stable step-size ratios computed by using different matrix norms in this thesis.

- 2-norm

  We cannot choose the 2-norm for A_j^*, since in the constant step-size case $\omega_j = 1$ for all j, and the characteristic polynomial $\rho_n(z)$ of degree three for VSSBDF3 scheme is given by

  $$\rho_n(z) = z^3 - \frac{18}{11}z^2 + \frac{9}{11}z - \frac{2}{11}. \tag{3.66}$$

  From (3.40), the companion matrix is

  $$A^* = \begin{bmatrix} \frac{7}{11} & -\frac{2}{11} \\ 1 & 0 \end{bmatrix}, \tag{3.67}$$

  which has 2-norm $\|A^*\|_2 = 1.1894 > 1$. Because SBDF3 is zero-stable, it follows that the 2-norm is not useful for generating the step-size ratio variations.

- Grigorieff matrix norm [5]

  The Grigorieff matrix norm [2] for A_n^* is defined and denoted by

  $$\|A_n^*\|_G \equiv \|Q_2^{-1} A_n^* Q_2\|_\infty, \tag{3.68}$$

[2] This matrix norm was originally designed in the stability analysis of BDFs methods on variable grids in [5].

Table 3.1: Ranges of zero-stable step-size ratios $[q, Q]$ for various third-order, three-step VSIMEX schemes using ∞-norm. † refers to the results in Grigorieff [5].

	$\|A_n^*\|_\infty$		R
(γ, θ, c)	q	Q	
(0, -2.036, -0.876)	0	1.3394	2.1628
(0.5, -1.25, -0.52)	0	1.2992	1.6180
(0.75, -0.43, -0.17)	0	1.2481	1.8667
VSSBDF3 (1, 0, 0)	0	1.0954	1.6180
$(1, 0, 0)^\dagger$	0	1.08	1.618
(1.5, 3.04, 1.26)	0	1.1915	1.6379
($\sqrt{3}$, 5.075, 2.105)	0	1.2151	1.6437
(1.75, 5.14, 2.13)	0	1.2112	1.6391
(2, 7.72, 3.2)	0	1.2283	1.6424
(2.5, 14.02, 5.81)	0	1.2485	1.6438

where A_n^* is defined in (3.62), and the 2×2 matrix Q_2 is chosen so that its columns form a basis of eigenvectors for A_n^* for equispaced grids.

Let $\tau_{j3}, j = 1, 2$ denote the roots of $\rho_n^*(z)$ for the equispaced grid case. Then the vector $u_{j3} = \begin{bmatrix} \tau_{j3}, 1 \end{bmatrix}^T$ is the corresponding eigenvector, a fact which can be easily verified by the definition of eigenvalues and eigenvectors of matrices[3]. Thus

$$Q_2 = \begin{bmatrix} \tau_{13} & \tau_{23} \\ 1 & 1 \end{bmatrix}, \tag{3.69}$$

and

$$Q_2^{-1} = \frac{1}{\tau_{13} - \tau_{23}} \begin{bmatrix} 1 & -\tau_{23} \\ -1 & \tau_{13} \end{bmatrix}. \tag{3.70}$$

Also direct calculation yields

$$Q_2^{-1} A_n^* Q_2 = \begin{bmatrix} \tau_{13} + \frac{\rho_n^*(\tau_{13})}{\tau_{23} - \tau_{13}} & \frac{\rho_n^*(\tau_{23})}{\tau_{23} - \tau_{13}} \\ \frac{\rho_n^*(\tau_{13})}{\tau_{13} - \tau_{23}} & \tau_{23} + \frac{\rho_n^*(\tau_{23})}{\tau_{13} - \tau_{23}} \end{bmatrix}. \tag{3.71}$$

Defining

$$\Pi_{k3} = (\tau_{l3} - \tau_{k3})^{-1}, \qquad k, l = 1, 2 \text{ and } k \neq l, \tag{3.72}$$

[3]The second subscript 3 in τ_{j3} and u_{j3} is used to indicate that the arguments in this section are associated with the third-order, three-step VSIMEX schemes. In fact, the same arguments also work for the other order-m, m-step VSIMEX schemes.

and making use of (3.71), we derive

$$\begin{aligned} \|A_n^*\|_G &= \|Q_2^{-1} A_n^* Q_2\|_\infty \\ &= \max\{|\tau_{k3} + \rho_n^*(\tau_{k3}) \cdot \Pi_{k3}| + |\Pi_{k3}| \cdot |\rho_n^*(\tau_{l3})|, \quad k,l = 1,2 \ \ \& \ \ k \neq l\}. \end{aligned} \tag{3.73}$$

Imposing $\|A_n^*\|_G \leq 1$ yields the range of zero-stable step-size ratios. Numerical results are shown in Table 3.2 for various third-order, three-step VSIMEX schemes.

In [5], two upper bounds K_{n3}^*, K_{n3} for $\|A_n^*\|_G$ are also given and used as two criteria to generate step-size ratio restrictions. In this thesis, we derive these two upper bounds, and calculate the ranges of zero-stable step-size ratios by imposing $K_{n3}^* \leq 1$ and $K_{n3} \leq 1$. Our numerical results show that the largest range of zero-stable step-size ratios is generated by carrying out the analysis using $\|A_n^*\|_G$. The next largest range is occurred for K_{n3}^* and the smallest one for K_{n3}. See Table 3.2.

The derivation of the upper bound K_{n3}^* for $\|A_n^*\|_G$ proceeds as follows:

Let ρ^* denote the characteristic polynomial ρ_n^* for the equispaced grids,

$$\rho^*(z) = z^2 + \lambda_1^* z + \lambda_0^*. \tag{3.74}$$

Since

$$\rho^*(\tau_{k3}) = 0, \qquad k = 1,2, \tag{3.75}$$

it follows that

$$\begin{aligned} \rho_n^*(\tau_{k3}) &= \rho_n^*(\tau_{k3}) - \rho^*(\tau_{k3}) \\ &= (\lambda_{1n}^* - \lambda_1^*)\tau_{k3} + \lambda_{0n}^* - \lambda_0^*. \end{aligned} \tag{3.76}$$

Similarly,

$$\rho_n^*(\tau_{l3}) = (\lambda_{1n}^* - \lambda_1^*)\tau_{l3} + \lambda_{0n}^* - \lambda_0^*. \tag{3.77}$$

Thus

$$\begin{aligned} \|A_n^*\|_G &= \|Q_2^{-1} A_n^* Q_2\|_\infty \\ &= \max\{|\tau_{k3} + \rho_n^*(\tau_{k3}) \cdot \Pi_{k3}| + |\Pi_{k3}| \cdot |\rho_n^*(\tau_{l3})|, \quad k,l = 1,2, \quad k \neq l\} \end{aligned} \tag{3.78}$$

$$\leq \max\{|\tau_{k3}| + |\Pi_{k3}| \cdot (|\rho_n^*(\tau_{k3})| + |\rho_n^*(\tau_{l3})|)\} \tag{3.79}$$

$$\leq \max\{|\tau_{k3}| + |\Pi_{k3}| \sum_{l=1}^{2} \sum_{r=0}^{1} |\lambda_{rn}^* - \lambda_r^*| \cdot |\tau_{l3}|^r, \quad k = 1,2\} = K_{n3}^* \tag{3.80}$$

using $\rho_n = (z-1)\rho_n^*$, we can easily obtain the other upper bound K_{n3} for $\|A_n^*\|$:

Recall

$$\begin{aligned} \rho_n^*(\tau_{k3}) &= \frac{\rho_n(\tau_{k3})}{\tau_{k3}-1}, \\ &= \frac{\tau_{k3}^3 + \sum_{j=0}^{2} \lambda_{jn} \cdot \tau_{k3}^j}{\tau_{k3}-1}. \end{aligned} \tag{3.81}$$

Thus

$$\begin{aligned} \rho_n^*(\tau_{k3}) &= \rho_n^*(\tau_{k3}) - \rho^*(\tau_{k3}) \\ &= \frac{1}{\tau_{k3}-1}\left[\sum_{j=0}^{2} \tau_{k3}^j \cdot (\lambda_{jn} - \lambda_j)\right], \end{aligned} \tag{3.82}$$

and

$$|\rho_n^*(\tau_{k3})| \le \frac{1}{|\tau_{k3}-1|}\left\{ \sum_{r=0}^{2} |\lambda_{rn} - \lambda_r| \cdot |\tau_{k3}|^r, \quad k = 1,2 \right\}. \tag{3.83}$$

Hence, from (3.80), we derive

$$\|A_n^*\|_G \le \max\left\{ |\tau_{k3}| + |\Pi_{k3}| \cdot \sum_{l=1}^{2}\sum_{r=0}^{2} |\lambda_{rn} - \lambda_r| \cdot \frac{|\tau_{l3}|^r}{|1-\tau_{l3}|}, \quad k = 1,2 \right\} = K_{n3} \tag{3.84}$$

This verifies the formulas 5.(16) and 5.(18) for $m = 3$ appearing in Grigorieff [5].

We conclude by calculating the restrictions required to achieve zero-stable step-size ratios ω_{n+1} and ω_{n+2} based on the Grigorieff matrix norm and its two upper bounds. To proceed we use (3.78), (3.80) and (3.84) to calculate $\|A_n^*\|_G$, K_{n3}^* and K_{n3}, and plot the contours of these quantities with respect to the step-size ratios for different third-order, three-step VSIMEX schemes. The ranges of step-size ratios which ensure zero-stability can be obtained by these contour plots.

Stability contours for various third-order VSIMEX schemes are given in Figures 3.17 to 3.22 to visualize the stability regions for general variable step-size ratios.

We remark that our study confirms the numerical stability ranges of step-size ratios for the three-step BDFs method in Grigorieff [5]; see Table 3.2.

Table 3.2: Ranges of zero-stable step-size ratios $[q, Q]$ for various third-order, three-step VSIMEX schemes using Grigorieff matrix norm and its two upperbounds. † refers to the results in Grigorieff [5].

	$\|A_n^*\|_G(3.78)$		$K_{n3}^*(3.80)$		$K_{n3}(3.84)$	
(γ, θ, c)	q	Q	q	Q	q	Q
(0, -2.036, -0.876)	0	1.4234	0.7010	1.2111	0.8394	1.1318
(0.5, -1.25, -0.52)	0	1.2631	0.5818	1.1892	0.7887	1.1355
(0.75, -0.43, -0.17)	0	1.3455	0.7343	1.1836	0.8622	1.1123
VSSBDF3 (1, 0, 0)	0	1.2613	0.8351	1.1273	0.9210	1.0693
$(1, 0, 0)^\dagger$	0	1.261	0.836	1.127	0.922	1.069
(1.5, 3.04, 1.26)	0	1.2714	0.7565	1.1612	0.8738	1.1003
($\sqrt{3}$, 5.075, 2.105)	0	1.2738	0.7321	1.1689	0.8597	1.1081
(1.75, 5.14, 2.13)	0	1.2722	0.7361	1.1674	0.8619	1.1068
(2, 7.72, 3.2)	0	1.2735	0.7166	1.1728	0.8509	1.1125
(2.5, 14.02, 5.81)	0	1.2740	0.6900	1.1789	0.8365	1.1193

In summary, we remark that for third-order, three-step VSIMEX schemes

- Analytical results are difficult to obtain due to the fact that two step-size ratios are involved at each step. This leads us to derive numerical results using Theorem 3.2.
- Numerical results are dependent on the (suitably chosen) matrix norm.
- The third-order, three-step VSIMEX with parameters $(\gamma, \theta, c) = (0, -2.036, -0.876)$ possesses the largest (zero-stable) range of step-size ratios of all schemes considered in Table 3.1 and Table 3.2. The constant step-size version of this scheme also has good linear stability as shown in Figure 3.12.

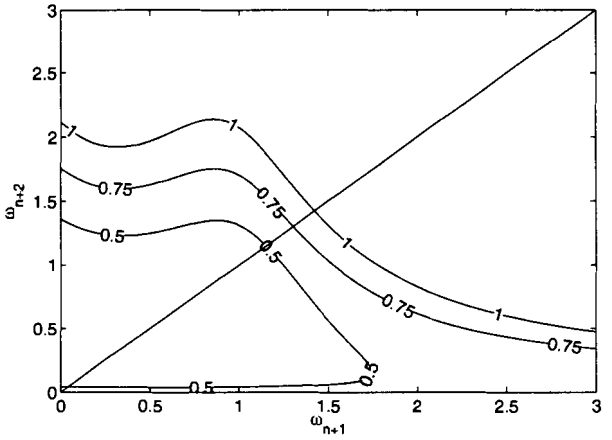


Figure 3.17: Stability contours based on $\|A_n^*\|_G$ (3.78) for the third-order VSIMEX scheme with $(\gamma, \theta, c) = (0, -2.036, -0.876)$.

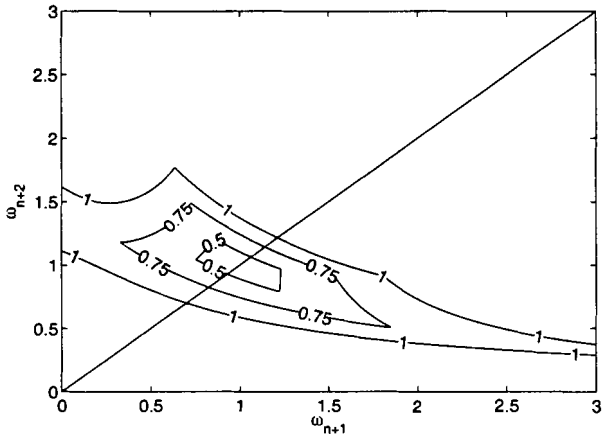


Figure 3.18: Stability contours based on K_{n3}^* (3.80) with $(\gamma, \theta, c) = (0, -2.036, -0.876)$.

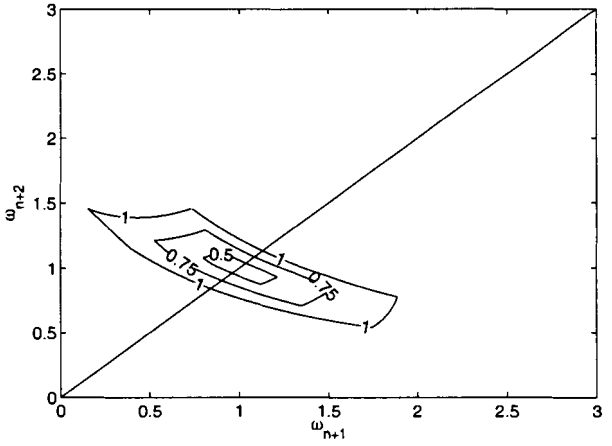


Figure 3.19: Stability contours based on K_{n3} (3.84) with $(\gamma, \theta, c) = (0, -2.036, -0.876)$.

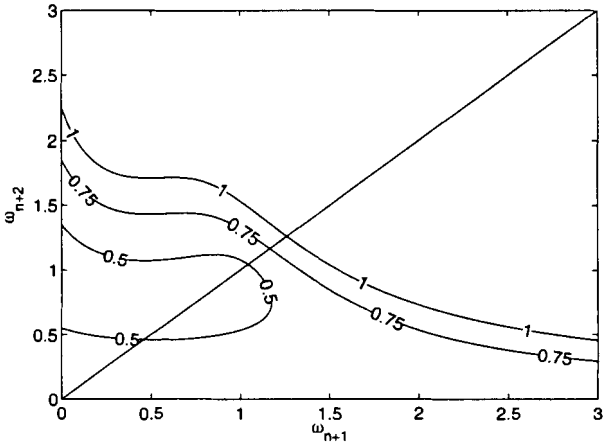


Figure 3.20: Stability contours based on $\|A_n^*\|_G$ (3.78) for VSSBDF3 $(\gamma, \theta, c) = (1, 0, 0)$.

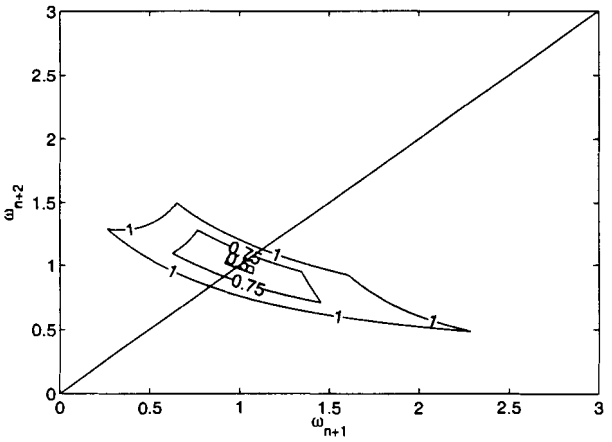


Figure 3.21: Stability contours based on K^*_{n3} (3.80) for VSSBDF3 $(\gamma, \theta, c) = (1, 0, 0)$.

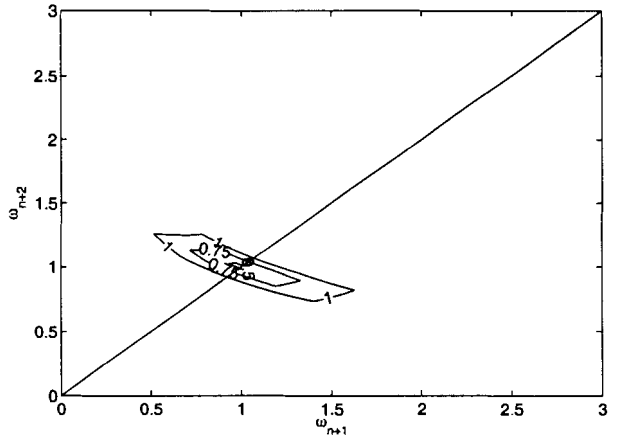


Figure 3.22: Stability contours based on K_{n3} (3.84) for VSSBDF3 $(\gamma, \theta, c) = (1, 0, 0)$.

3.2.4 The VSSBDF4 Scheme

To study the zero-stability of VSSBDF4 scheme, we adopt the ideas developed in Section 3.2.3. Applying our VSSBDF4 scheme (2.29) to $\dot{u} = 0$, and rearranging to make the coefficient of U^{n+4} equal to 1, yields

$$U^{n+4} + \sum_{j=0}^{3} \lambda_{j,n} U^{n+j} = 0, \qquad n = 0, 1, 2, \ldots, N-4, \tag{3.85}$$

where

$$\lambda_{j,n} = \frac{\alpha_{j,n}}{\alpha_{4,n}}, \qquad j = 0, 1, 2, 3 \ \ \& \ \ n = 0, 1, 2, \ldots, N-4, \tag{3.86}$$

N is the number of total nodes in time interval $[0, T]$ and $\alpha_{j,n}$, $\alpha_{4,n}$ are defined in (2.30).

According to (3.36), the characteristic polynomial $\rho_n(z)$ of degree four for the n-th step of (3.85) is given by

$$\rho_n(z) = z^4 + \sum_{j=0}^{3} \lambda_{j,n} z^j, \qquad n = 0, 1, \ldots, N. \tag{3.87}$$

For constant step-sizes,

$$\rho(z) = z^4 - \frac{48}{25} z^3 + \frac{36}{25} z^2 - \frac{16}{25} z + \frac{3}{25}. \tag{3.88}$$

Also

$$\rho(1) = 0 \Longrightarrow \rho(z) = (z-1)\rho^*(z). \tag{3.89}$$

By long division, we have

$$\rho^*(z) = z^3 - \frac{23}{25} z^2 + \frac{13}{25} z - \frac{3}{25}. \tag{3.90}$$

It is clear that the ∞-norm cannot be used to calculate the step-size ratio constraints, since $\|A^*\|_\infty = \frac{23}{25} + \frac{13}{25} + \frac{3}{25} = \frac{39}{25} > 1$ for the (constant step-size) SBDF4 scheme. Here, we adopt Grigorieff matrix norm defined in [5].

Let τ_{i4}, $i = 1, 2, 3$ denote the roots of $\rho^*(z)$. A short calculation shows that τ_{i4} are all different. Similar to the matrix Q_2 in Section 3.2.3, we choose the 3×3 matrix Q_3 such that its columns form a basis of eigenvectors of the companion matrix corresponding to ρ^*, i.e.

$$Q_3 = \begin{bmatrix} \tau_{14}^2 & \tau_{24}^2 & \tau_{34}^2 \\ \tau_{14} & \tau_{24} & \tau_{34} \\ 1 & 1 & 1 \end{bmatrix}. \tag{3.91}$$

Also, G-norm for A_n^* is defined by ([5])

$$\|A_n^*\|_G \equiv \|Q_3^{-1} A_n^* Q_3\|_\infty. \tag{3.92}$$

A straightforward calculation leads to

$$\|A_n^*\|_G = \max\left\{ |\tau_{k4} - \Pi_{k4} \cdot \rho_n^*(\tau_{k4})| + |\Pi_{k4}| \sum_{l \neq k} |\rho_n^*(\tau_{l4})|, \quad k,l = 1,2,3 \ \ \& \ \ k \neq l \right\}, \tag{3.93}$$

where

$$\begin{aligned} \Pi_{k4} &= \prod_{l \neq k} (\tau_{l4} - \tau_{k4})^{-1}, \qquad k,l = 1,2,3, \\ \rho_n^*(\tau_{k4}) &= \tau_{k4}^3 + \sum_{j=0}^{2} \lambda_{j,n}^* \tau_{k4}^j, \qquad n = 0,1,2,\ldots,N-4, \\ \lambda_{2,n}^* &= 1 + \lambda_{3,n}, \\ \lambda_{1,n}^* &= -(\lambda_{0,n} + \lambda_{1,n}), \\ \lambda_{0,n}^* &= -\lambda_{0,n}, \end{aligned} \tag{3.94}$$

and $\lambda_{0,n}$, $\lambda_{1,n}$, and $\lambda_{3,n}$ are defined in (3.86).

Because there are three step-size ratios involved in the computation of $\|A_n^*\|$ in (3.93), we plot the 3-D isosurface of $\|A_n^*\| = 1$ with respect to three step-size ratios $\omega_{n+i}, i = 1,2,3$. See Figure 3.23. Here, it is more challenging to concisely describe the range of zero-stable step-size ratios than it is for the second-order, two-step VSIMEX case described in Section 3.2.2. However, it is straightforward to plot the numerical step-size ratio constraints $[q, Q] = [0.8016, 1.0818]$ for the case of constant step-size ratios. See Figure 3.24 for the corresponding plot.

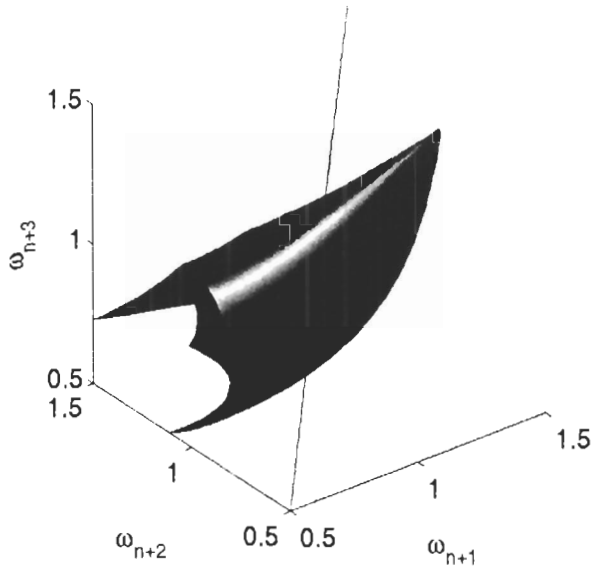


Figure 3.23: Isosurface contour ($\|A_n^*\|_G = 1.0$) for VSSBDF4

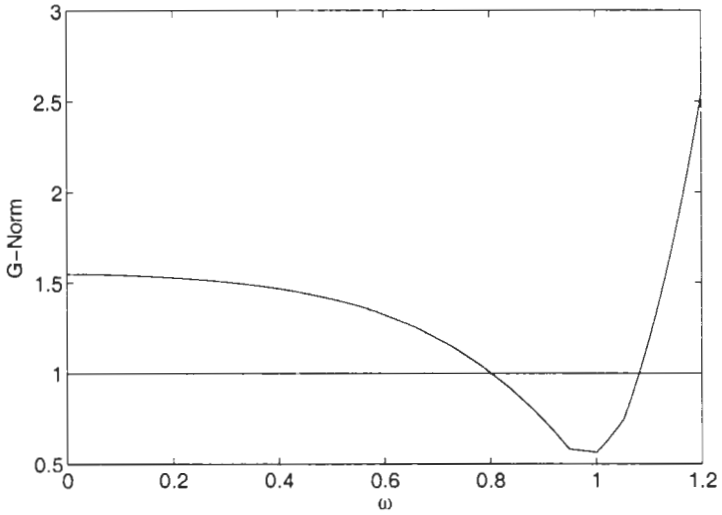


Figure 3.24: Stability region of VSSBDF4 vs the constant step-size ratios

Chapter 4

Numerical Experiments

In the previous two chapters, we presented second-, third- and fourth-order VSIMEX schemes, and studied the stability properties of these schemes and their corresponding IMEX schemes. In this chapter, we carry out numerical experiments which verify the expected orders of convergence of our various VSIMEX schemes. Our test problems are the constant coefficient advection-diffusion equation and Burgers' equation.

To calculate the starting values associated with the different VSIMEX schemes, we use the following methods:

For second-order, two-step VSIMEX schemes, we use the first-order SBDF1 method with a very small temporal step-size.

For the third-order, three-step VSIMEX case, we use the third-order implicit-explicit Runge-Kutta method IMEX RK(3,4,3) presented in Ascher, Ruuth and Spiteri [2]. This method applies a third-order, three-stage diagonally-implicit Runge-Kutta (DIRK) method for the stiff term, and a third-order, four-stage explicit Runge-Kutta (ERK) method for the nonstiff term.

For the fourth-order, four-step VSIMEX case, we use the fourth-order implicit-explicit Additive Runge-Kutta method ARK4(3)6L[2] presented in Kennedy and Carpenter [7]. This method applies a fourth-order, six-stage stiffly-accurate, explicit, singly diagonally implicit Runge-Kutta (ESDIRK) method for the stiff term, and a fourth-order, six-stage ERK method for the nonstiff term.

In the following experiments, we test various second-order, two-step VSIMEX schemes. For third- and fourth-order, we mainly focus on the third-order, three-step VSSBDF3 and fourth-order, four-step VSSBDF4 schemes.

For the advection-diffusion equation, numerical results show that some VSIMEX schemes possess better stability properties than their corresponding IMEX methods. Furthermore, our results demonstrate the superiority of VSIMEX schemes over the classical IMEX schemes in solving Burgers' equation.

4.1 Advection-Diffusion Equation

Our first test problem is the one-dimensional constant coefficient advection-diffusion equation,

$$u_t + cu_x = \lambda u_{xx}, \tag{4.1}$$

subject to periodic boundary conditions on the interval [0, 1] and initial conditions

$$u(x, 0) = \sin(2\pi x), \tag{4.2}$$

where c, λ are constant coefficients.

The analytical solution of this initial-boundary value problem is

$$u(x, t) = \exp(-4\pi^2\lambda t)\sin[2\pi(x - ct)]. \tag{4.3}$$

In the following sections, we present computational results for various VSIMEX schemes applied to this test problem.

4.1.1 Second-Order VSIMEX Schemes

Consider the test problem (4.1) and (4.2) with $c = 1$, $\lambda = \frac{1}{100}$.

In the following computations, we carry out the spatial discretization using central finite difference schemes to approximate u_x and u_{xx}:

$$u_x(x_j, t_n) = \frac{U^n_{j+1} - U^n_{j-1}}{2\Delta x} + O((\Delta x)^2), \tag{4.4}$$

$$u_{xx}(x_j, t_n) = \frac{U^n_{j+1} - 2U^n_j + U^n_{j-1}}{(\Delta x)^2} + O((\Delta x)^2). \tag{4.5}$$

Notice, in theory, second-order accuracy in space is obtained. Here, we take step-size $\Delta x = \frac{1}{1000}$ in space and compute the value u at $t = 1$. To measure the error, the maximum norm of the absolute error is evaluated, i.e., $\|U^N_j - u(x_j, 1)\|_\infty$ or simply

written as $\|U - u_e\|_\infty$, where $u_e(x) \equiv u(x, 1)$. From (4.3), the exact solution u_e at $t = 1$ is given by

$$u_e(x) = \exp(-0.04\pi^2)\sin[2\pi(x-1)]. \tag{4.6}$$

To choose the constant temporal step-size, we simply divide the interval [0, 1] by 100, i.e., $\Delta t = \frac{1}{100}$. To approximate the evolution, we compute the value U at time $t_j = \frac{j}{100}, j = 1, 2, \cdots, 100$.

To illustrate the stability and convergence properties of VSIMEX schemes, we generate variable temporal step-sizes by taking $t_j = (\frac{j}{100})^2$. It is clear that $t_{100} = 1$ and $\Delta t_j = t_{j+1} - t_j = \frac{2j+1}{10000}$.

We summarize the computational results for different second-order VSIMEX schemes in Table 4.1 and Table 4.2. Furthermore, the absolute errors for different IMEX and VSIMEX schemes are plotted in Figure 4.1 for $\Delta x = \frac{1}{1000}$.

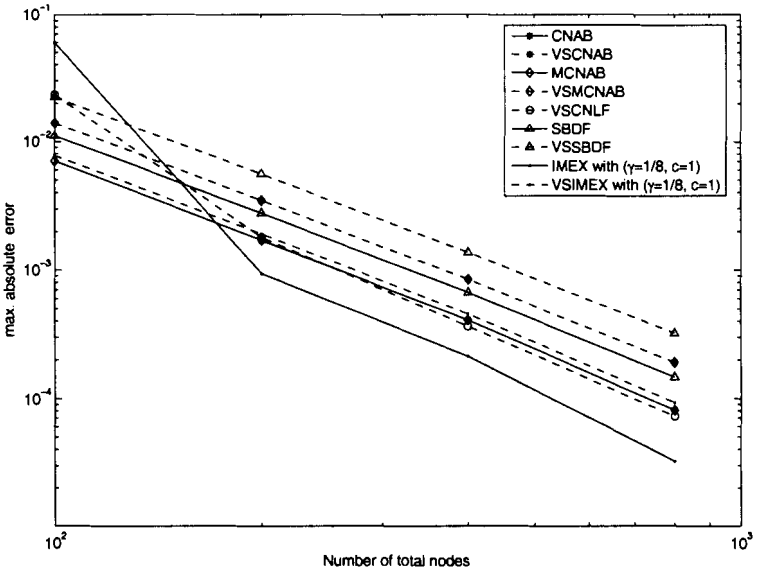


Figure 4.1: Absolute errors for different second-order VSIMEX and IMEX schemes in solving the advection-diffusion equation. The slopes are very close to -2.0 (using logarithmic scales), indicating second-order convergence.

We make the following observations:

- Second-order in time is observed for all the VSIMEX schemes tested.
- Using this particular choice of variable step-size, some VSIMEX schemes have better stability properties than their corresponding IMEX schemes. For example,

Table 4.1: Numerical results for the advection-diffusion equation using various second-order IMEX and VSIMEX schemes ($\Delta x = \frac{1}{1000}$)

		Second-Order IMEX			Second-Order VSIMEX		
Scheme	Nodes	$\|U - u_e\|_\infty$	Ratio	Order	$\|U - u_e\|_\infty$	Ratio	Order
CNAB	100	Unstable			1.40e-2		
$\gamma = \frac{1}{2}$	200	1.71e-3			3.47e-3	4.035	2.01
$c = 0$	400	4.08e-4	4.191	2.07	8.46e-4	4.102	2.04
	800	8.10e-5	5.037	2.33	1.90e-4	4.453	2.15
Modified	100	7.04e-3			1.40e-2		
CNAB	200	1.71e-3	4.117	2.04	3.47e-3	4.035	2.01
$\gamma = \frac{1}{2}$	400	4.08e-4	4.191	2.07	8.46e-4	4.102	2.04
$c = \frac{1}{8}$	800	8.10e-5	5.037	2.33	1.90e-4	4.453	2.15
CNLF	100	Unstable			2.33e-2		
$\gamma = 0$	200	Unstable			1.80e-3	12.944	3.69
$c = 1$	400	Unstable			3.65e-4	4.932	2.30
	800	1.55e-4			7.26e-5	5.028	2.33
SBDF	100	1.11e-2			2.23e-2		
$\gamma = 1$	200	2.76e-3	4.022	2.01	5.55e-3	4.018	2.01
$c = 0$	400	6.69e-4	4.126	2.04	1.37e-3	4.051	2.02
	800	1.46e-4	4.582	2.20	3.21e-4	4.268	2.09
	100	5.96e-2			7.76e-3		
$\gamma = \frac{1}{8}$	200	9.38e-4			1.91e-3	4.063	2.02
$c = 1$	400	2.13e-4	4.404	2.14	4.56e-4	4.189	2.07
	800	3.24e-5	6.574	2.72	9.29e-5	4.909	2.30

Table 4.2: Numerical results for the advection-diffusion equation using various second-order IMEX and VSIMEX schemes ($\Delta x = \frac{1}{2000}$).

Scheme	Nodes	Second-Order IMEX			Second-Order VSIMEX		
		$\|U - u_e\|_\infty$	Ratio	Order	$\|U - u_e\|_\infty$	Ratio	Order
CNAB	100	Unstable			1.41e-2		
$\gamma = \frac{1}{2}$	200	1.73e-3			3.49e-3	4.040	2.01
$c = 0$	400	4.29e-4	4.033	2.01	8.66e-4	4.030	2.01
	800	1.02e-4	4.206	2.07	2.11e-4	4.104	2.04
Modified	100	7.10e-3			1.40e-2		
CNAB	200	1.73e-3	4.104	2.04	3.49e-3	4.012	2.00
$\gamma = \frac{1}{2}$	400	4.29e-4	4.033	2.01	8.67e-4	4.025	2.01
$c = \frac{1}{8}$	800	1.02e-4	4.206	2.07	2.11e-4	4.109	2.04
CNLF	100	Unstable			2.97e-2		
$\gamma = 0$	200	Unstable			4.11e-3	7.226	2.85
$c = 1$	400	Unstable			4.44e-4	9.257	3.21
	800	Unstable			9.09e-5	4.885	2.29
SBDF	100	1.11e-2			2.23e-2		
$\gamma = 1$	200	2.78e-3	3.993	2.00	5.57e-3	4.004	2.00
$c = 0$	400	6.90e-4	4.029	2.01	1.39e-3	4.007	2.00
	800	1.67e-4	4.132	2.05	3.42e-4	4.064	2.02
	100	6.92e-2			7.78e-3		
$\gamma = \frac{1}{8}$	200	9.60e-4			1.93e-3	4.031	2.01
$c = 1$	400	2.34e-4	4.103	2.04	4.77e-4	4.046	2.02
	800	5.33e-5	4.390	2.13	1.14e-4	4.184	2.06

when $\Delta t = \frac{1}{200}$, CNLF scheme is unstable while VSCNLF is stable, and when $\Delta t = \frac{1}{100}$, CNAB is unstable while VSCNAB is stable.

- This particular variable temporal step-size choice does not give optimal accuracy since the IMEX schemes have better accuracy than the VSIMEX schemes for this test problem.

4.1.2 Third- and Fourth-Order VSIMEX Schemes

Consider the test problem

$$u_t + u_x = 0.1u_{xx}, \tag{4.7}$$

subject to the periodic boundary conditions on the interval [-1, 1] and initial condition

$$u(x, 0) = \sin(\pi x). \tag{4.8}$$

Here, we focus on the third-order, three-step VSSBDF3 and fourth-order, four-step VSSBDF4 schemes. For the spatial discretization, we fix $\Delta x = \frac{1}{100}$ and apply fourth-order accurate finite difference approximations of u_x and u_{xx}, i.e. we use the 5-point formulas

$$u_x(x_j, t_n) = \frac{1}{12\Delta x}[U^n_{j-2} - 8U^n_{j-1} + 8U^n_{j+1} - U^n_{j+2}] + O((\Delta x)^4), \tag{4.9}$$

$$u_{xx}(x_j, t_n) = -\frac{1}{12(\Delta x)^2}[U^n_{j-2} - 16U^n_{j-1} + 30U^n_j - 16U^n_{j+1} + U^n_{j+2}] + O((\Delta x)^4). \tag{4.10}$$

Here, we compute the solution u at time $t = 2$. To choose the variable temporal step-size, we first break the interval [0, 2] into 5 subintervals with equal length of 0.4, then split each subinterval into smaller subintervals with different sizes, say subinterval [0, 0.4] is divided by 6, [0.4, 0.8] by 4 etc. See partitioning scheme no. 2 in Table 4.3.

Initially, we take 25 nodes over the time interval [0, 2.0], as shown in Table 4.3. We then double the nodes, while keeping the ratios of nodes between consecutive subintervals unchanged, and continue this partition pattern until we take 400 nodes in time. E.g. for 50 nodes, we partition the time interval [0, 2.0] as shown in Table 4.4.

As mentioned in Section 4.1.1, we use the maximum norm $\|U^N_j - u_e(j\Delta x, 2)\|_\infty$ to measure the computational error.

Table 4.3: Different partitioning schemes for the time interval [0, 2] (total nodes=25)

Scheme No.	$[0, 0.4]$	$[0.4, 0.8]$	$[0.8, 1.2]$	$[1.2, 1.6]$	$[1.6, 2.0]$
1	8	7	3	3	4
2	6	4	3	7	5
3	3	3	4	7	8
4	1	1	5	8	10
5	3	7	2	5	8

Table 4.4: Different partitioning schemes for the time interval [0, 2] (total nodes=50)

Scheme No.	$[0, 0.4]$	$[0.4, 0.8]$	$[0.8, 1.2]$	$[1.2, 1.6]$	$[1.6, 2.0]$
1	16	14	6	6	8
2	12	8	6	14	10
3	6	6	8	14	16
4	2	2	10	16	20
5	6	14	4	10	16

We summarize the computational results for VSSBDF3 in Table 4.5 and for VSSBDF4 in Table 4.6. The corresponding rates of convergence for VSSBDF3 and VSSBDF4 are plotted in Figures 4.2 and 4.3 respectively.

Table 4.5: Numerical results for the advection-diffusion equation using the VSSBDF3 scheme.

Scheme No.	Nodes in time	$\|U - u_e\|_\infty$	Ratio	Order
Constant	25	1.194e-2		
step-size	50	1.520e-3	7.86	2.97
	100	1.895e-4	8.02	3.00
	200	2.362e-5	8.02	3.00
	400	2.962e-6	7.97	3.00
	25	2.986e-2		
	50	3.655e-3	8.17	3.03
1	100	4.527e-4	8.07	3.01
	200	5.614e-5	8.06	3.01
	400	6.999e-6	8.02	3.00
	25	1.974e-2		
	50	2.557e-3	7.72	2.95
2	100	3.212e-4	7.96	2.99
	200	4.016e-5	8.00	3.00
	400	5.032e-6	7.98	3.00
	25	2.560e-2		
	50	3.389e-3	7.55	2.92
3	100	4.365e-4	7.76	2.96
	200	5.515e-5	7.91	2.98
	400	6.937e-6	7.95	2.99
	25	1.490e-2		
	50	4.180e-2		
4	100	7.920e-3	5.28	2.40
	200	1.087e-3	7.29	2.87
	400	1.423e-4	7.64	2.93
	25	2.716e-2		
	50	5.278e-3	5.15	2.36
5	100	7.512e-4	7.03	2.81
	200	9.858e-5	7.62	2.93
	400	1.261e-5	7.82	2.97

Table 4.6: Numerical results for the advection-diffusion equation using the VSSBDF4 scheme.

Scheme No.	Nodes in time	$\|U - u_e\|_\infty$	Ratio	Order
Constant	25	3.389e-3		
step-size	50	2.132e-4	15.90	3.99
	100	1.329e-5	16.04	4.00
	200	8.431e-7	15.76	3.98
	25	1.222e-2		
	50	7.922e-4	15.43	3.95
1	100	4.911e-5	16.13	4.01
	200	3.060e-6	16.05	4.00
	25	7.172e-3		
	50	4.882e-4	14.69	3.88
2	100	3.111e-5	15.69	3.97
	200	1.971e-6	15.78	3.98
	25	9.269e-3		
	50	6.979e-4	13.28	3.73
3	100	4.631e-5	15.07	3.91
	200	2.975e-6	15.57	3.96
	25	2.761e-3		
	50	9.677e-3		
4	100	2.316e-3	4.18	2.06
	200	1.796e-4	12.90	3.69
	25	1.021e-2		
	50	1.337e-3	7.64	2.93
5	100	1.079e-4	12.39	3.63
	200	7.409e-6	14.56	3.86

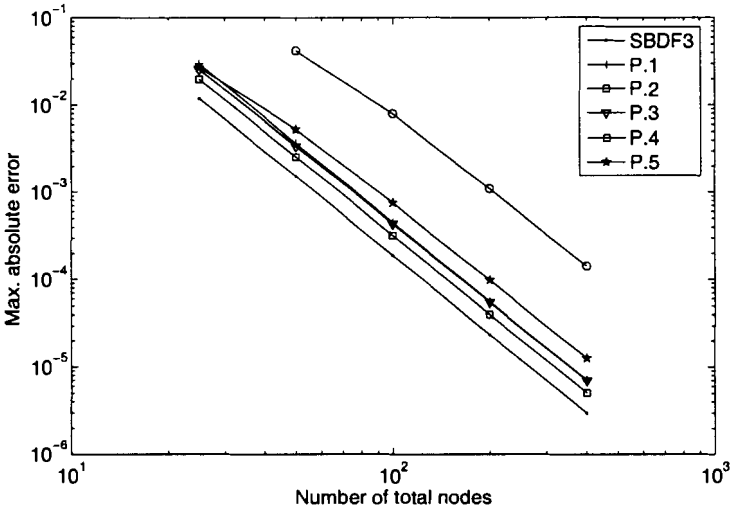


Figure 4.2: Absolute errors for the third-order VSSBDF3 scheme with different partitions in solving the advection-diffusion equation. The slopes are approximately -3.0 (using logarithmic scales), which indicates third-order convergence.

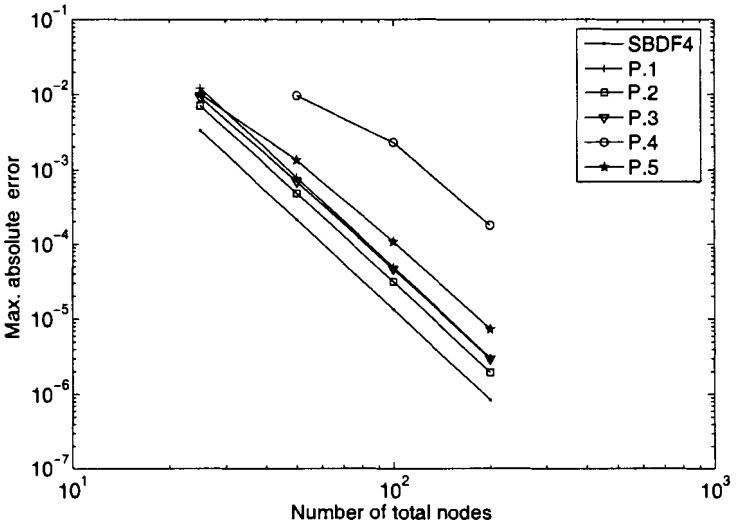


Figure 4.3: Absolute errors for the fourth-order VSSBDF4 scheme with different partitions in solving the advection-diffusion equation. The slopes are approximately -4.0 (using logarithmic scales), which indicates fourth-order convergence.

4.2 Burgers' Equation

Our second test problem is the one-dimensional Burgers' equation

$$u_t + uu_x = \lambda u_{xx}, \tag{4.11}$$

subject to periodic boundary conditions on the interval [-1, 1] and initial conditions

$$u(x, 0) = \sin(\pi x), \tag{4.12}$$

where λ is a constant coefficient.

In the following subsections, we present the computational results from applying different orders of VSIMEX schemes to this test problem.

4.2.1 Second-Order VSIMEX Schemes

Consider the test problem (4.11) and (4.12) with $\lambda = \frac{1}{100}$.

For the spatial discretization, as in Section 4.1.1, we apply the second-order accurate finite difference approximations of u_x and u_{xx}. See equations (4.4) and (4.5).

In the following experiments, we compute the solution u at time $t = 2$ and fix $\Delta x = \frac{1}{25000}$. To approximate the exact solution, u_e, we use SBDF2 scheme with a very small temporal step-size $\Delta t = \frac{1}{25000}$, i.e. 50,000 nodes in the time interval [0, 2].

To choose the variable temporal step-size, we use the methods given in Section 4.1.2. Specifically, we divide the time interval [0, 2.0] into 5 equal subintervals of length 0.4 with equal step-sizes in each subinterval. This leaves the number of nodes distributed in each subinterval unequal to give a variable step-size computation.

For 100 nodes in the time interval [0, 2.0], the distribution of these nodes is as shown in Table 4.7. We double the nodes, while keeping the ratio of nodes between consecutive subintervals unchanged and continue this partition pattern until we have 1,600 nodes in time. E.g. for 200 nodes, we have the node distribution shown in Table 4.8.

Table 4.7: Nodes distribution in time interval [0, 2] (total nodes=100)

Subinterval	$[0, 0.4]$	$[0.4, 0.8]$	$[0.8, 1.2]$	$[1.2, 1.6]$	$[1.6, 2.0]$
Nodes	18	28	22	17	15

We summarize the computational results for different second-order, two-step VSIMEX schemes in Table 4.9.

Table 4.8: Nodes distribution in time interval [0, 2] (total nodes=200)

Subinterval	$[0, 0.4]$	$[0.4, 0.8]$	$[0.8, 1.2]$	$[1.2, 1.6]$	$[1.6, 2.0]$
Nodes	36	56	44	34	30

4.2.2 Third- and Fourth-Order VSIMEX Schemes

Consider the test problem (4.11) and (4.12) with $\lambda = \frac{1}{10}$. We compute the solution u at time $t = 2$. For the spatial discretization, we apply fourth-order accurate finite difference approximations of u_x and u_{xx}. See equations (4.9) and (4.10) of Section 4.1.2.

Our approximations using the VSSBDF3 scheme consider a fixed $\Delta x = \frac{1}{250}$. A high resolution solution is used to approximate the exact solution u_e. This is obtained using the SBDF3 scheme with $\Delta t = \frac{1}{500}$, i.e. 1000 nodes in the time interval [0, 2].

For the VSSBDF4 scheme, we choose $\Delta x = \frac{1}{350}$. The approximation to the exact solution is computed using the SBDF4 scheme with $\Delta t = \frac{1}{500}$.

To choose the variable temporal step-sizes, we adopt the partition methods introduced in Section 4.1.2. See Table 4.3 and Table 4.4.

We summarize the computational results for VSSBDF3 in Table 4.10 and for VSSBDF4 in Table 4.11. The corresponding absolute errors for VSSBDF3 and VSSBDF4 are plotted in Figures 4.4 and 4.5 respectively.

Table 4.9: Numerical results for Burgers' equation using various second-order IMEX and VSIMEX schemes.

		Second-Order IMEX			Second-Order VSIMEX		
Scheme	Nodes in time	$\|U-u_e\|_\infty$	Ratio	Order	$\|U-u_e\|_\infty$	Ratio	Order
CNAB	100	1.477e-4			1.256e-4		
$\gamma=\frac{1}{2}$	200	3.740e-5	3.949	1.98	3.218e-5	3.903	1.96
$c=0$	400	9.399e-6	3.979	1.99	8.142e-6	3.952	1.98
	800	2.355e-6	3.991	2.00	2.047e-6	3.978	1.99
	1,600	5.890e-7	3.998	2.00	5.127e-7	3.993	2.00
Modified	100	1.440e-4			1.218e-4		
CNAB	200	3.648e-5	3.947	1.98	3.125e-5	3.898	1.96
$\gamma=\frac{1}{2}$	400	9.170e-6	3.978	1.99	7.909e-6	3.951	1.98
$c=\frac{1}{8}$	800	2.298e-6	3.990	2.00	1.989e-6	3.976	1.99
	1,600	5.746e-7	3.999	2.00	4.983e-7	3.992	2.00
CNLF	100	Unstable			Unstable		
$\gamma=0$	200	1.946e-3			Unstable		
$c=1$	400	6.634e-6			2.213e-5		
	800	1.659e-6	3.999	2.00	3.603e-6	6.142	2.62
	1,600	4.148e-7	4.000	2.00	9.086e-7	3.965	1.99
	3,200	1.038e-7	3.996	2.00	3.437e-7	2.644	1.40
SBDF	100	2.339e-4			1.965e-4		
$\gamma=1$	200	6.004e-5	3.896	1.96	5.176e-5	3.796	1.92
$c=0$	400	1.517e-5	3.958	1.98	1.325e-5	3.906	1.97
	800	3.811e-6	3.981	1.99	3.349e-6	3.956	1.98
	1,600	9.540e-7	3.995	2.00	8.414e-7	3.980	1.99
AMAB	100	Unstable			Unstable		
$\gamma=\frac{1}{2}$	200	Unstable			Unstable		
$c=-\frac{1}{6}$	400	Unstable			Unstable		
	800	Unstable			Unstable		
	1,600	Unstable			Unstable		

Table 4.10: Numerical results for Burgers' equation using the VSSBDF3 scheme.

Scheme No.	Nodes in time	$\|U - u_e\|_\infty$	Ratio	Order
Constant	25	7.418e-4		
step-size	50	1.066e-4	6.96	2.80
	100	1.447e-5	7.37	2.88
	200	1.881e-6	7.69	2.94
	400	2.273e-7	8.28	3.05
	25	2.445e-4		
	50	2.152e-5	11.36	3.51
1	100	2.191e-6	9.82	3.30
	200	2.514e-7	8.72	3.12
	400	3.874e-8	6.48	2.70
	25	4.403e-4		
	50	5.201e-5	8.47	3.08
2	100	6.702e-6	7.76	2.96
	200	8.506e-7	7.88	2.98
	400	9.471e-8	8.98	3.17
	25	3.117e-3		
	50	4.711e-4	6.62	2.73
3	100	6.586e-5	7.15	2.84
	200	8.790e-6	7.49	2.91
	400	1.127e-6	7.80	2.96
	25	1.174e-3		
	50	8.486e-3		
4	100	1.484e-3	5.72	2.52
	200	2.149e-4	6.91	2.79
	400	2.928e-5	7.34	2.88
	25	1.084e-3		
	50	3.577e-4	3.03	1.60
5	100	5.460e-5	6.55	2.71
	200	7.546e-6	7.24	2.86
	400	9.794e-7	7.70	2.94

Table 4.11: Numerical results for Burgers' equation using the VSSBDF4 scheme.

Partition No.	Nodes in time	$\|U - u_e\|_\infty$	Ratio	Order
Constant	25	5.112e-4		
step-size	50	4.209e-5	12.14	3.60
	100	3.160e-6	13.32	3.74
	200	2.196e-7	14.39	3.85
	25	7.461e-5		
	50	3.556e-6	20.98	4.39
1	100	2.469e-7	14.40	3.85
	200	1.667e-8	14.81	3.89
	25	5.221e-4		
	50	2.972e-5	17.57	4.14
2	100	1.898e-6	15.66	3.97
	200	1.230e-7	15.43	3.95
	25	3.152e-3		
	50	2.739e-4	11.51	3.52
3	100	2.188e-5	12.52	3.65
	200	1.601e-6	13.67	3.77
	25	1.075e-4		
	50	4.415e-3		
4	100	1.084e-3	4.07	2.03
	200	9.731e-5	11.14	3.48
	25	6.783e-5		
	50	1.900e-4		
5	100	1.806e-5	10.52	3.40
	200	1.403e-6	12.87	3.69

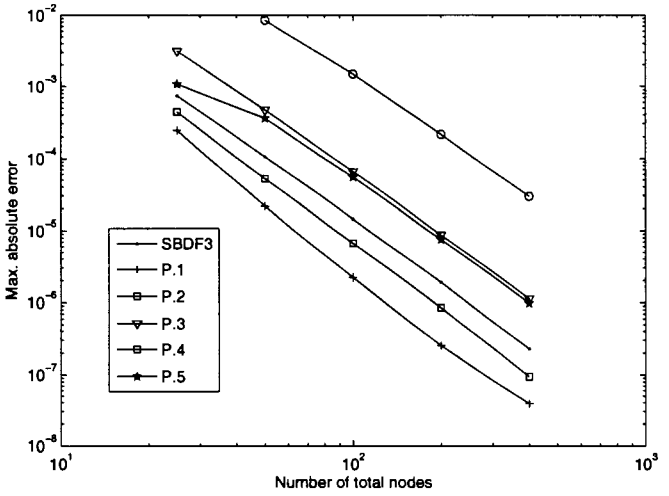


Figure 4.4: Absolute errors for the VSSBDF3 scheme in solving the Burgers' equation. The slopes are approximately -3.0 (using logarithmic scales), which indicates third-order convergence.

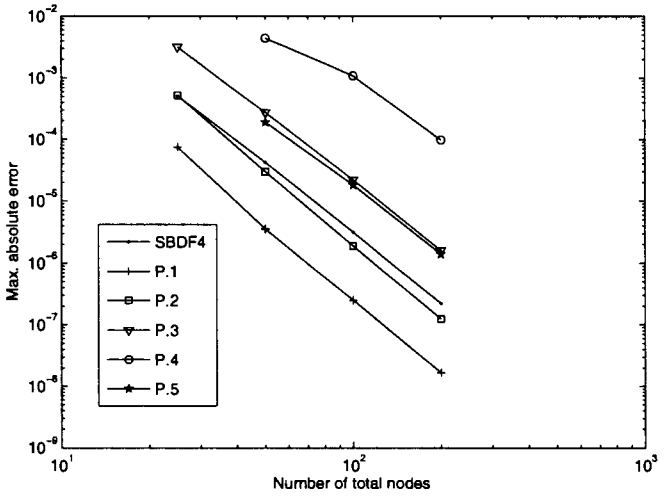


Figure 4.5: Absolute errors for the VSSBDF4 scheme in solving the Burgers' equation. The slopes are approximately -4.0 (using logarithmic scales), which indicates fourth-order convergence.

Chapter 5

Conclusions

5.1 Summary

In this thesis, we have successfully constructed a variety of new variable step-size IMEX linear multistep schemes up to fourth-order. All our VSIMEX schemes are order-s, s-step linear multistep methods.

First-order, one-step IMEX schemes are also VSIMEX schemes. A family of such schemes with one free parameter is given. The family of second-order, two-step VSIMEX schemes with two free parameters is derived. Included in this family of schemes are VSCNLF (2.17), VSCNAB (2.13), Modified VSCNAB (2.20) and VSSBDF2 (2.15), whose corresponding IMEX schemes are popular in practice.

A particular parameterization of third-order, three-step VSIMEX schemes is also provided, which admits three free parameters. In particular, we recommend VSSBDF3 (2.27) because of its superiority in solving stiff problems. A fourth-order, four-step VSIMEX scheme, VSSBDF4 (2.29) is also given.

The linear stability analysis of IMEX schemes is reviewed, and stability contours for different order IMEX schemes are plotted.

The zero-stability of VSIMEX schemes is systematically analyzed. This imposes restrictions on the step-size variations required to ensure VSIMEX schemes are stable as step-sizes approach zero. Based on this analysis, analytical results on restrictions of the step-size ratios for general second-order VSIMEX schemes are obtained and presented (see Corollary 3.4). In particular, VSIMEX schemes with $\gamma = \frac{1}{2}$, e.g. VSCNAB and Modified VSCNAB are zero-stable for any step-size sequences.

Zero-stability is also used to determine step-size ratios for third- and fourth-order VSIMEX schemes. These ratios are numerical results and dependent on the (suitably chosen) matrix norm.

Numerical experiments for the linear advection-diffusion equation and Burgers' equation are carried out using various IMEX and VSIMEX schemes. In our tests, the expected orders of convergence for VSIMEX schemes are achieved and accurate approximate solutions are obtained.

For a linear advection-diffusion problem, some stability improvements were observed using VSIMEX schemes. For a Burgers' equation example, it is demonstrated that VSIMEX schemes give improved accuracy over classical IMEX schemes when variable step-sizes are suitably chosen. In particular, when the time-stepping partitioning scheme No.1 is chosen in Table 4.3 and Table 4.4, the error declines by 90% when VSSBDF4 is used instead of its IMEX counterpart, SBDF4.

5.2 Future Work

There are several promising research opportunities related to VSIMEX schemes.

A study of the truncation error constants for VSIMEX schemes will be interesting, since it can help to determine which VSIMEX scheme is more accurate, especially for the second-order, two-step VSIMEX schemes.

In this thesis, we only consider the zero-stability analysis of VSIMEX schemes. This analysis measures how computational errors propagate as the computation proceeds and as the temporal step-size approaches zero. To better understand the behaviour of our schemes for large time-steps, it would be interesting to carry out a linear stability analysis of our VSIMEX schemes. We remark, however, that the freedom introduced by the step-size variations makes this a challenging problem.

This thesis provides some useful order-s, s-step VSIMEX schemes. We have not considered error control strategies. A natural continuation of our work is to design and implement an automatic time-stepping strategy for our VSIMEX schemes.

Bibliography

[1] Uri M. Ascher, Steven J. Ruuth, and Brian T.R. Wetton. Implicit-explicit methods for time-dependent partial differential equations. *SIAM J. Numerical Analysis*, 32(3):797-823, 1995.

[2] Uri Ascher, Steven Ruuth, and Raymond Spiteri. Implicit-explicit Runge-Kutta methods for time-dependent partial differential equations. *Applied Numerical Mathematics* 25(1997) 151-167.

[3] Carl M. Bender & Steven A. Orszag. *Advanced Mathematical Methods for Scientists and Engineers - Asymptotic Methods and Perturbation Theory.* Springer-Verlag New York, Inc. 1999.

[4] Richard L. Burden and J. Douglas Faires. *Numerical Analysis.* Brooks/Cole, seventh edition, 2001.

[5] Rolf D. Grigorieff. Stability of multistep-methods on variable grids. *Numer. Math.* 42, 359-377 (1983).

[6] E. Hairer, S.P. Norsett, and G. Wanner. *Solving Ordinary Differential Equation I.* Springer-Verlag, 1993.

[7] Christopher Kennedy and Mark Carpenter. Additive Runge-Kutta schemes for convection-diffusion-reaction equations. *Applied Numerical Mathematics* 44(2003) 139-181.

[8] J.D. Lambert. *Numerical Methods for Ordinary Differential Systems - the Initial Value Problem.* John Wiley & Sons Ltd. 2000.

[9] A. Nordsieck. On the numerical integration of ordinary differential equations. *Math. Comp.* 16(1962), pp. 22-49.

[10] John C. Strikwerda. *Finite Difference Schemes and Partial Differential Equations.* Wadsworth & Brooks/Cole, 1989.

[11] L.N. Trefethen, *Spectral methods in MATLAB, ser. Software, Environments, and Tools.* Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM), 2000.