Stability Analysis of Implicit-Explicit Runge-Kutta Discontinous Galerkin Methods for Convection-Dispersion Equations

A Thesis

Presented in Partial Fulfillment of the Requirements for the Degree Master of Mathematical Sciences in the Graduate School of The Ohio State University

By

Joseph Hunter, B.S.

Graduate Program in Mathematical Sciences

The Ohio State University

2021

Master's Examination Committee:

Dr. Yulong Xing, Advisor

Dr. Yuan Lou

Abstract

The purpose of this thesis is to analyze the stability of implicit-explicit Runge-Kutta (IMEX RK) methods when paired with the discontinuous Galerkin method. The analysis of this full discrete spatial-temporal discretization is performed on a linear convection-dispersion equation. The stability analysis is done on a uniform mesh using periodic boundary conditions with the aid of the Fourier method. Five different second or third order IMEX RK methods were studied in this thesis, and each of them is numerically stable under the CFL condition $\Delta t/\Delta x \leq C$ with some suitable constant C. Different IMEX RK methods lead to different value of this constant C. In addition, how the size of this time-step restriction constant C changed as the dispersion coefficient changes was also studied. The expectation was as the dispersion coefficient became large, so would the time-step restriction. Numerical results showed the value of the time-step restriction could approach zero, become constant, or become increasingly large, depending on the specific IMEX RK method.

Acknowledgments

I would like to thank my advisor Dr. Yulong Xing and my co-advisor Dr. Zheng Sun for the help and guidance they gave me in researching for and writing this thesis. Without their support the completion of this thesis would not have been possible. I would also like to thank Dr. Yuan Lou for serving as a committee member for the defense of this thesis.

Finally I would like to thank my family for their love and support throughout all the years and years to come.

Vita

2018 ..... B.S. Mathematics,
Univeristy of Tennessee

2019-present ..... Graduate Associate,
The Ohio State University

Fields of Study

Major Field: Mathematical Sciences

Contents

List of Tables

List of Figures

Chapter 1: Introduction

The goal of this thesis is to study the stability region of the high order finite element discontinuous Galerkin (DG) method when coupled with implicit-explicit (IMEX) Runge-Kutta (RK) methods. The partial differential equation (PDE) to be studied is a linear convection-dispersion equation, which can be viewed as a linearized Korteweg–de Vries (KdV) equation, taking the form

$$u_t + u_x + du_{xxx} = 0. \tag{1.1}$$

With the aid of the Fourier method, we will show that the IMEX DG methods are stable under the CFL condition of the type $\Delta t/\Delta x \leq c(d/\Delta x^2)$ where the value of c depends on the ratio $d/\Delta x^2$ of the dispersion coefficient and the mesh size. We will analyze how the value c changes with respect to the dispersive coefficient.

When discretizing the time derivative, we want to choose a method which will be computationally efficient. The ease of implementation should also be taken into consideration. A natural starting point is to use an explicit method. However the time-step restriction for an explicit method will be dominated by the highest order derivative term, which can cause the time-step restriction to become very small. Since the convection-dispersion equation has a third order spatial derivative term, when using an explicit method we can expect a time-step restriction of $\Delta t \leq C\Delta x^3$. Such

a small time-step restriction will increase the computational cost. The next natural consideration is an implicit method. Most implicit methods will have a relaxed time-step restriction if not unconditional stability. However because the method is implicit, a system of equations must be solved at every step. This can become costly since in practice the convection term is often nonlinear, meaning a nonlinear system of equations will need to be solved at every time step.

After seeing the pros and cons of both an explicit and implicit method it is clear why an IMEX method will be preferred. By treating the dispersion term implicitly we can avoid imposing a strict time-step restriction, and by treating the convection term explicitly we can avoid solving a nonlinear system of equations (if the dispersion term is linear). This is exactly what IMEX methods achieve. For equation (1.1) with dispersion coefficient d, we aim to determine for what value of c does $\tau \leq ch$ make the method stable, where τ is the timestep and h is the meshsize. Additionally, it is pointed out in [13] that in cases where the dispersion term is nonlinear, the nonlinear system that arises from this may have symmetry or positive definite properties which make obtaining the solutions easier.

The idea of using an IMEX method comes naturally from wanting to gain a less restrictive time-step restriction from the implicit method, while avoiding the computational costs associated with treating a nonlinear term implicitly. Developing methods like this began as early as 1980 in [8] where a linear multi-step IMEX method is used for parabolic equations. IMEX methods began to gain popularity after the publication of [1, 2] in which IMEX methods are studied and developed for use on convection-diffusion problems. In [2] the types of IMEX schemes being used were

multi-step. The authors noted that up to that point in time a combination of Adams-Bashforth and Crank-Nicholson was the most popular form of an IMEX scheme. In [1], Ascher, Ruuth and Spiteri noted that the use of high order multi-step IMEX schemes resulted in small time-step restrictions, which led them to investigate IMEX RK schemes. Further research into IMEX RK schemes was carried out by Pareschi and Russo [9] where they developed strong-stability preserving (SSP) IMEX RK schemes for hyperbolic systems with relaxation.

DG methods were first proposed in the 1970's by Reed and Hill [10] for solving hyperbolic equations. Since then the method has been developed and applied to a range of PDEs. One major development in the late 1990's is that of Cockburn and Shu [6] setting up the framework for the LDG method, based on previous work by Bassi and Rebay [3], and applying it to convection-diffusion problems. Advantages of the DG method outlined in [5] include high order accuracy when using approximating polynomial of a suitable degree, the DG method's ability to handle complex geometries, and $h-p$ adaptivity. Since the basis functions in the DG method are discontinuous, this makes the method ideal for equations with discontinuous solutions.

The study of the stability region of DG methods with various temporal discretizations has been carried out for many PDE models. For purely explicit RK methods, Cockburn and Shu [7] determined the CFL numbers for linear hyperbolic conservation laws when using a DG method with polynomials of degree k and an RK method of order v. In a recent paper [13], the IMEX method, coupled with a finite difference (FD) spatial discretization, on convection-diffusion and convection-dispersion PDEs has been studied. In this paper the authors were able to determine the lower bound of the CFL condition for the IMEX methods studied. In [14], the stability and error

estimates are studied for a convection-diffusion equation when the IMEX DG methods are used. The authors showed the time-step restriction is bounded by a constant proprotional to d/c^2, the diffusion and convection coefficients respectively.

The development and use of IMEX methods have also been applied to different PDEs such as hyperbolic equations with relaxation terms [9], convection-diffusion equations [14], hyperbolic-parabolic convection-diffusion equations [1], and convection-dispersion equations [13]. IMEX DG methods have been popularly used to solve the convection-dispersion equation, but there does not exist much study to investigate the stability property of such a method.

In this thesis, the IMEX schemes used are all L-stable IMEX schemes, meaning the implicit tableau is L-stable. The implicit methods are also diagonally implicit (DIRK) schemes. The main focus is to determine the time step constraint of certain IMEX methods when coupled with the DG spatial discretization for the linear convection-dispersion equation. The stability analysis is carried out using the Fourier method which assumes uniform mesh partition and periodic boundary conditions. In addition to determining these values of c, the two main interests of this thesis are as follows.

1. What is the relationship between c, d, and h? As $d \to \infty$ (or as $h \to 0$) can the behaviour of c be generalized? How is this behaviour determined by the IMEX method and/or the degree of the polynomials in the DG method?

2. For the same IMEX method, how do these results compare when a finite difference (FD) spatial method is used? Are the curve shapes for the FD generally the same as the DG method? Is there a relationship between curve shape for the FD and DG method?

To draw the contrast between the FD and DG method, the third order FD method considered in [13] is used for comparison. Based on the results in that paper, we might expect for the FD method as $d \to \infty$ the value of c should increase, or at the very least always be greater than or equal to the value of c for when $d = 0$.

The rest of the thesis is organized as follows. Chapter 2 will present the DG method and its application to the PDE of interest. The FD discretization will also be summarized. Chapter 3 will give the Butcher tableaus and information on the IMEX methods used, as well as a brief description on how to apply an IMEX method. Chapter 4 will present the Fourier analysis and the equations which arise from using the Fourier method on each specific IMEX method. Chapter 5 will present the results of the Fourier analysis, and the general observations. Chapter 6 will present some numerical experiments carried out guided by the results of the Fourier analysis. The concluding remarks and some future work will be discussed in Chapter 7.

Chapter 2: Spatial Discretizations: DG and FD methods

The time step constraint of an IMEX method will vary depending on the spatial discretization applied to the mathematical model. The spatial discretization to be used in this paper is the Local Discontinuous Galerkin (LDG) method. For comparison, we also consider the finite difference method, discussed in [13].

2.1 Discontinuous Galerkin method

In the DG method, the computational domain is discretized into N cells. Here we assume uniform meshes are used. The cells are denoted as $I_j = [x_{j-1/2}, x_{j+1/2}]$ for $1 \leq j \leq N$. The length of the cell will be denoted as $h = x_{j+1/2} - x_{j-1/2}$. In the DG method, the numerical solution is defined on the finite element space V_h^k consisting of piecewise polynomials [11]

$$V_h^k = \{v : v|_{I_j} \in P^k(I_j), 1 \leq j \leq N\},$$

where $P^k(I_j)$ is the linear space on I_j spanned by polynomials of degree less or equal to k. We will take the following basis of V_h for our later analysis

$$V_h^k = \text{span}\{\psi_{j,0}, ..., \psi_{j,k} : 1 \leq j \leq N\}. \tag{2.1}$$

Here $\psi_{j,k}$ refers to the normalization of the k^{th} Legendre polynomial on the cell I_j. A visualization of the solution space for the cell I_j is provided in Figure 2.1.

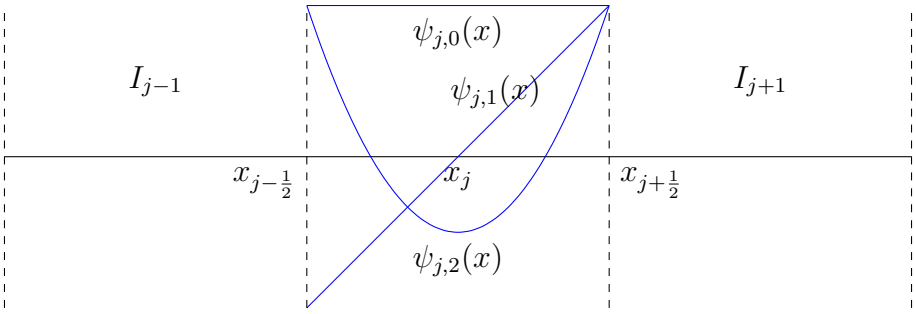

Figure 2.1: Visualization of the solution cells

The LDG method is related to the DG method by first converting the PDE of interest into a system of first-order equations, and then applying the DG method to discretize each equation. Taking the following convection-dispersion scalar PDE

$$u_t + u_x + u_{xxx} = 0, \tag{2.2}$$

as an example to introduce the LDG method. This equation's equivalent first-order system is

$$\begin{aligned} u_t + (u+p)_x &= 0, \\ p - q_x &= 0, \\ q - u_x &= 0. \end{aligned} \tag{2.3}$$

After applying the DG method to the first-order system (2.3), the semi-discrete form of the equation is defined as follows [11]: Find $u, p, q \in V_h^k$ such that for all test

functions $v, w, z \in V_h^k$ we have[1]

$$\begin{aligned}
\int_{I_j} u_t v \, dx - \int_{I_j} (u+p) v_x \, dx + (\hat{u}+\hat{p})_{j+1/2} v_{j+1/2}^- - (\hat{u}+\hat{p})_{j-1/2} v_{j-1/2}^+ = 0, \\
\int_{I_j} pw \, dx + \int_{I_j} q w_x \, dx - \hat{q}_{j+1/2} w_{j+1/2}^- + \hat{q}_{j-1/2} w_{j-1/2}^+ = 0, \\
\int_{I_j} qz \, dx + \int_{I_j} u z_x \, dx - \hat{u}_{j+1/2} z_{j+1/2}^- + \hat{u}_{j-1/2} z_{j-1/2}^+ = 0,
\end{aligned} \tag{2.4}$$

where all hat terms are numerical fluxes, which take the values of

$$\hat{u} = u^-, \qquad \hat{q} = q^+, \qquad \hat{p} = p^+.$$

Here $u_{j+1/2}^{\pm} = u(x_{j+1/2}^{\pm}) = \lim_{x \to x_{j+1/2}^{\pm}} u(x)$, denotes the left or right limit at the cell interfaces. Note that $x_{j+1/2}^-$ would indicate using the solution on cell I_j, and similarly $x_{j+1/2}^+$ would indicate using the solution on cell I_{j+1}.

Since the solutions $u,\ p,\ q \in V_h^k$, they can be written as a linear combination of the basis elements of the finite element space V_h^k. Therefore, the solution on the cell I_j can be written as

$$u(x) = u_{j,0}\psi_{j,0} + u_{j,1}\psi_{j,1} + ... + u_{j,k}\psi_{j,k}. \tag{2.5}$$

The Legendre polynomials have the property that they are orthogonal to each other. Also, each basis functions $\psi_{j,k}$ has been normalized so that $\int_{I_j} \psi_{j,k}^2 = 1$. The formulas

[1]Typically, the notations, u_h, p_h and q_h, are used for numerical quantities in the literature to distinguish with the exact solutions. Here we suppress the subscripts h for the ease of notations.

for the Legendre polynomials $\psi_{j,k}$ up to $k = 4$ are given below

$$\begin{aligned}
\psi_{j,0} &= \frac{1}{\sqrt{h}}, \\
\psi_{j,1} &= \frac{2\sqrt{3}}{\sqrt{h}} \frac{x - x_j}{h}, \\
\psi_{j,2} &= \frac{\sqrt{5}}{2\sqrt{h}} \left(-1 + \frac{12(x - x_j)^2}{h^2} \right), \\
\psi_{j,3} &= \frac{\sqrt{7}}{2\sqrt{h}} \left(\frac{40(x - x_j)^3}{h^3} - \frac{6(x - x_j)}{h} \right), \\
\psi_{j,4} &= \frac{\sqrt{9}}{8\sqrt{h}} \left(3 - \frac{120(x - x_j)^2}{h^2} + \frac{560(x - x_j)^4}{h^4} \right).
\end{aligned} \tag{2.6}$$

In general the k^{th} Legendre polynomial on the interval $[a, b]$ is given by $P_k\left(\frac{2x-a-b}{b-a}\right)$ where P_k is the Legendre polynomial of degree k. For the cell I_j, this general formula is $P_k\left(\frac{2(x-x_j)}{h}\right)$. Note that $\psi_{j,k}$ is a normalization of $P_k\left(\frac{2(x-x_j)}{h}\right)$, i.e., $\psi_{j,k} = \frac{1}{\alpha} P_k\left(\frac{2(x-x_j)}{h}\right)$ where $\alpha^2 = \int_{I_j} P_k\left(\frac{2(x-x_j)}{h}\right)^2 \mathrm{dx}$.

By replacing u, p and q in equation (2.4) with their solution coefficients on the cell I_j, one can formally rewrite the DG method in the form of a finite difference scheme. Because equation (2.4) holds for all the test functions $v, w, z \in V_h^k$, we replace v with $\psi_{j,0}, \psi_{j,1}, ..., \psi_{j,k}$ respectively. The same can be done for w and z. Making these replacements will result in three systems of equations. They include one system which can solve for q in terms of u, one system which can solve for p in terms of q, and a system of ordinary differential equations which updates u in terms of p and u. After solving the first two systems and backwards substituting the solutions of q and p, a semi-discrete scheme entirely in terms of u can be derived.

2.1.1 Semi-Discrete form with $k = 1$

As an example to illustrate this process in general, the details for $k = 1$ are given in this section. Further details are provided in the Appendix. Take the last equation

in (2.4) and replace q and u with their solutions on the cell I_j and replace z with $\psi_{j,0}$ and $\psi_{j,1}$ respectively. This results in the following 2×2 system

$$\begin{bmatrix} \int_{I_j} \psi_{j,0}^2 & \int_{I_j} \psi_{j,1}\psi_{j,0} \\ \int_{I_j} \psi_{j,0}\psi_{j,1} & \int_{I_j} \psi_{j,1}^2 \end{bmatrix} \begin{bmatrix} q_{j,0} \\ q_{j,1} \end{bmatrix} + \begin{bmatrix} \int_{I_j} \psi_{j,0}\psi'_{j,0} & \int_{I_j} \psi_{j,1}\psi'_{j,0} \\ \int_{I_j} \psi_{j,0}\psi'_{j,1} & \int_{I_j} \psi_{j,1}\psi'_{j,1} \end{bmatrix} \begin{bmatrix} u_{j,0} \\ u_{j,1} \end{bmatrix}$$
$$- \begin{bmatrix} \psi_{j,0}^2(x^-_{j+1/2}) & \psi_{j,1}(x^-_{j+1/2})\psi_{j,0}(x^-_{j+1/2}) \\ \psi_{j,0}(x^-_{j+1/2})\psi_{j,1}(x^-_{j+1/2}) & \psi_{j,1}^2(x^-_{j+1/2}) \end{bmatrix} \begin{bmatrix} u_{j,0} \\ u_{j,1} \end{bmatrix}$$
$$+ \begin{bmatrix} \psi_{j-1,0}(x^-_{j-1/2})\psi_{j,0}(x^+_{j-1/2}) & \psi_{j-1,1}(x^-_{j-1/2})\psi_{j,0}(x^+_{j-1/2}) \\ \psi_{j-1,0}(x^-_{j-1/2})\psi_{j,1}(x^+_{j-1/2}) & \psi_{j-1,1}(x^-_{j-1/2})\psi_{j,1}(x^+_{j-1/2}) \end{bmatrix} \begin{bmatrix} u_{j-1,0} \\ u_{j-1,1} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Due to the orthonormality of the basis functions $\psi_{j,k}$, the first matrix reduces to the identity, and the second matrix becomes a sparse, banded, lower-triangular matrix after the assembly. Solving this system for $q_{j,0}$ and $q_{j,1}$ gives

$$\begin{bmatrix} q_{j,0} \\ q_{j,1} \end{bmatrix} = \frac{1}{h} \left(\begin{bmatrix} 1 & \sqrt{3} \\ -\sqrt{3} & 3 \end{bmatrix} \begin{bmatrix} u_{j,0} \\ u_{j,1} \end{bmatrix} + \begin{bmatrix} -1 & -\sqrt{3} \\ \sqrt{3} & 3 \end{bmatrix} \begin{bmatrix} u_{j-1,0} \\ u_{j-1,1} \end{bmatrix} \right).$$

This process can be repeated on the second and first equation of (2.4) to get solutions for $p_{j,0}$ and $p_{j,1}$ and for $u'_{j,0}$ and $u'_{j,1}$. By backwards substituting the solution for q, the numerical scheme for u_t is given by

$$\begin{bmatrix} u'_{j,0} \\ u'_{j,1} \end{bmatrix} = \frac{1}{h} \left(\begin{bmatrix} -1 & -\sqrt{3} \\ \sqrt{3} & -3 \end{bmatrix} \begin{bmatrix} u_{j,0} \\ u_{j,1} \end{bmatrix} + \begin{bmatrix} 1 & \sqrt{3} \\ -\sqrt{3} & -3 \end{bmatrix} \begin{bmatrix} u_{j-1,0} \\ u_{j-1,1} \end{bmatrix} \right)$$
$$+ \frac{1}{h^3} \left(\begin{bmatrix} 8 & -4\sqrt{3} \\ 8\sqrt{3} & -12 \end{bmatrix} \begin{bmatrix} u_{j+2,0} \\ u_{j+2,1} \end{bmatrix} + \begin{bmatrix} -6 & -18\sqrt{3} \\ 18\sqrt{3} & -90 \end{bmatrix} \begin{bmatrix} u_{j+1,0} \\ u_{j+1,1} \end{bmatrix} \right.$$
$$\left. + \begin{bmatrix} -12 & 12\sqrt{3} \\ -24\sqrt{3} & -108 \end{bmatrix} \begin{bmatrix} u_{j,0} \\ u_{j,1} \end{bmatrix} + \begin{bmatrix} 10 & 10\sqrt{3} \\ -2\sqrt{3} & -6 \end{bmatrix} \begin{bmatrix} u_{j-1,0} \\ u_{j-1,1} \end{bmatrix} \right). \quad (2.7)$$

This can be made more compact by introducing vector and matrix notations. Let $u_j = [u_{j,0}, u_{j,1}]^T$. Then the semi-discrete form of (2.2) for $k = 1$ can be written as

$$u'_j = \frac{1}{h} \left(C_1 u_j + C_2 u_{j-1} \right) + \frac{1}{h^3} \left(D_1 u_{j+2} + D_2 u_{j+1} + D_3 u_j + D_4 u_{j-1} \right). \quad (2.8)$$

The structure of this equation stays the same for different values of k. All that changes is the size of the matrices and vectors and the coefficients in the matrices.

The matrices are size $(k+1)\times(k+1)$ and the vectors are size $(k+1)\times 1$. The coefficient matrices for different values of k are given in the Appendix.

When periodic boundary conditions are assumed, as they are for the Fourier analysis, the periodicity can be taken advantage of so that all values of u_j' can be solved for in one step. Equation (2.8) can rewritten as

$$U' = \frac{1}{h}CU + \frac{1}{h^3}DU, \tag{2.9}$$

where $U = [u_1, u_2, \ldots, u_N]^T$ and

$$C = \begin{bmatrix} C_1 & & & & C_2 \\ C_2 & C_1 & & & \\ & C_2 & C_1 & & \\ & & \ddots & \ddots & \\ & & & C_2 & C_1 \end{bmatrix}, \; D = \begin{bmatrix} D_3 & D_2 & D_1 & & & D_4 \\ D_4 & D_3 & D_2 & D_1 & & \\ & \ddots & \ddots & \ddots & \ddots & \\ & & D_4 & D_3 & D_2 & D_1 \\ D_1 & & & D_4 & D_3 & D_2 \\ D_2 & D_1 & & & D_4 & D_3 \end{bmatrix}.$$

Due to the periodicity, $u_0 = u_N$, $u_{N+1} = u_1$, and $u_{N+2} = u_2$.

2.2 Finite Difference Method

As mentioned previously, the stability of IMEX methods paired with a finite difference spatial discretization was studied on a linear convection-dispersion PDE in [13]. In that paper the semidiscrete finite difference method is given as follows

$$u_j' = \mathcal{L}(t,u)_j + N(t,u)_j. \tag{2.10}$$

$\mathcal{L}(t,u)_j$ comes from the discretization of the dispersion term du_{xxx} and is given by the third order one-point upwind biased scheme [13]

$$\mathcal{L}(t,u)_j = -d\frac{-u_{j+3} + 7u_{j+2} - 14u_{j+1} + 10u_j - u_{j-1} - u_{j-2}}{4h^3}. \tag{2.11}$$

$N(t,u)_j$ comes from the discretization of the convection term u_x and is given by the third order upwind biased finite difference scheme [13]

$$N(t,u)_j = -\frac{3u_j + 2u_{j+1} - 6u_{j-1} + u_{j-2}}{6h}. \tag{2.12}$$

Chapter 3: IMEX Methods

In this chapter, we briefly review several second order and third order IMEX methods that will be analyzed later.

3.1 A General IMEX Method

An arbitrary IMEX method can be represented by the following Butcher tableau:

$$\begin{array}{c|c} c_E & A_E \\ \hline & b_E^T \end{array} \qquad \begin{array}{c|c} c_I & A_I \\ \hline & b_I^T \end{array}$$

The one on the left is the explicit tableau, and the one on the right is the implicit tableau of the IMEX method. The elements in the i^{th} row and j^{th} column of A_E and A_I are $a_{ij}^{(E)}$ and $a_{ij}^{(I)}$, respectively. A_E is typically one row and column larger in size than A_I. In order to make the matrices the same size, and so the methods can work in tandem A_I is padded with a row and column of zeroes. In a similar way, one can define $c_i^{(E)}$, $c_i^{(I)}$, $b_i^{(E)}$, and $b_i^{(I)}$. The order conditions for an IMEX method include those needed for a Runge-Kutta method plus additional coupling conditions [9]. The amount of coupling conditions grow very large as the order increases. However these coupling conditions can be drastically reduced if $c_E = c_I$, if $b_E = b_I$, or if both of those vectors are equal [9]. For this reason, methods where one or both of these conditions hold are often used. The implicit part of the methods are often taken to be singly

diagonally implicit (SDIRK) methods, which means A_I is lower triangular and the diagonal elements are all the same. This simplifies the process because there is only one term with implicit dependence, and the coefficient of this term is the same at every stage. So there is only one matrix which will need to be inverted. It can be inverted once at the beginning and stored. When the implicit method is padded with zeros it becomes an explicit first stage singly diagonally implicit method (ESDIRK).

Applying an IMEX method whose matrices are size $s \times s$ to (2.9) results in the following sets of equations

$$\begin{aligned}
&U^{n,0} = U^n, \\
&U^{n,i} = U^n + \frac{\tau}{h}\sum_{j=1}^{i-1} a_{i,j}^{(E)} C U^{n,j} + \frac{\tau}{h^3}\sum_{j=1}^{i} a_{i,j}^{(I)} D U^{n,j} \qquad \text{for } 1 \le i \le s, \\
&U^{n+1} = U^n + \frac{\tau}{h}\sum_{j=1}^{s} b_i^{(E)} C U^{n,j} + \frac{\tau}{h^3}\sum_{j=1}^{s} b_i^{(I)} D U^{n,j}.
\end{aligned}$$

Here τ is the size of the time step, defined as $t^{n+1} - t^n = \tau$, and the implicit method is assumed to be diagonally implicit.

The second and third order methods to be studied are described below.

3.2 Second Order IMEX Methods

L-stable second order DIRK

$$\begin{array}{c|ccc|ccc}
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\gamma & \gamma & 0 & 0 & 0 & \gamma & 0 \\
1 & \delta & 1-\delta & 0 & 0 & 1-\gamma & \gamma \\
\hline
 & \delta & 1-\delta & 0 & 0 & 1-\gamma & \gamma
\end{array} \tag{3.1}$$

$$\gamma = 1 - \frac{\sqrt{2}}{2} \text{ and } \delta = 1 - \frac{1}{2\gamma}.$$

The Butcher tableau for the second order IMEX (3.1) in [14] is displayed above. The explicit part is on the left and the implicit part is on the right. The implicit

scheme is L-stable and is stiffly accurate [1]. This IMEX method will be paired with a second order DG discretization which uses polynomials up to degree $k = 1$, and a third order DG discretization ($k = 2$).

3.3 Third Order IMEX Methods

Third order IMEXSSP3

$$\begin{array}{c|cccc} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1/2 & 0 & 1/4 & 1/4 & 0 \\ \hline & 0 & 1/6 & 1/6 & 2/3 \end{array} \qquad \begin{array}{c|cccc} \alpha & \alpha & 0 & 0 & 0 \\ 0 & -\alpha & \alpha & 0 & 0 \\ 1 & 0 & 1-\alpha & \alpha & 0 \\ 1/2 & \beta & \eta & 1/2-\beta-\eta-\alpha & \alpha \\ \hline & 0 & 1/6 & 1/6 & 2/3 \end{array} \tag{3.2}$$

$\alpha = 0.24219426078821$, $\beta = 0.06042356519705$, and $\eta = 0.12915286960590$

The Butcher tableau for the third order IMEXSSP3 method is displayed above [9]. The explicitt method is a strong-stability preserving method. The implicit scheme is also L-stable. This scheme was considered with a DG discretization of $k = 2$.

Third order combination

$$\begin{array}{c|ccc|ccc} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \gamma & \gamma & 0 & 0 & 0 & \gamma & 0 \\ 1-\gamma & \gamma-1 & 2(1-\gamma) & 0 & 0 & 1-2\gamma & \gamma \\ \hline & 0 & 1/2 & 1/2 & 0 & 1/2 & 1/2 \end{array} \tag{3.3}$$

$$\gamma = \tfrac{3+\sqrt{3}}{6}$$

The Butcher tableau for the third order combination IMEX method considered in [1] is displayed above. Both the implicit and explicit parts by themselves are third order accurate as well. This IMEX method was only considered when using a DG discretization of $k = 2$.

L-stable third order DIRK

$$\begin{array}{c|cccc|cccc} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \gamma & \gamma & 0 & 0 & 0 & 0 & \gamma & 0 & 0 \\ \frac{1+\gamma}{2} & a_{31} & a_{32} & 0 & 0 & 0 & \frac{1-\gamma}{2} & \gamma & 0 \\ 1 & a_{41} & a_{42} & a_{43} & 0 & 0 & b_1 & b_2 & \gamma \\ \hline & 0 & b_1 & b_2 & \gamma & 0 & b_1 & b_2 & \gamma \end{array} \tag{3.4}$$

$\gamma = 0.4358665215$, $a_{31} = 0.3212788860$, $a_{32} = 0.3966543747$, $a_{41} = -0.105858296$,

$a_{42} = a_{43} = 0.5529291479$, $b_1 = 1.208496649$, $b_2 = -0.644363171$

The Butcher tableau for the L-stable third order DIRK is displayed above [1]. If higher precision for the coefficients is needed, the formulas for them can be found in Ascher, Ruuth, and Spiteri's classical paper [1] on IMEX methods for time-dependent PDEs. What is notable about this method is the four-stage third order explicit tableau has the same stability region as all four-stage fourth order explicit RK methods. The two previous third order methods did not have this property. The implicit part is still third order accurate and L-stable. This IMEX method was considered when using a DG discretization of $k = 2$ and $k = 3$.

Alternate L-stable third order DIRK

$$\begin{array}{c|cccc|cccc} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \gamma & \gamma & 0 & 0 & 0 & 0 & \gamma & 0 & 0 \\ \frac{1+\gamma}{2} & \frac{1+\gamma}{2} - a_1 & a_1 & 0 & 0 & 0 & \frac{1-\gamma}{2} & \gamma & 0 \\ 1 & 0 & 1 - a_2 & a_2 & 0 & 0 & b_1 & b_2 & \gamma \\ \hline & 0 & b_1 & b_2 & \gamma & 0 & b_1 & b_2 & \gamma \end{array} \tag{3.5}$$

$$\gamma = 0.4358665215,\ a_1 = -0.35,\ a_2 = \frac{1/3 - 2\gamma^2 - 2b_2 a_1 \gamma}{\gamma(1-\gamma)},\ b_1 = 1.208496649,$$
$$b_2 = -0.644363171$$

The Butcher tableau for this alternate L-stable third order DIRK is displayed above [14]. This method is of the same form as (3.4). The only difference is the choice

of parameters for $a_{31}, a_{32}, a_{41}, a_{42}, a_{43}$ (see [4] for more details on these parameter choices). However due to the choice of these parameters, the explicit scheme no longer has the same stability region as the classical third order explicit RK method (Figure 5.13). The value of the CFL number for the advection equation (i.e., when $d = 0$) was not the same as the one in [7, Table 2.2] when paired with a DG discretization with $k = 2$.

Chapter 4: Fourier Analysis

In order to analyze the time step constraint of the IMEX methods when paired with a DG space discretization, the Fourier method is used. After applying the DG method to the equation

$$u_t + u_x + du_{xxx} = 0, \tag{4.1}$$

the semi-discrete DG form of equation (4.1) takes the following form

$$u_j' = \frac{1}{h}(C_1 u_j + C_2 u_{j-1}) + \frac{d}{h^3}(D_1 u_{j+2} + D_2 u_{j+1} + D_3 u_j + D_4 u_{j-1}). \tag{4.2}$$

In equation (4.2), the vectors u_j are of size $(k+1) \times 1$ whose coefficients determine the value of u on the cell I_j, the matrices C_i and D_i are $(k+1) \times (k+1)$ matrices of coefficients, and $h = x_{j+1} - x_j$ which is the same for all j. Here k refers to the polynomial degree of the DG method used. A uniform mesh and periodic boundary conditions are assumed.

In Fourier analysis, the ansatz of the form

$$u_j = \tilde{u} e^{i\omega x_j}$$

with $i^2 = -1$ is made in the semi-discrete form. This substitution yields the following

$$\tilde{u}' = C\tilde{u} + D\tilde{u}. \tag{4.3}$$

Here $C = \frac{1}{h}(C_1 + C_2 e^{-iz})$ and $D = \frac{d}{h^3}(D_1 e^{2iz} + D_2 e^{iz} + D_3 + D_4 e^{-iz})$ with $z = \omega h$. The IMEX method of interest is then applied to equation (4.3). The matrix C arises from the convection term of the equation, so this part of the equation is the one treated explicitly. While the matrix D arises from the dispersion term, so it is treated implicitly.

Once the IMEX method is applied, the value of u at the next time step, t_{n+1}, can be written as

$$\tilde{U}^{n+1} = K\tilde{U}^n. \tag{4.4}$$

Here $\tilde{U}^n$ represents the Fourier expansion of the local truncation error for the numerical solution at time t^n. The matrix K is defined to be the amplification factor of the error in going from time-step t^n to t^{n+1}. The method will be strongly stable [12] if $\|K\|_2 \leq 1$ for all ω. If this is true, at each time step the error does not increase since $\|\tilde{U}^{n+1}\|_2 \leq \|\tilde{U}^n\|_2$. The matrix K is a function of τ, h, and ω, and the norm of K being less than one depends on the choice of τ and h, since for stability we require this to hold for all ω. However strong stability is a strict requirement which may pose a severe time step restriction. Instead the relaxed condition of $\rho(K) \leq 1$ was used, where $\rho(K)$ is the spectral radius of the amplification matrix K. Using this relaxed condition, it was found that the scheme was numerically stable and the solution did not blow up in finite time. Similar conclusions for the CFL number of $u_t + u_x = 0$ were studied in [7].

Regarding the finite difference scheme given by (2.10), an equation similar to (4.3) can be derived

$$\tilde{u}' = \tilde{\mathcal{L}}\tilde{u} + \tilde{N}\tilde{u} \tag{4.5}$$

with $\tilde{\mathcal{L}} = \frac{-d}{4h^3}(-e^{3iz} + 7e^{2iz} - 14e^{iz} + 10 - e^{-iz} - e^{-2iz})$ and $\tilde{N} = \frac{-1}{6h}(3 + 2e^{iz} - 6e^{-iz} + e^{-2iz})$ being scalars instead of matrices. Therefore the amplification factor matrix K reduces to a scalar, leading to an easier calculation of the time step constraint. For each IMEX method, the calculation of the amplification factor is the same for DG or finite difference scheme. The only differences in the following equations to determine K are to switch D with $\tilde{\mathcal{L}}$, switch C with $\tilde{N}$, and to interpret the matrix inverse as the reciprocal and the Identity matrix as 1.

4.1 L-stable second order DIRK

Applying (3.1) to the semi-discrete form derived from the Fourier analysis (4.3) yields the following equations to calculate the matrix K

$$\begin{aligned} K_1 &= (I - \gamma\tau D)^{-1}, \\ K_2 &= K_1(I + \gamma\tau C), \\ K &= K_1(I + \delta\tau C + (1-\delta)\tau C, K_2 + (1-\gamma)\tau D K_2). \end{aligned} \tag{4.6}$$

This method has the property that the last internal stage is the same as the final stage of the IMEX method. This is why the formula for the final K matrix is multiplied by K_1. Whereas the other IMEX methods in this paper do not have this property, so the final stage is needed to compute K.

4.2 Third order IMEXSSP3

Applying (3.2) to the semi-discrete form derived from the Fourier analysis (4.3) yields the following equations to calculate the matrix K

$$\begin{aligned}
K_1 &= (I - \alpha\tau D)^{-1}, \\
K_2 &= K_1(I - \alpha\tau D K_1), \\
K_3 &= K_1(I + (\tau C + (1-\alpha)\tau D)K_2), \\
K_4 &= K_1(I + \beta\tau D K_1 + (\frac{\tau}{4}C + \eta\tau D)K_2 + (\frac{\tau}{4}C + (\frac{1}{2} - \beta - \eta - \alpha)\tau D)K_3), \\
K &= I + \frac{\tau}{6}(C+D)K_2 + \frac{\tau}{6}(C+D)K_3 + \frac{2\tau}{3}(C+D)K_4.
\end{aligned} \tag{4.7}$$

4.3 Third order combination

Applying (3.2) to the semi-discrete form derived from the Fourier analysis (4.3) yields the following equations to calculate the matrix K

$$\begin{aligned}
K_1 &= (I - \gamma\tau D)^{-1}, \\
K_2 &= K_1(I + \gamma\tau C), \\
K_3 &= K_1(I + (\gamma-1)\tau C + 2(1-\gamma)\tau C K_2 + (1-2\gamma)\tau D K_2), \\
K &= I + \frac{\tau}{2}(C+D)K_2 + \frac{\tau}{2}(C+D)K_3.
\end{aligned} \tag{4.8}$$

4.4 L-stable third order DIRK

Applying (3.4) to the semi-discrete form derived from the Fourier analysis (4.3) yields the following equations to calculate the matrix K

$$\begin{aligned}
K_1 &= (I - \gamma\tau D)^{-1}, \\
K_2 &= K_1(I + \gamma\tau C), \\
K_3 &= K_1(I + a_{31}\tau C + a_{32}\tau C K_2 + \frac{1-\gamma}{2}\tau D K_2), \\
K_4 &= K_1(I + a_{41}\tau C + a_{42}\tau C K_2 + a_{43}\tau C K_3 + b_1\tau D K_2 + b_2\tau D K_3), \\
K &= I + b_1\tau(C + D)K_2 + b_2\tau(C + D)K_3 + \gamma\tau(C + D)K_4.
\end{aligned} \tag{4.9}$$

4.5 Alternate L-stable third order DIRK

Applying (3.5) to the semi-discrete form derived from the Fourier analysis (4.3) yields the following equations to calculate the matrix K

$$\begin{aligned}
K_1 &= (I - \gamma\tau D)^{-1}, \\
K_2 &= K_1(I + \gamma\tau C), \\
K_3 &= K_1(I + (\frac{1+\gamma}{2} - a_1)\tau C + a_1\tau C K_2 + \frac{1-\gamma}{2}\tau D K_2), \\
K_4 &= K_1(I + (1 - a_2)\tau C K_2 + a_2\tau C K_3 + b_1\tau D K_2 + b_2\tau D K_3), \\
K &= I + b_1\tau(C + D)K_2 + b_2\tau(C + D)K_3 + \gamma\tau(C + D)K_4.
\end{aligned} \tag{4.10}$$

4.6 Algorithm to Determine the CFL Condition

The stability of these IMEX methods was initially analyzed by determining if there exists a value c such that when $\tau \leq ch$ we have that $\rho(K) \leq 1$ for all ω and any h with $d = 1$. When this value c was found it was noted that this value wasn't necessarily the largest c could be for a specific h in order for $\rho(K) \leq 1$ for all ω. This maximal value of c depends on the values of d, the dispersion coefficient, and h. By

fixing h and varying d larger values of c could be achieved. In general it seemed that the value of c determined for all ω and h was a lower bound for achieving stability. A relationship of $\dfrac{d}{h^2}$ was found to be related to this maximal value of c. That is to say if the value of $\dfrac{d}{h^2}$ is held constant, the maximal value of c is the same no matter the choice of d and h (see Figure 4.1 and 4.2 for some comparisons).

A closer investigation shows that $K = K(\tau/h, z, d/h^2)$, depends on the three quantities, τ/h, z and d/h^2. To ensure the stability condition $\rho(K) \leq 1$ for a fixed value of d/h^2, c is taken as

$$c = c(d/h^2) = \min_z \left(\max\{c' : \rho(K(c', z, d/h^2)) \leq 1\} \right).$$

The algorithm to numerically determine c is presented below. In this algorithm *pts* refers to how densely $z = \omega h$ is sampled. Typically this value was taken to be 1001 and seemed to be dense enough for most methods. However for the IMEXSSP3 method 1001 and 10001 points were not dense enough and caused spurious oscillations to appear in the value of c. The value of *pts* was increased to 40001, at which point the oscillations disappeared. The value of *precision* refers to how many accurate decimal digits c will have. So *precision* is taken to be a power of 10, typically 100 for two decimal digit accuracy. A denser sampling of d can be taken as well to produce a fuller graph. Sometimes $d = 2^k,\ (15/10)^k,\ (13/10)^k,\ (11/10)^k$ were used to produce the graphs in Chapter 5.

Algorithm 1 CFL number

```
pts ← n
precision ← 100
Initialize coords
set value of h
for a suitable range of k do
  d ← 2^k
  bool ← TRUE
  c ← 0
  while bool = TRUE do
    τ ← c/precision
    for i=0 : pts − 1 do
      z = 2πi/(pts − 1)
      Compute ρ(K)
      if ρ(K) > 1 then
        Break out of for loop
      end if
    end for
    if bool ≠ 0 then
      c = c + 1
    else
      c = c − 1
      Break out of while loop
    end if
  end while
  Store (d, c) in coords
end for
return coords
```

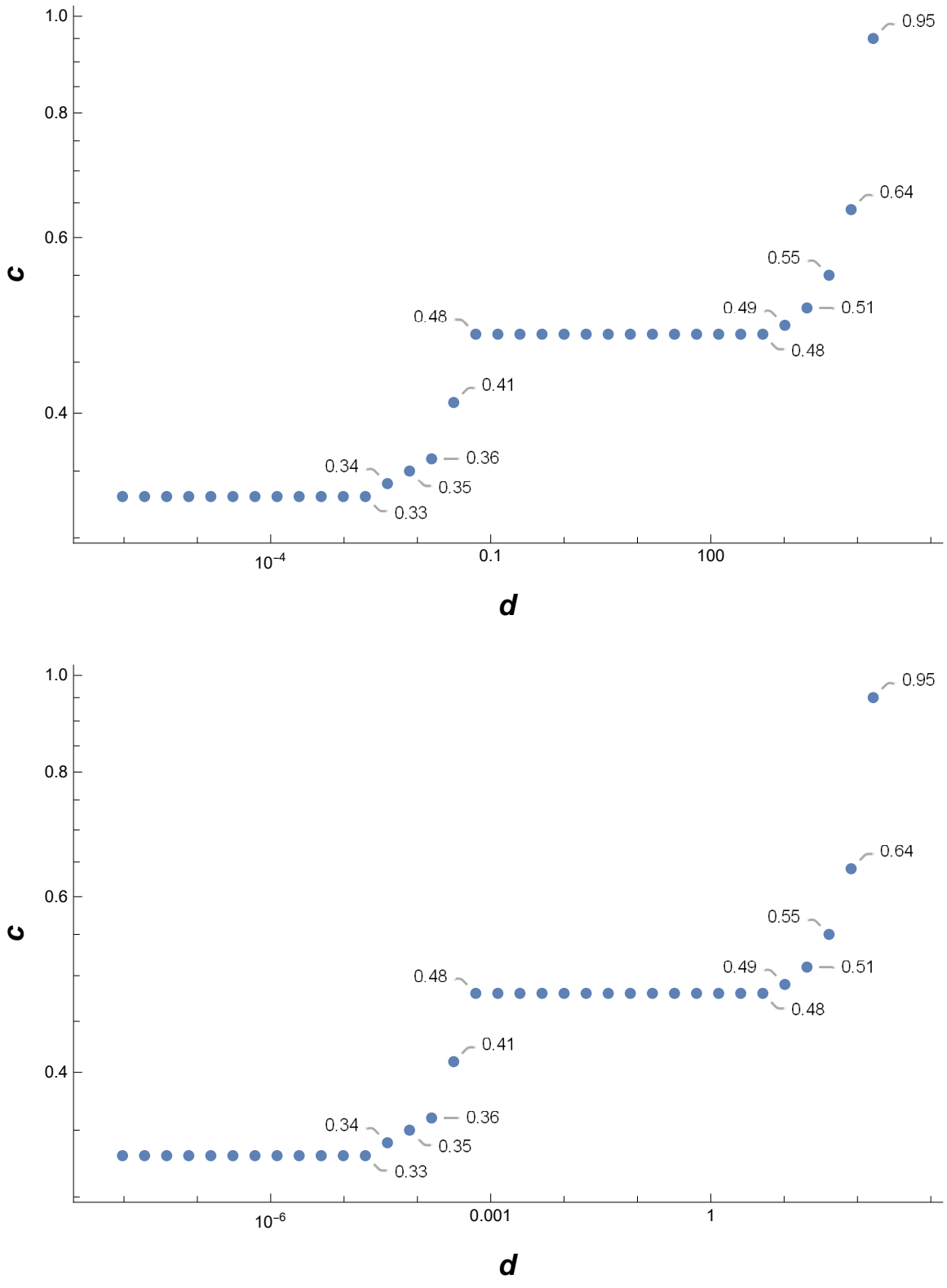

Figure 4.1: Relationship of $\frac{d}{h^2}$. Top: $h = 1$. Bottom: $h = 0.1$

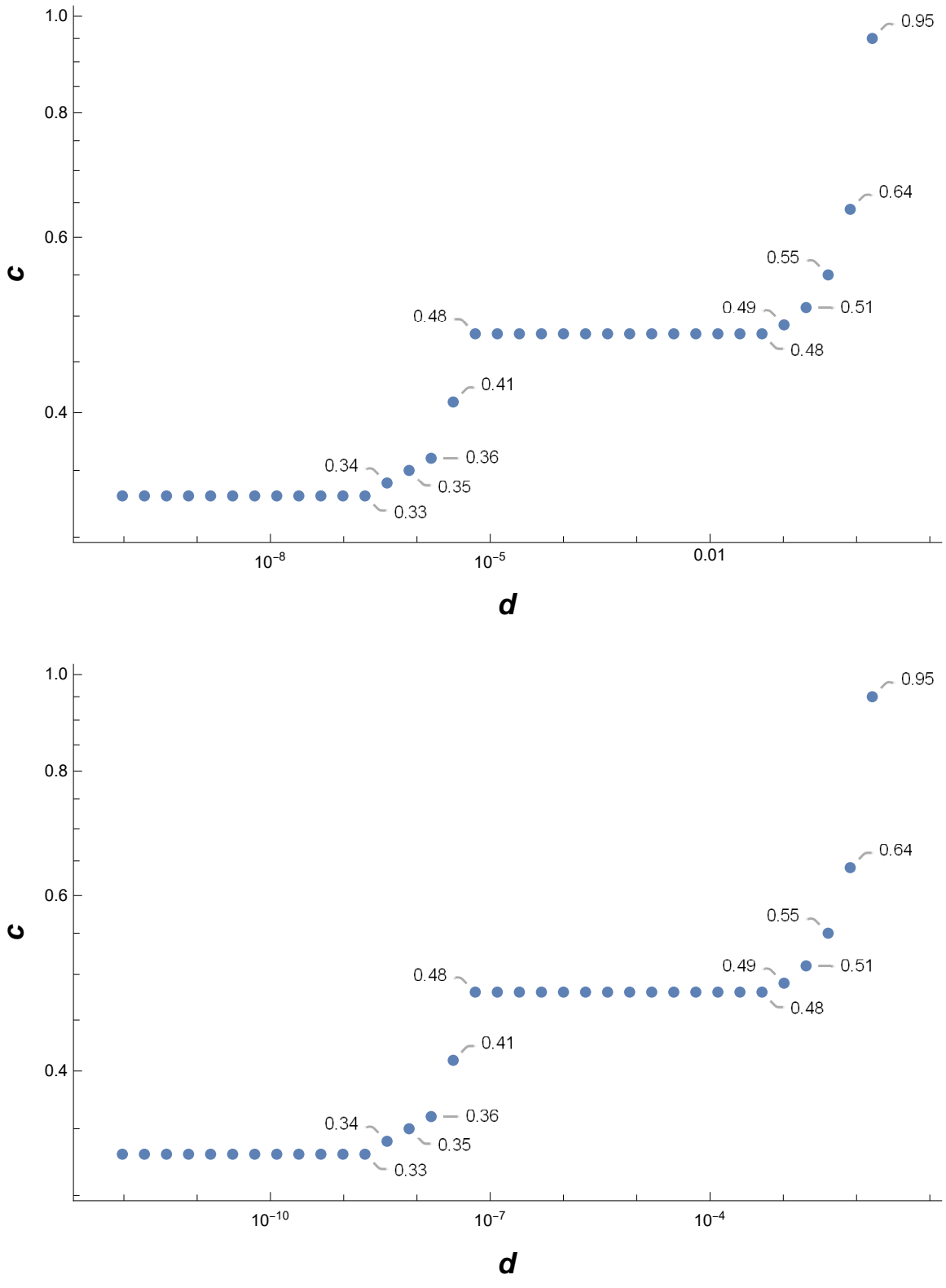

Figure 4.2: Relationship of $\frac{d}{h^2}$. Top: $h = 0.01$. Bottom: $h = 0.001$

Chapter 5: Stability Analysis

In this chapter we present the results of the stability analysis for the IMEX RK methods presented in Chapter 3. Each IMEX RK method is considered with an appropriate DG discretization, and these results are compared with the results of using the third order finite difference scheme. In [13] Fourier analysis is similarly applied on a linear convection-dispersion PDE to determine the CFL number of IMEX method coupled with the finite difference method. The authors show the numerical method is stable under the condition $\tau \leq ch$. The main result from the paper is that the value of c when $d = 0$ is a lower bound of c for stability analysis. For notational purposes, let's call c_0 this specific value of c. The amplification factor K which results from the Fourier analysis satisfies $|K| \leq 1$ when $\tau \leq c_0 h$ for any d. However the authors noticed that for values of $d > 0$ the scheme was still stable for $\tau \leq (c_0 + 0.01)h$. That is for this choice of τ the norm of the amplification factor was still less than or equal to one. The authors demonstrated this with an error table for a few small values of d, which shows that a little bit of dispersion helps the time-step restriction. The aim of using a finite difference discretization in this thesis is not only to verify the results from [13], but to take a deeper look as well as to provide a comparison for the results when a DG discretization is used.

We start by summarizing some observations of the Fourier analysis. When the DG method was used, we observed that this value of c_0 was not necessarily the lower bound. The minimum value of c was at times lower than the value of c_0. The value of c always initially increased, but for some methods it could reach a peak and start to decrease. The expectation was that as more dispersion was added this would help the stability of the equation and the value of c would increase. This was an expectation, not only because of the results in [13], but because as the dispersive term became large enough, the method would be dominated by the implicit part, which is generally unconditionally stable. This end behaviour of unconditional stability (or large values of c) was not observed for all IMEX schemes. Some did appear to become unconditionally stable in the asymptotic limit as d increased, but for other IMEX methods the value of c settled down and remained constant. For two methods this value of c was less than the value of c_0, meaning large dispersion was not helpful. In one of these cases we observed that c_0 was not the minimal value for the finite difference scheme, meaning the curve shapes for the DG method and finite difference scheme matched. For the other case, we observed that c_0 was the minimal value for the finite difference method, meaning the curve shapes for the DG method and finite difference method did not match. For the remaining methods the curve of c values with respect to d for the finite difference scheme follows a similar pattern as the curve of c values when a DG discretization is used.

The time step constraints of course depend on the IMEX method used and the order of the DG discretization. But is there a general pattern that can be observed? Is there a common characteristic of these stability regions when the order of the IMEX method and DG discretization is kept constant? Are similar patterns observed if the

difference between these two orders is kept constant? Do all v^{th} order IMEX methods share the same lower bound for c? After presenting and analyzing the data gathered, the hope is to find some sort of common thread or defining characteristic for these time-step restrictions.

5.1 L-stable second order DIRK

We start with the L-stable second order DIRK method paired with a DG discretization for $k = 1$. The value of its CFL number c as a function of d, computed with the algorithm discussed in Section 4.6, is plotted in Figure 5.1. Adding more dispersion helped and caused an increase in the value of c. The asymptotic behaviour of c as $d \to \infty$ appears to be $c \to \infty$. Much like the paper on finite difference methods paired with IMEX methods, the data showed that the lower bound of c is its value when $d = 0$. This value being $c = \frac{1}{3}$ [7]. This result was promising, and the hope was that if we looked at a third order IMEX and third order DG ($k = 2$) this same trend would be observed.

A limitation however was for how large a value of d data was gathered for. The values of d being tested stop for some value of d larger than 10^4. It was assumed that c became extremely large or unbounded after this point. It would be worthwhile to go back and put a maximum threshold on c to double check that the increasing trend continues. This is mainly because it was observed for third order IMEX methods that a decrease in c was possible.

When using the finite difference method, similar results were observed, and the numerically evaluated CFL number c is shown in Figure 5.2. Although it should be noted that the finite difference method used is third order, yet when $k = 1$ the DG

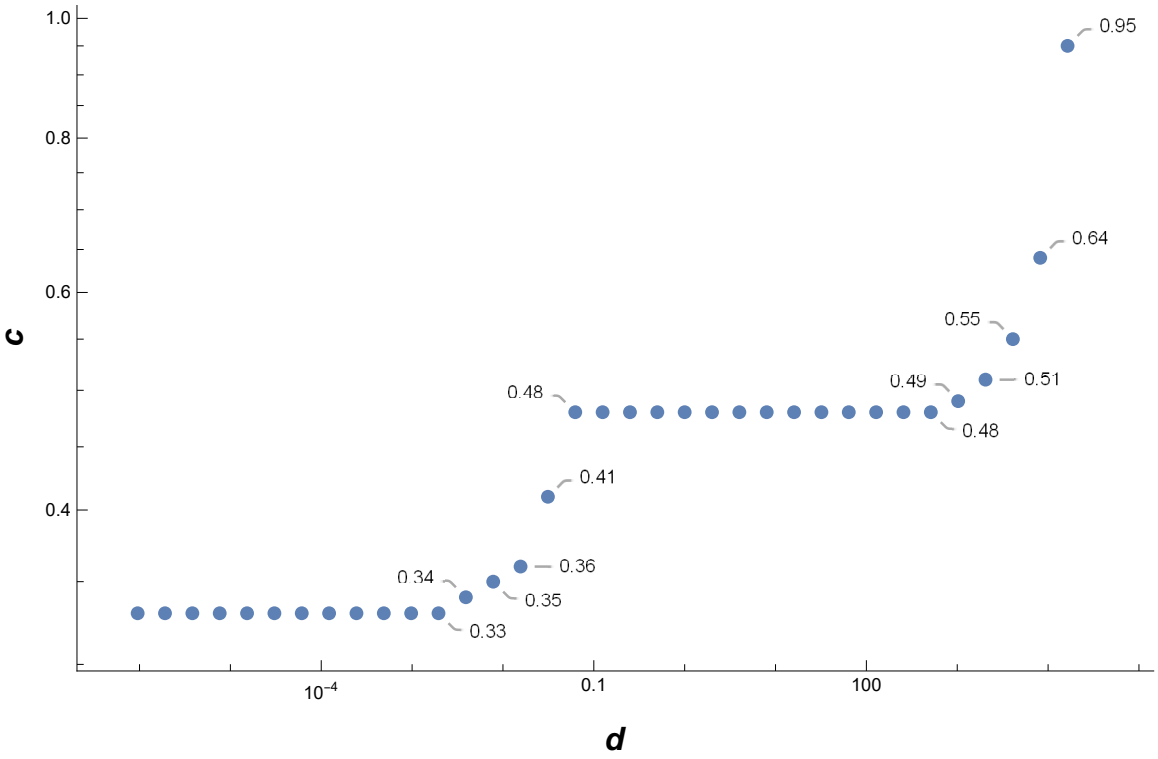

Figure 5.1: 2nd order IMEX. DG with $k = 1$. h is fixed to be 1.

discretization is second order. The asymptotic behaviour of c still seemed to become unbounded as $d \to \infty$, but in the figure the value of c was allowed to only increase to 10. It is also worth noting that somewhere between $d = 10^4$ and $d = 10^5$ is when the value of c drastically increases, just as it is for the case of $k = 1$. This is a nice result because so far it appears the asymptotic behaviour is the same between using FD and using DG. It should also be noted that the results of [13] are upheld by this data. That is the value of c when $d = 0$ is the lower bound.

This same second order IMEX was also analyzed briefly when using a third order DG discretization ($k = 2$). The numerically evaluated CFL number c of this method is shown in Figure 5.3. Based on previous results a significant value of c was not

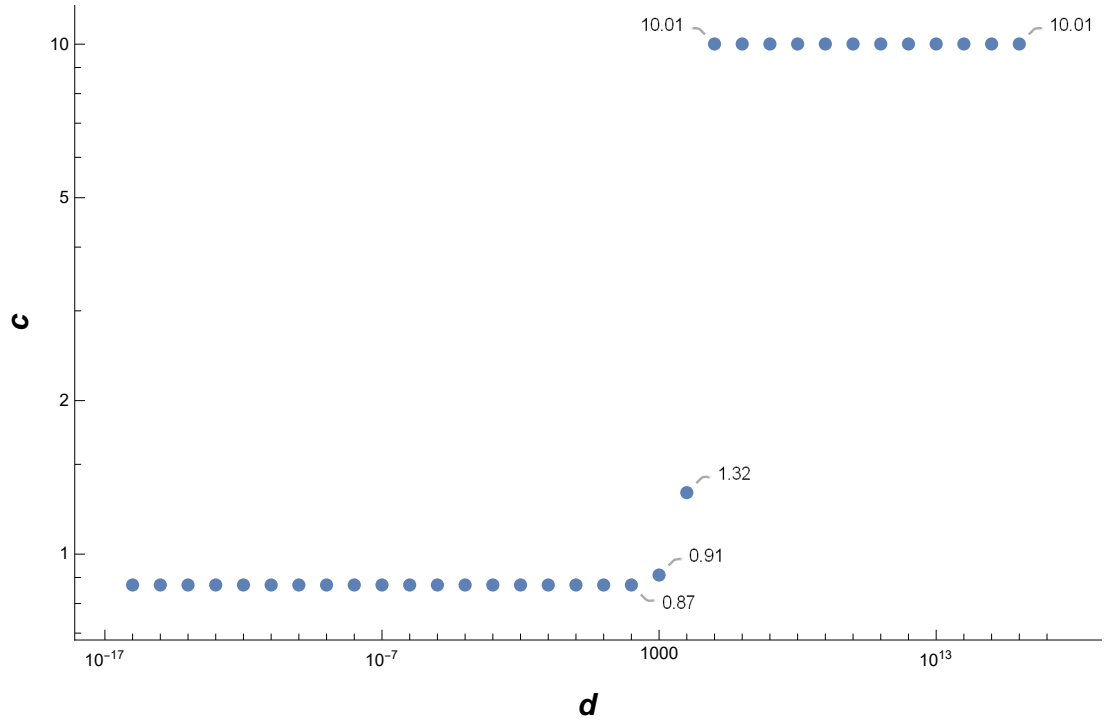

Figure 5.2: 2nd order IMEX with FD. h is fixed to be 1. The value of c is capped at 10.

expected because when $d = 0$ this pairing is unstable if $\frac{\tau}{h}$ is held constant [7], and we do observe a very small value of c for a small d. This suggests that this combination would not be a good choice numerically due to the strict time step constraint. It was not until a very large value of d, that a slightly larger value of c was observed. With h fixed as 1, the value of c did not increase from its lower bound until d was approximately 4000. Although it should be noted that this curve of values shifts left as h decreases, so a "large value of d" is relative. A similar threshold for d ($d \approx 3\times10^4$) was observed before the value of c became increasingly large or unbounded. Because c appeared to become so large, a threshold should be set to determine if this trend indeed continues.

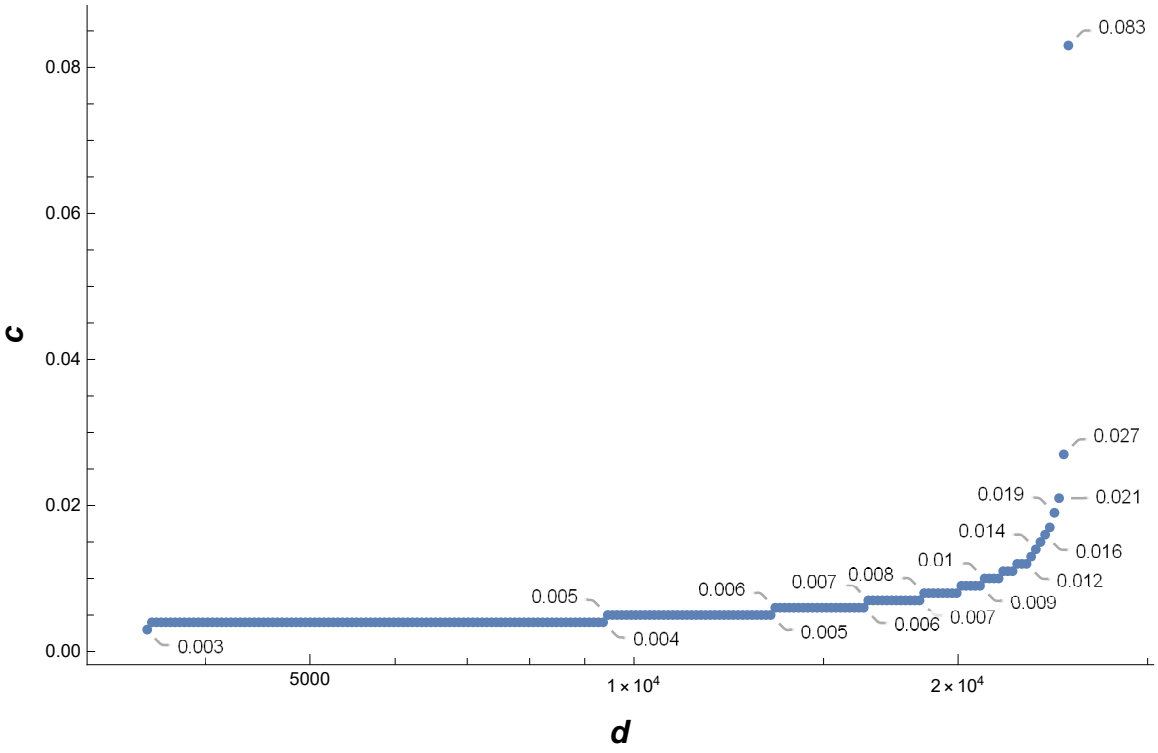

Figure 5.3: 2nd order IMEX. DG $k = 2$. h is fixed to be 1

Though it was not tested, it would be interesting to see what the asymptotic behaviour for the curve of c values would be if a DG discretization with $k = 3$ was used. Would we still have $c \to \infty$ as $d \to \infty$? Would we also observe the same behaviour if the the DG discretization was lowered to $k = 0$? Another interesting question to answer would be see how other second order IMEX methods behaved. As will be shown with the third order methods, the order of the IMEX method did not seem to be a unique identifier for the asymptotic behaviour of c.

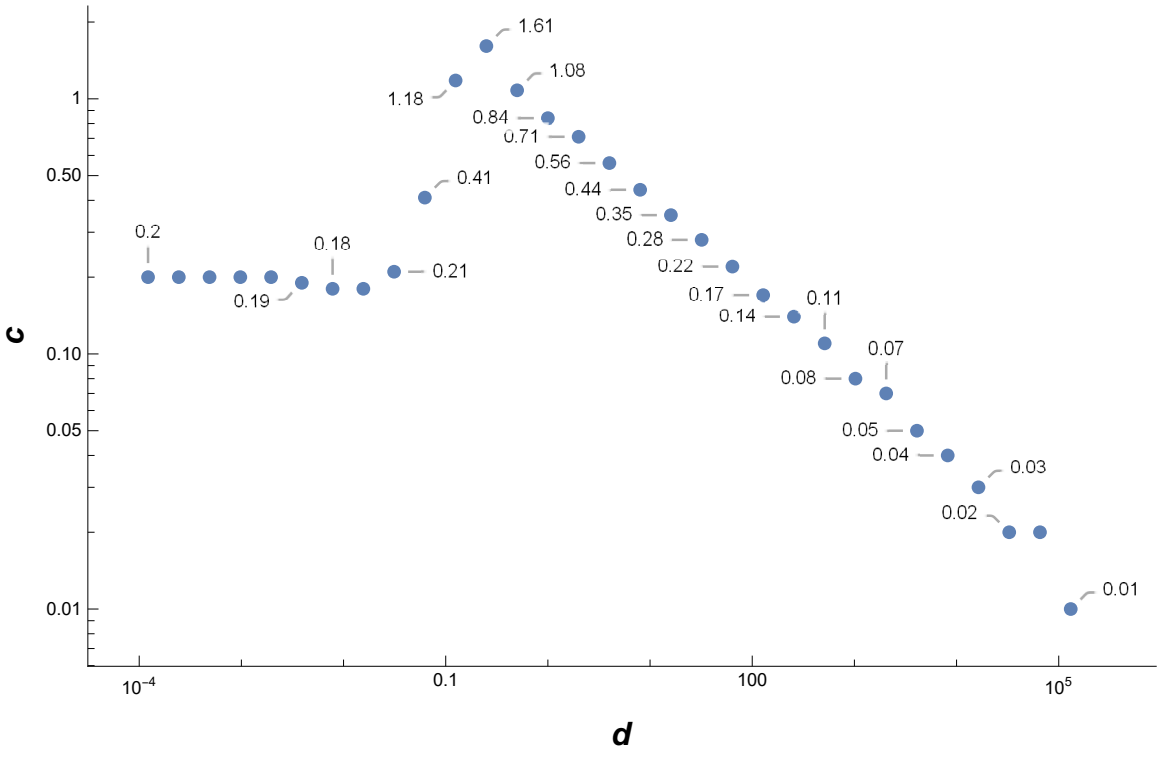

Figure 5.4: 3rd order IMEXSSP. DG $k = 2$. h is fixed to be 1.

5.2 Third order IMEXSSP3

The behaviour of this scheme when paired with a DG discretization for $k = 2$ was studied. The value of its CFL number c as a function of d is shown in Figure 5.4. One would expect that this method behaves similarly to the second order IMEX scheme, namely, as more dispersion was added, this would significantly alleviate the time-step restriction. But it does not seem to be the case. What was observed was that $c \to 0$ as $d \to \infty$. The value of c does not decrease to zero monotonically. It experiences a slight decrease near $d = 0.1$, and then increases up to $c \approx 1.61$ before heading to zero.

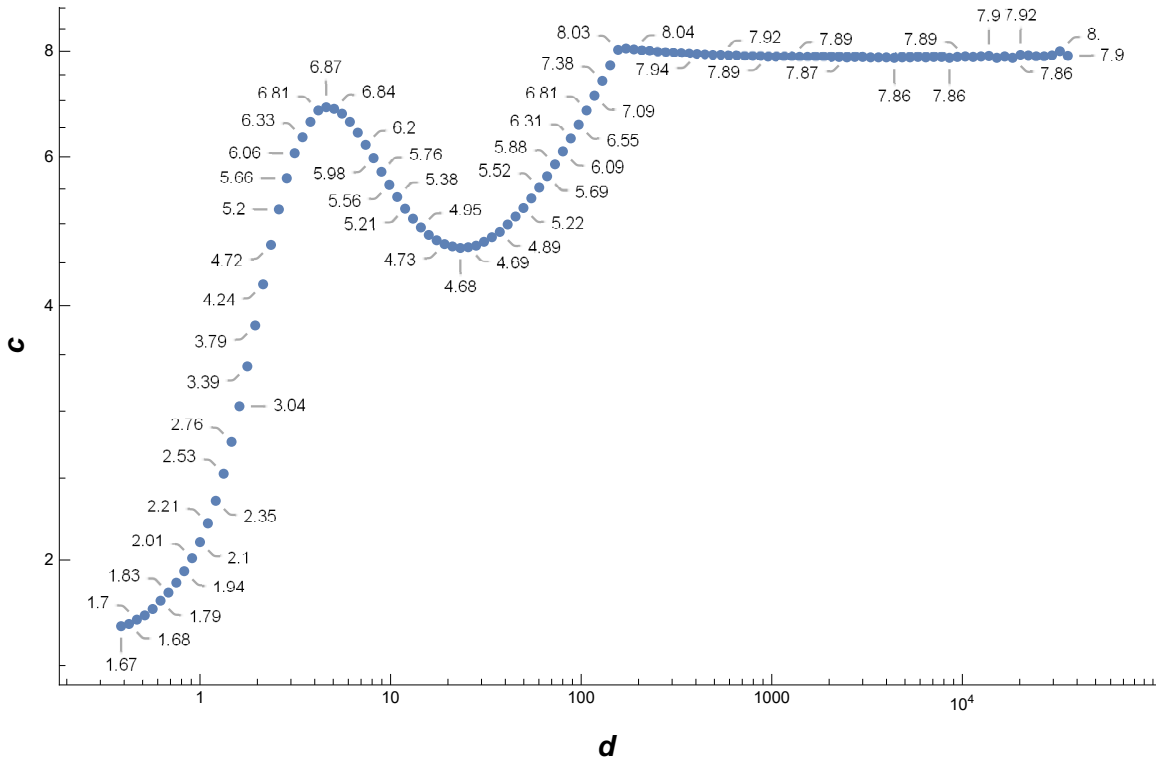

Figure 5.5: 3rd order IMEXSSP with 3rd order Finite Difference. h is fixed to be 1.

When the third order finite difference method was used instead of the DG method with $k = 2$, the asymptotic behaviour of c was starkly different. Of all the IMEX methods we studied in this thesis, this third order IMEXSSP3 method was the only one where this disconnect between the curve shapes was observed. For the finite difference, the asymptotic behaviour of c is that it will approach a fixed value between 7.9 and 8 as $d \to \infty$.

Since the finite difference method and the DG discretization have such starkly different curves for c values, one natural followup is to investigate the mechanism behind this, which will be studied in the future.

5.3 Third order combination

This third order combination IMEX method has a similar curve as the IMEXSSP scheme, and the value of its CFL number c as a function of d is shown in Figure 5.6. For both methods, the values of c start from the one when $d = 0$ (same as the one studied in [7]), and increase as the value of d increases. They reach a peak and then decrease below their starting value. The major difference is for the third order combination, the value of c does not asymptotically approach zero. Rather c asymptotically approaches 0.179. The drawback to this scheme is it doesn't allow much room for improvement with the maximal value of c. It only increases from 0.209 to 0.266, and its lower bound is 0.179. An advantage over the IMEXSSP3 scheme is that its c does not approach zero as d becomes large.

A common characteristic also seems to be that the curve of values for c will follow a similar pattern when a finite difference method is used instead of DG. The value of its CFL number c as a function of d is shown in Figure 5.7. The curve of values for this IMEX method also started at the value of c when $d = 0$, increased, and then decreased and stabilized at a lower value of c. This behaviour is very similar to the behaviour when a DG discretization was used. One interesting observation is the following. In [13], it was observed that, for several IMEX methods coupled with the finite difference scheme, the value of c when $d = 0$ is the lower bound of c for all values of d. However, this is not the case for this IMEX method (it should be noted this method was not studied in [13]). For values of d larger than approximately 0.1, $\rho(K) > 1$ for $\tau = 1.62h$.

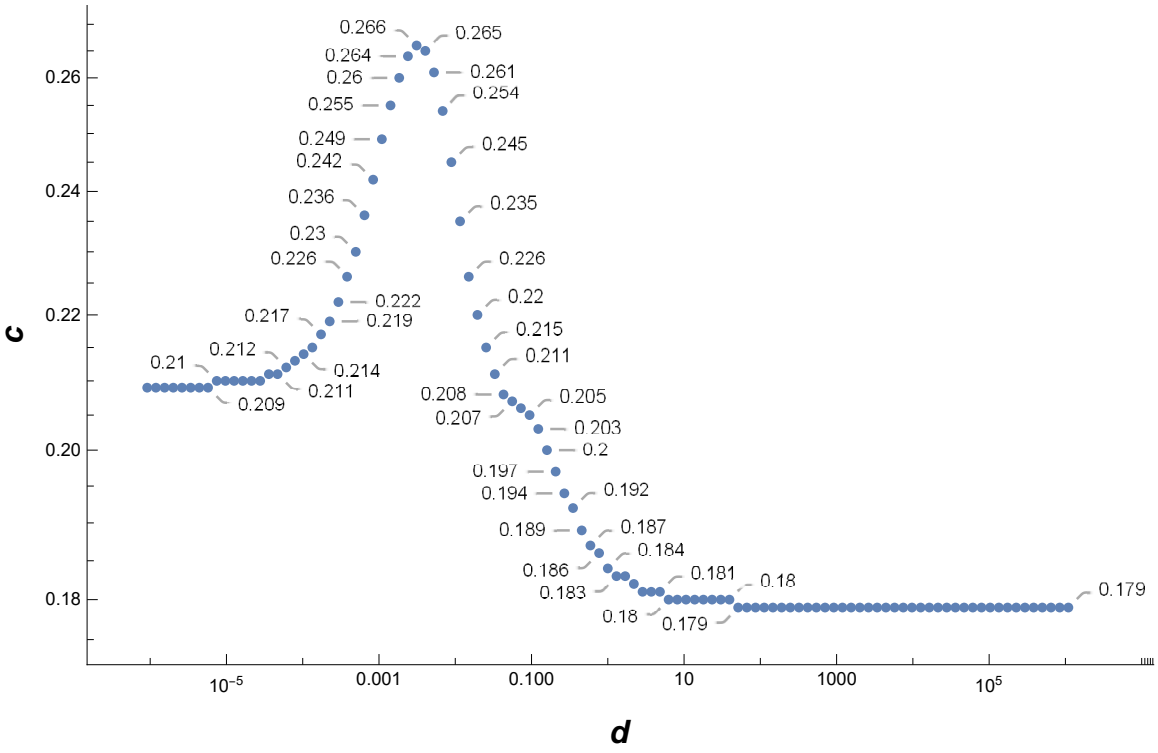

Figure 5.6: 3rd order combination. DG with $k = 2$. h is fixed to be 1.

5.4 L-stable third order DIRK

Next, we consider the L-stable third order DIRK method coupled with the third order DG method with $k = 2$. The value of its CFL number c as a function of d is shown in Figure 5.8. Of the third order methods seen so far, this scheme produces the nicest behaviour in the asymptotics $d \to \infty$, although it lacks the unconditional stability that comes with large values of d. The asymptotic behaviour of c is as $d \to \infty$ then $c \to 0.79$. Similar to the observations about finite difference methods, this IMEX scheme also has the property that the value of c when $d = 0$ is the lower bound for c. It should be noted that when $d = 0$ the value of c is 0.23, while for the other two third order methods seen this same value of c is 0.2. This difference is due

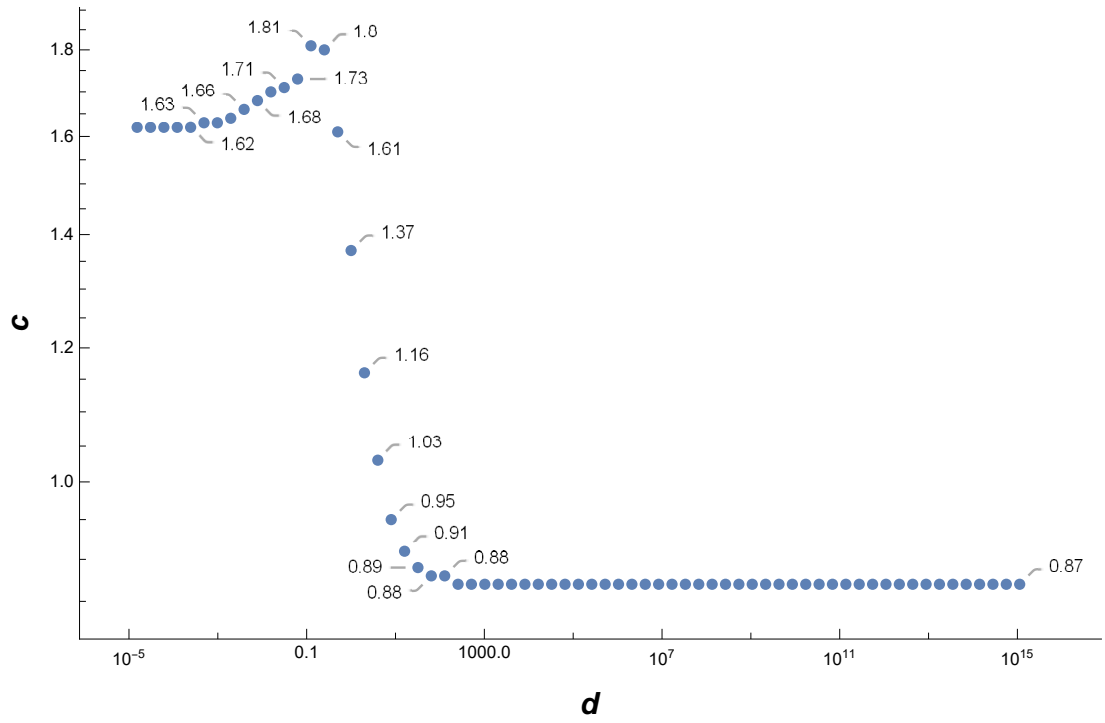

Figure 5.7: 3rd order combination with Finite Difference. h is fixed to be 1.

to the explicit method of the L-stable third order DIRK method having the stability region of a fourth order explicit Runge-Kutta method.

When the finite difference method is used, the asymptotic behaviour of c is the same when a DG discretization for $k = 2$ is used (see Figure 5.9). As $d \to \infty$ the value of c approaches a set value of 6.29. Both methods also produced a similar S-shaped curve.

The two previous third order IMEX methods examined in Sections 5.2 and 5.3 had third order explicit parts, while this L-stable DIRK IMEX method has a fourth order explicit part. Hence, we also analyzed this third order method when combined with a fourth order DG discretization with $k = 3$. The value of its CFL number c as a function of d is shown in Figure 5.10. We can observe that, raising the DG method

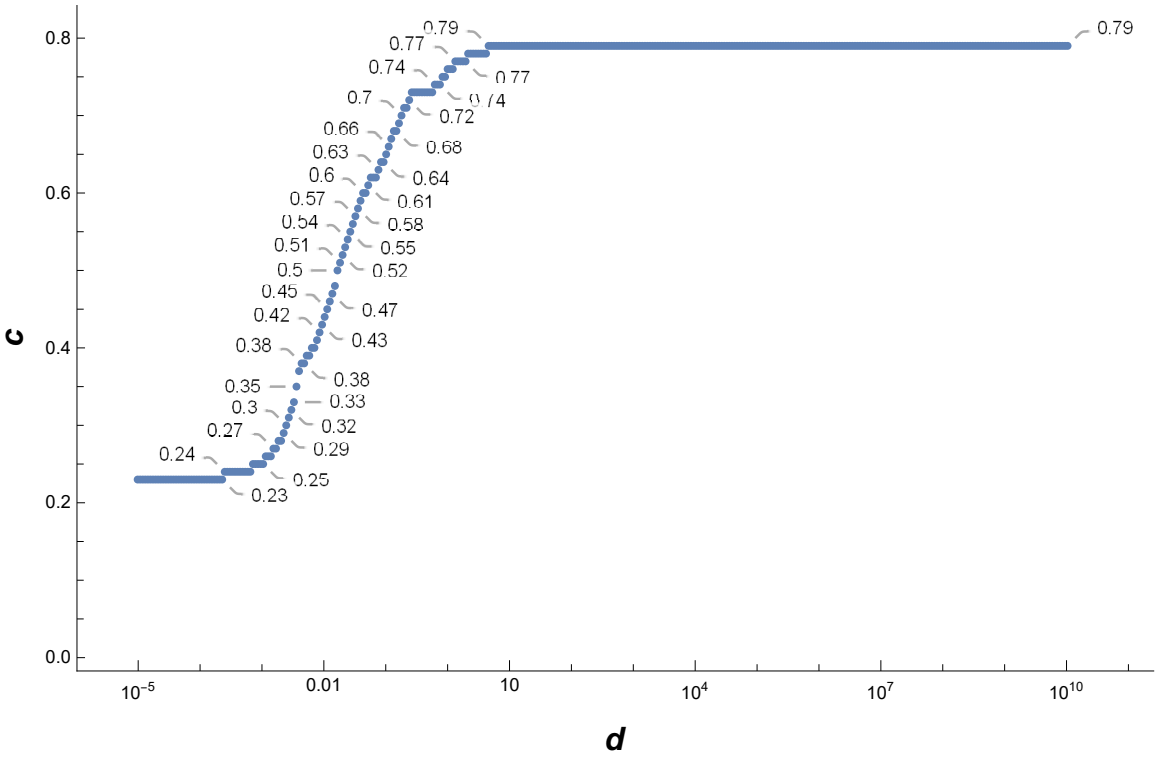

Figure 5.8: 3rd order DIRK. DG with $k = 2$. h is fixed to be 1.

to fourth order to match the explicit scheme did not change the asymptotic behaviour of c. The curve still approached a set value of $c = 0.49$ (even smaller than 0.79 when third order DG is used) as $d \to \infty$. The characteristics of the curve were also the same. The asymptotic value of c was the the maximum value of c and the curve was always increasing monotonically.

5.5 Alternate L-stable third order DIRK

Recall that this method is similar to the other L-stable third order DIRK discussed in Section 5.4. Its only difference is the choice of the coefficients for the explicit tableau. It shares the same weight vectors and implicit tableau as the other L-stable

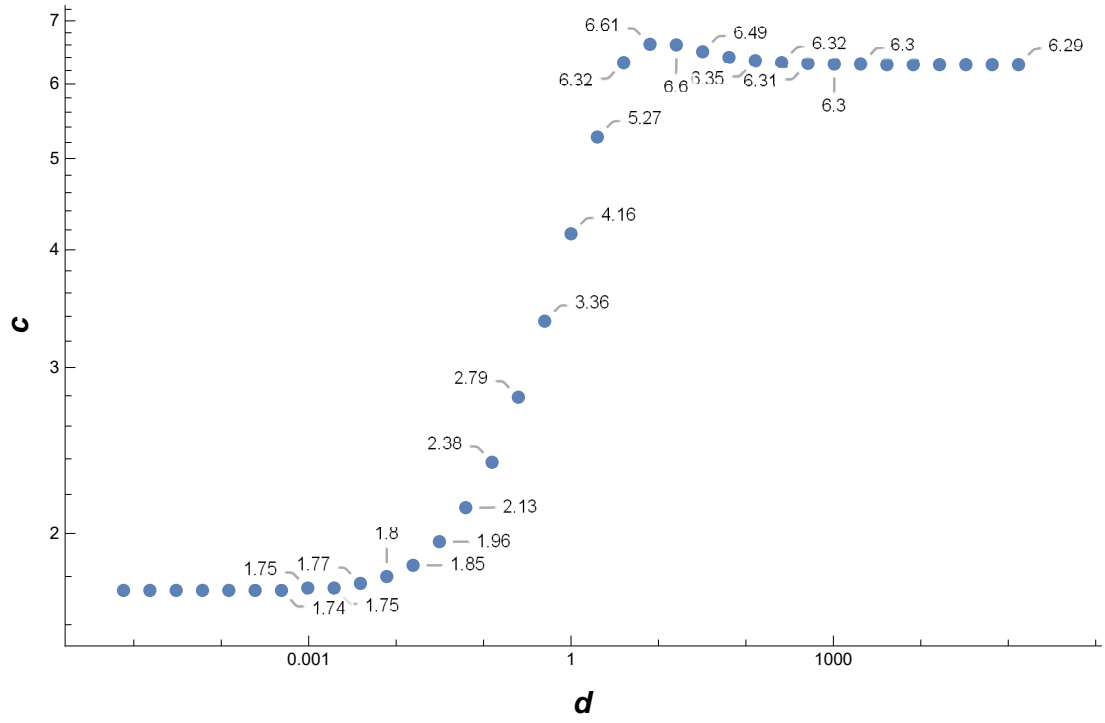

Figure 5.9: 3rd order DIRK Finite Difference. h is fixed to be 1.

third order DIRK. The expectation was this method would behave similarly and have the same asymptotic behaviour for c. Instead this IMEX method paired with a DG discretization for $k = 2$ seemed to have the asymptotic behaviour of when $d \to \infty$ then $c \to \infty$. Though this behaviour may not be clear from Figure 5.11, which stops at $d = 2^{16}$, for $d = 2^{20}$ and larger the value of c appeared to become extremely large if not unbounded. It is worth noting that the stability region for the explicit part is not that of a third order explicit Runge-Kutta method (see Figure 5.13). When $d = 0$ (or $d \approx 0$), the value of c is not 0.2, but instead $c = 0.01$. This value is shockingly low, and further work needs to be done to understand why the Fourier analysis would produce such a low value. However the explicit method is indeed third order. It

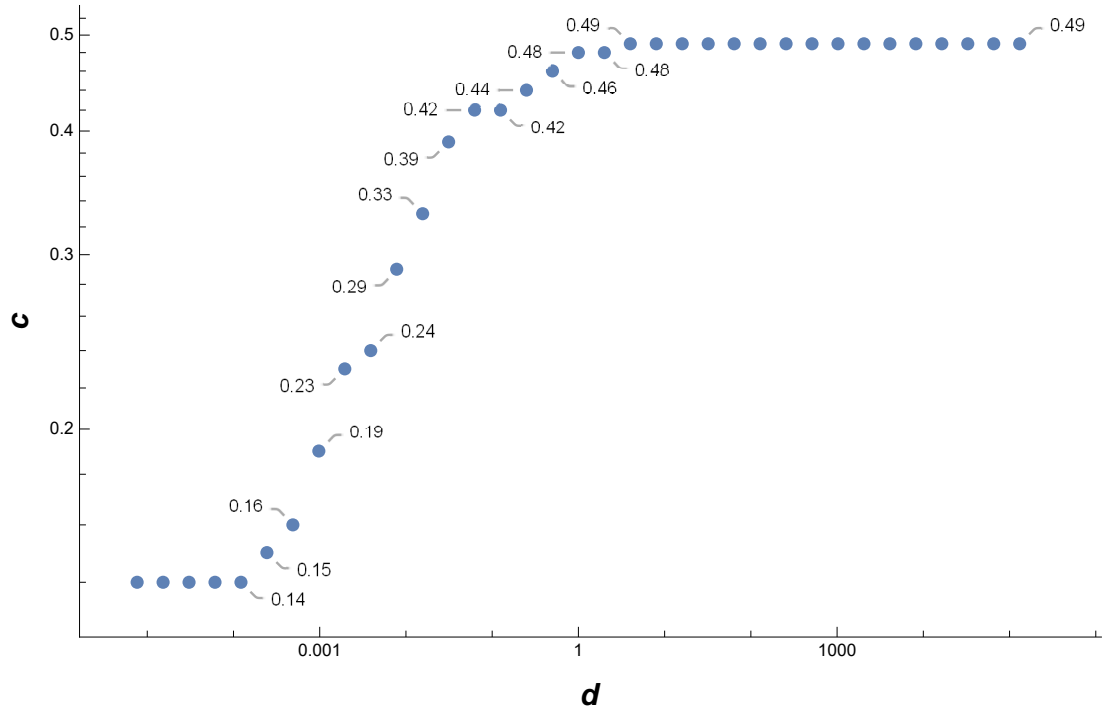

Figure 5.10: 3rd order DIRK. DG with $k = 3$. h is fixed to be 1.

satisfies all the order conditions for a third order RK method which can be found in [9].

When the finite difference is used on this method, the asymptotic behaviour of c is the same, which can be seen from Figure 5.12. As $d \to \infty$ so does $c \to \infty$. The most interesting aspect of the finite difference curve is the values of c for small d are not shockingly low. Instead they are comparable to the other third order methods in this paper. It is still not clear why using a DG discretization drastically lowers this value of c, but using a finite difference discretization seems to have no ill effect.

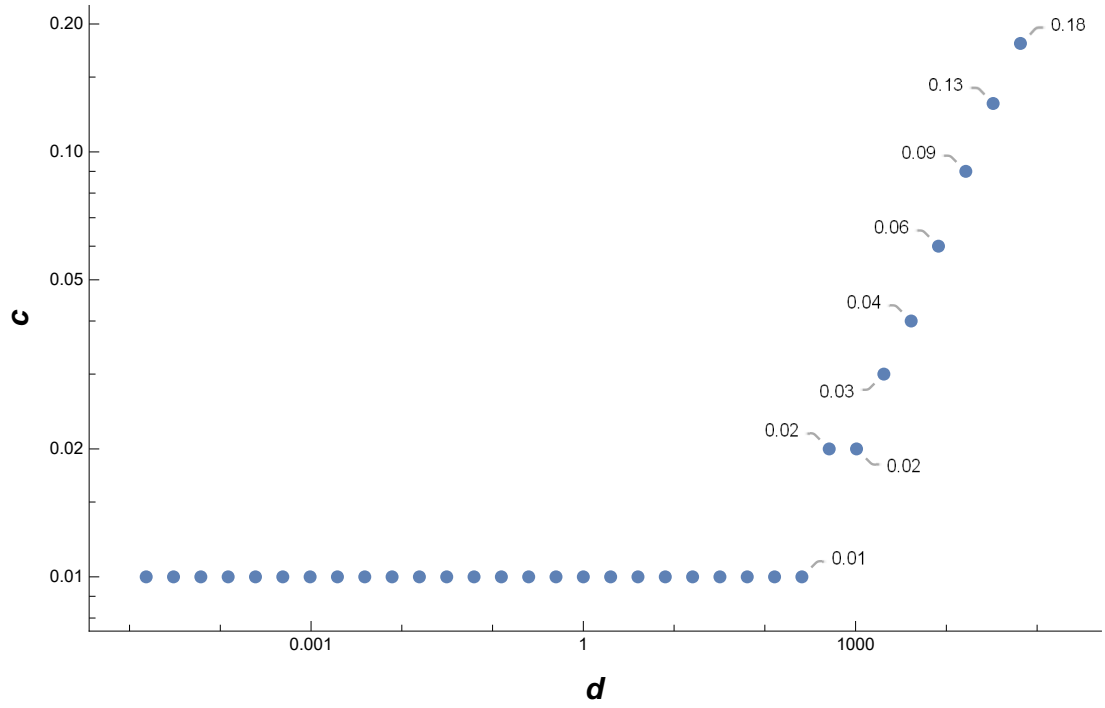

Figure 5.11: Alternate 3rd order DIRK. DG with $k = 2$. h is fixed to be 1. c is capped at 20.

5.6 General Observations

The asymptotic behaviour of the curve of c values fell into three categories as $d \to \infty$, which is equivalent to $h \to 0$ for a fixed d.

1. The value of c became very large or unbounded

2. The value of c approached a set value

3. The value of c approached zero

It is more beneficial to think of the asymptotic behaviour in terms of mesh refinement when considering which IMEX method will be useful. For a specific problem the value of d can't necessarily be changed, however the mesh size can always be refined.

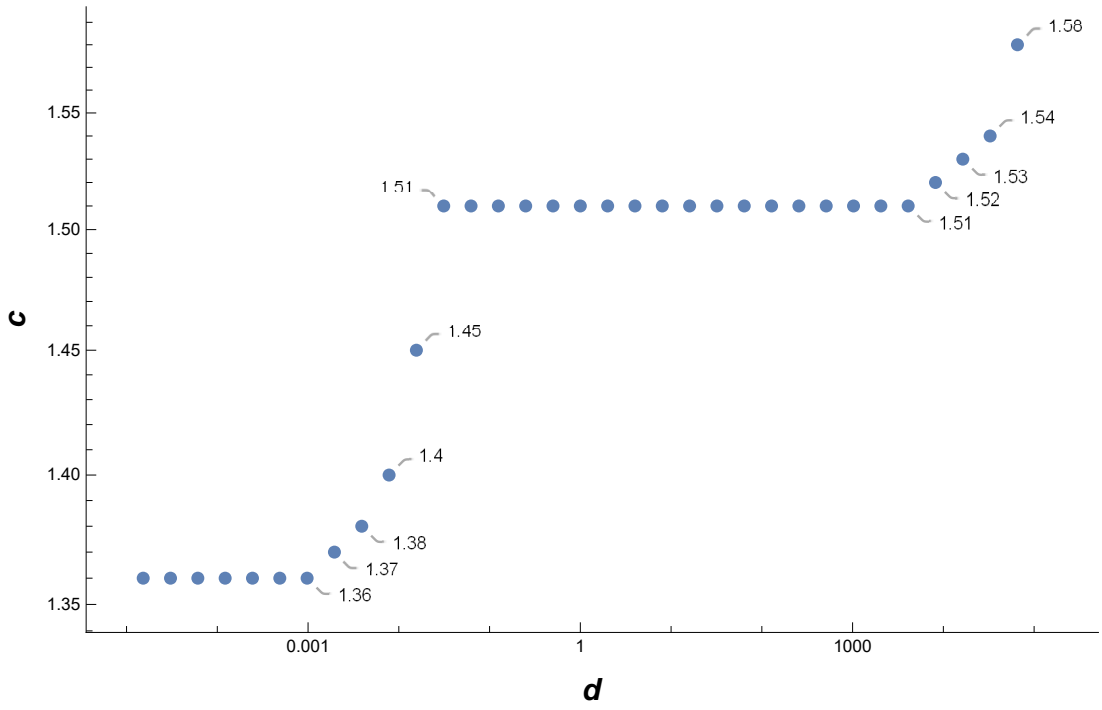

Figure 5.12: Alternate 3rd order DIRK Finite Difference. h is fixed to be 1.

order	method	k		
		1	2	3
2	L-stable	$c \to \infty$	$c \to \infty$	*
3	IMEXSSP3	*	$c \to 0$	*
3	Combination	*	$c \to 0.17$	*
3	L-stable	*	$c \to 0.79$	$c \to 0.49$
3	Alt. L-stable	*	$c \to \infty$	*

Table 5.1: The asymptotic behaviour of each method as $d \to \infty$, or equivalently as $h \to 0$ for a fixed d. * indicates the pairing was not tested.

The determination of this behaviour appears to be linked to the IMEX method itself. For the four third order IMEX methods analyzed, each of these three asymptotic behaviours was observed. However there are at least two variables, if not three, which play a role in the asymptotic behaviour: the order of the IMEX scheme, possibly the stability region for the explicit part in the IMEX scheme, and the order of the DG discretization. The behaviour could also be linked to the IMEX method itself. Table 5.1 presents these results. For methods that were tested for more than one k value, the asymptotic behaviour stayed the same. This could indicate that the DG discretization does not affect the asymptotic behaviour, but the table would need to be "filled in" for more weight to be put behind this hypothesis. Only one second order IMEX RK scheme was tested, so in the future it will be worthwhile to see how other second order schemes behave, i.e., will we see different asymptotic behaviour for different schemes. It will also be worthwhile to see how a second order SSP scheme would perform such as the ones presented in [9]. Would the c value also approach zero as it did for the third order SSP scheme?

An interesting observation is the two L-stable third order DIRKS only differed by their explicit tableaus, yet each method had two different asymptotic behaviours. The order of the IMEX method was the same, and the order of the DG discretization was the same. So perhaps the explicit scheme plays a strong role in determining the asymptotic behaviour of c as $h \to 0$ for a fixed value of d. To expand on this connection a little more, the explicit tableau of the alternate L-stable third order DIRK does not have the stability region of a third order explicit Runge-Kutta method. When $d = 0$ and a DG discretization with $k = 2$ is used, the value of $c \neq 0.209$. Instead this value of c is roughly 0.014. This suggests its stability region is not that of a third order

explicit Runge-Kutta method. In fact using the formula for the stability function

$$\phi(z) = \frac{\det(I - zA - zeb^T)}{\det(I - zA)},$$

where $z \in \mathbb{C}$, A is the explicit tableau of the method, e is a column vector of ones, and b^T is the weight vector for the explicit tableau, we can check the stability region of the explicit tableaus. Checking where $\phi(z) \leq 1$ for the alternate L-stable third order DIRK and comparing it to the stability region for third order explicit RK methods produced Figure 5.13. The other third order methods matched up with the third order explicit, or fourth order in the case of the L-stable third order DIRK, stability regions for RK methods.

These IMEX RK methods have been analyzed, but there is still the question of which method to choose. The choice could be made based on what happens as $h \to 0$ for a fixed value of d. In this case the larger value of c we can obtain the better. This would make the L-stable second order DIRK or the Alternate L-stable third order DIRK ideal. For these methods as $h \to 0$ for a fixed value of d the value of c will keep increasing. Although it should be noted that for larger h, the Alternate L-stable third order DIRK has a restrictive time-step. The L-stable DIRK would also be a good choice because for $k = 2$ as the mesh becomes refined c can be taken as large as 0.79.

Another way to decide which method to choose could be based on which method produces the largest c for a modest sized d. Although this becomes tricky because the figures present the values of c with $h = 1$. So as the value of h decreases these curves move to the left, which could cause the value of c to become drastically lower. If this approach is taken it would be better to consider the ratio of d/h^2. Then for a ratio of d/h^2 near the interval [0.1,10], the IMEXSSP3 method is ideal, having c values

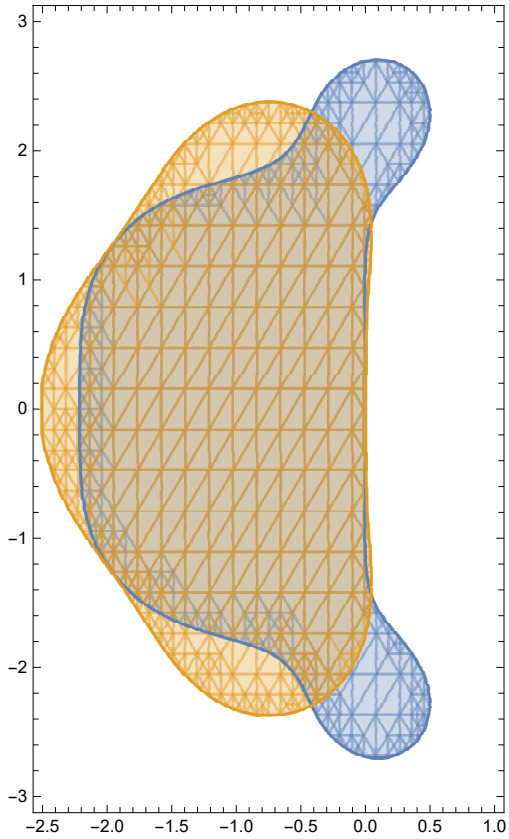

Figure 5.13: Stability region (in blue) of the explicit tableau for the Alternate L-stable DIRK compared to RK3 methods (in yellow)

between 0.41 and 1.61. The L-stable third order DIRK is also a good contender, achieving c values close to its maximum in this range. The other two third order methods are at their lowest values for d/h^2 near the interval [0.1,10].

A last consideration could be which method produces the largest value $C = \min(c)$. This should be close to if not the CFL condition $\tau \leq Ch$ which holds for all choices of d and h. In this case the L-stable third order DIRK is ideal. It produces the largest value of C being $C = 0.23$. The third order combination is also an option since it has $C = 0.17$. However the remaining third order methods would not be good

order	method	k		
		1	2	3
2	L-stable	0.33	0.003	*
3	IMEXSSP3	*	≈0	*
3	Combination	*	0.17	*
3	L-stable	*	0.23	0.14
3	Alt. L-stable	*	0.01	*

Table 5.2: The value $C = \min(c)$ for each method. Values highlighted match the value found in Table 2.2 of [7]. * indicates the pairing was not tested

choices. The IMEXSSP has its c values approaching 0, so we would expect C to be very small, and the Alternate L-stable third order DIRK has $C = 0.01$.

One last observation is that when using the finite difference method given by (2.11) and (2.12), the asymptotic behaviour of the c values seems to be the same as the asymptotic behaviour when a DG discretization was used. There was only one exception to this and that was with the IMEXSSP3 method. For finite difference c began to approach some value between 7.9 and 8, but for the DG discretization c went to zero. It should also be noted that using the third order finite difference gave less strict time-step restrictions than using a DG discretization, which is consistent with the case when explicit Runge-Kutta method is used.

Chapter 6: Numerical Experiments

Some numerical tests were carried out to determine the validity of the numerical results obtained from the Fourier analysis. We consider the following equation

$$\begin{cases} u_t + u_x + du_{xxx} = 0, & (x,t) \in [0, 2\pi] \times [0, T], \\ u_0 = \sin(x), & x \in [0, 2\pi], \end{cases} \tag{6.1}$$

with periodic boundary condition. This initial value boundary problem has the exact solution of the form $u(x,t) = \sin(x - (1-d)t)$.

After running some numerical experiments on the test equation (6.1), it seems to be the case that for small values of d the maximal value of c observed is a very strict upper bound. Increasing this c by 0.01 would cause the solution to blow up very quickly. When keeping the same h value but making d increasingly large, taking values of c larger than its maximal value did not appear to cause blowups in finite time, although it's possible that the final time T was not large enough for the blowup to be observed.

We start with testing the L-stable second order DIRK method, coupled with second order DG method with $k = 1$. In Table 6.1, we report the numerical error in L^2 norm and the convergence rate for $d = 0.5$, $h = \frac{2\pi}{N}$ and the final time $T = 100$. As can be seen from Table 6.1 the values of $\dfrac{d}{h^2}$ fall within the plateau where $c = 0.48$ for

the L-stable second order DIRK. Therefore this c value should be numerically stable, which appears to be validated from the table.

N	$\frac{d}{h^2}$	L^2 error	order
10	1.2665	3.1674	–
20	5.0661	0.8929	1.8267
40	20.264	0.22209	2.0074
80	81.057	0.05496	2.0147
160	324.23	0.013665	2.0079

Table 6.1: L-stable second order DIRK time-step restriction $\tau \leq ch$, $c = 0.48$

Table 6.2 presents similar results for the third order IMEXSSP3 using third order DG method with $k = 2$. We set $d = 0.000001$ and the final time as $T = 100$. As can be seen from Table 6.2 the values of $\frac{d}{h^2}$ fall within the range of c values which are all 0.2 for the IMEXSSP3 scheme. These values of c appear to be validated since the third order accuracy of the scheme is maintained.

N	$\frac{d}{h^2}$	L^2 error	order
10	2.533e-06	0.05633	–
20	1.0132e-05	0.0062108	3.181
40	4.0528e-05	0.0007486	3.0525
80	0.00016211	9.2608e-05	3.015
160	0.00064846	1.1538e-05	3.0048

Table 6.2: Third order IMEXSSP3 time-step restriction $\tau \leq ch$, $c = 0.2$

Table 6.3 presents the results for the third order combination scheme using third order DG method with $k = 2$. In the table $d = 0.00001$ and the final time is $T = 100$.

The values of $\frac{d}{h^2}$ all fall within a range where $c = 0.21$ would be a suitable value to use. The value appears to be validated since the method maintains third order accuracy. The values near the end of this curve were also validated in Table 6.4 using $d = 0.5$. The table shows that taking a value of c slightly above the one give in the curve results with blowup in the numerical solution.

N	$\frac{d}{h^2}$	L^2 error	order
10	2.533e-05	0.063899	–
20	0.00010132	0.0071294	3.1639
40	0.00040528	0.00086197	3.0481
80	0.0016211	0.0001067	3.0141
160	0.0064846	1.3295e-05	3.0046

Table 6.3: Third order combination time-step restriction $\tau \leq ch$, $c = 0.21$

	N	$\frac{d}{h^2}$	L^2 error	order
$c = 0.18$	10	1.2665	0.043239	–
	20	5.0661	0.0044106	3.2933
	40	20.264	0.00051829	3.0891
	80	81.057	6.3694e-05	3.0245
	160	324.23	7.9061e-06	3.0101
$c = 0.19$	10	1.2665	1.4193e+08	–
	20	5.0661	1.1443e+27	-62.806
	40	20.264	5.4828e+58	-105.24
	80	81.057	3.8008e+122	-212.07
	160	324.23	4.0283e+258	-451.87

Table 6.4: Third order combination. Comparison between $c = 0.18$ and $c = 0.19$ at $T = 100$

Table 6.5 presents the results for the L-stable third order DIRK using third order DG method with $k = 2$. In the table $d = 0.5$ and the final time is $T = 100$. Using $c = 0.7$ should be numerically stable based on the results presented in Chapter 5 and this table verifies that since the method maintains its third order accuracy. One interesting feature about this when method compared with its results when a finite difference discretization is used is the finite difference and DG curve differ slightly in shape. It was observed that for nearly all methods the finite difference and DG curve had the same general shape. However for this method, the finite difference curve asymptotically approaches its set value from above, while the DG curve asymptotically approaches its set value from below. Table 6.6 presents some evidence that the DG curve might approach from above as the finite difference curve does. It's possible there was some sort of bad approximation of a value or not enough accuracy being used which could've caused the Fourier analysis to undershoot the value of c when DG was used. It's also possible T was not large enough for the blowup to be observed.

N	$\frac{d}{h^2}$	L^2 error	order
10	1.2665	0.35354	–
20	5.0661	0.042605	3.0528
40	20.264	0.0052476	3.0213
80	81.057	0.00065304	3.0064
160	324.23	8.1471e-05	3.0028

Table 6.5: L-stable third order DIRK time-step restriction $\tau \leq ch$, $c = 0.7$

Table 6.7 and 6.8 present the results of the alternate L-stable third order DIRK using third order DG method with $k = 2$. Because the results of the Fourier analysis

	N	$\frac{d}{h^2}$	L^2 error	order
	10	1.2665	0.50304	–
	20	5.0661	0.061198	3.0391
$c = 0.79$	40	20.264	0.0075404	3.0208
	80	81.057	0.00093861	3.006
	160	324.23	0.00011727	3.0007
	10	1.2665	0.54117	–
	20	5.0661	0.065966	3.0363
$c = 0.81$	40	20.264	0.0081275	3.0208
	80	81.057	1.1614	-7.1589
	160	324.23	3.8355e+10	-34.943
	10	1.2665	0.58333	–
	20	5.0661	0.0709	3.0405
$c = 0.83$	40	20.264	328.29	-12.177
	80	81.057	1.0843e+16	-44.909
	160	324.23	6.0073e+42	-88.84

Table 6.6: Comparison of $c = 0.79$, $c = 0.81$ and $c = 0.83$ with $d = 0.5$ and $T = 100$

showed for small d or $d = 0$ that $c = 0.01$, we tested to see if numerically c could be larger. In Table 6.7 we take $d = 0$, $c = 0.13$ and a final time of $T = 10000$. We then do the same in Table 6.8 but with $d = 0.0001$. Keeping in mind the behaviour of c

N	$\frac{d}{h^2}$	L^2 error	order
10	0	0.014828	–
20	0	0.069002	-2.2183
40	0	0.010795	2.6762
80	0	0.0014168	2.9297
160	0	0.00017917	2.9833

Table 6.7: Alternate third order DIRK time-step restriction $\tau \leq ch$, $c = 0.13$

as $h \to 0$ for a fixed d, these two tables make sense in that respect. As N increase h becomes smaller so we should be able to take a larger value of c. This seems to

N	$\frac{d}{h^2}$	L^2 error	order
10	0.0002533	0.014215	–
20	0.0010132	0.070476	-2.3097
40	0.0040528	0.010891	2.694
80	0.016211	0.0014221	2.937
160	0.064846	0.00017945	2.9864

Table 6.8: Alternate third order DIRK time-step restriction $\tau \leq ch$, $c = 0.13$

explain the negative order with $N = 20$ compared to $N = 10$. But what is yet to be understood is why for this method there is not extreme blowup when when c it taken to be much larger than 0.01. Under similar conditions, small d and large T, the other methods would experience blowup by taking $c = c + 0.01$. This lack of blow-up could be because the initial condition of $\sin(x)$ does not correspond to an ansatz which would cause blow-up. Based on the Fourier analysis, there should be an inital condition corresponding to an ansatz for which $\rho(K) > 1$ if $c > 0.01$.

Tables 6.9, 6.10, 6.11, and 6.12 present a comparison of each third order method using a DG discretization with $k = 2$ where the value of $\frac{d}{h^2}$, N, and c are fixed. In each table $N = 80$ and the value of d is chosen such that at most three of the four methods will be stable for the choice of c. Then taking c as $c = c + 0.01$, at least one of the methods will become unstable. And we have chosen different parameter setup so that every one of these four methods has demonstrated the instability. Table 6.12 is of particular interest because it seems to suggest that for the Alternate L-stable third order DIRK c does not need to be taken as $c = 0.01$ for the method to be stable.

c	N	$\frac{d}{h^2}$	Method	L^2 error
0.18	80	0.01	IMEXSSP3	6.8506e-05
			Combination	6.8525e-05
			L-stable	1.7232e-05
			Alt. L-stable	4.1904e-05
0.19	80	0.01	IMEXSSP3	2.9262e+81
			Combination	7.9865e-05
			L-stable	1.7227e-05
			Alt. L-stable	4.8129e-05

Table 6.9: Numerical Blowup in the IMEXSSP3 scheme. Final time is $T = 100$

c	N	$\frac{d}{h^2}$	Method	L^2 error
0.18	80	0.5	IMEXSSP3	6.8457e-05
			Combination	6.93e-05
			L-stable	1.7203e-05
			Alt. L-stable	4.1482e-05
0.19	80	0.5	IMEXSSP3	7.9767e-05
			Combination	0.64312
			L-stable	1.7199e-05
			Alt. L-stable	4.7642e-05
0.20	80	0.5	IMEXSSP3	9.2425e-05
			Combination	2.866e+148
			L-stable	1.7193e-05
			Alt. L-stable	5.4631e-05

Table 6.10: Numerical Blowup in the third order combination scheme. Final time is $T = 100$

c	N	$\frac{d}{h^2}$	Method	L^2 error
0.79	80	10	IMEXSSP3	0.049237
			Combination	NaN
			L-stable	0.0033485
			Alt. L-stable	0.024861
0.80	80	10	IMEXSSP3	0.05111
			Combination	NaN
			L-stable	9.2987e+37
			Alt. L-stable	0.025815

Table 6.11: Numerical Blowup in L-stable third order DIRK. Final time is $T = 1000$

c	N	$\frac{d}{h^2}$	Method	L^2 error
0.18	80	0.0001	IMEXSSP3	6.858e-05
			Combination	6.858e-05
			L-stable	1.7264e-05
			Alt. L-stable	4.1866e-05
0.19	80	0.0001	IMEXSSP3	7.9918e-05
			Combination	7.9918e-05
			L-stable	1.7258e-05
			Alt. L-stable	1.2378e+139

Table 6.12: Numerical Blowup in Alternate L-stable third order DIRK. Final time is $T = 100$

Chapter 7: Conclusions and Future Work

The aim of this thesis was to analyze the stability regions of IMEX methods when paired with the DG spatial discretization. One initial hope was to create a table similar to Table 2.2 in [7]. This would be like a table which would give the maximum time-step restriction for a given v^{th} order IMEX method paired with a DG discretization using polynomials up to degree k. However because the PDE being analyzed was (4.1), it didn't make much sense to pin down a single value for the time-step restriction. Since the value of d could vary, so could the value of the time-step restriction. So instead a curve of these values for a selection of d was found for each IMEX method and DG pairing. The other initial hope was that as more dispersion was added and the value of d became larger, the time-step restriction would increase. If the explicit part in the IMEX scheme matched those in [7], the hope was the value for the pairing with $d = 0$ found in Table 2.2 of [7] would be the lower bound of the curve. Then as d increased, so would the values on the curve. Yet it was discovered that this was not the case for most IMEX methods and DG pairings. The asymptotic behaviour of the curve as $d \to \infty$ was not necessarily unbounded. It could also approach zero or another set value. These results were interesting because they went against the expectation, and because it was found that different IMEX methods of the same order with the same DG discretization had different asymptotic

behaviours for the curve of time-step restriction values. The difference in asymptotic behaviour seem to come from the IMEX methods themselves.

In order to figure out which variables have an influence on the asymptotic behaviour more research is needed regarding the time step constraints of IMEX methods paired with a DG discretization. It will be beneficial to figure out how the order of the IMEX method, the order of the DG discretization, and the explicit and implicit methods themselves play a role in the asymptotic behaviour of the curve of time-step restrictions. Once this analysis is carried out, organizing the data into a chart would be useful for identifying any patterns. A chart for this purpose has been used presented in this thesis, however it is not fully filled in. The data for $k = 1$ or $k = 3$ is missing for most of the methods. Filling in and expanding this chart might make it clear if the IMEX order or the DG order has an effect on the time step constraint.

Another aspect worth looking into might be a more careful selection of IMEX methods. A lot of these IMEX methods appear to have been designed with convection-diffusion equations or hyperbolic equations with stiff sources in mind. It will be worthwhile to search the literature on IMEX methods to see if any have been designed for convection-dispersion PDEs. These may have better results (possibly the expected results) than the ones analyzed here. This thesis could also be a springboard for designing IMEX methods for convection-dispersion PDEs, i.e., how can one avoid having results similar to the IMEXSSP3 method or the third order combination IMEX method?

Other future work should be directed at determining if the results found in this thesis will hold for a nonlinear convection-dipserion PDE. The analysis here is carried

out on a linear PDE since the Fourier analysis is utilized, and the numerical experiments seemed to work for a test linear PDE. Whether these time-step restrictions work for a nonlinear PDE could be investigated numerically.

Bibliography

[1] Uri M. Ascher, Steven J. Ruuth, and Raymond J. Spiteri. Implicit-explicit runge-kutta methods for time-dependent partial differential equations. *Applied Numerical Mathematics*, 25:151–167, 1997.

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

[3] F. Bassi and S. Rebay. A high-order accurate discontinuous finite element method for the numerical solution of the compressible navier–stokes equations. *Journal of Computational Physics*, 131(2):267–279, 1997.

[4] M.P. Calvo, J. de Frutos, and J. Novo. Linearly implicit runge-kutta methods for advection-reaction-diffusion equations. *Applied Numerical Mathematics*, 37(4):535–549, 2001.

[5] Bernardo Cockburn, George E. Karniadakis, and Chi-Wang Shu. *The Development of Discontinuous Galerkin Methods*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2000.

[6] Bernardo Cockburn and Chi-Wang Shu. The local discontinuous galerkin method for timedependent convection-diffusion systems. *SIAM Journal on Numerical Analysis*, 35(6):2440–2463, 1998.

[7] Bernardo Cockburn and Chi-Wang Shu. Runge–kutta discontinuous galerkin methods for convection-dominated problems. *Journal of Scientific Computing*, 16(3):173–261, 2001.

[8] Michel Crouzeix. Une méthode multipas implicite-explicite pour l'approximation des équations d'évolution paraboliques. *Numerische Mathematik*, 35:257–276, 1980.

[9] Lorenzo Pareschi and Giovanni Russo. Implicit-explicit runge-kutta schemes and applications to hyperbolic systems with relaxation. *Journal of Scientific Computing*, 25:129–155, 2005.

[10] W.H. Reed and T. Hill. Triangular mesh methods for the neutron transport equation. Tech. report la-ur-73-479, Los Alamos Scientific Laboratory, 1973.

[11] Chi-Wang Shu. Discontinuous galerkin methods: General approach and stability. In S. Bertoluzza, S. Falletta, G. Russo, and C.-W. Shu, editor, *Numerical Solutions of Partial Differential Equations*, pages 149–201. Advanced Courses in Mathematics CRM Barcelona, Birkhauser, Basel, 2008.

[12] Zheng Sun and Chi-Wang Shu. Strong stability of explicit Runge–Kutta time discretizations. *SIAM Journal on Numerical Analysis*, 57(3):1158–1182, 2019.

[13] Meiqi Tan, Juan Cheng, and Chi-Wang Shu. Stability of high order finite difference schemes with implicit-explicit time-marching for convection-diffusion and convection-dispersion equations. *International Journal of Numerical Analysis and Modeling*, 18(3):362–383, 2021.

[14] Haijin Wang, Chi-Wang Shu, and Qiang Zhang. Stability and error estimates of local discontinuous galerkin methods with implicit-explicit time-marching for convection-diffusion problems. *SIAM Journal on Numerical Analysis*, 53(1):206–227, 215.

Appendix A: Coefficient Matrices for DG Space Discretization

The way to derive these matrices is very formulaic. One can simply write a code which can derive these matrices quickly. The final matrices are dependent upon five matrices (actually six, but that one is the identity), which come from the semi-discrete form of the PDE. There are two matrices from the integral terms, and four matrices resulting from the different combination of numerical flux terms. The identity matrix comes from the first integral term which involves a time derivative or no derivatives. The matrix M_1 comes from the second integral term which involves a spacial derivative of the test function. M_2, M_3, M_4, and M_5 come from the multiplication of flux terms.

Letting $a = x_{j-1/2}$, $b = x_{j+1/2}$, and recalling that $\psi_{j,k}(x^-_{j-1/2}) = \psi_{j-1,k}(x_{j-1/2})$ and $\psi_{j,k}(x^+_{j+1/2}) = \psi_{j+1,k}(x_{j+1/2})$ we have the following formulas for these matrices.

$$M_1 = \begin{bmatrix} \int_{I_j} \psi_{j,0}\psi'_{j,0}\,dx & \int_{I_j} \psi_{j,1}\psi'_{j,0}\,dx & \dots & \int_{I_j} \psi_{j,n}\psi'_{j,0}\,dx \\ \int_{I_j} \psi_{j,0}\psi'_{j,1}\,dx & \int_{I_j} \psi_{j,1}\psi'_{j,1}\,dx & \dots & \int_{I_j} \psi_{j,n}\psi'_{j,1}\,dx \\ \vdots & \vdots & \ddots & \vdots \\ \int_{I_j} \psi_{j,0}\psi'_{j,n}\,dx & \int_{I_j} \psi_{j,1}\psi'_{j,n}\,dx & \dots & \int_{I_j} \psi_{j,n}\psi'_{j,n}\,dx \end{bmatrix},$$

$$M_2 = \begin{bmatrix} \psi_{j,0}(b)\psi_{j,0}(b) & \psi_{j,1}(b)\psi_{j,0}(b) & \dots & \psi_{j,n}(b)\psi_{j,0}(b) \\ \psi_{j,0}(b)\psi_{j,1}(b) & \psi_{j,1}(b)\psi_{j,1}(b) & \dots & \psi_{j,n}(b)\psi_{j,1}(b) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{j,0}(b)\psi_{j,n}(b) & \psi_{j,1}(b)\psi_{j,n}(b) & \dots & \psi_{j,n}(b)\psi_{j,n}(b) \end{bmatrix},$$

$$M_3 = \begin{bmatrix} \psi_{j-1,0}(b)\psi_{j,0}(a) & \psi_{j-1,1}(b)\psi_{j,0}(a) & \dots & \psi_{j-1,n}(b)\psi_{j,0}(a) \\ \psi_{j-1,0}(b)\psi_{j,1}(a) & \psi_{j-1,1}(b)\psi_{j,1}(a) & \dots & \psi_{j-1,n}(b)\psi_{j,1}(a) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{j-1,0}(b)\psi_{j,n}(a) & \psi_{j-1,1}(b)\psi_{j,n}(a) & \dots & \psi_{j-1,n}(b)\psi_{j,n}(a) \end{bmatrix},$$

$$M_4 = \begin{bmatrix} \psi_{j+1,0}(a)\psi_{j,0}(b) & \psi_{j+1,1}(a)\psi_{j,0}(b) & \dots & \psi_{j+1,n}(a)\psi_{j,0}(b) \\ \psi_{j+1,0}(a)\psi_{j,1}(b) & \psi_{j+1,1}(a)\psi_{j,1}(b) & \dots & \psi_{j+1,n}(a)\psi_{j,1}(b) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{j+1,0}(a)\psi_{j,n}(b) & \psi_{j+1,1}(a)\psi_{j,n}(b) & \dots & \psi_{j+1,n}(a)\psi_{j,n}(b) \end{bmatrix},$$

$$M_5 = \begin{bmatrix} \psi_{j,0}(a)\psi_{j,0}(a) & \psi_{j,1}(a)\psi_{j,0}(a) & \dots & \psi_{j,n}(a)\psi_{j,0}(a) \\ \psi_{j,0}(a)\psi_{j,1}(a) & \psi_{j,1}(a)\psi_{j,1}(a) & \dots & \psi_{j,n}(a)\psi_{j,1}(a) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{j,0}(a)\psi_{j,n}(a) & \psi_{j,1}(a)\psi_{j,n}(a) & \dots & \psi_{j,n}(a)\psi_{j,n}(a) \end{bmatrix}.$$

For a DG discretization of degree k, these matrices only need to be taken as the $(k+1) \times (k+1)$ upper-left sub-matrix.

From the semi-discrete form of the PDE there are three separate equations. One in which q can be solved for in terms of u; one in which p can be solved in terms of q; and one in which u' can be solved in terms of p and u. These three systems have the following general solutions

$$\begin{aligned} q_j &= \frac{1}{h}Q_1 u_j + \frac{1}{h}Q_2 u_{j-1}, \\ p_j &= \frac{1}{h}P_1 q_{j+1} + \frac{1}{h}P_2 q_j, \\ u'_j &= \frac{1}{h}U_1 u_j + \frac{1}{h}U_2 u_{j-1} + \frac{1}{h}U_3 p_{j+1} + \frac{1}{h}U_4 p_j. \end{aligned}$$

Here $Q_1 = M_2 - M_1$, $Q_2 = -M_3$, $P_1 = M_4$, $P_2 = -M_5 - M_1$, $U_1 = M_1 - M_2$, $U_2 = M_3$, $U_3 = -M_4$, and $U_4 = M_1 + M_5$.

Performing back substitution to get u'_j written entirely in terms of u results with the equation

$$\begin{aligned} u'_j =& \frac{1}{h}(U_1 u_j + U_2 u_{j-1}) + \\ & \frac{1}{h^3}(U_3 P_1 Q_1 u_{j+2} + (U_3 P_1 Q_2 + U_3 P_2 Q_1 + U_4 P_1 Q_1) u_{j+1} \\ & + (U_3 P_2 Q_2 + U_4 P_1 Q_2 + U_4 P_2 Q_1) u_j + U_4 P_2 Q_2 u_{j-1}) \end{aligned}$$

Thus making the matrices from (4.2) be $C_1 = U_1$, $C_2 = U_2$, $D_1 = U_3P_1Q_1$, $D_2 = U_3P_1Q_2 + U_3P_2Q_1 + U_4P_1Q_1$, $D_3 = U_3P_2Q_2 + U_4P_1Q_2 + U_4P_2Q_1$, and $D_4 = U_4P_2Q_2$, Below are these matrices for DG discretizations with $k = 0, 1, 2, 3, 4$.

$k = 0$

$C_1 = -1$, $C_2 = 1$, $D_1 = -1$, $D_2 = 3$, $D_3 = -3$, and $D_4 = 1$

$k = 1$

$$C_1 = \begin{bmatrix} -1 & -\sqrt{3} \\ \sqrt{3} & -3 \end{bmatrix}, C_2 = \begin{bmatrix} 1 & \sqrt{3} \\ -\sqrt{3} & -3 \end{bmatrix}, D_1 = \begin{bmatrix} 8 & -4\sqrt{3} \\ 8\sqrt{3} & -12 \end{bmatrix},$$
$$D_2 = \begin{bmatrix} -6 & -18\sqrt{3} \\ 18\sqrt{3} & -90 \end{bmatrix}, D_3 = \begin{bmatrix} -12 & 12\sqrt{3} \\ -24\sqrt{3} & -108 \end{bmatrix}, \text{ and } D_4 = \begin{bmatrix} 10 & 10\sqrt{3} \\ -2\sqrt{3} & -6 \end{bmatrix}.$$

$k = 2$

$$C_1 = \begin{bmatrix} -1 & -\sqrt{3} & -\sqrt{5} \\ \sqrt{3} & -3 & -\sqrt{15} \\ -\sqrt{5} & \sqrt{15} & -5 \end{bmatrix}, C_2 = \begin{bmatrix} 1 & \sqrt{3} & \sqrt{5} \\ -\sqrt{3} & -3 & -\sqrt{15} \\ \sqrt{5} & \sqrt{15} & 5 \end{bmatrix},$$
$$D_1 = \begin{bmatrix} -27 & 21\sqrt{3} & -9\sqrt{5} \\ -27\sqrt{3} & 63 & -9\sqrt{15} \\ -27\sqrt{5} & 21\sqrt{15} & -45 \end{bmatrix}, D_2 = \begin{bmatrix} 99 & -33\sqrt{3} & 87\sqrt{5} \\ 63\sqrt{3} & -15 & 75\sqrt{15} \\ 153\sqrt{5} & -75\sqrt{15} & 525 \end{bmatrix},$$
$$D_3 = \begin{bmatrix} -117 & -33\sqrt{3} & -123\sqrt{5} \\ -9\sqrt{3} & -183 & 9\sqrt{15} \\ -129\sqrt{5} & -141\sqrt{15} & -855 \end{bmatrix}, \text{ and } D_4 = \begin{bmatrix} 45 & 45\sqrt{3} & 45\sqrt{5} \\ -27\sqrt{3} & -81 & -27\sqrt{15} \\ 3\sqrt{5} & 3\sqrt{15} & 15 \end{bmatrix}.$$

$k = 3$

$$C_1 = \begin{bmatrix} -1 & -\sqrt{3} & -\sqrt{5} & -\sqrt{7} \\ \sqrt{3} & -3 & -\sqrt{15} & -\sqrt{21} \\ -\sqrt{5} & \sqrt{15} & -5 & -\sqrt{35} \\ \sqrt{7} & -\sqrt{21} & \sqrt{35} & -7 \end{bmatrix},$$
$$C_2 = \begin{bmatrix} 1 & \sqrt{3} & \sqrt{5} & \sqrt{7} \\ -\sqrt{3} & -3 & -\sqrt{15} & -\sqrt{21} \\ \sqrt{5} & \sqrt{15} & 5 & \sqrt{35} \\ -\sqrt{7} & -\sqrt{21} & -\sqrt{35} & -7 \end{bmatrix},$$
$$D_1 = \begin{bmatrix} 64 & -56\sqrt{3} & 40\sqrt{5} & -16\sqrt{7} \\ 64\sqrt{3} & -168 & 40\sqrt{15} & -16\sqrt{21} \\ 64\sqrt{5} & -56\sqrt{15} & 200 & -16\sqrt{35} \\ 64\sqrt{7} & -56\sqrt{21} & 40\sqrt{35} & -112 \end{bmatrix},$$

$$D_2 = \begin{bmatrix} 8 & -152\sqrt{3} & -4\sqrt{5} & -172\sqrt{7} \\ 168\sqrt{3} & -876 & 96\sqrt{15} & -212\sqrt{21} \\ 104\sqrt{5} & -236\sqrt{15} & 280 & -196\sqrt{35} \\ 328\sqrt{7} & -432\sqrt{21} & 196\sqrt{35} & -1764 \end{bmatrix},$$

$$D_3 = \begin{bmatrix} -208 & -72\sqrt{3} & -172\sqrt{5} & 52\sqrt{7} \\ -128\sqrt{3} & -1044 & -152\sqrt{15} & -348\sqrt{21} \\ -220\sqrt{5} & -120\sqrt{15} & -1100 & -80\sqrt{35} \\ -388\sqrt{7} & -428\sqrt{21} & -472\sqrt{35} & -3696 \end{bmatrix}, \text{ and}$$

$$D_4 = \begin{bmatrix} 136 & 136\sqrt{3} & 136\sqrt{5} & 136\sqrt{7} \\ -104\sqrt{3} & -312 & -104\sqrt{15} & -104\sqrt{21} \\ 52\sqrt{5} & 52\sqrt{15} & 260 & 52\sqrt{35} \\ -4\sqrt{7} & -4\sqrt{21} & -4\sqrt{35} & -28 \end{bmatrix}.$$

$k = 4$

$$C_1 = \begin{bmatrix} -1 & -\sqrt{3} & -\sqrt{5} & -\sqrt{7} & -3 \\ \sqrt{3} & -3 & -\sqrt{15} & -\sqrt{21} & -3\sqrt{3} \\ -\sqrt{5} & \sqrt{15} & -5 & -\sqrt{35} & -3\sqrt{5} \\ \sqrt{7} & -\sqrt{21} & \sqrt{35} & -7 & -3\sqrt{7} \\ -3 & 3\sqrt{3} & -3\sqrt{5} & 3\sqrt{7} & -9 \end{bmatrix},$$

$$C_2 = \begin{bmatrix} 1 & \sqrt{3} & \sqrt{5} & \sqrt{7} & 3 \\ -\sqrt{3} & -3 & -\sqrt{15} & -\sqrt{21} & -3\sqrt{3} \\ \sqrt{5} & \sqrt{15} & 5 & \sqrt{35} & 3\sqrt{5} \\ -\sqrt{7} & -\sqrt{21} & -\sqrt{35} & -7 & -3\sqrt{7} \\ 3 & 3\sqrt{3} & 3\sqrt{5} & 3\sqrt{7} & 9 \end{bmatrix},$$

$$D_1 = \begin{bmatrix} -125 & 115\sqrt{3} & -95\sqrt{5} & 65\sqrt{7} & -75 \\ -125\sqrt{3} & 345 & -95\sqrt{15} & 65\sqrt{21} & -75\sqrt{3} \\ -125\sqrt{5} & 115\sqrt{15} & -475 & 65\sqrt{35} & -75\sqrt{5} \\ -125\sqrt{7} & 115\sqrt{21} & -95\sqrt{35} & 455 & -75\sqrt{7} \\ -375 & 345\sqrt{3} & -285\sqrt{5} & 195\sqrt{7} & -225 \end{bmatrix},$$

$$D_2 = \begin{bmatrix} 575 & -305\sqrt{3} & 563\sqrt{5} & -145\sqrt{7} & 1185 \\ 375\sqrt{3} & -363 & 411\sqrt{15} & -41\sqrt{21} & 1065\sqrt{3} \\ 725\sqrt{5} & -443\sqrt{15} & 3385 & -233\sqrt{35} & 1275\sqrt{5} \\ 625\sqrt{7} & -351\sqrt{21} & 601\sqrt{35} & -1197 & 1215\sqrt{7} \\ 3225 & -2295\sqrt{3} & 2829\sqrt{5} & -1215\sqrt{7} & 4455 \end{bmatrix},$$

$$D_3 = \begin{bmatrix} -775 & -135\sqrt{3} & -793\sqrt{5} & -245\sqrt{7} & -2085 \\ 25\sqrt{3} & -1557 & 55\sqrt{15} & -429\sqrt{21} & -45\sqrt{3} \\ -787\sqrt{5} & -435\sqrt{15} & -4205 & -485\sqrt{35} & -2481\sqrt{5} \\ -415\sqrt{7} & -599\sqrt{21} & -445\sqrt{35} & -4263 & -1125\sqrt{7} \\ -2865 & -2985\sqrt{3} & -3279\sqrt{5} & -3555\sqrt{7} & -11475 \end{bmatrix}, \text{ and}$$

$$D_4 = \begin{bmatrix} 325 & 325\sqrt{3} & 325\sqrt{5} & 325\sqrt{7} & 975 \\ -275\sqrt{3} & -825 & -275\sqrt{15} & -275\sqrt{21} & -825\sqrt{3} \\ 187\sqrt{5} & 187\sqrt{15} & 935 & 187\sqrt{35} & 561\sqrt{5} \\ -85\sqrt{7} & -85\sqrt{21} & -85\sqrt{35} & -595 & -255\sqrt{7} \\ 15 & 15\sqrt{3} & 15\sqrt{5} & 15\sqrt{7} & 45 \end{bmatrix}.$$