FINITE ELEMENT BASED STABILIZED METHODS FOR TIME DEPENDENT CONVECTION–DIFFUSION EQUATION AND THEIR ANALYSIS

A Thesis Submitted to
the Graduate School of Engineering and Sciences of
İzmir Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCE

in Mathematics

by
Kemal Cem YILMAZ

December 2016
İZMİR

We approve the thesis of **Kemal Cem YILMAZ**

Examining Committee Members:

Prof. Dr. Gamze TANOĞLU
Department of Mathematics, İzmir Institute of Technology

Assoc. Prof. Dr. Ali KONURALP
Department of Mathematics, Celal Bayar University

Assist. Prof. Dr. Olha IVANYSHYN YAMAN
Department of Mathematics, İzmir Institute of Technology

26 December 2016

Prof. Dr. Gamze TANOĞLU
Supervisor, Department of Mathematics
İzmir Institute of Technology

Prof. Dr. Engin BÜYÜKAŞIK
Head of the Department of
Mathematics

Prof. Dr. Bilge KARAÇALI
Dean of the Graduate School of
Engineering and Sciences

ACKNOWLEDGMENTS

This work is a result of a one-year intense effort with a fully distressed second half. I would like to express my deepest appreciation to Prof. Dr. Gamze TANOĞLU and Prof. Dr. Engin BÜYÜKAŞIK for their valuable moral support during this unfortunate period.

My deepest thanks and regards goes to the committee members Assoc. Prof. Dr. Ali KONURALP and Assist. Prof. Dr. Olha IVANYSHYN-YAMAN for their valuable time and helpful comments on my possible further study.

I would also like to thank Assist. Prof. Dr. Faruk TEMUR for devoting his valuable time for my several questions.

Still, the first prize goes to my mother. This thesis is dedicated to her.

ABSTRACT

FINITE ELEMENT BASED STABILIZED METHODS FOR TIME DEPENDENT CONVECTION–DIFFUSION EQUATION AND THEIR ANALYSIS

This study is focused on a Fourier stability and accuracy analysis of the time integration algorithms using generalized trapezioidal family of methods of scalar unsteady convection–diffusion equation with periodic boundary conditions. The discretization in space dimension is performed by standard Galerkin finite element formulation for low Peclet numbers and stabilized finite element formulation for large Peclet numbers. The stability analysis is performed namely by von-Neumann stability analysis. Accuracy is measured in terms of damping errors and phase speed errors. The behaviour of these temporal errors of the particular time stepping algorithms, i.e. forward Euler, Crank-Nicolson and backward Euler methods are compared with each other. Particular attention is given to the stabilized finite element formulation, that is the case where we consider high Peclet numbers. For this case, it is concluded that the Crank-Nicolson time stepping represents a better approximate solution compared to the other time integrators on transport process of an initial wave profile. Finally, at the end of the study, we derive a stabilization parameter under a particular condition on Courant number, which provides the relative phase speed error being almost equivalent to its optimal level, that is, the waves with different Fourier modes propagate almost in the same speed. Theoretical results are confirmed by a number of numerical experiments.

ÖZET

ZAMANA BAĞLI KONVEKSİYON–DİFÜZYON DENKLEMİ İÇİN SONLU ELEMANLAR TABANLI KARARLI YÖNTEMLER VE BUNLARIN ANALİZİ

Bu çalışma, durağan olmayan konveksiyon–difüzyon denkleminin uzaysal ayrıklaştırmada sonlu elemanlar, zamansal ayrıklaştırmada θ-yöntemi kullanılarak elde edilen sayısal çözümlerinin Fourier kararlılık ve kesinlik analizi üzerinedir. Peclet sayısının küçük olduğu durumlarda uzaysal ayrıklaştırma için standart Galerkin, büyük olduğu durumlarda ise uzaysal ayrıklaştırma için dengelenmiş sonlu elemanlar yöntemi uygulandı. Kararlılık analizi von-Neumann yöntemi kullanılarak gerçekleştirildi. Kesinlik analizi, zamansal olarak, anlık sönüm hatası ve anlık faz hızı hatası türünden ölçüldü. Bu hatalar, θ-yönteminin özel halleri olan ileri fark, merkezi fark ve geri fark yöntemleri için ayrı ayrı elde edilip, birbirleriyle karşılaştırıldı. Özel olarak, kararlı sonlu elemanlar yönteminin kullanıldığı $\mathrm{Pe} > 1$ durumu için, merkezi fark yönteminin, başlangıç koşulu dalga profili olan taşınım problemlerinde diğer yöntemlerden daha iyi yaklaşık çözüm ürettiği sonucuna varıldı. Son olarak, Courant sayısının özel bir seçimi altında, göreli faz hızı hatasını hemen hemen ideal kılan, diğer bir ifade ile farklı dalga numaralarına sahip dalgaların hemen hemen aynı hız ile yayılımını sağlayan bir parametre elde edildi ve bu parametre seçime yönelik yapılan analizler, sayısal örnekler ile desteklendi.

TABLE OF CONTENTS

LIST OF FIGURES

Figure **Page**

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1. Introduction

The convection–diffusion equation generally models a change in substance in a medium occurs by the convective and diffusive process. There are numerous examples of the applications of this model in science and engineering. To exemplify pollutant dispersal in a river estuary, vorticity transport in the incompressible Navier–Stokes equations, groundwater flow and solute transport, Black-Scholes model for option pricing, modeling of the various stages of tumour development and so on [27]. The common crux of these models is that, they contain a small parameter so that, if we consider the limiting case of this parameter, the behaviour of the solution will totally be different compared to the solutions obtained by finite quantities of the parameter, so called *singularly perturbation problem*, first used by Friedrich and Wasow in their study [15].

1.1.1. A Brief Background of Stabilization Techniques on the Numerical Solution of Elliptic Problems

For elliptic problems, where the coefficient of the second order term is not close to zero, diffusive process plays the leading role compared to the convective process, whereas if the coefficient of second order term is smaller than the coefficient of first order term as we described in the above part, i.e.

$$\frac{|\text{Coefficient of Convection}|}{|\text{Coefficient of Diffusion}|} \gg 1,$$

then the convective process starts capturing the dominance of the problem. Such a differential operator, despite satisfying the ellipticity condition, is living dangerously by flirting with the non-elliptic world, Professor Martin Stynes describes [36]. In fact while con-

vective dominated convection–diffusion problems exhibits a solution with a convective nature, some abrupt changes on the solution and its derivative occur due to the boundary conditions, except some trivial cases. This type of behaviour is expressed as the solution has a *boundary layer*. For the numerical point of view, applying standard Galerkin finite element method (SGFEM) with an inadequate discretization to resolve these regions that includes the layers yields an incapable numerical solution that exhibits large numerical oscillations spreading all the domain. Background of this situation may become more visible and apprehensible by obtaining the discrete solution of the following simple model

$$-\kappa u'' + \beta u' = 0, \quad \text{in } (0,1)$$

where $\kappa > 0$ and without loss of generality $\beta > 0$. Approximating to this model by using SGFEM with a uniform discretization so that each element has a size of Δx with N elements, one can show that such discretization is equivalent to apply finite difference method using a central difference operator as

$$-\kappa \frac{u_{i+1} - 2u_i + u_{i-1}}{\Delta x^2} + \beta \frac{u_{i+1} - u_{i-1}}{2\Delta x} = 0, \quad i = 1, \ldots, N-1.$$

Solution of this linear difference equation is given by

$$u_i = c_1 + c_2 \left(\frac{1 + \mathrm{Pe}}{1 - \mathrm{Pe}} \right)^i, \quad c_1, c_2 \in \mathbb{R}$$

where $\mathrm{Pe} = \frac{\beta \Delta x}{2\kappa}$ is the dimensionless Peclet number. [1]. From the above form, one can immediatelly observe that choice of $\mathrm{Pe} > 1$ causes to appear a negative base and therefore yields a node-to-node oscillatory solution. In other words, to obtain a reasonable approximate solution, the Peclet number must be chosen as $\mathrm{Pe} < 1$ or equivalently the discretization is required to be done such that, the element size Δx must be of the same size as the ratio between the coefficient of diffusion and convection terms. In convective

[1] We note that $\mathrm{Pe} = 1$ is not the case for the solution given above as this choice simplifies the difference equation to $u_i - u_{i-1} = 0$.

dominated problems, i.e. $0 < \kappa \ll 1$, this choice will cause a computational cost, make the method impractical and therefore it becomes essential to derive cheap, practical, yet efficient numerical method.

In order to cure this instability, a number of approaches, commonly known as stabilized finite element methods have been developed which enhance the stability without compromising the consistency. One common way is achieved by adding residual of the differential operator to the basic Galerkin formulation weighted by a parameter, called stabilization parameter. There are several methods based on this idea, starting with Hughes and Brooks [8, 26], called streamline upwind/Petrov–Galerkin (SUPG) method. The idea rises by considering the physical role of the first order term, which is proceeded by adjusting the finite element basis functions by considering the convective velocity field. This modification shows itself as an additional term in the variational form added to the diffusion coefficient, called artificial diffusion. This additional dissipation occured in the formulation is responsible to damp the spurious oscillations caused by the space discretization while consistency is preserved. Many variations of this idea are proposed in the literature also for convection–diffusion–reaction equation and two of them are refered as Galerkin/least squares (GLS) method [24], unusual stabilized finite element method (USFEM) [14]. As we briefly introduce in Appendix A by skipping the technical details, if we exclude the reaction term, these two approximations become equivalent with the SUPG method. On the other hand, another well-known way to obtain a stabilized method is based on enriching the finite element space. An example to this kind of stabilized method is Residual–Free Bubble (RFB) method where the idea is to represent the numerical solution not only by the linear combination of standard finite element basis functions, but also the linear combination of *bubble* type functions, that are zero along the element boundary [5, 7]. Again in the same studies, if we consider the convective–diffusive problem, then bubble type modification of SGFEM can be expressed as in the same structure as the SUPG method, i.e. addition of a stabilization term of streamline diffusion type finite element method. In other words, the underlying principle and the mathematical structure for all these methods are same for convection–diffusion problems. Therefore, from now on, we will address them all by streamline diffusion finite element method (SDFEM).

1.1.2. Extension to Transient Problems

The stabiliation techniques proposed for steady problems could appropriately be applied for parabolic problems by combining them with time stepping methods. This can be defined through a process by seperating spatial and temporal discretizations. Indeed the model problem we will consider has two kind of differentiation: the spatial convection–diffusion operator and the first order time derivative which is the evolution of convection–diffusion process in time. Therefore one shall start discretizing first in space than in time or first in time than in space. The first technique, which we will prefer using in our study, is known as method of lines [31]. The computational part is then simple: once we obtain a semi-discrete form, i.e. our equation is discretized in space but still continuous in time, then the probleem is reduced to a system of coupled ordinary differential equation in time that can be solved by a number of time-marching schemes. In other words, a steady-state problem is required to be solved in each time step. For that reason, as we face while solving the elliptic problems numerically, standard numerical approaches like SGFEM will fail to solve unsteady problems when either the diffusion coefficient becomes small compared to convection coefficient or small time steps are employed in time discretizations [17, 18]. As a concequence, it is indispensable to apply a spatially stabilized numerical method for the space discretization.

One way to overcome this trouble is to extend the spatial stabilization techniques for unsteady models. This can be done by combining spatially stabilized discretization with a time discretization scheme. There are several studies dealed with this combination to obtain effective approximate solutions while providing their stability and convergence. For further details, see [1, 3, 4, 9, 12, 29] and references therein.

1.1.3. Time Accuracy Analysis

In transient problems, which are in our interest in the present study, the basic issue is not merely a question of deriving a stabilized method. The analysis of the numerical methods can be based on a number of metrics such as truncation error, rate of convergence, dispersive and dissipative behaviours and so on. It is demonstrated that Fourier analysis provides an automatic mechanism to quantify the spectral behaviour of the discrete convective–diffusive operator by seperating itself to its symmetric and skew-symmetric components which are responsible for dissipative and dispersive behaviours

respectively. We say that a numerical scheme is dissipative, if it reduces the amplitude more than as it should be, i.e. it erroneously decrease the energy and consequently causes a change on the true solution. The error in damping can be measured by comparing the exact diffusivity and apparent, i.e. numerical diffusivity. The phenomeon of waves of different frequencies traveling with different speed is called dispersion. To be more precise, if the numerical counterparts of the waves whose linear combination forms the exact solution according to the superposition principle propagates with different speeds, then there will be seen a distortion on the shape of the initial condition which is called as dispersion. Contrary if the speed of the waves are all same for each wave number, i.e. the phase speed is independent of wave number, then we say that the solution is non-dispersive. But this is not the case for any discrete scheme except some trivial cases. For this reason, we would like to point out that the non-dispersive behaviour of the continuous problem does not necessary mean any discrete model of it will be non-dispersive too. Therefore it is essential to check the wavenumber dependency of the speed of the waves in order to charactize the error in phase speed. Similarly, this can be done by comparing the phase speed with its numerical counterpart.

There are several sources in the literature related with the temporal errors. In [35, 37] dispersive and dissipative properties of various finite difference approximations to unsteady problems are given. An extensive study on the phase and damping errors of semi-discrete and fully discrete approximations to convection–diffusion equation can be found in [16]. In addition to these analogue studies, stability and accuracy properties for multidimensional and non-linear problems are briefly given in [22]. See [38] for a detailed study on phase and group velocity. Last but not least, a comprensive study of the Fourier analysis of hyperbolic equations is presented in [39].

The use of Fourier analysis to study on the dissipative and dispersive behaviours is not new and has been used by many researchers to make an investigation on the related errors of numerical methods. In [34] Shakib and Hughes presented Fourier analysis of a space-time GLS method for the solution of time dependent convective–diffusive problem, where constant-in-time and linear-in-time approximations are considered for time stepping. Dettmer and Peric in [13] combined the same spatial stabilization technique with several time stepping methods on unsteady convection–diffusion equation in order to make a comparison between them. In the light of the conclusions, they extend the time integration schemes for incompressible Navier–Stokes equations. In [23], Huerta and Donea considered spatially stabilized methods SUPG, GLS, sub-grid scale stabilization (SGS) and least-squares stabilization (LS) for time dependent convection–diffusion–reaction model and presented the time accuracy properties corresponding to each method.

Christon and his co-workers in [10] concerned with the spatial errors associated with standard and stabilized methods of finite difference, finite element and finite volume semi-discretizations of scalar convection–diffusion equation. In addition, they probe the group velocity errors and effect of the artificial diffusivity occured via the stabilized methods. In [40] they extend their study to two dimensional domains by also considering the wave propagation direction. In a series of studies [19, 20, 21] Hauke and Doweidar presented a Fourier analysis for unsteady convection–diffusion–reaction equation using SGFEM, SUPG, SGS and subgrid scale/gradient subgrid scale (SGS/GSGS) methods as well as predictor multi-corrector algorithms with various time integrators such as trapezoidal rule, constand-in-time and linear-in-time approximations.

1.2. Preliminaries and Notation

Let $\Omega = (0,1)$ be the spatial domain and define $\Omega_t := \Omega \times \{t > 0\}$. Our model problem is a linear, constant coefficient, parabolic initial–boundary value problem in one space dimension given as

$$\begin{aligned} u_t + \mathcal{L}u &= f, \quad \text{in } \Omega_t \\ u(x,0) &= u_0(x), \quad \text{in } \Omega \times \{t = 0\} \end{aligned}$$

with periodic boundary conditions

$$u(0,t) = u(1,t), \quad u_x(0,t) = u_x(1,t), \quad t > 0$$

where $\mathcal{L}$ is a linear and elliptic differential operator consists of first and second order differential operators which corresponds to convective and diffusive processes respectively. Coefficient of convective term will be β which is without loss of generality taken as a non-negative real quantitiy and coefficient of diffusive term is κ where $\kappa > 0$ must be held due to the well-posedness of the problem in the Hadamard sense. Particular to the Chapter 5, where we consider the convective dominated case, it is set to be a small number, i.e. $0 < \kappa \ll 1$.

The initial condition u_0, which is assumed to be periodic, belongs to the space of

square integrable functions on Ω, that is

$$L_2(\Omega) := \left\{ v : \Omega \to \mathbb{R} \mid v \text{ is Lebesgue measurable and } \int_\Omega |v|^2 dx \text{ is finite} \right\}.$$

We denote $(\cdot, \cdot)$ as the inner product and $\|\cdot\|_{L_2(\Omega)}$ as the associated norm in $L^2(\Omega)$. Further, we consider the Sobolev space, $\mathcal{H}^1(\Omega)$, of square integrable functions with derivatives up to the order 1, that is,

$$\mathcal{H}^1(\Omega) = \left\{ v \in L_2(\Omega) : \frac{\partial v}{\partial x} \in L^2(\Omega) \right\}$$

and in addition, functions that exhibit periodicity on Ω as

$$\mathcal{H}^1_{\text{per}}(\Omega) := \left\{ v \in \mathcal{H}^1(\Omega) : v(0, \cdot) = v(1, \cdot) \right\}.$$

The norm that is admitted by the space $\mathcal{H}^1(\Omega)$ is denoted by $\|\cdot\|_{\mathcal{H}^1(\Omega)}$ Now consider a functiov $v = v(x, t)$ not as a function of x and t, but as a mapping of $\mathbf{v}$ of t into the space $\mathcal{H}^1_{\text{per}}(\Omega)$. Then the space $L^2(0, T, \mathcal{H}^1_{\text{per}}(\Omega))$ denotes the set of all measurable functions $\mathbf{v} : [0, T] \to \mathcal{H}^1_{\text{per}}(\Omega)$ with

$$\int_0^T \|\mathbf{v}(t)\|^2_{\mathcal{H}^1_{\text{per}}(\Omega)} dt < \infty.$$

By ℓ_2, we denote the space of infinite dimensional sequences on $\mathbb{R}$ that are square summable with norm

$$\|\mathbf{u}\|_{\ell_2} = \left(\sum_{m=-\infty}^{\infty} |u_m|^2 \right)^{1/2}$$

and with the energy norm

$$\|\mathbf{u}\|_{\ell_{2,\Delta x}} = \left(\Delta x \sum_{m=-\infty}^{\infty} |u_m|^2 \right)^{1/2}.$$

The discretization shall be begun by observing that our model problem has two kind of differentiation: a spatial differential operator and evolution of this spatial operator. As we stated previously, we prefer using method of lines, i.e spatial discretization is considered to be first. For this purpose, we consider a uniform partition $\mathcal{T}^h$ of Ω and $\Omega^e \in \mathcal{T}^h$, $e = 1, \ldots, N$ be an arbitrary element out of N elements of this partition. The length of each element is measured by Δx. $\mathcal{V}^h$ denotes a finite dimensional subspace of Ω that consists of space of functions $v \in \mathcal{H}^1(\Omega)$, whose restriction into an element Ω^e belong to the space of polynomials up to the degree 1 and periodic on Ω, that is,

$$\mathcal{V}^h = \left\{ v^h \in \mathcal{H}_0^1(\Omega) : v^h|_{\Omega^e} \in \mathbb{P}_1(\Omega^e), \forall \Omega^e \in \mathcal{T}_h, v^h(x_0, \cdot) = v^h(x_N, \cdot) \right\}.$$

The basis functions of this space are defined as follows: For any node x_i, where $i = 0, \ldots, N$

$$\varphi_i(x) = \begin{cases} 1, & x = x_i \\ 0, & x \neq x_i \\ \text{linear}, & x \in \Omega \end{cases}$$

provided that $\varphi_0(x) = \varphi_N(x)$ as $x_0 = x_N$. See Figure 1.1 for the sketch of basis functions. For the time discretization, generalized trapezoidal family of methods, which is

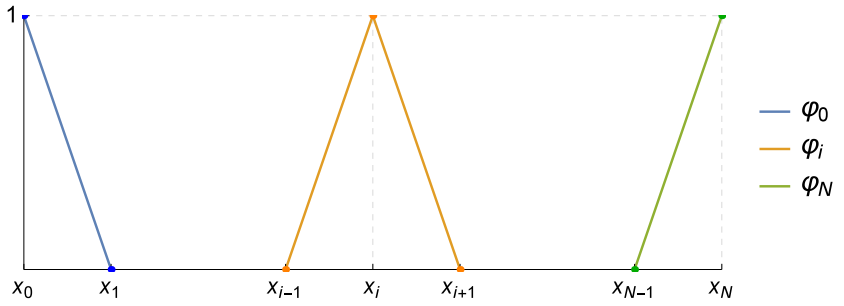


Figure 1.1. Basis functions for the piecewise linear finite element space.

also known as θ-method where $\theta \in [0, 1]$, is going to be applied. In particular, we will be interested in three specific choices of θ values which are $\theta = 0$, $\theta = \frac{1}{2}$ and $\theta = 1$ that yield forward Euler, Crank-Nicolson and backward Euler operators respectively.

The dimensionless numbers that characterize the numerical solution are the Courant number $\mathrm{C} = \frac{\beta \Delta t}{\Delta x}$, the diffusion number $\mu = \frac{\kappa \Delta t}{\Delta x^2}$ and the Peclet number $\mathrm{Pe} = \frac{\beta \Delta x}{2\kappa}$.

Lastly, as we mentioned above, using SGFEM with piecewise linear basis functions to solve an elliptic boundary value problem yields incapable approximate solution if the Peclet number is chosen as $\mathrm{Pe} > 1$. Therefore a stabilization is required which brings out stabilized methods and whenever we mention about *stabilized method*, we are pointing out a spatial stabilization that occurs due to the spatial discretization. On the other hand, we note that subsections related with stability analysis is devoted for *time stability* that occurs due to the time discretization operators.

1.3. Thesis Layout

This study consists of six chapters including introduction and conclusion as the first chapter and the last chapter. The layout of the rest of the thesis is as follows.

In Chapter 2, we start with the statement of our model problem. Then, we begin to discretize our problem by performing SDFEM in space, where note that choosing the stabilization term as zero allows us to recover the standard Galerkin finite element formulation. For the fully discretized form, we proceed with θ-method. To introduce and study the idea that we want to focus on the further parts, we re-express the fully discretized form by its finite difference counterpart.

In Chapter 3, we present some preliminary information related with Fourier analysis. We first introduce two important tools; discrete Fourier transform and Parseval's identity, in order to define the stability in transform space and derive a stability condition based on the numerical amplification, namely von-Neumann stability condition. Then, using the same tools, we define temporal behaviours, i.e. the numerical temporal damping and numerical temporal frequency. Finally, we make a comparison of them with their exact counterparts to define the relative damping error and relative phase speed error.

In Chapter 4, in order to derive the standard Galerkin finite element discretization in space, we set the stabilization parameter to zero on the discrete scheme we derived on Chapter 2. Using the results related with stability and accuracy that we presented on the previous chapter, we study the stability and temporal behaviours of numerical solutions of each time integrator. We compare the errors in damping and in phase speed that we derived for different choice of Courant numbers and for particular time integrators that are forward Euler, Crank-Nicolson and backward Euler methods. Note that due to the spatial instability, we consider only $\mathrm{Pe} \leq 1$ case.

In Chater 5, we consider the same problem except that now the the diffusion term is set to be a small quantity, that is, now convective process plays the major role in our

model problem. Therefore, spatial discretization is now performed by using SDFEM that we briefly introduce in Appendix A. We perform the same work, now for $\mathrm{Pe} > 1$ case, to examine the stability and temporal accuracy properties, i.e. phase and damping responses to observe possible changes and make a comparison between the time integrators. At the end of the chapter, we derive a stabilization parameter under a particular condition on Courant number that brings the phase error nearly to its optimal level. We corroborate this result by giving some numerical experiments and conclude our work.

CHAPTER 2

STABILIZED METHOD FOR UNSTEADY CONVECTION–DIFFUSION EQUATION

In this section we present our model problem, scalar time dependent convection–diffusion equation with periodic boundary conditions. As we have mentioned previously, we start discretizing our model problem by considering the method of lines technique. After deriving the fully discrete form, we convert our discrete problem to a finite difference scheme, to make it ready to perform Fourier analysis.

2.1. Problem Statement

Define $\Omega := (0,1)$ and $\Omega_t := \Omega \times (0,T]$ where T is a finite quantity. Our scalar model problem consists of finding $u(x,t)$ such that

$$\begin{aligned} u_t + \mathcal{L}u &= f, \quad \text{in } \Omega_t && (2.1)\\ u(0,t) &= u(1,t), \quad \text{on } \partial\Omega \times (0,T] && (2.2)\\ u_x(0,t) &= u_x(1,t), \quad \text{on } \partial\Omega \times (0,T] && (2.3)\\ u(x,0) &= u_0(x), \quad \text{in } \Omega \times \{t=0\} && (2.4) \end{aligned}$$

where f is a given source, assumed to be a piecewise constant function, u_0 is an initial condition that belongs to $L_2(\Omega)$ which can be represented as a series

$$u_0(x) = \sum_{j=0}^{\infty} b_j e^{ik_j x}, \quad k_j = 2\pi j, \tag{2.5}$$

and $\mathcal{L}$ is a linear, elliptic differential operator defined as

$$\mathcal{L}u := -\kappa u_{xx} + \beta u_x.$$

Here $\kappa > 0$ represents the diffusivity and without loss of generality β, stands for the scalar velocity field, is chosen to be a non-negative constant.

Now let $v \in \mathcal{H}^1_{per}(\Omega)$ be a test function. Then the variational form of the above model reads: Find $u \in L_2(0,T;\mathcal{H}^1_{\text{per}}(\Omega))$ such that

$$\begin{aligned} (u_t, v) + \mathcal{A}_0(u,v) &= (f,v), \quad \forall v \in \mathcal{H}^1_{\text{per}}(\Omega) && (2.6)\\ u(0) &= u_0(x) && (2.7) \end{aligned}$$

where $\mathcal{A}_0(\cdot,\cdot)$ is a coercive and bounded bilinear operator on $\mathcal{H}^1_{\text{per}}(\Omega)$ defined as

$$\mathcal{A}_0(u,v) := \kappa(u_x, v_x) + (\beta u_x, v). \tag{2.8}$$

2.2. Fully Discretized Model by SDFEM/θ-Method

As we have mentioned in Section 1.1.2, the discretization will be carried out first in space dimension, then in time dimension. The semi-discrete form, i.e. the case where the problem is still continuous in time, is obtained by performing SDFEM. To derive a fully discrete expression, we go on with θ-method for time stepping.

2.2.1. Discretization in Space Dimension Using SDFEM

We start by forming the finite element space. Let us consider a uniform partition $\mathcal{T}^h$ of Ω into N elements, where $\Omega^e \in \mathcal{T}^h$ stands for an arbitrary element, that has a size of $\Delta x = \frac{1}{N}$. Further, let us consider a finite dimensional subspace $\mathcal{V}^h$ that is spanned by

$$\mathcal{V}^h = \text{span}\,\{\varphi_i\}_{i=0}^{N}$$

with $\varphi_0(x) = \varphi_N(x)$. Then, the spatial finite element discretization of the variational form in $\mathcal{V}^h$ using SDFEM consists of finding $u^h(t) : (0, T] \to \mathcal{V}^h$ such that,

$$(\partial_t u^h, v^h) + \mathcal{A}_\tau(u^h, v^h) = (f, v^h), \quad \forall v^h \in \mathcal{V}^h \tag{2.9}$$

together with

$$\mathcal{A}_\tau(u^h, v^h) = \mathcal{A}_0(u^h, v^h) + \sum_{\Omega^e \in \mathcal{T}^h} \left((\partial_t + \mathcal{L})u^h - f, \tau\beta v_x^h\right)_{\Omega^e}$$

where $u^h(0)$ is chosen as a suitable approximation of $u_0(x)$.

Before going further with the time discretization, we first need to represent the semi-discrete formulation (2.9) in matrix form. This requires to express u^h in terms of the global finite element shape functions, $\{\varphi_i(x)\}_{i=0}^N$ of $\mathcal{V}^h$ and is given by

$$u^h(x, t) = \sum_{i=0}^{N} d_i(t)\varphi_i(x). \tag{2.10}$$

Here we find it necessary to mention that time dependency is carried by the nodal values only. Now substituting (2.10) into the semi-discrete form given by (2.9) and considering the problem in element wise, we deduce the following local matrices

$$\begin{aligned}
\mathbf{m}_\tau^e &= \mathbf{m}_0^e + \frac{\tau\beta}{2}\begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix} \\
\mathbf{k}_\tau^e &= \mathbf{k}_0^e + \frac{\tau\beta^2}{\Delta x}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \\
\mathbf{f}_\tau^e &= \mathbf{f}_0^e + \frac{\tau\beta}{\Delta x}\begin{bmatrix} -1 \\ 1 \end{bmatrix}
\end{aligned}$$

where the matrices on the right hand sides are obtained through the standard Galerkin

finite element discretization and are given by

$$\begin{aligned}
\mathbf{m}_0^e &= \frac{1}{6}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \\
\mathbf{k}_0^e &= \frac{\kappa}{\Delta x}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} + \frac{\beta}{2}\begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix} \\
\mathbf{f}_0^e &= \frac{\Delta x}{2}\begin{bmatrix} 1 \\ 1 \end{bmatrix}.
\end{aligned}$$

We leave the details of this process and refer to [25]. Assembling these matrices as carried out again in same reference, we conclude with the following matrix problem: Given $\mathbf{F}_\tau \in \mathbb{R}^{n_{eq}}$, find $\mathbf{d} : (0, T] \to \mathbb{R}^{n_{eq}}$ such that

$$\begin{aligned}
\mathbf{M}_\tau \dot{\mathbf{d}} + \mathbf{K}_\tau \mathbf{d} &= \mathbf{F}_\tau \\
\mathbf{d}(\mathbf{0}) &= \mathbf{d}_0
\end{aligned} \tag{2.11}$$

where $\mathbf{M}_\tau$ and $\mathbf{K}_\tau$ are named as global mass matrix and global stiffness matrix respectively, and $\mathbf{F}_\tau$ is called global load vector.

2.2.2. Discretization in Time Dimension Using θ-Method

Let $0 \leq n \leq T$ be the time index so that $t_n = n\Delta t$ and $\theta \in [0, 1]$ be a parameter. Applying θ-method on (2.11), we derive the fully discretized matrix problem: For a given $\mathbf{d}^n \in \mathbb{R}^{n_{eq}}$, find $\mathbf{d}^{n+1} \in \mathbb{R}^{n_{eq}}$ such that

$$\begin{aligned}
\mathbf{M}_\tau \frac{\mathbf{d}^{n+1} - \mathbf{d}^n}{\Delta t} + \mathbf{K}_\tau \mathbf{d}^{n+\theta} &= \mathbf{F}_\tau^{n+\theta}, \quad n = 0, 1, 2, \ldots \\
\mathbf{d}^{n+\theta} &= \theta \mathbf{d}^{n+1} + (1-\theta)\mathbf{d}^n, \\
\mathbf{F}_\tau^{n+\theta} &= \theta \mathbf{F}_\tau^{n+1} + (1-\theta)\mathbf{F}_\tau^n.
\end{aligned}$$

Above equations can be expressed in one equation as

$$(\mathbf{M}_\tau + \theta\Delta t\mathbf{K}_\tau)\,\mathbf{d}^{n+1} = (\mathbf{M}_\tau - (1-\theta)\Delta t\mathbf{K}_\tau)\,\mathbf{d}^n + \Delta t\left(\theta\mathbf{F}_\tau^{n+1} + (1-\theta)\mathbf{F}_\tau^n\right). \quad (2.12)$$

2.3. Finite Difference Counterpart of Fully Discrete Form

Now we are ready to transform our numerical scheme to its finite difference counterpart, as our interest is to study the stability and accuracy properties related with Fourier techniques. So from now on, we shall ignore the nonhomogeneous part of our model problem, i.e. we take $f \equiv 0$. Define

$$\begin{aligned} \mathbf{A}_\tau &:= \mathbf{M}_\tau + \theta\Delta t\mathbf{K}_\tau \\ \mathbf{B}_\tau &:= \mathbf{M}_\tau - (1-\theta)\Delta t\mathbf{K}_\tau. \end{aligned}$$

Then the homogeneous part of the fully discretized matrix form (2.12) can be expressed as

$$\mathbf{A}_\tau\mathbf{d}^{n+1} = \mathbf{B}_\tau\mathbf{d}^n.$$

Corresponding local matrices of $\mathbf{A}_\tau$ and $\mathbf{B}_\tau$ are given as

$$\mathbf{a}_\tau^e = \mathbf{a}_0^e + \frac{\tau\beta}{2}\begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix} + \theta\Delta t\frac{\tau\beta^2}{\Delta x}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

and

$$\mathbf{b}_\tau^e = \mathbf{b}_0^e + \frac{\tau\beta}{2}\begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix} - (1-\theta)\Delta t\frac{\tau\beta^2}{\Delta x}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

respectively, where

$$\mathbf{a}_0^e = \frac{1}{6}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} + \theta\Delta t\left(\frac{\kappa}{\Delta x}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} + \frac{\beta}{2}\begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix}\right)$$

and

$$\mathbf{b}_0^e = \frac{1}{6}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} - (1-\theta)\Delta t\left(\frac{\kappa}{\Delta x}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} + \frac{\beta}{2}\begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix}\right).$$

Now considering the m-th row of the homogeneous fully discretized matrix equation where $1 < m < n_{eq}$, we therefore deduce the finite difference scheme restricted to the space index m and with time index n as

$$\sum_{j=-1}^{1} A_{\tau,m+j} d_{m+j}^{n+1} = \sum_{j=-1}^{1} B_{\tau,m+j} d_{m+j}^{n}. \tag{2.13}$$

where

$$A_{\tau,m-1} = \left[\frac{\Delta x}{6} + \frac{\tau\beta}{2} + \theta\Delta t\left(-\frac{\kappa+\tau\beta^2}{\Delta x} - \frac{\beta}{2}\right)\right] \tag{2.14a}$$

$$A_{\tau,m} = \left[\frac{2\Delta x}{3} + \theta\Delta t\frac{2(\kappa+\tau\beta^2)}{\Delta x}\right] \tag{2.14b}$$

$$A_{\tau,m+1} = \left[\frac{\Delta x}{6} - \frac{\tau\beta}{2} + \theta\Delta t\left(-\frac{\kappa+\tau\beta^2}{\Delta x} + \frac{\beta}{2}\right)\right] \tag{2.14c}$$

and

$$B_{\tau,m-1} = \left[\frac{\Delta x}{6} + \frac{\tau\beta}{2} - (1-\theta)\Delta t\left(-\frac{\kappa+\tau\beta^2}{\Delta x} - \frac{\beta}{2}\right)\right] \tag{2.15a}$$

$$B_{\tau,m} = \left[\frac{2\Delta x}{3} - (1-\theta)\Delta t\frac{2(\kappa+\tau\beta^2)}{\Delta x}\right] \tag{2.15b}$$

$$B_{\tau,m+1} = \left[\frac{\Delta x}{6} - \frac{\tau\beta}{2} - (1-\theta)\Delta t\left(-\frac{\kappa+\tau\beta^2}{\Delta x} + \frac{\beta}{2}\right)\right]. \tag{2.15c}$$

Lastly, as might be expected, we note to echo that setting $\tau = 0$ on the local matrices or equivalently on the finite difference scheme will allow us to recover the standard Galerkin finite element formulation.

CHAPTER 3

PRELIMINARIES ON FOURIER ANALYSIS

Applying Fourier transform to a parabolic initial value problem results in a first order ordinary differential equation in the space of transformed functions. Also it is well-known that, the temporal behaviour of the solutions derived by the original problem and derived by the transformed problem are equivalent by dint of the equivalence of the norms of the functions in their respective spaces in L_2 sense. Consequently, one may switch to the space of transformed functions, study at that space and moreover, do not require to apply a back transformation. To be more precise, let us consider the following initial value problem

$$\begin{aligned} u_t - \kappa u_{xx} + \beta u_x &= 0, \quad \text{in } \mathbb{R} \times \{t > 0\} && (3.1) \\ u(x,0) &= u_0(x), \quad \text{on } \mathbb{R} \times \{t = 0\} && (3.2) \end{aligned}$$

where all of the parameters are set to be same as in our model problem (2.1)-(2.4) and let us start by giving the following definition.

Definition 3.1 *The Fourier transform of an integrable function u is defined by*

$$\hat{u}(k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u(x) e^{-ikx} dx, \quad k \in \mathbb{R}.$$

Assume that u and u_x are sufficiently nice and vanishing as $x \to \mp\infty$. Applying Fourier transform to the problem (3.1)-(3.2), one can derive

$$\begin{aligned} \hat{u}_t + (\kappa k^2 + i\beta k)\hat{u} &= 0, \quad t > 0 \\ \hat{u}(k,0) &= \hat{u_0}(k) \end{aligned}$$

which has a solution of

$$\hat{u}(k,t) = \hat{u_0}(k)e^{\nu t}, \quad \nu = -\kappa k^2 - i\beta k. \tag{3.3}$$

Here, ν stands for the temporal behaviour of the solution that shows, a Fourier mode with a wavenumber k is decaying monotonically an moving in the flow direction in time.

In order to study on the temporal behaviours of the approximate solution, an analogous work may be carried out for the discrete case. For this purpose, let us begin by giving the definition of the discrete analogue of Fourier transform that is the first fundamental tool for our study.

Definition 3.2 *The discrete Fourier transform of* $\mathbf{u}^h \in \ell_2$ *is the function* $\hat{u}^h \in L_2\left(\left[-\frac{\pi}{\Delta x}, \frac{\pi}{\Delta x}\right]\right)$ *and defined by*

$$\hat{u}^h(k) = \frac{\Delta x}{\sqrt{2\pi}} \sum_{m=-\infty}^{\infty} u_m e^{-imk\Delta x}, \quad k \in \left[-\frac{\pi}{\Delta x}, \frac{\pi}{\Delta x}\right].$$

Definition 3.3 *Let* $\mathbf{u}^h \in \ell_2$ *be a discrete function defined for all integers* m *and* $\hat{u}^h(k) \in L_2\left(\left[-\frac{\pi}{\Delta x}, \frac{\pi}{\Delta x}\right]\right)$ *be the discrete Fourier transform of* $\mathbf{u}^h$*. Then, the inverse discrete Fourier transform of* $\hat{u}^h(k)$ *is given by*

$$u_m = \frac{\Delta x}{\sqrt{2\pi}} \int_{-\frac{\pi}{\Delta x}}^{\frac{\pi}{\Delta x}} \hat{u}(k) e^{imk\Delta x} dk.$$

The second fundamental tool for our analysis is the equivalence of the norms of $\mathbf{u}^h$ and its discrete Fourier transform $\hat{u}^h$ in their respective spaces. The identity, namely Parseval's identity, is given by the following proposition.

Proposition 3.1 (Parseval's Identity) *If* $\mathbf{u} \in \ell_2$ *and* $\hat{u}$ *is the discrete Fourier transform of* $\mathbf{u}$*, then*

$$\|\hat{u}\|_{L_2} = \|\mathbf{u}\|_{\ell_2}$$

where the first norm is defined on $\left[-\frac{\pi}{\Delta x}, \frac{\pi}{\Delta x}\right]$.

Proof Starting with L_2 norm of $\hat{u}^h$ on $\left[-\frac{\pi}{\Delta x}, \frac{\pi}{\Delta x}\right]$ and seperating the modulus to its complex conjugates, we have

$$\begin{aligned}
\|\hat{u}\|_{L_2}^2 &= \int_{-\frac{\pi}{\Delta x}}^{\frac{\pi}{\Delta x}} |\hat{u}(k)|^2 dk \\
&= \int_{-\frac{\pi}{\Delta x}}^{\frac{\pi}{\Delta x}} \overline{\hat{u}(k)}\hat{u}(k) \\
&= \int_{-\frac{\pi}{\Delta x}}^{\frac{\pi}{\Delta x}} \overline{\hat{u}(k)} \left(\frac{\Delta x}{\sqrt{2\pi}} \sum_{m=-\infty}^{\infty} u_m e^{-imk\Delta x} \right) dk \\
&= \sum_{m=-\infty}^{\infty} u_m \frac{\Delta x}{\sqrt{2\pi}} \int_{-\frac{\pi}{\Delta x}}^{\frac{\pi}{\Delta x}} \overline{\hat{u}(k)} e^{-imk\Delta x} dk \\
&= \sum_{m=-\infty}^{\infty} u_m \overline{\frac{\Delta x}{\sqrt{2\pi}} \int_{-\frac{\pi}{\Delta x}}^{\frac{\pi}{\Delta x}} \hat{u}(k) e^{imk\Delta x} dk} \\
&= \sum_{m=-\infty}^{\infty} u_m \overline{u_m} \\
&= \|\mathbf{u}\|_{\ell_2}^2
\end{aligned}$$

where we applied discrete Fourier transform and inverse discrete Fourier transform on the third and fifth lines respectively. □

The reason why the Parseval's identity is fundamental for our analysis is because, the approximate solution and its discrete Fourier transform are reciprocally uniformly bounded in the sense of the associated norms. This situation allows us to study the stability properties in the space of transform functions instead of in our original finite dimensional domain.

3.1. On the Stability Analysis

Definition 3.4 *A two level difference scheme, $\mathbf{A}_\tau \mathbf{d}^{n+1} = \mathbf{B}_\tau \mathbf{d}^n$, is said to be stable with respect to a norm $\|\cdot\|$, if there exist non-negative constants K and α so that*

$$\|\mathbf{d}^n\| \leq K e^{\alpha t} \|\mathbf{d}^0\|, \quad t = n\Delta t, \quad n = 0, 1, \ldots,$$

holds true for sufficiently small Δx and Δt.

Combining the Parseval's identity with the Definition 3.4, we give the following proposition as a summary.

Proposition 3.2 *The sequence* $\{\mathbf{d}^n\}$ *is stable in* $\ell_{2,\Delta x}$ *if and only if the sequence* $\{\hat{d}^n\}$ *is stable in* $L_2\left(\left[-\frac{\pi}{\Delta x}, \frac{\pi}{\Delta x}\right]\right)$.

Proof Using the Parseval's identity and definition of $\ell_{2,\Delta x}$ norm, we have

$$\|\mathbf{d}^n\|_{\ell_{2,\Delta x}} = \Delta x \|\mathbf{d}^n\|_{\ell_2} = \Delta x \|\hat{d}^n\|_{L_2}.$$

On the other hand, assuming that K and α are non-negative constants and considering the above identity, we derive

$$\|\mathbf{d}^n\|_{\ell_{2,\Delta x}} \leq K e^{\alpha t_n} \|\mathbf{d}^0\|_{\ell_{2,\Delta x}} \quad \Leftrightarrow \quad \|\hat{d}^n\|_{L_2} \leq \Delta x K e^{\alpha t_n} \|\hat{d}^0\|_{L_2},$$

for $n = 0, 1, \ldots$ and $t_n = n\Delta t$. Thus, we provide the equivalence of the stability in our original domain and in the transform domain in the sense of associated norms. □

An important consequence of this proposition is given by the following remark.

Remark 3.1 *As a result of Proposition 3.2, once we guarantee the stability in transform space, we do not require to apply an inverse transformation to return back to our original equation.*

Before proceeding with transforming our finite difference scheme (2.13), we introduce two useful properties related with discrete Fourier transform, which makes our work easier. Let $\mathcal{F}^h : \ell_2 \to L_2\left(\left[-\frac{\pi}{\Delta x}, \frac{\pi}{\Delta x}\right]\right)$ be the discrete Fourier transform operator and $\mathbf{u}, \mathbf{v} \in \ell_2$.

i. *Linearity.* For arbitrary scalars α and β,

$$\mathcal{F}^h(\alpha\mathbf{u} + \beta\mathbf{v}) = \alpha\mathcal{F}^h(\mathbf{u}) + \beta\mathcal{F}^h(\mathbf{v}).$$

ii. *Shifting Property.* Define the shifting operator $\mathcal{S}_{\mp}\mathbf{u} = \{v_m\}$, where $v_m = u_{m\mp 1}$, $m = 0, \mp 1, \mp 2, \ldots$. Then,

$$\mathcal{F}^h\left(\mathcal{S}_{\mp}\mathbf{u}\right) = e^{\mp ik}\mathcal{F}^h(\mathbf{u}).$$

Lastly, we find it necessary to state that, eventhough we consider the initial value problem (3.1)-(3.2) for a moment, our finite difference scheme (2.13), which is derived by considering initial–boundary value problem (2.1)-(2.4), still remains same except that the space index m now takes all integers. After we derive the necessary and sufficient stability condition for the initial value problem, then we will take into account the boundary conditions as well, in order to derive a stability condition for initial–boundary value problem.

Now we are ready to transform our original discrete equation. So going on from (2.13) and applying discrete Fourier transform yields

$$\mathcal{F}^h\left(\sum_{j=-1}^{1} A_{\tau,m+j} d_{m+j}^{n+1}\right) = \mathcal{F}^h\left(\sum_{j=-1}^{1} B_{\tau,m+j} d_{m+j}^{n}\right).$$

It follows from using linearty and shifting properties of $\mathcal{F}^h$, we derive

$$\mathcal{F}^h(\mathbf{d}^{n+1}) \sum_{j=-1}^{1} A_{\tau,m+j} e^{jik\Delta x} = \mathcal{F}^h(\mathbf{d}^{n}) \sum_{j=-1}^{1} B_{\tau,m+j} e^{jik\Delta x}$$

or equivalently

$$\hat{d}^{n+1} = \left(\frac{B_{\tau,m-1}e^{-ik\Delta x} + B_{\tau,m} + B_{\tau,m+1}e^{ik\Delta x}}{A_{\tau,m-1}e^{-ik\Delta x} + A_{\tau,m} + A_{\tau,m+1}e^{ik\Delta x}}\right)\hat{d}^{n}.$$

Define

$$g(k) := \frac{B_{\tau,m-1}e^{-ik\Delta x} + B_{\tau,m} + B_{\tau,m+1}e^{ik\Delta x}}{A_{\tau,m-1}e^{-ik\Delta x} + A_{\tau,m} + A_{\tau,m+1}e^{ik\Delta x}}.$$

that is a continuous function of k. Then our discrete problem in the transform space reads: For a given $\hat{d}^n$, find $\hat{d}^{n+1}$ such that

$$\hat{d}^{n+1} = g\hat{d}^{n}, \quad n = 0, 1, \ldots .$$

Note that just how the continuous problem (3.1)-(3.2) is transformed to an ordinary differential equation with time derivative, corresponding discrete equation (2.13) is transformed to a discrete problem in time index. Solution of the above requrrence relation at the n-th time level is given by

$$\hat{d}^n = g^n \hat{d}^0 \tag{3.4}$$

where the superscript on g is a multiplicative exponent. Observe that the discrete solution is represented by in terms of the transform of the initial data and $g = g(k, \Delta x, \Delta t)$. Comparing the continous solution (3.3) and the discrete solution (3.4), we deduce that the quantity g^n is the discrete analogue of $e^{\nu t}$ so that, it gives us an information about the amplification of the approximate solution. Now we are ready to give the following definition.

Definition 3.5 *Let $\nu^h \in \mathbb{C}$ be the discrete counterpart of ν. Then,*

$$g(k, \Delta x, \Delta t) := e^{\nu^h \Delta t} = \frac{G_{B_\tau}}{G_{A_\tau}} \tag{3.5}$$

is called as the amplification factor of the finite difference equation (2.13)*, where*

$$G_{A_\tau} = A_{\tau,m-1} e^{-ik\Delta x} + A_{\tau,m} + A_{\tau,m+1} e^{ik\Delta x} \tag{3.6}$$

and

$$G_{B_\tau} = B_{\tau,m-1} e^{-ik\Delta x} + B_{\tau,m} + B_{\tau,m+1} e^{ik\Delta x}. \tag{3.7}$$

As the name describes its own role, the amplification factor presents an information about the growth of the approximate solution in transform space. Pursuing the idea that we stated on Proposition 3.2, we give the following proposition.

Proposition 3.3 *A two level difference scheme, $\mathbf{A}_\tau \mathbf{d}^{n+1} = \mathbf{B}_\tau \mathbf{d}^n$, is said to be stable with respect to $\ell_{2,\Delta x}$ norm if and only if there exist non-negative constants K and α so*

that

$$|g(k)|^n \leq Ke^{\alpha n\Delta t}, \quad \forall k \in \left[-\frac{\pi}{\Delta x}, \frac{\pi}{\Delta x}\right] \tag{3.8}$$

holds true for sufficiently small Δx and Δt.

Proof

$\Rightarrow$ Arguing with contradiction, assume that for each K and α, there exists an interval $I_{K,\alpha} \subset \left[-\frac{\pi}{\Delta x}, \frac{\pi}{\Delta x}\right]$ such that

$$|g(k)|^n > Ke^{\alpha n\Delta t}, \quad \forall k \in I_{K,\alpha.} \tag{3.9}$$

Indeed such interval exists as g is a continuous function of k. Now choose a function $\hat{u}_0$ which is equivalently zero outside of the interval $I_{K,\alpha}$. Multiplying this with the above inequality, taking square of both sides and then integrating over $\left[-\frac{\pi}{\Delta x}, \frac{\pi}{\Delta x}\right]$ yields

$$\int_{-\frac{\pi}{\Delta x}}^{\frac{\pi}{\Delta x}} |g^n \hat{u}_0|^2 dk > \left(Ke^{\alpha n\Delta t}\right)^2 \int_{-\frac{\pi}{\Delta x}}^{\frac{\pi}{\Delta x}} |\hat{u}_0|^2 dk$$

or

$$\|g^n \hat{u}_0\|_{L_2}^2 > \left(Ke^{\alpha n\Delta t}\right)^2 \|\hat{u}_0\|_{L_2}^2. \tag{3.10}$$

As (3.10) holds true for all non-negative K and α, this contradicts with Proposition 3.2. Therefore. our assumption (3.6) can not be valid.

$\Leftarrow$ Multiplying (3.8) with $\hat{u}_0$, taking square of both sides and integrating over $\left[-\frac{\pi}{\Delta x}, \frac{\pi}{\Delta x}\right]$, we derive

$$\|g^n \hat{u}_0\|_{L_2}^2 \leq \left(Ke^{\alpha n\Delta t}\right)^2 \|\hat{u}_0\|_{L_2}^2. \tag{3.11}$$

Thus, we gain the stability at the transform space which, by Proposition 3.2, pro-

vides the stability at our original domain as well.

□

Theorem 3.1 (von Neumann) *A two level difference scheme, $\mathbf{A}_\tau \mathbf{d}^{n+1} = \mathbf{B}_\tau \mathbf{d}^n$, is said to be stable with respect to $\ell_{2,\Delta x}$ norm if and only if there exists a non-negative K that is independent of k, Δx and Δt so that*

$$|g(k)| \leq 1 + K\Delta t, \quad \forall k \in \left[-\frac{\pi}{\Delta x}, \frac{\pi}{\Delta x}\right] \tag{3.12}$$

holds true for sufficiently small Δx and Δt.

Proof

$\Rightarrow$ Let us assume the opposite, that is, for all Δx, Δt and K, there exists a $k_C \in \left[-\frac{\pi}{\Delta x}, \frac{\pi}{\Delta x}\right]$ such that

$$|g(k_C)|^n > 1 + K\Delta t. \tag{3.13}$$

Consider a sequence $\{K_i\}$, such that $K_i \to \infty$ as $i \to \infty$. Then we get a sequence $\{k_i\}$ for each $\{K_i\}$ and according to our assumption (3.13), we have

$$|g(k_i)|^n \to \infty, \quad \text{as } k \to \infty,$$

which means $|g(k)|^n$ is not bounded and this contradicts with Proposition (3.3).

$\Leftarrow$ Following from $|g(k)| \leq 1 + K\Delta t \leq e^{K\Delta t}$ or $|g(k)|^n \leq e^{Kn\Delta t}$ and then Proposition (3.3), we end up with gaining the stability with respect to $\ell_{2,\Delta x}$ norm.

□

Recall that the problem we consider is still an initial value problem and the difference between the finite difference schemes approximated to an initial value problem and an initial–boundary value problem is, the latter case, in addition to the partial differential equation, consists of a suitable approximation through the boundaries. We may interpret the relation of the stability between initial value problems and initial–boundary value problems as, if a finite difference scheme approximated to an initial value problem is stable, then that finite difference scheme will also be stable in particular for some

squence of nodes that belong to an open interval, i.e. finite difference scheme approximated to an initial–boundary value problem remains stable only for the interior points. On the other hand, the discrete Fourier transform does not take into account the boundary points. Thereefore if the boundary conditions cause an instability then the scheme will be unstable as well. As a conclusion, a two level difference scheme, $\mathbf{A}_\tau \mathbf{d}^{n+1} = \mathbf{B}_\tau \mathbf{d}^n$, approximated for an initial–boundary value problem is necessarily stable, if it is stable for an initial value problem. Hence, we give the following result.

Remark 3.2 *The von Neumann stability analysis for a two level finite difference scheme,* $\mathbf{A}_\tau \mathbf{d}^{n+1} = \mathbf{B}_\tau \mathbf{d}^n$*, considered as an initial value problem yields a necessary stability condition for the difference scheme considered as an initial–boundary value problem.*

To recover the sufficiency on the stability condition, we require to ensure that the boundary conditions do not cause an instability. Recall that our model problem (2.1)-(2.4) consists of a periodic boundary conditions with an appropriate initial condition (2.5) and so the solution of the problem is therefore a periodic soution with a period of Ω. Thus, if our approximate solution is stable on the whole real line, then it will be stable in particular in one period. In other words, the von Neumann condition provides also the sufficiency on the stability of the approximate solution for our model problem by dint of the initial and boundary conditions.

Lastly we note that, the second term of the von Neumann estimate (3.12) is to control the exponential growth but the analytical solution of our problem does not have a such behavior with increasing time. Therefore the stability condition shall be replaced by

$$|g(k)| \leq 1 \tag{3.14}$$

to provide the sufficiency [33].

3.2. On the Accuracy Analysis

It is well-known that according to the superposition principle, the analytical solution of our model problem is infinite combination of trigonometric functions. By decomposing this infinite sum in orthogonal Fourier modes, exact solution may also be assumed

to be one of these combinations, where each of them has the form of

$$u(x,t) = e^{\nu t + ikx}. \tag{3.15}$$

Indeed, the periodic boundary conditions (2.2)-(2.3) and the initial condition that has the form (2.5) enable us to express the solution as (3.15). Here $k \in \mathbb{R}$ is the wavenumber, i is the imaginary unit and $\nu \in \mathbb{C}$ determines the temporal behaviour that can be written in the form $\nu := -\xi + i\omega$. Definition of the real and imaginary parts are as following.

Definition 3.6 *In the previous expression, ξ and ω are called as exact temporal damping and exact temporal frequency respectively.*

Employing (3.15) into our target equation (2.1) yields

$$\nu = -\kappa k^2 - i\beta k$$

and so exact temporal damping and exact temporal frequency are given by

$$\xi = \kappa k^2 \tag{3.16}$$

$$\omega = -\beta k \tag{3.17}$$

respectively.

Now we repeat the analogous process for the discrete case. For this purpose, consider the discrete counterpart of (3.15) as a solution of our finite difference equation (2.13) at n-th time level and m-th space level

$$u_m^n = e^{\nu^h t_n + ikx_m} = \left(e^{\nu^h \Delta t}\right)^n e^{ikm\Delta x}. \tag{3.18}$$

Recall that the first factor on the last expression is nothing but the amplification factor. Now define $\nu^h := -\xi^h + i\omega^h$ as in terms of numerical counterparts of exact temporal damping and exact temporal frequency. We first give the following definition.

Definition 3.7 *ξ^h and ω^h are called as numerical temporal damping and numerical temporal frequency respectively.*

Following from (3.18) and leaving ν^h alone, we therefore have

$$\nu^h = \frac{1}{\Delta t}\left[\ln|g| + i\arg(g)\right].$$

Considering the real and imaginary parts, numerical temporal damping and numerical temporal frequency are given by

$$\xi^h = -\frac{1}{\Delta t}\ln|g| \tag{3.19}$$

and

$$\omega^h = \frac{1}{\Delta t}\arg(g) \tag{3.20}$$

respectively. Now a comparison between between the exact temporal damping and the exact temporal frequency with their numerical counterparts will allow us to investigate how good or how bad the character of damping and frequency of our numerical scheme is. In consideration of (3.16) and (3.19), we give the following definition.

Definition 3.8 *The error in decay, called relative damping error, is defined by the ratio of computed damping and the exact damping which is given by*

$$\varepsilon_D := \frac{\xi^h}{\xi} = -\frac{\ln|g|}{\kappa k^2 \Delta t}. \tag{3.21}$$

Remark 3.3 *It could be interpreted by looking at* (3.21) *that* $0 < \varepsilon_D < 1$ *leads an underdamped numerical solution and* $\varepsilon_D > 1$ *leads an overdamped one. The aim is* $\varepsilon_D = 1$.

Remark 3.4 *As* ξ *is a positive quantity,* $\varepsilon_D < 0$ *implies* $\xi^h < 0$ *that is the numerical method behaves like anti-diffusive. Thus, in the case of* $\varepsilon_D < 0$*, one can conclude that the numerical solution is unstable.*

Next, comparing (3.17) with (3.20), we give the following definition.

Definition 3.9 *The error in speed, called relative phase speed error, is defined by the ratio of computed frequency and the exact frequency that is given by*

$$\varepsilon_S := \frac{\omega^h}{\omega} = -\frac{1}{k\beta\Delta t}\arg(g). \tag{3.22}$$

Remark 3.5 *One can observe from* (3.22) *that* $\varepsilon_S < 1$ *causes a lagging error and* $\varepsilon_S > 1$ *causes leading error. The aim is ofcourse* $\varepsilon_S = 1$.

CHAPTER 4

STABILITY AND ACCURACY ANALYSIS OF SGFEM/θ-METHOD

This chapter is devoted to the stability and accuracy properties of the numerical solution by using the conclusions we derived on Chapter 3. The numerical solution is obtained by using SGFEM in space, i.e. the stabilization parameter is set to be zero, and θ-method in time. Due to the spatial instability troubles of standard Galerkin finite element formulation as we emphasised on Chapter 1, the study we perform here is for the cases $\mathrm{Pe} \leq 1$. In particular, in order to observe the effect of increment of the Peclet number on the relative damping error, we also consider the $\mathrm{Pe} = 0$, i.e. $\beta = 0$, case which corresponds to the pure heat equation. For the time discretization, we concern with there particular time marching methods that belongs to the family of θ-methods obtained by choosing $\theta = 0$, $\theta = \frac{1}{2}$ and $\theta = 1$ which are namely forward Euler method, Crank-Nicolson method and backward Euler method respectively.

4.1. Stability Analysis

Recall from the Definition 3.5 that, the amplification factor for our approximate solution is given by (3.5). So proceeding with definition, together with (3.6)-(3.7) where the corresponding elements of the global matrices are given by the equations (2.14) and (2.15), one can derive

$$\begin{aligned} G_{A_0} &= (A_{0,m-1}e^{-i\phi} + A_{0,m} + A_{0,m+1}e^{i\phi}) \\ &= \frac{\Delta x}{6}\left(e^{-i\phi} + 4 + e^{i\phi}\right) + \theta\Delta t\left[\frac{-\kappa}{\Delta x}\left(e^{-i\phi} - 2 + e^{i\phi}\right) - \frac{\beta}{2}\left(-e^{-i\phi} + e^{i\phi}\right)\right] \\ &= \frac{\Delta x}{3}(2 + \cos\phi) + \theta\Delta t\left[\frac{2\kappa}{\Delta x}(1 - \cos\phi) + i\beta\sin\phi\right] \end{aligned}$$

and

$$\begin{aligned} G_{B_0} &= (B_{0,m-1}e^{-i\phi} + B_{0,m} + B_{0,m+1}e^{i\phi}) \\ &= \frac{\Delta x}{6}\left(e^{-i\phi} + 4 + e^{i\phi}\right) - (1-\theta)\Delta t\left[\frac{-\kappa}{\Delta x}\left(e^{-i\phi} - 2 + e^{i\phi}\right) - \frac{\beta}{2}\left(-e^{-i\phi} + e^{i\phi}\right)\right] \\ &= \frac{\Delta x}{3}(2+\cos\phi) - (1-\theta)\Delta t\left[\frac{2\kappa}{\Delta x}(1-\cos\phi) + i\beta\sin\phi\right]. \end{aligned}$$

Here ϕ is called as phase angle and is given by $\phi = k\Delta x$. Hence, the amplification factor $g = \frac{G_{B_0}}{G_{A_0}}$ is expressed as

$$g = \frac{1-(1-\theta)\Delta t\lambda}{1+\theta\Delta t\lambda} \tag{4.1}$$

together with

$$\lambda = \frac{\frac{2\kappa}{\Delta x}(1-\cos\phi) + i\beta\sin\phi}{\frac{\Delta x}{3}(2+\cos\phi)}. \tag{4.2}$$

To make a clearer observation through the amplification factor on our further study, let us state $\Delta t\lambda$, so the amplification factor, in terms of Peclet number and Courant number which will lead a more useful representation

$$\Delta t\lambda = 3\mathrm{C}\frac{\frac{1}{\mathrm{Pe}}(1-\cos\phi) + i\sin\phi}{\cos\phi + 2}. \tag{4.3}$$

Now the amplification factor (4.1) together with (4.3) is expressed in the desired form. Thus, we can continue to our work with the particular cases of the θ-method.

Forward Euler Case Employing $\theta = 0$ on the amplification factor (4.1) with (4.3) yields,

$$g = 1 - 3\mathrm{C}\frac{\frac{1}{\mathrm{Pe}}(1-\cos\phi) + i\sin\phi}{2+\cos\phi}.$$

The magnitude of this expression is

$$|g|^2 = \left(1 - \frac{3\mathrm{C}(1-\cos\phi)}{\mathrm{Pe}(2+\cos\phi)}\right)^2 + \left(\frac{3\mathrm{C}\sin\phi}{2+\cos\phi}\right)^2.$$

In order for requiring the von Neumann stability condition, we let $|g|^2 \leq 1$ and this leads

$$-6\mathrm{CPe}(1-\cos\phi)(2+\cos\phi) + 9\mathrm{C}^2(1-\cos\phi)^2 + 9\mathrm{C}^2\mathrm{Pe}^2(1-\cos^2\phi) \leq 0.$$

For $\phi = 0$, observe that $|g| = 1$. So considering $\phi \neq 0$ and leaving Courant number on the left hand side, we therefore have

$$\mathrm{C} \leq \frac{2\mathrm{Pe}}{3}\frac{2+\cos\phi}{(1-\cos\phi)+\mathrm{Pe}(1+\cos\phi)} := f(\mathrm{Pe},\phi).$$

To obtain a stability estimate, it follows from minimizing $f(\mathrm{Pe},\phi)$ with respect to ϕ that

$$\begin{aligned}\frac{\partial f}{\partial \phi} &= -\frac{2\mathrm{Pe}}{3}\frac{\sin\phi\left[1-\cos\phi+\mathrm{Pe}^2(1+\cos\phi)+(2+\cos\phi)(1-\mathrm{Pe}^2)\right]}{(1-\cos\phi+\mathrm{Pe}^2(1+\cos\phi))^2} \\ &= -\frac{2\mathrm{Pe}}{3}\frac{\sin\phi(3-\mathrm{Pe}^2)}{(1-\cos\phi+\mathrm{Pe}^2(1+\cos\phi))^2}.\end{aligned}$$

Now observe that $f(\mathrm{Pe},\phi)$ is constant along $\mathrm{Pe} = \sqrt{3}$. So this choice is the critical level to minimize $f(\mathrm{Pe},\phi)$.

Case 1. For any fix $\mathrm{Pe} < \sqrt{3}$, f is decreasing and so it attains the minimum at

$\phi = \pi$. Consequently

$$\mathrm{C} \leq f(\mathrm{Pe}, \pi) = \frac{\mathrm{Pe}}{3} \tag{4.4}$$

or

$$\Delta t \leq \frac{\Delta x^2}{6\kappa} \tag{4.5}$$

which is the stability result for this case.

Case 2. For any fix $\mathrm{Pe} > \sqrt{3}$, f is increasing. So to minimize $f(\mathrm{Pe}, \phi)$, we let $\phi \to 0$ and conclude that

$$\mathrm{C} \leq f(\mathrm{Pe}, \phi) = \frac{1}{\mathrm{Pe}} \tag{4.6}$$

or equivalently

$$\Delta t \leq \frac{2\kappa}{\beta^2}. \tag{4.7}$$

Note that for $\mathrm{Pe} = \sqrt{3}$, the estimates (4.5) and (4.7) are identical.

Finally, stability region for both cases with respect to the Courant number and Peclet number are given in the Figure 4.1. Blue and red regions stand for the conditions (4.4) and (4.6) respectively.

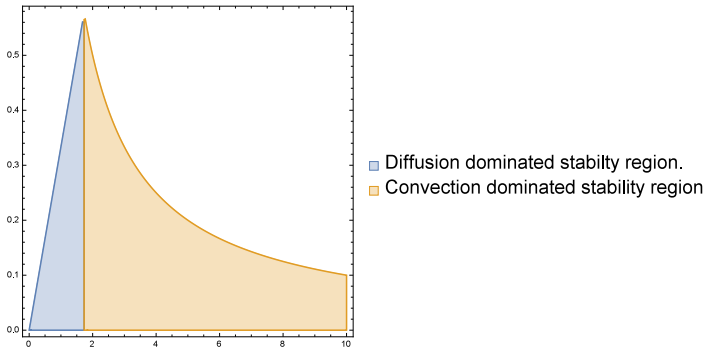
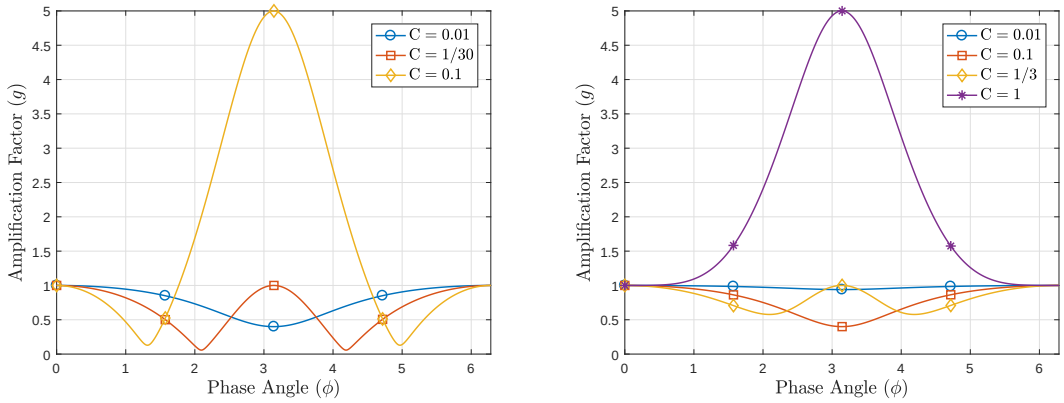


Figure 4.1. Stability region using SGFEM in space and forward Euler method in time.

In Figure 4.2, we give the plots of the amplification factors for various values of

Courant numbers for $\mathrm{Pe} = 0.1$ and $\mathrm{Pe} = 1$ cases. For the choices lesser than $\mathrm{C} = \frac{1}{30}$ and $\mathrm{C} = \frac{1}{3}$ which are the stability limits for $\mathrm{Pe} = 0.1$ case and $\mathrm{Pe} = 1$ case respectively, observe that amplification factors vary below than $g = 1$ level. On the other hand, choosing $\mathrm{C} = 0.1$ for $\mathrm{Pe} = 0.1$ case and $\mathrm{C} = 1$ for $\mathrm{Pe} = 1$ case that are greater than the corresponding stability limits, give rise to exceed the $|g| = 1$ level which ends up with an instability.



(a) $\mathrm{Pe} = 0.1$. (b) $\mathrm{Pe} = 1$.

Figure 4.2. Amplification factor using SGFEM in space and forward Euler method in time for various choices of Courant numbers.

We give the following remarks.

Remark 4.1 *Observe that the stability estimate* (4.5) *does not depend on the coefficient of convection term. In other words, considering the case* $\beta = 0$ *which corresponds to the heat equation, one shall still derive the same stability estimate, that is,* $\mu \leq \frac{1}{6}$ *where* μ *is the diffusion number. Therefore, up to* $\mathrm{Pe} = \sqrt{3}$*, diffusive process plays the leading role for the stability requirement.*

Remark 4.2 *As the Peclet number gets higher values, we visualise from Figure 4.1 that the selection field of Courant number, so* Δt*, gets decreased. Consequently, looking at* (4.6)*, the solution becomes unconditionally unstable as* $\mathrm{Pe} \to \infty$ *and this coincides the fact that forward Euler time integration yields an unconditionally unstable numerical solution for pure convective problem.*

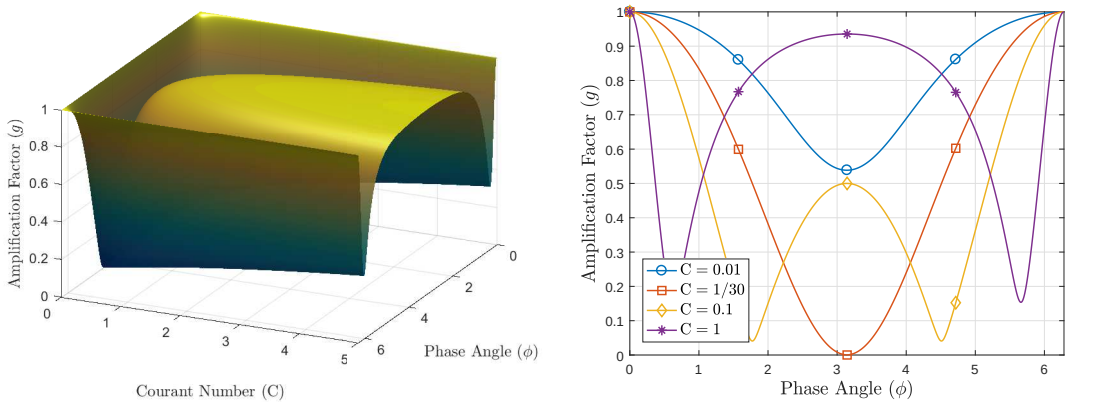
Crank-Nicolson Case For $\theta = \frac{1}{2}$, the amplification factor (4.1) together with (4.3) is given by

$$g = \frac{1 - \dfrac{3\mathrm{C}}{2}\dfrac{\dfrac{1}{\mathrm{Pe}}(1-\cos\phi) + i\sin\phi}{\cos\phi + 2}}{1 + \dfrac{3\mathrm{C}}{2}\dfrac{\dfrac{1}{\mathrm{Pe}}(1-\cos\phi) + i\sin\phi}{\cos\phi + 2}}. \tag{4.8}$$

A careful calculation leads

$$|g|^2 = \frac{\big(2\mathrm{Pe}(\cos\phi + 2) - 3\mathrm{C}(1-\cos\phi)\big)^2 + 9\mathrm{C}^2\mathrm{Pe}^2\sin^2\phi}{\big(2\mathrm{Pe}(\cos\phi + 2) + 3\mathrm{C}(1-\cos\phi)\big)^2 + 9\mathrm{C}^2\mathrm{Pe}^2\sin^2\phi}.$$

One can conclude that the above expression remains bounded by 1. In other words, the von Neumann stability condition holds true for any choice of Peclet number and Courant number which gives the unconditional stability. This fact could be visualized by observing the Figure 4.3 and Figure 4.4 for some choices of Courant numbers.

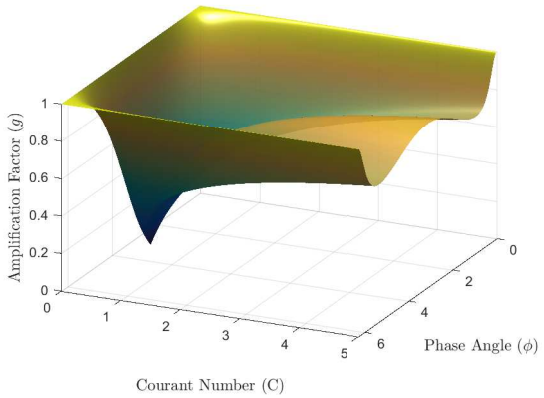


(a) 3D plot of the amplification factor.

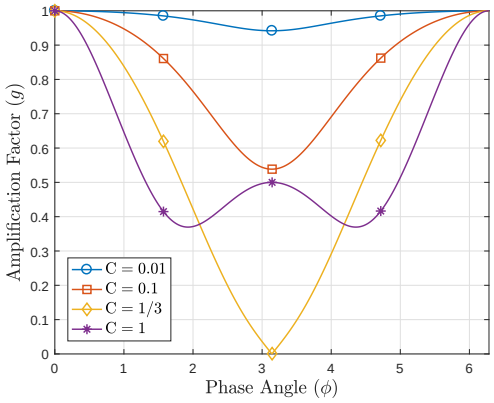
(b) Plots of the amplification factor for various choices of Courant numbers.

Figure 4.3. Amplification factor for $\mathrm{Pe} = 0.1$ using SGFEM in space and Crank-Nicolson method in time.

In spite of the unconditional stability of the Crank-Nicolson method, choice of Δt, or Courant number, should not be an arbitrary process, since choosing sufficiently large time step may cause the method has a tendency to oscillate. This situation requires an extra attention, so we give the following remark.



(a) 3D plot of the amplification factor.



(b) Plots of the amplification factor for various choices of Courant numbers.

Figure 4.4. Amplification factor for $\mathrm{Pe} = 1$ using SGFEM in space and Crank-Nicolson method in time.

Remark 4.3 *Looking at the expression* (4.8)*, observe that the complex part of the amplification factor vanishes for $\phi = \pi$. In other words, this choice annihilates the effect of the contribution of the convective process to the numerical solution, that is to say, only diffusive process plays the role on the approximate solution for this mode. On the other hand, looking at the plots of the amplification factors for the case $\mathrm{C} = \frac{1}{30}$ on Figure 4.3 and for the case $\mathrm{C} = \frac{1}{3}$ on Figure 4.4, we see that the amplification factor touches to zero for $\phi = \pi$, which means there is neither growth nor decay on the approximate solution for that Fourier mode. This is indicative of potential node-to-node oscillatory behaviour of the numerical solution [34].*

Now in order to obtain the oscillating limit, it follows from employing $\phi = \pi$ on the amplification factor (4.8) *and forcing $g(\pi) = 0$ that*

$$g(\pi) = \frac{1 - \dfrac{3\mathrm{C}}{\mathrm{Pe}}}{1 + \dfrac{3\mathrm{C}}{\mathrm{Pe}}} = 0.$$

Solving this equation for C, we therefore have the oscillating limit

$$\mathrm{C} = \frac{\mathrm{Pe}}{3}.$$

Corollary 4.1 *Let the problem* (2.1)-(2.4) *be discretized by SGFEM in space and Crank-Nicolson method in time. Then, the oscillating limit for the corresponding approximate solution is given by*

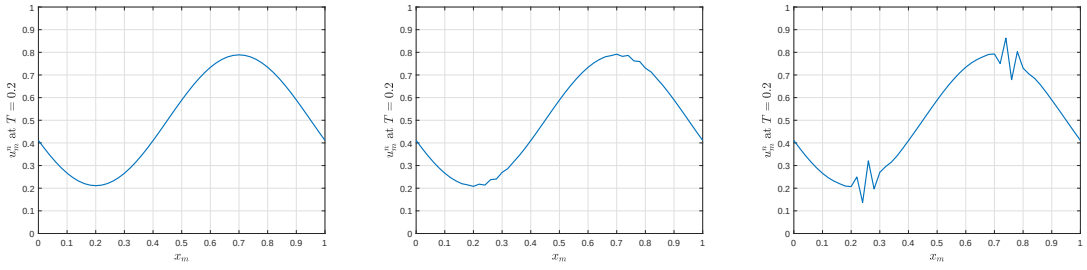
$$\mathrm{C} = \frac{\mathrm{Pe}}{3}. \tag{4.9}$$

We illustrate this with the following example.

Example 4.1 *Consider the unsteady convection–diffusion problem, where diffusion coefficient is set to* $\kappa = 0.1$ *and convective field is* $\beta = 1$*. No external forces are acting on the system and final time is set to* $T = 0.2$*. The initial data is defined as*

$$u_0(x) = \begin{cases} 1 & \text{if } |x - 0.5| \leq 0.25 \\ 0 & \text{otherwise} \end{cases}.$$

In Figure 4.5 numerical solutions derived by using SGFEM in space and Crank-Nicolson method in time for various choices of Courant numbers are represented. Peclet



(a) $\mathrm{C} = 1/30$. (b) $\mathrm{C} = 0.5$. (c) $\mathrm{C} = 1$.

Figure 4.5. Approximate solution of Example 4.1 for various choices of Courant numbers.

number is chosen as $\mathrm{Pe} = 0.1$ *and Courant numbers are chosen as* $\mathrm{C} = \frac{1}{30}$, $\mathrm{C} = 0.5$ *and* $\mathrm{C} = 1$ *respectively. Note that the first one corresponds for the oscillating limit* (4.9)*. We observe that the numerical solutions for the corresponding choices of Courant numbers remain stable. However, for* $\mathrm{C} = 0.5$*, we see from Figure 4.5(b) that small spurious oscillations start occuring near the points* $x = 0.25$ *and* $x = 0.75$ *where the initial condition has jump discontinuties. For* $\mathrm{C} = 1$*, we observe from Figure 4.5(c) that the oscillations become more visible around these points.*

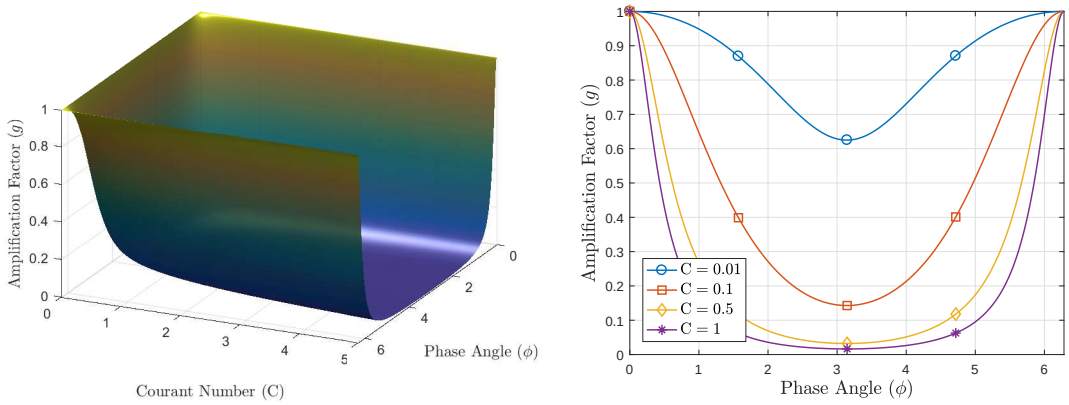
Backward Euler Case For $\theta = 1$, the amplification factor (4.1) in terms of Peclet number and Courant number is given by

$$g = \left(1 + 3\mathrm{C}\frac{\frac{1}{\mathrm{Pe}}(1-\cos\phi) + i\sin\phi}{\cos\phi + 2} \right)^{-1}.$$

Following from the above expression, we have

$$|g|^2 = \frac{\mathrm{Pe}^2(2+\cos\phi)^2}{\big(\mathrm{Pe}(2+\cos\phi) + 3\mathrm{C}(1-\cos\phi)\big)^2 + 9\mathrm{C}^2\mathrm{Pe}^2\sin^2\phi}$$

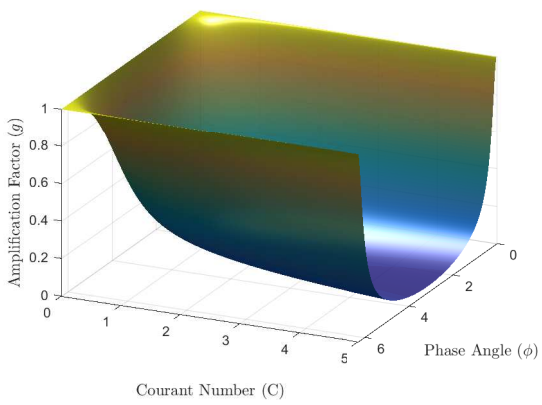
and this remains bounded by 1 for all choices of Peclet number and Courant number. Consequently, backward Euler method yields an unconditionally stable method similar to the Crank-Nicolson case. This conclusion can also be observed by looking to the figures 4.6 and 4.7.



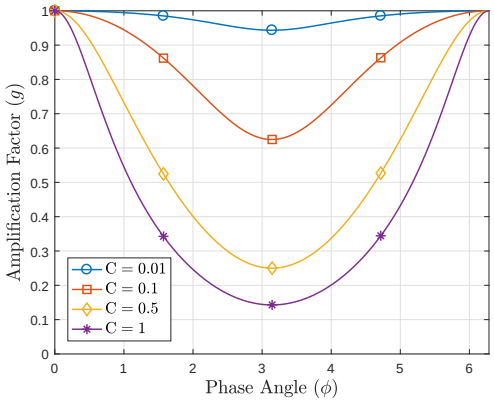
(a) 3D plot of the amplification factor.

(b) Plots of the amplification factor for various choices of Courant numbers.

Figure 4.6. Amplification factor for $\mathrm{Pe} = 0.1$ using SGFEM in space and backward Euler method in time.



(a) 3D plot of the amplification factor.



(b) Plots of the amplification factor for various choices of Courant numbers.

Figure 4.7. Amplification factor for $\mathrm{Pe} = 1$ using SGFEM in space and backward Euler method in time.

The summary of the stability conclusions given in the following.

Corollary 4.2 *Consider the fully discretized form of the model problem* (3.1)-(3.3) *obtained by using SGFEM in space and generalized trapezoidal rule in time.*

i. Let $\theta = 0$ *and assume that* $\mathrm{Pe} \leq \sqrt{3}$ *holds true. Then the time stability requirement is given by*

$$\mathrm{C} < \frac{\mathrm{Pe}}{3}.$$

ii. Let $\theta = 0$ *and assume that* $\mathrm{Pe} > \sqrt{3}$ *holds true. Then the time stability requirement is given by*

$$\mathrm{C} < \frac{1}{\mathrm{Pe}}.$$

iii. Let $\theta = \frac{1}{2}$ *or* $\theta = 1$. *Then the numerical solution is stable independent of choosing Courant number and Peclet number, i.e. space and time mesh sizes.*

4.2. Accuracy Analysis

In this section, using the formulations we derived on Section 3.2, we express the asymptotic expansions of numerical temporal damping and numerical temporal frequency in order to derive the accuracy in damping and accuracy in frequency with respect to space and time mesh. Then we plot the errors in damping and in phase speed, to visualise whether the computed temporal errors are more than enough or less than enough. As we stated at the begining of the chapter, we consider only the cases $\mathrm{Pe} \leq 1$ and in addition, $\mathrm{Pe} = 0$ case for relative damping error. For the latter case, note that, the Courant number is zero and therefore, the parameter we are going to use for the mesh ratio is the diffusion number, μ.

4.2.1. Relative Damping Error

Recall that the formulation of relative damping error is given in Definition 3.8. As we did for the amplification factor, we find it necessary to re-express the formulation in terms of Courant number and Peclet number as

$$\varepsilon_D = -\frac{2\mathrm{Pe}}{\mathrm{C}\phi^2} \ln |g| \tag{4.10}$$

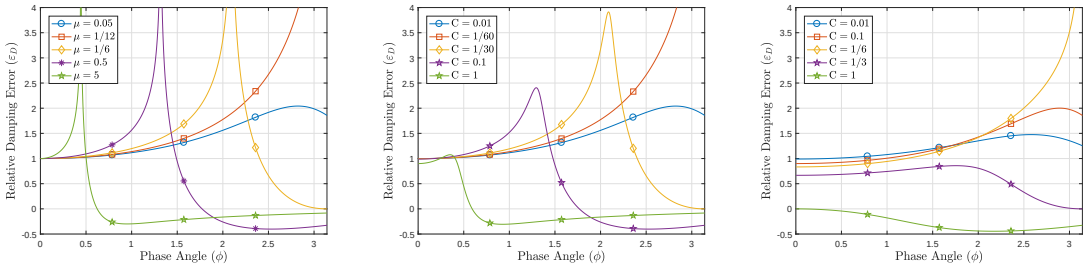
for our further analysis. We give the related results using forward Euler, Crank-Nicolson and backward Euler methods in the following.

Forward Euler Case Let $\theta = 0$. Using (3.19), series expansion of numerical damping factor in terms of ϕ is

$$\xi^h = \xi - \frac{\beta^2 \Delta t}{2\Delta x^2}\phi^2 + \frac{1}{12\Delta x^4}\left(6\kappa^2\Delta t - 12\kappa\beta^2\Delta t^2 + \kappa\Delta x^2 + 3\beta^4\Delta t^3\right)\phi^4 + \mathcal{O}(\phi^6).$$

In Figure 4.8, plots of the relative damping error for $\mathrm{Pe} = 0$, $\mathrm{Pe} = 0.1$ and $\mathrm{Pe} = 1$ cases are shown. As we stated in Remark 3.4, $\varepsilon_D < 0$ is interpreted as an unstable solution. In fact, observing from the figures that, choice of diffusion number above from $\frac{1}{6}$ for pure diffusive case, Courant numbers above from $\frac{1}{30}$ for $\mathrm{Pe} = 0.1$ case and above from $\frac{1}{3}$ for $\mathrm{Pe} = 1$ case, which are the corresponding stability limits, causes the error to drop to the

negative part. Another point of view which is an important and significant observation for $\mathrm{Pe} > 0$ cases is the underdamped behaviour in the low frequency range, i.e. in the range of $\phi \to 0$, that is indication of possible unstable behaviour of the approximate solution. It becomes more visible as we increase the Peclet number.



(a) $\mathrm{Pe} = 0$. (b) $\mathrm{Pe} = 0.1$. (c) $\mathrm{Pe} = 1$.

Figure 4.8. Relative damping error for various values of diffusion numbers ($\mathrm{Pe} = 0$) and Courant numbers ($\mathrm{Pe} \neq 0$) using SGFEM in space and forward Euler method in time.

Crank-Nicolson Case For $\theta = \frac{1}{2}$, series expansion of the numerical damping factor in terms of ϕ is

$$\xi^h = \xi \; - \; \frac{1}{12\Delta x^4}\left(-3\kappa\beta^2\Delta t^2 + \kappa\Delta x^2\right)\phi^4 + \frac{1}{720\Delta x^6}\left(60\kappa^3\Delta t^2 + 45\kappa\beta^4\Delta t^4 + 2\kappa\Delta x^4 - 15\kappa\beta^2\Delta x^2\Delta t^2\right)\phi^6 + \mathcal{O}(\phi^8).$$

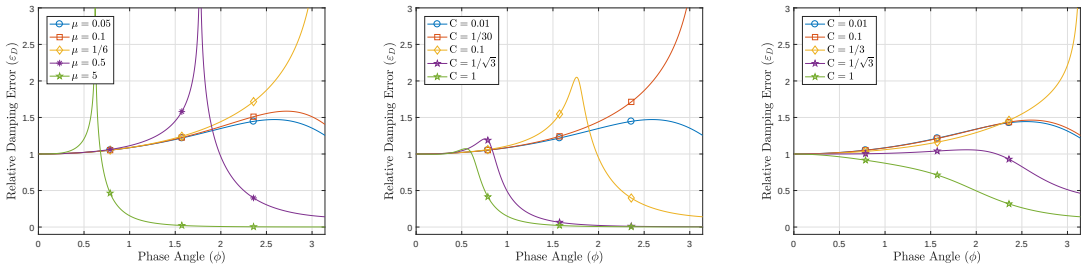
We first give the following remarks.

Remark 4.4 *Choosing the Courant number as* $\mathrm{C} = \frac{1}{\sqrt{3}}$ *results that the coefficient of the forth order term on the above expansion becomes zero. Consequently such choice will increase the accuracy of the damping of the method in space dimension to* $\mathcal{O}(\Delta x^4)$.

Remark 4.5 *Considering* $\mathrm{Pe} \to \infty$, *i.e.* $\kappa \to 0$, *we obtain* $\xi = \xi^h$ *which coincides the fact that Crank-Nicolson method is non-dissipative for pure convective problem.*

Relative damping errors of the numerical results for $\mathrm{Pe} = 0$, $\mathrm{Pe} = 0.1$ and $\mathrm{Pe} = 1$ cases are shown in Figure 4.9. For all cases, choice of low diffusion numbers or low Courant numbers results an overdamped numerical solution. Unlike the forward Euler case, damping error behaves as optimal in the low frequncy range. $\mu = \frac{1}{6}$ for pure diffusive case, $\mathrm{C} = \frac{1}{30}$ for $\mathrm{Pe} = 0.1$ case and $\mathrm{C} = \frac{1}{3}$ for $\mathrm{Pe} = 1$ case are the oscillating limits.

It can be interpreted from each figure that, the error has a tendency to the $\varepsilon_D = 0$ level, i.e. no damping level, for the choice of higher parameters than these oscillating limits.



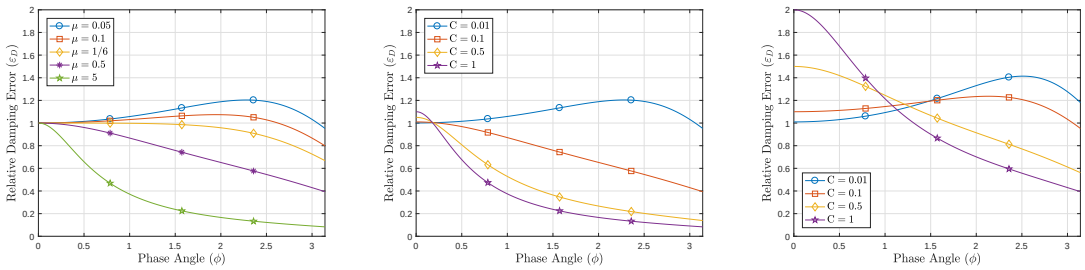
(a) $\mathrm{Pe} = 0$. (b) $\mathrm{Pe} = 0.1$. (c) $\mathrm{Pe} = 1$.

Figure 4.9. Relative damping error for various values of diffusion numbers ($\mathrm{Pe} = 0$) and Courant numbers ($\mathrm{Pe} \neq 0$) using SGFEM in space and Crank-Nicolson method in time.

Backward Euler Method Letting $\theta = 1$, series expansion of the numerical damping in terms of ϕ is

$$\xi^h = \xi + \frac{\beta^2 \Delta t}{2\Delta x^2}\phi^2 + \frac{1}{12\Delta x^4}\left(-6\kappa^2\Delta t - 12\kappa\beta^2\Delta t^2 + \kappa\Delta x^2 - 3\beta^4\Delta t^3\right)\phi^4 + \mathcal{O}(\phi^6).$$

Related plots for relative damping errors for $\mathrm{Pe} = 0$, $\mathrm{Pe} = 0.1$ and $\mathrm{Pe} = 1$ cases are shown in Figure 4.10. Comparing the figures, we see that the backward Euler method



(a) $\mathrm{Pe} = 0$. (b) $\mathrm{Pe} = 0.1$. (c) $\mathrm{Pe} = 1$.

Figure 4.10. Relative damping error for various values of diffusion numbers ($\mathrm{Pe} = 0$) and Courant numbers ($\mathrm{Pe} \neq 0$) using SGFEM in space and backward Euler method in time.

has a better damping property for pure diffusive case and for low Peclet numbers. For the $\mathrm{Pe} = 1$ case, we may comment that large Courant numbers are bad at both ends since it overdamps the long waves and underdamps the short waves. Especially the behaviour in

the low frequency range is important on making a comparison between Crank-Nicolson operator as it causes a serious damping even in the $\phi \to 0$ range.

Remark 4.6 *Despite being out of interest, we require it necessary to state that, for pure diffusive case, i.e.* $\beta = 0$*, choice of* $\mu = \frac{1}{6}$ *comes out with canceling the forth order term which increases the accuracy of damping of the method in time up to order 2.*

4.2.2. Relative Phase Speed Error

The expression for relative phase speed error that we derived on Definition 3.9 can be re-expressed in terms of Courant and Peclet numbers as

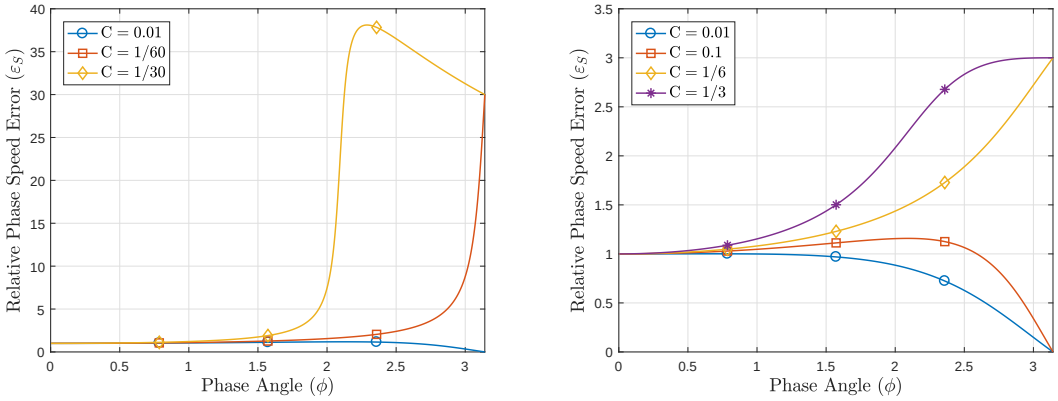
$$\varepsilon_S = -\frac{1}{\mathrm{C}\phi}\arg(g). \tag{4.11}$$

Related analysis for forward Euler, Crank-Nicolson and backward Euler cases are given in the following.

Forward Euler Case Taking $\theta = 0$, the series expansion of numerical temporal frequency in terms of ϕ is

$$\begin{aligned}\omega^h = \omega \quad &+ \quad \frac{1}{3\Delta x^3}\left(-3\kappa\beta\Delta t + \beta^3\Delta t^2\right)\phi^3 \\ &+ \quad \frac{1}{180\Delta x^5}\left(-180\kappa^2\beta\Delta t^2 - 15\kappa\beta\Delta t\Delta x^2 + 180\kappa\beta^3\Delta t^3\right. \\ &- \quad \left.36\beta^5\Delta t^4 + \beta\Delta x^4\right) + \mathcal{O}(\phi^7).\end{aligned}$$

Plots of the relative phase speed error for $\mathrm{Pe} = 0.1$ and $\mathrm{Pe} = 1$ cases are shown in Figure 4.11. $\mathrm{C} = \frac{1}{30}$ and $\mathrm{C} = \frac{1}{3}$ are the corresponding stability limits. Except for low Courant numbers, all choices result with a leading error. To capture nice phase behaviour, low Courant number is required to be chosen.

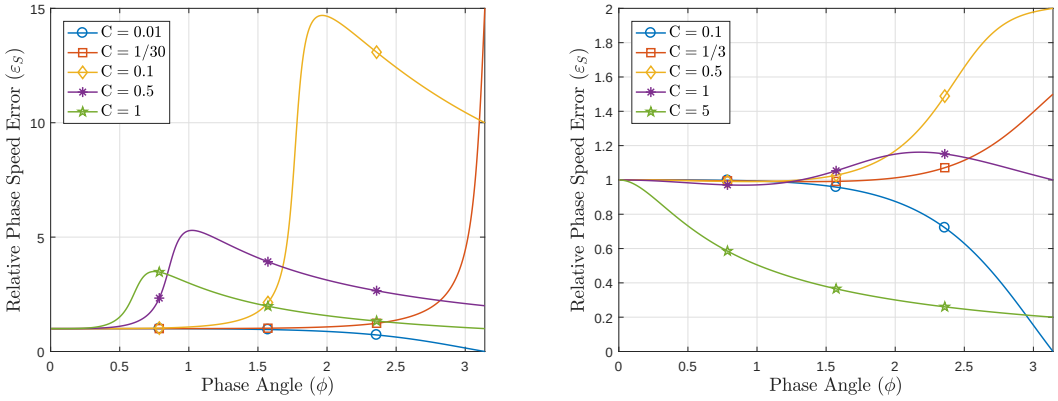


(a) Pe $= 0.1$. (b) Pe $= 1$.

Figure 4.11. Relative phase speed error for various values of Courant numbers, using SGFEM in space and forward Euler method in time.

Crank-Nicolson Case For $\theta = \frac{1}{2}$, series expansion of the numerical temporal frequency in terms of ϕ is given by

$$\omega^h = \omega + \frac{\beta^3 \Delta t^2}{12 \Delta x^3}\phi^3 + \frac{1}{720 \Delta x^5}\left(-180\kappa^2\beta\Delta t^2 - 9\beta^5\Delta t^4 + 4\beta\Delta x^4\right)\phi^5 + \mathcal{O}(\phi^7).$$



(a) Pe $= 0.1$. (b) Pe $= 1$.

Figure 4.12. Relative phase speed error for various values of Courant numbers, using SGFEM in space and Crank-Nicolson method in time.

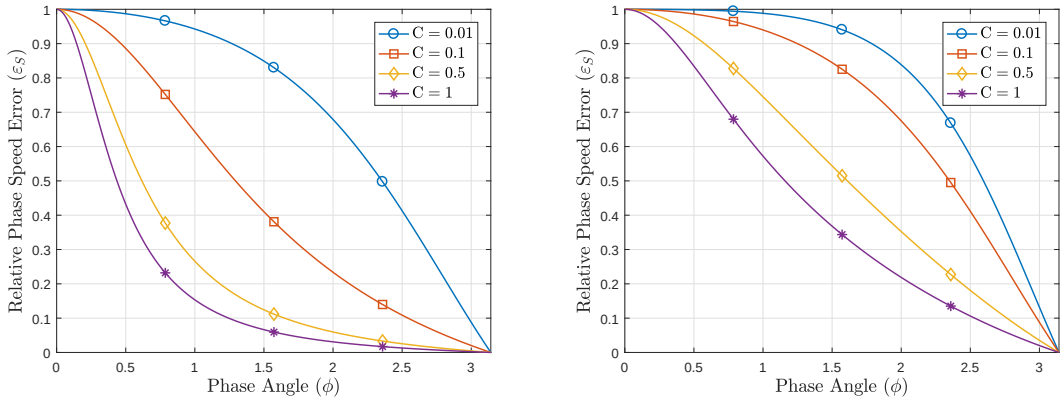
See Figure 4.12 for relative phase speed errors for $\mathrm{Pe} = 0.1$ and $\mathrm{Pe} = 1$ cases. For diffusion dominated case, numerical solutions have a leading error except for the low Courant numbers. As Pe increases, it seems that the method becomes more successful on phase speed accuracy. However, it is observed from Figure 4.12(b) that, large choice of Courant numbers causes a slow down on the waves. The behaviour is not monotonic so

that neither small nor large Courany number is required to catch a nice accuracy on phase speed.

Backward Euler Case Employing $\theta = 1$ on (4.11) and expanding in terms of ϕ, we have

$$\begin{aligned}\omega^h = \omega \quad &+ \quad \frac{1}{3\Delta x^3}\left(3\kappa\beta\Delta t + \beta^3\Delta t^2\right)\phi^3 \\ &+ \quad \frac{1}{180\Delta x^5}\left(-180\kappa^2\beta\Delta t^2 + 15\kappa\beta\Delta t\Delta x^2 - 180\kappa\beta^3\Delta t^3\right. \\ &- \quad \left.36\beta^5\Delta t^4 + \beta\Delta x^4\right) + \mathcal{O}(\phi^7).\end{aligned}$$

Figure 4.13 shows the relative phase speed error for the backward Euler method. It is obvious to observe that, the method causes a delay for any choice of Courant number and small Courant numbers are required in order to obtain a reasonable accuracy.



(a) Pe $= 0.1$. (b) Pe $= 1$.

Figure 4.13. Relative phase speed error for various values of Courant numbers, using SGFEM in space and backward Euler method in time.

4.2.3. Some Remarks and Conclusions

Consistency Looking through the series expansions of numerical damping and numerical frequency, difference of them with their exact counterparts tends to zero as $\Delta t \to 0$ and $\Delta x \to 0$. This provides the consistency.

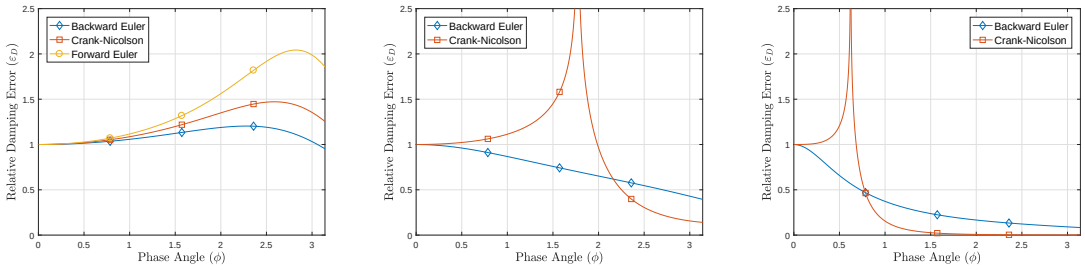
Accuracy In Table 4.1, order of the accuracies in damping and in frequency of each time stepping methods are shown. The orders are obtained by the looking at the lowest order terms of the corresponding series expansions. Also the orders correspond to a particular choice of Courant number, which is stated in Remark 4.4 are shown.

Looking through the first column of the table, all of the methods are equally accurate in space except the particular choice of Courant number using Crank-Nicolson method. On the other hand, this choice does not have a contribution on the accuracy of frequency in space dimension and all methods are fourth order accurate. For the time accuracy, Crank-Nicolson method has a second order accuracy in time for both damping and frequency, whereas others are first order accurate.

Time Integrator	ξ^h	ω^h
Forward Euler	$\mathcal{O}\left(\Delta t, \Delta x^2\right)$	$\mathcal{O}\left(\Delta t, \Delta x^4\right)$
Crank-Nicolson	$\mathcal{O}\left(\Delta t^2, \Delta x^2\right)$	$\mathcal{O}\left(\Delta t^2, \Delta x^4\right)$
Crank-Nicolson with C $= \frac{1}{\sqrt{3}}$	$\mathcal{O}\left(\Delta t^2, \Delta x^4\right)$	$\mathcal{O}\left(\Delta t^2, \Delta x^4\right)$
Backward Euler	$\mathcal{O}\left(\Delta t, \Delta x^2\right)$	$\mathcal{O}\left(\Delta t, \Delta x^4\right)$

Table 4.1. Accuracy of the numerical damping and numerical frequency of the numerical schemes using different time integrators.

Comparison of Relative Damping Error In the figures 4.14-4.15, relative damping errors using each time stepping method are compared with eachother.

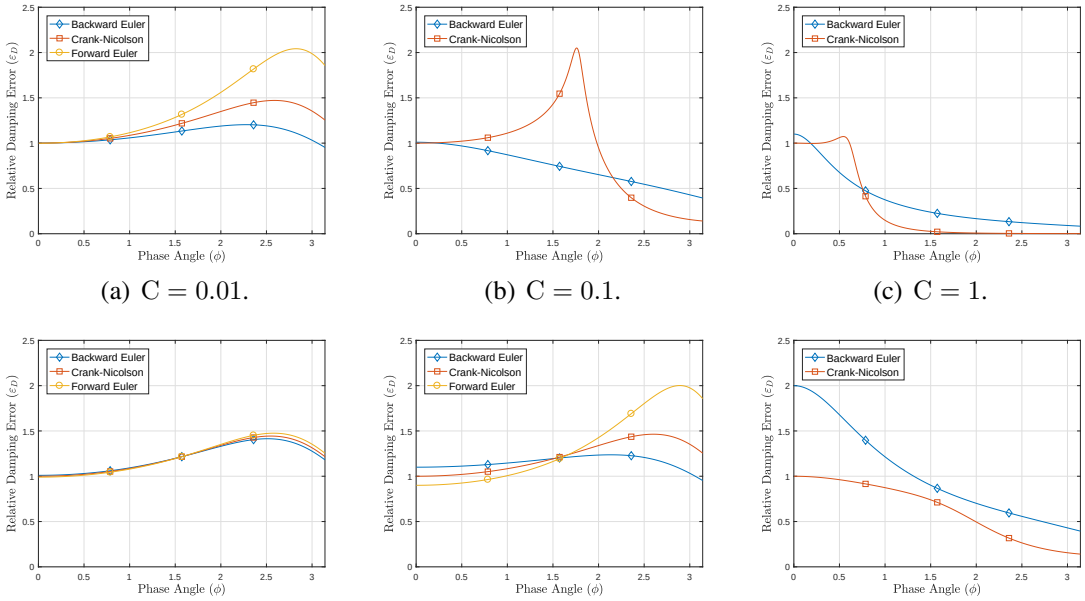


(a) $\mu = 0.05$. (b) $\mu = 0.5$. (c) $\mu = 5$.

Figure 4.14. Comparison of relative damping errors of the time integrators for pure diffusive case.

As the heat equation comes up with a smooth solution, we can focus on the low frequency range for pure diffusive case. For each choice, one can observe that backward Euler method has a better capability on the damping property.

Observing from 4.15(a)-4.15(c), this situation remains same. However, it starts changing in favor of Crank-Nicolson method, as we increase the Peclet number. In fact, backward Euler causes a serious damping in the range of $\phi \to 0$, whereas the behaviour



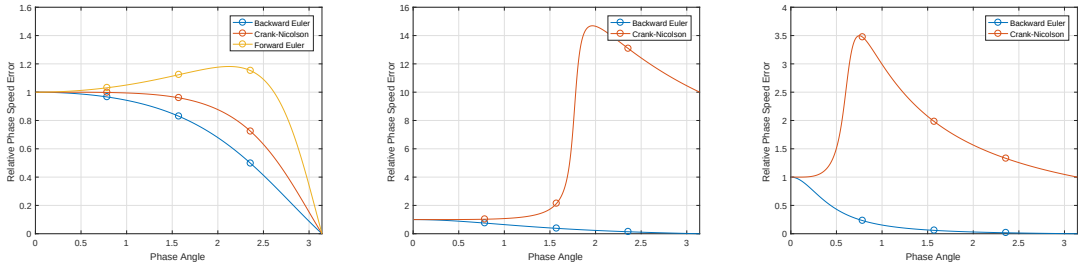
(a) $\mathrm{C} = 0.01$. (b) $\mathrm{C} = 0.1$. (c) $\mathrm{C} = 1$.

(d) $\mathrm{C} = 0.01$. (e) $\mathrm{C} = 0.1$. (f) $\mathrm{C} = 1$.

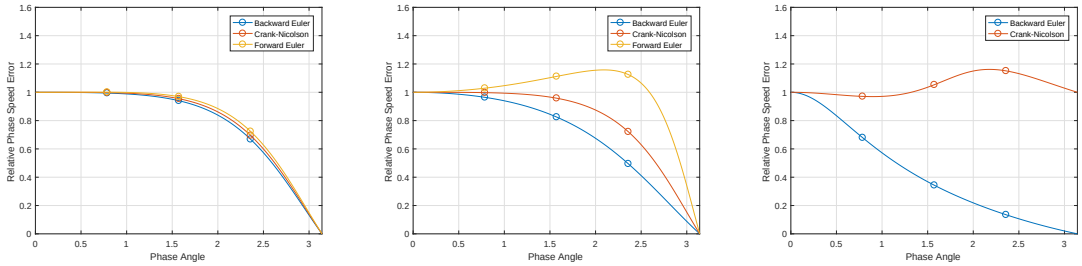
Figure 4.15. Comparison of relative damping errors of the time integrators for convective–diffusive case: $\mathrm{Pe} = 0.1$ on the above and $\mathrm{Pe} = 1$ on the below.

of Crank-Nicolson on damping is in the optimal range. Forward Euler method is most uncapable one out of three. Note that, the figures which do not contain the error obtained by forward Euler is because of the instability of the numerical solution for those choices of Courant and Peclet numbers.

Comparison of Relative Phase Speed Error In Figure 4.16, relative phase speed errors using each time integrator for $\mathrm{Pe} = 0.1$ case on the above and $\mathrm{Pe} = 1$ case on the below are shown. An immediate observation here is, all errors are close to the optimal level in the low frequency range. Secondly, as we increase the peclet number, we observe that Crank-Nicolson method has a better coherence out of three on the phase speed error. We again note that the plots that do not include the forward Euler case is due to the instable behaviour of the numerical solution.



(a) $\mathrm{C} = 0.01$. (b) $\mathrm{C} = 0.1$. (c) $\mathrm{C} = 1$.



(d) $\mathrm{C} = 0.01$. (e) $\mathrm{C} = 0.1$. (f) $\mathrm{C} = 1$.

Figure 4.16. Comparison of relative phase speed errors of the time integrators: $\mathrm{Pe} = 0.1$ on the above and $\mathrm{Pe} = 1$ on the below.

CHAPTER 5

STABILITY AND ACCURACY ANALYSIS OF SDFEM/θ-METHOD

As we have mentioned briefly on Section 1.1.1, SGFEM with low order piecewise polynomials develops spurious oscillations for Convection–Diffusion problems near boundary layer if the problem is convection dominated, i.e.

$$\frac{|\text{Coefficient of Convection}|}{|\text{Coefficient of Diffusion}|} \gg 1.$$

In order to avoid this undesired situation, now we set a stabilization term which we mention in Appendix B. We also briefly introduce the related discrete structure in Appendix A and express the corresponding finite difference formulation in Section 2.2. So, by considering the approximate solution for $\mathrm{Pe} > 1$ case, our main objective in this chapter is to observe and comment on the possible changes of the stability and temporal accuracy properties of the stabilized methods and also check whether there can be found any optimal choice of stabilization parameter that yields a well-behaved approximate solution with respect to the temporal errors.

5.1. Stability Analysis

Recall that the amplification factor, by Definition 3.5, is given by

$$g(\phi, \Delta x, \Delta t) = \frac{G_{B_\tau}}{G_{A_\tau}}$$

where, together with $\tau \neq 0$,

$$\begin{aligned} G_{A_\tau} &= \left(A_{\tau,m-1}e^{-i\phi} + A_{\tau,m} + A_{\tau,m+1}e^{i\phi}\right) \\ &= \frac{\Delta x}{6}\left(e^{-i\phi} + 4 + e^{i\phi}\right) + \frac{\tau\beta}{2}\left(e^{-i\phi} + e^{i\phi}\right) \\ &+ \theta\Delta t\left[-\frac{\kappa+\tau\beta^2}{\Delta x}\left(e^{-i\phi} - 2 + e^{i\phi}\right) - \frac{\beta}{2}\left(-e^{-i\phi} + e^{i\phi}\right)\right] \\ &= \frac{\Delta x}{3}(\cos\phi + 2) - i\tau\beta\sin\phi + \theta\Delta t\left[\frac{2(\kappa+\tau\beta^2)}{\Delta x}(1-\cos\phi) + i\beta\sin\phi\right] \end{aligned}$$

and

$$\begin{aligned} G_{B_\tau} &= \left(B_{\tau,m-1}e^{-i\phi} + B_{\tau,m} + B_{\tau,m+1}e^{i\phi}\right) \\ &= \frac{\Delta x}{6}\left(e^{-i\phi} + 4 + e^{i\phi}\right) + \frac{\tau\beta}{2}\left(e^{-i\phi} + e^{i\phi}\right) \\ &- (1-\theta)\Delta t\left[-\frac{\kappa+\tau\beta^2}{\Delta x}\left(e^{-i\phi} - 2 + e^{i\phi}\right) - \frac{\beta}{2}\left(-e^{-i\phi} + e^{i\phi}\right)\right] \\ &= \frac{\Delta x}{3}(\cos\phi + 2) - i\tau\beta\sin\phi - (1-\theta)\Delta t\left[\frac{2(\kappa+\tau\beta^2)}{\Delta x}(1-\cos\phi) + i\beta\sin\phi\right]. \end{aligned}$$

Proceeding with some basic operations, one may derive the amplification factor as

$$g = \frac{1-(1-\theta)\Delta t\lambda}{1+\theta\Delta t\lambda} \tag{5.1}$$

together with

$$\Delta t\lambda = \frac{\dfrac{2(\kappa+\tau\beta^2)}{\Delta x}(1-\cos\phi) + i\beta\sin\phi}{\dfrac{\Delta x}{3}(\cos\phi+2) - i\tau\beta\sin\phi}. \tag{5.2}$$

Re-expressing $\Delta t\lambda$, thus g, in terms of Peclet number and Courant number, we have

$$\Delta t\lambda = \mathrm{C}\frac{\left(\frac{1}{\mathrm{Pe}}+\frac{2\tau\beta}{\Delta x}\right)(1-\cos\phi)+i\sin\phi}{\frac{2+\cos\phi}{3}-i\frac{\tau\beta}{\Delta x}\sin\phi}. \tag{5.3}$$

In the following parts, we will continue with the stability analysis by considering the forward Euler, Crank-Nicolson and backward Euler cases.

Forward Euler Case Let $\theta = 0$. Then the amplification factor (5.1) with (5.3) is given by

$$g = 1 - \mathrm{C}\frac{\left(\frac{1}{\mathrm{Pe}}+\frac{2\tau\beta}{\Delta x}\right)(1-\cos\phi)+i\sin\phi}{\frac{2+\cos\phi}{3}-i\frac{\tau\beta}{\Delta x}\sin\phi}$$

Then,

$$|g|^2 = \frac{\left[\frac{2+\cos\phi}{3}-\mathrm{C}\left(\frac{1}{\mathrm{Pe}}+\frac{2\tau\beta}{\Delta x}\right)(1-\cos\phi)\right]^2+\left(\mathrm{C}+\frac{\tau\beta}{\Delta x}\right)^2(1-\cos^2\phi)}{\left(\frac{2+\cos\phi}{3}\right)^2+\left(\frac{\tau\beta}{\Delta x}\right)^2(1-\cos^2\phi)}. \tag{5.4}$$

Numerator of the last expression can be rewritten as

$$\begin{aligned}\left(\frac{2+\cos\phi}{3}\right)^2 &- 2\mathrm{C}\left(\frac{1}{\mathrm{Pe}}+\frac{2\tau\beta}{\Delta x}\right)\frac{(2+\cos\phi)(1-\cos\phi)}{3}\\ &+ \mathrm{C}^2\left(\frac{1}{\mathrm{Pe}}+\frac{2\tau\beta}{\Delta x}\right)^2(1-\cos\phi)^2+\left(\mathrm{C}+\frac{\tau\beta}{\Delta x}\right)^2(1-\cos^2\phi)\end{aligned}$$

or equivalently as

$$\begin{aligned}\left(\frac{2+\cos\phi}{3}\right)^2 &+ \left(\frac{\tau\beta}{\Delta x}\right)^2(1-\cos^2\phi)\\ &+ 2\mathrm{C}(1-\cos\phi)\left[-\left(\frac{1}{\mathrm{Pe}}+\frac{2\tau\beta}{\Delta x}\right)\frac{2+\cos\phi}{3}+\frac{\tau\beta}{\Delta x}(1+\cos\phi)\right]\\ &+ \mathrm{C}^2(1-\cos\phi)\left[\left(\frac{1}{\mathrm{Pe}}+\frac{2\tau\beta}{\Delta x}\right)^2(1-\cos\phi)+(1+\cos\phi)\right].\end{aligned}$$

Now dividing this with the denominator of (5.4), forcing the stability condition $|g|^2 < 1$ and simplifying the resulting expression yields

$$\begin{aligned}2\left[-\left(\frac{1}{\mathrm{Pe}}+\frac{2\tau\beta}{\Delta x}\right)\frac{2+\cos\phi}{3} &+ \frac{\tau\beta}{\Delta x}(1+\cos\phi)\right]\\ &+ \mathrm{C}\left[\left(\frac{1}{\mathrm{Pe}}+\frac{2\tau\beta}{\Delta x}\right)^2(1-\cos\phi)+(1+\cos\phi)\right]<0.\end{aligned}$$

If follows from leaving C on the left hand side of the last inequality that

$$\begin{aligned}\mathrm{C} &< \frac{2\left[\left(\frac{1}{\mathrm{Pe}}+\frac{2\tau\beta}{\Delta x}\right)\frac{2+\cos\phi}{3}-\frac{\tau\beta}{\Delta x}(1+\cos\phi)\right]}{\left(\frac{1}{\mathrm{Pe}}+\frac{2\tau\beta}{\Delta x}\right)^2(1-\cos\phi)+(1+\cos\phi)}\\ &= \frac{2}{3\Delta x\mathrm{Pe}}\frac{2\Delta x+\tau\mathrm{Pe}\beta+(\Delta x-\tau\mathrm{Pe}\beta)\cos\phi}{\left(\frac{1}{\mathrm{Pe}}+\frac{2\tau\beta}{\Delta x}\right)^2(1-\cos\phi)+(1+\cos\phi)}.\end{aligned}\tag{5.5}$$

Now define

$$f(\mathrm{Pe},\phi) := \frac{2}{3\Delta x\mathrm{Pe}}\frac{2\Delta x+\tau\mathrm{Pe}\beta+(\Delta x-\tau\mathrm{Pe}\beta)\cos\phi}{\left(\frac{1}{\mathrm{Pe}}+\frac{2\tau\beta}{\Delta x}\right)^2(1-\cos\phi)+(1+\cos\phi)}.\tag{5.6}$$

We will obtain the stability requirements through the inequality $\mathrm{C} \leq f(\mathrm{Pe}, \phi)$ by minimizing $f(\mathrm{Pe}, \phi)$ with respect to ϕ. But first, our expectation from $f(\mathrm{Pe}, \phi)$ is to be a positive function for all values of ϕ in order to obtain a reasonable inequality. So we give the following lemma.

Lemma 5.1 *f, defined by* (5.6)*, is a positive function of ϕ for any given positive real parameters Δx, β and τ, and for any choice of Peclet number.*

Proof It is enough to show that

$$2\Delta x + \tau \mathrm{Pe}\beta + (\Delta x - \tau \mathrm{Pe}\beta)\cos\phi > 0$$

or

$$2 + \frac{\tau \mathrm{Pe}\beta}{\Delta x} + \left(1 - \frac{\tau \mathrm{Pe}\beta}{\Delta x}\right)\cos\phi > 0$$

holds true as it is straightforward to see that the denominator of (5.6) is a positive function of ϕ. Define $\eta := \frac{\tau \mathrm{Pe}\beta}{\Delta x} > 0$.

Case 1: $0 < \eta \leq 1$. Then,

$$2 + \eta + (1 - \eta)\cos\phi = 1 + (1 + \cos\phi) + \eta(1 - \cos\phi) > 1.$$

Case 2. $\eta > 1$. Then,

$$2 + \eta + (1 - \eta)\cos\phi = \eta(1 - \cos\phi) + 2 + \cos\phi > 1.$$

Hence, we end up with positiveness of the numerator of f. Thus f is itself a positive function of ϕ. This provides us a reasonable inequality for $\mathrm{C} \leq f(\mathrm{Pe}, \phi)$, as Courant number is supposed to be a positive quantity. □

Based on the positiveness of f, we can now minimize $f(\mathrm{Pe}, \phi)$ with respect to ϕ and take the lowest of all choices of $f(\mathrm{Pe}, \phi)$ to obtain a stability requirement for the forward Euler method. This is given by the following lemma.

Lemma 5.2 *Let Δx, β and τ are given parameters. Then f, defined on* (5.6)*, attains its minimum at $\phi = \pi$ for $\mathrm{Pe} < \frac{3\tau\beta}{\Delta x} + \sqrt{3 + \left(\frac{3\tau\beta}{\Delta x}\right)^2}$ and at $\phi = 0$ for $\mathrm{Pe} > \frac{3\tau\beta}{\Delta x} + \sqrt{3 + \left(\frac{3\tau\beta}{\Delta x}\right)^2}$. The minimum values are given by*

$$f(\mathrm{Pe}, \pi) = \frac{1}{3}\left(\frac{2\tau\beta}{\Delta x} + \frac{1}{\mathrm{Pe}}\right)^{-1} \tag{5.7}$$

and

$$f(\mathrm{Pe}, 0) = \frac{1}{\mathrm{Pe}} \tag{5.8}$$

respectively.

Proof Differentiating (5.6) with respect to ϕ results in

$$\frac{\partial f}{\partial \phi} = \frac{\Delta x^2 \mathrm{Pe}(\Delta x + 2\tau\mathrm{Pe}\beta)\left[\Delta x(\mathrm{Pe}^2 - 3) - 6\tau\mathrm{Pe}\beta\right]\sin\phi}{3\left[\Delta x^2(\mathrm{Pe}^2+1) + (\Delta x(\mathrm{Pe}-1) - 2\tau\mathrm{Pe}\beta)(\Delta x(\mathrm{Pe}+1) + 2\tau\mathrm{Pe}\beta) + 4\tau\mathrm{Pe}\beta(\tau\mathrm{Pe}\beta+1)\right]^2}.$$

Observe that $f(\mathrm{Pe}, \phi)$ is constant along $\Delta x(\mathrm{Pe}^2 - 3) - 6\tau\mathrm{Pe}\beta = 0$. Solving this quadratic equation for Pe yields

$$\mathrm{Pe} = \frac{3\tau\beta}{\Delta x} + \sqrt{3 + \left(\frac{3\tau\beta}{\Delta x}\right)^2} \tag{5.9}$$

or

$$\mathrm{Pe} = \frac{3\tau\beta}{\Delta x} - \sqrt{3 + \left(\frac{3\tau\beta}{\Delta x}\right)^2}.$$

The latter is negative, so we choose (5.9) as a threshold level to minimize $f(\mathrm{Pe}, \phi)$.

Case 1. Let $\mathrm{Pe} < \frac{3\tau\beta}{\Delta x} + \sqrt{3 + \left(\frac{3\tau\beta}{\Delta x}\right)^2}$. Then $\frac{\partial f}{\partial \phi} < 0$ and so, for a fix Peclet number, f gets its minimum at $\phi = \pi$. Thus, following from (5.6),

$$\begin{aligned} f(\mathrm{Pe}, \pi) &= \frac{2}{3\Delta x \mathrm{Pe}} \frac{2\Delta x + \tau \mathrm{Pe}\beta - (\Delta x - \tau \mathrm{Pe}\beta)}{2\left(\frac{2\tau\beta}{\Delta x} + \frac{1}{\mathrm{Pe}}\right)^2} \\ &= \frac{1}{3}\left(\frac{2\tau\beta}{\Delta x} + \frac{1}{\mathrm{Pe}}\right)^{-1}. \end{aligned}$$

Case 2. Let $\mathrm{Pe} > \frac{3\tau\beta}{\Delta x} + \sqrt{3 + \left(\frac{3\tau\beta}{\Delta x}\right)^2}$. Then $\frac{\partial f}{\partial \phi} < 0$, i.e. f is increasing with respect to ϕ and consequently to get the minimum value, letting $\phi \to 0$ yields

$$f(\mathrm{Pe}, 0) = \frac{2}{3\Delta x \mathrm{Pe}} \frac{2\Delta x + \tau \mathrm{Pe}\beta + (\Delta x - \tau \mathrm{Pe}\beta)}{2} = \frac{1}{\mathrm{Pe}}.$$

One can deduce by choosing the Peclet number as $\mathrm{Pe} = \frac{3\tau\beta}{\Delta x} + \sqrt{3 + \left(\frac{3\tau\beta}{\Delta x}\right)^2}$ that, the minimum values (5.7) and (5.8) become equivalent. □

Along with obtaining the minimum values of $f(\mathrm{Pe}, \phi)$, we can now conclude with the stability requirements. Following from (5.5), for $\mathrm{Pe} < \frac{3\tau\beta}{\Delta x} + \sqrt{3 + \left(\frac{3\tau\beta}{\Delta x}\right)^2}$ we have

$$\mathrm{C} < f(\mathrm{Pe}, \pi) = \frac{1}{3}\left(\frac{2\tau\beta}{\Delta x} + \frac{1}{\mathrm{Pe}}\right)^{-1}. \tag{5.10}$$

Contrary for $\mathrm{Pe} > \frac{3\tau\beta}{\Delta x} + \sqrt{3 + \left(\frac{3\tau\beta}{\Delta x}\right)^2}$, then we have

$$\mathrm{C} < f(\mathrm{Pe}, 0) = \frac{1}{\mathrm{Pe}}$$

or equivalently

$$\Delta t < \frac{2\kappa}{\beta^2}. \tag{5.11}$$

We give the following remarks.

Remark 5.1 *When we compare the diffusion dominated stability limits obtained by standard Galerkin finite element discretization and the stabilized method which are given by* (4.4) *and* (5.10) *respectively, we see on the latter case that there is an additional term on the denominator of the condition which depends on the convective velocity and the element measure. So the price of removing the boundary layer problem rises from standard Galerkin finite element discretization is restricting the choice interval of time step. Also, employing* $\tau = 0$*, note that we recover the estimate obtained by using SGFEM.*

Remark 5.2 *Contrary, stability condition for the convection dominated case* (5.11) *is exactly same as the one we obtained by SGFEM which is given by* (4.7)*. Whence we may comment that the stabilization parameter does not have a positive, nor negative effect on choosing the time step so that, the condition still preserves its strongness especially when* $0 < \kappa \ll 1$.

Crank-Nicolson Case Taking $\theta = \frac{1}{2}$, the amplification factor (5.1) together with (5.3) is

$$g = \frac{1 - \dfrac{\mathrm{C}}{2}\dfrac{\left(\dfrac{1}{\mathrm{Pe}} + \dfrac{2\tau\beta}{\Delta x}\right)(1-\cos\phi) + i\sin\phi}{\dfrac{2+\cos\phi}{3} - i\dfrac{\tau\beta}{\Delta x}\sin\phi}}{1 + \dfrac{\mathrm{C}}{2}\dfrac{\left(\dfrac{1}{\mathrm{Pe}} + \dfrac{2\tau\beta}{\Delta x}\right)(1-\cos\phi) + i\sin\phi}{\dfrac{2+\cos\phi}{3} - i\dfrac{\tau\beta}{\Delta x}\sin\phi}} \tag{5.12}$$

Then,

$$|g|^2 = \frac{\left[\dfrac{2(2+\cos\phi)}{3} - \mathrm{C}\left(\dfrac{1}{\mathrm{Pe}} + \dfrac{2\tau\beta}{\Delta x}\right)(1-\cos\phi)\right]^2 + \left(\dfrac{2\tau\beta}{\Delta x} + \mathrm{C}\right)^2 \sin^2\phi}{\left[\dfrac{2(2+\cos\phi)}{3} + \mathrm{C}\left(\dfrac{1}{\mathrm{Pe}} + \dfrac{2\tau\beta}{\Delta x}\right)(1-\cos\phi)\right]^2 + \left(\dfrac{2\tau\beta}{\Delta x} - \mathrm{C}\right)^2 \sin^2\phi}$$

or equivalently

$$
\begin{aligned}
|g|^2 &= 1 - \frac{\dfrac{8\mathrm{C}}{3}\left(\dfrac{1}{\mathrm{Pe}}+\dfrac{2\tau\beta}{\Delta x}\right)(2+\cos\phi)(1-\cos\phi) - 8\mathrm{C}\dfrac{\tau\beta}{\Delta x}\sin^2\phi}{\left[\dfrac{2(2+\cos\phi)}{3}+\mathrm{C}\left(\dfrac{1}{\mathrm{Pe}}+\dfrac{2\tau\beta}{\Delta x}\right)(1-\cos\phi)\right]^2+\left(\dfrac{2\tau\beta}{\Delta x}-\mathrm{C}\right)^2\sin^2\phi} \\
&= 1 - \frac{\dfrac{8\mathrm{C}(1-\cos\phi)}{3\mathrm{Pe}}\left[2+\dfrac{\tau\beta\mathrm{Pe}}{\Delta x}+\left(1-\dfrac{\tau\beta\mathrm{Pe}}{\Delta x}\right)\cos\phi\right]}{\left[\dfrac{2(2+\cos\phi)}{3}+\mathrm{C}\left(\dfrac{1}{\mathrm{Pe}}+\dfrac{2\tau\beta}{\Delta x}\right)(1-\cos\phi)\right]^2+\left(\dfrac{2\tau\beta}{\Delta x}-\mathrm{C}\right)^2\sin^2\phi}.
\end{aligned}
$$

Numerator of the second term on the right hand side of the last expression is positive by Lemma 5.1. Thus, the amplification factor $|g|$ remains bounded by 1 for any parameter. As a result, the Crank-Nicolson method yields an unconditionally stable approximate solution.

Backward Euler Case For $\theta = 1$, the amplification factor (4.3) with (4.5) is represented as

$$
g = \frac{1}{1+\mathrm{C}\dfrac{\left(\dfrac{1}{\mathrm{Pe}}+\dfrac{2\tau\beta}{\Delta x}\right)(1-\cos\phi)+i\sin\phi}{\dfrac{2+\cos\phi}{3}-i\dfrac{\tau\beta}{\Delta x}\sin\phi}}. \tag{5.13}
$$

Then its magnitude is

$$
|g|^2 = \frac{\left(\dfrac{2+\cos\phi}{3}\right)^2+\left(\dfrac{\tau\beta}{\Delta x}\right)^2\sin^2\phi}{\left[\dfrac{2+\cos\phi}{3}+\mathrm{C}\left(\dfrac{1}{\mathrm{Pe}}+\dfrac{2\tau\beta}{\Delta x}\right)(1-\cos\phi)\right]^2+\left(\mathrm{C}-\dfrac{\tau\beta}{\Delta x}\right)^2\sin^2\phi}.
$$

which can be re-expressed as

$$
\begin{aligned}
|g|^2 &= 1 - \frac{\mathrm{C}^2(1-\cos\phi)\left[\left(\dfrac{1}{\mathrm{Pe}}+\dfrac{2\tau\beta}{\Delta x}\right)^2(1-\cos\phi)+(1+\cos\phi)\right]}{\left[\dfrac{2+\cos\phi}{3}+\mathrm{C}\left(\dfrac{1}{\mathrm{Pe}}+\dfrac{2\tau\beta}{\Delta x}\right)(1-\cos\phi)\right]^2+\left(\mathrm{C}-\dfrac{\tau\beta}{\Delta x}\right)^2\sin^2\phi} \\
&- \frac{\dfrac{2\mathrm{C}(1-\cos\phi)}{3\mathrm{Pe}}\left[2+\dfrac{\tau\beta\mathrm{Pe}}{\Delta x}+\left(1-\dfrac{\tau\beta\mathrm{Pe}}{\Delta x}\right)\cos\phi\right]}{\left[\dfrac{2+\cos\phi}{3}+\mathrm{C}\left(\dfrac{1}{\mathrm{Pe}}+\dfrac{2\tau\beta}{\Delta x}\right)(1-\cos\phi)\right]^2+\left(\mathrm{C}-\dfrac{\tau\beta}{\Delta x}\right)^2\sin^2\phi}.
\end{aligned}
$$

Looking on the right hand side, second term is positive. Again by using Lemma 5.1, third term is also positive. Thus, $|g| \leq 1$ holds true which yields the unconditionally stability of the approximate solution.

We summerize the stability conclusions as following.

Theorem 5.1 *Consider the fully discretized form of the model problem* (2.1)-(2.4) *by using SDFEM in space and generalized trapezoidal rule in time. Let* Δx, β *and* Pe *are given constants and* τ *be a stabilization parameter.*

i. Let $\theta = 0$ *and assume that* $\mathrm{Pe} \leq \frac{3\tau\beta}{\Delta x} + \sqrt{3+\left(\frac{3\tau\beta}{\Delta x}\right)^2}$ *holds true. Then the time stability requirement is given by*

$$\mathrm{C} < \frac{1}{3}\left(\frac{2\beta\tau}{\Delta x}+\frac{1}{\mathrm{Pe}}\right)^{-1}.$$

ii. Let $\theta = 0$ *and assume that* $\mathrm{Pe} > \frac{3\tau\beta}{\Delta x} + \sqrt{3+\left(\frac{3\tau\beta}{\Delta x}\right)^2}$ *holds true. Then the time stability requirement is given by*

$$\mathrm{C} < \frac{1}{\mathrm{Pe}} \quad \Rightarrow \quad \Delta t < \frac{2\kappa}{\beta}.$$

iii. Let $\theta = \frac{1}{2}$ *or* $\theta = 1$. *Then the numerical solution is stable independent of choosing Courant number and Peclet number, i.e. space and time mesh sizes.*

Lastly, note that setting $\tau = 0$, we recover the results given in the Corollary 4.2.

5.2. Accuracy Analysis

This section is devoted to the similar study done on Section 4.2. We are going to measure the temporal behaviours of the approximate solutions, again using the relations we derived on Section 3.2. Unlike the previous section, we now consider $\mathrm{Pe} > 1$ case and therefore a stabilization parameter is required to be set, which is given by (B.2) in Appendix B. Also let us state that, we now do not interested in forward Euler case as the stability condition we derived for convection dominated case (5.11) is too restrictive on choosing time step. Moreover choosing such small time steps does not only cause a computational cost, but also cause a rounding error which affects the accuracy of the numerical solution due to the finite precision of computer arithmetic [2].

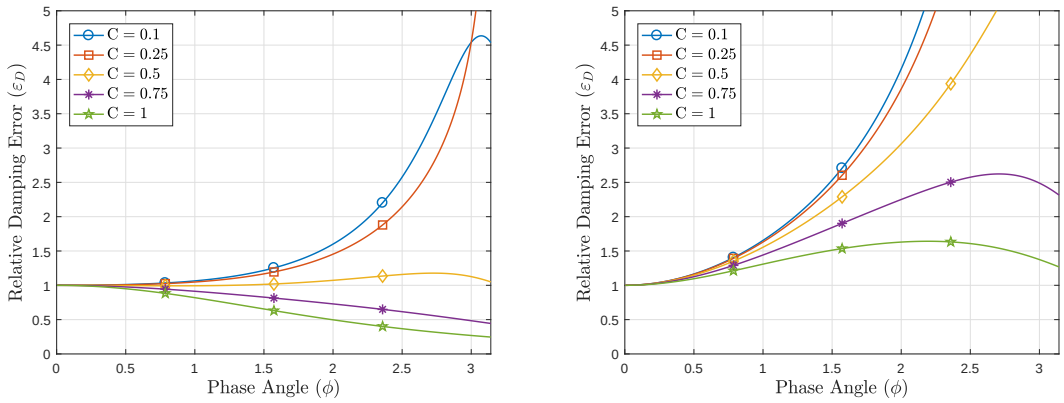
5.2.1. Relative Damping Error

In this part, we give the results of damping properties for Crank-Nicolson and backward Euler methods.

Crank-Nicolson Case Series expansion of numerical damping for $\theta = \frac{1}{2}$ is given by

$$\begin{aligned}\xi^h = \xi \quad &+ \quad \frac{1}{12\Delta x^4}\left(-3\kappa^2\Delta t^2 + (\kappa + \beta^2\tau)\Delta x^2 - 12\kappa\beta^2\tau^2\right)\phi^4 \\ &+ \quad \big(60\kappa(\kappa^2 - 6\kappa\beta^2\tau + 3\beta^4\tau^2)\Delta t^2 + 45\kappa\beta^4\Delta t^4 + 2(\kappa + 5\beta^2\tau)\Delta x^4 \\ &+ \quad 720\kappa\beta^4\tau^4 - 15\beta^2(\kappa + \beta^2\tau)\Delta t^2\Delta x^2 + 60\beta^2(-\kappa + \beta^2\tau)\Delta x^2\tau^2\big)\phi^6 + \mathcal{O}(\phi^8).\end{aligned}$$

The error in damping for various choices of Courant numbers are shown in Figure 5.1. In the range of long waves, i.e. $\phi \to 0$, all choices yields nearly same amount of dampings which are almost same as the optimal level, as in the case of SGFEM in section 4.2.1. This is due to the fact that the stabilization parameter (B.2) vanishes as $\phi \to 0$ and thus the behaviour of the approximate solutions derived by SDFEM and SGFEM are similar. As a result, the artificial diffusion does not shows up for long waves as it should be and consequently, we shall comment that, this is a positive effect of Crank-Nicolson method.



(a) Pe $= 2.5$. (b) Pe $= 10$.

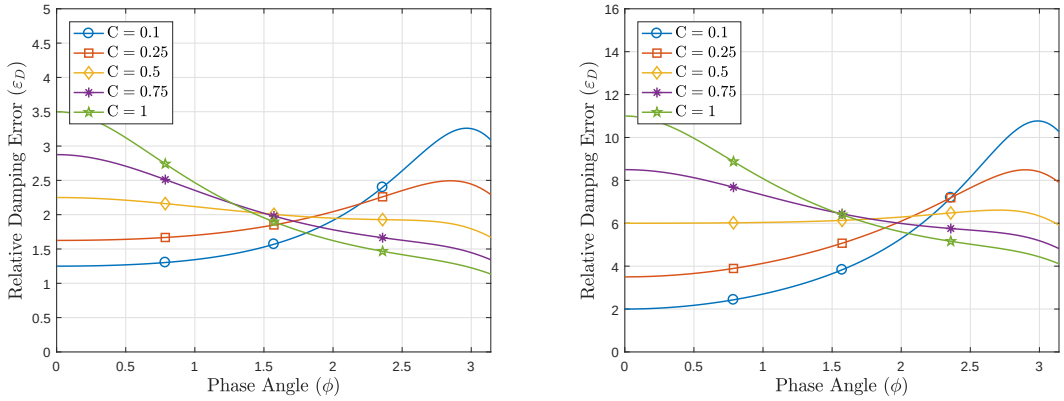
Figure 5.1. Relative damping error for various values of Courant numbers, using SD-FEM in space and Crank-Nicolson method in time.

On the other hand looking at the high frequency range, especially for $\mathrm{Pe} = 10$, all choices yield an additional dissipation on the numerical solution. This is very welcome, especially for the convection dominated problems where high frequency modes are responsible for the spatial oscillations, since the main duty of the artificial diffusion is to damp these spurious oscillations occured because of the standard spatial discretization.

Backward Euler Case Taking $\theta = 1$, series expansion of numerical damping in terms of ϕ is

$$\begin{aligned}\xi^h = \xi \quad &+ \quad \frac{\beta^2\Delta t}{2\Delta x^2}\phi^2 + \frac{1}{12\Delta x^4}\left(-6\kappa(\kappa - 2\beta^2\tau)\Delta t + (\kappa + \beta^2\tau)\Delta x^2 - 12\kappa\beta^2\tau^2\right. \\ &- \quad \left. 12\kappa\beta^2\Delta t^2 - 3\beta^4\Delta t^3\right)\phi^4 + \mathcal{O}(\phi^6).\end{aligned}$$

Plots of relative damping errors for various choices of Courant numbers are displayed in Figure 5.2. An immediate observation here is, in the range of long waves, $\varepsilon_D > 1$ and therefore even the long waves are being damped overmuch. As we increase the Peclet number, i.e. the coefficient of diffusive term becomes smaller, this situation becomes more evident and is not welcome. To catch a reasonable approximation, we require a low Courant number.



(a) Pe = 2.5. (b) Pe = 10.

Figure 5.2. Relative damping error for various values of Courant numbers, using SD-FEM in space and backward Euler method in time.

5.2.2. Relative Phase Speed Error

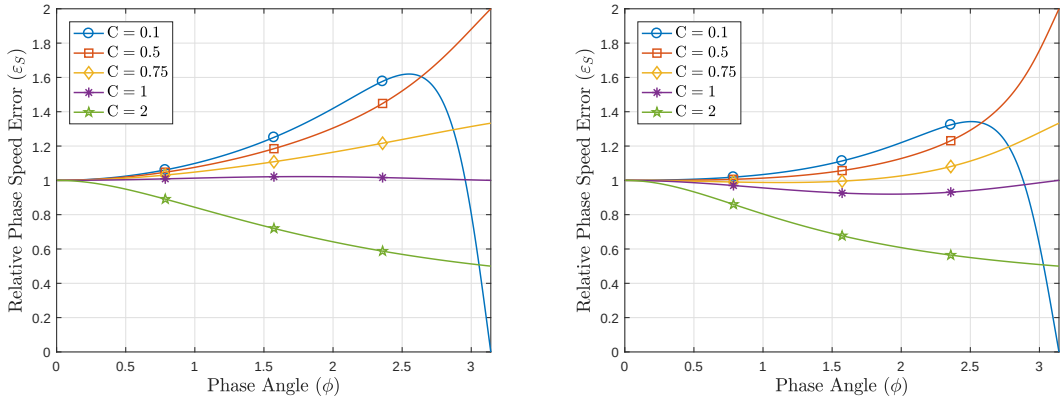
In this part, we give the series expansions and display the numerical results of the phase speed errors obtained by Crank-Nicolson method and backward Euler method.

Crank-Nicolson Case Let $\theta = \frac{1}{2}$. Series expansion of numerical temporal frequency in terms of ϕ is given by

$$\begin{aligned}\omega^h = \omega \quad &+ \quad \frac{1}{12\Delta x^3}\left(-12\kappa\beta\tau + \beta^3\Delta t^2\right)\phi^3 + \frac{1}{720\Delta x^5}\left(180\kappa\beta(-\kappa + \beta^2\tau)\Delta t^2\right. \\ &- \quad \left. 60\beta(\kappa + \beta^2\tau)\Delta x^2\tau + 720\kappa\beta^3\tau^3 - 9\beta^5\Delta t^4 + 4\beta\Delta x^4\right)\phi^5 + \mathcal{O}\left(\phi^7\right)\end{aligned}$$

Looking at Figure 5.3 choice of Courant number should be neihter large nor small to catch a good accuracy on phase speed. Sufficiently large Courant numbers yield a delayed numerical solution, whereas the small ones cause a leading error. The behaviour is not monotonic with respect to Courant number, or Δt, thus there may be expected a intermediate value close to the expected level $\varepsilon_S = 1$.

Remark 5.3 *To annihilate the third order term in the series expansion given above, we let* $\tau = \frac{\Delta x^2}{12\kappa}$ *under the choice* $\mathrm{C} = 1$. *Consequently, the method becomes forth order accurate in space. Related comments on the phase error and corresponding numerical experiments are shown in Section 5.2.3.*

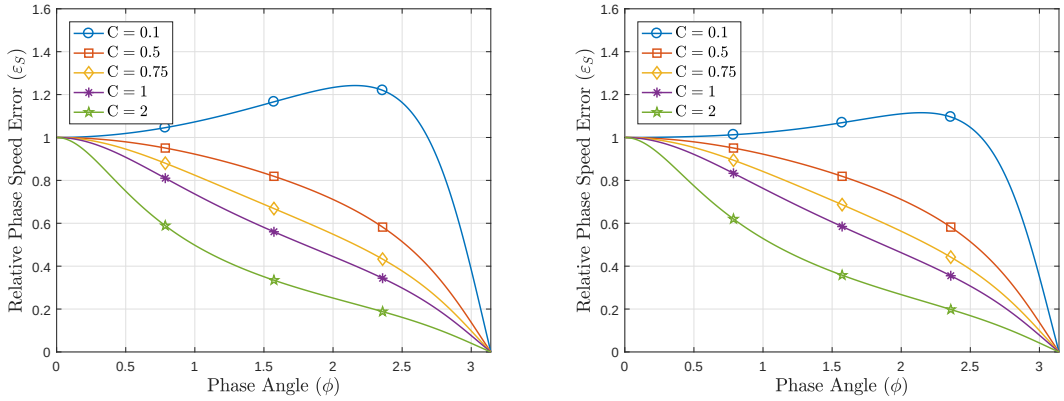


(a) Pe $= 2.5$. (b) Pe $= 10$.

Figure 5.3. Relative phase speed error for various values of Courant numbers using SDFEM in space and Crank-Nicolson method in time.

Backward Euler Case Taking $\theta = 1$, series expansion of numerical temporal frequency in terms of ϕ is given by

$$\begin{aligned}\omega^h = \omega \quad &+ \quad \frac{1}{12\Delta x^3}\left(-3\kappa\beta\Delta t - \beta^3\Delta t^2 + 3\kappa\beta\tau\right)\phi^3 + \frac{1}{180\Delta x^5}\left(180\kappa\beta(-\kappa+\beta^2\tau)\Delta t\tau\right.\\ &+ \quad 180\kappa\beta(\kappa-\beta^2\tau)\Delta t^2 + 15\beta(\kappa-\beta^2\tau)\Delta x^2\tau - 15\beta(\kappa+\beta^2\tau)\Delta t\Delta x^2\\ &+ \quad \left.180\kappa\beta^3\Delta t^3 - 180\kappa\beta^3\tau^3 + 36\beta^5\Delta t^4 - \beta\Delta x^4\right)\phi^5 + \mathcal{O}\left(\phi^7\right).\end{aligned}$$



(a) Pe $= 2.5$. (b) Pe $= 10$.

Figure 5.4. Relative phase speed error for various values of Courant numbers using SDFEM in space and backward Euler method in time.

Observing from the Figure 5.4, all choices except the low ones causes a delay on the numerical solution. We note that there is not any significant change on the phase speed behaviour of the solution as we increase Peclet number. This situation also holds

for higher Peclet numbers.

5.2.3. Some Remarks and Numerical Experiments

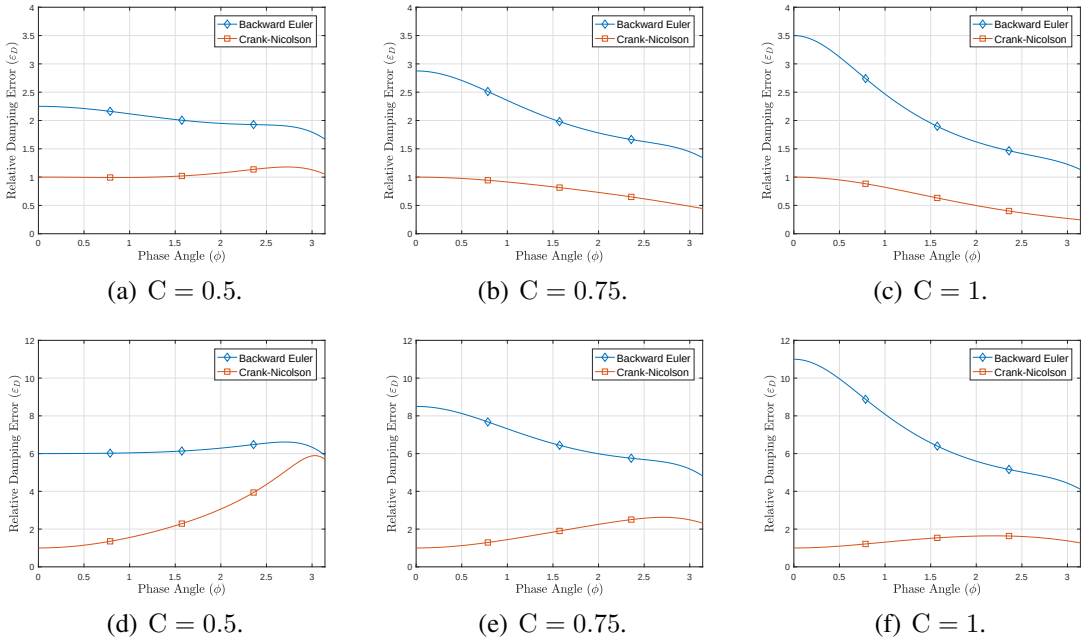
Consistency Each of the method we considered, the constant terms in the series expansions of ξ^h and ω^h are identical to their exact counterparts ξ and ω, as $\Delta x \to 0$ and $\Delta t \to 0$ provided that $\tau \to 0$ is automatically being satisfied according to our choice (B.2). Hence, consistency is recovered.

Accuracy In the Table 5.1 lowest order terms of Δx, Δt and τ are shown for both methods. They have the same accuracy in space: second order and mixed term involving second order for numerical damping and forth order or mixed term involving second order for numerical frequency. Observe for the numerical frequency that, the stabilization term τ occurs together with a second order spatial mesh. Therefore an inappropriate choice of stabilization term may cause a drop of the accuracy of numerical frequency in space. The difference of the errors occurs in time so that, Crank-Nicolson is second order accurate in both numerical damping and numerical frequency, whereas backward Euler is first order accurate.

Time Integrator	ξ^h	ω^h
Crank-Nicolson	$\mathcal{O}\left(\Delta t^2, \Delta x^2, \tau^2, \Delta t^2\tau, \Delta x^2\tau\right)$	$\mathcal{O}\left(\Delta t^2, \Delta x^4, \tau, \Delta t^2\tau, \Delta x^2\tau\right)$
Backward Euler	$\mathcal{O}\left(\Delta t, \Delta x^2, \tau^2, \Delta t\tau, \Delta x^2\tau\right)$	$\mathcal{O}\left(\Delta t, \Delta x^4, \tau, \Delta t\tau, \Delta x^2\tau\right)$

Table 5.1. Accuracy of the numerical damping and numerical frequency of the numerical schemes with different time integrators.

Comparison Between Time Integrators: Damping In the Figure 5.5 plots for relative damping errors of Crank-Nicolson Method and backward Euler method for $\mathrm{Pe} = 2.5$ case on the above and $\mathrm{Pe} = 10$ case on the below are displayed. Immediate observation from the figures is the overdamped behaviour of the backward Euler method in the low frequency range. On the other hand the error of Crank-Nicolson method varies around the optimal level and is better on damping according to backward Euler method in general as we mentioned in Section 5.2.1. Especially, for some particular choices on the parameters yield closer relative errors to the optimal level.



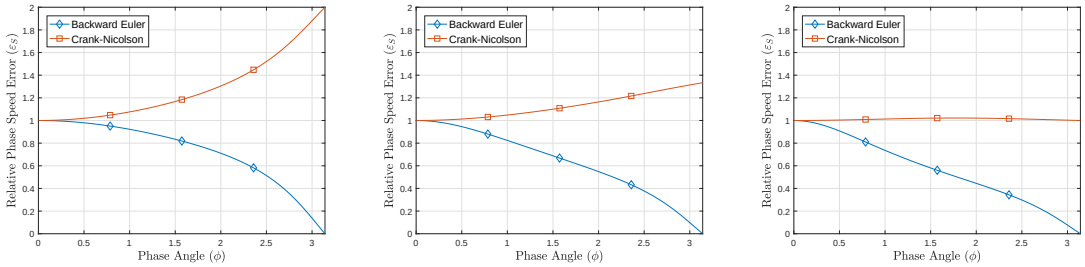
(a) $\mathrm{C} = 0.5$. (b) $\mathrm{C} = 0.75$. (c) $\mathrm{C} = 1$.

(d) $\mathrm{C} = 0.5$. (e) $\mathrm{C} = 0.75$. (f) $\mathrm{C} = 1$.

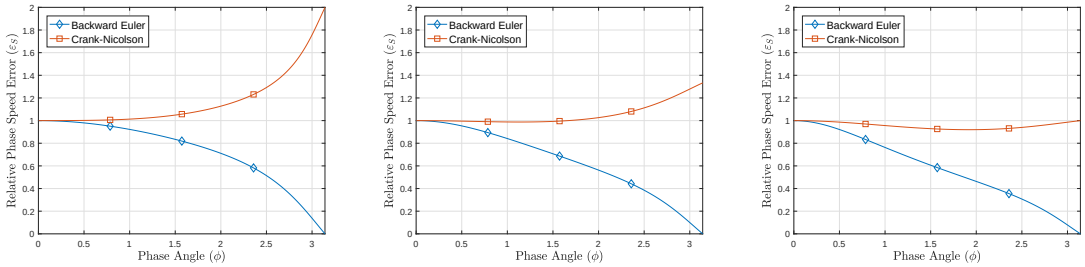
Figure 5.5. Comparison of relative damping errors of the time integrators for convective–diffusive case: $\mathrm{Pe} = 2.5$ on the above and $\mathrm{Pe} = 10$ on the below.

Comparison Between Time Integrators: Phase Speed For a comparison of relative phase speed errors, see Figure 5.6. As we stated previously, all choices causes a delay on the numerical solution obtained by backward Euler method. On the other hand, it is obvious to observe that Crank-Nicolson method is more capable on phase accuracy especially for some specific parameters.

Numerical Experiments for Different Choice of Stabilization Terms In this part, we test the numerical solutions using our stabilization term $\tau = \frac{\Delta x^2}{12\kappa}$, that we give in Remark 5.3, with different choices. These stabilization terms are given in (B.2) which is suggested by Shakib and Hughes [34] and in (B.5) which is suggested by Raymond and Garder [32]. Depending on our damping error and phase speed error comparisons on the previous part, we prefer using Crank-Nicolson time stepping. The problem we present is as following: The diffusion coefficient is taken as $\kappa = 10^{-6}$, the convective field is $\beta = 1$, there is no external source acting on the system and final time is set to be $T = 0.5$. Lastly, refering to [28], the Courant number is set to be $\mathrm{C} = 0.5$ for other stabilization parameters, since the numerical method with this choice has a better ability on reducing the spurious oscillations, whereas choice of Courant number for our stabilization term is set to be $\mathrm{C} = 1$.



(a) C = 0.5. (b) C = 0.75. (c) C = 1.



(d) C = 0.5. (e) C = 0.75. (f) C = 1.

Figure 5.6. Comparison of relative phase speed errors of the time integrators for convective–diffusive case: Pe = 2.5 on the above and Pe = 10 on the below.

Example 5.1 (Cosine Wave) *First we consider a transport process of a smooth function, a cosine profile, given by*

$$u_0(x) = \begin{cases} \dfrac{1+\cos(10\pi(x-0.2))}{2}, & \textit{if } |x-0.2| \le 0.1 \\ 0, & \textit{elsewhere} \end{cases}$$

and shown by the Figure 5.7. See Figure 5.8 for the related numerical results of the

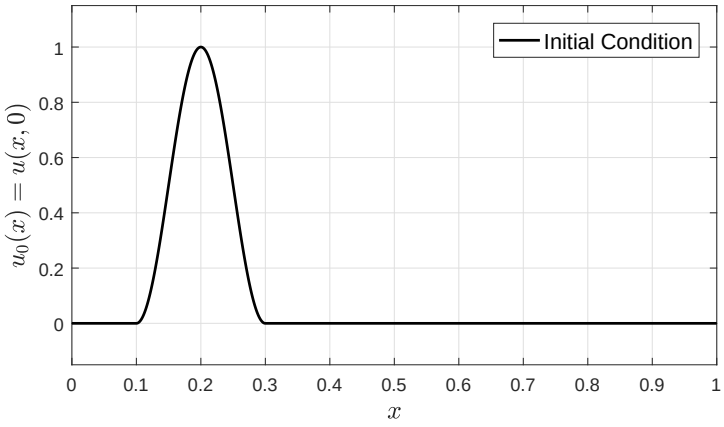
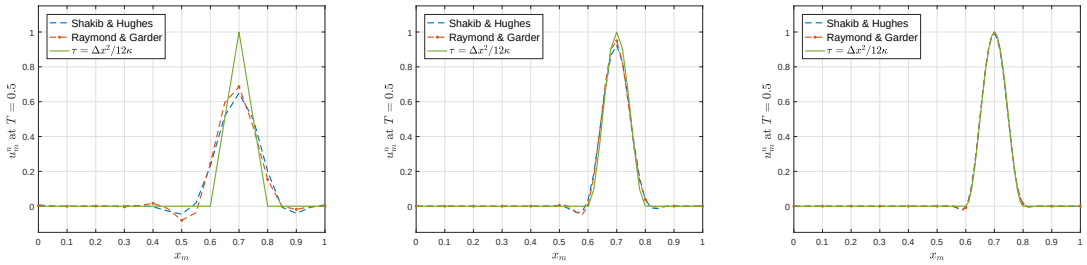


Figure 5.7. The initial condition of Example 5.1.

corresponding stabilization parameters, using various spatial mesh sizes.



(a) $\Delta x = 0.05$. (b) $\Delta x = 0.02$. (c) $\Delta x = 0.01$.

Figure 5.8. Numerical solutions of Example 5.1 for different choice of stabilization parameters.

Example 5.2 (Triangular Wave) *As a second example, we consider a triangular wave as an initial condition given as*

$$u_0(x) = \begin{cases} 1 - |10x - 2|, & \textit{if } |x - 0.2| \leq 0.1 \\ 0, & \textit{elsewhere} \end{cases}.$$

See Figure 5.9 for the plot of the initial condition and Figure 5.10 for the approximate

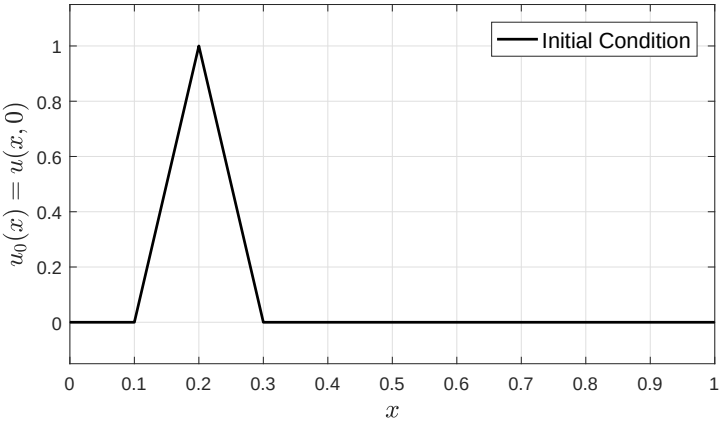
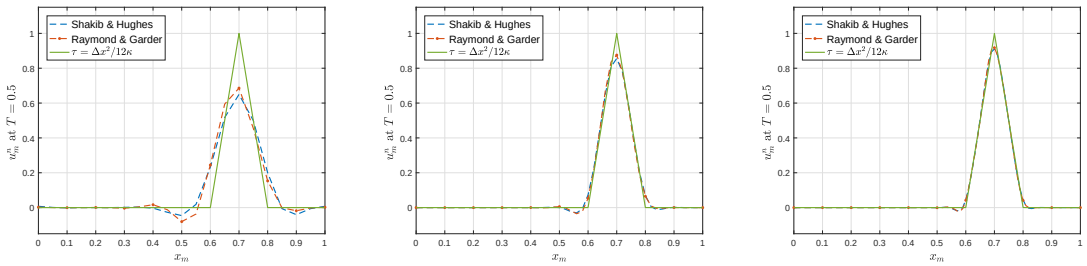


Figure 5.9. The initial condition of Example 5.2.

solutions for the specified stabilization parameters.



(a) $\Delta x = 0.05$. (b) $\Delta x = 0.02$. (c) $\Delta x = 0.01$.

Figure 5.10. Numerical solutions of Example 5.2 for different choice of stabilization parameters.

Example 5.3 (Square Wave) *The third example is devoted to a square pulse as an initial condition,*

$$u_0(x) = \begin{cases} 1, & \textit{if } |x - 0.2| \leq 0.1 \\ 0, & \textit{elsewhere} \end{cases},$$

and shown in Figure 5.11. For the approximate solutions, see Figure 5.12.

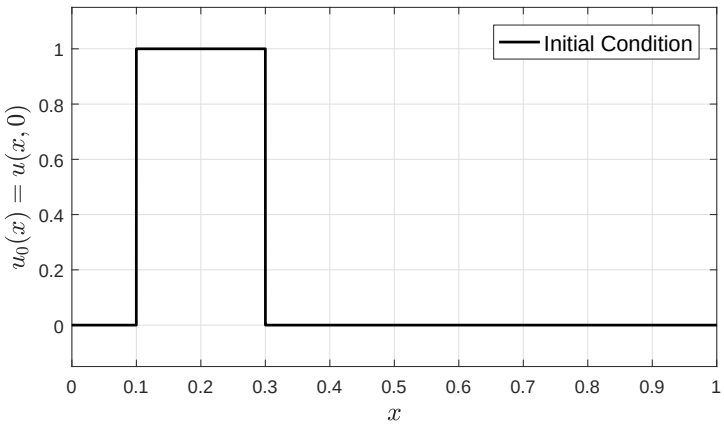
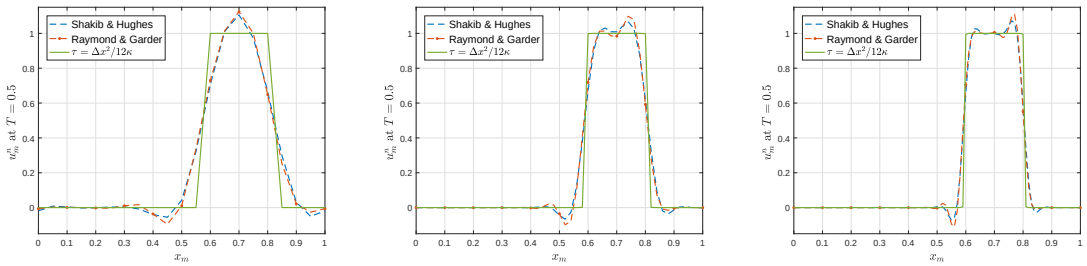


Figure 5.11. The initial condition of Example 5.3.

Immediate observation from the numerical experiments shown in Figures 5.8, 5.10 and 5.12 is that the other stabilization terms causes a dispersion on numerical solutions, whereas the stabilization parameter $\tau = \frac{\Delta x^2}{12\kappa}$ yields nearly a nondispersive numerical solution, i.e. transports the initial profile nearly without causing a defect on it. On the first problem, as we increase the number of elements, both methods become nearly equivalent. On the other hand, dispersion error still remains even on the third problem if fine mesh



(a) $\Delta x = 0.05$. (b) $\Delta x = 0.02$. (c) $\Delta x = 0.01$.

Figure 5.12. Numerical solutions of Example 5.3 for different choice of stabilization parameters.

is used, whereas choice of $\tau = \frac{\Delta x^2}{12\kappa}$ holds the shape of square pulse even on the coarse mesh.

Example 5.4 (Transport of a Wave Packet) *Just as the phase speed error, studying on group speed is essential in order to understand the dispersive behaviour of discrete models. Despite being out of our interest, we want to go one step further to check our stabilization parameter with an initial condition taken as a wave packet, as this choice allows us to make a direct examination and observation of group speed, i.e. the evolution of a group of waves containing several wave numbers. We skip and refer to [38, 39] for technical details. The initial condition here we take is a sine wave modulated by a Gaussian at* $x = 0.2$

$$u_0(x) = \Phi(x, 0)\sin(40\pi x).$$

where

$$\Phi(x, 0) = e^{-160(x-0.2)^2}$$

is defined as the envelope of the wave packet. See Figure 5.13 for the initial condition and corresponding envelope. Note that the group velocity, i.e. the velocity of the envelope, is identical to the phase speed for a constant velocity, β*. So our expectation from an approximate solution is to preserve its shape as far as possible so that the envelope still encloses the wave packet at the final time.*

All parameters are set to be same as the previous exapmples except that the final time is set to be $T = 0.5$ *and space mesh sizes we study now are* $\Delta x = \frac{1}{100}$, $\Delta x = \frac{1}{200}$

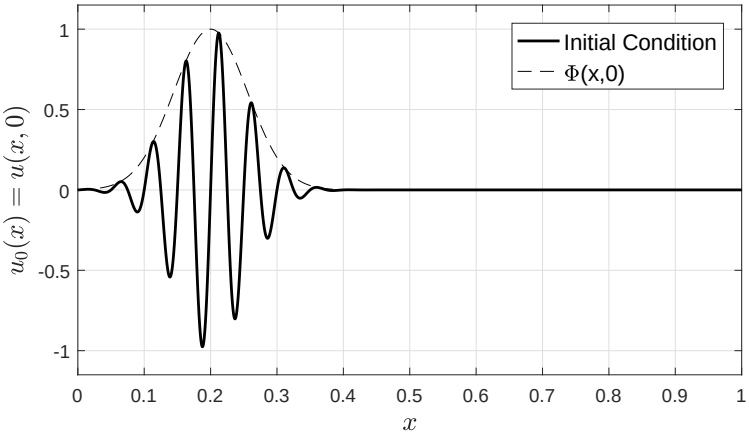
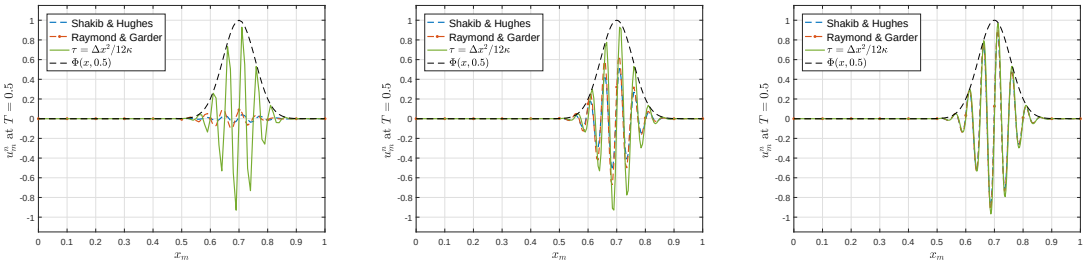


Figure 5.13. The initial condition of Example 5.4 and corresponding envelope.

and $\Delta x = \frac{1}{400}$. *See Figure 5.14 for the related approximate solutions.*



(a) $\Delta x = 0.05$. (b) $\Delta x = 0.02$. (c) $\Delta x = 0.01$.

Figure 5.14. Numerical solutions of Example 5.4 for different choice of stabilization parameters.

As we experienced on the previous examples, our stabilization term that is shown by green yields a nearly non-dispersive transport process, yet the method is coherent on the group velocity whereas, fine mesh is required to be used in order to catch a nice approximation for other parameters.

Numerical Experiments for Phase Speed Error Comparison Using $\tau = \frac{\Delta x^2}{12\kappa}$ The stabilization parameter $\tau = \frac{\Delta x^2}{12\kappa}$ that we give on Remark 5.3 ends up with nice approximate solutions as we illustrate on Examples 5.1-5.4. But this term is derived by regarding to a specific choice of Courant number, $\mathrm{C} = 1$ and now we are in search of to find a reasonable explanation for the sensitiveness of this quantity. For this purpose, we slightly increase and decrease the Courant number, and then give the corresponding relative phase speed errors of the numerical method as well as for $\mathrm{C} = 1$. We use the same problem with same parameters given previously. The spatial mesh is now set to be $\Delta x = 0.01$. Courant numbers are taken as $\mathrm{C} = 0.9$, $\mathrm{C} = 1$, $\mathrm{C} = 1.1$. To shed to the sensitiveness of Courant

number, we give the plots of relative phase speed errors in Figure 5.15. Observe through the plots that choice of $\mathrm{C} = 0.9$ causes a leading error and $\mathrm{C} = 1.1$ causes a lagging error on the numerical solutions. On the other hand the error with $\mathrm{C} = 1$ varies very close to the $\varepsilon_S = 1$ level which is the optimal level as we have a constant convective velocity, i.e. our expectation from the approximate solution is to be nondispersive.

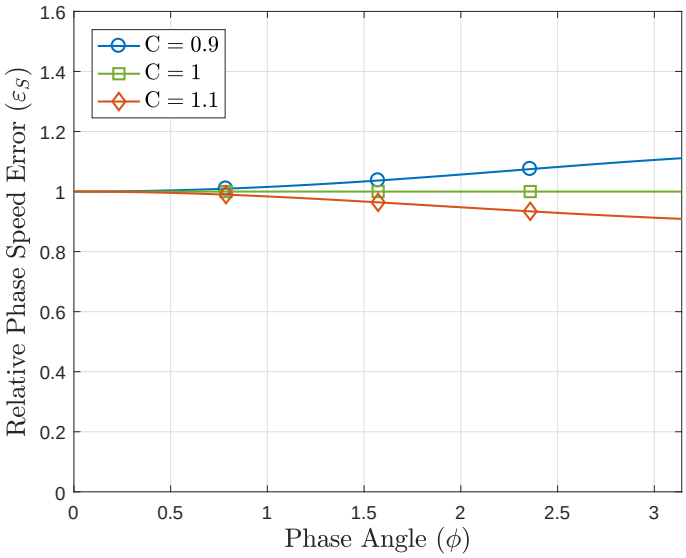


Figure 5.15. Relative phase speed error for stabilized method in space using $\tau = \frac{\Delta x^2}{12\kappa}$ and Crank-Nicolson method in time.

In the figures 5.16-5.18, numerical results that corresponds to the examples 5.1-5.3 are being displayed respectively. By a straight observation at the results obtained for the choice of $\mathrm{C} = 0.9$, spurious oscillations occur on the numerical solution and these stand beyond from where they should be. As we explained in the above, the reason behind this situation is the behaviour of relative phase speed error. Again looking at 5.15, the error is strictly greater than one, i.e. all modes are moving faster than they should move. Moreover the waves that has different wave numbers propagate with different speeds which causes a dispersion error on the numerical solution. As a result, this situation reflects and shows itself on the numerical solution as a spurious oscillations which propagates faster than the convective velocity. The reason of the oscillations occur on the numerical solutions obtained by $\mathrm{C} = 1.1$ is again depend on the same idea, except the fact that now there is a delay on the waves and this may be verified by looking at the Figure 5.15.

On the other hand, looking at the numerical solutions obtained by $\mathrm{C} = 1$, solutions almost preserve the shape of the corresponding initial conditions and this is again due to the fact that the relative phase speed error for this case is nearly optimal, i.e. both low and high frequency modes are propagating almost in same speed.

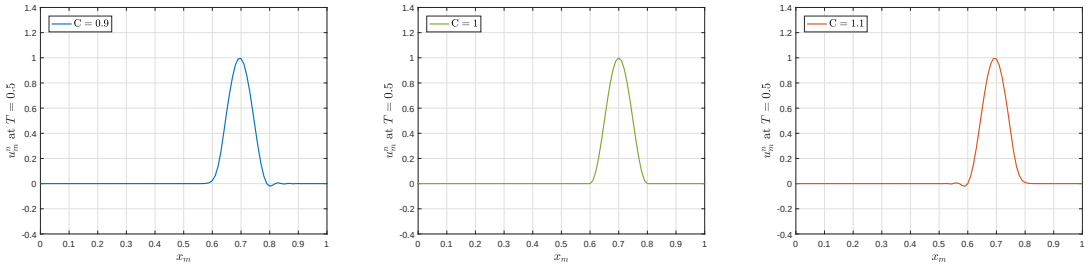


Figure 5.16. Numerical solutions of Example 5.1 using various Courant numbers.

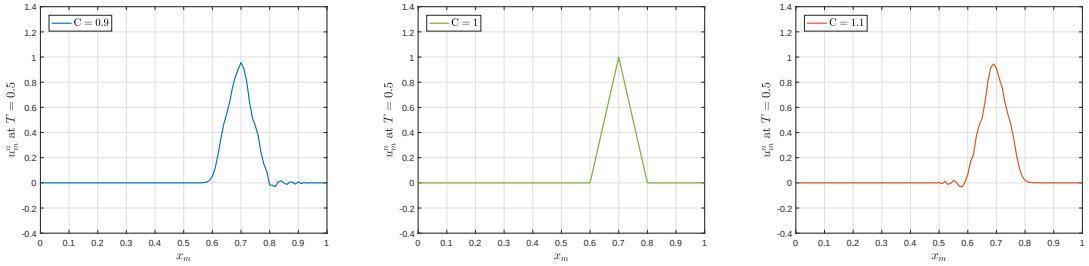


Figure 5.17. Numerical solutions of Example 5.2 using various Courant numbers.

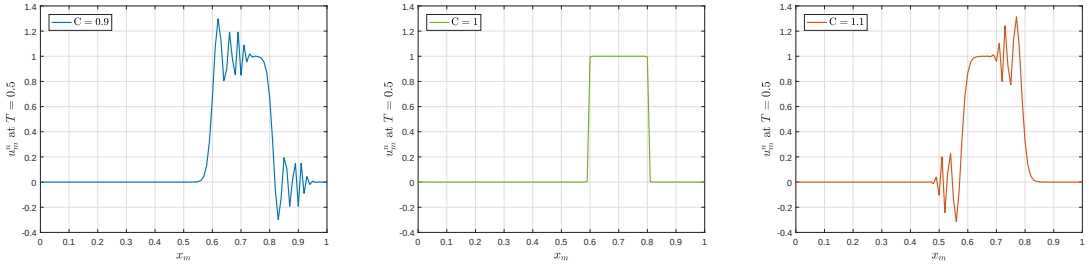


Figure 5.18. Numerical solutions of Example 5.3 using various Courant numbers.

CHAPTER 6

CONCLUSION

We have presented a Fourier stability and accuracy analysis of forward Euler, Crank-Nicolson and backward Euler time integration algorithms to scalar unsteady convection–diffusion equation, where standard and stabilized finite element method is applied for space discretization.

Crank-Nicolson and backward Euler time stepping methods yield unconditional stable numerical results. However, we demonstrated and experienced by numerical experiments for Crank-Nicolson method that, choosing sufficiently large Courant numbers may result an oscillatory numerical solution, eventhough the stable behaviour is still preserved. On the other hand numerical solution obtained by forward Euler method is stable only under a certain condition. For the unsteady convection–diffusion equation with diffusion dominated case, we deduced that the stability condition is

$$\Delta t \leq \frac{\Delta x^2}{6\kappa} \tag{6.1}$$

whereas, when convective process starts dominating the diffusive process, the stability estimate (6.1) changes to

$$\Delta t \leq \frac{2\kappa}{\beta^2}. \tag{6.2}$$

Moreover when we consider using a stabilized method for spatial discretization of convective dominated problem, we derived that the stability condition using forward Euler time stepping is exactly same as (6.2). As the duty of stabilized methods is to cure the undesired and poor numerical approximations that occurs due to the low diffusion coefficient, i.e. $0 < \kappa \ll 1$, we conclude that the stability estimate (6.2) is a strong condition on choosing time step so that, it restricts ourselves to choose a very small time step. This is not only bad because of the computational cost, but also may end up with a rounding error, and as a result, affect the accuracy of the numerical solution that occurs due to the finite precision of computer arithmetic [2].

Along the stability properties, we studied the temporal errors by use of Fourier analysis. This provided us an automatic process for seperating the terms, which are responsible for dissipation and dispersion, i.e. damping and frequency of discrete convective–diffusive operator. Using the relationship between Fourier analysis and classical truncation error analysis as given in [39], we express the asymptotic truncation errors associated with numerical damping and numerical frequency of each time marching method to obtain the order of accuracy with respect to space and time mesh. We obtained some particular choice of parameters that increases the order of accuracy. Then we formulate the relative errors which are expressed by comparing the exact damping and exact frequency with their numerical counterparts. Comparisons of the errors for each time integrator showed that, applying backward Euler method to the heat equation yields better approximate results on damping. However, as we consider the unsteady convection–diffusion model and increase the Peclet number, this situation changes in favour of Crank-Nicolson method. Especially for $\mathrm{Pe} > 1$ case which is in our particular interest, backward Euler operator damps the numerical solution much more than enough even in the low frequency range, whereas the error in damping for the Crank-Nicolson method is in the optimal range as $\phi \to 0$. On the other hand, for high wave numbers, we observed that Crank-Nicolson has an overdmaped behaviour which is very welcome and also expected, since the additional dissipation that occurs due to the artificial diffusion is responsible to damp the spurious high frequency oscillations related with the spatial discretization of convective dominated problems. For the frequency accuracy, the error plots for each Peclet number indicate that, backward Euler method slows down the waves that causes a delay on the numerical solution. On the contrary, for $\mathrm{Pe} \geq 1$ case Crank-Nicolson method yields better numerical results on phase speed than backward Euler. Especially by choosing Courant number within a certain range, we observed through the figures that, the error varies closer to the optimal level.

Then, based on the series expansion of numerical frequency, we derived a stabilization parameter

$$\tau = \frac{\Delta x^2}{12\kappa}$$

regarding to the choice of Courant number $\mathrm{C} = 1$. As we displayed on the Figure 5.15, such choice ensures the relative phase speed error to behave nearly as the optimal level $\varepsilon_S = 1$. In other words, different Fourier modes propagate almost in same speed which prevents the dispersion error. Ofcourse the error is not exactly $\varepsilon_S = 1$, therefore one may

expect even a little dispersion on the approximate solution. But in our opinion, these small oscillations that occurs due to the little dispersion is damped by the additional dissipative effect of the artificial diffusion. Thus, we may comment that, the artificial diffusion idea that comes through the discretization of singularly perturbed problems of elliptic differential equations, is responsible not only to damp the spurious oscillations caused by the space discretization, but also responsible to damp the oscillations that occur due to the dispersion error of the numerical method. In fact, numerical experiments verifies these comments on the nearly optimal phase error result so that, observing from the given examples, our choice on stabilization term prevents the distortion on the shape of the initial condition.

REFERENCES

[1] M. I. Asensio, B. Ayuso, and G. Sangalli. Coupling stabilized finite element methods with finite difference time integration for advection–diffusion–reaction problems. *Computer Methods in Applied Mechanics and Engineering*, 196(35):3475–3491, 2007.

[2] K. Atkinson, W. Han, and D. E. Stewart. *Numerical solution of ordinary differential equations*, volume 81. John Wiley & Sons, 2009.

[3] O. Baysal. *Stabilized finite element methods for time dependent convection–diffusion equation*. PhD thesis, Izmir Institute of Technology, Department of Mathematics, 2012.

[4] P. B. Bochev, M. D. Gunzburger, and J. N. Shadid. Stability of the supg finite element method for transient advection–diffusion problems. *ComputerMethods in Applied Mechanics and Engineering*, 193(23):2301–2323, 2004.

[5] F. Brezzi, M.-O. Bristeau, L. P. Franca, M. Mallet, and G. Rogé. A relationship between stabilized finite element methods and the galerkin method with bubble functions. *Computer Methods in Applied Mechanics and Engineering*, 96(1):117–129, 1992.

[6] F. Brezzi, D. Marini, and A. Russo. Applications of the pseudo residual–free bubbles to the stabilization of convection-diffusion problems. *Computer Methods in Applied Mechanics and Engineering*, 166(1):51–63, 1998.

[7] F. Brezzi and A. Russo. Choosing bubbles for advection–diffusion problems. *Mathematical Models and Methods in Applied Sciences*, 4(04):571–587, 1994.

[8] A. N. Brooks and T. J. R. Hughes. Streamline upwind/petrov–galerkin formulations for convection dominated flows with particular emphasis on the incompressible navier-stokes equations. *Computer Methods in Applied Mechanics and Engineering*, 32(1):199–259, 1982.

[9] E. Burman. Consistent supg-method for transient transport problems: Stability and convergence. *Computer Methods in Applied Mechanics and Engineering*,

199(17):1114–1123, 2010.

[10] M. A. Christon, M. J. Martinez, and T. E. Voth. Generalized fourier analyses of the advection-diffusion equation - part i: one-dimensional domains. *International Journal for Numerical Methods in Fluids*, 45(8):839–887, 2004.

[11] R. Codina. Comparison of some finite element methods for solving the diffusion-convection-reaction equation. *Computer Methods in Applied Mechanics and Engineering*, 156(1):185–210, 1998.

[12] J. de Frutos and J. Novo. Bubble stabilization of linear finite element methods for non-linear evolutionary convection–diffusion equations. *Computer Methods in Applied Mechanics and Engineering*, 197(45):3988–3999, 2008.

[13] W. Dettmer and D. Perić. An analysis of the time integration algorithms for the finite element solutions of incompressible navier–stokes equations based on a stabilised formulation. *Computer Methods in Applied Mechanics and Engineering*, 192(9):1177–1226, 2003.

[14] L. P. Franca, S. L. Frey, and T. J. R. Hughes. Stabilized finite element methods: I. application to the advective–diffusive model. *Computer Methods in Applied Mechanics and Engineering*, 95(2):253–276, 1992.

[15] K. O. Friedrich and W. R. Wasow. Singular perturbations of non-linear oscillations. 1946.

[16] P. M. Gresho and R. L. Sani. Incompressible flow and the finite element method. volume 1: Advection–diffusion and isothermal laminar flow. 1998.

[17] I. Harari. Stability of semidiscrete formulations for parabolic problems at small time steps. *Computer Methods in Applied Mechanics and Engineering*, 193(15):1491–1516, 2004.

[18] I. Harari and G. Hauke. Semidiscrete formulations for transient transport at small time steps. *International Journal for Numerical Methods in Fluids*, 54(6-8):731–743, 2007.

[19] G. Hauke and M. Doweidar. Fourier analysis of semi-discrete and space–time stabilized methods for the advective–diffusive–reactive equation: I. supg. *Computer Methods in Applied Mechanics and Engineering*, 194(1):45–81, 2005.

[20] G. Hauke and M. Doweidar. Fourier analysis of semi-discrete and space–time stabilized methods for the advective–diffusive–reactive equation: Ii. sgs. *Computer Methods in Applied Mechanics and Engineering*, 194(6):691–725, 2005.

[21] G. Hauke and M. Doweidar. Fourier analysis of semi-discrete and space–time stabilized methods for the advective–diffusive–reactive equation: Iii. sgs/gsgs. *Computer Methods in Applied Mechanics and Engineering*, 195(44):6158–6176, 2006.

[22] C. Hirsch. *Numerical Computation of Internal and External Flows. Volume 1: Fundamentals of Numerical Discretization*. John Wiley & Sons, 8 2001.

[23] A. Huerta, B. Roig, and J. Donea. Time-accurate solution of stabilized convection–diffusion–reaction equations: Ii-accuracy analysis and examples. *Communications in Numerical Methods in Engineering*, 18(8):575–584, 2002.

[24] T. J. Hughes, L. P. Franca, and G. M. Hulbert. A new finite element formulation for computational fluid dynamics: Viii. the galerkin/least–squares method for advective–diffusive equations. *Computer Methods in Applied Mechanics and Engineering*, 73(2):173–189, 1989.

[25] T. J. R. Hughes. *The Finite Element Method: Linear Static and Dynamic Finite Element Analysis.* Dover Publications, 2000.

[26] T. J. R. Hughes and A. Brooks. A multidimensional upwind scheme with no crosswind diffusion. *Finite Element Methods for Convection Dominated Flows*, 34:19–35, 1979.

[27] M. K. Kadalbajoo and V. Gupta. A brief survey on numerical methods for solving singularly perturbed problems. *Applied Mathematics and Computation*, 217(8):3641–3716, 2010.

[28] A. Kaya. Pseudo residual–free bubble functions for the stabilization of convection-diffusion-reaction problems. Master's thesis, Izmir Institute of Technology, De-

partment of Mathematics, 2012.

[29] G. Lube and D. Weiss. Stabilized finite element methods for singularly perturbed parabolic problems. *Applied Numerical Mathematics*, 17(4):431–459, 1995.

[30] A. I. Nesliturk. A stabilizing subgrid for convection–diffusion problem. *Mathematical Models and Methods in Applied Sciences*, 16(02):211–231, 2006.

[31] A. Quarteroni and A. Valli. *Numerical approximation of partial differential equations*, volume 23. Springer Science & Business Media, 2008.

[32] W. H. Raymond and A. Garder. Selective damping in a galerkin method for solving wave problems with variable grids. *Monthly Weather Review*, 104(12):1583–1590, 1976.

[33] R. D. Richtmyer and K. W. Morton. Difference methods for initial–value problems. 1967.

[34] F. Shakib and T. J. R. Hughes. A new finite element formulation for computational fluid dynamics: Ix. fourier analysis of space-time galerkin/least–squares algorithms. *Computer Methods in Applied Mechanics and Engineering*, 87(1):35–58, 1991.

[35] J. C. Strikwerda. *Finite difference schemes and partial differential equations*. Siam, 2004.

[36] M. Stynes. Steady-state convection–diffusion problems. *Acta Numerica*, 14:445–508, 2005.

[37] J. W. Thomas. *Numerical partial differential equations: finite difference methods*, volume 22. Springer Science & Business Media, 2013.

[38] L. N. Trefethen. Group velocity in finite difference schemes. *SIAM review*, 24(2):113–136, 1982.

[39] R. Vichnevetsky and J. B. Bowles. *Fourier analysis of numerical approximations of hyperbolic equations*, volume 5. Siam, 1982.

[40] T. E. Voth, M. J. Martinez, and M. A. Christon. Generalized fourier analyses of the advection–diffusion equation - part ii: two-dimensional domains. *International Journal for Numerical Methods in Fluids*, 45(8):889–920, 2004.

APPENDIX A

DESCRIPTION OF THE STABILIZED METHODS

As we gave briefly in Section 1.1.1, solving elliptic boundary value problems using SGFEM yields poor numerical approximations through the sharp changes of the analytical solution. To troubleshoot this situation, SGFEM must be improved in some way. Again as we mentioned in the same section, there are two well-known ideas to achive spatial stability, which are based on adding the residual of the differential operator to the standard Galerkin formulation and enriching the standard finite element space with bubble type functions. Here, we briefly introduce these methods and show how the homogeneous parts of the corresponding discrete schemes have equivalent structure for our model problem. For this purpose, let us first introduce a steady state convective–diffusive model,

$$-\kappa u'' + \beta u' = f, \quad \text{in } \Omega$$

where $\Omega = (0, 1)$ as usual with periodic boundary conditions and f is set to be same as in our model problem. Define a test space as

$$\mathcal{V}^h = \left\{ v^h \in \mathcal{H}^1_{\text{per}}(\Omega) : v^h|_{\Omega^e} \in \mathbb{P}_1(\Omega^e), \forall \Omega^e \in \mathcal{T}_h, \right\}.$$

Then the coresponding Galerkin method reads: find $u^h \in \mathcal{V}^h$ such that

$$\mathcal{A}_0(u^h, v^h) = (f, v^h), \quad \forall v^h \in \mathcal{V}^h \tag{A.1}$$

where $\mathcal{A}_0$ is the usual bilinear form given by (2.8).

A.1. Streamline Upwind/Petrov–Galerkin Method

Modifying the standard finite element basis functions with respect to the flow direction is equivalent to modifying the test space as

$$\mathcal{W}^h = \left\{ w^h : w^h = v^h + \tau\beta v_x^h, v^h \in \mathcal{V}^h \right\}.$$

This modification graphically shown in Figure A.1. Analytically, this modification shows

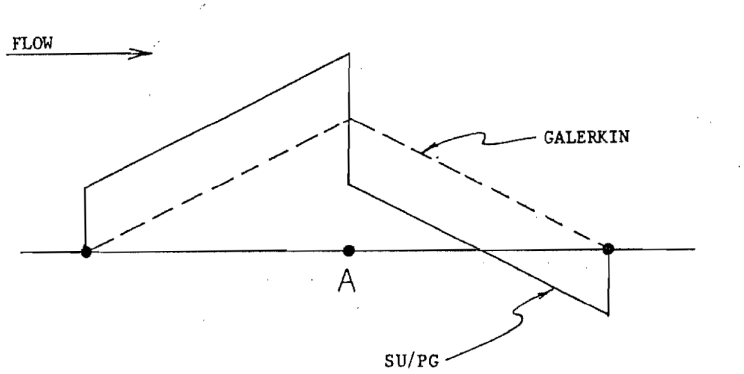


Figure A.1. Comparison of SUPG and SGFEM basis functions for a node $x = A$. Figure is taken from [8].

itself in the standard Galerkin formulation as an addition of a residual of the differential equation,

$$\sum_{\Omega^e \in \mathcal{T}_h} \left(\mathcal{L}u^h - f, \tau\beta v_x^h \right)_{\Omega^e}. \tag{A.2}$$

weighted by a parameter τ, called stabilization parameter, that is to be determined or suggested depending on the varios metrics such as truncation error, coercivity properties, order of convergence, temporal errors and so on. The subscript Ω^e here refers that the integration is considered in elementwise. Under this improvement, the SUPG discretization reads: find $u^h \in \mathcal{V}^h$ such that

$$\mathcal{A}_0(u^h, v^h) + \sum_{\Omega^e \in \mathcal{T}_h} \left(\mathcal{L}u^h - f, \tau\beta v_x^h \right)_{\Omega^e} = (f, v^h), \quad \forall v^h \in \mathcal{V}^h. \tag{A.3}$$

Skipping the details, we can re-express the last expression as

$$(\kappa + \tau\beta^2)\left(u_x^h, v_x^h\right) + \left(\beta u_x^h, v^h\right) = \left(f, v^h + \tau\beta v_x^h\right), \quad \forall v^h \in \mathcal{V}^h$$

which shows that such improvement shows itself as an artificial diffusion, that is responsible to damp the spurious oscillations.

A.2. Galerkin/Least–Squares Method

Similar to the SUPG method, this method is a residual based method except that now the additional residual, which is again weighted by the stabilization parameter, is given by

$$\sum_{\Omega^e \in \mathcal{T}_h} \left(\mathcal{L}u^h - f, \tau\mathcal{L}v^h\right)_{\Omega^e}.$$

Considering the fact that the basis functions are linear and differentiable on each element, one can re-express this term as (A.2). Consequently, discrete structure of the homogeneous parts of the methods become equivalent.

A.3. Unusual Stabilized Finite Element Method

This method, which is also known as adjoint stabilized Galerkin method, is based on adding again a residual to the standard Galerkin formulation, but now given by

$$-\sum_{\Omega^e \in \mathcal{T}_h} \left(\mathcal{L}u^h - f, \tau\mathcal{L}^* v^h\right)_{\Omega^e}.$$

where $\mathcal{L}^*$ is the adjoint operator of $\mathcal{L}$. Similarly, one can conclude that the structure of the scheme is same as the previous ones.

A.4. Residual–Free Bubble Method

For the purpose of enlarging the finite element space, define $B_K := \mathcal{H}_0^1(\Omega^e)$ for each $\Omega^e \in \mathcal{T}^h$ and enlarge the finite element space as $\mathcal{V}^h \oplus \mathcal{V}^B$ where

$$\mathcal{V}^B := \oplus_{\mathcal{T}_h} B_{\Omega^e} = \text{span}\left\{b^e\right\}.$$

Here the bubbles belong to the first and last element are supposed to be equivalent due to the periodic boundary conditions. Using this enlargement, we can decompose any function that belongs to $\mathcal{V}^h \oplus \mathcal{V}^B$, as a sum of linear part and a bubble part given by

$$v^h + v^B = v^h + \sum_{\Omega^e \in \mathcal{T}^h} c^e b^e, \quad c^e \in \mathbb{R}.$$

Now we will solve the steady state problem (A.1) on $\mathcal{V}^h \oplus \mathcal{V}^B$. So the RFB discretization of the problem (A.1) by splitting itself reads: Find $u^h + u^B \in \mathcal{V}^h \oplus \mathcal{V}^B$ such that

$$\begin{aligned} \mathcal{A}(u^h + u^B, v^h) &= (f, v^h), \quad \forall v^h \in \mathcal{V}^h && \text{(A.4)} \\ \mathcal{A}(u^h + u^B, b^e)_{\Omega^e} &= (f, b^e)_{\Omega^e} && \text{(A.5)} \end{aligned}$$

where the subscript Ω^e indicates that the integrals are restricted to the corresponding element. To obtain the numerical solution, one needs to derive c^e's, i.e. bubbles first. This could be done by solving (A.5) in elementwise. By following [7], the result is given by

$$c^e = (f - \beta v_x^h)_{\Omega^e} \frac{(1, b^e)_{\Omega^e}}{\kappa(b_x^h, b_x^e)_{\Omega^e}}.$$

Now substituting c^e back into the main equation (A.4), one can derive the problem: find $u^h \in \mathcal{V}^h$ such that

$$\mathcal{A}_0(u^h, v^h) + \sum_{\Omega^e \in \mathcal{T}_h} (\beta u_x^h - f, \tau^e \beta v_x^h)_{\Omega^e} = (f, v^h). \quad \text{(A.6)}$$

where

$$\tau^e \;=\; \frac{1}{|\Omega^e|}\frac{(b^e,1)^2_{\Omega^e}}{\kappa(b^h_x,b^e_x)_{\Omega^e}}. \tag{A.7}$$

For further details, we refer to [5, 7]. Now observe that (A.6) together with (A.7) has exactly the same structure with (A.3). Thus, once we obtain an explicit form for (A.7), we conclude that the RFB method for convective–diffusive model is equivalent to a stabilized method using a stabilization term with streamline diffusion type.

Now extending this steady-state structure to an unsteady problem, we therefore have the following semi-discrete form

$$(\partial_t u^h, v^h) + \mathcal{A}_0(u^h, v^h) + \sum_{\Omega^e\in\mathcal{T}_h}\left((\partial_t+\mathcal{L})u^h - f, \tau\beta v^h_x\right)_{\Omega^e} \;=\; (f, v^h), \quad \forall v^h \in \mathcal{V}^h.$$

APPENDIX B

CHOICE OF STABILIZATION PARAMETERS

Here we introduce some stabilization terms that are recommended and being used in the literature.

In [8] stabilization term is suggested as

$$\tau_1 = \frac{\Delta x}{2\beta}\alpha(\mathrm{Pe}), \quad \alpha(\mathrm{Pe}) = \coth \mathrm{Pe} - \frac{1}{\mathrm{Pe}}. \tag{B.1}$$

where, this choice makes the finite difference counterpart of the approximation for steady-state problem nodally exact.

In [34], authors tested the temporal behaviours of the numerical solution of unsteady convection–diffusion equation by using

$$\begin{aligned} \tau_2 &= \left(\left(\frac{2\beta}{\Delta x} \right)^2 + 9 \left(\frac{4\kappa}{\Delta x^2} \right)^2 \right)^{-1/2} \\ &= \frac{\Delta x}{2\beta} \left(1 + \left(\frac{3}{\mathrm{Pe}} \right)^2 \right)^{-1/2}. \end{aligned} \tag{B.2}$$

In [11], the author gives some requirements on choosing stabilization term and then deduces as

$$\begin{aligned} \tau_3 &= \left(\frac{4\kappa}{\Delta x^2} + \frac{2\beta}{\Delta x} \right)^{-1} \\ &= \frac{\Delta x}{2\beta} \left(1 + \frac{1}{\mathrm{Pe}} \right)^{-1} \end{aligned} \tag{B.3}$$

Pseudo Residual–Free Bubbles To determine the stabilization term that corresponds to the RFB method given by (A.7), bubble functions are needed to be determined. Technically, this requires the steady-state problem to be solved in each element [7] which is same as solving the original problem. Therefore, as stated in [6], this problem cannot be computable in an easy way. Again in the same reference, the authors construct a cheap

way to compute the approximate solution, called Pseudo Residual–Free Bubbles (PRFB). By using PRFB method, the author in [30] expressed the stabilization parameter as

$$\tau_4 = \begin{cases} \frac{\Delta x}{2\beta} - \frac{\kappa}{\beta^2}, & \text{if } \kappa \leq \frac{\Delta x \beta}{6} \\ \frac{\Delta x^2}{18\kappa}, & \text{if } \kappa > \frac{\Delta x \beta}{6} \end{cases}.$$

This can be re-expressed in terms of Peclet number as

$$\tau_4 = \begin{cases} \frac{\Delta x}{2\beta}\left(1 - \frac{1}{\mathrm{Pe}}\right), & \text{if } \mathrm{Pe} \geq 3 \\ \frac{\Delta x}{2\beta}\frac{2\mathrm{Pe}}{9}, & \text{if } \mathrm{Pe} < 3 \end{cases}. \tag{B.4}$$

In the Figure B.1 we give the plots of the stabilization terms given by (B.1)-(B.4) with respect to the change of Peclet number. Observe that all of the stabilization terms approach to the same level as we increase the Peclet number. So our expectation is that, there does not occur much difference in the temporal errors. However, according to our numerical experiments, numerical solution obtained by using τ_2 results in a better dispersive properties. So in our numerical tests, the stabilization term we use is (B.2).

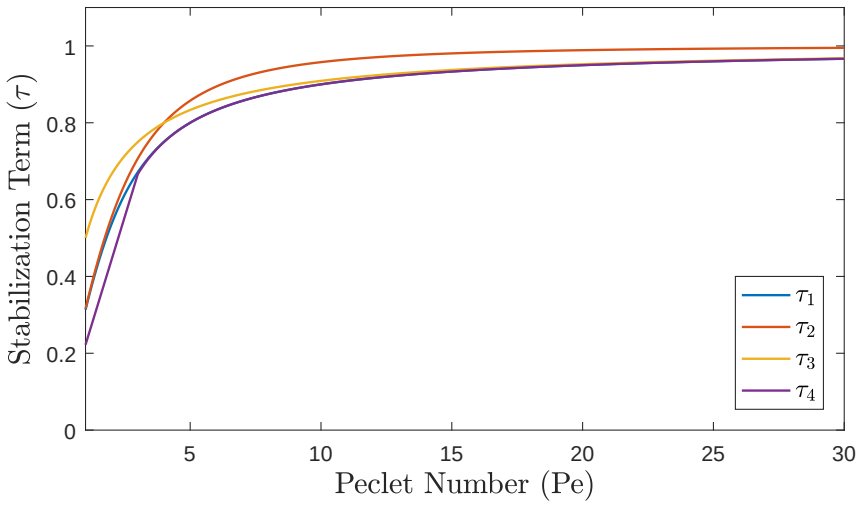


Figure B.1. Plots of the various stabilization terms.

A Choice on Maximixing Frequency Accuracy in Space In the limit case of $\Delta t \to 0$, i.e. the semi-discrete form, series expansion of numerical frequency for the generalized

trapezoidal rule is given by

$$\begin{aligned}\omega^h = \omega \quad &- \quad \frac{\kappa\beta\tau}{\Delta x^3}\phi^3 + \frac{1}{180\Delta x^5}\left(-15\kappa\beta\Delta x^2\tau + 180\kappa\beta^3\tau^3 - 15\beta^3\Delta x^2\tau^2\right. \\ &+ \quad \left.\beta\Delta x^4\right)\phi^5 + \mathcal{O}(\phi^7).\end{aligned}$$

Now considering the pure convection problem, i.e. letting $\kappa \to 0$, the third order term vanishes. In order to annihilate the fifth order term, the stabilization parameter is supposed to be chosen as

$$\tau_5 = \frac{\Delta x}{\sqrt{15}\beta} \tag{B.5}$$

which was shown by Raymond and Garder [32]. This choice of stabilization term increases the accuracy of the phase speed error of the method up to sixth order. We refer to [10, 16] for further details.