
Why classical finite difference approximations fail for a singularly perturbed system of convection-diffusion equations

A Thesis presented to the Department of Pure and Applied Mathematics
African University of Science and Technology, Abuja
In partial fulfilment of the requirements for

MASTER DEGREE IN PURE AND APPLIED MATHEMATICS

By

Aroh Innocent Tagbo

Supervised by

Prof J.K. Djoko (University of Pretoria)



Knowledge is Freedom

African University of Science and Technology

www.aust.edu.ng

P.M.B 681, Garki, Abuja F.C.T

Nigeria.

June 7, 2016

Certification

WHY CLASSICAL FINITE DIFFERENCE APPROXIMATIONS FAIL FOR
A SINGULARLY PERTURBED SYSTEM OF CONVECTION-DIFFUSION
EQUATIONS

A THESIS APPROVED

BY

DEPARTMENT OF PURE AND APPLIED MATHEMATICS

RECOMMENDED:

.....
Supervisor: PROF. J.K.Djoko

.....
H.O.D, Pure and Applied Mathematics: PROF. C.E.
CHIDUME(FTWAS,FAS,FNMS)

APPROVED:

.....
Vice President (Academic)

.....
Date

Abstract

Why Classical Finite Difference Approximations fail for a singularly perturbed system of convection-diffusion equations

Msc candidate: Aroh Innocent Tagbo

We consider classical Finite Difference Scheme for a system of singularly perturbed convection-diffusion equations coupled in their reactive terms, we prove that the classical SFD scheme is not a robust technique for solving such problem with singularities. First we prove that the discrete operator satisfies a stability property in the l_2 -norm which is not uniform with respect to the perturbation parameters, as the solution blows up when the perturbation parameters goes to zero. An error analysis also shows that the solution of the SFD is not uniformly convergent in the discrete l_2 -norm with respect to the perturbation parameters, i.e., the convergence is very poor for a sufficiently small choice of the perturbation parameters. Finally we present numerical results that confirm our theoretical findings.

Declaration

I, Aroh Innocent Tagbo. know the meaning of plagiarism and I declare that all of the work in the document, save for that which is properly acknowledged is my own.

Dedication

I dedicate this work to God almighty and to my family, especially my mum Mrs M.N Aroh(Dibu-ugwu-nwanyi), i love you.

Acknowledgements

First and foremost, a big thanks to almighty God who made all things possible, for giving me the grace, the will, emotional and psychological stability to succeed in AUST.

A special thanks to my supervisor and friend Prof Djoko Jules Kamdem for his patience, kindness, dedication, hard work and constructive criticism as regards to the success of this work, thank you very much sir. To my director and my teacher Prof C.E Chidume(FAS), thank you very much for the motivation and push towards learning and studying properly, your fatherly love and care, keep on the good work you are doing in Mathematics Institute and AUST at large, you are a good man.

I am grateful to all my lecturers in AUST, thank you all for teaching me good mathematics. And my senior friends Dr Usman Bello, Dr Buchi, Sister Amarakristi and Dr Patrice thank you all. I also appreciate the assistance of Korande Ngufor and Monday Ogudu during the simulation stage of my work.

My parents Chief and Mrs C.C Aroh, my sweet siblings Uzzy, Ikem and Ifesinachi, i love you all. My friends Dera, Uche, Mifa, Ifeekwue Odigwe, Joy, Edeh Calister and my colleagues Ukaa, Mark, MarkJoe, Kenny, Monday, Ogonaya, Igwe and Obinna, especially Nnyaba Ukamaka and Mark Uzochukwu, thank you for impacting my life positively. And to the special one, Nneka Ifeacho, thank you for your love, support and tolerance over the years.

Finally i am also grateful to the management of AUST, most especially my good friends Miss Amaka Udigwe and Miss Bolade Igbagbo and also to Nelson Mandela Institute for the wonderful opportunity to study in this great citadel of learning.

Contents

1	Introduction	1
1.1	Motivation	1
1.2	Formulation of the problem	2
2	Numerical Schemes	5
2.1	Finite difference approximation	5
3	Consistency-Stability	13
3.1	consistency analysis	13
3.2	Stability analysis	17
4	Convergence	25
5	Numerical simulations and future works	29
5.1	Numerical examples	29
5.2	Conclusion and Future Research	40

List of Figures

5.1	solution profile of (2.8)	36
5.2	solution profile of (2.8)	36
5.3	solution profile of (2.8)	37
5.4	solution profile of (2.8)	37
5.5	solution profile of (2.8)	38

5.6	solution profile of (2.8)	39
-----	-------------------------------------	----

List of Tables

5.1	Rate of convergence	32
5.2	Rate of convergence	33
5.3	Rate of convergence	33
5.4	Rate of convergence	34
5.5	Rate of convergence	34
5.6	Rate of convergence	35
5.7	Rate of convergence	35

CHAPTER 1

Introduction

The contents of this thesis fall within the general area of numerical methods for PDE, an area which has attracted the attention of prominent mathematicians due to its diverse applications in numerous fields of sciences

1.1 Motivation

Imagine a river - a river flowing strongly and smoothly, liquid pollution pours into the water at a certain point, which shape does the pollution stain form on the surface of the river? Two physical processes operate here: the pollution diffuses slowly through the water, but the dominant mechanism is the swift movement of the river which rapidly convects the pollution along a one - dimensional curve on the surface; diffusion gradually spreads that curve. When convection and diffusion are both present in a linear differential equation and convection dominates, we have a convection - diffusion problem. The simplest mathematical model of a convection - diffusion problem is a two point - point boundary value problem of the form,

$$\begin{cases} -\varepsilon u''(x) + a(x)u'(x) + b(x)u(x) = f(x) & 0 < x < 1 \\ u(0) = u(1) = 0, \end{cases} \quad (1.1)$$

where ε is a small positive parameter and $a(x)$, $b(x)$, $f(x)$ are some given functions. The term u'' corresponds to the diffusion and its coefficient ε is small, while the expression u' represents convection. Finally u and $f(x)$ play the roles of a source and driving term respectively.

Now, having known that the solutions of ODE's lives in $C[a, b]$, consider the problem

$$-\varepsilon u''(x) + u'(x) = 1 \quad \text{for } 0 < x < 1 \quad \dots \quad (1.2)$$

with $u(0) = u(1) = 0$ and $0 < \varepsilon < 1$.

suppose that we formally set $\varepsilon = 0$, here we get

$$\begin{cases} u'(x) = 1 & \text{for } 0 < x < 1 \quad \dots \\ u(0) = u(1) = 0. \end{cases} \quad (1.3)$$

The problem (1.3) has no solution in $C[0, 1]$ so we infer than when ε is near zero the solution of (1.3) is badly behaved. Problems like (1.3) are differential equations that depend on small positive parameter ε and whose solutions (or their derivatives) approach a discontinuous limit as ε approaches zero. We say that such problems are singularly perturbed where we regard ε as a perturbation parameter. In more technical terms , one cannot represent the solution of a singularly perturbed differential equation as an asymptotic expansion in the powers of ε . Moreover not every differential equation be it ODE or PDE can be solved analytically and singular Perturbations arise in several branches of engineering and applied mathematics, including fluid dynamics, so in investigating numerical skills for tackling such problems leads to the main objective of this thesis.

1.2 Formulation of the problem

Classical Finite Difference Scheme is one of the most frequently used method for numerical solution for both ordinary and partial differential equation. But on the contrary, in this work we study why classical SFD scheme fails to approximate a coupled system of singularly perturbed convection-diffusion. The governing equations of the problem are given by

$$\begin{cases} -\varepsilon u_{xx} - a_1(x)u_x + b_{11}(x)u + b_{12}(x)v = f(x), \\ -\mu v_{xx} - a_2(x)v_x + b_{21}(x)u + b_{22}(x)v = g(x), \\ u(0) = u(1) = v(0) = v(1) = 0. \end{cases} \quad (1.4)$$

where (u, v) is the solution of (1.4) above. In (1.4), we assume that

$$0 < \varepsilon \leq \mu < 1, \quad (1.5)$$

and

$$a_k(x) \geq \alpha > 0, b_{kk}(x) \geq 0, k = 1, 2. \quad (1.6)$$

The convection-diffusion equation (1.4) are considered as linearised version of the Navier-Stokes equation, they constitute an element of interest in the area of fluid dynamics and hydro dynamics. Although the equation (1.4) may not be applied directly to real applications, it is an important stage in investigation of many practical applications. There is a lot of work in literature dealing with the numerical solution of a single equation of (1.4) but systems of equations appear relatively rare.

In chapter 2, we introduced the notion of the classical SFD approximation accompanied with some basic definitions and results. Then we formulated the classical SFD for (1.4) and showed its consistency with the continuous problem (1.4), we gave an elegant proof of the existence and uniqueness of the solution of the discrete operator.

In chapter 3 and chapter 4, stability analysis and error analysis were both investigated respectively, and both turned out not to be uniform with respect to the perturbation parameters (ε, μ) . For the stability analysis, the solution blows up as (ε, μ) goes to zero, and there will no convergence at all as (ε, μ) goes to zero. Basically this is why the classical SFD fail to approximate (1.4), it couldn't take care of (ε, μ) and they found them selves in damaging positions.

In chapter 5, we wrote a computer program and simulate the method for several cases of interest and the numerical investigations corroborated with our theoretical findings.

CHAPTER 2

Numerical Schemes

The goal is to formulate the finite difference approximation associated with (1.4). We also discuss the existence and uniqueness of the approximation formulated.

2.1 Finite difference approximation

2.1.1 Basic definitions and results

We introduce approximation of derivatives, and introduce a mesh of domain. first we recall basics definitions and results for the definition of finite difference approximation.

The definitions and results we recall in this paragraph are standard and can be found in [2, 3, 4].

The fundamental idea of almost any numerical method for solving equations of the form (1.4) is to approximate the differential equation by a system of algebraic equations. The system of algebraic equations is set up in a way as to produce a “good solution” of the differential equation. The way to generating such a system is to replace the derivatives in the equation by finite differences. In fact the basic idea of any finite difference scheme is based on the very familiar definition of the derivative of a smooth function:

$$u'(x) = \lim_{h \rightarrow 0} \frac{u(x+h) - u(x)}{h}.$$

This clearly indicates that in order to get good approximations, h must be very small. Hence to compute good solutions, we need h approaching zero. This in

turn implies that we need to solve large system of algebraic equations. The solution is computed at the endpoints and determined by the differential equation in the interior solution domain.

The first step in deriving a finite difference approximation of (1.4) is to partition the unit domain $[0, 1]$ into a finite number of sub-intervals. We introduce the grids points $\{x_i\}_{i=0}^{n+1}$ given by $x_i = ih$, where $n \geq 1$ is an integer and the spacing h is the distance between two consecutive points, that is $h = x_{i+1} - x_i$. As we assume that the sub-division is uniform, then $h = 1/(n + 1)$. Typically n will be large so that h tends towards zero.

The second step, in the construction of finite difference scheme is the discrete operator. The following operator will be used throughout the work

$$D^+ \phi(x) = \frac{1}{h}(\phi(x + h) - \phi(x)), \quad (2.1)$$

$$D^- \phi(x) = \frac{1}{h}(\phi(x) - \phi(x - h)), \quad (2.2)$$

The operator operator D^+ is called “forward”, and D^- backward. From the definition of the derivative, it is clear that if ϕ is twice differentiable, then $D^+ \phi(x)$, and $D^- \phi(x)$ tend towards ϕ_x when $h \rightarrow 0$.

Definition 2.1. *An operator \tilde{A}_h is said to be a consistent approximation of A with respect to the discretization parameter h if*

$$\tilde{A}_h \phi - A\phi \rightarrow 0 \quad \text{if } h \rightarrow 0.$$

If furthermore the error $|\tilde{A}_h \phi - A\phi|$ committed is bounded, up to a multiplicative constant, by h^p , the approximation is said to be consistent of order p .

We easily verify that

Lemma 2.2. *1- If ϕ is twice differentiable on $[x, x + h]$, then $D^+ \phi$ is a consistent approximation of order one of $\partial_x \phi$.*

2- If ϕ is twice differentiable on $[x - h, x]$, then $D^- \phi$ is a consistent approximation of order one of $\partial_x \phi$.

Next, we claim that

Lemma 2.3. *If ϕ is four times continuously differentiable on $[x - h, x + h]$, then*

$$D^- D^+ \phi(x) = \frac{1}{h^2} [\phi(x + h) - 2\phi(x) + \phi(x - h)]$$

is a consistent, second order, approximation to ϕ_{xx}

Proof. . From Taylor's expansion,

$$\phi(x+h) = \phi(x) + h\phi'(x) + \frac{h}{2}\phi^{(2)}(x) + \frac{h^3}{6}\phi^{(3)}(x) + \frac{h^4}{24}\phi^{(4)}(x_1), \quad (2.3)$$

with $x_1 \in (x, x+h)$. Likewise

$$\phi(x-h) = \phi(x) - h\phi'(x) + \frac{h}{2}\phi^{(2)}(x) - \frac{h^3}{6}\phi^{(3)}(x) + \frac{h^4}{24}\phi^{(4)}(x_2), \quad (2.4)$$

with $x_2 \in (x-h, x)$ Adding (2.1) to (2.2) gives

$$\frac{\phi(x+h) - 2\phi(x) + \phi(x-h)}{h^2} - \phi^{(2)}(x) = \frac{h^2}{24} [\phi^{(4)}(x_1) + \phi^{(4)}(x_2)]. \quad (2.5)$$

To continue, we need the following result

Lemma 2.4. (*Mean Value for sum*). Let g_i , be functions such that $g_i \geq 0$, $\sum_i g_i = 1$. Let f a function defined on $(\min x_i, \max x_i)$. Then there is $z \in (\min x_i, \max x_i)$ such that for $y_i \in (\min x_i, \max x_i)$

$$\sum_i g_i(y_i)f(y_i) = f(z)$$

In (2.5), we take $g_1 = g_2 = 1/2$, and application of Lemma 2.4 leads to the existence of $z \in (x-h, x+h)$ such that

$$\frac{\phi(x+h) - 2\phi(x) + \phi(x-h)}{h^2} - \phi^{(2)}(x) = \frac{h^2}{12}\phi^{(4)}(z). \quad (2.6)$$

We also recall the following result

Lemma 2.5. If $f : [0, 1] \rightarrow \mathbb{R}$ is continuous, then

- (a) f is bounded
- (b) f attains its minimum at some point in $[0, 1]$
- (c) f attains its maximum at some point in $[0, 1]$

We then deduce that

$$|D^+D^-\phi(x) - \phi^{(2)}(x)| \leq \frac{h^2}{12} \sup_{x-h < z < x+h} |\phi^{(4)}(z)|.$$

Hence the approximation is consistent of order 2. It is noted that since $\phi^{(4)}$ is continuous, the maximum is well defined due to Lemma 2.5.

Remark 2.6. *By combining in different ways the operators D^+, D^- , it is possible to construct approximations to a partial derivative of any order; some are obviously better than others, meaning that their consistency order is higher.*

With the discrete operators D^+D^- , and D^-D^+ , we can define a finite difference approximation of (1.4).

From now and the rest of this work we let

$$\begin{aligned} u_i &\approx u(x_i), \\ v_i &\approx v(x_i), \\ f_i &= f(x_i), \quad g_i = g(x_i). \end{aligned} \tag{2.7}$$

With the consistent operators D^+, D^-D^+ introduced and (2.7), the finite difference scheme approximating (1.4) is defined as follows:

$$\begin{cases} \text{Find } (u_i, v_i) \text{ such that for all } i \in \{1, 2, 3, \dots, N\}, \\ -\epsilon D^-D^+u_i - a_1(x_i)D^+u_i + b_{11}(x_i)u_i + b_{12}(x_i)v_i = f_i, \\ -\mu D^-D^+v_i - a_2(x_i)D^+v_i + b_{21}(x_i)u_i + b_{22}(x_i)v_i = g_i, \\ u_0 = u_{N+1} = v_0 = v_{N+1} = 0, \end{cases} \tag{2.8}$$

as mentioned above, $h = 1/(N + 1)$ is the step size, $x_i = ih$ is the endpoint.

Regarding the discrete scheme (2.8), two fundamental arise:

- (a) does the discrete problem admit a unique solution?
- (b) is the method convergent, that is does it hold that

$$\|(u(x_i), v(x_i)) - (u_i, v_i)\| \rightarrow 0, \quad \text{when } h \rightarrow 0 ?$$

The first question, will be answered in the next paragraph while the second question will constitute is the object of the next Chapter, moreover, we want to see the influence of ϵ, μ on the convergence.

2.1.2 Existence and Uniqueness of solution of (2.8)

The problem (2.8) is a system of $2N \times 2N$ linear equations, which can be written in compact form as

$$\mathcal{B}\psi = F, \quad (2.9)$$

where \mathcal{B} is a $2N \times 2N$ matrix, $\psi^T = (u_1, v_1, u_2, v_2, \dots, u_N, v_N)$ is the unknown vector of size $2N \times 1$ and $F = (f_1, g_1, f_2, g_2, \dots, f_N, g_N)$ is given vector of size $2N \times 1$. Hence, it suffice to show that \mathcal{B} is invertible to claim the existence and uniqueness of solution of (2.8). From elementary linear algebra, since we are in finite dimension the following conditions are equivalent.

Lemma 2.7. (a) *the kernel of \mathcal{B} contains only the zero vector.*

(b) *\mathcal{B} is surjective*

(c) *\mathcal{B} is a bijection*

We claim that

theorem 2.8. *The kernel of \mathcal{B} contains only the zero vector. Hence the linear system of equations (2.8) is uniquely solvable.*

Proof. For $i = 1, 2, 3, \dots, N$, let (u_i, v_i) such that

$$\begin{cases} -\epsilon D^- D^+ u_i - a_1(x_i) D^+ u_i + b_{11}(x_i) u_i + b_{12}(x_i) v_i = 0, \\ -\mu D^- D^+ v_i - a_2(x_i) D^+ v_i + b_{21}(x_i) u_i + b_{22}(x_i) v_i = 0, \\ u_0 = u_{N+1} = v_0 = v_{N+1} = 0. \end{cases} \quad (2.10)$$

We would like to show that $u_i = v_i = 0$ for $i = 1, 2, 3, \dots, N$.

Before embarking on the proof, we first recall some facts.

• **discrete Holder inequality**

$$\sum_{i=1}^N x_i y_i \leq \left[\sum_{i=1}^N x_i^2 \right]^{1/2} \left[\sum_{i=1}^N y_i^2 \right]^{1/2} \quad (2.11)$$

• **Young inequality**

$$ab \leq \frac{\epsilon}{2} a^2 + \frac{1}{2\epsilon} b^2 \quad \text{for all } a, b \in \mathbb{R} \quad \text{and } \epsilon \in \mathbb{R}. \quad (2.12)$$

• **discrete Integration by parts.** Let $U = (u_i)_{0,\dots,N+1}$ and $V = (v_i)_{0,\dots,N+1}$ be two sequences. Then

$$\sum_{i=1}^N v_i D^- u_i + \sum_{i=1}^{N+1} u_i D^- v_i = \frac{1}{h} (u_{N+1} v_{N+1} - u_1 v_0)$$

If in particular $v_0 = v_{N+1} = 0$, then

$$\sum_{i=1}^N v_i D^- u_i = - \sum_{i=1}^{N+1} u_i D^- v_i \quad (2.13)$$

• **discrete Poincare's inequality.** Let $U = (u_i)_{i=0,\dots,N+1}$, with $u_0 = u_{N+1} = 0$. Then there is a positive constant \mathbf{c} independent of h such that

$$\mathbf{c} \sum_i u_i^2 \leq \sum_i |D^+ u_i|^2 = \sum_i |D^- u_i|^2. \quad (2.14)$$

• For a vector $\mathbf{u} = (u_i)_{i=1}^n$, we define the following norms

$$\|\mathbf{u}\|^2 = \sum_{i=1}^n u_i^2, \quad \|\mathbf{u}\|_\infty = \max_{1 \leq i \leq n} |u_i|,$$

from which we deduce that

$$\frac{1}{n} \|\mathbf{u}\|^2 \leq \|\mathbf{u}\|_\infty^2 \leq \|\mathbf{u}\|^2. \quad (2.15)$$

• **parallelogram identity.** Let $x = (x_i)$ and $y = (y_i)$ two sequences of vectors indexed for $i = 0, 1, 2, \dots, N$. Then

$$2 \sum_{i=0}^N (x_i - y_i) x_i = \|x\|^2 - \|y\|^2 + \|x - y\|^2. \quad (2.16)$$

We multiply the first equation of (2.10) by u_i , the second equation of (2.10) by v_i , sum the resulting equation over i and use (2.13). We obtain

$$\begin{aligned} \varepsilon \sum_i (D^+ u_i)^2 + \sum_i b_{11}(x_i) u_i^2 + \sum_i b_{12}(x_i) u_i v_i &= \sum_i a_1(x_i) u_i D^+ u_i, \\ \mu \sum_i (D^+ v_i)^2 + \sum_i b_{21}(x_i) u_i v_i + \sum_i b_{22}(x_i) v_i^2 &= \sum_i a_2(x_i) v_i D^+ v_i, \end{aligned}$$

Adding these relations gives

$$\begin{aligned} & \sum_i (\varepsilon(D^+u_i)^2 + \mu(D^+v_i)^2) + \sum_i (b_{11}(x_i)u_i^2 + b_{22}(x_i)v_i^2) \\ &= \sum_i (a_1(x_i)u_iD^+u_i + a_2(x_i)v_iD^+v_i) - \sum_i (b_{21}(x_i)u_iv_i + b_{12}(x_i)u_iv_i). \end{aligned} \quad (2.17)$$

We treat the right hand side of (2.17) with the help of (2.11) and (2.12) as follows

$$\begin{aligned} & \sum_i (a_1(x_i)u_iD^+u_i + a_2(x_i)v_iD^+v_i) - \sum_i (b_{21}(x_i)u_iv_i + b_{12}(x_i)u_iv_i) \\ &\leq \sum_i |a_1(x_i)||u_i||D^+u_i| + |a_2(x_i)||v_i||D^+v_i| + \sum_i (|b_{21}(x_i)| + |b_{12}(x_i)|) |u_i||v_i| \\ &\leq \sum_i \frac{1}{2\varepsilon} |a_1(x_i)|^2 |u_i|^2 + \frac{\varepsilon}{2} |D^+u_i|^2 + \sum_i \frac{1}{2\mu} |a_2(x_i)|^2 |v_i|^2 + \sum_i \frac{\mu}{2} |D^+v_i|^2 \\ &\quad + \sum_i (|b_{21}(x_i)| + |b_{12}(x_i)|) \left(\frac{\theta}{2} |u_i|^2 + \frac{1}{2\theta} |v_i|^2 \right), \end{aligned} \quad (2.18)$$

where θ is positive real number that must be made precise later. So, returning to (2.17) with (2.18) we have

$$\begin{aligned} & \sum_i \left(\frac{\varepsilon}{2} (D^+u_i)^2 + \frac{\mu}{2} (D^+v_i)^2 \right) + \sum_i \left(b_{11}(x_i) - \frac{1}{2\varepsilon} |a_1(x_i)|^2 - (|b_{21}(x_i)| + |b_{12}(x_i)|) \frac{\theta}{2} \right) u_i^2 \\ &+ \sum_i \left(b_{22}(x_i) - \frac{1}{2\mu} |a_2(x_i)|^2 - (|b_{21}(x_i)| + |b_{12}(x_i)|) \frac{1}{2\theta} \right) v_i^2 \leq 0, \end{aligned}$$

which by (2.14) gives

$$\begin{aligned} & \sum_i \left(c \frac{\varepsilon}{2} + b_{11}(x_i) - \frac{1}{2\varepsilon} |a_1(x_i)|^2 - (|b_{21}(x_i)| + |b_{12}(x_i)|) \frac{\theta}{2} \right) u_i^2 \\ &+ \sum_i \left(c \frac{\mu}{2} + b_{22}(x_i) - \frac{1}{2\mu} |a_2(x_i)|^2 - (|b_{21}(x_i)| + |b_{12}(x_i)|) \frac{1}{2\theta} \right) v_i^2 \leq 0. \end{aligned} \quad (2.19)$$

We then take θ such that the coefficient in front of u_i^2 and v_i^2 are positive. With that in mind, it is clear that for (2.19) to happen we need $u_i = 0 = v_i$. Thus the operator \mathcal{B} is injective, and the problem (2.8) has a unique solution.

CHAPTER 3

Consistency-Stability

3.1 consistency analysis

Our aim here is to discuss the consistency of the numerical scheme introduced in Chapter 2.

In general, any given partial differential equation, including its boundary/or initial conditions, can be written as an abstract operator equation

$$Au = f \tag{3.1}$$

with appropriately chosen functions spaces U, V , a mapping $A : U \rightarrow V$, and $f \in V$. The related discrete problem can be stated analogously as

$$A_h u_h = f_h \tag{3.2}$$

with $A_h : U_h \rightarrow V_h$, $f_h \in V_h$, and discrete spaces U_h, V_h .

Definition 3.1. *Let \tilde{u} be the exact solution at the grid points, and $\|\cdot\|_{V_h}$ the norm on V_h . Then the value $\|A_h \tilde{u} - f_h\|_{V_h}$ is called the consistency error relative to $u \in U$.*

It should be noted that the remainder term in Taylor's expansion gives a way of estimating the consistency error, provided that the solution u of (3.1) is smooth enough and f_h forms an appropriate discretization of f .

Definition 3.2. *A discretization of (3.1) is consistent if*

$$\|A_h \tilde{u} - f_h\|_{V_h} \rightarrow 0, \quad h \rightarrow 0.$$

If in addition the consistency error satisfies the more precise estimate

$$\|A_h \tilde{u} - f_h\|_{V_h} \leq ch^p$$

where c is a positive constant, then the discretization is said to be consistent of order p .

Remark 3.3. 2.1 is the definition of consistency for an operator and definition 3.2 is the definition of consistency for a discretization.

3.1.1 Consistency of finite difference scheme

A finite difference approximation is considered consistent if by reducing the mesh, the truncation error terms could be made to approach zero, in that case, the solution to the difference equation would approach the true solution to the continuous problem.

In this section, we show that our SFD scheme is consistent with our continuous problem.

theorem 3.4. Let (u, v) be the solution of (1.4), assumed to be four times continuously differentiable in $[0, 1]$. Then the following holds

$$\|\xi\|_\infty \leq \max(\varepsilon, \mu) \mathbf{c}_1 h^2 + \max_{0 \leq i \leq N} (|a_1(x_i)|, |a_2(x_i)|) \mathbf{c}_2 h,$$

where $\mathbf{c}_1, \mathbf{c}_2$ are positive constants independent of h, μ , and ε . $\xi = (\xi_1, \xi_2)$ is defined as follows

$$\begin{aligned} \xi_1 &= -\varepsilon D^+ D^- u(x_i) - a_1(x_i) D^+ u(x_i) + b_{11}(x_i) u(x_i) + b_{12}(x_i) v(x_i) - f(x_i), \\ \xi_2 &= -\mu D^+ D^- v(x_i) - a_2(x_i) D^+ v(x_i) + b_{21}(x_i) u(x_i) + b_{22}(x_i) v(x_i) - g(x_i). \end{aligned}$$

Proof. Since u is four times continuously differentiable, using Taylor's expansion as in Lemma 2.3, there are $x_1 \in (x_i, x_{i+1})$, $x_2 \in (x_{i-1}, x_i)$ and $x_3 \in (x_i, x_{i+1})$ such that

$$D^+ u(x_i) = u_x(x_i) + \frac{h}{2} u^{(2)}(x_1), \tag{3.3}$$

$$D^+ D^- u(x_i) = u_{xx}(x_i) + \frac{h^2}{24} [u^{(4)}(x_2) + u^{(4)}(x_3)].$$

Similarly, there are $y_1 \in (x_i, x_{i+1})$, $y_2 \in (x_{i-1}, x_i)$ and $y_3 \in (x_i, x_{i+1})$ such that

$$D^+v(x_i) = v_x(x_i) + \frac{h}{2}v^{(2)}(y_1), \quad (3.4)$$

$$D^+D^-v(x_i) = v_{xx}(x_i) + \frac{h^2}{24} [v^{(4)}(y_2) + v^{(4)}(y_3)].$$

With (3.3) and (3.4), ξ_1 and ξ_2 becomes respectively

$$\begin{aligned} \xi_1 = & -\varepsilon \left(u_{xx}(x_i) + \frac{h^2}{24} [u^{(4)}(x_2) + u^{(4)}(x_3)] \right) - a_1(x_i) \left(u_x(x_i) + \frac{h}{2}u^{(2)}(x_1) \right) \\ & + b_{11}(x_i)u(x_i) + b_{12}(x_i)v(x_i) - f(x_i), \end{aligned} \quad (3.5)$$

$$\begin{aligned} \xi_2 = & -\mu \left(v_{xx}(x_i) + \frac{h^2}{24} [v^{(4)}(y_2) + v^{(4)}(y_3)] \right) - a_2(x_i) \left(v_x(x_i) + \frac{h}{2}v^{(2)}(y_1) \right) \\ & + b_{21}(x_i)u(x_i) + b_{22}(x_i)v(x_i) - g(x_i). \end{aligned}$$

But (1.4) at the point x_i is

$$\begin{cases} -\varepsilon u_{xx}(x_i) - a_1(x_i)u_x(x_i) + b_{11}(x_i)u(x_i) + b_{12}(x_i)v(x_i) = f(x_i), \\ -\mu v_{xx}(x_i) - a_2(x_i)v_x(x_i) + b_{21}(x_i)u(x_i) + b_{22}(x_i)v(x_i) = g(x_i). \end{cases}$$

Thus, (3.5) is reduced to

$$\xi_1 = -\varepsilon \frac{h^2}{24} [u^{(4)}(x_2) + u^{(4)}(x_3)] - a_1(x_i) \frac{h}{2}u^{(2)}(x_1) \quad (3.6)$$

$$\xi_2 = -\mu \frac{h^2}{24} [v^{(4)}(y_2) + v^{(4)}(y_3)] - a_2(x_i) \frac{h}{2}v^{(2)}(y_1)$$

from which we deduce that

$$|\xi_1| \leq \varepsilon \frac{h^2}{24} [|u^{(4)}(x_2)| + |u^{(4)}(x_3)|] + |a_1(x_i)| \frac{h}{2}|u^{(2)}(x_1)| \quad (3.7)$$

$$|\xi_2| \leq \mu \frac{h^2}{24} [|v^{(4)}(y_2)| + |v^{(4)}(y_3)|] + |a_2(x_i)| \frac{h}{2}|v^{(2)}(y_1)|$$

Now applying Lemma 2.4, there are $x^* \in (x_{i-1}, x_{i+1})$ and $y^* \in (x_{i-1}, x_{i+1})$ such that (3.7) becomes

$$\begin{aligned} |\xi_1| &\leq \varepsilon \frac{h^2}{12} |u^{(4)}(x^*)| + |a_1(x_i)| |u^{(2)}(x_1)| \frac{h}{2} \\ |\xi_2| &\leq \mu \frac{h^2}{12} |v^{(4)}(y^*)| + |a_2(x_i)| |v^{(2)}(y_1)| \frac{h}{2} \end{aligned} \quad (3.8)$$

Thus

$$\begin{aligned} \|\xi\|_\infty &= \max(|\xi_1|, |\xi_2|) \\ &\leq |\xi_1| + |\xi_2| \\ &\leq \max(\varepsilon, \mu) (|u^{(4)}(x^*)| + |v^{(4)}(y^*)|) \frac{h^2}{12} \\ &\quad + \max(|a_1(x_i)|, |a_2(x_i)|) (|u^{(2)}(x_1)| + |v^{(2)}(y_1)|) \frac{h}{2}. \end{aligned} \quad (3.9)$$

Finally since $u^{(4)}$ and $v^{(4)}$ are continuous on $[0, 1]$, $|u^{(4)}(x^*)|$, $|v^{(4)}(y^*)|$, $|u^{(2)}(x_1)|$ and $|v^{(2)}(y_1)|$ are bounded and we have

$$\|\xi\|_\infty \leq \max(\varepsilon, \mu) \mathbf{c}_1 h^2 + \max(|a_1(x_i)|, |a_2(x_i)|) \mathbf{c}_2 h,$$

which is the result announced.

It is manifest that the consistency depend on $\max(\varepsilon, \mu)$, and we have the following

Corollary 3.5. *Let (u, v) be the solution of (1.4), assumed to be four times continuously differentiable in $[0, 1]$.*

- *If $\max(\varepsilon, \mu)$ is a fixed parameter independent of h , then the finite difference scheme (2.8) is consistent with respect to $\|\cdot\|_\infty$ norm, of order one with (1.4) in the following sense*

$$\|\xi\|_\infty \leq \mathbf{c}h,$$

where \mathbf{c} is a positive constant independent of h , μ , and ε .

- *If $\max(\varepsilon, \mu) = O(h^{-1})$, then the finite difference scheme (2.8) is consistent with respect to $\|\cdot\|_\infty$ norm, of order one with (1.4) in the following sense*

$$\|\xi\|_\infty \leq \mathbf{c}h,$$

where \mathbf{c} is a positive constant independent of h , μ , and ε .

- If $\max(\varepsilon, \mu) = 0(h^{-\alpha})$, with $\alpha \geq 2$ then we have

$$\|\xi\|_\infty \leq \mathbf{c}h^{2-\alpha},$$

with \mathbf{c} a positive constant. The scheme (2.8) is not consistent with respect to $\|\cdot\|_\infty$ norm, as $h^{2-\alpha}$ is not approaching zero as $h \rightarrow 0$.

3.2 Stability analysis

Definition 3.6. A discretization method (3.2) is stable if for some positive constant \mathbf{c} independent of h one has

$$\|u_h - v_h\|_{U_h} \leq \mathbf{c}\|A_h u_h - A_h v_h\|_{V_h} \quad \text{for all } u_h, v_h \in U_h.$$

Hence stability ensures that rounding errors occurring in the problem will not have excessive effect on the final result.

Remark 3.7. If the discrete operators A_h are linear, then stability of the discretization is equivalent to the existence of a constant $\mathbf{c} > 0$, independent of h , such that

$$\|A_h^{-1}\| \leq \mathbf{c}.$$

The stability condition (see definition 3.6) is replaced by

$$\|u_h\|_{U_h} \leq \mathbf{c}\|A_h u_h\|_{V_h} \quad \text{for all } u_h \in U_h.$$

But when A_h are nonlinear, the property mentioned in definition 3.6 is only needed in some neighborhood of the discrete solution u_h .

theorem 3.8. Let $u = (u_i)_i$, $v = (v_i)_i$ be the solution of (2.8), then there exists a constant \mathbf{c} independent of h , ε and μ such that,

$$\|u, v\| \leq \frac{\mathbf{c}}{\varepsilon^{1/2}} \|f, g\|$$

Proof. We multiply of (2.8) by u_i and the second one by v_i . We obtain

$$\begin{aligned}
-\varepsilon D^+ D^- u_i u_i - a_1(x_i) D^- u_i u_i + b_{11}(x_i) u_i u_i + b_{21}(x_i) v_i u_i &= f(x_i) u_i, \\
-\mu D^+ D^- v_i v_i - a_2(x_i) D^- v_i v_i + b_{12}(x_i) u_i v_i + b_{22}(x_i) v_i v_i &= g(x_i) v_i.
\end{aligned} \tag{3.10}$$

Summing (3.10) for $i = 1, 2, 3, \dots, N$ gives

$$\begin{aligned}
-\varepsilon \sum_{i=1}^N D^+ D^- u_i u_i - \sum_{i=1}^N a_1(x_i) D^- u_i u_i + \sum_{i=1}^N b_{11}(x_i) u_i u_i + \sum_{i=1}^N b_{12}(x_i) v_i u_i &= \sum_{i=1}^N f(x_i) u_i \\
-\mu \sum_{i=1}^N D^+ D^- v_i v_i - \sum_{i=1}^N a_2(x_i) D^- v_i v_i + \sum_{i=1}^N b_{21}(x_i) u_i v_i + \sum_{i=1}^N b_{22}(x_i) v_i^2 &= \sum_{i=1}^N g(x_i) v_i.
\end{aligned} \tag{3.11}$$

Applying (2.13) we get,

$$\begin{aligned}
\varepsilon \sum_{i=1}^{N+1} D^- u_i D^- u_i - \sum_{i=1}^{N+1} a_1(x_i) D^- u_i u_i + \sum_{i=1}^{N+1} b_{11}(x_i) u_i u_i + \sum_{i=1}^{N+1} b_{12}(x_i) v_i u_i &= \sum_{i=1}^{N+1} f(x_i) u_i \\
\mu \sum_{i=1}^{N+1} D^- v_i D^- v_i - \sum_{i=1}^{N+1} a_2(x_i) D^- v_i v_i + \sum_{i=1}^{N+1} b_{21}(x_i) u_i v_i + \sum_{i=1}^{N+1} b_{22}(x_i) v_i v_i &= \sum_{i=1}^{N+1} g(x_i) v_i.
\end{aligned} \tag{3.12}$$

Rearranging (3.12) we have,

$$\begin{aligned}
\varepsilon \sum_{i=1}^{N+1} (D^- u_i)^2 + \sum_{i=1}^{N+1} b_{11}(x_i) u_i u_i &= \sum_{i=1}^{N+1} a_1(x_i) D^- u_i u_i - \sum_{i=1}^{N+1} b_{12}(x_i) v_i u_i + \sum_{i=1}^{N+1} f(x_i) u_i \\
\mu \sum_{i=1}^{N+1} (D^- v_i)^2 + \sum_{i=1}^{N+1} b_{22}(x_i) v_i^2 &= \sum_{i=1}^{N+1} a_2(x_i) D^- v_i v_i - \sum_{i=1}^{N+1} b_{21}(x_i) u_i v_i + \sum_{i=1}^{N+1} g(x_i) v_i.
\end{aligned} \tag{3.13}$$

Adding the two equations that made up (3.13) we have

$$\begin{aligned}
& \varepsilon \|D^- u\|^2 + \mu \|D^- v\|^2 + \sum_{i=1}^{N+1} b_{11}(x_i) u_i^2 + \sum_{i=1}^{N+1} b_{22}(x_i) v_i^2 \\
&= \sum_{i=1}^{N+1} a_1(x_i) u_i D^- u_i + \sum_{i=1}^{N+1} a_2(x_i) v_i D^- v_i - \sum_{i=1}^{N+1} (b_{12}(x_i) + b_{21}(x_i)) u_i v_i \\
&+ \sum_{i=1}^{N+1} (f(x_i) u_i + g(x_i) v_i), \tag{3.14}
\end{aligned}$$

where

$$\|D^- u\|^2 = \sum_{i=1}^{N+1} (D^- u_i)^2, \quad \|D^- v\|^2 = \sum_{i=1}^{N+1} (D^- v_i)^2.$$

Now we would like to bound from above the right hand side of (3.14). First,

$$\begin{aligned}
& \sum_{i=1}^{N+1} a_1(x_i) u_i D^- u_i + \sum_{i=1}^{N+1} a_2(x_i) v_i D^- v_i \\
\leq & \sum_{i=1}^{N+1} a_1(x_i) |u_i| |D^- u_i| + \sum_{i=1}^{N+1} a_2(x_i) |v_i| |D^- v_i| \\
\leq & \sum_{i=1}^{N+1} (a_1^2(x_i) \frac{|u_i|^2}{2\delta_1} + \frac{\delta_1}{2} |D^- u_i|^2) + \sum_{i=1}^{N+1} (a_2^2(x_i) \frac{|v_i|^2}{2\delta_2} + \frac{\delta_2}{2} |D^- v_i|^2),
\end{aligned}$$

where (2.12) was used. Thus,

$$\begin{aligned}
\sum_{i=1}^{N+1} a_1(x_i) u_i D^- u_i + \sum_{i=1}^{N+1} a_2(x_i) v_i D^- v_i &\leq \sum_{i=1}^{N+1} (a_1^2(x_i) \frac{|u_i|^2}{2\delta_1}) + \sum_{i=1}^{N+1} (a_2^2(x_i) \frac{|v_i|^2}{2\delta_2}) \\
&+ \frac{\delta_1}{2} \|D^- u\|^2 + \frac{\delta_2}{2} \|D^- v\|^2, \tag{3.15}
\end{aligned}$$

where it is noted that $|D^+u_i| = |D^-u_i|$. Secondly,

$$\begin{aligned}
-\sum_{i=1}^{N+1} (b_{12}(x_i) + b_{21}(x_i))u_iv_i &\leq \left| \sum_{i=1}^{N+1} (b_{12}(x_i) + b_{21}(x_i))u_iv_i \right| \\
&\leq \sum_{i=1}^{N+1} (|b_{12}(x_i)| + |b_{21}(x_i)|)|u_i||v_i| \\
&\leq \sum_{i=1}^{N+1} (|b_{12}(x_i)| + |b_{21}(x_i)|) \left(\frac{1}{2\delta_3}|u_i|^2 + \frac{\delta_3}{2}|v_i|^2 \right)
\end{aligned}$$

where (2.12) was applied. Define $\|b\| = \max_i \{|b_{12}(x_i)|, |b_{21}(x_i)|\}$, hence

$$b_{12}(x_i) + b_{21}(x_i) \leq 2\|b\|.$$

We deduce that

$$\begin{aligned}
-\sum_{i=1}^{N+1} (b_{12}(x_i) + b_{21}(x_i))u_iv_i &\leq \frac{\|b\|}{\delta_3} \sum_{i=1}^{N+1} |u_i|^2 + \|b\|\delta_3 \sum_{i=1}^{N+1} |v_i|^2 \\
&\leq \frac{\|b\|}{\delta_3} \|u\|^2 + \|b\|\delta_3 \|v\|^2. \tag{3.16}
\end{aligned}$$

Finally

$$\begin{aligned}
\sum_{i=1}^{N+1} (f_i u_i + g_i v_i) &\leq \sum_{i=1}^{N+1} |f_i| |u_i| + |g_i| |v_i| \\
&\leq \left(\sum_{i=1}^{N+1} |f_i|^2 \right)^{\frac{1}{2}} \left(\sum_{i=1}^{N+1} |u_i|^2 \right)^{\frac{1}{2}} + \left(\sum_{i=1}^{N+1} |g_i|^2 \right)^{\frac{1}{2}} \left(\sum_{i=1}^{N+1} |v_i|^2 \right)^{\frac{1}{2}},
\end{aligned}$$

where (2.11) has been used. We let

$$\|f\|^2 = \sum_{i=1}^{N+1} |f_i|^2, \quad \|g\|^2 = \sum_{i=1}^{N+1} |g_i|^2.$$

Hence

$$\begin{aligned}
\sum_{i=1}^{N+1} (f_i u_i + g_i v_i) &\leq \|f\| \|u\| + \|g\| \|v\| \\
&\leq \frac{1}{2\delta_4} \|f\|^2 + \frac{\delta_4}{2} \|u\|^2 + \frac{1}{2\delta_5} \|g\|^2 + \frac{\delta_5}{2} \|v\|^2 \\
&\leq \frac{1}{2\delta_4} \|f\|^2 + \frac{1}{2\delta_5} \|g\|^2 + \frac{\delta_4}{2} \|u\|^2 + \frac{\delta_5}{2} \|v\|^2. \quad (3.17)
\end{aligned}$$

Substituting (3.17) , (3.16) and (3.15) in (3.14) we have,

$$\begin{aligned}
&\varepsilon \|D^- u\|^2 + \mu \|D^- v\|^2 + \sum_{i=1}^{N+1} b_{11}(x_i) u_i^2 + \sum_{i=1}^{N+1} b_{22}(x_i) v_i^2 \\
&\leq \frac{\delta_1}{2} \|D^- u\|^2 + \frac{\delta_2}{2} \|D^- v\|^2 + \frac{1}{2\delta_1} \sum_{i=1}^{N+1} a_1^2(x_i) |u_i|^2 + \frac{1}{2\delta_2} \sum_{i=1}^{N+1} a_2^2(x_i) |v_i|^2 \\
&\quad + \frac{\|b\|}{\delta_3} \|u\|^2 + \|b\| \delta_3 \|v\|^2 + \frac{1}{2\delta_4} \|f\|^2 + \frac{1}{2\delta_5} \|g\|^2 + \frac{\delta_4}{2} \|u\|^2 + \frac{\delta_5}{2} \|v\|^2.
\end{aligned}$$

Rearranging to get,

$$\begin{aligned}
&(\varepsilon - \frac{\delta_1}{2}) \|D^- u\|^2 + (\mu - \frac{\delta_2}{2}) \|D^- v\|^2 + (\varepsilon - \frac{\delta_1}{2}) \|D^-\|^2 + \sum_{i=1}^{N+1} b_{11}(x_i) u_i^2 + \sum_{i=1}^{N+1} b_{22}(x_i) v_i^2 \\
&\leq \frac{1}{2\delta_1} \sum_{i=1}^{N+1} a_1^2(x_i) |u_i|^2 + \frac{1}{2\delta_2} \sum_{i=1}^{N+1} a_2^2(x_i) |v_i|^2 + (\frac{\|b\|}{\delta_3} + \frac{\delta_4}{2}) \|u\|^2 + (\|b\| \delta_3 + \frac{\delta_5}{2}) \|v\|^2 \\
&\quad + \frac{1}{2\delta_4} \|f\|^2 + \frac{1}{2\delta_5} \|g\|^2. \quad (3.18)
\end{aligned}$$

Now, we need to choose $\delta_1, \delta_2, \delta_3, \delta_4$ and δ_5 .

First, we take δ_1 and δ_2 such that

$$\varepsilon - \frac{\delta_1}{2} > 0 \quad , \quad \mu - \frac{\delta_2}{2} > 0.$$

One possibility is $(\delta_1, \delta_2) = (\varepsilon, \mu)$. Replacing the values of δ_1 and δ_2 in (3.18), we

obtain (using $a_k(x_i) \geq \alpha$)

$$\begin{aligned} & \frac{\varepsilon}{2} \|D^- u\|^2 + \frac{\mu}{2} \|D^- v\|^2 + \sum_{i=1}^{N+1} (b_{11}(x_i) |u_i|^2 - \frac{\alpha^2}{2\varepsilon} |u_i|^2) - \left(\frac{\|b\|}{\delta_3} + \frac{\delta_4}{2} \right) \|u\|^2 \\ & + \sum_{i=1}^{N+1} (b_{22}(x_i) |v_i|^2 - \frac{\alpha^2}{2\mu} |v_i|^2) - \left(\|b\| \delta_3 + \frac{\delta_5}{2} \right) \|v\|^2 \leq \frac{1}{2\delta_4} \|f\|^2 + \frac{1}{2\delta_5} \|g\|^2 \end{aligned}$$

Rearranging to get,

$$\begin{aligned} & \frac{\varepsilon}{2} \|D^- u\|^2 + \frac{\mu}{2} \|D^- v\|^2 + \sum_{i=1}^{N+1} \left[b_{11}(x_i) - \frac{\alpha^2}{2\varepsilon} - \frac{\|b\|}{\delta_3} - \frac{\delta_4}{2} \right] |u_i|^2 \\ & + \sum_{i=1}^{N+1} \left[b_{22}(x_i) - \frac{\alpha^2}{2\mu} - \|b\| \delta_3 - \frac{\delta_5}{2} \right] |v_i|^2 \leq \frac{1}{2\delta_4} \|f\|^2 + \frac{1}{2\delta_5} \|g\|^2 \end{aligned} \tag{3.19}$$

we need δ_3 , δ_4 and δ_5 positive such that,

$$\begin{aligned} & \frac{\alpha^2}{2\varepsilon} + \frac{\|b\|}{\delta_3} + \frac{\delta_4}{2} \leq b_{11}(x_i) \leq \|b\|, \\ & \frac{1\alpha^2}{2\mu} + \|b\| \delta_3 + \frac{\delta_5}{2} \leq b_{22}(x_i) \leq \|b\|. \end{aligned}$$

which is reduced to

$$\frac{\|b\|}{\delta_3} + \frac{\delta_4}{2} \leq b_{11}(x_i) \leq \|b\|, \tag{3.20}$$

and

$$\|b\| \delta_3 + \frac{\delta_5}{2} \leq b_{22}(x_i) \leq \|b\|. \tag{3.21}$$

From (3.20) we have

$$\delta_3 \geq \frac{\|b\|}{\|b\| - \frac{\delta_4}{2}} \tag{3.22}$$

and from (3.21) we have

$$\delta_3 \leq \frac{\|b\| - \frac{\delta_5}{2}}{\|b\|} \quad (3.23)$$

Therefore (3.23) and (3.22) yields

$$\frac{\|b\|}{\|b\| - \frac{\delta_4}{2}} \leq \delta_3 \leq \frac{\|b\| - \frac{\delta_5}{2}}{\|b\|} \quad (3.24)$$

Therefore one needs δ_4 and δ_5 such that,

$$\frac{\|b\|}{\|b\| - \frac{\delta_4}{2}} \leq \frac{\|b\| - \frac{\delta_5}{2}}{\|b\|} \quad (3.25)$$

i.e.,

$$\|b\|^2 \leq (\|b\| - \frac{\delta_4}{2})(\|b\| - \frac{\delta_5}{2}) \quad (3.26)$$

(3.26) is one inequality with two unknowns, hence there infinitely many possibilities for (δ_4, δ_5) independent of ε and μ .

Returning to (3.19), with the way δ_3 , δ_4 , and δ_5 was chosen, that are;

$$\begin{aligned} b_{11}(x_i) - \frac{\alpha^2}{2\varepsilon} - \frac{\|b\|}{\delta_3} - \frac{\delta_4}{2} &\geq 0, \\ b_{22}(x_i) - \frac{\alpha^2}{2\mu} - \|b\|\delta_3 - \frac{\delta_5}{2} &\geq 0, \end{aligned}$$

we deduce that

$$\begin{aligned} (b_{11}(x_i) - \frac{\alpha^2}{2\varepsilon}, -\frac{\|b\|}{\delta_3} - \frac{\delta_4}{2})|u_i|^2 &\geq 0, \\ (b_{22}(x_i) - \frac{\alpha^2}{2\mu} - \|b\|\delta_3 - \frac{\delta_5}{2})|v_i|^2 &\geq 0. \end{aligned}$$

Hence there is positive constant \mathbf{c} independent of μ, ε such that the relation (3.19) implies that

$$\frac{\varepsilon}{2} \|D^-u\|^2 + \frac{\mu}{2} \|D^-v\|^2 \leq \mathbf{c} \|f, g\|^2. \quad (3.27)$$

It should be observed that the constant \mathbf{c} in (3.27) is independent of μ, ε because deduce from δ_4, δ_5 which are independent of μ, ε . Observing that

$$\min\left(\frac{\varepsilon}{2}, \frac{\mu}{2}\right) [\|D^-u\|^2 + \|D^-v\|^2] \leq \frac{\varepsilon}{2} \|D^-u\|^2 + \frac{\mu}{2} \|D^-v\|^2,$$

(3.27) becomes,

$$\min\left(\frac{\varepsilon}{2}, \frac{\mu}{2}\right) [\|D^-u\|^2 + \|D^-v\|^2] \leq \mathbf{c} \|f, g\|^2,$$

which by (2.14) leads to

$$\min\left(\frac{\varepsilon}{2}, \frac{\mu}{2}\right) \|u, v\|^2 \leq \mathbf{c} \|f, g\|^2$$

which is

$$\|u, v\| \leq \frac{\mathbf{c}}{\varepsilon^{1/2}} \|f, g\|. \quad (3.28)$$

Hence the proof is complete.

As a crucial consequence of Theorem 3.8, we have

Corollary 3.9. *The solution of (2.8) is not uniformly stable with respect to ε, μ .*

Indeed

$$\lim_{\varepsilon \rightarrow 0} \|u, v\| \leq \infty.$$

The lack of uniform stability with respect to (μ, ε) as we will see in the next Chapter is reason of the poor convergence when $\varepsilon \rightarrow 0$.

CHAPTER 4

Convergence

The convergence of the method is defined similarly as the stability. Indeed

Definition 4.1. *A discretization method is convergence if the error satisfies*

$$\|\tilde{u} - u_h\|_{U_h} \rightarrow 0 \quad h \rightarrow 0,$$

and convergent of order p if there exists a constant $\mathbf{c} > 0$ such that

$$\|\tilde{u} - u_h\|_{U_h} \leq \mathbf{c}h^p.$$

Convergence is proved by demonstrating consistency and stability of the discretization.

Indeed definition 3.2 and definition 3.6 imply immediately the following abstract convergence theorem

theorem 4.2. *Assume that both the continuous and the discrete problem have a unique solutions. If the discretization method is consistent and stable then the method is also convergent. Furthermore, the order of convergence is at least as large as the order of consistency of the method.*

theorem 4.3. *Let $e_i = \overrightarrow{u(x_i)} - \overrightarrow{u_i}$ be a mesh function where $\overrightarrow{u(x_i)}$ is the exact solution at x_i and $\overrightarrow{u_i}$ is the numerical approximation to $\overrightarrow{u(x_i)}$. Then there exists a generic constant \mathbf{c} independent of h such that*

$$\|u(x_i) - u_i, v(x_i) - v_i\| \leq \mathbf{c} [\mu h^2 + h] \frac{1}{\varepsilon^{1/2}}. \quad (4.1)$$

Proof. we start by setting

$$e_i = \overrightarrow{u(x_i)} - \overrightarrow{u_i} \quad (4.2)$$

$$\overrightarrow{u(x_i)} = (u(x_i), v(x_i)), \quad \overrightarrow{u_i} = (u_i, v_i) \text{ and } e_i = (e_{1i}, e_{2i}).$$

The equation (4.2) is re-written as follows,

$$\overrightarrow{u_i} = \overrightarrow{u(x_i)} - e_i \quad (4.3)$$

and (4.3) is equivalent to

$$u_i = u(x_i) - e_{1i} \quad , \quad v_i = v(x_i) - e_{2i}. \quad (4.4)$$

Substituting (4.4) in the SFD scheme (2.8) we have,

$$\begin{cases} -\epsilon D^- D^+(u(x_i) - e_{1i}) - a_1(x_i) D^+(u(x_i) - e_{1i}) + b_{11}(x_i)(u(x_i) - e_{1i}) + b_{12}(x_i)(v(x_i) - e_{2i}) \\ = f(x_i) \\ -\mu D^- D^+(v(x_i) - e_{2i}) - a_2(x_i) D^+(v(x_i) - e_{2i}) + b_{21}(x_i)(u(x_i) - e_{1i}) + b_{22}(x_i)(v(x_i) - e_{2i}) \\ = g(x_i) \end{cases}$$

Rearranging to get,

$$\begin{cases} -\epsilon D^- D^+ e_{1i} - a_1(x_i) D^+ e_{1i} + b_{11}(x_i) e_{1i} + b_{12}(x_i) e_{2i} = \xi_1 \\ -\mu D^- D^+ e_{2i} - a_2(x_i) D^+ e_{2i} + b_{21}(x_i) e_{1i} + b_{22}(x_i) e_{2i} = \xi_2 \\ e_{1(0)} = e_{2(0)} = 0 \\ e_{1(N+1)} = e_{2(N+1)} = 0 \end{cases} \quad (4.5)$$

Where,

$$\xi_1(x_i) = -\epsilon D^+ D^- u_i - a_1(x_i) D^+ u_i + b_{11}(x_i) u_i + b_{12}(x_i) v_i - f(x_i)$$

$$\xi_2(x_i) = -\mu D^+ D^- v_i - a_2(x_i) D^+ v_i + b_{12}(x_i) u_i + b_{22}(x_i) v_i - g(x_i)$$

From stability analysis, i.e., Theorem (3.8) we have that,

$$\|e_1, e_2\| \leq \frac{\mathbf{c}}{\varepsilon^{\frac{1}{2}}} \|\xi_1, \xi_2\| \leq \sqrt{N} \frac{\mathbf{c}}{\varepsilon^{\frac{1}{2}}} \|\xi\|_{\infty}. \quad (4.6)$$

From Theorem 3.8,

$$\|\xi\|_{\infty} \leq \mu \mathbf{c}_1 h^2 + \max_{1 \leq i \leq N+1} (|a_1(x_i)|, |a_2(x_i)|) \mathbf{c}_2 h,$$

where $\mathbf{c}_1, \mathbf{c}_2$ are positive constants independent of h, μ , and ε . Thus (4.6) becomes

$$\|e_1, e_2\| \leq \frac{\mathbf{c}}{\varepsilon^{\frac{1}{2}}} [\mu h + 1] h. \quad (4.7)$$

So the proof is complete.

it is clear from the above (4.3) that the convergence is not uniform with respect to ε and μ . As one can see:

- If μ, ε are independent of h ,

$$\|e_1, e_2\| \leq ch \quad (4.8)$$

- If μ goes to zero, then $\mu h^2 + h \leq ch$, and

$$\|e_1, e_2\| \leq \frac{c}{\varepsilon^{\frac{1}{2}}} h. \quad (4.9)$$

Hence the convergence is very poor as $\varepsilon^{1/2}$ is close to h , and it will blow up if $\varepsilon^{1/2} \ll h$

Corollary 4.4. *The solution to the SFD scheme (2.8) does not converge uniformly to the solution of the problem (1.4) with respect to the perturbation parameters ε and μ .*

Indeed for $\varepsilon^{1/2} \ll h$

$$\lim_{\varepsilon \rightarrow 0} \|e_1, e_2\| \leq \infty.$$

CHAPTER 5

Numerical simulations and future works

In this Chapter, we demonstrate computationally the theoretical results obtained in Chapter 4. More particularly, we show the poor convergence display in Theorem 4.3, and we also show that if the mesh size goes to zero faster than ε then good behaviour is obtained. We conclude this Chapter by reviewing our results, and indicating some future research to correct the bad behaviour of the numerical solution for (2.8).

5.1 Numerical examples

The focus in this section is to show that the parameters ε and μ are responsible for the poor/good convergence of the finite difference (2.8).

In this section, we first write down the system of equations (2.8) in the form

$$\mathcal{B}\psi = F, \tag{5.1}$$

with $N + 1$ the number of grid points in $[0, 1]$, \mathcal{B} is a $2N \times 2N$ matrix, $\psi^T = (u_1, v_1, u_2, v_2, \dots, u_N, v_N)$ is the unknown vector of size $2N \times 1$, and $F = (f_1, g_1, f_2, g_2, \dots, f_N, g_N)$ is given vector of size $2N \times 1$. We recall that the mesh size is

$$h = \frac{1}{N + 1}.$$

All the examples discussed are from [1], and we also recall that the rate of con-

vergence is computed using the formula

$$\alpha = \frac{\log e_2/e_1}{\log h_2/h_1}.$$

We consider the following test example

$$\begin{aligned} -\varepsilon u'' - u' + 2u - v &= f, \\ -\mu v'' - 2v' - u + 4v &= g, \\ u(0) = u(1) = v(0) = v(1) &= 0. \end{aligned}$$

Where comparing with (1.4) we have that $a_1(x_i) = 1$, $a_2(x_i) = 2$, $b_{11}(x_i) = 2$, $b_{12}(x_i) = -1$, $b_{21} = -1$, $b_{22}(x_i) = 4 \quad \forall i$.

The source term (f, g) are chosen so that the exact solution is

$$\begin{aligned} u(x) &= \frac{1 - \exp(-x/\varepsilon)}{1 - \exp(-1/\varepsilon)} + \frac{1 - \exp(-x/\mu)}{1 - \exp(-1/\mu)} - 2 \sin \frac{\pi}{2} x, \\ v(x) &= \frac{1 - \exp(-x/\mu)}{1 - \exp(-1/\mu)} - x \exp(x - 1). \end{aligned}$$

solving for the source term (f, g) we have,

$$\begin{aligned} f(x) &= \left[\frac{\varepsilon}{\mu^2} - \frac{1}{\mu} \right] \frac{\exp(-x/\mu)}{1 - \exp(-1/\mu)} - \left[\frac{\varepsilon \pi^2}{2} + 4 \right] \sin \frac{\pi}{2} x + 2 \left[\frac{1 - \exp(-x/\varepsilon)}{1 - \exp(-1/\varepsilon)} \right] - \left[\frac{1 - \exp(-x/\mu)}{1 - \exp(-1/\mu)} \right] + \\ &\pi \cos \frac{\pi}{2} x + x \exp(x - 1), \\ g(x) &= 3 \left[\frac{1 - \exp(-x/\mu)}{1 - \exp(-1/\mu)} \right] - \left[\frac{\exp(-x/\mu)}{\mu(1 - \exp(-1/\mu))} \right] - \left[\frac{1 - \exp(-x/\varepsilon)}{1 - \exp(-1/\varepsilon)} \right] + \\ &[\mu x + 2\mu + 2 - 2x] \exp(x - 1) + 2 \sin \frac{\pi}{2} x. \end{aligned}$$

Firstly, the system of equations (2.8) can be written in this form, for $N = 4$ we have ,

$$\left\{ \begin{array}{cccccccc}
A_1 & b_{12}(x_1) & C_1 & 0 & 0 & 0 & 0 & 0 \\
b_{21}(x_1) & B_1 & 0 & D_1 & 0 & 0 & 0 & 0 \\
\frac{-\epsilon}{h^2} & 0 & A_2 & b_{12}(x_2) & C_2 & 0 & 0 & 0 \\
0 & \frac{-\mu}{h^2} & b_{21}(x_2) & B_2 & 0 & D_2 & 0 & 0 \\
0 & 0 & \frac{-\epsilon}{h^2} & 0 & A_3 & b_{12}(x_3) & C_3 & 0 \\
0 & 0 & 0 & \frac{-\mu}{h^2} & b_{21}(x_3) & B_3 & 0 & D_3 \\
0 & 0 & 0 & 0 & \frac{-\epsilon}{h^2} & 0 & A_4 & b_{12}(x_4) \\
0 & 0 & 0 & 0 & 0 & \frac{-\mu}{h^2} & b_{21}(x_4) & B_4
\end{array} \right\} \begin{pmatrix} u_1 \\ v_1 \\ u_2 \\ v_2 \\ u_3 \\ v_3 \\ u_4 \\ v_4 \end{pmatrix} = \begin{pmatrix} f(x_1) \\ g(x_1) \\ f(x_2) \\ g(x_2) \\ f(x_3) \\ g(x_3) \\ f(x_4) \\ g(x_4) \end{pmatrix}$$

Where,

$$A_1 = \frac{2\epsilon}{h^2} + \frac{a_1(x_1)}{h},$$

$$B_1 = \frac{2\mu}{h^2} + \frac{a_2(x_1)}{h},$$

$$A_i = \frac{2\epsilon}{h^2} + \frac{a_1(x_i)}{h} + b_{11}(x_i),$$

$$B_i = \frac{2\mu}{h^2} + \frac{a_2(x_i)}{h} + b_{22}(x_i) \quad i \geq 2.$$

$$C_i = \frac{-\epsilon}{h^2} - \frac{a_1(x_i)}{h},$$

$$D_i = \frac{-\mu}{h^2} - \frac{a_2(x_i)}{h}, \quad i \geq 1.$$

Hence the general form of (2.8) is,

$$\left\{ \begin{array}{cccccccc}
A_1 & b_{12}(x_1) & C_1 & 0 & \cdot & \cdot & \cdot & 0 \\
b_{21}(x_1) & B_1 & 0 & D_1 & \cdot & \cdot & \cdot & \cdot \\
\frac{-\epsilon}{h^2} & 0 & A_2 & \cdot & \cdot & \cdot & \cdot & \cdot \\
0 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 & D_{N-1} \\
\cdot & \cdot & \cdot & \cdot & \frac{-\epsilon}{h^2} & 0 & A_N & b_{12}(x_N) \\
0 & \cdot & \cdot & \cdot & 0 & \frac{-\mu}{h^2} & b_{21}(x_N) & B_N
\end{array} \right\} \begin{pmatrix} u_1 \\ v_2 \\ \cdot \\ \cdot \\ \cdot \\ v_{N-1} \\ u_N \\ v_N \end{pmatrix} = \begin{pmatrix} f_1 \\ g_1 \\ \cdot \\ \cdot \\ \cdot \\ g_{N-1} \\ f_N \\ g_N \end{pmatrix}$$

5.1.1 example 1

Here we solve the matrix equation equivalent to the example in consideration so as to obtain the numerical solutions, and the exact solution was evaluated at the internal nodes for any N , μ , ϵ . chosen. Then the difference between the numerical solution obtained for any chosen N , μ , ϵ . and the exact solution evaluated at the internal nodes were taken under the maximum norm so as to obtain the maximum error of all the errors at each node, then the rate of convergence formula was used to obtain the rates of convergence.

h	$(\mu, \epsilon) = 10^{-6}$	Rates of Convergence
$h_1 = 1/11$	$0.7247 = E_1$	
$h_2 = 1/21$	$0.8124 = E_2$	-0.017666
$h_3 = 1/31$	$0.8431 = E_3$	-0.09524
$h_4 = 1/41$	$0.8586 = E_4$	-0.065516
$h_5 = 1/51$	$0.8680 = E_5$	-0.049889

Table 5.1: Rate of convergence

5.1.2 example 2

h	$(\mu, \varepsilon) = (10^{-2}, 10^{-6})$	Rates of Convergence
$h_1 = 1/11$	$0.7299 = E_1$	
$h_2 = 1/21$	$0.8470 = E_2$	-0.23011
$h_3 = 1/31$	$0.9356 = E_3$	-0.25545
$h_4 = 1/41$	$1.0183 = E_4$	-0.30296
$h_5 = 1/51$	$1.1122 = E_5$	-0.40414

Table 5.2: Rate of convergence

The results obtained in Table (5.1) and Table (5.2) corroborated with the theoretical findings, indeed from Corollary 4.4, when $\varepsilon^{1/2}$ is less than h , no convergence is expected. This translated computationally with negative rate of convergence.

5.1.3 example 3

h	$(\mu, \varepsilon) = 10^{-2}$	Rates of Convergence
$h_1 = 1/101$	$0.8998 = E_1$	
$h_2 = 1/111$	$0.8932 = E_2$	0.07798
$h_3 = 1/121$	$0.8893 = E_3$	0.05073
$h_4 = 1/131$	$0.8865 = E_4$	0.03971
$h_5 = 1/141$	$0.8831 = E_5$	0.052237

Table 5.3: Rate of convergence

5.1.4 example 4

h	$(\mu, \varepsilon) = (10^{-1}, 10^{-2})$	Rates of Convergence
$h_1 = 1/101$	$0.8126 = E_1$	
$h_2 = 1/111$	$0.8094 = E_2$	0.038171
$h_3 = 1/121$	$0.8071 = E_3$	0.03299
$h_4 = 1/131$	$0.8056 = E_4$	0.02343
$h_5 = 1/141$	$0.8045 = E_5$	0.01857

Table 5.4: Rate of convergence

In Table (5.3) and Table (5.4), the rate of convergence are very poor because $\varepsilon^{1/2} > h$, this situation was predicted in the theory (see equation 4.9).

5.1.5 example 5

h	$(\mu, \varepsilon) = (10^{-1}, 10^{-2})$	Rates of Convergence
$h_1 = 1/21$	$0.9014 = E_1$	
$h_2 = 1/31$	$0.8920 = E_2$	0.02692
$h_3 = 1/41$	$0.8713 = E_3$	0.08391
$h_4 = 1/51$	$0.8526 = E_4$	0.09941
$h_5 = 1/61$	$0.8412 = E_5$	0.075181

Table 5.5: Rate of convergence

Here, $\varepsilon < h < \mu$ so from table 5.5 there will be convergence but the convergence will still be poor since $\varepsilon^{1/2}$ is close to h .

5.1.6 example 6

h	$(\mu, \varepsilon) = (10^{-1}, 10^{-1})$	Rates of Convergence
$h_1 = 1/6$	$0.6855 = E_1$	
$h_2 = 1/7$	$0.6673 = E_2$	0.1746
$h_3 = 1/8$	$0.6413 = E_3$	0.2976
$h_4 = 1/9$	$0.6123 = E_4$	0.3929
$h_5 = 1/10$	$0.5826 = E_5$	0.4719

Table 5.6: Rate of convergence

5.1.7 example 7

h	$(\mu, \varepsilon) = (0.9, 0.8)$	Rates of Convergence
$h_1 = 1/6$	$0.2057 = E_1$	
$h_2 = 1/7$	$0.1962 = E_2$	0.3067
$h_3 = 1/8$	$0.1861 = E_3$	0.3958
$h_4 = 1/9$	$0.1813 = E_4$	0.2219
$h_5 = 1/10$	$0.1742 = E_5$	0.3791

Table 5.7: Rate of convergence

In Table (5.6), h is almost equal to ε and μ and in Table (5.7) h is small compare to ε and μ , and in both Tables we could see a good behaviour which translated in the rates of convergence.

5.1.8 Graphs

In getting the maximum error, one should focus on the solution axis(y-axis) .

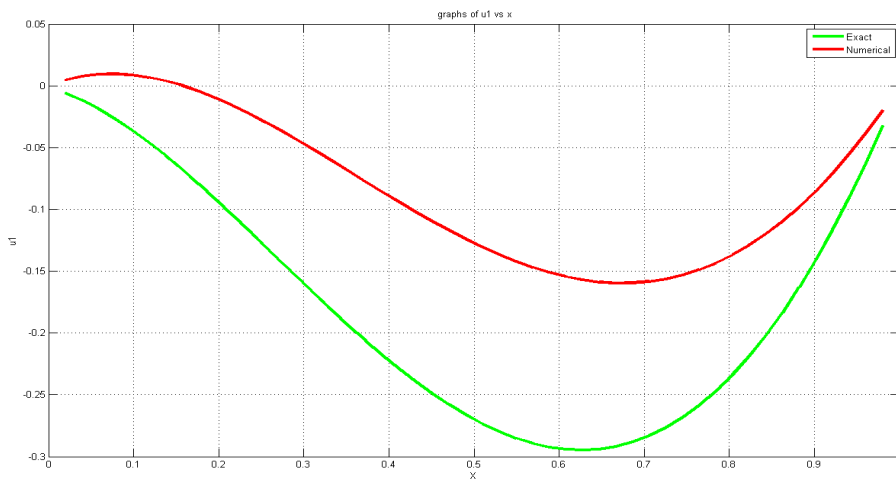


Figure 5.1: solution profile of (2.8)

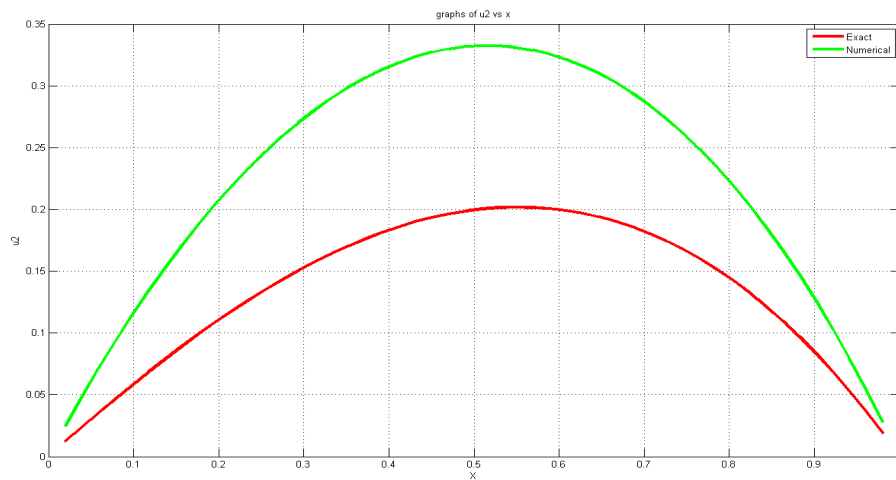


Figure 5.2: solution profile of (2.8)

For figure (5.1) and figure (5.2), $N = 50$, $\varepsilon = 0.8$, $\mu = 0.9$, (ε, μ) were chosen

to be so large compare to h , hence h goes to zero faster than (ε, μ) , so a good behaviour(i.e., relatively small maximum error) and a nice rate of convergence is expected which could be seen clearly in Table (5.6) and Table (5.7).

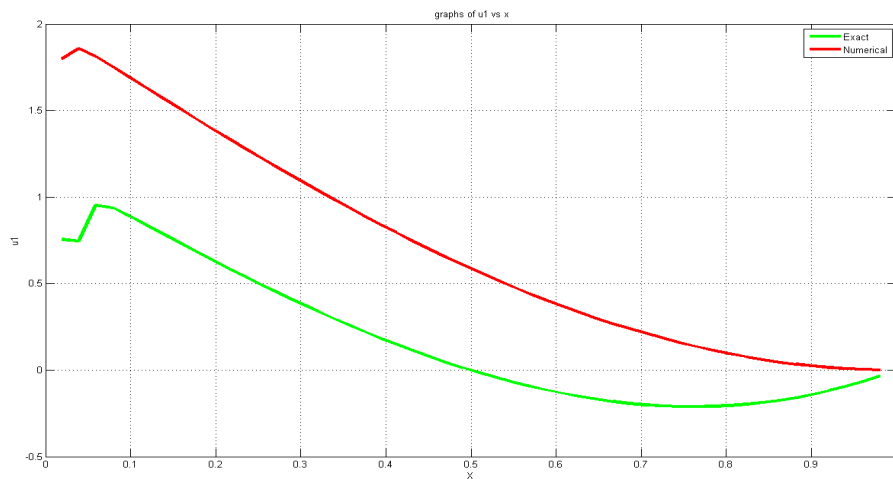


Figure 5.3: solution profile of (2.8)

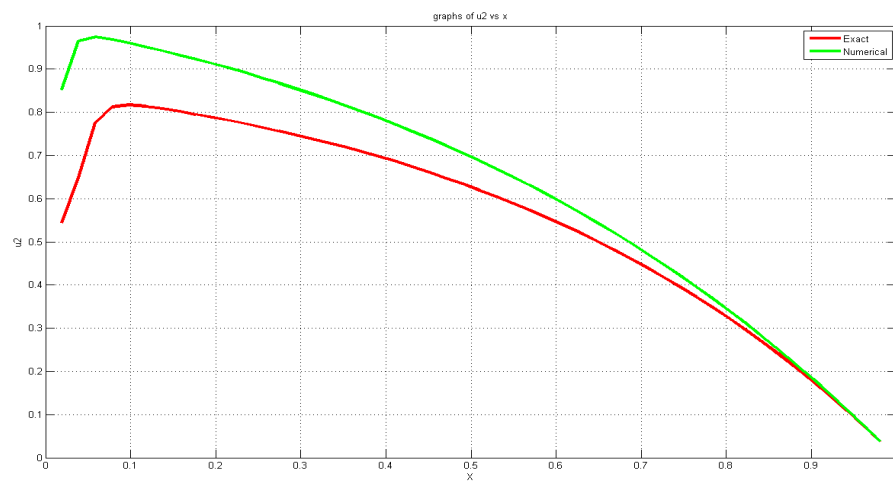


Figure 5.4: solution profile of (2.8)

For figure (5.3) and figure (5.4), $N = 50$, $\varepsilon = 1e - 6$, $\mu = 1e - 2$, the solution is badly behaved because ε is sufficiently small so the error is large, and also $\varepsilon^{1/2} < h$ so from corollary 4.4 no convergence is expected, Table (5.1) and Table (5.2) told us more.

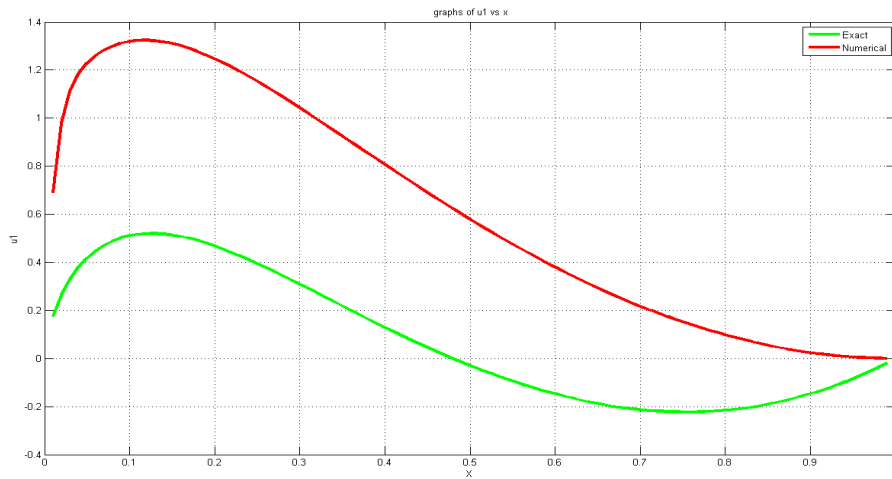


Figure 5.5: solution profile of (2.8)

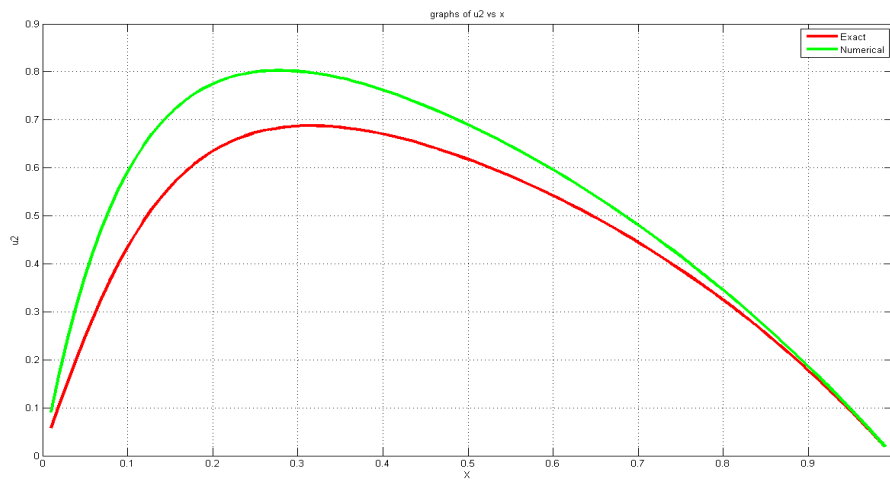


Figure 5.6: solution profile of (2.8)

For figure (5.5) and figure (5.6), $N = 100$, $\varepsilon = 1e - 2$, $\mu = 1e - 1$ the rate of convergence is very poor as can be seen in Table (5.3), Table (5.4) and Table (5.5) since $\varepsilon^{1/2} > h$

5.2 Conclusion and Future Research

In this thesis, we investigated the classical Finite Difference Scheme for a coupled system of singularly perturbed convection-diffusion equation, and we proved that the underlying scheme is not a good approximation for such problem. First we demonstrated the consistency of our proposed scheme to the continuous problem, we proved the existence and uniqueness of the solution of the discrete operator and analysed its stability in the l_2 -norm, it turned out that the solution of the discrete operator is not uniformly stable with respect to ε as the solution blows up when ε tends to zero. An error analysis also shows that the solution of the discrete operator does not converge uniformly to the solution of the continuous problem with respect to ε . Finally, the error between the exact solution and the solution of the discrete operator was simulated and investigated for several cases and we found that the error increases as the step size h gets smaller for sufficiently small values of ε and μ . We have observed that our theoretical findings support the numerical results that we have obtained.

Future research could be to investigate other numerical schemes such as the Non Standard Finite Difference Scheme (NSFD) or even the Finite Element Method (FEM) and others so as to see the one that will turn out to be robust. And this work could be extended to a singularly perturbed Convection-Diffusion system whose convective terms are coupled, as well as time dependent ones.

Bibliography

- [1] J.B. Munyakazi, A uniformly convergent nonstandard finite difference for a system of convection diffusion equations. *Computational and Applied Mathematics*, Vol 34, No. 3, 1153–1165, 2015.
- [2] B. Lucquin and O. Pironneau, Introduction to Scientific Computing. *Wiley*, 1998.
- [3] C. Grossmann, H. Gorg Roos, M. Stynes, Numerical Treatment of Partial Differential Equations. *Springer Verlag*, 2007.
- [4] A. Tveito and R. Winther, Introduction to Partial Differential Equations: A Computational Approach. *Springer Verlag*, 1998.
- [5] H.G. Roos, M. Stynes and L.Tobiska, Numerical Methods for Singularly Perturbed Differential Equations (Convection-Diffusion and Flow Problems), *Springer-Verlag Berlin Heidelberg* , 1996.