ANALYSIS OF MONOTONE NUMERICAL SCHEMES

by

Vladimir Nosov

A Thesis submitted to the Faculty of Graduate Studies of
The University of Manitoba
in partial fulfilment of the requirements of the degree of

MASTER OF SCIENCE

Department of Mathematics
University of Manitoba
Winnipeg

UNIVERSITY OF MANITOBA
DEPARTMENT OF DEPARTMENT OF MATHEMATICS

The undersigned hereby certify that they have read and recommend to the Faculty of Graduate Studies for acceptance a thesis entitled "*Analysis of Monotone Numerical Schemes*" by *Vladimir Nosov* in partial fulfillment of the requirements for the degree of *Master of Science*.

Dated: ______________________________

Supervisor: ______________________________
S. Kirkland & S.H. Lui

Reader : ______________________________
Ian Jeffrey

UNIVERSITY OF MANITOBA

Date:	*August 2016*
Author:	*Vladimir Nosov*
Title:	*Analysis of Monotone Numerical Schemes*
Department:	*Department of Mathematics*
Degree:	*M.Sc.*
Convocation:	*June 14*
Year:	*2016*

Signature of Author

Acknowledgements

I would like to express my gratitude to the following people who have helped me in the process of writing this thesis.

First I would to thank parents, for their unconditional love and support, who have always been there for me and put me on the path that lead to this day. Immigrating from Ukraine to Israel and then to Canada, always in pursuit for a better future for myself and my brother. I would not be here today without their guidance, teaching me that the key to success is education.

To my supervisors Dr. S.H. Lui and Dr. S. Kirkland for their guidance and help in the process of writing this thesis, for the financial support and the patience for all my questions and mistakes. I would not have been able to finish this thesis without your comments and corrections.

To my friends and fellow graduate students for all the good memories, help and advice. I am very grateful to all of you.

To the office staff for all the assistance and patience for all my questions and requests.

Last, I would like to express my gratitude to all the teachers and instructors who put their faith in me and motivated me to continue my education even further.

Abstract

In the study of partial differential equations (PDEs) one rarely finds an analytical solution. But a numerical solution can be found using different methods such as finite difference, finite element, etc. The main issue using such numerical methods is whether the numerical solution will converge to the “real” analytical solution and if so how fast will it converge as we shrink the discretization parameter .

In the first part of this thesis discrete versions of well known inequalities from analysis are used in proving the convergence of certain numerical methods for the one dimensional Poisson equation with Dirichlet boundary conditions and with Neumann boundary conditions.

A matrix is monotone if its inverse exists and is non-negative. In the second part of the thesis we will show that finite difference discretization of two PDEs result in monotone matrices. The monotonicity property will be used to demonstrate stability of certain methods for the Poisson and Biharmonic equations. Convergence of all schemes is also shown.

This thesis surveys known techniques to analyze numerical schemes.There are no original results demonstrated in the thesis other than proofs of monotonicity of several schemes.

Contents

Chapter 1

Introduction

In the modern world humans rely heavily on mathematical models to build, design and predict the outcome. Many mathematical models use partial differential equations (PDEs) to describe a certain system, such as the Navier-Stokes equations used to describe the flow of air around a wing. Nevertheless, even if we are given the PDE that describes a system, a “real” or exact solution is an elusive idea since it is usually impossible to find analytically. Thankfully, we have numerical methods that create numerical schemes to find the numerical solution of the system.

One of the oldest numerical methods is the finite difference (FD) method. The idea behind it is simple. Let us assume, for simplicity, that the domain is a rectangle. Set up a rectangular grid on the domain with a constant grid size h. Using finite differences to approximate derivatives at grid points, we end up with a finite system of equations to solve for the unknowns at the grid points.

Using analysis we can show that as the step size h becomes smaller, i.e we use more points, the numerical solution will converge to the analytical solution. In the second chapter of the thesis discrete inequalities equivalent to well known inequalities from analysis such as the Poincaré-Friedrichs and Poincaré inequalities will be proven. Afterwards the terms stability, consistency and convergence will be defined with respect to numerical schemes. Two numerical schemes solving the 1D case of the Poisson equation once with Dirichlet boundary conditions and once with Neumann boundary conditions will be introduced. Using the discrete inequalities proven in the beginning of the chapter we will demonstrate that the numerical schemes converge and at what rate they converge.

In the third chapter a family of matrices called monotone matrices will be introduced, such matrices are defined as matrices whose inverse is non-negative. The chapter also contains five equivalent conditions for a matrix to be monotone although many more can be found in [1]. Monotone matrices arise in many areas in mathematics, economy, engineering and scientific computing.

For example, in mathematics monotone matrices occur in the study of finite Markov chains in the field of probability theory. Even though the chapter is very short, its purpose is mainly to introduce the reader to monotone matrices which play an important role in showing stability of numerical schemes in later chapters.

In the fourth chapter, two monotone fourth order numerical schemes solving the 2D Poisson equation with Dirichlet boundary conditions will be defined. Unlike the previous schemes from the second chapter these schemes are fourth order meaning they converge faster to the solution. The first scheme has a structure such that for points next to the boundary we use the second order scheme from the first chapter and for interior point we use a fourth order scheme. Even though we use second order scheme next to the boundary, because there are so few points the overall convergence is still fourth order. The second scheme introduced in the chapter is proven to be monotone as well and using that property stability is proven. The restriction operator on the domain "samples" a function and turns it into a vector. In order to show fourth order convergence for this scheme a modified restriction operator is used.

In the fifth chapter a numerical scheme from [2] solving the 1D biharmonic equation will be introduced. It is a more complex scheme as it involves many steps all of which are shown and explained. As done before, monotonicity of the scheme is proven and later it is used in the proof of stability. Using the techniques from previous chapters a proof of first order convergence is shown, although in [2] a proof of fourth order convergence is shown using a different method.

Overall the thesis surveys known techniques to analyze numerical schemes.There are no original results demonstrated in the thesis other than proofs of monotonicity of several schemes. At the end of the thesis there is a page with all symbols used throughout the thesis and the first page where they appeared.

Chapter 2

Discrete Inequalities and Applications

2.1 Discrete Identities and Inequalities

In this chapter, we state well-known identities and inequalities for smooth functions, followed by the corresponding identities and inequalities for discrete functions. Their proofs follows those of the continues case. Throughout this thesis the terms c, c_1, c_2 represents a positive constant whose value might change in different places. For any positive integer N, let $h = \frac{1}{N}$. We denote the vector $[v_0, ..., v_N]$ by v_h. We define two finite difference schemes for approximating the derivative at a point i:

$$\delta_h v_i = \frac{v_{i+1} - v_i}{h},$$

$$\delta_{-h} v_i = \frac{v_i - v_{i-1}}{h}.$$

We now define three norms that will be used extensively later, given $v_h \in \mathbb{R}^{N+1}$.

The ℓ_∞ norm,	$\lvert v_h\rvert_\infty = \max_i \lvert v_i\rvert.$
The ℓ_2 vector norm,	$\lvert v_h\rvert_2 = \left(\sum_{i=0}^{N} v_i^2\right)^{1/2}.$
The discrete ℓ_2 norm,	$\lvert v_h\rvert_h = h^{1/2}\lvert v_h\rvert_2.$

The first fundamental theorem of calculus states that, given a function $v(x)$ such that v' is continuous on $L^2(0,1)$, then

$$\int_0^1 v'(x)dx = v(1) - v(0).$$

Proposition 2.1. [*First discrete fundamental theorem of calculus*] *For* $1 \leq i \leq N$, $h \sum_{j=1}^{i} \delta_{-h} v_j = v_i - v_0$.

Proof. Note that,

$$\begin{aligned} h\sum_{j=1}^{i} \delta_{-h} v_j &= h \sum_{j=1}^{i} \frac{v_j - v_{j-1}}{h} = \sum_{j=1}^{i} (v_j - v_{j-1}) \\ &= (v_1 - v_0) + (v_2 - v_1) + ... + (v_{i-1} - v_{i-2}) + (v_i - v_{i-1}) \\ &= v_i - v_0. \end{aligned}$$

□

The second fundamental theorem of calculus states that, given v a continuous function on a closed interval I with $0 \in I$, and x any point in I, then,

$$\frac{d}{dx}\int_0^x v(\tau)d\tau = v(x).$$

Proposition 2.2. [*Second discrete fundamental theorem of calculus*]. *Define* $V_i = h \sum_{j=0}^{i} v_j$, *for* $0 \leq i \leq N$. *Then,*

$$\delta_{-h} V_i = v_i, \quad 1 \leq i \leq N.$$

Proof. For $1 \leq i \leq N$

$$\delta_{-h} V_i = \frac{V_i - V_{i-1}}{h} = \frac{1}{h}\left(h\sum_{j=0}^{i} v_j - h\sum_{j=0}^{i-1} v_j \right) = \sum_{j=0}^{i} v_j - \sum_{j=0}^{i-1} v_j = v_i.$$

□

Definition 2.3. *We say* $v \in H_0^1(0,1)$ *if* $v, v' \in L^2(0,1)$ *and* $v(0) = v(1) = 0$.

Given $v \in H_0^1(0,1)$, then:

$$\sup_{x\in[0,1]} |v(x)|^2 \le \frac{1}{\pi}\left(\frac{1}{\pi}+2\right)\int_0^1 v'(x)^2 dx.$$

We wish to prove a discrete version of the above inequality.

Proposition 2.4. *Suppose $v_0 = 0 = v_N$. Then,*

$$|v_h|_\infty \le 2^{-1}h^{1/2}|\delta_h v_h|_2.$$

Proof. Assume $0 \le j \le N$. Then we have,

$$\begin{aligned}
v_j^2 &= (1-jh)v_j^2 + jhv_j^2 \\
&= (1-jh)h^2\left|\sum_{i=0}^{j-1}\frac{v_{i+1}-v_i}{h}\right|^2 + jhh^2\left|\sum_{i=j}^{N-1}\frac{v_{i+1}-v_i}{h}\right|^2 \\
&\le (1-jh)h^2\sum_{i=0}^{j-1}\left(\frac{v_{i+1}-v_i}{h}\right)^2\cdot\sum_{i=0}^{j-1}1^2 + jhh^2\sum_{i=j}^{N-1}\left(\frac{v_{i+1}-v_i}{h}\right)^2\cdot\sum_{i=j}^{N-1}1^2 \\
&= (1-jh)h^2\sum_{i=0}^{j-1}\left(\frac{v_{i+1}-v_i}{h}\right)^2 j + jhh^2\sum_{i=j}^{N-1}\left(\frac{v_{i+1}-v_i}{h}\right)^2(N-j) \\
&= \frac{N-j}{N}\frac{j}{N^2}\sum_{i=0}^{j-1}\left(\frac{v_{i+1}-v_i}{h}\right)^2 + \frac{j}{N}\frac{N-j}{N^2}\sum_{i=j}^{N-1}\left(\frac{v_{i+1}-v_i}{h}\right)^2 \\
&= \frac{j(N-j)}{N^3}\sum_{i=0}^{j-1}\left(\frac{v_{i+1}-v_i}{h}\right)^2 + \frac{j(N-j)}{N^3}\sum_{i=j}^{N-1}\left(\frac{v_{i+1}-v_i}{h}\right)^2 \\
&= \frac{j(N-j)}{N^3}|\delta_h v_h|_2^2. \qquad (2.1)
\end{aligned}$$

Hence,

$$\begin{aligned}
v_j^2 &\le \frac{j(N-j)}{N^3}|\delta_h v_h|_2^2 \\
&= \frac{4j(N-j)}{4N^3}|\delta_h v_h|_2^2 \\
&= \frac{N^2-(N-2j)^2}{4N^3}|\delta_h v_h|_2^2 \\
&\le \frac{N^2}{4N^3}|\delta_h v_h|_2^2 \\
&= 2^{-2}h|\delta_h v_h|_2^2.
\end{aligned}$$

Taking square root of both sides, and since the inequality holds for all j,

$$|v_h|_\infty \le 2^{-1}h^{1/2}|\delta_h v_h|_2. \qquad (2.2)$$

□

The Poincaré-Friedrichs inequality states that if $v \in H_0^1(0,1)$, then:

$$\int_0^1 v(x)^2 dx \leq \frac{1}{\pi^2} \int_0^1 v'(x)^2 dx.$$

Proposition 2.5. [*Discrete Poincaré-Friedrichs inequality*] *We consider a vector v_h so that $v_0 = 0 = v_N$ and as before $h = 1/N$. Then:*

$$|v_h|_2 \leq 6^{-1/2} |\delta_h v_h|_2.$$

Proof. We will use inequality (2.1) proven in the previous proposition. We have

$$\begin{aligned}
|v_h|_2^2 &= \sum_{j=0}^{N} v_j^2 \\
&\leq \sum_{j=0}^{N} \frac{j(N-j)}{N^3} |\delta_h v_h|_2^2 \\
&= \frac{|\delta_h v_h|_2^2}{N^3} \sum_{j=0}^{N} j(N-j) \\
&= \frac{|\delta_h v_h|_2^2}{N^3} \left(N \sum_{j=0}^{N} j - \sum_{j=0}^{N} j^2 \right) \\
&= \frac{|\delta_h v_h|_2^2}{N^3} \left(\left(N \frac{N^2+N}{2} \right) - \left(\frac{1}{3} N^3 + \frac{1}{2} N^2 + \frac{1}{6} N \right) \right) \\
&= \frac{|\delta_h v_h|_2^2}{N^3} \left(\frac{N^3}{2} + \frac{N^2}{2} - \frac{N^3}{3} - \frac{N^2}{2} - \frac{N}{6} \right) \\
&= \frac{|\delta_h v_h|_2^2}{N^3} \left(\frac{N^3}{6} - \frac{N}{6} \right) \\
&\leq \frac{|\delta_h v_h|_2^2}{N^3} \left(\frac{N^3}{6} \right) \\
&= \frac{|\delta_h v_h|_2^2}{6}.
\end{aligned}$$

We have shown,

$$|v_h|_2^2 \leq 6^{-1} |\delta_h v_h|_2^2.$$

Taking square root of both sides,

$$|v_h|_2 \leq 6^{-1/2} |\delta_h v_h|_2. \tag{2.3}$$

$\square$

The Poincaré inequality states that given $v \in H^1(0,1)$ with zero average i.e. $\int\limits_0^1 v(x) = 0$, then there exists a constant C, depending only on 1, so that.

$$\int\limits_0^1 v(x)^2 dx \leq C \int\limits_0^1 v'(x)^2 dx.$$

We shall now prove the discrete Poincaré inequality in one dimension.

Proposition 2.6. *[Discrete Poincaré inequality], Assume the vector v_h has zero average i.e, $\sum\limits_{j=0}^{N} v_j = 0$. Then $|v_h|_2 \leq 2^{-1/2}|\delta_h v_h|_2$.*

Proof. Let $u_i = v_i - v_0$ then $u_0 = 0$.

First we prove a small inequality to assist with the proof; assume $1 \leq i \leq N$. We have,

$$\begin{aligned}
u_i^2 &= h^2 \left| \sum_{j=0}^{i-1} \frac{u_{j+1} - u_j}{h} \right|^2 \\
&\leq h^2 \sum_{j=0}^{i-1} \left(\frac{u_{j+1} - u_j}{h} \right)^2 \cdot \sum_{j=0}^{i-1} 1 \\
&= i \cdot h^2 \sum_{j=0}^{i-1} \left(\frac{u_{j+1} - u_j}{h} \right)^2 \\
&\leq i \cdot h^2 \sum_{j=0}^{N-1} \left(\frac{u_{j+1} - u_j}{h} \right)^2.
\end{aligned}$$

Since the inequality above holds for all i,

$$\begin{aligned}
\sum_{i=0}^{N-1} u_i^2 &\leq h^2 \sum_{i=0}^{N-1} i \sum_{j=0}^{i-1} \left(\frac{u_{j+1} - u_j}{h} \right)^2 \\
&\leq h^2 \sum_{j=0}^{N-1} \left(\frac{u_{j+1} - u_j}{h} \right)^2 \cdot \sum_{i=0}^{N-1} i \\
&= h^2 \sum_{j=0}^{N-1} \left(\frac{u_{j+1} - u_j}{h} \right)^2 \frac{N(N-1)}{2} \\
&\leq h^2 \frac{N^2}{2} \sum_{j=0}^{N-1} \left(\frac{u_{j+1} - u_j}{h} \right)^2 \\
&= \frac{1}{2} \sum_{j=0}^{N-1} \left(\frac{u_{j+1} - u_j}{h} \right)^2.
\end{aligned}$$

We have proved that for the vector u_h we have defined, we have

$$\sum_{i=0}^{N-1} u_i^2 \leq \frac{1}{2} \sum_{i=0}^{N-1} \left(\frac{u_{i+1} - u_i}{h} \right)^2.$$

Now we switch back to the original vector v_h and get the inequality,

$$\sum_{i=0}^{N-1}(v_i - v_0)^2 \leq \frac{1}{2}\sum_{i=0}^{N-1}\left(\frac{v_{i+1}-v_i}{h}\right)^2.$$

We expand the left side of the above inequality,

$$\begin{aligned}
\sum_{i=1}^{N-1}(v_i - v_0)^2 &= \sum_{i=1}^{N-1} v_i^2 - 2v_0\sum_{i=1}^{N-1} v_i + (N-1)v_0^2 \\
&= \sum_{i=0}^{N-1} v_i^2 - 2v_0\sum_{i=1}^{N-1} v_i + (N-2)v_0^2 \\
&= \sum_{i=0}^{N-1} v_i^2 - 2v_0(-v_0 - v_N) + (N-2)v_0^2 \\
&= \sum_{i=0}^{N} v_i^2 - v_N^2 + 2v_0^2 + 2v_0v_N + (N-2)v_0^2 \\
&= \sum_{i=0}^{N} v_i^2 + v_N^2 + 2v_0v_N + v_0^2 - 2v_N^2 + (N-1)v_0^2 \\
&= \sum_{i=0}^{N} v_i^2 + (v_0 + v_N)^2 - 2v_N^2 + (N-1)v_0^2 \\
&\geq \sum_{i=0}^{N} v_i^2 - 2v_N^2 + (N-1)v_0^2.
\end{aligned}$$

We have shown,

$$\sum_{i=0}^{N} v_i^2 - 2v_N^2 + (N-1)v_0^2 \leq \frac{1}{2}\sum_{i=0}^{N-1}\left(\frac{v_{i+1}-v_i}{h}\right)^2.$$

Now we define a new vector $w_i = v_i - v_N$ Using the same operations as before, we get the similar inequality,

$$\sum_{i=0}^{N-1} w_i^2 \leq \frac{1}{2}\sum_{i=0}^{N-1}\left(\frac{w_{i+1}-w_i}{h}\right)^2.$$

Substituting vector w_i with $v_i - v_N$ we get,

$$\sum_{i=0}^{N-1}(v_i - v_N)^2 \leq \frac{1}{2}\sum_{i=0}^{N-1}\left(\frac{v_{i+1}-v_i}{h}\right)^2.$$

Expanding the left side of the inequality similarly as before we get

$$\sum_{i=0}^{N} v_i^2 - 2v_0^2 + (N-1)v_N^2 \leq \frac{1}{2}\sum_{i=0}^{N-1}\left(\frac{v_{i+1}-v_i}{h}\right)^2.$$

Adding the two inequalities we get

$$2\sum_{i=0}^{N} v_i^2 + (N-3)[v_0^2 + v_N^2] \leq \sum_{i=0}^{N-1}\left(\frac{v_{i+1}-v_i}{h}\right)^2.$$

Clearly for $N \geq 3$ we can conclude

$$\sum_{i=0}^{N} v_i^2 \leq \frac{1}{2} \sum_{i=0}^{N-1} \left(\frac{v_{i+1} - v_i}{h} \right)^2$$

or,

$$|v_h|_2^2 \leq \frac{1}{2} |\delta_h v_h|_2^2.$$

We will now prove the inequality for $N = 1$ and $N = 2$.

When $N = 1$,
$h = 1$, since we require 0 average on v_h and $v_0 + v_1 = 0$ so we get,

$$\begin{aligned}
|\delta_h v_h|_2^2 &= (v_1 - v_0)^2 \\
&= v_0^2 - 2v_0 v_1 + v_1^2 \\
&= v_0^2 + 2v_0^2 + v_0^2 \\
&= 4v_0^2 \\
&= 2(v_0^2 + v_1^2) \\
&= 2|v_h|_2^2.
\end{aligned}$$

When $N = 2$,
$h = \frac{1}{2}$, and $v_0 + v_1 + v_2 = 0$ since we require 0 average on v_h.

$$\begin{aligned}
|\delta_h v_h|_2^2 &= \left(\frac{v_1 - v_0}{1/2} \right)^2 + \left(\frac{v_2 - v_1}{1/2} \right)^2 \\
&= 4[(v_1 - v_0)^2 + (v_2 - v_1)^2] \\
&= 4[v_0^2 - 2v_0 v_1 + v_1^2 + v_1^2 - 2v_1 v_2 + v_2^2] \\
&= 2v_0^2 + 2v_1^2 + 2v_2^2 + 2v_0^2 + 6v_1^2 + 2v_2^2 - 8v_0 v_1 - 8v_1 v_2 \\
&= 2(v_0^2 + v_1^2 + v_2^2) + 2v_0^2 + 6v_1^2 + 2(v_0 + v_1)^2 - 8v_0 v_1 - 8v_1[-(v_0 + v_1)] \\
&= 2(v_0^2 + v_1^2 + v_2^2) + 2v_0^2 + 6v_1^2 + 2(v_0 + v_1)^2 - 8v_0 v_1 + 8v_1 v_0 + 8v_1^2 \\
&= 2(v_0^2 + v_1^2 + v_2^2) + 2v_0^2 + 14v_1^2 + 2(v_0 + v_1)^2 \\
&\geq 2(v_0^2 + v_1^2 + v_2^2) \\
&= 2|v_h|_2^2.
\end{aligned}$$

The inequality holds for $N = 1$ and $N = 2$ and for $N \geq 3$, thus we have proven that for all N,

$$|v_h|_2 \leq 2^{-1/2} |\delta_h v_h|_2. \tag{2.4}$$

□

Hardy's inequality says that if $v \in H_0^1(0,1)$, then there is some positive constant c such that

$$\int_0^1 \frac{v^2(x)}{d(x)^2}dx \le c\int_0^1 (v'(x))^2 dx,$$

where $d(x)$ is the minimum distance to the boundary, i.e., $d(x) = min(x, 1-x)$.

Proposition 2.7. [*Discrete Hardy's inequality*]. *Suppose* $v_0 = 0 = v_N$. *Then*

$$\sum_{j=1}^{N-1} \frac{v_j^2}{h^2 min^2(j, N-j)} \le 4|\delta_h v_h|_2^2.$$

Proof. We shall prove a small inequality which we will use later for proving the proposition. If $a_i \ge 0$ define $A_i = \sum\limits_{j=1}^{i} a_j$, then

$$\sum_{i=1}^{N}\left(\frac{A_i}{i}\right)^2 \le 4\sum_{i=1}^{N} a_i^2. \tag{2.5}$$

We begin by demonstrating two inequalities which will be used in the proof.

$$\begin{aligned}\left(\frac{A_i}{i}\right)^2 &= \left(a_i + \frac{A_i}{i} - a_i\right)^2 \\ &\le 2a_i^2 + 2\left(\frac{A_i}{i} - a_i\right)^2 \\ &= 4a_i^2 - \frac{4a_iA_i}{i} + \frac{2A_i^2}{i^2}.\end{aligned} \tag{2.6}$$

Also,

$$\begin{aligned}-2A_ia_i &= -(2A_i - a_i)a_i - a_i^2 \\ &= -(A_i + A_{i-1})(A_i - A_{i-1}) - a_i^2 \\ &= -(A_i^2 - A_{i-1}^2) - a_i^2 \\ &\le -(A_i^2 - A_{i-1}^2)\end{aligned} \tag{2.7}$$

Using the two inequalities (2.6) and (2.7) on (2.5) we get,

$$\begin{aligned}\sum_{i=1}^{N}\left(\frac{A_i}{i}\right)^2 &\le 4\sum_{i=1}^{N} a_i^2 - 4\sum_{i=1}^{N}\frac{a_iA_i}{i} + 2\sum_{i=1}^{N}\left(\frac{A_i}{i}\right)^2 \\ &\le 4\sum_{i=1}^{N} a_i^2 - 2\sum_{i=1}^{N}\frac{A_i^2 - A_{i-1}^2}{i} + 2\sum_{i=1}^{N}\left(\frac{A_i}{i}\right)^2 \\ &= 4\sum_{i=1}^{N} a_i^2 - 2\left(\frac{A_1^2}{1} + \frac{A_2^2}{2\cdot 3} + ... + \frac{A_{n-1}^2}{(n-1)n} + \frac{A_n^2}{n^2}\right) + 2\sum_{i=1}^{N}\left(\frac{A_i}{i}\right)^2 \\ &\le 4\sum_{i=1}^{N} a_i^2 - 2\sum_{i=1}^{N}\frac{A_i^2}{i(i+1)} + 2\sum_{i=1}^{N}\left(\frac{A_i}{i}\right)^2 \\ &\le 4\sum_{i=1}^{N} a_i^2.\end{aligned}$$

Now we start the proof of the discrete Hardy's inequality. First we state a simple equality:

$$v_i^2 = h^2 \left| \sum_{j=1}^{i} \frac{v_j - v_{j-1}}{h} \right|^2, \qquad 1 \le i \le N.$$

We sum the first $\left\lfloor \frac{N-1}{2} \right\rfloor$ indices. Using (2.5)

$$\begin{aligned}
\sum_{i=1}^{\left\lfloor \frac{N-1}{2} \right\rfloor} \frac{v_i^2}{h^2 i^2} &\le \sum_{i=1}^{\left\lfloor \frac{N-1}{2} \right\rfloor} \frac{\left| \sum_{j=1}^{i} \frac{v_j - v_{j-1}}{h} \right|^2}{i^2} \\
&\le \sum_{i=1}^{\left\lfloor \frac{N-1}{2} \right\rfloor} \sum_{j=1}^{i} \frac{\left| \frac{v_j - v_{j-1}}{h} \right|^2}{i^2} \\
&\le 4 \sum_{i=1}^{\left\lfloor \frac{N-1}{2} \right\rfloor} \left(\frac{v_i - v_{i-1}}{h} \right)^2.
\end{aligned}$$

We get,

$$\sum_{i=1}^{\left\lfloor \frac{N-1}{2} \right\rfloor} \frac{v_i^2}{h^2 i^2} \le 4 \sum_{i=1}^{\left\lfloor \frac{N-1}{2} \right\rfloor} \left(\frac{v_i - v_{i-1}}{h} \right)^2. \tag{2.8}$$

Now we sum the last $\left\lfloor \frac{N-1}{2} \right\rfloor - 1$ indices.

$$\begin{aligned}
\sum_{i=\left\lfloor \frac{N-1}{2} \right\rfloor + 1}^{N-1} \frac{v_i^2}{h^2 (N-i)^2} &= \sum_{i=1}^{N - \left\lfloor \frac{N-1}{2} \right\rfloor - 1} \frac{v_{N-i}^2}{h^2 i^2} \\
&\le \sum_{i=1}^{N - \left\lfloor \frac{N-1}{2} \right\rfloor - 1} 4 \left(\frac{v_{N-i+1} - v_{N-i}}{h} \right)^2 \\
&= 4 \sum_{i=\left\lfloor \frac{N-1}{2} \right\rfloor + 1}^{N} \left(\frac{v_i - v_{i-1}}{h} \right)^2.
\end{aligned}$$

We get,

$$\sum_{i=\left\lfloor \frac{N-1}{2} \right\rfloor + 1}^{N-1} \frac{v_i^2}{h^2 (N-i)^2} \le 4 \sum_{i=\left\lfloor \frac{N-1}{2} \right\rfloor + 1}^{N} \left(\frac{v_i - v_{i-1}}{h} \right)^2. \tag{2.9}$$

Adding (2.8) and (2.9) together we get the discrete Hardy's inequality,

$$\sum_{i=1}^{N-1} \frac{v_j^2}{h^2 min^2(i, N-i)} \le 4 \sum_{i=1}^{N} \left(\frac{v_i - v_{i-1}}{h} \right)^2. \tag{2.10}$$

$\square$

Given $v, v' \in L^2(0,1)$ we can get the inequality

$$\sup_{x\in[0,1]} v(x)^2 \leq c \int_0^1 (v(x)^2 + v'(x)^2)dx.$$

where c is a positive real number independent of v and v'.

Proposition 2.8. [*Inequality between the discrete infinity norm and the discrete H^1 - norm*]. *We have,*

$$|v_h|_\infty^2 \leq \frac{2h}{\sqrt{5}-1}|v_h|_{H^1}^2, \quad \textit{where} \quad |v_h|_{H^1}^2 = |v_h|_2^2 + |\delta_h v_h|_2^2.$$

Proof. Assume that,

$$|v_h|_\infty^2 = v_k^2, \quad \text{and} \quad \min_{0\leq i\leq N} v_i^2 = v_m^2, \qquad \text{where} \quad 0 \leq k, m \leq N.$$

Without loss of generality assume $m \leq k$. Let ϵ be any positive number. Then

$$\begin{aligned}
|v_h|_\infty^2 &= v_k^2 = v_m^2 + h\sum_{i=m}^{k-1} \frac{v_{i+1}^2 - v_i^2}{h} \\
&= v_m^2 + h\sum_{i=m}^{k-1}\left(\frac{v_{i+1}-v_i}{h}\right)(v_{i+1}+v_i) \\
&\leq v_m^2 + h\Bigg[\sum_{i=m}^{k-1}\left(\frac{v_{i+1}-v_i}{h}\right)v_{i+1} + \sum_{i=m}^{k-1}\left(\frac{v_{i+1}-v_i}{h}\right)v_i\Bigg] \\
&\leq v_m^2 + h\Bigg[\frac{1}{2\epsilon}\sum_{i=m}^{k-1}\left(\frac{v_{i+1}-v_i}{h}\right)^2 + \frac{\epsilon}{2}\sum_{i=m}^{k-1} v_{i+1}^2 + \frac{1}{2\epsilon}\sum_{i=m}^{k-1}\left(\frac{v_{i+1}-v_i}{h}\right)^2 + \frac{\epsilon}{2}\sum_{i=m}^{k-1} v_i^2\Bigg] \\
&\leq h\sum_{i=0}^{N} v_i^2 + h\Bigg[\frac{1}{2\epsilon}\sum_{i=0}^{N-1}\left(\frac{v_{i+1}-v_i}{h}\right)^2 + \frac{\epsilon}{2}\sum_{i=0}^{N} v_i^2 + \frac{1}{2\epsilon}\sum_{i=0}^{N-1}\left(\frac{v_{i+1}-v_i}{h}\right)^2 + \frac{\epsilon}{2}\sum_{i=0}^{N} v_i^2\Bigg] \\
&= h\Bigg[\frac{1}{\epsilon}\sum_{i=0}^{N-1}\left(\frac{v_{i+1}-v_i}{h}\right)^2 + (1+\epsilon)\sum_{i=0}^{N} v_i^2\Bigg]
\end{aligned}$$

For (2.8) to hold we require,

$$1+\epsilon \leq \frac{2}{\sqrt{5}-1} \quad \text{and} \quad \frac{1}{\epsilon} \leq \frac{2}{\sqrt{5}-1}.$$

For the inequalities above to hold we find that,

$$\epsilon = \frac{\sqrt{5}-1}{2}.$$

Choosing such ϵ will give the smallest constant and prove the proposition.

$$|v_h|_\infty^2 \leq \frac{2h}{\sqrt{5}-1}\Bigg[\sum_{j=0}^{N} v_j^2 + \sum_{j=0}^{N-1}\left(\frac{v_{j+1}-v_j}{h}\right)^2\Bigg].$$

$\square$

For any $v \in H^1(0,1)$ define,

$$\|v\|^2 = \int_0^1 v'(x)^2 dx + v(0)^2 + v(1)^2.$$

It is known that $\|\cdot\|$ is equivalent to the H^1 norm.

We will prove a discrete version of the inequality above.

Proposition 2.9. *We have,*

$$|v_h|_{H^1}^2 \leq 2(|v_h|_{H^1}^*)^2, \quad \textit{where} \quad (|v_h|_{H^1}^*)^2 = \frac{v_0^2 + v_N^2}{h} + |\delta_h v_h|_2^2.$$

Proof. Assume $0 \leq i \leq N$. For any $\epsilon > 0$,

$$\begin{aligned}
v_i^2 &= v_0^2 + h\sum_{j=0}^{i-1} \frac{v_{j+1}^2 - v_j^2}{h} \\
&= v_0^2 + h\sum_{j=0}^{i-1} \left(\frac{v_{j+1} - v_j}{h}\right)(v_j + v_{j+1}) \\
&= v_0^2 + h\sum_{j=0}^{i-1} v_j \left(\frac{v_{j+1} - v_j}{h}\right) + h\sum_{j=0}^{i-1} v_{j+1}\left(\frac{v_{j+1} - v_j}{h}\right) \\
&\leq v_0^2 + \frac{h}{2}\left[\epsilon\sum_{j=0}^{i-1} v_j^2 + \frac{1}{\epsilon}\sum_{j=0}^{i-1}\left(\frac{v_{j+1} - v_j}{h}\right)^2\right] + \frac{h}{2}\left[\epsilon\sum_{j=0}^{i-1} v_{j+1}^2 + \frac{1}{\epsilon}\sum_{j=0}^{i-1}\left(\frac{v_{j+1} - v_j}{h}\right)^2\right] \\
&\leq v_0^2 + \frac{h}{2}\left[\epsilon\sum_{j=0}^{i-1} v_j^2 + \epsilon\sum_{j=0}^{i-1} v_{j+1}^2 + \frac{2}{\epsilon}\sum_{j=0}^{i-1}\left(\frac{v_{j+1} - v_j}{h}\right)^2\right].
\end{aligned}$$

We get,

$$v_i^2 \leq v_0^2 + \frac{h}{2}\left[\epsilon\sum_{j=0}^{i-1} v_j^2 + \epsilon\sum_{j=0}^{i-1} v_{j+1}^2 + \frac{2}{\epsilon}\sum_{j=0}^{i-1}\left(\frac{v_{j+1} - v_j}{h}\right)^2\right]. \tag{2.11}$$

Similarly, from the fact that

$$v_i^2 = v_N^2 - h\sum_{j=i}^{N-1} \frac{v_{j+1}^2 - v_j^2}{h},$$

we can get

$$v_i^2 \leq v_N^2 + \frac{h}{2}\left[\epsilon\sum_{j=i}^{N-1} v_j^2 + \epsilon\sum_{j=i}^{N-1} v_{j+1}^2 + \frac{2}{\epsilon}\sum_{j=i}^{N-1}\left(\frac{v_{j+1} - v_j}{h}\right)^2\right]. \tag{2.12}$$

Adding (2.11) and (2.12) we get,

$$\begin{aligned}
2v_i^2 &\le v_0^2 + \frac{h}{2}\left[\epsilon\sum_{j=0}^{i-1} v_j^2 + \epsilon\sum_{j=0}^{i-1} v_{j+1}^2 + \frac{2}{\epsilon}\sum_{j=0}^{i-1}\left(\frac{v_{j+1}-v_j}{h}\right)^2\right] \\
&\quad + v_N^2 + \frac{h}{2}\left[\epsilon\sum_{j=i}^{N-1} v_j^2 + \epsilon\sum_{j=i}^{N-1} v_{j+1}^2 + \frac{2}{\epsilon}\sum_{j=i}^{N-1}\left(\frac{v_{j+1}-v_j}{h}\right)^2\right] \\
&= v_0^2 + v_N^2 + \frac{h}{2}\left[\epsilon\sum_{j=0}^{N-1} v_j^2 + \epsilon\sum_{j=0}^{N-1} v_{j+1}^2 + \frac{2}{\epsilon}\sum_{j=0}^{N-1}\left(\frac{v_{j+1}-v_j}{h}\right)^2\right] \\
&\le v_0^2 + v_N^2 + \frac{h}{2}\left[\epsilon\sum_{j=0}^{N} u_j^2 + \epsilon\sum_{j=0}^{N} v_j^2 + \frac{2}{\epsilon}\sum_{j=0}^{N-1}\left(\frac{v_{j+1}-v_j}{h}\right)^2\right] \\
&= v_0^2 + v_N^2 + \frac{h}{2}\left[2\epsilon\sum_{j=0}^{N} v_j^2 + \frac{2}{\epsilon}\sum_{j=0}^{N-1}\left(\frac{v_{j+1}-v_j}{h}\right)^2\right].
\end{aligned}$$

Dividing by two we get the inequality,

$$v_i^2 \le \frac{v_0^2+v_N^2}{2} + \frac{h}{2}\left[\epsilon\sum_{j=0}^{N} v_j^2 + \frac{1}{\epsilon}\sum_{j=0}^{N-1}\left(\frac{v_{j+1}-v_j}{h}\right)^2\right]. \tag{2.13}$$

Notice that inequality (2.13) is independent of i, we will use this property to prove the original proposition.

$$\begin{aligned}
\sum_{i=0}^{N} v_i^2 &= v_0^2 + \sum_{i=1}^{N} v_i^2 \\
&\le v_0^2 + \left(\frac{v_0^2+v_N^2}{2} + \frac{h}{2}\left[\epsilon\sum_{j=0}^{N} v_j^2 + \frac{1}{\epsilon}\sum_{j=0}^{N-1}\left(\frac{v_{j+1}-v_j}{h}\right)^2\right]\right)\sum_{i=1}^{N} 1 \\
&\le v_0^2 + \left(\frac{v_0^2+v_N^2}{2} + \frac{h}{2}\left[\epsilon\sum_{j=0}^{N} v_j^2 + \frac{1}{\epsilon}\sum_{j=0}^{N-1}\left(\frac{v_{j+1}-v_j}{h}\right)^2\right]\right)\frac{1}{h} \\
&= \frac{(2h+1)v_0^2+v_N^2}{2h} + \frac{1}{2}\left[\epsilon\sum_{j=0}^{N} v_j^2 + \frac{1}{\epsilon}\sum_{j=0}^{N-1}\left(\frac{v_{j+1}-v_j}{h}\right)^2\right].
\end{aligned}$$

We can change index i to j on the left side of the inequality for simplicity,

$$\sum_{j=0}^{N} v_j^2 \le \frac{(2h+1)v_0^2+v_N^2}{2h} + \frac{1}{2}\left[\epsilon\sum_{j=0}^{N} v_j^2 + \frac{1}{\epsilon}\sum_{j=0}^{N-1}\left(\frac{v_{j+1}-v_j}{h}\right)^2\right]. \tag{2.14}$$

Similarly, from the fact that

$\sum\limits_{i=0}^{N} v_i^2 = v_N^2 + \sum\limits_{i=0}^{N-1} v_i^2,$

we can get

$$\sum_{j=0}^{N} v_j^2 \le \frac{v_0^2+(2h+1)v_N^2}{2h} + \frac{1}{2}\left[\epsilon\sum_{j=0}^{N} v_j^2 + \frac{1}{\epsilon}\sum_{j=0}^{N-1}\left(\frac{v_{j+1}-v_j}{h}\right)^2\right]. \tag{2.15}$$

Adding (2.14) and (2.15) together, we have

$$2\sum_{i=0}^{N} v_i^2 \leq \frac{(2h+2)v_0^2 + (2h+2)v_N^2}{2h} + \epsilon \sum_{j=0}^{N} v_j^2 + \frac{1}{\epsilon}\sum_{j=0}^{N-1}\left(\frac{v_{j+1}-v_j}{h}\right)^2.$$

Dividing by two we get the inequality,

$$\sum_{i=0}^{N} v_i^2 \leq \frac{(h+1)v_0^2 + (h+1)v_N^2}{2h} + \frac{1}{2}\left[\epsilon \sum_{j=0}^{N} v_j^2 + \frac{1}{\epsilon}\sum_{j=0}^{N-1}\left(\frac{v_{j+1}-v_j}{h}\right)^2\right].$$

Moving the ℓ_2 norm term from the right side to the left we get,

$$\left(1-\frac{\epsilon}{2}\right)\sum_{j=0}^{N} v_j^2 \leq \frac{(h+1)(v_0^2+v_N^2)}{2h} + \frac{1}{2\epsilon}\sum_{j=0}^{N-1}\left(\frac{v_{j+1}-v_j}{h}\right)^2.$$

From that we may get,

$$\sum_{j=0}^{N} v_j^2 \leq \left(1-\frac{\epsilon}{2}\right)^{-1}\left[\frac{(h+1)(v_0^2+v_N^2)}{2h} + \frac{1}{2\epsilon}\sum_{j=0}^{N-1}\left(\frac{v_{j+1}-v_j}{h}\right)^2\right].$$

In order to get the desired inequality constant we require,

$$\frac{h+1}{2}\left(1-\frac{\epsilon}{2}\right)^{-1} \leq 2 \quad \text{and} \quad \frac{1}{2\epsilon}\left(1-\frac{\epsilon}{2}\right)^{-1} \leq 1.$$

We can get this constant when $\epsilon = 1$ and $h \leq 1$ (which holds in the general case), so the inequality becomes,

$$\sum_{j=0}^{N} v_j^2 \leq 2\frac{v_0^2+v_N^2}{h} + \sum_{j=0}^{N-1}\left(\frac{v_{j+1}-v_j}{h}\right)^2.$$

Adding $|\delta_h v_h|_2^2$ to both sides we get,

$$\sum_{j=0}^{N} v_j^2 + \sum_{j=0}^{N-1}\left(\frac{v_{j+1}-v_j}{h}\right)^2 \leq 2\left[\frac{v_0^2+v_N^2}{h} + \sum_{j=0}^{N-1}\left(\frac{v_{j+1}-v_j}{h}\right)^2\right]. \tag{2.16}$$

□

For a smooth function f on $[0,1]$ we cannot bound $\int_0^1 f'(x)^2dx$ in terms of $\int_0^1 f(x)^2dx$ by a constant independent of f. For example if we take the function $f(x) = \sin(nx)$ the derivative is $f'(x) = n \cdot \cos(nx)$, and we can easily see that as n grows so will the bounding constant of the inequality. Therefore it cannot be independent of f. However, such an inequality is possible in the discrete case with a constant depending on h. These so called inverse inequalities are useful in estimating the condition numbers of discrete differential operators, as in the following results some of which can be found in [7].

Proposition 2.10. [*Inverse estimates*]. *We have,*

(*a*) $|\delta_h v_h|_2 \leq 2h^{-1}|v_h|_2$.

(*b*) $|\delta_h v_h|_\infty \leq \sqrt{2}h^{-1}|v_h|_2$.

(*c*) $|\delta_h v_h|_\infty \leq 2h^{-1}|v_h|_\infty$.

Proof.

(*a*) We have,

$$
\begin{aligned}
|\delta_h v_h|_2^2 &= \sum_{j=0}^{N-1}\left(\frac{v_{j+1}-v_j}{h}\right)^2 \\
&= \frac{1}{h^2}\left(\sum_{j=0}^{N-1} v_{j+1}^2 - 2\sum_{j=0}^{N-1} v_{j+1}v_j + \sum_{j=0}^{N-1} v_j^2\right) \\
&= \frac{1}{h^2}\left(\sum_{j=0}^{N} v_j^2 - v_0^2 - 2\sum_{j=0}^{N-1} v_{j+1}v_j + \sum_{j=0}^{N} v_j^2 - v_N^2\right) \\
&= \frac{1}{h^2}\left(2\sum_{j=0}^{N} v_j^2 - (v_N^2+v_0^2) - 2\sum_{j=0}^{N-1} v_{j+1}v_j\right) \\
&= \frac{1}{h^2}\left(2\sum_{j=0}^{N} v_j^2 - (v_N^2+v_0^2) - 2\sum_{j=0}^{N-1} v_{j+1}v_j + 2\sum_{j=0}^{N} v_j^2 - 2\sum_{j=0}^{N} v_j^2\right) \\
&= \frac{1}{h^2}\left(4\sum_{j=0}^{N} v_j^2 - \left[v_N^2+v_0^2 + 2\sum_{j=0}^{N} v_j^2 + 2\sum_{j=0}^{N-1} v_{j+1}v_j\right]\right) \\
&= \frac{1}{h^2}\left(4\sum_{j=0}^{N} v_j^2 - \left[v_N^2+v_0^2 + \sum_{j=1}^{N} v_j^2 + v_0^2 + \sum_{j=0}^{N-1} v_j^2 + v_N^2 + 2\sum_{j=0}^{N-1} v_{j+1}v_j\right]\right) \\
&= \frac{1}{h^2}\left(4\sum_{j=0}^{N} v_j^2 - \left[2(v_0^2+v_N^2) + \sum_{j=0}^{N-1} v_{j+1}^2 + \sum_{j=0}^{N-1} v_j^2 + 2\sum_{j=0}^{N-1} v_{j+1}v_j\right]\right) \\
&= \frac{1}{h^2}\left(4\sum_{j=0}^{N} v_j^2 - \left[2(v_0^2+v_N^2) + \sum_{j=0}^{N-1} (v_{j+1}-v_j)^2\right]\right) \\
&\leq \frac{4}{h^2}\sum_{j=0}^{N} v_j^2.
\end{aligned}
$$

We proved,

$$|\delta_h v_h|_2^2 \leq 4h^{-2}|v_h|_2^2.$$

Taking the square root of both sides we have proved,

$$|\delta_h v_h|_2 \leq 2h^{-1}|v_h|_2. \tag{2.17}$$

(b) Assume that the infinity norm occurs at index k. Then

$$\begin{aligned}
|\delta_h v_h|_\infty^2 &= max_j \left(\frac{v_{j+1} - v_j}{h}\right)^2 \\
&= \frac{1}{h^2}(v_{k+1} - v_k)^2 \\
&= \frac{1}{h^2}(v_{k+1}^2 + v_k^2 - 2v_{k+1}v_k) \\
&= \frac{1}{h^2}(2(v_{k+1}^2 + v_k^2) - (2v_{k+1}v_k + v_{k+1}^2 + v_k^2)) \\
&= \frac{1}{h^2}(2(v_{k+1}^2 + v_k^2) - (v_{k+1} + v_k)^2) \\
&\leq \frac{2}{h^2}(v_{k+1}^2 + v_k^2) \\
&\leq \frac{2}{h^2}|v_h|_2^2.
\end{aligned}$$

Taking the square root of both sides,

$$|\delta_h v_h|_\infty \leq \sqrt{2}h^{-1}|v_h|_2. \tag{2.18}$$

We can see the inequality is sharp if we choose the vector $v_h = (-1, 1, 0..., 0)$.

(c) Note that from (b),

$$\begin{aligned}
|\delta_h v_h|_\infty^2 &\leq \frac{2}{h^2}(v_{k+1}^2 + v_k^2) \\
&\leq \frac{4}{h^2}|v_h|_\infty^2.
\end{aligned}$$

Taking the square root of both sides,

$$|\delta_h v_h|_\infty \leq \ 2h^{-1}|v_h|_\infty. \tag{2.19}$$

We can see the inequality is sharp if we choose the vector $v_h = (-1, 1, 0, ..., 0)$. □

2.2 Second Order Numerical Schemes for the 1D Poisson equation

When we wish to solve a partial differential equation there is no guarantee we can find an analytical solution so we create a numerical scheme that mimic the continues operator and since we transformed the continues problem into a discrete one we can use numerical methods to approximate the solution. For a finite difference numerical scheme to be successful it needs to be convergent, which is relatively easy to prove if we know a scheme is stable and consistent. We now define each of those terms with respect to the Poisson equation.
Consider the simplest finite difference scheme with a uniform grid size h

$$-\Delta_h^* u_h = f_h = R_h f,$$

where $(-\Delta_h^*)$ is any finite difference discretization of the laplacian and R_h is the restriction operator on Ω_h that "samples" a function such as f on Ω_h and converts it into a vector. As in the previous chapter the term c is a positive constant independent of u and h.

Definition 2.11. *The term $\|\cdot\|^*_{C^r(\overline{\Omega})}$ is defined as,*

$$\|u\|^*_{C^r(\overline{\Omega})} = \max_{|\alpha|=r} \sup_{x\in\overline{\Omega}} |D^\alpha u(x)|,$$

where $D^\alpha u$ represent a derivative of u of order r, with α a multi index such that $r = |\alpha| = \alpha_1 + \alpha_2$.

For example, if $\alpha = (3, 1)$ then $D^\alpha u$ represents the fourth derivative u_{xxxy}.
While $||\cdot||^*_{C^r(\overline{\Omega})}$ is not a norm, it is a useful quantity in the analysis of finite difference schemes.

Definition 2.12. *The discretization $(-\Delta_h^*)$ is said to be **consistent of order r** if*

$$|\Delta_h^* R_h u - R_h \Delta u|_\infty \leq c\|u\|^*_{C^{r+2}(\overline{\Omega})} h^r, \quad \textit{for} \quad u \in C^{r+2}(\overline{\Omega}), \quad \textit{where } u = 0 \textit{ on } \partial\Omega.$$

Definition 2.13. *The ℓ_2 norm of a matrix A is defined as $|A|_2 = \sup_{x\neq 0} \frac{|Ax|_2}{|x|_2}$.*

Definition 2.14. *The discretization $(-\Delta_h^*)$ is said to be **stable** with respect to $|\cdot|_2$ if $|(\Delta_h^*)^{-1}|_2$ is bounded independently of h.*

Definition 2.15. *Given the scheme $-\Delta_h^* u_h = R_h f$ for solving the Poisson equation $-\Delta u = f$ with boundary condition $u = 0$ on $\partial\Omega$, we say $(-\Delta_h^*)$ to be **convergent of order r** if*

$$|R_h u - u_h|_\infty \leq c\|u\|^*_{C^{r+2}(\overline{\Omega})} h^r.$$

Definition 2.16. *The **condition number** of a numerical scheme represented as a square non-singular matrix A is defined as:*

$$\kappa(A) = |A^{-1}|_2 |A|_2.$$

In the next two subsections we use the inequalities of the last section to show convergence of the finite difference schemes.

2.2.1 Solving the 1D Poisson Equation with Dirichlet Boundary Conditions.

Consider the boundary value problem on domain $\Omega = (0,1)$:

$$-u'' = f \quad \text{on} \quad (0,1) \quad \text{with} \quad u(0) = 0 = u(1).$$

Consider the simplest finite difference scheme with a uniform grid size h

$$-\Delta_h u_h = f_h = R_h f.$$

Since $u(0) = 0 = u(1)$ we only need to find u_i for $1 \leq i \leq N-1$.
We define the grid as

$$\Omega_h = \Big\{ ih, \quad 1 \leq i \leq N-1 \Big\}.$$

For $1 \leq i \leq N-1$ the equivalent $(N-1) \times (N-1)$ matrix to the discretization grid is simply,

$$-\Delta_h = \frac{1}{h^2} \begin{bmatrix} 2 & -1 & & & \\ -1 & 2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{bmatrix}. \tag{2.20}$$

Theorem 2.17. *The Δ_h scheme is second order consistent with respect to $|\cdot|_\infty$.*

Proof. For $1 \leq i \leq N-1$, $u_i = u(x_i)$ can be expanded using Taylor's theorem,

$$u_{i\pm1} = u_i \pm \frac{du_i}{dx}h + \frac{d^2u_i}{dx^2}\frac{h^2}{2} \pm \frac{d^3u_i}{dx^3}\frac{h^3}{6} + \frac{d^4u(\xi^\pm)}{dx^4}\frac{h^4}{24},$$

for some $\xi^\pm \in \Omega$. Adding the two equalities above we get

$$u_{i+1} + u_{i-1} = 2u_i + \frac{d^2u_i}{dx^2}h^2 + \frac{d^4u(\xi^\pm)}{dx^4}\frac{h^4}{12}.$$

By rearranging the terms we get

$$\frac{-2u_i + u_{i+1} + u_{i-1}}{h^2} - \frac{d^2u_i}{dx^2} =: E, \quad \text{and} \quad \|E\|^*_{C^4(\overline{\Omega})} \leq c\|u\|^*_{C^4(\overline{\Omega})}h^2,$$

where c is a positive constant. Since we chose any $1 \leq i \leq N-1$ we may conclude that,

$$|\Delta_h R_h u - R_h \Delta u|_\infty \leq c\|u\|^*_{C^4(\overline{\Omega})}h^2. \tag{2.21}$$

$\square$

Theorem 2.18. *The Δ_h scheme is second order convergent with respect to $|\cdot|_\infty$, $|\cdot|_h$, and convergent of order $\frac{3}{2}$ with respect to $|\cdot|_{H^1}$ and $|\cdot|^*_{H^1}$.*

Proof. Let $e_h = R_h u - u_h$. It is easily seen that $e_0 = 0 = e_N$. Now we prove a summation by parts equality. Note that,

$$\begin{aligned}
-h^2 e_h^T \Delta_h e_h &= \sum_{i=1}^{N-1} e_i(2e_i - e_{i-1} - e_{i+1}) \\
&= \sum_{i=1}^{N-1} e_i^2 - \sum_{i=1}^{N-1} e_i e_{i-1} + \sum_{i=1}^{N-1} e_i^2 - \sum_{i=1}^{N-1} e_i e_{i+1} \\
&= \sum_{i=0}^{N-1} e_i^2 - 2\sum_{i=0}^{N-1} e_{i+1} e_i + \sum_{i=0}^{N-1} e_{i+1}^2 \\
&= \sum_{i=0}^{N-1} (e_{i+1} - e_i)^2.
\end{aligned}$$

We can summarize the summation by parts equality as

$$-e_h^T \Delta_h e_h = |\delta_h e_h|_2^2.$$

From that we may get the inequality,

$$|\delta_h e_h|_2^2 \leq |e_h|_2 |\Delta_h e_h|_2.$$

Using (2.3) we get

$$|\delta_h e_h|_2^2 \leq \frac{1}{6} |\delta_h e_h|_2 |\Delta_h e_h|_2,$$

$$|\delta_h e_h|_2 \leq \frac{1}{6} |\Delta_h e_h|_2,$$

or

$$|\delta_h e_h|_h \leq \frac{1}{6} |\Delta_h e_h|_h. \tag{2.22}$$

Since $\Delta_h e_h = \Delta_h R_h u - R_h \Delta u$, from (2.21) can write,

$$|\Delta_h e_h|_\infty \leq c\|u\|^*_{C^4(\overline{\Omega})} h^2.$$

Clearly,

$$|\Delta_h e_h|_2 \leq N^{1/2} |\Delta_h e_h|_\infty \leq h^{-1/2} c\|u\|^*_{C^4(\overline{\Omega})} h^2.$$

From the above equality we can see that

$$h^{1/2} |\Delta_h e_h|_2 \leq c\|u\|^*_{C^4(\overline{\Omega})} h^2.$$

We may conclude that,

$$|\Delta_h e_h|_h \leq c\|u\|^*_{C^4(\overline{\Omega})} h^2.$$

Plugging that into equation (2.22) we get

$$|\delta_h e_h|_h \leq c\|u\|^*_{C^4(\overline{\Omega})} h^2.$$

Using Proposition 2.5, again we can bound $|e_h|_h$ as follows,

$$|e_h|_h \leq \frac{1}{6}|\delta_h e_h|_h \leq c\|u\|^*_{C^4(\overline{\Omega})} h^2. \tag{2.23}$$

In addition, using (2.2) we can bound $|e_h|_\infty$ as follows,

$$|e_h|_\infty \leq\ 2^{-1}h^{1/2}|\delta_h e_h|_2 = 2^{-1}|\delta_h e_h|_h \leq c\|u\|^*_{C^4(\overline{\Omega})} h^2. \tag{2.24}$$

Equations (2.23) and (2.24) are proofs that the scheme Δ_h is convergent of second order with respect to $|\cdot|_h$ and $|\cdot|_\infty$ respectively.

We wish to prove now that the scheme is convergent with respect to the discrete H^1 norm and the equivalent H^1 norm from equation (2.16).

$$(|e_h|^*_{H^1})^2 = \frac{e_0^2 + e_N^2}{h} + \sum_{j=0}^{N-1}\left(\frac{e_{j+1} - e_j}{h}\right)^2.$$

As was stated before, $e_0 = 0 = e_N$, giving us

$$\begin{aligned}
|e_h|^*_{H^1} &= \left[\sum_{j=0}^{N-1}\left(\frac{e_{j+1} - e_j}{h}\right)^2\right]^{1/2} \\
&= |\delta_h e_h|_2 \\
&= h^{-1/2}|\delta_h e_h|_h \\
&\leq h^{-1/2} c\|u\|^*_{C^4(\overline{\Omega})} h^2 \\
&= c\|u\|^*_{C^4(\overline{\Omega})} h^{3/2}.
\end{aligned}$$

We have shown that

$$|e_h|^*_{H^1} \leq c\|u\|^*_{C^4(\overline{\Omega})} h^{3/2}. \tag{2.25}$$

Using (2.16) we can see that,

$$|e_h|_{H^1} \leq \sqrt{2}|e_h|^*_{H^1} \leq c\|u\|^*_{C^4(\overline{\Omega})} h^{3/2}.$$

These demonstrate that

$$|e_h|_{H^1} \leq c\|u\|^*_{C^4(\overline{\Omega})} h^{3/2}. \tag{2.26}$$

We can easily see that equations (2.25) and (2.26) prove that the scheme Δ_h is convergent of order 3/2 with respect to $|\cdot|^*_{H^1}$ and $|\cdot|_{H^1}$ respectively. □

Theorem 2.19. *The condition number of the Δ_h scheme is at most $\frac{2}{3h^2}$.*

Proof. In order to estimate the condition number of Δ_h, we need to estimate $|\Delta_h|_2$ and $|\Delta_h^{-1}|_2$. Let u_h be a vector such that $u_0 = 0 = u_N$.

It is easily seen that $\Delta_h u_h = \delta_{-h}(\delta_h u_h)$. Using inequality (2.17) twice, we get

$$|\Delta_h u_h|_2 \leq \frac{2}{h}|\delta_h u_h|_2 \leq \frac{4}{h^2}|u_h|_2.$$

Using the minimax characterization for eigenvalues we may conclude that

$$|\Delta_h|_2 \leq \frac{4}{h^2}. \tag{2.27}$$

Applying (2.3) on the summation by parts we have shown previously we get,

$$u_h^T(-\Delta_h)u_h = |\delta_h u_h|_2^2 \geq \ 6|u_h|_2^2.$$

Therefore the smallest eigenvalue of $(-\Delta_h)$ is $\lambda_{min} \geq 6$. Since $(-\Delta_h)$ is symmetric

$$|\Delta_h^{-1}|_2 \leq \frac{1}{6}.$$

Now we can estimate the condition number using the $|\cdot|_2$ norm,

$$\kappa(\Delta_h) = |\Delta_h^{-1}|_2|\Delta_h|_2 \leq \frac{1}{6} \cdot \frac{4}{h^2} = \frac{2}{3h^2}.$$

□

Since Δ_h is a symmetric positive definite tridiagonal Toeplitz matrix we can know the values of all of its eigenvalues explicitly,

$$\lambda_i = \alpha - 2\sqrt{\beta\gamma}\cos\left(\frac{a}{N}\pi\right), \quad a = 1, 2, ..., N-1.$$

α represents the value on the main diagonal, β the value on the upper diagonal and γ the value on the lower diagonal. We also know that $|\Delta_h|_2$ will be equal to its largest eigenvalue and $|\Delta_h^{-1}|_2$ will be equal to the inverse of its smallest eigenvalue. Using that we find the condition number to be

$$\kappa(\Delta_h) \approx \frac{4}{h^2\pi^2}.$$

2.2.2 Solving the 1D Poisson Equation with Neumann Boundary Conditions.

Consider the Poisson equation

$$-u'' = f \quad \text{on} \quad (0,1) \quad \text{with} \quad u'(0) = 0 = u'(1).$$

Given that f has zero average and letting u be the unique solution with zero average, we define the offset grid as

$$\widetilde{\Omega}_h = \left\{ \left(i + \frac{1}{2} \right) h, 0 \le i \le N-1 \right\}.$$

For $1 \le i \le N-2$ the discretization is simple, but at the boundaries the discretization requires points that do not exist , i.e. $u_{\frac{1}{2}-1}$ and $u_{N-\frac{1}{2}+1}$. For those we use the Neumann boundary conditions,

$$0 = u'(0) \approx \frac{u_{1/2} - u_{-1/2}}{h}$$

and so we get

$$u_{-1/2} = u_{1/2}. \tag{2.28}$$

Similarly from the second boundary condition we get

$$u_{N-1/2} = u_{N+1/2}. \tag{2.29}$$

Let us define the discretization matrix to be,

$$-\Delta_h^{(3)} = \frac{1}{h^2} \begin{bmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 1 \end{bmatrix}.$$

The scheme was taken from [7] using the same notation. Note that $-\Delta_h^{(3)}$ is singular with a one-dimensional null space given by scalar multiples of $\mathbf{1}$, where $\mathbf{1}$ represents a vector of ones.
Let $V = \{u_h \in R^{\mathbf{N}}, \quad \sum_{i=1}^{N} u_i = 0\}$ and $A_h : V \longrightarrow V$, defined by $A_h u_h = (-\Delta_h^{(3)}) u_h$ for $u_h \in V$. Then A_h is invertible on V.

Theorem 2.20. *The A_h scheme is stable and $|A_h^{-1}|_2 \leq \frac{1}{2}$.*

Proof. First we prove a summation by parts equality to be used later. Let $u_h \in V$ and $u_k = u_{i+\frac{1}{2}}$. We have,

$$\begin{aligned}
-h^2 u_h^T A_h u_h &= \sum_{i=1}^{N-2} u_k(2u_k - u_{k-1} - u_{k+1}) + u_{\frac{1}{2}}(u_{\frac{1}{2}} - u_{\frac{3}{2}}) + u_{N-\frac{1}{2}}(u_{N-\frac{1}{2}} - u_{N-\frac{3}{2}}) \\
&= \sum_{i=1}^{N-2} u_k(u_k - u_{k-1}) + \sum_{i=1}^{N-2} u_k(u_k - u_{k+1}) + u_{\frac{1}{2}}(u_{\frac{1}{2}} - u_{\frac{3}{2}}) + u_{N-\frac{1}{2}}(u_{N-\frac{1}{2}} - u_{N-\frac{3}{2}}) \\
&= \sum_{i=1}^{N-2} u_k^2 - \sum_{i=1}^{N-2} u_k u_{k-1} + \sum_{i=1}^{N-2} u_k^2 - \sum_{i=1}^{N-2} u_k u_{k+1} + u_{\frac{1}{2}}^2 - u_{\frac{1}{2}} u_{\frac{3}{2}} + u_{N-\frac{1}{2}}^2 - u_{N-\frac{1}{2}} u_{N-\frac{3}{2}} \\
&= \sum_{i=0}^{N-2} u_k^2 - \sum_{i=1}^{N-1} u_k u_{k-1} + \sum_{i=1}^{N-1} u_k^2 - \sum_{i=0}^{N-2} u_k u_{k+1} \\
&= \sum_{i=0}^{N-2} u_k^2 - \sum_{i=0}^{N-2} u_k u_{k+1} + \sum_{i=0}^{N-2} u_{k+1}^2 - \sum_{i=0}^{N-2} u_k u_{k+1} \\
&= \sum_{i=0}^{N-2} u_k^2 - 2\sum_{i=0}^{N-2} u_k u_{k+1} + \sum_{i=0}^{N-2} u_{k+1}^2 \\
&= \sum_{i=0}^{N-2} (u_{k+1} - u_k)^2 \\
&= \sum_{i=0}^{N-1} (u_{k+1} - u_k)^2.
\end{aligned}$$

From the above equality we can see that

$$|\delta_h u_h|_2^2 = u_h^T A_h u_h. \tag{2.30}$$

Using the discrete Poincaré inequality (2.4) we obtain,

$$2|u_h|_2^2 \leq u_h^T A_h u_h,$$

leading to the conclusion

$$|A_h^{-1}|_2 \leq \frac{1}{2}. \tag{2.31}$$

□

Theorem 2.21. *The A_h scheme is second order consistent.*

Proof. Let u be any smooth function with zero average and $\tilde{R}_h$ be the restriction to $\tilde{\Omega}_h$. In order to show consistency we simply need to prove that

$$|A_h\tilde{R}_h u - \tilde{R}_h\Delta u|_h \leq c\|u\|^*_{C^4(\overline{\Omega})}h^2.$$

For $1 \leq i \leq N-2$, using the previous notation where $k = i + 1/2$, $u_k = u(x_k)$ can be expanded using Taylor's theorem,

$$u_{k\pm1} = u_k \pm \frac{du_k}{dx}h + \frac{d^2u_k}{dx^2}\frac{h^2}{2} \pm \frac{d^3u_k}{dx^3}\frac{h^3}{6} + \frac{d^4u(\xi^\pm)}{dx^4}\frac{h^4}{24},$$

for some $\xi^\pm \in \Omega$. Adding the two equalities above we get

$$u_{k+1} + u_{k-1} = 2u_k + \frac{d^2u_k}{dx^2}h^2 + \frac{d^4u(\xi^\pm)}{dx^4}\frac{h^4}{12}.$$

By rearranging the terms we get

$$\frac{-2u_k + u_{k+1} + u_{k-1}}{h^2} - \frac{d^2u_k}{dx^2} = E, \quad \text{and} \quad \|E\|^*_{C^4(\overline{\Omega})} \leq c\|u\|^*_{C^4(\overline{\Omega})}h^2, \tag{2.32}$$

where c is a positive constant. Since we chose any $1 \leq i \leq N-2$ we may conclude that the inequality is true for all those indices. We now need to prove the inequality for the boundaries.

Let us start with the left side boundary

$$u_{\frac{1}{2}\pm1} = u_{1/2} \pm \frac{du_{1/2}}{dx}h + \frac{d^2u_{1/2}}{dx^2}\frac{h^2}{2} \pm \frac{d^3u_{1/2}}{dx^3}\frac{h^3}{6} + \frac{d^4u(\xi^\pm)}{dx^4}\frac{h^4}{24},$$

for some $\xi^\pm \in \Omega$. Adding the two equalities above we get

$$u_{3/2} + u_{-1/2} = 2u_{1/2} + \frac{d^2u_{1/2}}{dx^2}h^2 + \frac{d^4u(\xi^\pm)}{dx^4}\frac{h^4}{12}.$$

Using equation (2.28) and rearranging the terms we get

$$\frac{-u_{1/2} + u_{3/2}}{h^2} - \frac{d^2u_{1/2}}{dx^2} = E, \quad \text{and} \quad \|E\|^*_{C^4(\overline{\Omega})} \leq c\|u\|^*_{C^4(\overline{\Omega})}h^2, \tag{2.33}$$

where c is a positive constant.
We now check the right side boundary

$$u_{N-\frac{1}{2}\pm1} = u_{N-\frac{1}{2}} \pm \frac{du_{N-\frac{1}{2}}}{dx}h + \frac{d^2u_{N-\frac{1}{2}}}{dx^2}\frac{h^2}{2} \pm \frac{d^3u_{N-\frac{1}{2}}}{dx^3}\frac{h^3}{6} + \frac{d^4u(\xi^\pm)}{dx^4}\frac{h^4}{24}.$$

Following a similar procedure as the left side boundary and using equation (2.29) we get

$$\frac{-u_{N-\frac{1}{2}} + u_{N-\frac{3}{2}}}{h^2} - \frac{d^2u_{N-\frac{1}{2}}}{dx^2} = E, \quad \text{and} \quad \|E\|^*_{C^4(\overline{\Omega})} \leq c\|u\|^*_{C^4(\overline{\Omega})}h^2. \tag{2.34}$$

where c is a positive constant.

Adding the three inequalities (2.32), (2.33) and (2.34) together demonstrates that the scheme is consistent, i.e.

$$|A_h \tilde{R}_h u - \tilde{R}_h \Delta u|_\infty \leq c\|u\|^*_{C^4(\overline{\Omega})} h^2.$$

□

Theorem 2.22. *The A_h scheme is convergent with respect to the norms $|\cdot|_h$, $|\cdot|_{H^1}$, $|\cdot|^*_{H^1}$, $|\cdot|_\infty$.*

Proof. Let

$$-\Delta u = f, \quad A_h u_h = \tilde{R}_h f, \quad e_h = \tilde{R}_h u - u_h.$$

Since $A_h e_h = A_h \tilde{R}_h u - \tilde{R}_h \Delta u$ we can write,

$$|A_h e_h|_\infty \leq c\|u\|^*_{C^4(\overline{\Omega})} h^2. \tag{2.35}$$

From Theorem 2.21 we know the scheme is consistent with respect to $|\cdot|_\infty$. It can be easily seen that

$$|A_h e_h|_h = h^{1/2}|A_h e_h|_2 \leq h^{1/2} \cdot h^{-1/2}|A_h e_h|_\infty \leq c\|u\|^*_{C^4(\overline{\Omega})} h^2,$$

demonstrating that the scheme is consistent with respect to $|\cdot|_h$ as well, i.e.

$$|A_h e_h|_h \leq c\|u\|^*_{C^4(\overline{\Omega})} h^2. \tag{2.36}$$

Now we can demonstrate that the scheme is convergent.
For that we need to show that $|e_h|_h \leq c\|u\|^*_{C^4(\overline{\Omega})} h^2$; we have

$$\begin{aligned}
|e_h|_h &= |(A_h)^{-1} A_h e_h|_h \\
&\leq |A_h^{-1}|_h |A_h e_h|_h \\
&= |A_h^{-1}|_2 |A_h e_h|_h \\
&\leq \frac{1}{2} c\|u\|^*_{C^4(\overline{\Omega})} h^2 \\
&\leq c\|u\|^*_{C^4(\overline{\Omega})} h^2.
\end{aligned}$$

Using (2.31) and (2.36) above we have proven that,

$$|e_h|_h \leq c\|u\|^*_{C^4(\overline{\Omega})} h^2. \tag{2.37}$$

Demonstrating that the scheme A_h is convergent of second order with respect to $|\cdot|_h$.

We will demonstrate that $\Delta_h^{(3)} u_h$ is equivalent to $\delta_{-h}(\delta_h u_h)$ and we will use the matrix representation to show that. Let us first find $\delta_h v_h$ on the interior points,

$$\frac{1}{h}[u_{3/2} - u_{1/2}, ..., u_{N-1/2} - u_{N-3/2}].$$

Applying the Neumann boundary condition we get,

$$\delta_h u_h = \frac{1}{h}[0, u_{3/2} - u_{1/2}, ..., u_{N-1/2} - u_{N-3/2}, 0].$$

Notice that after applying the boundary condition, $\delta_h u_h$ has $(N+1)$ components. If we wish to write δ_h in a matrix representation we get an $(N+1) \times N$ matrix:

$$\delta_h = \frac{1}{h}\begin{bmatrix} 0 & 0 & & \\ -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \\ & & 0 & 0 \end{bmatrix}.$$

We apply the normal backwards discrete derivative as we defined it before, i.e.

$$\begin{aligned} \delta_{-h}(\delta_h u_h) = {} & \frac{1}{h^2}[(u_{3/2} - u_{1/2}) - 0, (u_{5/2} - u_{3/2}) - (u_{3/2} - u_{1/2}), ... \\ & ..., (u_{N-1/2} - u_{N-3/2}) - (u_{N-3/2} - u_{N-5/2}), 0 - (u_{N-1/2} - u_{N-3/2})] \\ = {} & \frac{1}{h^2}[-u_{1/2} + u_{3/2}, -2u_{3/2} + u_{5/2} + u_{1/2}, ..., -2u_{N-3/2} + u_{N-1/2} + u_{N-5/2}, u_{N-3/2} - u_{N-1/2}]. \end{aligned}$$

We can easily see that $\delta_{-h}(\delta_h v_h)$ has (N) components Putting δ_{-h} in matrix form we can easily see that it is required to be an $(N) \times (N+1)$ matrix,

$$\delta_{-h} = \frac{1}{h}\begin{bmatrix} 0 & 1 & & & & \\ & -1 & 1 & & & \\ & & \ddots & \ddots & & \\ & & & -1 & 1 & \\ & & & & -1 & 0 \end{bmatrix}.$$

Notice that if we multiply matrix δ_{-h} by δ_h we get a $(N \times N)$ matrix which is exactly $\Delta_h^{(3)}$,

$$-\delta_{-h}(\delta_h) = -\frac{1}{h^2}\begin{bmatrix} 0 & 1 & & & & \\ & -1 & 1 & & & \\ & & \ddots & \ddots & & \\ & & & -1 & 1 & \\ & & & & -1 & 0 \end{bmatrix}\begin{bmatrix} 0 & 0 & & \\ -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \\ & & 0 & 0 \end{bmatrix} = \frac{1}{h^2}\begin{bmatrix} 1 & -1 & & & \\ -1 & 2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 1 \end{bmatrix} = -\Delta_h^{(3)}.$$

This gives us,

$$\delta_{-h}(\delta_h) = \Delta_h^{(3)}. \tag{2.38}$$

We wish to prove now that the scheme is also convergent with respect to the discrete H^1 norm and the equivalent H^1 norm from equation (2.16). We have

$$
\begin{aligned}
(|e_h|^*_{H^1})^2 &= \frac{e^2_{1/2} + e^2_{N-1/2}}{h} + \sum_{j=0}^{N-2} \left(\frac{e_{(j+1/2)+1} - e_{j+1/2}}{h} \right)^2 \\
&\leq \frac{|e_h|^2_2}{h} + |\delta_h e_h|^2_2 \\
&\leq \frac{|e_h|^2_h}{h^2} + \frac{1}{6} |\Delta^{(3)}_h e_h|^2_2 \\
&\leq \frac{|A^{-1}_h|^2_h |\Delta^{(3)}_h e_h|^2_h}{h^2} + \frac{1}{6} |\Delta^{(3)}_h e_h|^2_2 \\
&= \frac{|A^{-1}_h|^2_2 |\Delta^{(3)}_h e_h|^2_h}{h^2} + \frac{1}{6h} |\Delta^{(3)}_h e_h|^2_h \\
&\leq \frac{|\Delta^{(3)}_h e_h|^2_h}{2h^2} + \frac{1}{6h} |\Delta^{(3)}_h e_h|^2_h \\
&= \frac{2}{3h^2} |\Delta^{(3)}_h e_h|^2_h.
\end{aligned}
$$

In order to get the above inequality we have used inequalities (2.3), (2.38) and (2.31). Taking the square root of both sides and using the fact the scheme is consistent we can show that,

$$|e_h|^*_{H^1} \leq c\|u\|^*_{C^4(\overline{\Omega})} h^{\cdot} \tag{2.39}$$

Using inequality (2.16) we can easily see that,

$$|e_h|_{H^1} \leq 2|e_h|^*_{H^1} \leq c\|u\|^*_{C^4(\overline{\Omega})} h^{\cdot}$$

We have proven that,

$$|e_h|_{H^1} \leq c\|u\|_{C^4(\overline{\Omega})} h. \tag{2.40}$$

Using inequality (2.8) we can see that

$$|e_h|_\infty \leq \sqrt{\frac{2h}{\sqrt{5}-1}} |e_h|_{H^1} \leq c\|u\|_{C^4(\overline{\Omega})} h^{3/2}.$$

We can get the inequality,

$$|e_h|_\infty \leq c\|u\|_{C^4(\overline{\Omega})} h^{3/2}. \tag{2.41}$$

Equations (2.39), (2.40) and (2.41) are proofs that the scheme $\Delta^{(3)}_h$ is convergent of first order with respect to $|\cdot|^*_{H^1}$ and $|\cdot|_{H^1}$ respectively and order 3/2 convergent with respect to $|\cdot|_\infty$. $\square$

Theorem 2.23. *The condition number of the A_h scheme is at most $\frac{2}{h^2}$.*

Proof. In order to find the condition number using the $|\cdot|_2$ norm we need to estimate $|A_h|_2$. If we apply inequality (2.17) to (2.30) we get,

$$u_h^T A_h u_h = |\delta_h u_h|_2^2 \leq \frac{4}{h^2}|u_h|_2^2, \quad u_h \in V.$$

From that we may conclude that

$$|A_h|_2 \leq \frac{4}{h^2}.$$

Now we can bound the condition number,

$$\kappa(A_h) = |A_h^{-1}|_2|A_h|_2 \leq \frac{1}{2} \cdot \frac{4}{h^2} = \frac{2}{h^2}.$$

Another method of finding a bound on $|A_h|_2$ is by eigenvalue perturbation.

Eigenvalue Perturbations. Let $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_n$ be the eigenvalues of a Hermitian $A \in \mathbb{R}^{n\times n}$, and suppose that A is perturbed by a Hermitian matrix E with eigenvalues $\epsilon_1 \geq \epsilon_2 \geq ... \geq \epsilon_n$ to produce $B = A + E$ which is also Hermitian. Let β_i be the eigenvalues of B, If $\beta_1 \geq \beta_2 \geq ... \geq \beta_n$ then,

$$\max_i \beta_i \leq \max_i \lambda_i + \epsilon_1.$$

A proof of the above theorem is available on page 551 in [3]. Let E_h be a $(N+1)\times(N+1)$ matrix defined as

$$E_h = \frac{1}{h^2}\begin{bmatrix} -1 & 0 & & & \\ 0 & 0 & 0 & & \\ & \ddots & \ddots & \ddots & \\ & & 0 & 0 & 0 \\ & & & 0 & -1 \end{bmatrix}.$$

We can easily see that $\Delta_h^{(3)} = \frac{(h')^2}{h^2}\Delta_{h'} + E_h$, where $\Delta_{h'}$ is the scheme defined in 2.20 with grid size of (N+2) points. If we denote β_i, λ_i and ϵ_i denote the eigenvalues of $\Delta_h^{(3)}$, $\Delta_{h'}$ and E_h respectively. From (2.27) we can find that,

$$\max_i \lambda_i \leq \frac{4}{(h')^2}.$$

It is easily seen that $\epsilon_1 = 0$. From that we can find,

$$\max_i \beta_i \leq \frac{(h')^2}{h^2}\frac{4}{(h')^2} + 0 = \frac{4}{h^2}.$$

As we stated before, $|\Delta_h^{(3)}|_2 = \max_i \beta_i$.

Now we can estimate the condition number,

$$\kappa(A_h) = |A_h^{-1}|_2|A_h|_2 \leq \frac{1}{2} \cdot \frac{4}{h^2} = \frac{2}{h^2}.$$

$\square$

Chapter 3

Monotone matrices

In this chapter I will introduce a family of matrices called monotone matrices, some equivalent conditions for monotonicity will be introduced in this chapter as well. There are more equivalent conditions for monotonicity that are not introduced but can be found in [1]. In chapter four and five we will use the fact (after we prove it of course) that the numerical schemes are monotone to prove stability.

Definition 3.1. *A vector v is said to be $v > 0$ when all entries are nonnegative and at least one element is positive.*

Definition 3.2. *A vector v is said to be $v \gg 0$ when all entries are positive.*

Definition 3.3. *A vector v is said to be $v \geq 0$ when all entries are nonnegative.*

Definition 3.4. *A matrix M is said to be $M \geq 0$ when all entries are nonnegative.*

Definition 3.5. *A real square matrix M is said to be* ***monotone*** *if $Mv \geq 0$ implies $v \geq 0$.*

From the definition of a monotone matrix one can claim the following:

Claim 3.6. A monotone matrix is non-singular.

Proof. Given a monotone matrix M and vector v, assume there is $v \neq 0$ such that $Mv = 0$. Then $M(-v) = 0$ as well. Since M is monotone, if $Mv = 0$ we can conclude $v \geq 0$. Taking the equation $M(-v) = 0$ we can conclude $(-v) \geq 0$.
The only possible vector obeying both inequalities is $v = 0$, a contradiction to the assumption. We have proved that the only solution of the equation $Mv = 0$ is the trivial solution $v = 0$, and from that we can conclude that the matrix M is non-singular. □

Claim 3.7. Suppose that M is non-singular and real. Then M is monotone iff $(M^{-1})_{ij} \geq 0\ \forall i, j$.

Proof. ($\Leftarrow$) Assume that $(M^{-1})_{ij} \geq 0$ and $Mv \geq 0$. Then
$0 \leq (M^{-1})(Mv) = v$, so $v \geq 0$; hence M is monotone.

($\Rightarrow$) Assume that M is monotone. Let e_j denote j^{th} standard unit vector.
Since M is non-singular $\exists v$ such that $Mv = e_j \geq 0$.
As M is monotone, $v \geq 0$. But v is just the j^{th} column of M^{-1}.
It now folllows that $M^{-1} \geq 0$. □

Claim 3.8. A matrix M is monotone if and only if there exists matrix R such that $B = M + R$ is monotone, $B^{-1}R \geq 0$ and $\rho(B^{-1}R) < 1$, where ρ denotes the spectral radius.

Proof. ($\Rightarrow$) Assume that M is monotone. Choose $R = \mathbf{0}$, the null matrix. We get,
$B = M + \mathbf{0} = M$.
Clearly,

$$B^{-1} = M^{-1} \geq 0, \quad B^{-1}R = \mathbf{0} \geq \mathbf{0}, \quad \rho(B^{-1}R) = 0 < 1.$$

($\Leftarrow$) Suppose R to be a matrix such that $B = M + R$ is monotone, $B^{-1}R \geq 0$ and $\rho(B^{-1}R) < 1$. We can see that,

$$M = B - R = B(I - B^{-1}R).$$

We now state two well known results.

Let A be an $N \times N$ matrix such that $\rho(A) < 1$. Then,

$$\lim_{j \to \infty} A^j = 0, \tag{3.1}$$

and,

$$(I - A)^{-1} = \sum_{j=0}^{\infty} A^j. \tag{3.2}$$

We are given $\rho(B^{-1}R) < 1$, using (3.2) we get,

$$(I - B^{-1}R)^{-1} = \sum_{j=0}^{\infty} (B^{-1}R)^j.$$

From (3.1) we know this sum will converge, we may conclude

$$M^{-1} = (I - B^{-1}R)^{-1} B^{-1}.$$

and since we are given that $B^{-1}R \geq 0$ we conclude,

$$(I - B^{-1}R)^{-1} = \sum_{j=0}^{\infty} (B^{-1}R)^j \geq 0$$

Since we know B is monotone, $B^{-1} \geq 0$. Thus,

$$M^{-1} = (I - B^{-1}R)^{-1} B^{-1} \geq 0.$$

Using Claim 3.7 we conclude that M is monotone.

$\square$

Claim 3.9. A matrix M is monotone if and only if there exists a monotone matrix $B \geq M$ and a vector $x \gg 0$ such that $Mx \gg 0$.

Proof. ($\Leftarrow$) We will use Claim 3.8 to show the above claim is true. Let us define matrix $R = B - M$, since $B \geq M$ we may conclude $R \geq 0$. Since B is monotone we know $B^{-1} \geq 0$, using that we can find $B^{-1}R \geq 0$. We already have two of the conditions for monotonicity from claim (3.8), all we have left to show is that $\rho(B^{-1}R) < 1$. Let us define matrix $C = B^{-1}R = B^{-1}(B-M) = I - B^{-1}M \geq 0$.

We prove now that,

$$\rho(C) \leq \max_i \frac{(Cx)_i}{x_i}.$$

Let,

$$\tilde{C} = \begin{bmatrix} \frac{1}{x_1} & & \\ & \ddots & \\ & & \frac{1}{x_n} \end{bmatrix} C \begin{bmatrix} x_1 & & \\ & \ddots & \\ & & x_n \end{bmatrix} \geq 0.$$

The spectrum of $\tilde{C}$ is the same as matrix C, $\Lambda(\tilde{C}) = \Lambda(C)$.

Notice,

$$\tilde{c}_{ij} = \frac{1}{x_i} c_{ij} x_j.$$

By the Gershgorin circle theorem,

$$\Lambda(\tilde{C}) \subset \bigcup_i B_{r_i}(\tilde{c}_{ii}), \quad r_i = \sum_{i \neq j} \tilde{c}_{ij},$$

where $B_r(x)$ is the ball of radius r with centre x. Using the above,

$$\rho(\tilde{C}) \leq \max_i \tilde{c}_{ii} + r_i = \max_i \frac{c_{ii} x_i}{x_i} + \sum_{i \neq j} \frac{c_{ij} x_j}{x_i} = \max_i \frac{(Cx)_i}{x_i}.$$

Remembering $\Lambda(\tilde{C}) = \Lambda(C)$,

$$\rho(C) \leq \max_i \frac{(Cx)_i}{x_i}.$$

We now prove $\rho(C) < 1$,

$$\rho(C) \leq \max_i \frac{(Cx)_i}{x_i} = \max_i \frac{[(I - B^{-1}M)x]_i}{x_i} = 1 - \min_i \frac{(B^{-1}Mx)_i}{x_i} < 1.$$

We have a monotone matrix B such that, $B^{-1}R \geq 0$ and $\rho(B^{-1}R) < 1$. Using (3.8) we conclude M is a monotone matrix.

($\Rightarrow$) Given a monotone matrix M, let us define matrix $B = M$, clearly B is a monotone matrix. We define a vector $x = M^{-1}\mathbf{1}$, and we can easily see $x \gg 0$ since by Claim 3.7 $M^{-1} \geq 0$ and no entire row of M^{-1} can have 0 otherwise M^{-1} is not invertible. We calculate $Mx = M(M^{-1}\mathbf{1}) = \mathbf{1} \gg 0$. Given a monotone matrix M we have found a monotone matrix $B \geq M$ and a vector $x \gg 0$ such that $Mx \gg 0$.

$\square$

Claim 3.10. Suppose that for the real square matrix M, $M_{ii} > 0$ and $M_{ij} \leq 0$ for every $i \neq j$. If M is irreducible and diagonally dominant where at least in one row it is strictly diagonally dominant, then M is monotone.

Proof. The matrix M can be written as

$$M = sI - B,$$

where $B_{i,j} \geq 0$ and $s > \rho(B)$.
Such an s exists since the matrix M is irreducible and strictly diagonally dominant in one of the rows. We can write

$$M = sI - B = s\left(I - \frac{B}{s}\right)$$

then

$$M^{-1} = \frac{1}{s}\left(I - \frac{B}{s}\right)^{-1}.$$

Let us define $A = \frac{B}{s}$, so that $\rho(A) < 1$. Now the equality becomes

$$M^{-1} = \frac{1}{s}(I - A)^{-1}$$

Applying (3.2) we get,

$$M^{-1} = \frac{1}{s}(I - A)^{-1} = \frac{1}{s}\sum_{j=0}^{\infty} A^j = \frac{1}{s}\sum_{j=0}^{\infty}\left(\frac{B}{s}\right)^j$$

From (3.1) we know this sum will converge and since $B_{i,j} \geq 0$ and $s > 0$ we can conclude that

$$M^{-1} = \frac{1}{s}\sum_{j=0}^{\infty}\left(\frac{B}{s}\right)^j \geq 0.$$

Hence M is a monotone matrix by Claim 3.7. □

Claim 3.11. A matrix M is monotone if and only if there exist monotone matrices B_1 and B_2 such that
$B_1 \leq M \leq B_2$.

Proof. ($\Rightarrow$) If M is monotone, let $B_1 = M$ and $B_2 = M$. It can be easily shown B_1 and B_2 are monotone since $B_1^{-1} = B_2^{-1} = M^{-1} \geq 0$.
Given a monotone matrix M we found two monotone matrices B_1 and B_2 such that $B_1 \leq M \leq B_2$.

($\Leftarrow$) If $B_1 = M$, then we conclude immediately M is monotone.
If $B_{1_{i,j}} \leq M_{i,j}$, let v be a vector defined as $v = B_1^{-1}\mathbf{1} \gg 0$ where $\mathbf{1}$ represents a vector where all entries are 1. Therefore $B_1 v = \mathbf{1}$. We find, $0 < 1 = B_1 v \leq Mv$.

We have found a vector $v \gg 0$ such that $Mv \gg 0$ and a monotone matrix $B_2 \geq M$.
Using Claim 3.9 we conclude M is monotone. □

Claim 3.12. Suppose that L_h is a monotone matrix so that whenever $v_h \geq 0$ and $|v_h|_\infty = 1$ there is some positive constant α such that $L_h v_h \geq \alpha \mathbf{1}$ for all $h \in (0,1)$. Then $|L_h^{-1}|_\infty \leq \alpha^{-1}$ for all $h \in (0,1)$.

Proof. Since the matrix L_h is monotone, we know that $L_h^{-1} \geq 0$. Let $v_h \geq 0$ so that $|v_h|_\infty = 1$. From $L_h v_h \geq \alpha \mathbf{1}$ and $L_h^{-1} \geq 0$ we can get,

$$v_h \geq \alpha L_h^{-1} \mathbf{1}.$$

Taking the infinity norm we get the inequality,

$$|v_h|_\infty \geq |\alpha L_h^{-1} \mathbf{1}|_\infty.$$

Using $|v_h|_\infty = 1$ we get,

$$1 \geq \alpha |L_h^{-1} \mathbf{1}|_\infty. \tag{3.3}$$

Using $L_h^{-1} \geq 0$ and the definition of infinity norm of a matrix we get the equality,

$$|L_h^{-1} \mathbf{1}|_\infty = |L_h^{-1}|_\infty. \tag{3.4}$$

Applying (3.4) on (3.3) gives us,

$$1 \geq \alpha |L_h^{-1}|_\infty.$$

Dividing both sides by α gives us the desired result,

$$|L_h^{-1}|_\infty \leq \alpha^{-1}.$$

□

Claim 3.13. Consider the PDE $Lv = f$ with v vanishing on the boundary and finite difference scheme $L_h v_h = f_h \coloneqq R_h f$. Assume L_h satisfies the hypotheses of claim 3.12.
Let $E(v) \coloneqq R_h L v - L_h R_h u$ be the consistency error for any smooth function v vanishing on the boundary. We claim that $|v_h - R_h v|_\infty \leq |E(v)|_\infty \alpha^{-1}$.

Proof. We have

$$\begin{aligned}
|v_h - R_h v| &= |L_h^{-1} L_h v_h - R_h v| = |L_h^{-1} R_h L v - R_h v| = |L_h^{-1} R_h L v - L_h^{-1} L_h R_h v| \\
&= |L_h^{-1}(R_h L v - L_h R_h v)| \leq |L_h^{-1}|_\infty |R_h L v - L_h R_h v|_\infty = |L_h^{-1}|_\infty |E(v)|_\infty \\
&\leq |E(v)|_\infty \alpha^{-1}.
\end{aligned}$$

□

Claim 3.14. The matrix $(-\Delta_h)$ defined in (2.20) is a monotone matrix.

Proof. We first prove that $(-\Delta_h)$ is irreducible. Let us define a matrix A defined as,

$$A_{-\Delta_h} = \begin{cases} 0 & (-\Delta_h)_{ij} = 0 \\ 1 & (-\Delta_h)_{ij} \neq 0 \end{cases}.$$

We can see matrix $A_{-\Delta_h}$ will be,

$$A_{-\Delta_h} = \begin{bmatrix} 1 & 1 & & & \\ 1 & 1 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & 1 & 1 \\ & & & 1 & 1 \end{bmatrix}.$$

We can treat matrix $A_{-\Delta_h}$ as the adjacency matrix of a directed graph, which will be the associated graph of the matrix $(-\Delta_h)$.
We wish to prove now that $(-\Delta_h)$ is irreducible. It is well known that if the directed graph of a matrix is strongly connected then it is irreducible.
If we set all N vertices in a line we can easily see that they all are connected to their nearest neighbours. Clearly, one can move from from any vertex to any other vertex on the line.

Thus we may conclude that the matrix $(-\Delta_h)$ is irreducible.

We can easily see that $(-\Delta_h)$ is a diagonally dominant matrix, strictly diagonally dominant in the first and last row, and $(-\Delta_h)_{ii} > 0$ and $(-\Delta_h)_{ij} \leq 0$ for every $i \neq j$.

Using Claim 3.10 we can conclude that $(-\Delta_h)$ is a monotone matrix. $\square$

Theorem 3.15. *The scheme $-\Delta_h$ is stable and $|-\Delta_h|_\infty \leq \frac{1}{8}$.*

Proof. We wish to show that,

$$\frac{|-\Delta_h^{-1} f_h|_\infty}{|f_h|_\infty} = \frac{|u_h|_\infty}{|f_h|_\infty} \leq \frac{1}{8}.$$

We now define the vector $w_h = [w_1, w_2, ..., w_{n-1}]^T$,
where,

$$w_i = \frac{(ih)(1-ih)}{2},$$

and the function,

$$w = \frac{x(1-x)}{2} \quad 0 \leq x \leq 1.$$

We consider w as a function and w_h the discretization of it, defined as $w_h = R_h w$. We now show that $-\Delta_h w_h = \mathbf{1}$. By simple calculation we can see that $-\Delta w = \mathbf{1}$ and $\mathbf{1} - \Delta_h w_h = R_h \Delta w - \Delta_h R_h w = 0$, since we have shown the consistency error involves fourth derivatives of w, which vanish. This proves the claim $-\Delta_h w_h = \mathbf{1}$.

It is easily seen that $|w_h|_\infty \leq \frac{1}{8}$. Now,

$$-\Delta_h(|f_h|_\infty w_h + u_h) = |f|_\infty \mathbf{1} + f_h \geq 0.$$

Since $-\Delta_h$ is monotone,

$$|f_h|_\infty w_h + u_h \geq 0.$$

From here we have,

$$-u_h \leq |f_h|_\infty w_h \leq |w_h|_\infty |f_h|_\infty \mathbf{1} \leq \frac{1}{8}|f_h|_\infty \mathbf{1}. \tag{3.5}$$

Similarly,

$$-\Delta_h(|f_h|_\infty w_h - u_h) = |f|_\infty \mathbf{1} - f_h \geq 0.$$

Since S_h is monotone,

$$|f_h|_\infty w_h - u_h \geq 0.$$

From here we find that,

$$u_h \leq |f_h|_\infty w_h \leq |w_h|_\infty |f_h|_\infty \mathbf{1} \leq \frac{1}{8}|f_h|_\infty \mathbf{1}. \tag{3.6}$$

Combining (3.5) and (3.6) we get

$$|u_h|_\infty \leq \frac{1}{8}|f_h|_\infty.$$

We conclude

$$|-\Delta_h^{-1}|_\infty \leq \frac{1}{8}. \tag{3.7}$$

Using the monotonicity property of $-\Delta_h$ we have proved the scheme to be stable. $\square$

Chapter 4

Fourth Order Numerical Schemes for the 2D Poisson Equation

In this chapter we will define two fourth order numerical schemes for the 2D Poisson equation, both of this numerical schemes can be found in [7] and previous work on the first scheme can be found in [6]. As now we work with the 2D case we will use the notation of block matrices. These schemes are well known, although the fact that they are monotone was not used to demonstrate stability previously. Throughout this chapter the term c represents a positive constant whose value might change in different places.

4.1 The $-\hat{\Delta}_h$ scheme

Unlike the two previous schemes, $-\hat{\Delta}_h$ uses nine points to estimate the Laplacian. This creates a problem for points adjacent to the boundary, which can be easily solved by using the second order scheme $-\Delta_h$ for those points.
We denote the interior of a unit square 2D discretization to be,

$$\Omega_h = \left\{ \Big(ih, jh\Big), \quad 1 \leq i, j \leq N-1 \right\},$$

and we denote the boundary points as,

$$\partial\Omega_h = \left\{ (0, ih), (1, ih), (ih, 0)(ih, 1), \quad 1 \leq i \leq N-1 \right\}.$$

We define $\overline{\Omega}_h = \Omega_h \cup \partial\Omega$, and notice $\overline{\Omega}_h$ does not include the corners.

Given the boundary condition that $u_h = 0$ on $\partial\Omega_h$ we define the matrix,

$$-\hat{\Delta}_h = \frac{1}{12h^2}\begin{bmatrix} M^* & -12I & & & & & \\ M_1 & M_2 & M_1 & I^* & & & \\ I^* & M_1 & M_2 & M_1 & I^* & & \\ & \ddots & \ddots & \ddots & \ddots & \ddots & \\ & & I^* & M_1 & M_2 & M_1 & I^* \\ & & & I^* & M_1 & M_2 & M_1 \\ & & & & & -12I & M^* \end{bmatrix}.$$

Here we have,

$$M^* = \begin{bmatrix} 48 & -12 & & & \\ -12 & 48 & -12 & & \\ & \ddots & \ddots & \ddots & \\ & & -12 & 48 & -12 \\ & & & -12 & 48 \end{bmatrix}, \quad M_1 = \begin{bmatrix} -12 & & & & \\ & -16 & & & \\ & & \ddots & & \\ & & & -16 & \\ & & & & -12 \end{bmatrix}.$$

$$M_2 = \begin{bmatrix} 48 & -12 & & & & & \\ -16 & 60 & -16 & 1 & & & \\ 1 & -16 & 60 & -16 & 1 & & \\ & \ddots & \ddots & \ddots & \ddots & \ddots & \\ & & 1 & -16 & 60 & -16 & 1 \\ & & & 1 & -16 & 60 & -16 \\ & & & & & -12 & 48 \end{bmatrix} \quad \text{and} \quad I^* = \begin{bmatrix} 0 & & & & \\ & 1 & & & \\ & & \ddots & & \\ & & & 1 & \\ & & & & 0 \end{bmatrix}.$$

The matrix I represents the identity matrix. All block matrices M^*, M_1, M_2, I and I^* are $(n-1) \times (n-1)$.

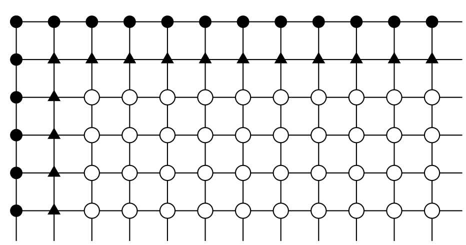
Let us define Ω_h^0 as the set of points $P \in \Omega_h$ that has all nearest neighbours of P in Ω_h, leaving points in $\Omega_h \backslash \Omega_h^0$ be those with at least one neighbour in $\partial\Omega$. For point $P \in \Omega_h \backslash \Omega_h^0$ points we use the 2D version of the scheme $-\Delta_h$ introduced in chapter one. We can split how $-\hat{\Delta}_h$ acts at a point P into two cases

$$(-\hat{\Delta}_h)(P) = \begin{cases} (-\Delta_h)(P) & \text{if } P \in \Omega_h \backslash \Omega_h^0 \\ (-\Delta_h^{(4)})(P) & \text{if } P \in \Omega_h^0 \end{cases}$$

Where $-\Delta_h$ and $-\Delta_h^{(4)}$ have the molecules

$$-\Delta_h^{(4)} = \frac{1}{12h^2} \begin{bmatrix} & & 1 & & \\ & & -16 & & \\ 1 & -16 & 60 & -16 & 1 \\ & & -16 & & \\ & & 1 & & \end{bmatrix}, \quad -\Delta_h = \frac{1}{h^2} \begin{bmatrix} & -1 & \\ -1 & 4 & -1 \\ & -1 & \end{bmatrix}.$$

Below is a figure demonstrating how the grid is separated into three layers. First layer is the boundary $\partial\Omega$ denoted by filled circles where we are given that the solution $u = 0$, the second layer is $\Omega_h \backslash \Omega_h^0$ denoted by triangles adjacent to $\partial\Omega$ where we use the second order scheme $-\Delta_h$, last is the third layer Ω_h^0 denoted by empty circles where we apply the fourth order scheme $(-\Delta_h^{(4)})$.



●$-\partial\Omega$,
▲$-\Omega_h \backslash \Omega_h^0$,
○$-\Omega_h^0$.

Theorem 4.1. *The scheme $-\hat{\Delta}_h$ is fourth order consistent in the interior.*

Proof. One can see that for points $P \in \Omega_h \backslash \Omega_h^0$ the scheme $(-\hat{\Delta}_h)$ reduces to the scheme $(-\Delta_h)$, thus we get second order consistency for those points.

$$|\hat{\Delta}_h(R_h u) - R_h \Delta u|_{\infty,\Omega_h \backslash \Omega_h^0} \leq c\|u\|^*_{C^4(\overline{\Omega})} h^2. \tag{4.1}$$

Where c is an independent constant. We now prove that for points $P \in \Omega_h^0$ the scheme is fourth order consistent

Let $u_{i,j} = u(x_{i,j})$ such that $x_{i,j} \in \Omega_h^o$. Expanding using Taylor's Theorem we get,

$$u_{i\pm1,j} = u_{i,j} \pm \frac{\partial u_{i,j}}{\partial x} h + \frac{\partial^2 u_{i,j}}{\partial x^2} \frac{h^2}{2} \pm \frac{\partial^3 u_{i,j}}{\partial x^3} \frac{h^3}{6} + \frac{\partial^4 u_{i,j}}{\partial x^4} \frac{h^4}{24} \pm \frac{\partial^5 u_{i,j}}{\partial x^5} \frac{h^5}{120} + \frac{\partial^6 u(\xi^\pm)}{\partial x^6} \frac{h^6}{720},$$

for some $\xi^\pm \in \Omega_h$. After simplification we get,

$$-2u_{i,j} + u_{i+1,j} + u_{i-1,j} - \frac{\partial^4 u_{i,j}}{\partial x^4} \frac{h^4}{12} - \frac{\partial^2 u_{i,j}}{\partial x^2} h^2 = E_1 h^2 \quad \text{where} \quad |E_1\|^*_{C^6(\overline{\Omega})} \leq c\|u\|^*_{C^6(\overline{\Omega})} h^4. \tag{4.2}$$

Expanding using Taylor's Theorem again jumping 2 steps we get,

$$u_{i\pm2,j} = u_{i,j} \pm \frac{\partial u_{i,j}}{\partial x}(2h) + \frac{\partial^2 u_{i,j}}{\partial x^2} \frac{(2h)^2}{2} \pm \frac{\partial^3 u_{i,j}}{\partial x^3} \frac{(2h)^3}{6} + \frac{\partial^4 u_{i,j}}{\partial x^4} \frac{(2h)^4}{24} \pm \frac{\partial^5 u_{i,j}}{\partial x^5} \frac{(2h)^5}{120} + \frac{\partial^6 u(\xi^\pm)}{\partial x^6} \frac{(2h)^6}{720},$$

for some $\xi^\pm \in \Omega_h$. After simplification we get,

$$\frac{\partial^4 u_{i,j}}{\partial x^4} h^4 = \frac{3}{4}\left[-2u_{i,j} + u_{i+2,j} + u_{i-2,j} - \frac{\partial^2 u_{i,j}}{\partial x^2} 4h^2 - E_2 h^2 \right] \quad \text{where} \quad |E_2\|^*_{C^6(\overline{\Omega})} \leq c\|u\|^*_{C^6(\overline{\Omega})} h^4. \tag{4.3}$$

Plugging (4.3) into (4.2) and simplifying we get,

$$\frac{-30u_{i,j} + 16u_{i+1,j} + 16u_{i-1,j} - u_{i-2,j} - u_{i+2,j}}{12h^2} - \frac{\partial^2 u_{i,j}}{\partial x^2} = E_3 \quad \text{where} \quad \|E_3\|^*_{C^6(\overline{\Omega})} \leq c\|u\|^*_{C^6(\overline{\Omega})} h^4.$$

Adding a similar equation for the second y derivative we get the desired result,

$$|\hat{\Delta}_h R_h u - R_h \Delta u|_{\infty,\Omega_h^0} \leq c\|u\|^*_{C^6(\overline{\Omega})} h^4. \tag{4.4}$$

□

Theorem 4.2. *The matrix $-\hat{\Delta}_h$ is a monotone matrix.*

Proof. In the following proof we will use Claim 3.9. We can easily construct the matrix B by multiplying two monotone matrices. Let

$$B_1 = \begin{bmatrix} 8I & & & & \\ -I^* & B_1^* & -I^* & & \\ & \ddots & \ddots & \ddots & \\ & & -I^* & B_1^* & -I^* \\ & & & & 8I \end{bmatrix}, \quad B_2 = \begin{bmatrix} B_2^* & -I & & & \\ -I & B_2^* & -I & & \\ & \ddots & \ddots & \ddots & \\ & & -I & B_2^* & -I \\ & & & -I & B_2^* \end{bmatrix},$$

where,

$$B_1^* = \begin{bmatrix} 8 & & & & \\ -1 & 8 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 8 & -1 \\ & & & & 8 \end{bmatrix}, \quad B_2^* = \begin{bmatrix} 8 & -1 & & & \\ -1 & 8 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 8 & -1 \\ & & & -1 & 8 \end{bmatrix}.$$

Using the same idea as in Claim 3.14 one can see the matices B_1 and B_2 are irreducible, applying Claim 3.10 shows matrices B_1 and B_2 are monotone, meaning their inverses are non-negative. Defining $B = B_1B_2$, we get $B^{-1} = B_2^{-1}B_1^{-1} \geq 0$ from Claim 3.7. Hence B is monotone as well. By calculating the matrix B one can see $-\hat{\Delta}_h \leq \frac{1}{12h^2}B$. We have

$$B = \begin{bmatrix} 8B_2^* & -8I & & & & & \\ -I^*B_2^* - B_1^* & B_1^*B_2^* + 2I^* & -I^*B_2^* - B_1^* & I^* & & & \\ I^* & -I^*B_2^* - B_1^* & B_1^*B_2^* + 2I^* & -I^*B_2^* - B_1^* & I^* & & \\ & \ddots & \ddots & \ddots & \ddots & \ddots & \\ & & I^* & -I^*B_2^* - B_1^* & B_1^*B_2^* + 2I^* & -I^*B_2^* - B_1^* & I^* \\ & & & I^* & -I^*B_2^* - B_1^* & B_1^*B_2^* + 2I^* & -I^*B_2^* - B_1^* \\ & & & & & -8I & 8B_2^* \end{bmatrix},$$

where,

$$8B_2^* = \begin{bmatrix} 64 & -8 & & & \\ -8 & 64 & -8 & & \\ & \ddots & \ddots & \ddots & \\ & & -8 & 64 & -8 \\ & & & -8 & 64 \end{bmatrix}, \quad -I^*B_2^* - B_1^* = \begin{bmatrix} -8 & & & & \\ 2 & -16 & 2 & & \\ & \ddots & \ddots & \ddots & \\ & & 2 & -16 & 2 \\ & & & & -8 \end{bmatrix}.$$

Note that, $B_1^* B_2^* + 2I^* = \begin{bmatrix} 64 & -8 & & & & & \\ -16 & 68 & -16 & 1 & & & \\ 1 & -16 & 68 & -16 & 1 & & \\ & \ddots & \ddots & \ddots & \ddots & \ddots & \\ & & 1 & -16 & 68 & -16 & 1 \\ & & & 1 & -16 & 68 & -16 \\ & & & & & -8 & 64 \end{bmatrix}$ and $I^* = \begin{bmatrix} 0 & & & & \\ & 1 & & & \\ & & \ddots & & \\ & & & 1 & \\ & & & & 0 \end{bmatrix}$.

Matrix I represents the identity matrix. All block matrices are $(n-1) \times (n-1)$.

We now define the vector $w_h = [w_{1,1}, w_{2,1}, ..., w_{n-1,1}, w_{1,2}, w_{2,2}, ..., w_{n-1,2}, ...w_{n-1,n-1}]^T$,

$$w_{i,j} = ih(1-ih)jh(1-jh).$$

Notice that vector $w_h = R_h w$, where

$$w = x(1-x)y(1-y) \quad 0 \leq x, y \leq 1.$$

Notice w_h does follow the boundary conditions, we can easily see that $-\Delta w = 2y(1-y)+2x(1-x)$, which is strictly positive on the domain $0 < x, y < 1$. Since the consistency error for points $P \in \Omega_h^0$ involve sixth order derivative which vanish we can conclude that $(-\hat{\Delta}_h w_h)_{P \in \Omega_h^0} = -\Delta w \gg 0$. For points $P \in \Omega_h \backslash \Omega_h^0$ the consistency error involves the fourth derivatives which do not vanish. For those points we can show, by direct computation, that $(-\hat{\Delta}_h w_h)_{P \in \Omega_h \backslash \Omega_h^0} \gg 0$.

We found a monotone matrix B such that $B \geq (-\hat{\Delta}_h)$ and $w_h \gg 0$ such that $(-\hat{\Delta}_h) w_h \gg 0$, thus by Claim 3.9 we have shown that $(-\hat{\Delta}_h)$ is monotone. □

Theorem 4.3. *The scheme $(-\hat{\Delta}_h)$ is fourth order convergent.*

Proof. We wish to show that $|R_h u - u_h|_\infty \leq c\big(\|u\|^*_{C^6(\overline{\Omega})} + \|u\|^*_{C^4(\overline{\Omega})}\big)h^4$ where c is an independent constant. To prove that we will use the discrete Green's Function. Let us first define the inner product for two vectors x and y

$$< x, y >_h := \sum_{i=0}^{N} x_i y_i.$$

It is well known that Green's Functions can be used to solve partial differential equations. We now define a discrete analog. Let the values of the discrete Green's function for $(-\hat{\Delta}_h)$ be $\hat{G}_h(P,Q)$ defined as:

$$\begin{aligned} -\hat{\Delta}_h^T \hat{G}_h(P,\cdot) &= h^{-2}\delta(P,\cdot), \\ \hat{G}_h(P,Q) &= 0 \quad \text{on} \quad \partial\Omega_h. \end{aligned}$$

In the above, the point $P \in \Omega$ is fixed, and,

$$\delta(P,Q) = \begin{cases} 1 & \text{for} \quad P = Q \\ 0 & \text{for} \quad P \neq Q \end{cases}$$

It should be noted that due to the monotonicity of the scheme, the discrete Green's function is non-negative.
Let $e_h = R_h u - u_h$. Then

$$\begin{aligned} |e_h(P)| =& h^2 |e_h^T(-\hat{\Delta}_h)^T G(\cdot,P)| \\ &\leq h^2 \max_{Q\in\Omega_h^0} |(\hat{\Delta}_h e_h)_Q| \sum_{Q\in\Omega_h^0} \hat{G}_h(P,Q) \\ &+ h^2 \max_{Q\in\Omega_h\setminus\Omega_h^0} |(\hat{\Delta}_h e_h)_Q| \sum_{Q\in\Omega_h\setminus\Omega_h^0} \hat{G}_h(P,Q). \end{aligned}$$

We have shown that,

$$|e_h(P)| \leq h^2 \max_{Q\in\Omega_h^0} |(\hat{\Delta}_h e_h)_Q| \sum_{Q\in\Omega_h^0} |\hat{G}_h(P,Q)| + h^2 \max_{Q\in\Omega_h\setminus\Omega_h^0} |(\hat{\Delta}_h e_h)_Q| \sum_{Q\in\Omega_h\setminus\Omega_h^0} |\hat{G}_h(P,Q)|. \quad (4.5)$$

In [7] page 17 equation (1.20) it is shown that,

$$h^2 \sum_{Q\in\Omega_h^0} \hat{G}_h(P,Q) = < \hat{G}(P,\cdot), 1 >_h \leq \frac{1}{8}. \quad (4.6)$$

In [6] Lemma 3.3 it is shown that,

$$\sum_{Q\in\Omega_h\setminus\Omega_h^0} \hat{G}_h(P,Q) \leq 2. \quad (4.7)$$

Even though equation (4.7) was proven in [6] Lemma 3.3 for $(-\hat{\Delta}_h)$, a similar proof can be made to show it is true for $(-\hat{\Delta}_h)^T$.

Applying (4.6) and (4.7) in (4.5) we get,

$$|e_h(P)| \leq \frac{1}{8} \max_{Q\in\Omega_h^0} |(\hat{\Delta}_h e_h)_Q| + 2h^2 \max_{Q\in\Omega_h\setminus\Omega_h^0} |(\hat{\Delta}_h e_h)_Q|.$$

Applying the consistency results from (4.1) and (4.4) we get the desired result, i.e.

$$|R_h u - u_h|_\infty \leq c\Big[\|u\|^*_{C^6(\overline{\Omega})} + \|u\|^*_{C^4(\overline{\Omega})}\Big]h^4. \tag{4.8}$$

□

Theorem 4.4. *The scheme $-\hat{\Delta}_h$ is stable, with $|(-\hat{\Delta}_h)^{-1}|_\infty \leq c$.*

Proof. Theorem 4.3 demonstrated the scheme is fourth order convergent. Using a well known result that a consistent convergent scheme is stable, we get that there is a constant c independent of h such that $|(-\hat{\Delta}_h)^{-1}|_\infty \leq c$.

□

4.2 The $-\tilde{\Delta}_h$ scheme

We will use the standard 2D grid as defined in the $-\hat{\Delta}_h$ discretization.
Given the boundary condition that $u_h = 0$ on $\partial\Omega_h$ we use the scheme,

$$-\tilde{\Delta}_h = \frac{1}{6h^2}\begin{bmatrix} T & E & & & \\ E & T & E & & \\ & \ddots & \ddots & \ddots & \\ & & E & T & E \\ & & & E & T \end{bmatrix},$$

where,

$$T = \begin{bmatrix} 20 & -4 & & & \\ -4 & 20 & -4 & & \\ & \ddots & \ddots & \ddots & \\ & & -4 & 20 & -4 \\ & & & -4 & 20 \end{bmatrix} \quad \text{and} \quad E = \begin{bmatrix} -4 & -1 & & & \\ -1 & -4 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & -4 & -1 \\ & & & -1 & -4 \end{bmatrix}.$$

Both T and E are $(n-1) \times (n-1)$. The molecule of $-\tilde{\Delta}_h$ is

$$-\tilde{\Delta}_h = \frac{1}{6h^2}\begin{bmatrix} -1 & -4 & -1 \\ -4 & 20 & -4 \\ -1 & -4 & -1 \end{bmatrix}.$$

Theorem 4.5. *The matrix $-\tilde{\Delta}_h$ is a monotone matrix.*

Proof. Doing the same steps as in Claim 3.14 we can show $-\tilde{\Delta}_h$ is irreducible.

We can easily see that $(-\tilde{\Delta}_h)$ is a diagonally dominant matrix, strictly diagonally dominant in the first and last row in every block, $(-\tilde{\Delta}_h)_{ii} > 0$ and $(-\tilde{\Delta}_h)_{ij} \leq 0$ for every $i \neq j$.

Using Claim 3.10 we can conclude that $(-\tilde{\Delta}_h)$ is a monotone matrix. $\square$

Theorem 4.6. *The scheme $\tilde{\Delta}_h$ is stable and $|\tilde{\Delta}_h^{-1}|_\infty \leq 4$.*

Proof. For any matrix A of size $n \times n$ we define the quantities,

$$r_i(A) = \sum_{j=1}^{n} a_{i,j}, \quad i = 1, ..., n$$

$$r_*(A) = \min_{1 \leq i \leq n} r_i(A), \quad r^*(A) = \max_{1 \leq i \leq n} r_i(A).$$

From [8] we know that if A is a monotone matrix and D is a positive diagonal matrix , then

$$\frac{1}{r^*(AD)}|D|_\infty \leq |A^{-1}|_\infty \leq \frac{1}{r_*(AD)}|D|_\infty, \quad \text{provided that} \quad r_*(AD) > 0.$$

We define a block diagonal matrix D of the form,

$$D = \begin{bmatrix} \overline{D_1} & & & & \\ & \overline{D_2} & & & \\ & & \ddots & & \\ & & & \overline{D_{n-1}} & \\ & & & & \overline{D_n} \end{bmatrix} \quad \text{where} \quad \overline{D_k} = \sqrt{k} \cdot I.$$

and where I is the identity matrix with size $(n-1) \times (n-1)$. For proving the stability of $-\tilde{\Delta}_h$ we are only interested in the upper bound, i.e

$$|\tilde{\Delta}_h^{-1}|_\infty \leq \frac{1}{r_*(-\tilde{\Delta}_h D)}|D|_\infty.$$

Since D is diagonal, multiplying the scheme by D will simply multiply each column of $-\tilde{\Delta}_h$ by the corresponding diagonal entry of D. Due to the structure of $-\tilde{\Delta}_h$ and D we can prove all row sums are positive and decreasing. If we take the sum of any non-corner row k of $-6h^2\tilde{\Delta}_h D$ we get,

$$(20\sqrt{k} - 4\sqrt{k} - 4\sqrt{k} - 4\sqrt{k+1} - 4\sqrt{k-1} - \sqrt{k+1} - \sqrt{k+1} - \sqrt{k-1} - \sqrt{k-1}).$$

After simplifying we get

$$6(2\sqrt{k} - \sqrt{k+1} - \sqrt{k-1}).$$

To prove this is always positive and decreasing we treat it as a continuous function for $x \geq 2$,

$$f(x) = 2\sqrt{x} - \sqrt{x+1} - \sqrt{x-1}.$$

By simplifying the function we can get,

$$f(x) = \frac{2}{(\sqrt{x} + \sqrt{x-1})(\sqrt{x} + \sqrt{x+1})(\sqrt{x+1} + \sqrt{x-1})}.$$

Clearly $f(x)$ is always positive and the horizontal asymptote is $y = 0$. All we need to prove now is that it is always decreasing using the derivative.

$$f'(x) = -\frac{\sqrt{x}(\sqrt{x+1}+\sqrt{x-1})}{\sqrt{x}\sqrt{x^2-1}(\sqrt{x}+\sqrt{x-1})(\sqrt{x}+\sqrt{x+1})(\sqrt{x+1}+\sqrt{x-1})}.$$

Clearly the derivative is always negative. We have proven that for $x \geq 2$, $f(x)$ is always positive and decreasing, thus it is also true for the discrete case.

One should be able to see that the minimum row sum will not happen in the last block but rather in the block before it since in the last block we subtract fewer elements from the diagonal entry. For the same reason as above, the minimum in every block will happen in a non-corner row.

Now can we find $r_*(-\tilde{\Delta}_h D)$, which happens at any non-corner row in the $(n-1)$ block. We have

$$\begin{aligned}
r_*(\tilde{\Delta}_h D) &= \frac{1}{6h^2}(20\sqrt{n-1} - 4\sqrt{n-1} - 4\sqrt{n-1} - 4\sqrt{n} - 4\sqrt{n-2} - \sqrt{n} - \sqrt{n} - \sqrt{n-2} - \sqrt{n-2}) \\
&= \frac{1}{h^2}(2\sqrt{n-1} - \sqrt{n} - \sqrt{n-2}) \\
&= \frac{1}{h^2}\frac{2}{(\sqrt{n-1}+\sqrt{n-2})(\sqrt{n-1}+\sqrt{n})(\sqrt{n}+\sqrt{n-2})}.
\end{aligned}$$

Now we can find the upper bound. Note that

$$\begin{aligned}
|\tilde{\Delta}_h^{-1}|_\infty &\leq \frac{1}{r_*(-\tilde{\Delta}_h D)}|D|_\infty \\
&= \frac{1}{\frac{1}{h^2}\frac{2}{(\sqrt{n-1}+\sqrt{n-2})(\sqrt{n-1}+\sqrt{n})(\sqrt{n}+\sqrt{n-2})}} \cdot \sqrt{n} \\
&= h^2\frac{(\sqrt{n-1}+\sqrt{n-2})(\sqrt{n-1}+\sqrt{n})(\sqrt{n}+\sqrt{n-2})}{2}\sqrt{n} \\
&\leq h^2\frac{(\sqrt{n}+\sqrt{n})(\sqrt{n}+\sqrt{n})(\sqrt{n}+\sqrt{n})}{2}\sqrt{n} \\
&\leq h^2\frac{8n^{3/2}}{2}\sqrt{n} \\
&\leq h^2 \cdot 4n^2 \\
&= 4.
\end{aligned}$$

We have found a constant upper bound on $|\tilde{\Delta}_h^{-1}|_\infty$ proving it is stable, i.e.

$$|\tilde{\Delta}_h^{-1}|_\infty \leq 4. \tag{4.9}$$

□

Theorem 4.7. *The scheme $\tilde{\Delta}_h$ is fourth order consistent.*

Proof. We begin by following the same steps as before where c is an independent constant ,using Taylor's Theorem. Let $u_{i,j} = u(x_{i,j})$ be such that $x_{i,j} \in \Omega_h$. Expanding using Taylor's Theorem we get,

$$u_{i\pm 1,j} = u_{i,j} \pm \frac{\partial u_{i,j}}{\partial x}h + \frac{\partial^2 u_{i,j}}{\partial x^2}\frac{h^2}{2} \pm \frac{\partial^3 u_{i,j}}{\partial x^3}\frac{h^3}{6} + \frac{\partial^4 u_{i,j}}{\partial x^4}\frac{h^4}{24} \pm \frac{\partial^5 u_{i,j}}{\partial x^5}\frac{h^5}{120} + \frac{\partial^6 u(\xi^\pm)}{\partial x^6}\frac{h^6}{720},$$

for some $\xi^\pm \in \Omega_h$. After simplification we get,

$$-2u_{i,j} + u_{i-1,j} + u_{i+1,j} = \frac{\partial^2 u_{i,j}}{\partial x^2}h^2 + \frac{\partial^4 u_{i,j}}{\partial x^4}\frac{h^4}{12} + E_1 h^2 \quad \text{where} \quad \|E_1\|^*_{C^6(\overline{\Omega})} \leq c\|u\|^*_{C^6(\overline{\Omega})}h^4. \tag{4.10}$$

Doing the same for the second y derivative we get,

$$-2u_{i,j} + u_{i-1,j} + u_{i+1,j} = \frac{\partial^2 u_{i,j}}{\partial y^2}h^2 + \frac{\partial^4 u_{i,j}}{\partial y^4}\frac{h^4}{12} + E_2 h^2 \quad \text{where} \quad \|E_2\|^*_{C^6(\overline{\Omega})} \leq c\|u\|^*_{C^6(\overline{\Omega})}h^4. \tag{4.11}$$

We can see that unlike before, we use the corners for the $-\tilde{\Delta}_h$ scheme. We apply Taylor's Theorem for multivariable functions at the corners, where $k_1, k_2 = \pm 1$.

$$\begin{aligned}
u_{i+k_1,j+k_2} =& u_{i,j} + \left[\frac{\partial u_{i,j}}{\partial x}k_1 + \frac{\partial u_{i,j}}{\partial y}k_2\right]h + \left[\frac{\partial^2 u_{i,j}}{\partial x^2}k_1^2 + 2\frac{\partial u_{i,j}}{\partial x \partial y}k_1 k_2 + \frac{\partial^2 u_{i,j}}{\partial y^2}k_2^2\right]\frac{h^2}{2} + \\
& \left[\frac{\partial^3 u_{i,j}}{\partial x^3}k_1^3 + 3\frac{\partial^3 u_{i,j}}{\partial x^2 \partial y}k_1^2 k_2 + 3\frac{\partial^3 u_{i,j}}{\partial x \partial y^2}k_1 k_2^2 + \frac{\partial^3 u_{i,j}}{\partial y^3}k_2^3\right]\frac{h^3}{6} + \\
& \left[\frac{\partial^4 u_{i,j}}{\partial x^4}k_1^4 + 4\frac{\partial^4 u_{i,j}}{\partial x^3 \partial y}k_1^3 k_2 + 6\frac{\partial^4 u_{i,j}}{\partial x^2 \partial y^2}k_1^2 k_2^2 + 4\frac{\partial^4 u_{i,j}}{\partial x \partial y^3}k_1 k_2^3 + \frac{\partial^4 u_{i,j}}{\partial y^4}k_2^4\right]\frac{h^4}{24} + \\
& \left[\frac{\partial^5 u_{i,j}}{\partial x^5}k_1^5 + 5\frac{\partial^5 u_{i,j}}{\partial x^4 \partial y}k_1^4 k_2 + 10\frac{\partial^5 u_{i,j}}{\partial x^3 \partial y^2}k_1^3 k_2^2 + 10\frac{\partial^5 u_{i,j}}{\partial x^2 \partial y^3}k_1^2 k_2^3 + 5\frac{\partial^5 u_{i,j}}{\partial x \partial y^4}k_1 k_2^4 + \frac{\partial^5 u_{i,j}}{\partial y^5}k_2^5\right]\frac{h^5}{120} + \\
& \left[\frac{\partial^6 u_{i,j}}{\partial x^6}k_1^6 + 6\frac{\partial^6 u_{i,j}}{\partial x^5 \partial y}k_1^5 k_2 + 15\frac{\partial^6 u_{i,j}}{\partial x^4 \partial y^2}k_1^4 k_2^2 + 40\frac{\partial^6 u_{i,j}}{\partial x^3 \partial y^3}k_1^3 k_2^3 + 15\frac{\partial^6 u_{i,j}}{\partial x^2 \partial y^4}k_1^2 k_2^4 + \right. \\
& \left. + 6\frac{\partial^6 u_{i,j}}{\partial x \partial y^5}k_1 k_2^5 + \frac{\partial^6 u_{i,j}}{\partial y^6}k_2^6\right]\frac{h^6}{720}.
\end{aligned} \tag{4.12}$$

Combining (4.10),(4.11) and all four possibilities of (4.12) we get,

$$\tilde{\Delta}_h R_h u = R_h \Delta u + \frac{h^2}{12}R_h \Delta^2 u + \frac{h^4}{72}\left[\frac{1}{5}\left(\frac{\partial^6 u}{\partial x^6} + \frac{\partial^6 u}{\partial y^6}\right) + \frac{\partial^6 u}{\partial x^4 \partial y^2} + \frac{\partial^6 u}{\partial x^2 \partial y^4}\right], \tag{4.13}$$

where the h^4 terms are evaluated at some point in Ω_h. Thus the $\tilde{\Delta}_h$ scheme is second order consistent. To prove it is fourth order consistent we have to modify the restriction operator R_h.

Let us define the new restriction operator as a $(N-1)^2 \times (N-1)^2$ matrix,

$$\tilde{R}_h = \frac{1}{12}\begin{bmatrix} & 1 & \\ 1 & 8 & 1 \\ & 1 & \end{bmatrix} = \begin{bmatrix} & & \\ & 1 & \\ & & \end{bmatrix} + \frac{1}{12}\begin{bmatrix} & 1 & \\ 1 & -4 & 1 \\ & 1 & \end{bmatrix}.$$

For example $(\tilde{R}_h u_h)_{ij} = u_{ij} + \frac{1}{12}(u_{i+1,j} + u_{i-1,j} - 4u_{i,j} + u_{i,j+1} + u_{i,j-1})$.
Notice that, $\tilde{R}_h u = R_h u + \frac{h^2}{12}\Delta_h R_h u$, from which we may conclude that,

$$\tilde{R}_h \Delta u = R_h \Delta_h + \frac{h^2}{12} R_h \Delta^2 u + E, \quad \text{where} \quad \|E_3\|^*_{C^6(\overline{\Omega})} \le c\|u\|^*_{C^6(\overline{\Omega})} h^4. \tag{4.14}$$

Using (4.14) in (4.13) we get,

$$\tilde{\Delta}_h R_h u = \tilde{R}_h \Delta u - E_3 + \frac{h^4}{72}\left[\frac{1}{5}\left(\frac{\partial^6 u}{\partial x^6} + \frac{\partial^6 u}{\partial y^6}\right) + \frac{\partial^6 u}{\partial x^4 \partial y^2} + \frac{\partial^6 u}{\partial x^2 \partial y^4}\right].$$

Since E_3 involves sixth order deriatives, when we take the infinity norm the h^4 terms will be included in E. Thus we may conclude,

$$|\tilde{\Delta}_h R_h u - \tilde{R}_h \Delta u|_\infty \le c\|u\|^*_{C^6(\overline{\Omega})} h^4. \tag{4.15}$$

We have proven the scheme to be fourth order consistent.

□

Theorem 4.8. *The scheme $-\tilde{\Delta}_h u_h = \tilde{R}_h f$ is fourth order convergent.*

Proof. We wish to show $|R_h u - u_h|_\infty \le c\|u\|^*_{C^6(\overline{\Omega})} h^4$.

We have,

$$\begin{aligned} |R_h u - u_h|_\infty &= |(-\tilde{\Delta}_h^{-1})(-\tilde{\Delta}_h)(R_h u - u_h)|_\infty \\ &\le |-\tilde{\Delta}_h^{-1}|_\infty| - \hat{\Delta}_h(R_h u - u_h)|_\infty. \\ &\le 4| - \tilde{\Delta}_h R_h u - \tilde{R}_h f|_\infty \\ &= 4| - \tilde{\Delta}_h R_h u + \tilde{R}_h \Delta u|_\infty. \end{aligned}$$

Using (4.15), we may conclude the desired result showing the scheme is fourth order convergent.

i.e.,

$$|R_h u - u_h|_\infty \le c\|u\|^*_{C^6(\overline{\Omega})} h^4.$$

□

Chapter 5

Fourth Order Numerical Schemes for the 1D Biharmonic Equation

All the numerical schemes showed before were used to solve the Poisson equation. The numerical scheme to solve the biharmonic equation defined in this book was developed in [2], they have done an extensive analysis of the scheme in the book. We introduce the biharmonic equation defined as:

$$\Delta^2 u = f.$$

Since only the 1D case will be shown here we can rewrite the equation as,

$$\frac{d^4u}{dx^4} = f \quad \text{with boundary conditions} \quad u(0) = u(1) = u'(0) = u'(1) = 0.$$

Since $u(0) = 0 = u(1)$ we only need to find u_i for $1 \leq i \leq N-1$.

We define the grid as

$$\Omega_h = \left\{ ih, \quad 1 \leq i \leq N-1 \right\}.$$

We denote $v = u'$, the first derivative of u. This scheme requires a detailed explanation as to how it is constructed. Let $x_i = ih$ for $0 \leq i \leq N$, and u_h be a vector with entries $u_i = u(x_i)$.

We define two new finite difference schemes for approximation of the first and second derivative respectively at point u_i

$$\delta_h^* u_i = \frac{u_{i+1} - u_{i-1}}{2h}, \quad \text{and} \quad \delta_h^2 u_i = \frac{u_{i+1} - 2u_i + u_{i+1}}{h^2}.$$

Let us take a fourth order polynomial,

$$Q(x) = a_0 + a_1(x - x_i) + a_2(x - x_i)^2 + a_3(x - x_i)^3 + a_4(x - x_i)^4.$$

In order for $Q(x)$ to fit the given vector u_h and v_h we require,

$$Q(x_{i-1}) = u_{i-1}, \quad Q(x_i) = u_i, \quad Q(x_{i+1}) = u_{i+1}, \quad Q'(x_{i-1}) = v_{i-1}, \quad Q'(x_{i+1}) = v_{i+1}.$$

Under the conditions above, $Q(x)$ has a unique solution,

$$\begin{aligned}
a_0 &= u_i,\\
a_1 &= \frac{3}{2}\delta_h^* u_i - \frac{1}{4}\big[v_{i+1} + v_{i-1}\big],\\
a_2 &= \delta_h^2 u_i - \frac{1}{2}\delta_h^* v_i,\\
a_3 &= \frac{1}{h^2}\big(\delta_h^* u_i - v_i\big),\\
a_4 &= \frac{1}{2h^2}\big(\delta_h^* v_i - \delta_h^2 u_i\big).
\end{aligned}$$

We can approximate then $\frac{d^4u}{dx^4}$ by taking the fourth order derivative of $Q(x)$,

$$\left(\frac{d^4u}{dx^4}\right)_i \simeq \left(\frac{d^4Q}{dx^4}\right)_i = 24a_4 = \frac{12}{h^2}\big(\delta_h^* v_i - \delta_h^2 u_i\big) := \delta_h^{(4)} u_i. \tag{5.1}$$

We found that the approximation $\delta_h^{(4)} u_h$ depends on both u_h and v_h, and we wish to estimate using only u_h. For that we need to find the relationship between u_h and v_h. An intuitive way of doing it is setting $v_i = a_1$. This gives us,

$$v_i = \frac{3}{2}\delta_h^* u_i - \frac{1}{4}\big[v_{i+1} + v_{i-1}\big], \quad 1 \le i \le N-1$$

or

$$\frac{1}{6}v_{i-1} + \frac{2}{3}v_i + \frac{1}{6}v_{i+1} = \delta_h^* u_i, \quad 1 \le i \le N-1. \tag{5.2}$$

Combining all we have shown, we can now find a discrete solution for the biharmonic equation using the scheme,

$$\delta_h^{(4)} u_i = f(x_i), \quad 1 \le i \le N-1, \tag{5.3}$$

$$\delta_h^* u_i = \frac{1}{6}v_{i-1} + \frac{2}{3}v_i + \frac{1}{6}v_{i+1}, \quad 1 \le i \le N-1, \tag{5.4}$$

$$u_0 = u_N = v_0 = v_N = 0. \tag{5.5}$$

As we have done before, we wish to use a matrix representation for the scheme.

$$\text{The operator} \quad \delta_h^* \quad \text{is equivalent to the matrix} \quad \frac{1}{2h}K, \tag{5.6}$$

where K is defined as,

$$K = \begin{bmatrix} 0 & 1 & & & \\ -1 & 0 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 0 & 1 \\ & & & -1 & 0 \end{bmatrix}.$$

Equation (5.2) can be represented as $\frac{1}{2h}Ku_h = \frac{1}{6}Pv_h$ and after simplification we get

$$v_h = \frac{3}{h}P^{-1}Ku_h, \tag{5.7}$$

where matrix P is defined as,

$$P = \begin{bmatrix} 4 & 1 & & & \\ 1 & 4 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & 4 & 1 \\ & & & 1 & 4 \end{bmatrix}.$$

It is easy to find that,

$$|P|_\infty = 6. \tag{5.8}$$

$$\text{The matrix representation of} \quad \delta_h^2 \quad \text{is} \quad -\frac{1}{h^2}T, \tag{5.9}$$

with T defined as

$$T = \begin{bmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{bmatrix}.$$

The matrix T should be familiar, as it is $(-h^2\Delta_h)$, defined in (2.20) to solve the 1D case of the Poisson equation with Dirichlet boundary conditions.

The matrix representation of (5.1) is given by combining (5.6), (5.7), (5.9) as follows:

$$S_h = \frac{12}{h^2}\left[\frac{1}{2h}K\frac{3}{h}P^{-1}K - \left(-\frac{1}{h^2}T\right)\right].$$

Simplifying the above we get,

$$S_h = \frac{6}{h^4}\left[3KP^{-1}K + 2T\right]. \tag{5.10}$$

Notice that K and P commute almost everywhere except the corners, and by simple calculation it can be shown that,

$$PK - KP = \begin{bmatrix} -2 & & & & \\ & 0 & & & \\ & & \ddots & & \\ & & & 0 & \\ & & & & 2 \end{bmatrix}. \tag{5.11}$$

We define an operator σ_h on u_h,

$$\sigma_h u_i = \frac{1}{6}u_{i-1} + \frac{2}{3}u_i + \frac{1}{6}u_{i+1}, \quad 1 \leq i \leq N-1. \tag{5.12}$$

We will refer to σ_h as the Simpson operator, and its matrix representation is $\frac{1}{6}P$.

Lemma 5.1. *Given a vector u_h satisfying the boundary conditions, then*

$$\sigma_h \delta^{(4)} u_i = \delta_h^2 \delta_h^2 u_i \quad 2 \leq i \leq N-2.$$

Proof. Using (5.4) we can establish the relationship

$$\delta_h^* u_i = \sigma_h v_i, \quad 1 \leq i \leq N-1. \tag{5.13}$$

As noted above the matrix representation for σ_h is $\frac{P}{6}$. We may also observe that $P = 6I - T$ where I is the identity, and we may conclude that,

$$\sigma_h = \frac{P}{6} = I - \frac{T}{6} = I + \frac{h^2}{6}\delta_h^2. \tag{5.14}$$

From (5.11) we can establish that for $2 \leq i \leq N-2$,

$$\delta_h^* \sigma_h u_i = \frac{K}{2h}\frac{P}{6}u_i = \frac{P}{6}\frac{K}{2h}u_i = \sigma_h \delta_h^* u_i. \tag{5.15}$$

We now investigate the effect of taking $\sigma_h \delta_h^{(4)}$ at the interior points. We have

$$\sigma_h \delta_h^{(4)} u_i = \frac{1}{6}\delta_h^{(4)} u_{i-1} + \frac{2}{3}\delta_h^{(4)} u_i + \frac{1}{6}\delta_h^{(4)} u_{i+1}, \quad 2 \leq i \leq N-2. \tag{5.16}$$

From (5.1) we see that,

$$\delta_h^{(4)} u_i = \frac{12}{h^2}\left(\delta_h^* v_i - \delta_h^2 u_i\right), \tag{5.17}$$

and applying that to (5.16) we get,

$$\sigma_h \delta_h^{(4)} u_i = \frac{1}{6}\delta_h^{(4)} u_{i-1} + \frac{2}{3}\delta_h^{(4)} u_i + \frac{1}{6}\delta_h^{(4)} u_{i+1} \tag{5.18}$$

$$= \frac{12}{h^2}\left(\left[\frac{1}{6}\delta_h^* v_{i-1} + \frac{2}{3}\delta_h^* v_i + \frac{1}{6}\delta_h^* v_{i+1}\right] - \left[\frac{1}{6}\delta_h^2 u_{i-1} + \frac{2}{3}\delta_h^2 u_i + \frac{1}{6}\delta_h^2 u_{i+1}\right]\right). \tag{5.19}$$

Investigating the first term we get,

$$\begin{aligned}
\frac{1}{6}\delta_h^* v_{i-1} + \frac{2}{3}\delta_h^* v_i + \frac{1}{6}\delta_h^* v_{i+1} &= \sigma_h \delta_h^* v_i \\
&= \delta_h^* \sigma_h v_i \\
&= \delta_h^* \delta_h^* u_i \\
&= \frac{1}{4h^2}(u_{i-2} - 2u_i + u_{i+2}), \quad 2 \leq i \leq N-2.
\end{aligned}$$

In the above we have used (5.13) and (5.15). We got,

$$\frac{1}{6}\delta_h^* v_{i-1} + \frac{2}{3}\delta_h^* v_i + \frac{1}{6}\delta_h^* v_{i+1} = \frac{1}{4h^2}(u_{i-2} - 2u_i + u_{i+2}), \quad 2 \leq i \leq N-2. \tag{5.20}$$

Applying the definition of δ_h^2 to the second term in (5.19) we get,

$$\frac{1}{6}\delta_h^2 u_{i-1} + \frac{2}{3}\delta_h^2 u_i + \frac{1}{6}\delta_h^2 u_{i+1} = \frac{1}{6h^2}(u_{i-2} + 2u_{i-1} - 6u_i + 2u_{i+1} + u_{i+2}), \quad 2 \le i \le N-2. \tag{5.21}$$

Plugging (5.20) and (5.21) in (5.19) we get for $2 \le i \le N-2$,

$$\sigma_h \delta^{(4)} u_i = \frac{1}{h^4}(u_{i-2} - 4u_{i-1} + 6u_i - 4u_{i+1} + u_{i+2}) = \delta_h^2 \delta_h^2 u_i. \tag{5.22}$$

□

Theorem 5.2. *The scheme S_h is fourth order consistent in the interior and first order on the near boundary points.*

Proof. From Lemma 5.1 we know that

$$\sigma_h \delta^{(4)} u_i = \delta_h^2 \delta_h^2 u_i \quad 2 \le i \le N-2.$$

We expand the term $\delta_h^2 \delta_h^2$ using Taylor series. We begin by,

$$u_{i\pm 1} = u_i \pm \frac{du_i}{dx} h + \frac{d^2 u_i}{dx^2}\frac{h^2}{2} \pm \frac{d^3 u_i}{dx^3}\frac{h^3}{6} + \frac{d^4 u_i}{dx^4}\frac{h^4}{24} \pm \frac{d^5 u_i}{dx^5}\frac{h^5}{120} + \frac{d^6 u(\xi^\pm)}{dx^6}\frac{h^6}{720},$$

for some $\xi^\pm \in \Omega_h$. Adding the two equalities above and simplifying we get

$$\delta_h^2 u_i = \frac{d^2 u_i}{dx^2} + \frac{d^4 u_i}{dx^4}\frac{h^2}{12} + \frac{d^2 u(\xi^\pm)}{dx^6}\frac{h^4}{360}. \tag{5.23}$$

Applying δ_h^2 to (5.23) at the interior points for $2 \le i \le N-2$, we get after simplification,

$$\sigma_h \delta_h^{(4)} u_i = \delta_h^2 \delta_h^2 u_i = \frac{d^4 u_i}{dx^4} + \frac{d^6 u_i}{dx^6}\frac{h^2}{6} + E_1, \quad \text{and} \quad \|E_1\|^*_{C^8(\overline{\Omega})} \le c_1 \|u\|^*_{C^8(\overline{\Omega})} h^4. \tag{5.24}$$

Now we expand the term $\sigma_h \frac{d^4 u_i}{dx^4}$ around the interior points $2 \le i \le N-2$,

$$\begin{aligned} \sigma_h \frac{d^4 u_i}{dx^4} &= (I + \frac{h^2}{6}\delta_h^2)\frac{d^4 u_i}{dx^4} \\ &= \frac{d^4 u_i}{dx^4} + \frac{d^6 u_i}{dx^6}\frac{h^2}{6} + E_2, \quad \text{and} \quad \|E_2\|^*_{C^8(\overline{\Omega})} \le c_2 \|u\|^*_{C^8(\overline{\Omega})} h^4. \end{aligned}$$

We have shown that,

$$\sigma_h \frac{d^4 u_i}{dx^4} = \frac{d^4 u_i}{dx^4} + \frac{d^6 u_i}{dx^6}\frac{h^2}{6} + E_2, \quad \text{where} \quad \|E_2\|^*_{C^8(\overline{\Omega})} \le c_2 \|u\|^*_{C^8(\overline{\Omega})} h^4. \tag{5.25}$$

Taking the $|\cdot|_2$ norm of the difference between (5.24) and (5.25) gives us the desired result, namely

$$\left| \sigma_h \delta^{(4)} u_i - \sigma_h \frac{d^4 u_i}{dx^4} \right|_2 \le c \|u\|^*_{C^8(\overline{\Omega})} h^4, \quad 2 \le i \le N-2. \tag{5.26}$$

We will now work on the consistency errors of near boundary points. We will demonstrate the error at u_1, as the proof for the error at u_{N-1} is similar. Below, the terms c, c_1, c_2 are positive

independent constants
We set $\delta_h^{(4)} u_0 = (\frac{d^4 u_0}{dx^4})$. Using the definition of $\sigma_h \delta_h^{(4)}$ we get

$$\sigma_h \delta_h^{(4)} u_1 - \sigma_h \frac{d^4 u_1}{dx^4} = \left(\frac{2}{3}\delta_h^{(4)} u_1 + \frac{1}{6}\delta_h^{(4)} u_2\right) - \left(\frac{2}{3}\frac{d^4 u_1}{dx^4} + \frac{1}{6}\frac{d^4 u_2}{dx^4}\right) \tag{5.27}$$

$$= \frac{2}{3}\left(\delta_h^{(4)} u_1 - \frac{d^4 u_1}{dx^4}\right) + \frac{1}{6}\left(\delta_h^{(4)} u_2 - \frac{d^4 u_2}{dx^4}\right). \tag{5.28}$$

Using the boundary condition $u_0 = v_0 = 0$ and Lemma 10.1 on page 152 of [2] we get,

$$\delta_h^* v_1 = \frac{v_2}{2h} = \frac{d^2 u_1}{dx^2} + \frac{d^4 u_1}{dx^4}\frac{h^2}{6} + R_1, \quad \text{where} \quad \|R_1\|^*_{C^5(\overline{\Omega})} \le c_1 \|u\|^*_{C^5(\overline{\Omega})} h^3, \tag{5.29}$$

as well as,

$$\delta_h^2 u_1 = \frac{d^2 u_1}{dx^2} + \frac{d^4 u_1}{dx^4}\frac{h^2}{12} + R_2, \quad \text{where} \quad \|R_2\|^*_{C^5(\overline{\Omega})} \le c_2 \|u\|^*_{C^5(\overline{\Omega})} h^3. \tag{5.30}$$

Plugging (5.29) and (5.30) into the definition of $\delta_h^{(4)}$from (5.17) we get,

$$\delta_h^{(4)} u_1 = \frac{12}{h^2}\big(\delta_h^* v_1 - \delta_h^2 u_1\big) = \frac{d^4 u_1}{dx^4} + R_3, \quad \text{where} \quad \|R_3\|^*_{C^5(\overline{\Omega})} \le c\|u\|^*_{C^5(\overline{\Omega})} h.$$

We have shown that,

$$\delta_h^{(4)} u_1 - \frac{d^4 u_1}{dx^4} = R_3, \quad \text{where} \quad \|R_3\|^*_{C^5(\overline{\Omega})} \le c\|u\|^*_{C^5(\overline{\Omega})} h. \tag{5.31}$$

Following the same steps, inequalities (5.29), (5.30) and (5.31) can be applied to point u_2,that is

$$\delta_h^{(4)} u_2 - \frac{d^4 u_2}{dx^4} = S_3, \quad \text{where} \quad \|S_3\|^*_{C^5(\overline{\Omega})} \le c\|u\|^*_{C^5(\overline{\Omega})} h. \tag{5.32}$$

Plugging (5.31) and (5.32) into (5.28) we get the desired result, i.e.

$$|\sigma_h \delta_h^{(4)} u_1 - \sigma_h(\frac{d^4 u_1}{dx^4})|_2 \le c\|u\|^*_{C^5(\overline{\Omega})} h. \tag{5.33}$$

As stated before, a similar estimate can be made at u_{N-1}, i.e.

$$|\sigma_h \delta_h^{(4)} u_{N-1} - \sigma_h(\frac{d^4 u_{N-1}}{dx^4})|_2 \le c\|u\|^*_{C^5(\overline{\Omega})} h. \tag{5.34}$$

From [2] page 163 equation 10.87 we find $\left|\left(\frac{P}{6}\right)^{-1}\right|_2 \leq 3$, applying that to equations (5.26), (5.33) and (5.34) we get:

for $2 \leq i \leq N-2$,

$$\begin{aligned}
\left|\delta^{(4)} u_i - \frac{d^4 u_i}{dx^4}\right|_2 &= \left|(\sigma_h^{-1})\left(\sigma_h \delta^{(4)} u_i - \sigma_h \frac{d^4 u_i}{dx^4}\right)\right|_2 \\
&\leq |(\sigma_h^{-1})|_2 \left|\sigma_h \delta^{(4)} u_i - \sigma_h \frac{d^4 u_i}{dx^4}\right|_2 \\
&\leq 3\left|\sigma_h \delta^{(4)} u_i - \sigma_h \frac{d^4 u_i}{dx^4}\right|_2 \\
&\leq c\|u\|^*_{C^8(\overline{\Omega})} h^4.
\end{aligned}$$

For $i = 1$ and $i = N-1$, following similar steps we get

$$|\delta_h^{(4)} u_1 - (\frac{d^4 u_1}{dx^4})|_2 \leq c\|u\|^*_{C^5(\overline{\Omega})} h,$$

$$|\delta_h^{(4)} u_{N-1} - (\frac{d^4 u_{N-1}}{dx^4})|_2 \leq c\|u\|^*_{C^5(\overline{\Omega})} h.$$

If we wish to write a consistency error equation involving all points, we would have to take the lowest estimate which happens at the boundary points, we may conclude that

$$|\delta_h^{(4)} u_h - (\frac{d^4 u_h}{dx^4})|_2 \leq ch(\|u\|^*_{C^5(\overline{\Omega})} + \|u\|^*_{C^8(\overline{\Omega})}). \tag{5.35}$$

If we use the $|\cdot|_h$ we get,

$$|\delta_h^{(4)} u_h - (\frac{d^4 u_h}{dx^4})|_h \leq ch^{3/2}(\|u\|^*_{C^5(\overline{\Omega})} + \|u\|^*_{C^8(\overline{\Omega})}),$$

and using similar notation as with previous schemes, the above becomes

$$|S_h R_h u - R_h \Delta^2 u|_h \leq ch^{3/2}(\|u\|^*_{C^5(\overline{\Omega})} + \|u\|^*_{C^8(\overline{\Omega})}), \tag{5.36}$$

$\square$

Theorem 5.3. *The matrix S_h is a monotone matrix*

Proof. By personal correspondence with Haggai Katriel, who is working with Matania Ben-Artzi, one of the authors of [2] I acquired a preprint result for the explicit formula for the solution. For $1 \leq i \leq N-1$

$$u_i = \frac{h}{6}\left[\sum_{k=1}^{i-1}(x_i - x_k)^3 f_k + x_i^2 \sum_{k=1}^{N-1}(1-x_k)^2(2(1-x_i)x_k + x_k - x_i)f_k\right], \tag{5.37}$$

where $u_i = u(x_i)$. We can rewrite it as,

$$u_i = \frac{h}{6}\left[(1-x_i)^2 \sum_{k=1}^{i-1} x_k^2(2x_i(1-x_k) + x_i - x_k)f_k + x_i^2 \sum_{k=i}^{N-1}(1-x_k)^2(2x_k(1-x_i) + x_k - x_i)f_k\right],$$

(remembering that $x_i = ih$, from which we can see, that $1 - x_i = x_{N-i}$).This gives us,

$$u_i = \frac{h}{6}\left[x_{N-i}^2 \sum_{k=1}^{i-1} x_k^2(2x_i x_{N-k} + x_i - x_k)f_k + x_i^2 \sum_{k=i}^{N-1} x_{N-k}^2(2x_k x_{N-i} + x_k - x_i)f_k\right]. \tag{5.38}$$

In other words, defining

$$M_{ik} = \frac{h}{6}x_{N-i}^2 x_k^2(2x_i x_{N-k} + x_i - x_k), \quad 1 \leq k \leq i \leq N-1, \tag{5.39}$$

and using (5.39) we can rewrite (5.38) as,

$$u_i = \sum_{k=1}^{i-1} M_{ik} f_k + \sum_{k=i}^{N-1} M_{ki} f_k. \tag{5.40}$$

Equation (5.40) can be written as

$$u_h = M_h f_h. \tag{5.41}$$

We now show all elements of the matrix M_h are nonnegative.
Without loss of generality we take the lower triangular part of M_h, M_{ik} where $1 \leq k \leq i \leq N-1$. From (5.39),

$$\begin{aligned} M_{ik} =& \frac{h}{6}x_{N-i}^2 x_k^2(2x_i x_{N-k} + x_i - x_k) \\ =& \frac{h}{6}x_{N-i}^2 x_k^2(2(ih)(N-k)h + ih - kh) \\ =& \frac{h}{6}x_{N-i}^2 x_k^2(2ih^2(N-k) + h(i-k)) \\ \geq& 0. \end{aligned}$$

In the above we have used that $k \leq i \leq N-1$. From (5.41) it is easy to see that matrix M_h is in fact S_h^{-1}, and since we have shown matrix $M_h \geq 0$ we can conclude that so is $S_h^{-1} \geq 0$. Using Claim 3.7 we conclude S_h to be monotone. □

Theorem 5.4. *The scheme S_h is stable and* $|S_h^{-1}|_\infty \leq \frac{1}{384}$.

Proof. We wish to show that,

$$\frac{|S_h^{-1} f_h|_\infty}{|f_h|_\infty} = \frac{|u_h|_\infty}{|f_h|_\infty} \leq \frac{1}{384}.$$

We now define the vector $w_h = [w_{1,1}, w_{2,1}, ..., w_{n-1,1}, w_{1,2}, w_{2,2}, ..., w_{n-1,2}, ...w_{n-1,n-1}]^T$, where,

$$w_{i,j} = \frac{(ih)^2(1-ih)^2}{24},$$

and the function,

$$w = \frac{x^2(1-x)^2}{24} \quad 0 \leq x \leq 1.$$

We consider w as a function and w_h the discretization of it, defined as $w_h = R_h w$. We now show that $S_h w_h = \mathbf{1}$. By simple calculation we can see that $(\Delta)^2 w = \mathbf{1}$ and $\mathbf{1} - S_h w_h = R_h(\Delta)^2 w - S_h R_h w = 0$, since we have shown the consistency error involves fifth and eighth derivatives of w, all of which vanish. This proves the claim $S_h w_h = \mathbf{1}$.

It is easily seen that $|w_h|_\infty \leq \frac{1}{384}$. Now,

$$S_h(|f_h|_\infty w_h + u_h) = |f|_\infty \mathbf{1} + f_h \geq 0.$$

Since S_h is monotone,

$$|f_h|_\infty w_h + u_h \geq 0.$$

From here we have,

$$-u_h \leq |f_h|_\infty w_h \leq |w_h|_\infty |f_h|_\infty \mathbf{1} \leq \frac{1}{384} |f_h|_\infty \mathbf{1}. \tag{5.42}$$

Similarly,

$$S_h(|f_h|_\infty w_h - u_h) = |f|_\infty \mathbf{1} - f_h \geq 0.$$

Since S_h is monotone,

$$|f_h|_\infty w_h - u_h \geq 0.$$

From here we find that,

$$u_h \leq |f_h|_\infty w_h \leq |w_h|_\infty |f_h|_\infty \mathbf{1} \leq \frac{1}{384} |f_h|_\infty \mathbf{1}. \tag{5.43}$$

Combining (5.42) and (5.43) we get

$$|u_h|_\infty \leq \frac{1}{384} |f_h|_\infty.$$

We have proved the scheme to be stable and

$$|S_h^{-1}|_\infty \leq \frac{1}{384}. \tag{5.44}$$

□

Theorem 5.5. *The scheme S_h is convergent.*

Proof. We wish to demonstrate the scheme is convergent with respect to the $|\cdot|_h$ and the infinity norms. Using (5.44) and (5.36) we obtain

$$\begin{aligned}
|R_h u - u_h|_h &= |S_h^{-1} S_h (R_h u - u_h)|_h \\
&\leq |S_h^{-1}|_h |S_h (R_h u - u_h)|_h. \\
&\leq \frac{1}{384} |S_h R_h u - R_h f|_h \\
&\leq c h^{3/2} (\|u\|^*_{C^5(\overline{\Omega})} + \|u\|^*_{C^8(\overline{\Omega})}).
\end{aligned}$$

As before, the term c is a positive constant independent of u. We have shown convergence of order 3/2 in the $|\cdot|_h$,

$$|R_h u - u_h|_h \leq c h^{3/2} (\|u\|^*_{C^5(\overline{\Omega})} + \|u\|^*_{C^8(\overline{\Omega})}).$$

We can also show convergence of first order in the infinity norm, since

$$\begin{aligned}
|R_h u - u_h|_\infty &\leq |R_h u - u_h|_2 \\
&= h^{-1/2} |R_h u - u_h|_h \\
&\leq h^{-1/2} c h^{3/2} (\|u\|^*_{C^5(\overline{\Omega})} + \|u\|^*_{C^8(\overline{\Omega})}) \\
&\leq c h (\|u\|^*_{C^5(\overline{\Omega})} + \|u\|^*_{C^8(\overline{\Omega})}).
\end{aligned}$$

□

Chapter 6

Conclusions

There are many ways to show a numerical scheme is convergent. Throughout this thesis we have established a scheme is consistent using Taylor series expansion and proved it is stable. Many methods exist to show stability but we have focused mainly on using the fact that the matrix of the scheme is monotone and applying the properties of monotone matrices.

The scope of this thesis is limited to five numerical schemes, two for for the 1D Poisson equation, two for the 2D Poisson equation and one for the Biharmonic equation. No work previously has used that the numerical schemes introduced in chapters three and four are monotone to prove stability. The ideas in this thesis can be expanded to some other numerical schemes.

Unfortunately, establishing whether a matrix of a scheme is a monotone matrix or not could be difficult or simply impossible since it might even be false. Nevertheless, in cases where a scheme is proven to be monotone it could be a powerful tool to assist in showing stability and subsequently, convergence.

This page includes all notation used in the thesis and on what page they appeared first.

δ_h – forward finite difference scheme to approximate the first derivative. $p2$

δ_{-h} – backward finite difference scheme to approximate the first derivative. $p2$

$|\cdot|_\infty$ – ℓ_∞ norm for vectors. $p2$

$|\cdot|_2$ – ℓ_2 vector norm. $p2$

$|\cdot|_h$ – discrete ℓ_2 norm. $p2$

$H^1_0(o,\ell)$ – Soboloev space. $p3$

$|\cdot|_{H^1}$ – discrete H^1norm. $p11$

$|\cdot|^*_{H^1}$ – a norm equivalent to the discrete H^1 norm. $p12$

Δ^*_h – a general scheme approximating the laplacian operator. $p17$

R_h – restriction operator. $p17$

$\|\cdot\|^*_{C^r(\overline{\Omega})}$ – a quantity used in proving consistency of numerical schemes. $p17$

$\kappa(M)$ – condition number. $p17$

Ω – problem domain. $p18$

Ω_h – discrete domain applied in $ch1.2.1$. $p18$

Δ_h – specific scheme approximating the laplacian operator applied in $ch1.2.1$. $p18$

$\tilde{\Omega}_h$ – discrete domain applied in $ch1.2.2$. $p22$

$\Delta^{(3)}_h$ – specific scheme approximating the laplacian operator applied in $ch1.2.2$. $p12$

$v > 0$ – all entries of vector v are nonnegative and at least one element is positive. $p29$

$v \gg 0$ – all entries of vector v are positive. $p29$

$v \geq 0$ – all entries of vector v are nonnegative. $p29$

$M \geq 0$ – all entries of matrix M are nonnegative. $p29$

$\rho(M)$ – spectral radius. $p31$

$\hat{\Delta}_h$ – specific scheme approximating the laplacian operator applied in $ch3.1$. $p37$

$\Delta^{(4)}_h$ – nine point approximate for the laplacian used inside $\hat{\Delta}_h$. $p39$

$\tilde{\Delta}_h$ – specific scheme approximating the laplacian operator applied in $ch3.2$. $p46$

$\tilde{R}_h$ – Special restriction operator used in $ch3.2$. $p50$

δ^*_h – central finite difference scheme to approximate the first derivative. $p51$

δ^2_h – central finite difference scheme to approximate the second derivative. $p51$

$\delta^{(4)}_h$ – scheme approximating the fourth derivative. $p52$

S_h – specific scheme approximating the biharmonic operator applied in $ch4$. $p53$

σ_h – Simpson operator used in $ch4$. $p54$

$\mathbf{1}$ – vector of ones. $p22$

Bibliography

[1] Abraham Berman and Robert J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences*, Academic Press, 1979.

[2] Ben-Artzi Matania, Jean-Pierre Croisille, and Dalia Fishelov, *Navier-Stokes equations on planar domains*, Imperial College Press, 2013.

[3] Carl D. Meyer, *Matrix Analysis and Applied Linear Algebra*, S.I.A.M, 2000.

[4] Harvey S. Price, *Monotone and oscillation matrices applied to finite difference approximations*, Mathematics of Computation **22** (1968), 489-516.

[5] J. R. Kuttler, *A fourth order finite-difference approximation for the fixed membrane eigenproblem*, Mathematics of Computation **25** (1971), 237-256.

[6] James H. Bramble and Bert E. Hubbard, *New monotone type approximations for elliptic problems*, Mathematics of Computation **18** (1964), 349-367.

[7] S.H. Lui, *Numerical Analysis of Partial Differential Equations*, Wiley, 2011.

[8] Yu S. Volkov and V. L. Miroshnichenko, *Norm estimates for the inverses of matrices of monotone type and totally positive matrices*, Siberian Mathematical Journal **50** (2009), 982-987.