STABLE AND HIGH ORDER ACCURATE FINITE DIFFERENCE METHOD FOR THE INCOMPRESSIBLE LAMINAR BOUNDARY LAYER EQUATIONS

Presented by

MOJALEFA PRINCE NCHUPANG
Industrial Computational Fluid Dynamics Research Group
Department of Mechanical Engineering
University of Cape Town

Supervised by

Prof. Arnaud G. Malan
Industrial Computational Fluid Dynamics Research Group
Department of Mechanical Engineering
University of Cape Town

Prof. Jan Nordström
Computational Mathematics
Department of Mathematics
Linköping University, *Sweden*

A dissertation submitted to the University of Cape Town in fulfillment of the requirements for the degree of Master of Science in Engineering

January 2020

University of Cape Town

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

University of Cape Town

Declaration of authorship

I, Mojalefa Prince Nchupang, declare that this dissertation titled 'Stable and high order accurate finite difference method for the incompressible laminar boundary layer equations' and the work presented in it are my own. I can confirm that:

- This work was done wholly or mainly while in candidature for a masters degree at the University of Cape Town (UCT);
- Where any part of this dissertation has previously been submitted for a degree or any other academic qualification at UCT or any other institution, this has been clearly stated;
- Where I have consulted the published work of others, this is always clearly attributed;
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this dissertation is entirely my own work;
- I have acknowledged all main sources of help;

Signed by candidate

Signed: .....

Date: .....

This work is based on research supported in part by the National Research Foundation of South Africa (Grant Numbers: 89916). The opinions, findings and conclusions or recommendations expressed is that of the authors alone, and the NRF accepts no liability whatsoever in this regard.

Abstract

Numerical simulations of incompressible flows are unequivocally important due to their numerous industrial applications. These applications ranges from the large-scale fluid's flow modelling such as aerodynamics [1], atmospheric-ocean modelling [2] to a simple pipe flows in the petroleum industry [3]. This study is devoted to develop a provably stable and high order approximation for the incompressible laminar boundary layer equations. A new set of energy-stable boundary conditions are derived using the energy method. It is shown that both the weak and strong implementation of these boundary conditions yields an energy estimate. The semi-discrete problem is formulated by discretizing the continuous spatial derivatives using high order finite difference approximations on summation-by-parts form. The boundary conditions are implemented weakly using the simultaneous approximation terms methods. The discrete energy estimate is derived by mimicking the continuous analysis and hence, the numerical approximation is proved to be stable. The accuracy and linear stability of the developed scheme is also validated by solving the celebrated laminar flat plate flow problem. This is done by injecting the Blasius solution into the coefficient matrix as well as weak boundary conditions.

Acknowledgments

I would like to express my sincere gratitude and deep appreciation to my supervisors Prof. Arnaud Malan and Prof. Jan Nordström for their immense support throughout this study. Their reviews, valuable and constructive criticism, and guide made this journey possible.

I extend my appreciation to Prof. Malan for financial assistance through South African Research Chair Initiatives, the University of Cape Town's postgraduate financial aid and lastly, the North West's Department of Education and Sports Development.

My gratitude also goes to my mentor Dr. Tomas Lundquist, who taught me and clarified some concepts in computational mathematics. To my fellow InCFD research group mates, it has been a pleasure to work with you and thank you for creating a conducive working environment. Lastly, I would like to thank my family, friends and UCT mountain and Ski club members for their love and support.

Baie dankie! Tack så mycket! Ke a leboga!

Contents

Nomenclature

Abbreviations

CFD	Computational Fluid Dynamics
HOFDMs	High Order Finite Difference Methods
HOMs	High Order Methods
IBVP	Initial-Boundary Value Problem
LHS	Left-Hand-Side
N-S	Navier-Stokes
ODE	Ordinary Differential Equation
PDE	Partial Differential Equation
RHS	Right-Hand Side
SAT	Simultaneous Approximation Term
SBP	Summation-By-Parts

Greek Symbol

λ	Eigenvalue
$\Lambda^{\pm}$	Positive and negative eigenvalue matrix
μ	Dynamical viscosity
ν	Kinematic viscosity
Ω	Domain
$\partial\Omega$	Boundary of Ω
ρ	Density
ε	Diffusion constant

Mathematical operators and symbols

$\mathcal{O}$	Order
∇	Gradient operator
$\otimes$	Kronecker product
$\partial_{\mathbf{x}}$	Derivative operator
$\det(\cdot)$	Determinant

Set notation

$\in$	Element of
$\mathbb{R}$	Field of real numbers
$\subset$	Subset
$M_{n\times m}(\mathbb{R})$	Set of $n \times m$ real matrices

Roman Symbol

$\mathbf{n}$	Unit normal vector
$\mathbf{x} = (x, y)$	Spatial coordinates
$D_{x,y}$	SBP first derivative operators
$D_{xx,yy}$	SBP second derivative operators
n-D	n dimensions
$n_{\mathbf{x}}$	Components of a normal vector
$P_{x,y}$	Diagonal-norm matrix
$u_{\mathbf{n}}$	Normal velocity

Chapter 1

Introduction

The complexity of modelling incompressible flows with the Navier-Stokes (N-S) equations arises as a results of the fluid's density being invariant with respect to pressure. The incompressible N-S equations are coupled system of nonlinear partial differential equations (PDEs). In a general 3-dimensional flow domain with the isothermal conditions, this system consists of four equations viz. the conservation of mass and three momentum conservation equations with four flow variables. Unlike compressible flow modelling where the equation of state defines the relationship between density and pressure, incompressible flow contains no such relation.

One common practice in Computational Fluid Dynamics (CFD), a subfield of fluid mechanics that uses physics conservation laws and numerical methods to describe fluid's motions, is to decouple the pressure from the momentum equations [4, 5, 6]. In these so-called fractional-step methods, the pressure is solved implicitly from the pressure Poisson equation obtained by taking the divergence of the momentum equations. The constant density is assumed to ensure the divergence-free solution. Another way to compute the pressure, is to introduce an artificial compressibility term [7, 8, 9] in the continuity equation. Thus, a fictitious pressure derivative term is added to the continuity equation and the system is solved such that the artificial pressure term vanishes.

In the present study, we however consider the nonlinear incompressible N-S equations in the original form (i.e. the pressure is not decoupled from the momentum equations). We further simplify this system using the so-called flat-plate theory [10] assumptions to obtain the laminar boundary layer equations. We then construct an energy stable and high order finite difference method (HOFDM) [11, 12] to compute an approximate solutions of these equations. This analysis will provide a basic framework for the solution procedure required to solve the general incompressible N-S equations.

High order methods (HOMs) are efficient and computationally inexpensive since for a fixed number of degrees of freedom, the accuracy of the numerical solution to the initial-boundary

value problem (IBVP) can be increased. Also, the number of degrees of freedom can be reduce significantly for a given error tolerance. The concerns that arises when using these methods have to do with the difficulty to guarantee stability. The stability of the numerical approximations ties back to the well-posedness of the continuous problem which fundamentally relies on the choice of the boundary conditions. In particular, for linear IBVPs, the correct minimal number and appropriate form of boundary conditions which guarantee a unique and bounded solution must be determined prior the discretization. Once this has been established, the discrete problem approximating the solution of continuous problem is formulated. The numerical implementation of these boundary conditions such that there is no non-physical growth in the solution is challenging.

Although HOMs are efficient and flexible for problems defined on complex spatial domains, it is difficult to accurately and stably approximate the solution in the vicinity of the boundaries. These concerns were addressed in [13] when Carpenter *et al.* introduced a penalty-like boundary treatment called the Simultaneous Approximation Term (SAT). This method enforces the boundary or interface conditions by weakly combining them with the governing PDEs. The stability of the numerical approximations are now reachable when the discrete differential operators on summation-by-parts (SBP) form are augmented with the SAT boundary treatment. The SBP-SAT technique continues to gain popularity in CFD for numerous reasons. Initially, this framework was formulated to construct provably stable high order approximations, however it has developed beyond that. The developments include hybridizing different numerical methods [14, 15] in an accurate and stable manner, improving the efficiency of numerical schemes by accelerating their convergence [16, 17] and recently it was extended to the time domain [18].

In this thesis, we derive a new set of energy-stable boundary conditions and stability condition for different orders of accuracy for the incompressible laminar boundary layer equations. The rest of the thesis is organized as follow: In Chapter 2, the concepts of well-posedness and stability are formulated. The energy method is introduced as a technique to construct well-posed problems and applied to the linear advection equation. The SBP-SAT technique, the structures and properties of the SBP matrices are outlined in Chapter 3. We also discuss their design orders of accuracy and end the chapter by extending these operators to the two dimensional (2-D) domain using the Kronecker product. The IBVPs are considered in Chapter 4. In particular, the linear advection-diffusion equation with constant and variable coefficients respectively is considered. Computational results are also presented to illustrate the performance of the SBP-SAT schemes. In Chapter 5, we present the novel contribution of this thesis i.e. a stable and high order-accurate numerical scheme for the laminar boundary layer equation. The summary, concluding remarks and on-going investigations are discussed in Chapter 6.

Chapter 2

Mathematical definitions

In this chapter, our objective is to introduce the mathematical definitions and concepts used in the present study. The theory of linear partial differential equations [19, 20] is presented. In particular, we focus on the well-posedness of the continuous problems, the stability of the corresponding discrete formulations and the energy method for constructing the energy estimates. As a motivational example, the linear advection equation with boundary conditions imposed both in a strong and weak sense is considered. The energy method is used to determine the required number and form of the boundary conditions such the continuous problem is well-posed. The Kronecker products which will be referenced in the upcoming analysis are defined.

2.1 Preliminaries

Let $\mathbb{R}$ be a field of real numbers equipped with the usual addition and multiplication operations. We define a set of real-valued continuous functions as

$$\mathcal{F}(\Omega, \mathbb{R}) = \{f : \Omega \to \mathbb{R} | f \text{ is continuous}\},$$

where $\Omega \subset \mathbb{R}^n$ and n is the number of space dimensions. The space of square-integrable functions is denoted by $L^2(\mathbb{R})$.
For any two vector functions $U(\mathbf{x})$, $V(\mathbf{x}) \in \mathcal{F} \subset L^2$, the inner products and the corresponding norms are defined by

$$\begin{aligned} (U,V)_\Omega &= \int_\Omega U^T H V dV, \qquad ||U||_\Omega^2 = (U,U)_\Omega, \\ (U,V)_{\partial\Omega} &= \oint_{\partial\Omega} U^T H V ds, \qquad ||U||_{\partial\Omega}^2 = (U,U)_{\partial\Omega}, \end{aligned} \tag{2.1}$$

where $H = H(\mathbf{x})$ is a symmetric, positive definite matrix and $\mathbf{x} = (x_1, x_2, \dots x_n)^T$. Further, $\int_\Omega$ and $\oint_{\partial\Omega}$ are volume and surface integral operating on the interior and boundary of Ω respectively. The volume and boundary surface area elements are indicated by dV and ds respectively.

Analogously to (2.1), we introduce the discrete inner product and norm for any real-valued

vectors $\mathbf{u}$, $\mathbf{v}$ with elements u_l, v_l as

$$(\mathbf{u},\mathbf{v})_{\tilde{H}_\Omega} = \sum_{l=1}^{N} u_l v_l h_{\Omega_l} \delta V_l = \mathbf{u}^T \tilde{H}_{\Omega_l} \mathbf{v}, \qquad ||\mathbf{u}||^2_{\tilde{H}_\Omega} = (\mathbf{u},\mathbf{u})_{\tilde{H}_\Omega}, \tag{2.2}$$
$$(\mathbf{u},\mathbf{v})_{\tilde{H}_{\partial\Omega}} = \sum_{l=1}^{N} u_l v_l h_{\partial\Omega_l} \delta s_l = \mathbf{u}^T \tilde{H}_{\partial\Omega_l} \mathbf{v}, \qquad ||\mathbf{u}||^2_{\tilde{H}_{\partial\Omega}} = (\mathbf{u},\mathbf{u})_{\tilde{H}_{\partial\Omega}}.$$

Here, h_{Ω_l}, $h_{\partial\Omega_l}$ denotes the elements of the diagonal positive definite matrices $\tilde{H}_\Omega$ and $\tilde{H}_{\partial\Omega}$ operating on the interior and boundary points of Ω_l. Moreover, they may be interpreted as discretely mimicking $\int_\Omega$ and $\oint_{\partial\Omega}$ respectively if $\tilde{H}_\Omega$ and $\tilde{H}_{\partial\Omega}$ are unit matrices (more on this later). In (2.2), δV_l and δs_l are the volume and area element of the lth computational cell respectively.

For later reference, we introduce the integration-by-parts rule in $n-$space dimension

$$(u, \partial_{\mathbf{x}} v)_\Omega = (u, n_{\mathbf{x}} v)_{\partial\Omega} - (\partial_{\mathbf{x}} u, v)_\Omega \tag{2.3}$$

and Gauss' theorem

$$(\nabla, u)_\Omega = (u, \mathbf{n})_{\partial\Omega}, \tag{2.4}$$

for $u, v \in \mathcal{F}$. In (2.3) and (2.4), $\partial_{\mathbf{x}}$ is a derivative operator, ∇ denotes a gradient operator and $n_{\mathbf{x}}$ is a scalar component of the unit outward pointing normal vector $\mathbf{n}$ on $\partial\Omega$. Closely related to Gauss' theorem (2.4), is a *lifting operator* $\mathcal{L}$ [21, 22] defined such that for any two vector functions $\psi, \xi \in \mathcal{F}$,

$$\big(\psi, \mathcal{L}(\xi)\big)_\Omega = (\psi, \xi)_{\partial\Omega} \tag{2.5}$$

holds. The operator $\mathcal{L}(\cdot)$ restricts the solution to the boundary points. The discrete analogue of (2.3) is called the summation-by-parts rule and it is defined for any two real-valued vectors $\mathbf{u}$, $\mathbf{v}$ as

$$(\mathbf{u}, \mathcal{D}_x \mathbf{v})_{\tilde{H}_\Omega} = (\mathbf{u}, \mathbf{v})_{\tilde{H}_{\partial\Omega}} - (\mathcal{D}_x \mathbf{u}, \mathbf{v})_{\tilde{H}_\Omega} \tag{2.6}$$

where $\mathcal{D}_x$ approximates $\partial\mathbf{x}$ discretely. The inner products $(\cdot,\cdot)_{\tilde{H}_\Omega}$ and $(\cdot,\cdot)_{\tilde{H}_{\partial\Omega}}$ are given in (2.2).

2.2 Well-posed problems

Consider the following IBVP posed on the domain $\Omega \subset \mathbb{R}^n$ with boundary $\partial\Omega$

$$\begin{aligned} u_t + D(\mathbf{x}, \partial_{\mathbf{x}}) u &= \mathbf{f}(t, \mathbf{x}), \qquad \mathbf{x} \in \Omega, \qquad t > 0 \\ Bu &= g(t), \qquad \mathbf{x} \in \partial\Omega, \quad t > 0 \\ u(0, \mathbf{x}), &= f(\mathbf{x}) \qquad \mathbf{x} \in \Omega, \quad t = 0. \end{aligned} \tag{2.7}$$

In (2.7), u is a dependent variable, D is a linear differential operator in space and the boundary operator denoted by B together with u defines the set of boundary conditions imposed on $\partial\Omega$. The forcing function, boundary and initial data are known and denoted by the continuous functions F, g and f respectively. In order to guarantee a smooth solution u, we assume that all three are compatible and sufficiently smooth.

Definition 1. The IBVP (2.3) with zero boundary data and no forcing function is *well-posed* if there exist a unique solution u such that the energy estimate

$$||u(t,\cdot)||_\Omega^2 \leq K(t)||f||_\Omega^2$$

holds where $K(t)$ is bounded for a finite time t and independent of f.

Definition 2. The IBVP (2.7) is *strongly well-posed* with nonzero data if it is well-posed and satisfies the estimate

$$||u(t,\cdot)||_\Omega^2 \leq K(t)\big(||f||_\Omega^2 + \int_0^t (||F(\tau,\cdot)||_\Omega^2 + ||g(\tau)||_{\partial\Omega}^2)d\tau\big)$$

where $K(t)$ is bounded and independent of F, g and f.

Although not shown yet, the energy estimate depends almost only on the boundary conditions if the corresponding Cauchy problem is well-posed. Therefore, in order for a unique and bounded solution to exist, the correct number and appropriate form of the boundary conditions are required. Over or under- specifying of the boundary conditions will not yield well-posedness since neither the existence nor uniqueness of the solution is possible. Next, we demonstrate that a small variation on the forcing function, initial and boundary data leads to a small variation in the solution.

Let's consider problem (2.7) with perturbed data $F + \delta F$, $g + \delta g$ and $f + \delta f$. By denoting the solution of the perturbed problem with v, we get

$$\begin{aligned} v_t + D(\mathbf{x}, \partial_{\mathbf{x}})v = F(t,\mathbf{x}) + \delta F, \qquad & \mathbf{x} \in \Omega, \qquad t > 0 \\ Bv = g(t) + \delta g, \qquad & \mathbf{x} \in \partial\Omega, \quad t > 0 \\ v(0,\mathbf{x}) = f(\mathbf{x}) + \delta f, \qquad & \mathbf{x} \in \Omega, \qquad t = 0. \end{aligned} \tag{2.8}$$

Next, we subtract (2.7) from (2.8) and define the difference $w = v - u$ to obtain

$$w_t + D(\mathbf{x}, \partial_{\mathbf{x}})w = \delta F \tag{2.9}$$

with the boundary and initial conditions defined as $Bw = \delta g$ and $w(0,x) = \delta f$ respectively. In (2.9), we used the assumption that both the differential and boundary operator are linear. The energy estimate of (2.9), using Definition 2, follows as

$$||w(t,\cdot)||_\Omega^2 \leq K(t)\big(||\delta f||_\Omega^2 + \int_0^t (||\delta F(\tau,\cdot)||_\Omega^2 + ||\delta g(\tau)||_{\partial\Omega}^2)d\tau\big).$$

Hence, for small deviations in data, the solution u and v are almost the same. Uniqueness of the solution follows directly by requiring δF, δf and , δg to turn zero.

2.3 Stability

Let $\mathbf{u}$ be a vector containing an approximation solution of (2.7). The semi-discrete formulation of (2.7) is

$$\begin{aligned} \mathbf{u}_t + \mathcal{D}\mathbf{u} &= \mathbf{F}, \qquad \mathbf{x}_i \in \Omega, \quad t > 0 \\ \mathcal{B}\mathbf{u} &= \mathbf{g}, \qquad \mathbf{x}_i \in \partial\Omega, \quad t > 0 \\ \mathbf{u} &= \mathbf{f}, \qquad \mathbf{x}_i \in \Omega, \quad t = 0 \end{aligned} \tag{2.10}$$

where F, $\mathbf{g}$ and $\mathbf{f}$ are vectors of the same size as $\mathbf{u}$ containing pointwise forcing, boundary and initial data respectively. The discrete differential and boundary operators mimicking D and B in (2.7) are denoted by $\mathcal{D}$ and $\mathcal{B}$. The concept of stability may be viewed as a discrete imitation of well-posedness. Roughly speaking, a numerical approximation scheme is stable if small perturbations of the data in the continuous problem leads to a small change in the numerical solution. We formalize this in the next two definitions.

Definition 3. The approximation (2.10) with $\mathbf{F} = \mathbf{0}$ and $\mathbf{g} = \mathbf{0}$ is *stable* if the estimate

$$||\mathbf{u}(t)||^2_{\tilde{H}_\Omega} \leq K(t)||\mathbf{f}||^2_{\tilde{H}_\Omega}$$

holds. Here, $K(t)$ is bounded for finite time domain and independent of $\mathbf{f}$. The norm $||\cdot||_{\tilde{H}_\Omega}$ is defined in (2.2).

Definition 4. The approximation (2.10) with nonzero data is *strongly stable* if it is stable and satisfies

$$||\mathbf{u}(t)||^2_{\tilde{H}_\Omega} \leq K(t)\big(||\mathbf{f}(t)||^2_{\tilde{H}_\Omega} + \max_{t\in[0,\tau]}||\mathbf{F}(\tau)||^2_{\tilde{H}_\Omega} + \max_{t\in[0,\tau]}||\mathbf{g}(\tau)||^2_{\partial\tilde{H}_\Omega}\big),$$

where $K(t)$ is bounded and independent of $\mathbf{F}, \mathbf{g}$ and $\mathbf{f}$.

2.4 The energy method

A well-posed problem as we have already seen, is characterized by the existence of a unique and bounded solution. The existence of such a solution depends on the correct minimal number of boundary conditions. The solution bound depends continuously on the boundary data which should be specified by an appropriate form of boundary conditions. In addition to this, we must know where to impose them on the boundary. There are at least three techniques available to construct the well-posed problems namely; the Fourier technique, the normal mode analysis (also known as Laplace transform analysis) and the energy method [19, 23]. These techniques possess different strengths and limitations, for example, existence, uniqueness and boundedness of the solution can be trivially proven when using the Fourier technique (which is based on the Fourier transforms). However, this technique is limited to Cauchy problems and problems defined on the periodic domains. The Laplace analysis method is suitable for PDEs

with high order derivatives. The question about the required number of boundary conditions and where to impose them can be answered using this technique, however, it does not provide information about the form that the boundary conditions must have.

The energy method is applicable to general space-time PDEs on finite domains. This technique is able to tackle all the three key questions regarding well-posedness. It involves multiplying the continuous governing equation with the solution and integrating over the domain. Because of this, it heavily relies on the integration-by-parts rule which introduce two requirements. First requirement, all the coefficient matrices involved in the problem are required to be symmetric. Secondly, a scalar product and the corresponding norm needs to be carefully chosen since not all norms will yield an estimate. Compared to Fourier and Laplace analysis, the energy method is a more suitable method for the construction of well-posed problems. If the coefficient matrices of the continuous problem are not symmetric, we symmetrize them if possible, for example see [24, 25]. The cited work involves the shallow water and linearized compressible Navier-Stokes equations which are preconditioned by the non-singular symmetrizing matrices respectively. We illustrate the energy method in the continuous case using the linear advection equation.

Let's consider the linear advection equation in 1-D with the advection speed $a > 0$

$$\begin{aligned} u_t + au_x = 0 \qquad & x \in [0,1], \quad t \geq 0 \\ Lu = g(t) \qquad & x \in 0 \cup 1, \quad t \geq 0 \\ u(x,0) = f(x) \qquad & x \in [0,1], \quad t = 0, \end{aligned} \tag{2.11}$$

where $f, g \subset \mathcal{F}$ defines initial and boundary data. Further, L is a linear boundary operator and its form or the boundary on which it is imposed is not yet known. By multiplying (2.11) with u and integrating over the domain, we obtain

$$\int_0^1 \frac{d}{dt}(\frac{u^2}{2})dx + \frac{1}{2}\int_0^1 (\frac{u^2}{2})_x dx = 0. \tag{2.12}$$

Next, we use norm (2.1) (with $H = 1$) and Gauss' theorem (2.4) to simplify (2.12) to

$$\frac{d}{dt}||u||^2_{x\in[0,1]} = au^2(0,t) - au^2(1,t), \quad t \geq 0. \tag{2.13}$$

Based on the energy rate (2.13), the question about the number of boundary conditions required, the form they must take and finally where to specify them on the boundary can now be answered. The right boundary term on the right-hand side (RHS) of (2.13) is squared and preceded by the desired sign i.e. it allows the energy rate to decrease as time progresses. Therefore, there is no need to prescribe boundary condition on the right boundary. However, the left boundary term is positive and adds growth to the estimate. To ensure bounded growth, a boundary condition must be imposed on the left boundary ($x = 0$). We have so far answered two questions i.e. precisely one boundary condition is needed and it must be

specified at the inlet boundary. The appropriate form of that boundary condition such that the energy estimate is obtained is still outstanding. Since the boundary term in (2.13) involves only the solution, we impose the Dirichlet boundary condition i.e. $u(0,t) = g(t)$. The energy rate becomes

$$\begin{aligned} \frac{d}{dt}||u||^2 &= ag(t)^2 - au^2(1,t) \\ &\leq ag(t)^2, \end{aligned} \tag{2.14}$$

which is bounded by data. Temporal integration over a finite time domain leads to

$$||u||^2_{x\in[0,1]} \leq ||f||^2_{x\in[0,1]} + a\int_0^t ||g(\tau)||^2_{x=0} d\tau,$$

which shows that (2.11) is strongly well-posed in the case of $F = 0$.

2.4.1 Weak boundary conditions

In (2.11), the solution is forced to satisfy the boundary condition at $x = 0$. This boundary treatment is known as *strong* implementation of the boundary conditions. Although we were able to obtain the estimate, this procedure often leads to stability issues in the numerical setting. To address this, an alternative way to impose boundary conditions, is to combine them with the governing equation by posing them as a penalty term. This involves forcing the solution at the boundary points to satisfy the boundary conditions by subtracting the exact amount it deviates from the boundary data. This boundary treatment is known as a *weak* implementation of the boundary conditions.

Using (2.11) as an example, we show that the weakly implemented boundary conditions also yields an estimate. Let's consider problem (2.11) with weakly imposed boundary condition at $x = 0$

$$\begin{aligned} u_t + au_x &= \mathcal{L}\big(\sigma(u - g(t))\big) \quad x \in [0,1], \quad t \geq 0 \\ u(x,0) &= f(x) \qquad\qquad\qquad\quad x \in [0,1], \quad t = 0. \end{aligned} \tag{2.15}$$

In (2.15), $\mathcal{L}(\cdot)$ is a lifting operator defined in (2.5) and σ denote a penalty coefficient to be determined such that the discrepancy between the solution and boundary data tends to zero. The energy method applied to (2.15) yields

$$\frac{d}{dt}||u||^2_{x\in[0,1]} = -au^2(1,t) + (a + 2\sigma)u^2(0,t) - 2\sigma u(0,t)g(t). \tag{2.16}$$

Proposition 2.1. *The energy estimate (2.16) is bounded by data if $\sigma = -a$.*

Proof. By substituting $\sigma = -a$, adding and subtracting $ag(t)^2$ on the RHS of (2.16), we get

$$\begin{aligned} \frac{d}{dt}||u||^2_{x\in[0,1]} &= ag(t)^2 - au^2(1,t) - \big(au(0,t) - g(t)\big)^2 \\ &\leq ag(t)^2. \end{aligned}$$

Temporal integration completes the proof. □

Remark 1. Note that the choice of σ in Proposition 2.1 is not unique.

The linear advection equation (2.15) with weakly imposed boundary conditions is hence strongly well-posed. In the next chapter, we will imitate the energy method in the discrete setting and construct a provably stable numerical scheme that approximates the solution of (2.11).

2.5 The Kronecker product

Let $M_{n\times m}(\mathbb{R})$ be a set of real matrices with $n-$ rows and $m-$columns. For arbitrary matrices $A \in M_{n\times m}$ and $B \in M_{p\times q}$, the Kronecker product of A and B is a block matrix of size $np \times mq$ denoted by

$$A \otimes B = \begin{bmatrix} a_{11}B & \dots & \dots & \dots & a_{1m}B \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & a_{ii}B & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ a_{n1}B & \dots & \dots & \dots & a_{nm}B \end{bmatrix}$$

where $a_{i,j}$ denotes the elements of A and $i \in \{1, \dots, n\}$, $j \in \{1, \dots, m\}$ respectively. The following Kronecker product properties are essential in this dissertation and will be referenced frequently in the upcoming analysis

1. $(A \otimes B)^T = A^T \otimes B^T$
2. $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$
3. $(A \otimes B)(C \otimes D) = AC \otimes BD$
4. $(ABC \otimes DEF) = (A \otimes D)(B \otimes E)(C \otimes F)$.

The last two properties holds only if the standard matrix multiplications hold.

Chapter 3

Summation-by-parts finite difference operators

Finite difference operators which satisfy the summation-by-parts rule (2.6) are considered in this chapter. These operators are constructed based on central-difference stencils and closed at the boundary points with one-sided finite difference approximations. Their structures and underlying properties are outlined. We proceed to show, using the model problem (2.15), that SBP operators when augmented with the simultaneous approximation terms yields discrete energy estimate that mimics the continuous one. We also study the orders of accuracy of these operators and close the chapter by extending them to 2-D for the purpose of the upcoming analysis.

3.1 The finite difference method

Finite difference methods are well-known and widely used to approximate the solutions of PDEs. They are favoured for their simplicity, computational efficiency and they can be modified locally to handle problems with discontinuities. High order approximations can also be easily derived [11]. The finite difference SBP operators considered in this study are constructed based on central-difference stencils. However, it is possible to construct them on non-centered stencils [26, 27] as well. The drawback with central-difference stencils is encountered near the boundary since they make use of solution values outside the computational domain. There are several procedures that can be used to correct this including using non-centered difference stencils [27] or one-sided difference stencils. In this work, we consider one-sided finite difference approximations near the boundaries. This often requires a drop in accuracy of the scheme near the boundary. To demonstrate this, we construct a second-order accurate central finite difference approximation for the first-derivative.

Let $u(x,t) \in C^{\infty}(\Omega)$ i.e. u is continuously differentiable in $\Omega \subset \mathbb{R}$. The Taylor expansion

of u around point $x_0 \in \Omega$ is

$$u(x_0+h,t) \approx u(x_0,t) + h\frac{\partial u}{\partial x}\Big|_{x_0} + \frac{h^2}{2}\frac{\partial^2 u}{\partial x^2}\Big|_{x_0} + \frac{h^3}{6}\frac{\partial^3 u}{\partial x^3}\Big|_{x_0} + \frac{h^4}{24}\frac{\partial^4 u}{\partial x^4}\Big|_{x_0} + \mathcal{O}(h^5) \tag{3.1}$$

where h is an infinitesimal distance from x to x_0. By replacing h with $-h$, we obtain

$$u(x_0-h,t) \approx u(x_0,t) - h\frac{\partial u}{\partial x}\Big|_{x_0} + \frac{h^2}{2}\frac{\partial^2 u}{\partial x^2}\Big|_{x_0} - \frac{h^3}{6}\frac{\partial^3 u}{\partial x^3}\Big|_{x_0} + \frac{h^4}{24}\frac{\partial^4 u}{\partial x^4}\Big|_{x_0} - \mathcal{O}(h^5). \tag{3.2}$$

A traditional second-order central difference approximating $\frac{\partial}{\partial x}$ is obtained by subtracting (3.2) from (3.1)

$$\frac{u(x_0+h,t)-u(x_0-h,t)}{2h} = \frac{\partial u}{\partial x}\Big|_{x_0} + \frac{h^2}{6}\frac{\partial^2 u}{\partial x^2}\Big|_{x_0} + \dots. \tag{3.3}$$

where $h \ll 1$ is preferable. The order of accuracy of the difference operator is q if the leading discretization error term is $\mathcal{O}(h^q)$. In (3.3), the leading term is of $\mathcal{O}(h^2)$ which makes the difference operator second-order accurate.

Notice that (3.3) approximate the derivative of u at x_0 only. In order to approximate the derivative of u on the close interval $[a,b] \subseteq \Omega$, we need to compute (3.3) at each grid point. Let N be a positive integer, we partition $x \in [a,b]$ into N equidistant subintervals given by $x_i = x_0 + ih$ where $h = (b-a)/N$ is a length of each subinterval and $i = 0, 1, \dots, N$. The left and right boundary points are $x_0 = a$ and $x_N = b$ respectively. Equation (3.3) for a general point x_i is

$$\frac{u_{i+1}-u_{i-1}}{2h} = \frac{\partial u_i}{\partial x} + \mathcal{O}(h^2). \tag{3.4}$$

Here, u_i is a short-hand notation for $u(x_i,t)$ and should not be confused with the partial derivatives. For arbitrary order of accuracy $2s$ ($s = 1, 2, \dots$), the form of the central finite difference approximation is

$$\frac{1}{h}\sum_{k=1}^{s} \alpha_k (u_{i+k} - u_{i-k}) = \frac{\partial u_i}{\partial x} + \mathcal{O}(h^{2s}) \tag{3.5}$$

where

$$\alpha_k = \frac{(-1)^{k+1}(s!)^2}{(s+k)!(s-k)!}.$$

In (3.4), we encounter a difficulty when $i = 0$ or $i = N$ since both u_{-1} and u_{N+1} lies outside the domain. As indicated previously, one-sided finite difference is used to approximate the solution near the boundary. Rearranging (3.1) and (3.2) respectively yields

$$\begin{aligned} \frac{u_{i+1}-u_i}{h} &= \frac{\partial u_i}{\partial x} + \mathcal{O}(h), \\ \frac{u_i-u_{i-1}}{h} &= \frac{\partial u_i}{\partial x} + \mathcal{O}(h). \end{aligned} \tag{3.6}$$

The approximations in (3.6) are called first-order Forward and Backward finite difference approximation respectively. The order of accuracy of (3.4) drops by order 1 at the boundaries. However, it can be shown [28] that the overall approximation becomes second-order accurate if the underlying IBVP is well-posed. By putting together (3.4) and (3.6), a second-order central finite difference approximating the first-derivative is

$$\begin{aligned} \frac{u_{i+1}-u_i}{h} &= \frac{\partial u_i}{\partial x} + \mathcal{O}(h), \qquad \text{if} \quad i = 0 \\ \frac{u_{i+1}-u_{i-1}}{2h} &= \frac{\partial u_i}{\partial x} + \mathcal{O}(h^2), \qquad \text{if} \quad i = 1, \dots, N-1 \\ \frac{u_i-u_{i-1}}{h} &= \frac{\partial u_i}{\partial x} + \mathcal{O}(h), \qquad \text{if} \quad i = N. \end{aligned} \tag{3.7}$$

3.2 The summation-by-parts operators

Even though finite difference SBP operators are efficient and flexible, obtaining stability proofs is not trivial. It was shown in [13] that, when posed in SBP form and used in conjunction with SAT boundary treatment, a discrete energy estimate can be obtained. As a motivation, we consider the discrete formulation of the model problem (2.15).

Approximating u_x using (3.7) yields

$$\begin{aligned} u_{i,t} &= g_0(t), && i = 0 \\ u_{i,t} + a\frac{u_{i+1}-u_{i-1}}{2h} &= 0 && i = 1, \dots, N-1 \\ u_{i,t} + a\frac{u_i-u_{i-1}}{h} &= 0 && i = N \\ u_{i,0} &= f_i && i = 1, \dots, N \end{aligned} \tag{3.8}$$

where g_i and f_i are grid functions projected on the points x_i indicating the boundary and initial data respectively. Here, the left boundary condition is implemented strongly. In (2.15), we were able to obtain the estimate with the boundary condition implemented weakly. We imitate this here by modifying the left boundary treatment to

$$u_{0,t} + a\frac{u_1-u_0}{h} = \sigma(u_0 - g_0).$$

Equation (3.8) with modified right boundary condition can now be rewritten in a matrix-vector form as

$$\frac{d}{dt}\begin{pmatrix} u_0 \\ u_1 \\ u_2 \\ \vdots \\ u_{N-2} \\ u_{N-1} \\ u_N \end{pmatrix} + \frac{a}{2h}\begin{pmatrix} -2 & 2 & 0 & 0 & 0 & \dots & 0 \\ -1 & 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & -1 & 0 & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots & \\ 0 & \dots & 0 & -1 & 0 & 1 & 0 \\ 0 & \dots & 0 & 0 & -1 & 0 & 1 \\ 0 & \dots & 0 & 0 & 0 & -2 & 2 \end{pmatrix}\begin{pmatrix} u_0 \\ u_1 \\ u_2 \\ \vdots \\ u_{N-2} \\ u_{N-1} \\ u_N \end{pmatrix} = \frac{1}{h}\begin{pmatrix} \sigma(u_0 - g_0) \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

or compactly as

$$\begin{aligned} \mathbf{u}_t + aD\mathbf{u} &= \mathbb{S} \qquad t \geq 0 \\ \mathbf{u} &= \mathbf{f} \qquad t = 0 \end{aligned} \tag{3.9}$$

where $\mathbf{f} = [f_1, f_2, \ldots, f_N]^T$ and $\mathbb{S} = \sigma(u_0 - g_0)\mathbf{e}_0$ is a SAT penalty term. The projection vector $\mathbf{e}_0 = [1, 0, \ldots, 0]^T$ ensures that penalty term is confined on the left boundary node. The value of u_0 is not enforced, it is calculated as part of the solution process just like other grid values. In general, we have $u_0 \neq g(t)$. The SAT boundary treatment forces u_0 towards g_0.

The difference operator (3.8) admits the multiplicative decomposition

$$D = P^{-1}Q \tag{3.10}$$

where

$$P = h\text{diag}(\frac{1}{2}, 1, \ldots, 1, \frac{1}{2}); \quad Q = \frac{1}{2}\begin{bmatrix} -1 & 1 & & & \\ -1 & 0 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 0 & 1 \\ & & & -1 & 1 \end{bmatrix}.$$

Inspecting the matrices in (3.10) reveals that P is symmetric, positive definite and that Q is almost skew-symmetric

$$P = P^T > 0, \quad Q + Q^T = E_N - E_0 \tag{3.11}$$

where $E_0 = \text{diag}(1, 0, \ldots, 0)$ and $E_N = \text{diag}(0, \ldots, 0, 1)$. Next, we show that the difference operator D in (3.10) satisfies the SBP rule (2.6). Let $\tilde{H} = P$ in (2.2), which is allowed since P is symmetric and positive. For any two nonzero real-valued vectors $\mathbf{u}$ and $\mathbf{v}$, we have

$$\begin{aligned} (\mathbf{u}, D\mathbf{v})_P &= \mathbf{u}^T P P^{-1} Q\mathbf{v} = \mathbf{u}^T\big((E_N - E_0) - Q^T\big)\mathbf{v} = \mathbf{u}^T(E_N - E_0)\mathbf{v} - \mathbf{u}^T Q^T P^{-1} P\mathbf{v} \\ &= (uv)\big|_0^N - (D\mathbf{u}, \mathbf{v})_P \end{aligned}$$

which satisfies (2.6). Here, we have used the matrix properties (3.11).

Summarising the above, we constructed a difference operator D which is second-order accurate on the interior stencil and first order accurate at the boundary points. Subsequently, written in the form (3.10), we have shown that the operator D satisfies the SBP rule. Loosely speaking, an operator approximating the continuous derivatives is called SBP preserving if it can be written in the form (3.10) and satisfies the SBP rule. This allows SBP operators to be formulated outside the finite difference framework as well , see for example; the discontinuous Galerkin method [29, 30], spectral element method[31] and finite volume method [32, 33].

Next, we provide a formal definition of the first derivative SBP operators. Let $u \in \mathcal{F}$ and $\mathbf{u} = (u_0, u_1, \ldots, u_N)^T$ where $u_i = u(x_i, t)$

Definition 5. A difference operator $D \in M_{N\times N}(\mathbb{R})$ is a first derivative SBP operator if

$$D\mathbf{u} \approx \frac{\partial u}{\partial x} \quad \text{and} \quad D = P^{-1}Q$$

where $P = P^T > 0$ and Q is almost skew-symmetric satisfying $Q+Q^T = B = \text{diag}(-1, 0, \ldots, 0, 1)$.

In addition, P defines the discrete L^2- norm $||\mathbf{u}||_P^2 = \mathbf{u}^T P\mathbf{u}$ and mimics the continuous integral

$$\mathbf{1}^T P\mathbf{u} \approx \int_\Omega u dx \tag{3.12}$$

where $\mathbf{1} = [1, \ldots, 1]^T$.

3.2.1 The structure of P and Q

In (3.10), we presented the SBP matrices P and Q explicitly for the 2nd-order accurate operator. In this case, the matrix P was diagonal, however in general, it is allowed to have block structures near the boundary points. The operator Q has block structures near the boundary points and is banded on the internal points. The bandwidth of the internal stencil is equal to the order of accuracy of the central finite difference operator plus one. In (3.10), the band of Q comprises of the repeated coefficients of central finite difference approximation with exception at the boundary points where the coefficients correspond to those of the one-sided operator.

The general forms of P and Q are

$$P = h\begin{bmatrix} P_0 & & \\ & \begin{matrix} 1 & & \\ & \ddots & \\ & & 1 \end{matrix} & \\ & & P_0^* \end{bmatrix}; \quad Q = \begin{bmatrix} Q_0 & & \\ & \begin{matrix} \alpha_k & & \\ & \ddots & \\ & & \alpha_k \end{matrix} & \\ & & -Q_0^* \end{bmatrix}. \tag{3.13}$$

where P_0^* and Q_0^* are the mirror images (rotated πrad anti-clockwise) of P_0 and Q_0 respectively. Here, α_k denotes the coefficients of the central finite difference approximation defined in (3.5). For a $2s$ ($s > 0$) accurate finite difference SBP operator in the interior stencil, the bandwidth of the internal band is $2s + 1$. The exact form of the block matrix Q_0 is determined by a one-sided difference stencil used to close the boundary. In (3.10), $Q_0 = [-1/2 \quad 1/2; -1/2 \quad 0]$ and $Q_0^* = [0 \quad 1/2; -1/2 \quad 1/2]$. The matrix P_0 can be diagonal or a block matrix. However, both structures informs the accuracy of the approximation near the boundary as well as the global accuracy. We state the following theorems and the proofs can be found in [11].

Theorem 3.1. *For the interior order of accuracy $2s$ ($s > 0$), there exist an SBP operator $D = P^{-1}Q$ such that*

i. P_0 is a block matrix

ii. the order of accuracy is $2s - 1$ near the boundaries.

Theorem 3.2. *For the interior order of accuracy $2s$ ($s \in [1, 4]$), there exist an SBP operator $D = P^{-1}Q$ such that*

i. P_0 is a diagonal matrix

ii. the order of accuracy is s near the boundaries.

The global order of accuracy of the numerical scheme depends on the norm matrix P. If P_0 in (3.13) is a block matrix then the global accuracy is $2s$. If P_0 is diagonal then the global accuracy is $s + 1$. In this study, we will only consider P_0 with diagonal structure for stability reasons. The notation $(2s, s)$ will be used to denote the accuracy of the SBP operator where $2s$ is the accuracy of the interior stencil and s denotes accuracy near the boundary points. SBP operators (up to 8^{th} order) can be found in [19, 34].

3.2.2 Second derivative SBP operators

Second derivatives in space appears in parabolic PDEs such as the incompressible Navier-Stokes and advection-diffusion equations. The finite difference SBP operators approximating these derivatives can be constructed by applying the first derivative operator twice or directly from a compact difference stencil. The general form of these operators is $D_2 = P^{-1}R$ where P is a diagonal-norm matrix. If both D and D_2 are used then same P must be used. The matrix R can be decomposed as $R = -M + S$ where $M = M^T \geq 0$ is a contribution of the central difference stencil and S is a sparse matrix with the coefficients of one-sided stencil on the first and last row.

Discretizing the second derivative by applying the first SBP operator D twice yields

$$D_2 = DD = P^{-1}QP^{-1}Q = P^{-1}(B - Q^T)P^{-1}Q = P^{-1}(-M + S) \tag{3.14}$$

where $M = Q^T P^{-1} Q$ and $S = BD$. The resulting stencil when using this approach is wider compared to the stencil of the first derivative operator. To illustrate this, we consider a second-order accurate operator and recall that the internal elements of D comes from (3.4). Let $u \in C^\infty$, applying (3.4) twice to approximate u_{xx} at grid point x_i yields

$$\frac{u_{i+2} - 2u_i + u_{i-2}}{(2h)^2} = \frac{\partial^2 u}{\partial x^2}\Big|_{x=x_i} + \mathcal{O}(2h)^2. \tag{3.15}$$

In a 1-D mesh, the computational stencil becomes

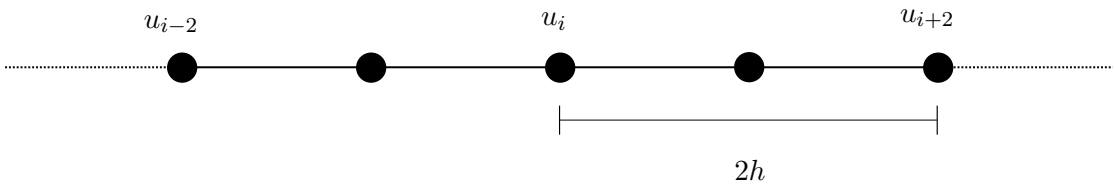

Figure 3.1: The wide computational stencil.

The width of the internal band of D_2 is $4s+1$ which is wider than the width of D. To get a more compact operator, we can construct D_2 directly from the difference stencils. By adding (3.1) and (3.2), a central finite difference approximating the second derivatives is obtained

$$\frac{u_{i+1}-2u_i+u_{i-1}}{h^2} = \frac{\partial^2 u}{\partial x^2}\Big|_{x=x_i} + \mathcal{O}(h^2).$$

and the resulting stencil is narrow with the internal bandwidth of D_2 equals to $2s+1$.

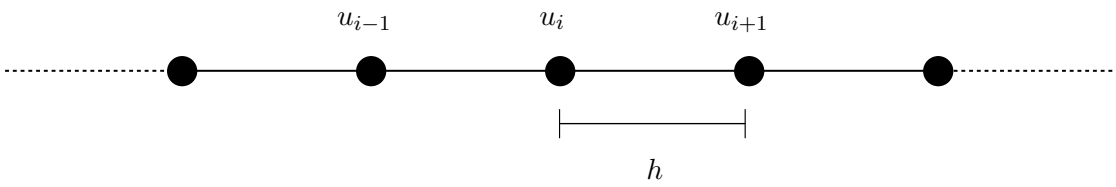

Figure 3.2: The compact computational stencil.

When defining the second derivative SBP operator D_2 according to (3.14) with a diagonal-norm P matrix, the global order of accuracy of the numerical solution to a well-posed IBVP is $s+1$ where $s=1,..,4$. This order can be increase to $s+2$ by constructing D_2 according to (3.15) and carefully choosing S [28]. For the case $s=1$, the matrix M and S becomes

$$M = \begin{bmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & -1 & 2 & -1 & \\ & & \ddots & \ddots & \ddots \end{bmatrix} \qquad S = \frac{1}{h}\begin{bmatrix} \frac{3}{2} & -2 & \frac{1}{2} & & \\ & 0 & & & \\ & & 0 & & \\ & & & \ddots & \end{bmatrix}.$$

Here, the elements of S comes from the 2nd order Backward finite difference operator. The SBP framework can also be extended to approximate third and higher derivatives [35].

3.2.3 Two-dimensional SBP operators

The SBP operators can be extended to multiple dimensions using the Kronecker products (2.5). To be able to differentiate between the operators approximating $x-$ and $y-$derivatives, we denote them with subscripts x and y respectively

$$D_x\mathbf{u} = P_x^{-1}Q_x\mathbf{u} \approx \frac{\partial u}{\partial x}, \quad D_y\mathbf{u} = P_y^{-1}Q_y\mathbf{u} \approx \frac{\partial u}{\partial y}.$$

Let N and M be the number of nodes along the horizontal and vertical axis respectively. In 2-D, the SBP operators become

1. $\mathcal{P} = P_x \otimes P_y$
2. $\mathcal{D}_x = D_x \otimes I_M, \quad \mathcal{D}_y = I_N \otimes D_y$
3. $\mathcal{Q}_x = Q_x \otimes I_M, \quad \mathcal{Q}_y = I_N \otimes Q_y$
4. $\mathcal{D}_{xx} = D_{xx} \otimes I_M, \quad \mathcal{D}_{yy} = I_N \otimes D_{yy}$
5. $\mathcal{B}_x = (E_N - E_0) \otimes P_y, \quad \mathcal{B}_y = P_x \otimes (E_M - E_0),$

where $I_{N,M}$ is a unit matrix, $E_{N,M} = \text{diag}(0, \dots, 0, 1)$ and $E_0 = \text{diag}(1, 0, \dots, 0)$ are matrices of size $N \times N$ and $M \times M$ respectively.

3.3 The discrete energy method

To demonstrate that the approximation (3.9) is strongly stable, we apply the energy method in the discrete setting. Let's consider the discrete problem (3.9)

$$\begin{aligned} \mathbf{u}_t + aD\mathbf{u} &= P^{-1}E_0(\mathbf{u} - g) \qquad t \geq 0. \\ \mathbf{u} &= \mathbf{f} \qquad\qquad\quad t = 0 \end{aligned} \tag{3.16}$$

where $P^{-1}E_0$ is a discrete analogue of the lifting operator $\mathcal{L}(\cdot)$ in (2.5) and σ is a penalty coefficient yet to be determined. We imitate the continuous energy method discretely by multiplying (3.16) on the left with $\mathbf{u}^T P$. In (3.12), we defined the diagonal-norm matrix P as a discrete integration operator. Equation (3.16) becomes

$$\mathbf{u}^T P \mathbf{u}_t + a\mathbf{u}^T Q\mathbf{u} = \sigma u_0(u_0 - g) \tag{3.17}$$

By taking the transpose of (3.17), adding to itself and using SBP property $Q + Q^T = B$, we obtain

$$\frac{d}{dt}||\mathbf{u}||_p^2 = -au_N^2 + (a + 2\sigma)u_0^2 - 2\sigma u_0 g. \tag{3.18}$$

The relation (3.18) is identical to the continuous relation (2.16), therefore, we state a similar proposition.

Proposition 3.3. *The approximation* (3.16) *is strongly stable if* $\sigma = -a$.

Proof. Adding and subtracting $ag(t, 0)$ to the LHS of (3.18) with $\sigma = -a$ yields

$$\frac{d}{dt}||\mathbf{u}||_p^2 = ag^2 - au_N^2 - (au_0 - g)^2 \leq ag^2.$$

By integrating over finite time $[0, t]$, we obtain the estimate

$$||\mathbf{u}||_P^2 \leq ||\mathbf{f}||_P^2 + \int_0^t ag(\tau, 0)d\tau,$$

therefore, the approximation (3.9) is strongly stable. □

Chapter 4

Initial boundary value problems

In this chapter, we apply the theory presented in Chapter 2 and 3. The linear advection-diffusion equations in one-dimension with constant coefficients and two-dimensions with variable coefficients are considered for this purpose.

4.1 Linear advection-diffusion in one dimension

The 1-D linear advection-diffusion equation reads

$$\begin{aligned} u_t + au_x &= \varepsilon u_{xx}, \quad x \in [0,1], \quad t \geq 0 \\ L_0 u(t,0) &= g_0(t) \\ L_1 u(t,1) &= g_1(t) \\ u(0,x) &= f(x) \end{aligned} \tag{4.1}$$

where a, $\varepsilon > 0$ and $\varepsilon \ll a$. Both a, ε are constants. The boundary and initial data are given by continuous functions g_0, g_1 and f respectively. In (4.1), L_0 and L_1 are linear boundary operators and their exact form will be determined using the energy method such that the continuous problem is well-posed.

4.1.1 The continuous problem

The energy method applied to (4.1) yields

$$\frac{d}{dt}||u||^2 + 2\varepsilon||u_x||^2 = -(au - 2\varepsilon u u_x)\Big|_{x=0}^{x=1}. \tag{4.2}$$

It is not apparent from (4.2) whether $||u||^2$ is bounded. Therefore, the boundary operators L_0 and L_1 must be specified such that the estimate is bounded by data. Since the boundary term in (4.2) comprises of the solution and its derivatives, it is natural to seek the form of boundary conditions that includes both. Robin boundary condition imposed at $x = 0$ and Neumann boundary condition imposed at $x = 1$ leads to well-posedness [18, 36]. In this study we however consider Dirichlet boundary conditions at both boundaries i.e. $L_0 = L_1 = 1$ [37]. Next, we show that (4.1) is well-posed under these conditions.

Proposition 4.1. *The IBVP (4.1) with $g_0 = g_1 = 0$ is well-posed.*

Proof. Substituting boundary conditions into (4.1) yields

$$\frac{d}{dt}||u||^2 + 2\varepsilon||u_x||^2 = 0.$$

Therefore, by integrating in time, we obtain

$$||u||^2 \leq ||f||^2 \tag{4.3}$$

which is bounded. Uniqueness follows directly from (4.3), hence (4.1) is well-posed. □

4.1.2 The semi-discrete problem

The semi-discrete formulation of (4.1) is considered. For brevity (without limitation), the SBP operator approximating the second derivative term is constructed by applying the first derivative operator twice. We divide the interval $[0, 1]$ into N non-overlapping equidistant subintervals with end points $x_i = ih$, $i = 0, 1, \ldots N$ where h is the length of each subinterval. The numerical approximation of u at x_i is denoted by u_i and $\mathbf{u}$ is a vector containing the numerical solution at each grid point.

The semi-discrete problem together with weakly imposed boundary conditions becomes

$$\begin{aligned} \mathbf{u}_t + aP^{-1}Q\mathbf{u} &= \varepsilon P^{-1}Q\mathbf{u}_x + \sigma_0 P^{-1}(u_0 - g)\mathbf{e}_0 + \sigma_1 P^{-1}(u_N - g)\mathbf{e}_1 \\ \mathbf{u}(0) &= \mathbf{f} \end{aligned} \tag{4.4}$$

where σ_0, σ_1 are penalty coefficients to be determined in the interest of stability. The notation $\mathbf{u}_x$ is short-hand notation for $D\mathbf{u}$ and $\mathbf{f}$ is the vector containing initial data at each grid point. The boundary projection vectors $\mathbf{e}_0 = (1, 0, \ldots, 0)^T$ and $\mathbf{e}_1 = (0, \ldots, 0, 1)^T$ restricts the SAT terms to the left and right boundary node respectively.

To derive stability conditions, we employ the discrete energy method. Multiplying (4.4) with $\mathbf{u}^T P$ yields

$$\mathbf{u}^T P\mathbf{u}_t + a\mathbf{u}^T Q\mathbf{u} = \varepsilon Q\mathbf{u}_x + \sigma_o u_0(u_0 - g_0) + \sigma_1 u_N(u_N - g_N). \tag{4.5}$$

Taking the transpose of (4.5), adding to itself and setting $g_0 = g_1 = 0$ yields

$$\frac{d}{dt}||\mathbf{u}||_P^2 = (a + 2\sigma_0)u_0^2 - (a - 2\sigma_1)u_N^2 + \varepsilon \underbrace{(\mathbf{u}^T Q\mathbf{u}_x + \mathbf{u}_x^T Q^T\mathbf{u})}_{①}. \tag{4.6}$$

By using SBP property $Q + Q^T = B$ for the diffusion terms denoted by ①, we obtain

$$\begin{aligned} \mathbf{u}^T Q\mathbf{u}_x + \mathbf{u}_x^T Q^T\mathbf{u} &= \mathbf{u}^T(B - Q^T)\mathbf{u}_x + \mathbf{u}_x^T(B - Q)\mathbf{u} \\ &= -\mathbf{u}^T Q^T P^{-1} P\mathbf{u}_x - \mathbf{u}_x^T P P^{-1} Q\mathbf{u} + 2\mathbf{u}^T B\mathbf{u}_x \\ &= -2\mathbf{u}_x^T P\mathbf{u}_x + 2u_N(\mathbf{u}_x)_N - 2u_0(\mathbf{u}_x)_0. \end{aligned} \tag{4.7}$$

Above, we inserted unit matrices PP^{-1} and $P^{-1}P$ to simplify the indefinite terms. By substituting (4.7) into (4.6), we obtain

$$\frac{d}{dt}||\mathbf{u}||_P^2 + 2\varepsilon||\mathbf{u}_x||_P^2 = (a+2\sigma_0)u_0^2 - 2\varepsilon u_0(\mathbf{u}_x)_0 - (a-2\sigma_1)u_N^2 + 2\varepsilon u_N(\mathbf{u}_x)_N. \tag{4.8}$$

The signs of $(\mathbf{u}_x)_{0,N}$ are unknown and the boundedness of (4.8) is not clear.

Next, we use the *borrowing* technique [38]. Notice that $||\mathbf{u}_x||_P^2$ can be written as

$$||\mathbf{u}_x||_P^2 = \mathbf{u}_x^T P \mathbf{u}_x = (\mathbf{u}_x)_0^2 hP_0 + h\sum_{i=1}^{N-1}(\mathbf{u}_x)_i^2 P_i + (\mathbf{u}_x)_N^2 hP_N.$$

Substituting the results above into (4.8) yields

$$\begin{aligned}\frac{d}{dt}||\mathbf{u}||_P^2 + 2\varepsilon||\mathbf{u}_x||_{\bar{P}}^2 &= (a+2\sigma_0)u_0^2 - 2\varepsilon u_0(\mathbf{u}_x)_0 - 2\varepsilon(\mathbf{u}_x)_0^2 hP_0 \\ &\quad - (a-2\sigma_1)u_N^2 + 2\varepsilon u_N(\mathbf{u}_x)_N - 2\varepsilon(\mathbf{u}_x)_N^2 hP_N \\ &= \begin{bmatrix} u_0 \\ (\mathbf{u}_x)_0 \end{bmatrix}^T \begin{bmatrix} 2\sigma_0 + a & -\varepsilon \\ -\varepsilon & -2h\varepsilon P_0 \end{bmatrix} \begin{bmatrix} u_0 \\ (\mathbf{u}_x)_0 \end{bmatrix} \\ &\quad + \begin{bmatrix} u_N \\ (\mathbf{u}_x)_N \end{bmatrix}^T \begin{bmatrix} 2\sigma_1 - a & \varepsilon \\ \varepsilon & -2h\varepsilon P_N \end{bmatrix} \begin{bmatrix} u_N \\ (\mathbf{u}_x)_N \end{bmatrix}\end{aligned} \tag{4.9}$$

where

$$\bar{P} = P - \mathrm{diag}(P_0, 0, \ldots, 0, P_N).$$

The energy rate (4.9) is bounded if the boundary matrices on the RHS are negative semi-definite.

Proposition 4.2. *The approximation* (4.1) *is stable if*

$$\sigma_0 \leq -\frac{a}{2} - \frac{\varepsilon}{4hP_0}, \qquad \sigma_1 \leq \frac{a}{2} - \frac{\varepsilon}{4hP_N}.$$

Proof. Consider the left boundary matrix

$$A = \begin{bmatrix} 2\sigma_0 + a & -\varepsilon \\ -\varepsilon & -2h\varepsilon P_0 \end{bmatrix}.$$

We compute the eigenvalues of A and determine σ_0 such that they are non-positive. The eigenvalues of A denoted by $\lambda = \lambda_1, \lambda_2$ are obtained by solving the characteristic polynomial

$$\det(A - \lambda I_2) = \lambda^2 - 2(\alpha - \varepsilon hP_0)\lambda - (\varepsilon^2 + 4\alpha\varepsilon P_0) = 0 \tag{4.10}$$

where I_2 is a 2×2 unit matrix and $2\alpha = 2\sigma_0 + a$. The roots of (4.10) are

$$\lambda_{1,2} = (\alpha - \varepsilon hP_0) \mp \sqrt{(\alpha - \varepsilon hP_0)^2 + (\varepsilon^2 + 4\alpha\varepsilon hP_0)}$$

and they are non-positive if

$$\alpha - \varepsilon h P_0 \leq 0 \quad \text{and} \quad \varepsilon^2 + 4\alpha\varepsilon h P_0 \leq 0.$$

Therefore, by substituting 2α into the inequalities above, we obtain

$$\sigma_0 \leq -\frac{a}{2} + \varepsilon h P_0 \quad \text{and} \quad \sigma_0 \leq -\frac{a}{2} - \frac{\varepsilon}{4hP_0} \quad \Rightarrow \quad \sigma_0 \leq -\frac{a}{2} - \frac{\varepsilon}{4hP_0}.$$

Similarly, the right boundary matrix is negative semi-definite if

$$\sigma_1 \leq \frac{a}{2} - \frac{\varepsilon}{4hP_N}.$$

The energy rate (4.9) is therefore bounded and temporal integration yields the estimate

$$||\mathbf{u}||_P^2 \leq ||\mathbf{f}||_P^2$$

which mimics the continuous counterpart (4.3). Hence, the approximation (4.4) is stable. □

4.1.3 Numerical test

In this section, the computational results illustrating the accuracy of the approximation (4.4) are presented. We compute the spatial derivatives using $(2s, s)$ accurate finite difference SBP operators where $2s$ is the accuracy of the internal stencil and s denotes the accuracy near the boundaries. The diffusion term is discretized using the wide and narrow stencil operators. The convergence rate $q^{(2s,s)}$ is computed as

$$q^{(2s,s)} = \log_{10}\left(\frac{||\mathbf{u} - \mathbf{v}^{h_1}||_P}{||\mathbf{u} - \mathbf{v}^{h_2}||_P}\right) \Big/ \log_{10}\left(\frac{h_1}{h_2}\right)^{1/d}, \quad s = 1, \ldots, 4 \tag{4.11}$$

where $\mathbf{u}$, $\mathbf{v}$ denotes the numerical and analytical solution respectively, $||\mathbf{u} - \mathbf{v}^{h_1}||_P$ is discretely equivalent to the L_2 norm of error and d is the space dimension. The coarse and fine mesh sizes are denoted by h_1 and h_2 respectively. The accuracy of (4.4) with $g_0 = 1$, $g_1 = 0$ is verified by considering the analytical steady solution

$$v(x) = 1 - \frac{e^{ax/\varepsilon} - 1}{e^{a/\varepsilon} - 1}.$$

Table 4.1: The $\log L_2$ norm of errors and convergence rates of the SBP operators (up to 8th order) with *wide stencil* second derivative operators are compared.

N	$\log L_2$	$q^{(2,1)}$	$\log L_2$	$q^{(4,2)}$	$\log L_2$	$q^{(6,3)}$	$\log L_2$	$q^{(8,4)}$
20	-0.831	-	-1.533	-	-1.692	-	-2.148	-
40	-1.736	3.007	-2.543	3.355	-2.82	3.748	-3.552	4.666
80	-2.407	2.228	-3.454	3.026	-3.929	3.684	-4.944	4.622
160	-3.028	2.062	-4.347	2.969	-5.066	3.778	-6.369	4.736
320	-3.636	2.021	-5.243	2.974	-6.232	3.873	-7.83	4.851
640	-4.241	2.008	-6.141	2.985	-7.416	3.933	-9.311	4.922

Table 4.2: The $\log L_2$ norm of errors and convergence rates of the SBP operators (up to 8th order) with *narrow stencil* second derivative operators are compared.

N	$\log L_2$	$q^{(2,1)}$	$\log L_2$	$q^{(4,2)}$	$\log L_2$	$q^{(6,3)}$	$\log L_2$	$q^{(8,4)}$
20	-2.026	-	-2.82	-	-2.894	-	-2.366	-
40	-2.725	2.325	-3.91	3.622	-4.2	4.338	-3.8	4.763
80	-3.445	2.39	-5.059	3.817	-5.599	4.646	-5.418	5.375
160	-4.145	2.327	-6.237	3.911	-7.048	4.812	-7.129	5.684
320	-4.813	2.217	-7.428	3.957	-8.523	4.903	-8.887	5.841
640	-5.452	2.125	-8.625	3.978	-10.017	4.961	-10.551	5.525

The numerical steady solution is computed by disregarding the temporal term and reducing (4.4) to a matrix problem. The coefficient matrix is then inverted to obtain the solution. The advection speed and diffusion constant are taken as $a = 1$ and $\varepsilon = 0.1$.

The convergence rates and logarithmic-scaled L_2 norms are presented in Table. 4.1 for wide and Table. 4.2 for narrow stencil second derivative operators. From the results above, the global orders of accuracy for the calculations using the wide stencil operator is $s+1$ while for the narrow operator is $s+2$. The efficiency of high order methods can be seen from the L_2 norms. For example, in Table. 4.1, the log-scaled error norm is below -4 with 80 nodes when using 8^{th} order accurate approximation while 2^{nd} order approximation requires 640 nodes to achieve the same accuracy.

4.2 Linear advection-diffusion with variable coefficients

The linear advection-diffusion equation with variable coefficients is considered on $\Omega \subset \mathbb{R}^n$

$$\begin{aligned} u_t + \mathbf{a}(\mathbf{x})^T \nabla u &= \varepsilon \nabla^T \nabla u \qquad && \mathbf{x} \in \Omega, \quad t \geq 0 \\ \frac{1}{2}\big(|\mathbf{a}^T\mathbf{n}| - \mathbf{a}^T\mathbf{n}\big)u + \varepsilon \mathbf{n}^T \nabla u &= g(t) && \mathbf{x} \in \partial\Omega, \quad t \geq 0 \\ u(\mathbf{x}, 0) &= f(\mathbf{x}) && \mathbf{x} \in \Omega, \quad t \geq 0 \end{aligned} \tag{4.12}$$

where u is an unknown scalar field, $\mathbf{a}$ is a vector of variable coefficients (advection speed) and $\mathbf{n}$ denotes the unit outward pointing normal vector on $\partial\Omega$. We assume that $\mathbf{a}$ is smooth such that the maximum norm is bounded. The normal component of $\mathbf{a}$ on the boundary and normal derivative operator are denoted by $\mathbf{a}^T\mathbf{n}$ and $\mathbf{n}^T\nabla$ respectively. Notice that if $\mathbf{a}^T\mathbf{n} > 0$, we use well-posed Neumann boundary conditions while $\mathbf{a}^T\mathbf{n} \leq 0$ leads to Robin boundary conditions. The boundary and initial data are given by $g, f \in \mathcal{F}$ respectively and $\varepsilon > 0$ is a constant. Next, we show that the (4.12) is well-posed.

Proposition 4.3. *The problem (4.12) with non-homogeneous boundary conditions is strongly well-posed.*

Proof. The energy method and Gauss theorem applied to (4.12) yields

$$\frac{d}{dt}||u||^2 + 2\varepsilon||\nabla u||^2 = \int_\Omega \nabla^T \mathbf{a} u^2 dV + \oint_{\partial\Omega} (2\varepsilon u \mathbf{n}^T \nabla u - \mathbf{a}^T \mathbf{n} u^2) ds. \tag{4.13}$$

The indefinite integral term on the RHS of (4.13) adds growth to the energy rate, however, it is a bounded growth since

$$\int_\Omega \nabla^T \mathbf{a} u^2 dV \leq \int_\Omega |\nabla^T \mathbf{a}||u|^2 dV \leq |\nabla^T \mathbf{a}|_{\max}||u||^2 \tag{4.14}$$

where $|\cdot|_{\max}$ is a maximum norm. Next, we substitute boundary conditions (4.12) and (4.14) into (4.13) to obtain

$$\frac{d}{dt}||u||^2 + 2\varepsilon||\nabla u||^2 \leq |\nabla^T \mathbf{a}|_{\max}||u||^2 + \oint_{\partial\Omega} \big(2ug - |\mathbf{a}^T \mathbf{n}|u^2\big) ds. \tag{4.15}$$

Since the choice of boundary data $g(t)$ is arbitrary, we choose it such that $g = |\mathbf{a}^T \mathbf{n}|\hat{g}$ holds for some $\hat{g} \in \mathcal{F}$. By adding and subtracting $|\mathbf{a}^T \mathbf{n}|\hat{g}^2$ to the RHS of (4.15), the energy rate become

$$\begin{aligned} \frac{d}{dt}||u||^2 + 2\varepsilon||\nabla u||^2 &\leq |\nabla^T \mathbf{a}|_{\max}||u||^2 + \oint_{\partial\Omega} \big(|\mathbf{a}^T \mathbf{n}|\hat{g}^2 - |\mathbf{a}^T \mathbf{n}|(u-\hat{g})^2\big) ds \\ &\leq |\nabla^T \mathbf{a}|_{\max}||u||^2 + \oint_{\partial\Omega} \big(|\mathbf{a}^T \mathbf{n}|\hat{g}^2\big) ds. \end{aligned} \tag{4.16}$$

which is bounded. Multiplying (4.16) with the integration factor $e^{-|\nabla^T \mathbf{a}|_{\max} t}$ and integrating over finite time leads to

$$||u||^2 \leq e^{|\nabla^T \mathbf{a}|_{\max} t}\Big(||f||^2 + \int_0^t \big(e^{|\nabla^T \mathbf{a}|_{\max} \xi} \oint_{\partial\Omega} |\mathbf{a}^T \mathbf{n}|\hat{g}^2(\xi) ds\big) d\xi\Big). \tag{4.17}$$

Time integration shows that the energy estimate is bounded, therefore, (4.12) is strongly well-posed. □

4.2.1 Weakly implemented boundary conditions

To prepare for the numerical implementation, we impose the boundary conditions in (4.12) weakly and determine the penalty coefficient such that the continuous problem is well-posed. Equation (4.12) with weak boundary conditions become

$$u_t + \mathbf{a}^T \nabla u = \varepsilon \nabla^T \nabla u + \mathcal{L}\Big(\Sigma\big(\frac{1}{2}(|\mathbf{a}^T \mathbf{n}| - \mathbf{a}^T \mathbf{n})u + \varepsilon \mathbf{n}^T \nabla u - g(t)\big)\Big) \tag{4.18}$$

where $\mathcal{L}(\cdot)$ is a lifting operator (2.5) and Σ denotes the penalty coefficient yet to be determined. The energy method applied to (4.18) yields

$$\frac{d}{dt}||u||^2 + 2\varepsilon||\nabla u||^2 \leq |\nabla^T \mathbf{a}|_{\max}||u||^2 + BT \tag{4.19}$$

where

$$BT = \oint_{\partial\Omega} \Big(2\varepsilon u \mathbf{n}^T \nabla u - \mathbf{a}^T \mathbf{n} u^2 + \Sigma\big(|\mathbf{a}^T \mathbf{n}|u^2 - \mathbf{a}^T \mathbf{n} u^2 + 2\varepsilon u \mathbf{n}^T \nabla u - 2ug\big)\Big) ds$$

are boundary terms. We used the integral bound (4.14) in (4.19).

Proposition 4.4. *The energy rate* (4.19) *is bounded if* $\Sigma = -1$.

Proof. The boundary term BT in (4.19) with $\Sigma = -1$ becomes

$$BT = \oint_{\partial\Omega} \big(2ug - |\mathbf{a}^T\mathbf{n}|u^2\big)ds$$

which is identical to the boundary term on the RHS of (4.15). Here too, we scale data as $g = |\mathbf{a}^T\mathbf{n}|\hat{g}$. Therefore, by adding and subtracting $g = |\mathbf{a}^T\mathbf{n}|\hat{g}^2$, the BT becomes

$$BT = \oint_{\partial\Omega} \big(|\mathbf{a}^T\mathbf{n}|\hat{g}^2 - |\mathbf{a}^T\mathbf{n}|(u-\hat{g})^2\big)ds \leq \oint_{\partial\Omega} \big(|\mathbf{a}^T\mathbf{n}|\hat{g}^2\big)ds.$$

The weakly imposed boundary conditions yields an estimate, hence (4.18) is strongly well-posed. □

It was pointed out in [39] that the convective term with variable coefficients leads to stability concerns since the discrete energy estimate cannot be obtained explicitly. To avoid this, we rewrite it in a skew-symmetric form

$$\mathbf{a}^T\nabla u = \frac{1}{2}\big(\mathbf{a}^T\nabla u + \nabla^T(\mathbf{a}u) - (\nabla^T\mathbf{a})u\big). \tag{4.20}$$

By substituting (4.20) into (4.18), we obtain

$$u_t + \frac{1}{2}\big(\mathbf{a}^T\nabla u + \nabla^T(\mathbf{a}u) - (\nabla^T\mathbf{a})u\big) = \varepsilon\nabla^T\nabla u + \mathcal{L}\Big(\Sigma\big(\frac{1}{2}(|\mathbf{a}^T\mathbf{n}| - \mathbf{a}^T\mathbf{n})u + \varepsilon\mathbf{n}^T\nabla u - g(t)\big)\Big). \tag{4.21}$$

4.2.2 The semi-discrete problem

The semi-discrete formulation of (4.21) is considered. For simplicity of presentation and brevity, we consider a regular quadrilateral domain $\Omega \subset \mathbb{R}^2$ on the Cartesian plane. The outward pointing unit normal vectors on the east, west, north and south (E, W, N, S) boundary are denoted by $\mathbf{n}_E$, $\mathbf{n}_W$, $\mathbf{n}_N$, $\mathbf{n}_S$ respectively.

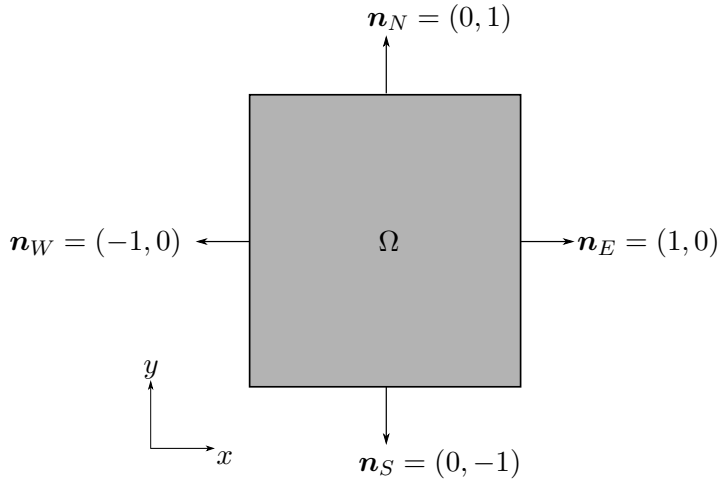

Figure 4.1: 2-D domain representation with outward pointing normal vectors

To discretize (4.21) in space, we consider $N \times M$ equidistant computational grids $\Omega_{ij} = \Omega(x_i, y_j)$ where (x_i, y_j) are the Cartesian grid coordinates given by

$$x_i = ih_x, \quad y_j = jh_y$$

for $i = 0, 1, \ldots, N$ and $j = 0, 1, \ldots, M$. Here, h_x, h_y denotes the length and width of each grid. In total, there are $N+1$ and $M+1$ nodes along the horizontal and vertical axis, however we contain the notation by referring to them as N and M. Consequently, all the matrices associated with x and y terms will be taken from the sets $M_{N\times N}$ and $M_{M\times M}$ respectively.

Let U be an approximation solution of (4.21) and A, B be diagonal matrices containing the variable coefficients. To keep the notation cleaner and concise, we introduce the following notation

$$\mathcal{D}_{\mathbf{x}} = [\mathcal{D}_x \quad \mathcal{D}_y]^T, \quad \mathcal{A} = [A \quad B]^T, \quad \mathcal{A}_{\mathbf{x}} = \mathcal{D}_{\mathbf{x}}^T \mathcal{A} = \mathcal{D}_x A + \mathcal{D}_y B, \quad \mathcal{D}_{\mathbf{xx}} = \mathcal{D}_{\mathbf{x}}^T \mathcal{D}_{\mathbf{x}} = \mathcal{D}_{xx} + \mathcal{D}_{yy}.$$

The semi-discrete SBP-SAT formulation of (4.21) becomes

$$\begin{aligned} U_t + \frac{1}{2}\big(\mathcal{A}^T \mathcal{D}_{\mathbf{x}} U + \mathcal{D}_{\mathbf{x}}^T \mathcal{A} U - \mathcal{A}_{\mathbf{x}} U\big) - \varepsilon \mathcal{D}_{\mathbf{xx}} U = \mathbb{SAT}, \\ U(0) = F \end{aligned} \tag{4.22}$$

where $\mathbb{SAT}$ denotes the boundary penalty terms defined on each boundary as

$$\begin{aligned} \mathbb{SAT} = &(P_x^{-1} E_N \otimes I_M)\Big(\Sigma_E\big(\frac{1}{2}(|\mathcal{A}_{\mathbf{n}_E}| - \mathcal{A}_{\mathbf{n}_E})U + \varepsilon \mathcal{D}_{\mathbf{n}_E} - G_E\big)\Big) \\ &+ (P_x^{-1} E_0 \otimes I_M)\Big(\Sigma_W\big(\frac{1}{2}(|\mathcal{A}_{\mathbf{n}_W}| - \mathcal{A}_{\mathbf{n}_W})U + \varepsilon \mathcal{D}_{\mathbf{n}_W} - G_W\big)\Big) \\ &+ (I_N \otimes P_x^{-1} E_M)\Big(\Sigma_N\big(\frac{1}{2}(|\mathcal{A}_{\mathbf{n}_N}| - \mathcal{A}_{\mathbf{n}_N})U + \varepsilon \mathcal{D}_{\mathbf{n}_N} - G_N\big)\Big) \\ &+ (I_N \otimes P_x^{-1} E_0)\Big(\Sigma_S\big(\frac{1}{2}(|\mathcal{A}_{\mathbf{n}_S}| - \mathcal{A}_{\mathbf{n}_S})U + \varepsilon \mathcal{D}_{\mathbf{n}_S} - G_S\big)\Big). \end{aligned}$$

In (4.22), the matrices $\{P_x^{-1} E_N \otimes I_M; P_x^{-1} E_0 \otimes I_M; I_N \otimes P_x^{-1} E_M; I_N \otimes P_x^{-1} E_0\}$ are the discrete analogues of the lifting operator $\mathcal{L}(\cdot)$ in (2.5) acting on the E, W, N, S boundary respectively. The penalty coefficient matrices $\Sigma_{E,W,N,S}$ will be determined such that the approximation (4.22) is stable. The discrete analogues of the normal component of the advection speed and normal derivative along the ith boundary are denoted by $\mathcal{A}_{\mathbf{n}_i}$ and $\mathcal{D}_{\mathbf{n}_i}$ respectively. The vector F contains initial data given by grid grid function $f_{ij} = f(x_i, y_j)$. Lastly, $G_{E,W,N,S}$ are sparse vectors of the same size as U with boundary data injected on the places corresponding to a particular boundary nodes.

The discrete energy methods (multiplying the approximation with $U^T \mathcal{P}$, taking the transpose and adding to itself) applied to (4.22) yields

$$\frac{d}{dt}||U||_P^2 + 2\varepsilon||\mathcal{D}_{\mathbf{x}} U||_P^2 \leq |\mathcal{A}_{\mathbf{x}}|_{\max}||U||_{\mathcal{P}}^2 + BT. \tag{4.23}$$

In (4.23), we used that $\mathcal{P}$ and $\mathcal{A}_{\mathbf{x}}$ commutes with respect to matrix multiplication to bound $U^T\mathcal{P}\mathcal{A}_{\mathbf{x}}U$

$$U^T\mathcal{P}\mathcal{A}_{\mathbf{x}}U = U^T\mathcal{A}_{\mathbf{x}}\mathcal{P}U = (U\mathcal{A}_{\mathbf{x}}, U)_{\mathcal{P}} \leq |\mathcal{A}_{\mathbf{x}}|_{\max}||U||^2_{\mathcal{P}}.$$

The boundary terms denoted by $BT = BT_E + BT_W + BT_N + BT_S$ are

$$BT_E = U^T(E_N \otimes P_y)\Big(- \mathcal{A}_{\mathbf{n}_E}U + 2\varepsilon\mathcal{D}_{\mathbf{n}_E}U + \Sigma_E\big((|\mathcal{A}_{\mathbf{n}_E}| - \mathcal{A}_{\mathbf{n}_E})U + 2\varepsilon\mathcal{D}_{\mathbf{n}_E}U - 2G_E\big)\Big), \tag{4.24}$$

$$BT_W = U^T(E_0 \otimes P_y)\Big(- \mathcal{A}_{\mathbf{n}_W}U + 2\varepsilon\mathcal{D}_{\mathbf{n}_W}U + \Sigma_W\big((|\mathcal{A}_{\mathbf{n}_W}| - \mathcal{A}_{\mathbf{n}_W})U + 2\varepsilon\mathcal{D}_{\mathbf{n}_W}U - 2G_W\big)\Big),$$

$$BT_N = U^T(P_x \otimes E_M)\Big(- \mathcal{A}_{\mathbf{n}_N}U + 2\varepsilon\mathcal{D}_{\mathbf{n}_N}U + \Sigma_N\big((|\mathcal{A}_{\mathbf{n}_N}| - \mathcal{A}_{\mathbf{n}_N})U + 2\varepsilon\mathcal{D}_{\mathbf{n}_N}U - 2G_N\big)\Big),$$

$$BT_S = U^T(P_x \otimes E_0)\Big(- \mathcal{A}_{\mathbf{n}_S}U + 2\varepsilon\mathcal{D}_{\mathbf{n}_S}U + \Sigma_S\big((|\mathcal{A}_{\mathbf{n}_S}| - \mathcal{A}_{\mathbf{n}_S})U + 2\varepsilon\mathcal{D}_{\mathbf{n}_S}U - 2G_S\big)\Big).$$

The matrices $\{E_N \otimes P_y; E_0 \otimes P_y; P_x \otimes E_M; P_x \otimes E_0\}$ mimics the continuous line integrals along the E, W, N, S boundary respectively, for example

$$\mathbf{1}^T(E_N \otimes P_y)U \approx \oint_{\partial\Omega_{\mathbf{n}_E}} uds.$$

Remark 2. The boundary terms in (4.24) all have similar structure. Therefore, to keep the derivations cleaner, we focus on the east boundary.

Proposition 4.5. *The approximation* (4.22) *is strongly stable if* $\Sigma_{E,W,N,S} = -I_{N\times M}$.

Proof. Let's consider BT_E in (4.22) with boundary data scaled as $G_E = \mathcal{A}_{\mathbf{n}_E}\hat{G}_E$ for some smooth vector function $\hat{G}_E$. By adding and subtracting $\hat{G}_E^T(E_N \otimes P_y)\mathcal{A}_{\mathbf{n}_E}\hat{G}_E$ to the RHS of BT_E, we get

$$\begin{aligned} BT_E &= \hat{G}_E^T(E_N \otimes P_y)\mathcal{A}_{\mathbf{n}_E}\hat{G}_E - (U - \hat{G}_E)^T(E_N \otimes P_y)\mathcal{A}_{\mathbf{n}_E}(U - \hat{G}_E) \\ &\leq \hat{G}_E^T(E_N \otimes P_y)\mathcal{A}_{\mathbf{n}_E}\hat{G}_E. \end{aligned}$$

By treating other boundaries in a similar manner, the energy rate (4.23) becomes

$$\begin{aligned} \frac{d}{dt}||U||^2_P + 2\varepsilon||\mathcal{D}_{\mathbf{x}}U||^2_P \leq |\mathcal{A}_{\mathbf{x}}|_{\max}||U||^2_{\mathcal{P}} &+ \hat{G}_E^T(E_N \otimes P_y)\mathcal{A}_{\mathbf{n}_E}\hat{G}_E + \hat{G}_W^T(E_0 \otimes P_y)\mathcal{A}_{\mathbf{n}_W}\hat{G}_W \\ &+ \hat{G}_N^T(P_x \otimes E_M)\mathcal{A}_{\mathbf{n}_N}\hat{G}_N + \hat{G}_S^T(E_x \otimes E_0)\mathcal{A}_{\mathbf{n}_S}\hat{G}_S \end{aligned}$$

which discretely mimics (4.16). Therefore, time integration yields discrete energy estimate

$$\begin{aligned} ||U||^2_P \leq e^{|\mathcal{A}_{\mathbf{x}}|_{\max}t}\Big(||F||^2_P + \int_0^t e^{|\mathcal{A}_{\mathbf{x}}|_{\max}\xi}\big[&\hat{G}_E^T(E_N \otimes P_y)\mathcal{A}_{\mathbf{n}_E}\hat{G}_E + \hat{G}_W^T(E_0 \otimes P_y)\mathcal{A}_{\mathbf{n}_W}\hat{G}_W \\ &+ \hat{G}_N^T(P_x \otimes E_M)\mathcal{A}_{\mathbf{n}_N}\hat{G}_N + \hat{G}_S^T(E_x \otimes E_0)\mathcal{A}_{\mathbf{n}_S}\hat{G}_S\big]\,d\xi\Big) \end{aligned}$$

which is analogue to (4.17). Hence, the approximation (4.22) is strongly-stable which concludes the proof. □

4.2.3 Numerical test

The accuracy of the approximation (4.22) is validated by employing the method of manufactured solution. We consider the solution $u(x, y) = \cos(2\pi x)\sin(4\pi y)$ and the variable coefficients

$$a(x, y) = \cosh(2\pi y)e^{-3x}$$
$$b(x, y) = \sin(2\pi y + 1)\cos(xy).$$

The maximum norm of a, b is bounded on the domain $(x, y) \in [0, 1] \times [0, 1]$. Forcing function, boundary data are obtained by substituting the exact solution into (4.12) and boundary conditions are implemented weakly using SAT technique. The diffusion term is discretized using narrow stencil operator and the diffusion constant is $\varepsilon = 0.1$. The numerical solution procedure is the same as in the previous case, the coefficient matrix independent of u is constructed and inverted to obtained the approximation solution.

Table 4.3: L_2 norm of errors and convergence rates are compared for the advection-diffusion equation with variable coefficients

N = M	$\log L_2$	$q^{(2,1)}$	$\log L_2$	$q^{(4,2)}$	$\log L_2$	$q^{(6,3)}$	$\log L_2$	$q^{(8,4)}$
40	-1.981	2.104	-3.679	4.161	-4.052	5.665	-3.392	5.216
60	-2.342	2.051	-4.436	4.302	-4.993	5.343	-4.358	5.487
80	-2.596	2.034	-4.97	4.275	-5.659	5.329	-5.137	6.228
100	-2.793	2.025	-5.381	4.235	-6.182	5.394	-5.761	6.441

The convergence rates of the approximation (4.22) and the L_2 norms of errors are presented in Table. 4.3 for the $(2s, s)$-accurate SBP operators. The global orders of accuracy asymtotes towards $s + 2$ where $s = 1, \ldots, 4$. Therefore, the scheme is provably stable and accurate up to 8^{th} order.

Chapter 5

The laminar boundary layer

The concept of boundary layer is encountered in various engineering fields such as golf-ball enhancement [40, 41], the automobile industry [42] and atmospheric boundary layer modeling [43, 44]. This refers to the fluids in the immediate neighborhood adjacent to the solid boundary surface where viscous forces dominates the inertial forces. This chapter is concerned with a development of an energy-stable high order approximation scheme for the incompressible laminar boundary layer equations. The incompressible Navier-Stokes equations are used as an original model of the problem. Using order-of-magnitude analysis, these equations are simplified to boundary layer equations specific to the so-called flat-plate theory.

Consider the laminar incompressible flow over a flat plate of finite length L which is aligned with a free-stream velocity U_∞. This is shown schematically in Fig. 5.1. Due to shear forces between the layers of the fluid and wall's no-slip velocity condition, a boundary layer is formed in the immediate vicinity of the solid boundary. The thickness of the boundary layer, denoted by $\delta(x)$, is defined as a distance from the solid boundary to a point in the flow where the $x-$component of the flow velocity is approximately $0.99U_\infty$. The boundary layer thickness grows rapidly in the vicinity of the plate leading edge resulting in the largest gradients in all velocity components.

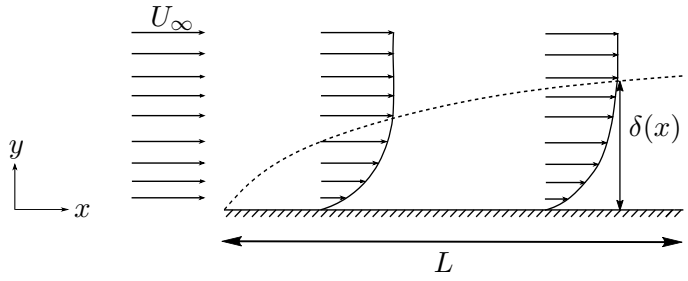

Figure 5.1: Boundary-layer over a flat plate.

The incompressible Navier-Stokes equation under isothermal conditions in 2-D are

$$\begin{aligned}\frac{\partial u}{\partial t}+u\frac{\partial u}{\partial x}+v\frac{\partial u}{\partial y}&=-\frac{1}{\rho}\frac{\partial p}{\partial x}+\nu\Big(\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}\Big),\\ \frac{\partial v}{\partial t}+u\frac{\partial v}{\partial x}+v\frac{\partial v}{\partial y}&=-\frac{1}{\rho}\frac{\partial p}{\partial y}+\nu\Big(\frac{\partial^2 v}{\partial x^2}+\frac{\partial^2 v}{\partial y^2}\Big),\\ \frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}&=0,\end{aligned} \tag{5.1}$$

where u, v are the components of velocity in x and y direction respectively while the pressure and density are denoted by $p=p(x,y)$ and ρ. The incompressible isothermal flow condition implies that ρ is constant. The kinematic viscosity ν is defined as $\nu=\frac{\mu}{\rho}$, where μ is the dynamical viscosity. To derive the boundary layer equations, we first assume that the streamwise length scale L is much larger than the transverse length scale δ and that it varies by $\mathcal{O}(Re)^{1/2}$

$$\delta \ll L, \quad \frac{L}{\delta}\sim\sqrt{Re}.$$

Here, Re is a Reynolds number expressed as a ratio of inertial forces to viscous forces inside the boundary layer, $Re=\dfrac{U_\infty L}{\nu}$. We define the following dimensionless quantities denoted by *

$$\begin{aligned} x^*&=\frac{x}{L}, & y^*&=\frac{y}{L}\sqrt{Re}, & t^*&=\frac{U_\infty}{L}t,\\ u^*&=\frac{u}{U_\infty}, & v^*&=\frac{v}{U}\sqrt{Re}, & p^*&=\frac{p-p_0}{\rho U_\infty^2},\end{aligned} \tag{5.2}$$

where p_0 is the atmospheric pressure. Substituting (5.2) into (5.1) yields a dimensionless governing system

$$\begin{aligned}\frac{\partial u^*}{\partial t^*}+u^*\frac{\partial u^*}{\partial x^*}+v^*\frac{\partial u^*}{\partial y^*}&=-\frac{\partial p^*}{\partial x^*}+\frac{1}{Re}\frac{\partial^2 u^*}{\partial x^{*2}}+\frac{\partial^2 u^*}{\partial y^{*2}},\\ \frac{1}{Re}\Big(\frac{\partial v^*}{\partial t^*}+u^*\frac{\partial v^*}{\partial x^*}+v^*\frac{\partial v^*}{\partial y^*}\Big)&=-\frac{\partial p^*}{\partial y^*}+\frac{1}{Re^2}\frac{\partial^2 v^*}{\partial x^{*2}}+\frac{1}{Re}\frac{\partial^2 v^*}{\partial y^{*2}},\\ \frac{\partial u^*}{\partial x^*}+\frac{\partial v^*}{\partial y^*}&=0.\end{aligned} \tag{5.3}$$

For significantly large Re, equations in (5.3) reduces to

$$\begin{aligned}\frac{\partial u^*}{\partial t^*}+u^*\frac{\partial u^*}{\partial x^*}+v^*\frac{\partial u^*}{\partial y^*}&=-\frac{\partial p^*}{\partial x^*}+\frac{\partial^2 u^*}{\partial y^{*2}},\\ 0&=-\frac{\partial p^*}{\partial y^*},\\ \frac{\partial u^*}{\partial x^*}+\frac{\partial v^*}{\partial y^*}&=0.\end{aligned} \tag{5.4}$$

Notice that the continuity equation is invariant to this dimensions scaling and pressure is a function of x only inside the boundary layer. The Bernoulli equation can be used to relate it to the stream velocity

$$p\propto-\frac{U_\infty^2}{2}$$

and for constant U_∞, the pressure gradients in (5.4) can be neglected. To keep the notation cleaner, we drop the asterisks. The system (5.4) is accurate and valid only for large Re, however, it must be moderated since for very large Re, the flow transition to turbulent flow.

5.1 The continuous problem

Boundary layer equations (5.4) can be written in a matrix-vector form as

$$\tilde{I}U_t + AU_x + BU_y = \tilde{I}U_{yy} \qquad (x,y) \in \Omega,\ t > 0, \tag{5.5a}$$

$$LU = \mathbf{g}(x,y,t) \quad (x,y) \in \partial\Omega, t > 0, \tag{5.5b}$$

$$U = \mathbf{f}(x,y) \qquad (x,y) \in \Omega,\ t = 0, \tag{5.5c}$$

where

$$U = \begin{bmatrix} u \\ v \\ p \end{bmatrix}, \quad A = \begin{bmatrix} u & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} v & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \quad \text{and} \quad \tilde{I} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

The boundary operator is denoted by L while $\mathbf{g}, \mathbf{f} \in \mathcal{F}$ specify the boundary and initial data respectively. We assume these are smooth and compatible such that the solution is smooth.

To construct a well-posed IBVP and the corresponding stable SBP-SAT approximation, we follow [23] and [45] for the construction of energy-stable boundary conditions for the incompressible N-S equations. As we have already seen, well-posedness depends almost entirely on the boundary conditions. In the linear case, if the solution exist and is bounded then uniqueness follows directly from the bound. However, this is generally not the case with nonlinear problems and hence we shall not refer to problem (5.5a) – (5.5c) as a well-posed problem even if we have an estimate. We will employ the energy method to construct the boundary operator L in (5.5b) such that it implements the correct minimal number of boundary conditions with the appropriate form at the correct boundary locations.

For the upcoming discrete analysis, we split the convective terms in (5.5a) into a skew-symmetric form

$$AU_x = \frac{1}{2}(AU)_x + \frac{1}{2}AU_x - \frac{1}{2}A_xU, \quad BU_y = \frac{1}{2}(BU)_y + \frac{1}{2}BU_y - \frac{1}{2}B_yU. \tag{5.6}$$

Notice that $A_x + B_y = \tilde{I}(u_x + v_y) = \mathbf{0}$ using the divergence-free velocity condition. The coefficient matrices A and B are symmetric as required by the integration-by-parts rule, therefore, the energy method can be applied directly. Equation (5.5a) with the convective terms written as in (5.6) becomes

$$\tilde{I}U_t + \frac{1}{2}\Big[(AU)_x + AU_x + (BU)_y + BU_y\Big] = \tilde{I}U_{yy}. \tag{5.7}$$

We define the inner product $(\cdot,\cdot)_{\tilde{I}}$ and the corresponding semi-norm as

$$(U,V)_{\tilde{I}} = \int_\Omega U^T\tilde{I}V\,d\Omega, \quad ||U||^2_{\tilde{I}} = (U,U)_{\tilde{I}} \tag{5.8}$$

for any two vector functions $U, V \in \mathcal{F}$.

5.1.1 The energy method

We next apply the energy method to (5.7) posed on the regular quadrilateral domain depicted in Fig. 5.2. The west boundary is aligned with the plate's leading edge while the east boundary is such that L ensures laminar flow. The south boundary is on the plate's surface and the north boundary is chosen such that it is significantly larger than $\delta(x)$.

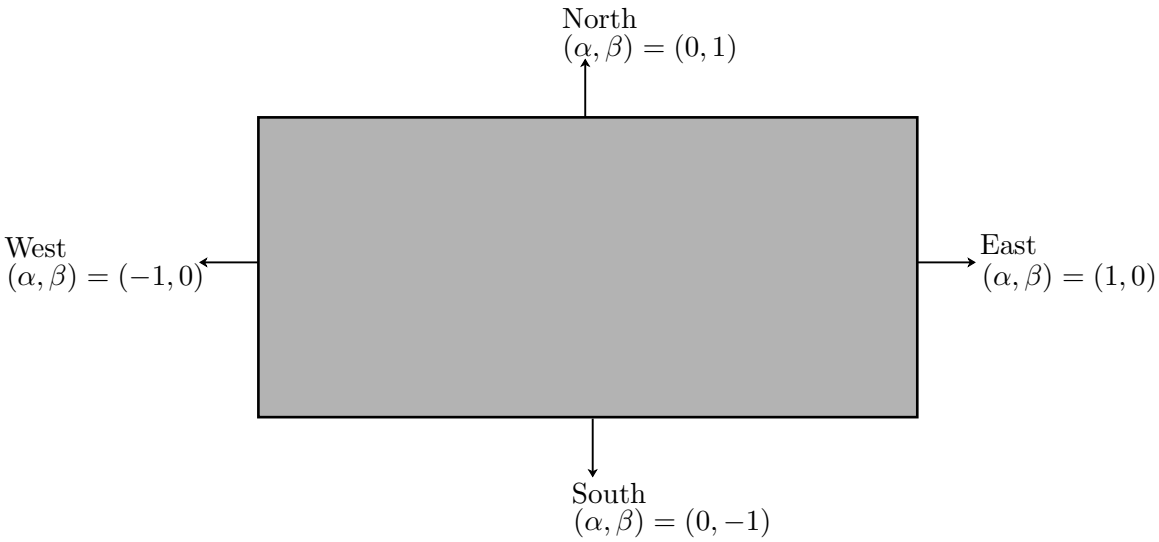

Figure 5.2: The two dimensional computational domain showing normal vectors for the boundary layer equations.

Applying the energy method and Gauss theorem to (5.7) yields

$$\frac{d}{dt}||U||_{\tilde{I}}^2 + 2||U_y||_{\tilde{I}}^2 = -\oint_{\partial\Omega} \left[U^T(\alpha A + \beta B)U - 2\beta U^T \tilde{I} U_y\right] ds \tag{5.9}$$

where α, β are the x and y components of the unit normal vector $\mathbf{n}$ and $ds = \sqrt{dx^2 + dy^2}$ denotes the infinitesimal line element as before. The quadratic boundary term on the RHS of (5.9) can be written in matrix-vector form as

$$\frac{d}{dt}||U||_{\tilde{I}}^2 + 2||U_y||_{\tilde{I}}^2 = -\oint_{\partial\Omega} \mathbf{q}^T M \mathbf{q} ds \tag{5.10}$$

where

$$\mathbf{q} = \begin{bmatrix} u \\ v \\ p \\ u_y \end{bmatrix}, \quad M = \begin{bmatrix} u_{\mathbf{n}} & 0 & \alpha & -\beta \\ 0 & 0 & \beta & 0 \\ \alpha & \beta & 0 & 0 \\ -\beta & 0 & 0 & 0 \end{bmatrix}.$$

Here, $u_{\mathbf{n}} = \alpha u + \beta v$ is the boundary normal velocity.

The boundary matrix M provides all the information we need in order to bound the energy rate (5.10).

Remark 3. The minimal number of boundary condition that leads to the energy estimate is equal to the number of negative eigenvalues of the boundary matrix M.

The eigenvalues of M denoted by λ are obtained by solving the characteristic equation

$$\det(M - \lambda I_4) = \lambda^4 - u_{\mathbf{n}}\lambda^3 - (\alpha^2 + 2\beta^2)\lambda^2 + \beta^2 u_{\mathbf{n}}\lambda + \beta^4 = 0 \tag{5.11}$$

where I_4 is a 4×4 unit matrix.

Computing the zeros of (5.11) is a nontrivial task in general. However, in our case it simplifies since we consider a regular quadrilateral domain where the normal vectors are constant as shown in Fig. 5.2. For each boundary, (5.11) becomes

$$\begin{aligned}
\lambda^4 - u_{\mathbf{n}}\lambda^3 - \lambda^2 = 0, &\qquad \text{east boundary} \\
\lambda^4 - u_{\mathbf{n}}\lambda^3 - \lambda^2 = 0, &\qquad \text{west boundary} \\
\lambda^4 - u_{\mathbf{n}}\lambda^3 - 2\lambda^2 + u_{\mathbf{n}}\lambda + 1 = 0, &\qquad \text{north boundary} \\
\lambda^4 - u_{\mathbf{n}}\lambda^3 - 2\lambda^2 + u_{\mathbf{n}}\lambda + 1 = 0, &\qquad \text{south boundary.}
\end{aligned} \tag{5.12}$$

The eigenvalues λ_i's in an increasing order obtained by solving (5.12) for each boundary and the associated orthonormal basis of eigenvectors indicated by X are

i. East boundary

$$\lambda_{1,4} = \frac{u_{\mathbf{n}}}{2} \mp \sqrt{\left(\frac{u_{\mathbf{n}}}{2}\right)^2 + 1}, \quad \lambda_{2,3} = 0, \tag{5.13a}$$

$$X = \begin{bmatrix} \lambda_1 & 0 & 0 & \lambda_4 \\ 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \quad N = \operatorname{diag}\Big(\frac{1}{\sqrt{\lambda_1^2+1}}, 1, 1, \frac{1}{\sqrt{\lambda_4^2+1}}\Big)$$

ii. West boundary

$$\lambda_{1,4} = \frac{u_{\mathbf{n}}}{2} \mp \sqrt{\left(\frac{u_{\mathbf{n}}}{2}\right)^2 + 1}, \quad \lambda_{2,3} = 0, \tag{5.13b}$$

$$X = \begin{bmatrix} \lambda_1 & 0 & 0 & \lambda_4 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad N = \operatorname{diag}\Big(\frac{1}{\sqrt{\lambda_1^2+1}}, 1, 1, \frac{1}{\sqrt{\lambda_4^2+1}}\Big)$$

iii. North boundary

$$\lambda_{1,4} = \frac{u_{\mathbf{n}}}{2} \mp \sqrt{\left(\frac{u_{\mathbf{n}}}{2}\right)^2 + 1}, \quad \lambda_{2,3} = \mp - 1, \tag{5.13c}$$

$$X = \begin{bmatrix} \lambda_1 & 0 & 0 & \lambda_4 \\ 0 & 1 & 1 & 0 \\ 0 & -1 & 1 & 0 \\ -1 & 0 & 0 & -1 \end{bmatrix}, \quad N = \operatorname{diag}\Big(\frac{1}{\sqrt{\lambda_1^2+1}}, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{\lambda_4^2+1}}\Big)$$

iv. South boundary

$$\lambda_{1,4} = \frac{u_{\mathbf{n}}}{2} \mp \sqrt{\left(\frac{u_{\mathbf{n}}}{2}\right)^2 + 1}, \quad \lambda_{2,3} = \mp - 1, \tag{5.13d}$$

$$X = \begin{bmatrix} \lambda_1 & 0 & 0 & \lambda_4 \\ 0 & 1 & 1 & 0 \\ 0 & -1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}, \quad N = \operatorname{diag}\Big(\frac{1}{\sqrt{\lambda_1^2+1}}, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{\lambda_4^2+1}}\Big).$$

In (5.13a) – (5.13d), N denotes a normalizing matrix.

Notice that for each boundary, $\lambda_1 < 0$ and $\lambda_4 > 0$ for all $u_{\mathbf{n}}$. Therefore, there is exactly one negative eigenvalue λ_1 at the east and west boundary. Similarly, there are exactly two negative eigenvalues $\lambda_{1,2}$ at the north and south boundary.

Remark 4. The sign of the eigenvalues imply that precisely one boundary condition must be specified at the east and west boundary and two boundary conditions for both north and south boundary.

Next, let's consider the eigenvalue decomposition of M

$$M(XN) = (XN)\Lambda_M, \qquad \Lambda_M = \operatorname{diag}(\lambda_1, \ldots, \lambda_4)$$

where XN is a normalized eigenvectors matrix defined in (5.13a) – (5.13d) for each boundary. Using the fact that the columns of XN are linearly independent i.e. XN is orthogonal ($(XN)^{-1} = (XN)^T$), we rewrite the decomposition above as

$$M = X\Lambda X^T, \qquad \Lambda = N\Lambda_M N^T. \tag{5.14}$$

The elements of a new scaled eigenvalue matrix Λ are bounded for all $|u_{\mathbf{n}}|$. Next, we arrange Λ as a block-diagonal matrix of three disjoint sub-matrices

$$\Lambda = \operatorname{diag}(\Lambda^+, \Lambda^0, \Lambda^-)$$

where the diagonal matrices Λ^+, Λ^0 and Λ^- contain the positive, zero and negative eigenvalues respectively. In a similar way, the eigenvector matrix X is partitioned as a column vector of

vectors $X = [X_+, X_0, X_-]$, where X_+, X_0 and X_- are sets of eigenvectors corresponding to the positive, zero and negative eigenvalues respectively. After substituting (5.14) into (5.10) and utilizing the splitting above, the energy rate becomes

$$\frac{d}{dt}||U||^2_{\tilde{I}} + 2||U_y||^2_{\tilde{I}} = -\oint_{\partial\Omega} \begin{bmatrix} X_+^T\mathbf{q} \\ X_0^T\mathbf{q} \\ X_-^T\mathbf{q} \end{bmatrix}^T \begin{bmatrix} \Lambda^+ & 0 & 0 \\ 0 & \Lambda^0 & 0 \\ 0 & 0 & \Lambda^- \end{bmatrix} \begin{bmatrix} X_+^T\mathbf{q} \\ X_0^T\mathbf{q} \\ X_-^T\mathbf{q} \end{bmatrix} ds \tag{5.15}$$

where $\mathbf{q}$ is given in (5.10). We simplify the notation above by defining $\mathbf{w}^+ = X_+^T\mathbf{q}$, $\mathbf{w}^0 = X_0^T\mathbf{q}$ and $\mathbf{w}^- = X_-^T\mathbf{q}$. The variables $\mathbf{w}^-$ and $\mathbf{w}^+$ are called the incoming and outgoing characteristics propagating in and out of the domain, $\mathbf{w}^0$ is not interesting and is never used since Λ^0 is zero. The energy rate (5.15) becomes

$$\frac{d}{dt}||U||^2_{\tilde{I}} + 2||U_y||^2_{\tilde{I}} = -\oint_{\partial\Omega} \left(\mathbf{w}^{+T}\Lambda^+\mathbf{w}^+ + \mathbf{w}^{-T}\Lambda^-\mathbf{w}^-\right) ds. \tag{5.16}$$

So far, we have determined the correct number of boundary conditions required such that an energy bound is obtained. We now turn the focus to the correct form of boundary conditions that will yield a bounded estimate. The desired form of boundary conditions must specify the incoming characteristics in terms of the outgoing ones and boundary data.

The general relation between the outgoing and incoming characteristics is

$$\mathbf{w}^- = R\mathbf{w}^+ + \mathbf{g} \tag{5.17}$$

where R is a matrix with the number of rows equal to the number of negative eigenvalues and $\mathbf{g}$ is boundary data (5.5b). Equation (5.17) defines the form of the boundary operator L in (5.5b) that will yield an estimate i.e.

$$LU\big|_{\partial\Omega} = \mathbf{w}^-\big|_{\partial\Omega} - R\mathbf{w}^+\big|_{\partial\Omega} = \mathbf{g}. \tag{5.18}$$

The second boundary term in (5.16) might add growth to the energy estimate. However, the following propositions shows that a strong imposition of boundary condition (5.18) leads to the energy bound.

Proposition 5.1. *The homogeneous form of the boundary condition* (5.18) *yields an estimate if*

$$\Lambda^+ + R^T\Lambda^- R \geq 0. \tag{5.19}$$

Proof. Substituting (5.18) with zero data into (5.16) to eliminate $\mathbf{w}^-$ yields

$$\begin{aligned} \frac{d}{dt}||U||^2_{\tilde{I}} + 2||U_y||^2_{\tilde{I}} &= -\oint_{\partial\Omega} (\mathbf{w}^{+T}\Lambda^+\mathbf{w}^+ + (R\mathbf{w}^+)^T\Lambda^- R\mathbf{w}^+) ds \\ &= -\oint_{\partial\Omega} \mathbf{w}^{+T}(\Lambda^+ + R^T\Lambda^- R)\mathbf{w}^+ ds. \end{aligned} \tag{5.20}$$

The energy-rate (5.20) is bounded if (5.19) holds, and hence temporal integration yields an estimate. □

The condition (5.19) is sufficient for an energy estimate. The boundary conditions must have the form (5.18) and matrix R must satisfy (5.19). In the next proposition, we prove that the non-homogeneous form of boundary condition (5.18) with a sharper requirement on condition (5.19) also yields an estimate.

Proposition 5.2. *The non-homogeneous form of boundary condition* (5.18) *yields an estimate if there exist a positive semi-definite matrix* G *satisfying* $(R\Lambda^-)(\Lambda^+ + R^T\Lambda^- R)^{-1}(R\Lambda^-)^T - \Lambda^- \leq G < \infty$ *and* $\Lambda^+ + R^T\Lambda^- R > 0$.

Proof. Let G be a positive semi-definite matrix. By inserting (5.18) with nonzero data into (5.16) and adding and subtracting $\oint_{\partial\Omega} \mathbf{g}^T G\mathbf{g} ds$, we obtain

$$\frac{d}{dt}||U||_{\tilde{I}}^2 + 2||U_y||_{\tilde{I}}^2 = -\oint_{\partial\Omega} \begin{bmatrix} \mathbf{w}^+ \\ \mathbf{g} \end{bmatrix}^T \underbrace{\begin{bmatrix} \Lambda^+ + R^T\Lambda^- R & R^T\Lambda^- \\ \Lambda^- R & \Lambda^- + G \end{bmatrix}}_{K} \begin{bmatrix} \mathbf{w}^+ \\ \mathbf{g} \end{bmatrix} ds + \oint_{\partial\Omega} \mathbf{g}^T G\mathbf{g} ds \tag{5.21}$$

which is bounded by data if K is positive semi-definite and $G < \infty$. We determine the signs of the eigenvalues of K using the rotation technique and find

$$K = Y^T P Y, \quad P = \begin{bmatrix} \Lambda^+ + R^T\Lambda^- R & 0 \\ 0 & \Lambda^- - (\Lambda^- R)(\Lambda^+ + R^T\Lambda^- R)^{-1}(\Lambda^- R)^T + G \end{bmatrix}$$

where the rotation matrix Y is given by the upper triangular matrix

$$Y = \begin{bmatrix} I & -(\Lambda^+ + R^T\Lambda^- R)^{-1}(\Lambda^- R)^T \\ 0 & I \end{bmatrix}.$$

The eigenvalues of P are not necessarily the same as those of K, however, they have the same signs [46]. The matrix P and hence K is positive semi-definite if P_{11} and P_{22} are non-negative matrices . From (5.19), we have that $P_{11} \geq 0$ and in order for the inverse matrix in P_{22} and Y_{12} to make sense, we require it to be strictly positive i.e. $P_{11} > 0$. The matrix element P_{22} is non-negative if a bounded positive semi-definite matrix G can be chosen such that

$$\Lambda^- - (\Lambda^- R)(\Lambda^+ + R^T\Lambda^- R)^{-1}(\Lambda^- R)^T + G \geq 0$$

which implies

$$G \geq (\Lambda^- R)(\Lambda^+ + R^T\Lambda^- R)^{-1}(\Lambda^- R)^T - \Lambda^-. \tag{5.22}$$

In (5.22), the matrix G is bounded from below by a positive matrix. In addition to this, we require it to be finite i.e. $G < \infty$ such that the boundary data term on the RHS of (5.21)

is bounded. Therefore, the energy rate (5.21) is bounded and time integration over a finite domain $[0, T]$ yields the estimate

$$||U||_{\tilde{I}}^2 + 2\int_0^T ||U_y||_{\tilde{I}}^2 dt \leq ||\mathbf{f}||_{\tilde{I}}^2 + \int_0^T \oint_{\partial\Omega} \mathbf{g}^T G \mathbf{g} ds dt. \tag{5.23}$$

□

Remark 5. Note that the estimate (5.23) only gives a bound on the streamwise velocity u.

5.1.2 The form of boundary operator L

The operator L can be decomposed as

$$LU = (L^- - RL^+)U = \mathbf{g}$$

where

$$L^-U = X_-^T TU = \mathbf{w}^-, \quad L^+U = X_+^T TU = \mathbf{w}^+ \tag{5.24}$$

and the transformation matrix T is defined as

$$T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ \partial/\partial y & 0 & 0 \end{bmatrix}$$

such that $\mathbf{q} = TU$. For each boundary, we then have

$$\begin{aligned} L^-U\big|_{\partial\Omega_E} &= \lambda_1 u - p, & L^+U\big|_{\partial\Omega_E} &= \lambda_4 u - p \\ L^-U\big|_{\partial\Omega_W} &= \lambda_1 u + p, & L^+U\big|_{\partial\Omega_W} &= \lambda_4 u + p \\ L^-U\big|_{\partial\Omega_N} &= \begin{bmatrix} \lambda_1 u - u_y \\ v - p \end{bmatrix}, & L^+U\big|_{\partial\Omega_N} &= \begin{bmatrix} v + p \\ \lambda_4 u - u_y \end{bmatrix} \\ L^-U\big|_{\partial\Omega_S} &= \begin{bmatrix} \lambda_1 u + u_y \\ v - p \end{bmatrix}, & L^+U\big|_{\partial\Omega_S} &= \begin{bmatrix} v + p \\ \lambda_4 u + u_y \end{bmatrix}. \end{aligned} \tag{5.25}$$

The matrix R uniquely defines the type of boundary condition imposed on a particular boundary and it must satisfy condition (5.19). The most commonly used boundary conditions in CFD are solid wall, Robin and Neumann boundary conditions. For the problem (5.4), we impose solid wall conditions (Dirichlet conditions) for both streamwise and transverse velocity components on the solid boundary. Dirichlet streamwise velocity and pressure conditions are specified on the inlet and outlet boundary respectively. Furthermore, at the north boundary, we specify u_y and p. We next demonstrate that these boundary conditions yields energy estimates and satisfies (5.19). Recall that the size of R is determined by the number of negative

eigenvalues and associated linearly independent eigenvectors. Therefore, from (5.13a)−(5.13d), the size of R is 1×1 for east, west boundary and 2×2 for the north and south boundary.

1. *Solid wall (Dirichlet) boundary condition at south boundary.*
We seek R and a non-singular coefficient matrix $C = \mathrm{diag}(c_1, c_2)$ such that

$$L^- U\big|_{\partial\Omega_S} - RL^+ U\big|_{\partial\Omega_S} = C \begin{bmatrix} u \\ v \end{bmatrix}_{\partial\Omega_S} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

After solving a system of linear equations, we find

$$R = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \quad c_1 = \lambda_1 - \lambda_4 = -\sqrt{u_{\mathbf{n}}^2 + 4}, \quad c_2 = 2$$

and R satisfies (5.19) since

$$\Lambda^+ + R^T \Lambda^- R = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{\lambda_4}{\lambda_4^2+1} \end{bmatrix} + \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{\lambda_1}{\lambda_1^2+1} & 0 \\ 0 & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

2. *Inlet boundary condition at west boundary.*
Considering the west boundary condition in (5.25)

$$L^- U\big|_{\partial\Omega_W} - RL^+ U\big|_{\partial\Omega_W} = (\lambda_1 - R\lambda_4)u + (1 - R)p = cu = \mathbf{g}_W$$

where c is a coefficient yet to be determined by specifying R such that the pressure vanishes. The obvious way to eliminate the pressure is to set $R = 1$. This choice also satisfies (5.19) since

$$\Lambda^+ + \Lambda^- = \frac{\lambda_4}{\lambda_4^2 + 1} + \frac{\lambda_1}{\lambda_1^2 + 1} = 0$$

and $c = \lambda_1 - \lambda_4 = -\sqrt{u_{\mathbf{n}}^2 + 4}$.

3. *Outlet boundary condition at east boundary.*
In a similar way, the pressure condition is specified as

$$L^- U\big|_{\partial\Omega_E} - RL^+ U\big|_{\partial\Omega_E} = (\lambda_1 - R\lambda_4)u + (R - 1)p = cp = \mathbf{g}_E$$

by setting $R = \dfrac{\lambda_1}{\lambda_4}$ which gives $c = \dfrac{\lambda_1}{\lambda_4} - 1$ and satisfies (5.19)

$$\Lambda^+ + \left(\frac{\lambda_1}{\lambda_4}\right)^2 \Lambda^- = \frac{\lambda_4}{\lambda_4^2 + 1} + \left(\frac{\lambda_1}{\lambda_4}\right)^2 \frac{\lambda_1}{\lambda_1^2 + 1} = \frac{u_{\mathbf{n}}(u_{\mathbf{n}}^2 + 1)}{(\lambda_1^2 + 1)(\lambda_4^2 + 1)\lambda_4^2} \geq 0$$

since $u_{\mathbf{n}} \geq 0$ at the outflow boundary.

4. *Free-stream boundary condition at north boundary.*
Consider the north boundary. We seek R and C such that

$$L^- U\big|_{\partial\Omega_N} - RL^+ U\big|_{\partial\Omega_N} = C \begin{bmatrix} u_y \\ p \end{bmatrix}_{\partial\Omega_N} = \mathbf{g}_N.$$

The matrix

$$R = \begin{bmatrix} 0 & \frac{\lambda_1}{\lambda_4} \\ 1 & 0 \end{bmatrix}$$

removes u and v from the boundary condition LU and satisfies

$$\Lambda^+ + R^T \Lambda^- R = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{\lambda_4}{\lambda_4^2+1} \end{bmatrix} + \begin{bmatrix} 0 & \frac{\lambda_1}{\lambda_4} \\ 1 & 0 \end{bmatrix}^T \begin{bmatrix} \frac{\lambda_1}{\lambda_1^2+1} & 0 \\ 0 & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} 0 & \frac{\lambda_1}{\lambda_4} \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & \beta \end{bmatrix}.$$

The constants are

$$\beta = \frac{\lambda_4}{\lambda_4^2+1} + \Big(\frac{\lambda_1}{\lambda_4}\Big)^2 \frac{\lambda_1}{\lambda_1^2+1} = \frac{u_{\mathbf{n}}(u_{\mathbf{n}}^2+1)}{(\lambda_1^2+1)(\lambda_4^2+1)\lambda_4^2} \geq 0 \quad \text{and} \quad C = \operatorname{diag}\Big(\frac{\lambda_1}{\lambda_4} - 1, -2\Big),$$

leading to an energy bound.

5.1.3 Weak boundary conditions

Next, we implement the boundary conditions (5.5b) weakly. Equation (5.7) together with boundary conditions becomes

$$\tilde{I} U_t + \frac{1}{2}\Big[(AU)_x + AU_x + (BU)_y + BU_y\Big] = \tilde{I} U_{yy} + \mathcal{L}\big(\Sigma(LU - \mathbf{g})\big) \tag{5.26}$$

where Σ is a penalty coefficient matrix and $\mathcal{L}(\cdot)$ denotes the lifting operator (2.5).

The energy method applied to (5.26) yields

$$\frac{d}{dt}||U||_{\tilde{I}}^2 + 2||U_y||_{\tilde{I}}^2 = -\oint_{\partial\Omega} \mathbf{q}^T M \mathbf{q} ds + \oint_{\partial\Omega} \big(U^T \Sigma(LU - \mathbf{g}) + (U^T \Sigma(LU - \mathbf{g}))^T\big) ds \tag{5.27}$$

where $\mathbf{q}$ and M are given in (5.10). By diagonalizing M as in (5.14) and substituting the correct form of LU as specified in (5.18), the energy rate (5.27) becomes

$$\frac{d}{dt}||U||_{\tilde{I}}^2 + 2||U_y||_{\tilde{I}}^2 = BT. \tag{5.28}$$

In (5.28), BT denotes the boundary terms which are

$$\begin{aligned} BT = &- \oint_{\partial\Omega} \begin{bmatrix} \mathbf{w}^+ \\ \mathbf{w}^- \end{bmatrix}^T \begin{bmatrix} \Lambda^+ & 0 \\ 0 & \Lambda^- \end{bmatrix} \begin{bmatrix} \mathbf{w}^+ \\ \mathbf{w}^- \end{bmatrix} ds \\ &+ \oint_{\partial\Omega} \big(U^T \Sigma(\mathbf{w}^- - R\mathbf{w}^+ - \mathbf{g}) + (U^T \Sigma(\mathbf{w}^- - R\mathbf{w}^+ - \mathbf{g}))^T\big) ds. \end{aligned}$$

Next, we will determine the penalty coefficient Σ such that BT is bounded by data. Let us first consider (5.28) with zero data and make the ansatz $U^T\Sigma = \mathbf{w}^{-T}\tilde{\Sigma}$. We find

$$BT = -\oint_{\partial\Omega} \begin{bmatrix} \mathbf{w}^+ \\ \mathbf{w}^- \end{bmatrix}^T \begin{bmatrix} \Lambda^+ & (\tilde{\Sigma}R)^T \\ \tilde{\Sigma}R & \Lambda^- - (\tilde{\Sigma} + \tilde{\Sigma}^T) \end{bmatrix} \begin{bmatrix} \mathbf{w}^+ \\ \mathbf{w}^- \end{bmatrix} ds. \tag{5.29}$$

Proposition 5.3. *The estimate* (5.28) *with homogeneous boundary conditions is bounded if* $\tilde{\Sigma} = \Lambda^-$.

Proof. Substituting $\tilde{\Sigma} = \Lambda^-$ in (5.29) yields

$$BT = -\oint_{\partial\Omega} \begin{bmatrix} \mathbf{w}^+ \\ \mathbf{w}^- \end{bmatrix}^T \begin{bmatrix} \Lambda^+ & R^T\Lambda^- \\ \Lambda^- R & -\Lambda^- \end{bmatrix} \begin{bmatrix} \mathbf{w}^+ \\ \mathbf{w}^- \end{bmatrix} ds.$$

Adding and subtracting $\oint_{\partial\Omega} \mathbf{w}^{+T} R^T\Lambda^- R\mathbf{w}^+ ds$ in the equation above leads to

$$\begin{aligned} BT &= \oint_{\partial\Omega} \begin{bmatrix} \mathbf{w}^+ \\ \mathbf{w}^- \end{bmatrix}^T \begin{bmatrix} R^T\Lambda^- R & -R^T\Lambda^- \\ -\Lambda^- R & \Lambda^- \end{bmatrix} \begin{bmatrix} \mathbf{w}^+ \\ \mathbf{w}^- \end{bmatrix} ds - \oint_{\partial\Omega} \mathbf{w}^{+T}\big(\Lambda^+ + R^T\Lambda^- R\big)\mathbf{w}^+ ds \\ &= \oint_{\partial\Omega} (\mathbf{w}^- - R\mathbf{w}^+)^T\Lambda^-(\mathbf{w}^- - R\mathbf{w}^+) ds - \oint_{\partial\Omega} \mathbf{w}^{+T}\big(\Lambda^+ + R^T\Lambda^- R\big)\mathbf{w}^+ ds. \end{aligned}$$

The first term above is negative and the second one is non-positive if (5.19) holds. □

Using Proposition 5.3, we can now compute the penalty coefficient matrix Σ as

$$U^T\Sigma = \mathbf{w}^{-T}\Lambda^- = (L^-U)^T\Lambda^- \quad \Rightarrow \quad \Sigma = L^{-T}\Lambda^- = T^T X_-\Lambda^-. \tag{5.30}$$

Next, we treat the case with boundary data.

Proposition 5.4. *The penalty coefficient* (5.30) *bounds* (5.28) *with nonzero boundary data.*

Proof. The boundary term (5.28) with Σ given by (5.30) becomes

$$\begin{aligned} BT = &-\oint_{\partial\Omega} \begin{bmatrix} \mathbf{w}^+ \\ \mathbf{w}^- \end{bmatrix}^T \begin{bmatrix} \Lambda^+ & 0 \\ 0 & \Lambda^- \end{bmatrix} \begin{bmatrix} \mathbf{w}^+ \\ \mathbf{w}^- \end{bmatrix} ds \\ &+ \oint_{\partial\Omega} \big[\mathbf{w}^{-T}\Lambda^-(\mathbf{w}^- - R\mathbf{w}^+ - \mathbf{g}) + \big(\mathbf{w}^{-T}\Lambda^-(\mathbf{w}^- - R\mathbf{w}^+ - \mathbf{g})\big)^T\big] ds. \\ = &-\oint_{\partial\Omega} \begin{bmatrix} \mathbf{w}^+ \\ \mathbf{w}^- \\ \mathbf{g} \end{bmatrix}^T \begin{bmatrix} \Lambda^+ & R^T\Lambda^- & 0 \\ \Lambda^- R & -\Lambda^- & \Lambda^- \\ 0 & \Lambda^- & 0 \end{bmatrix} \begin{bmatrix} \mathbf{w}^+ \\ \mathbf{w}^- \\ \mathbf{g} \end{bmatrix} ds. \end{aligned} \tag{5.31}$$

By the matrix splitting [23]

$$\begin{bmatrix} \Lambda^+ & R^T\Lambda^- & 0 \\ \Lambda^- R & -\Lambda^- & \Lambda^- \\ 0 & \Lambda^- & 0 \end{bmatrix} = \begin{bmatrix} \Lambda^+ + R^T\Lambda^- R & 0 & R^T\Lambda^- \\ 0 & 0 & 0 \\ \Lambda^- R & 0 & \Lambda^- + G \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -G \end{bmatrix} + \begin{bmatrix} -R^T\Lambda^- R & R^T\Lambda^- & -R^T\Lambda^- \\ \Lambda^- R & -\Lambda^- & \Lambda^- \\ -\Lambda^- R & \Lambda^- & -\Lambda^- \end{bmatrix}, \tag{5.32}$$

we find

$$BT = -\oint_{\partial\Omega} \begin{bmatrix} \mathbf{w}^+ \\ \mathbf{g} \end{bmatrix}^T \begin{bmatrix} \Lambda^+ + R^T\Lambda^- R & R^T\Lambda^- \\ \Lambda^- R & \Lambda^- + G \end{bmatrix} \begin{bmatrix} \mathbf{w}^+ \\ \mathbf{g} \end{bmatrix} ds + \oint_{\partial\Omega} \mathbf{g}^T G \mathbf{g} ds - \oint_{\partial\Omega} \bar{\mathbf{q}}^T D \bar{\mathbf{q}} ds. \tag{5.33}$$

In (5.33),

$$\bar{\mathbf{q}} = [\mathbf{w}^+, \mathbf{w}^-, \mathbf{g}]^T, \quad D = \begin{bmatrix} -R^T\Lambda^- R & R^T\Lambda^- & -R^T\Lambda^- \\ \Lambda^- R & -\Lambda^- & \Lambda^- \\ -\Lambda^- R & \Lambda^- & -\Lambda^- \end{bmatrix}$$

and G is a positive definite matrix satisfying the inequality (5.22). The first two boundary terms in (5.33) are identical to the one's in (5.21). However, in order to obtain an estimate, the last term must also be non-positive, i.e $D \geq 0$. Let's consider the multiplicative decomposition of D

$$D = Q^T(Z \otimes \Lambda^-)Q, \quad Q = \begin{bmatrix} R & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix}, \quad Z = \begin{bmatrix} -1 & 1 & -1 \\ 1 & -1 & 1 \\ -1 & 1 & -1 \end{bmatrix}. \tag{5.34}$$

The eigenvalue decomposition of Z is

$$Z = W\Lambda_z W^T, \quad W = \begin{bmatrix} 1 & -1 & 1 \\ -1 & 0 & 2 \\ 1 & 1 & 1 \end{bmatrix}, \quad \Lambda_z = \mathrm{diag}(-3, 0, 0). \tag{5.35}$$

Here, $\otimes$ denotes the Kronecker product (2.5). The term $Z \otimes \Lambda^-$ in (5.34) can be decomposed as

$$Z \otimes \Lambda^- = (W\Lambda_Z W^T) \otimes (I\Lambda^- I^T) = (W \otimes I)(\Lambda_Z \otimes \Lambda^-)(W \otimes I)^T \tag{5.36}$$

where

$$(\Lambda_Z \otimes \Lambda^-) = \mathrm{diag}(-3, 0, 0) \otimes \Lambda^- = 3 \begin{bmatrix} -\Lambda^- & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \geq 0.$$

By substituting (5.36) into (5.34) and defining $\bar{Q} = Q^T(W \otimes I)$, D simplifies to

$$D = \bar{Q}(\Lambda_Z \otimes \Lambda^-)\bar{Q}^T$$

which is positive semi-definite. Therefore, the energy rate (5.28) is bounded, and hence, the weak boundary conditions leads to an estimate. □

5.2 The semi-discrete formulation

The spatial discretization of (5.26) using the SBP-SAT method is considered next. The domain $\Omega \subset \mathbb{R}^2$ is discretized using $N \times M$ equidistant grid points $\Omega_{ij} = \Omega(x_i, y_j)$ where (x_i, y_j) are the Cartesian coordinates. Let $\boldsymbol{U} = [\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{p}]^T$ be an approximation solution of (5.26) where $\boldsymbol{u}$, $\boldsymbol{v}$ and $\boldsymbol{p}$ are $NM \times 1$ vectors denoting the numerical approximation of u, v and p on the grid Ω_{ij} arranged as $\boldsymbol{u} = (u_{11}, \dots, u_{1M}, \dots, u_{N1}, \dots, u_{NM})^T$, $\boldsymbol{v} = (v_{11}, \dots, v_{1M}, \dots, v_{N1}, \dots, v_{NM})^T$ and $\boldsymbol{p} = (p_{11}, \dots, p_{1M}, \dots, p_{n1}, \dots, p_{NM})^T$ respectively.

In order to ensure that the 2-D SBP matrices (3.2.3) operates on all the variables $\boldsymbol{u}$, $\boldsymbol{v}$ and $\boldsymbol{p}$, we arrange them as block-diagonal matrices of 3 disjoint sub-matrices using the Kronecker product

1. $\mathcal{P} = I_3 \otimes (P_x \otimes P_y)$
2. $\mathcal{D}_x = I_3 \otimes (D_x \otimes I_M)$, $\mathcal{D}_y = I_3 \otimes (I_N \otimes D_y)$
3. $\mathcal{Q}_x = I_3 \otimes (Q_x \otimes P_y)$, $\mathcal{Q}_y = I_3 \otimes (P_x \otimes Q_y)$
4. $\mathcal{D}_{yy} = I_3 \otimes (I_N \otimes D_y D_y)$
5. $\mathcal{B}_x = I_3 \otimes ((E_N - E_0) \otimes P_y)$, $\mathcal{B}_y = I_3 \otimes (P_x \otimes (E_M - E_0))$

where I_3 is a 3×3 identity matrix.

Let I_{NM} be an $NM \times NM$ identity matrix. Next, by introducing the following block-matrices

$$\boldsymbol{A} = \begin{bmatrix} U & \mathbf{0} & I_{NM} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ I_{NM} & \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} V & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & I_{NM} \\ \mathbf{0} & I_{NM} & \mathbf{0} \end{bmatrix}, \quad \tilde{\boldsymbol{I}} = \text{diag}(I_{NM}, \mathbf{0}, \mathbf{0}),$$
$$U = \text{diag}(\boldsymbol{u}), \quad V = \text{diag}(\boldsymbol{v}),$$

we formulate the semi-discrete approximation of (5.26) as

$$\tilde{\boldsymbol{I}}\,\boldsymbol{U}_t + \frac{1}{2}\Big[(\boldsymbol{A}\mathcal{D}_x + \mathcal{D}_x\boldsymbol{A}) + (\boldsymbol{B}\mathcal{D}_y + \mathcal{D}_y\boldsymbol{B})\Big]\,\boldsymbol{U} = \tilde{\boldsymbol{I}}\mathcal{D}_{yy}\,\boldsymbol{U} + \mathcal{S}(\boldsymbol{U}), \tag{5.37}$$
$$\boldsymbol{U}_0 = F.$$

In (5.37), $\mathcal{S}(\boldsymbol{U})$ denotes the weakly implemented boundary conditions for each boundary as

$$\begin{aligned}\mathcal{S} = (I_3 \otimes (I_N \otimes P_y^{-1} E_M))\boldsymbol{\Sigma}_{_N}(\boldsymbol{L}_{_N}\boldsymbol{U} - G_{_N}) + (I_3 \otimes (P_x^{-1} E_N \otimes I_M))\boldsymbol{\Sigma}_{_E}(\boldsymbol{L}_{_E}\boldsymbol{U} - G_{_E}) \\ + (I_3 \otimes (I_N \otimes P_y^{-1} E_0))\boldsymbol{\Sigma}_{_S}(\boldsymbol{L}_{_S}\boldsymbol{U} - G_{_S}) + (I_3 \otimes (P_x^{-1} E_0 \otimes I_M))\boldsymbol{\Sigma}_{_W}(\boldsymbol{L}_{_W}\boldsymbol{U} - G_{_W})\end{aligned}$$

where the penalty matrices for the north, east, south and west boundary are indicated by $\boldsymbol{\Sigma}_{_{N,E,S,W}}$ while the discrete analogue of the boundary operator L in (5.5b) is denoted by $\boldsymbol{L}_{_{N,E,S,W}}$ for each boundary. Here, $G_{_{N,E,S,W}}$ are $3NM \times 1$ vectors denoting boundary data whose elements are given by a grid function $\mathbf{g}_k = \mathbf{g}(\partial\Omega_{ij})$ and similarly, initial data projected on the grid points is denoted by F. The discrete lifting operators $\{I_3 \otimes (I_N \otimes P_y^{-1} E_M)$, $I_3 \otimes (P_x^{-1} E_N \otimes I_M)$, $I_3 \otimes (I_N \otimes P_y^{-1} E_0)$, $I_3 \otimes (P_x^{-1} E_0 \otimes I_M)\}$, as defined before in (4.22), ensures that the penalty terms are confined to the appropriate boundary points.

The discrete inner product and the corresponding semi-norm mimicking (5.8) are

$$(\boldsymbol{U}, \boldsymbol{V})_{\tilde{\mathcal{P}}} = \boldsymbol{U}^T \tilde{\mathcal{P}} \boldsymbol{V}, \quad ||\boldsymbol{U}||^2_{\tilde{\mathcal{P}}} = (\boldsymbol{U}, \boldsymbol{U})_{\tilde{\mathcal{P}}}, \quad \tilde{\mathcal{P}} = \mathcal{P}\tilde{\boldsymbol{I}}.$$

5.2.1 The discrete energy method

Applying the discrete energy method to (5.37) yields

$$\frac{d}{dt}||U||^2_{\tilde{\boldsymbol{I}}} + \mathbf{conv} = \mathbf{diff} + \mathbf{pen} \tag{5.38}$$

where **conv**, **diff** and **pen** denotes the convective, diffusion and penalty terms respectively. In the interest of readability, we simplify the terms separately.

Considering the convective terms

$$\begin{aligned}\mathbf{conv} &= \frac{1}{2}\boldsymbol{U}^T\Big[(\boldsymbol{A}\mathcal{Q}_x + \mathcal{Q}_x\boldsymbol{A}) + (\mathcal{Q}_x^T\boldsymbol{A} + \boldsymbol{A}\mathcal{Q}_x^T) + (\boldsymbol{B}\mathcal{Q}_y + \mathcal{Q}_y\boldsymbol{B}) + (\mathcal{Q}_y^T\boldsymbol{B} + \boldsymbol{B}\mathcal{Q}_y^T)\Big]\boldsymbol{U} \qquad (5.39)\\ &= \frac{1}{2}\boldsymbol{U}^T\Big[\boldsymbol{A}(\mathcal{Q}_x + \mathcal{Q}_x^T) + (\mathcal{Q}_x + \mathcal{Q}_x^T)\boldsymbol{A} + \boldsymbol{B}(\mathcal{Q}_y + \mathcal{Q}_y^T) + (\mathcal{Q}_y + \mathcal{Q}_y^T)\boldsymbol{B}\Big]\boldsymbol{U}\\ &= \boldsymbol{U}^T\big[\mathcal{B}_x\boldsymbol{A} + \mathcal{B}_y\boldsymbol{B}\big]\boldsymbol{U}.\end{aligned}$$

In (5.39), we used the SBP properties $\mathcal{Q}_{x,y} + \mathcal{Q}_{x,y}^T = \mathcal{B}_{x,y}$. Next, we consider the diffusion terms

$$\begin{aligned}\mathbf{diff} &= \boldsymbol{U}^T\tilde{\boldsymbol{I}}\mathcal{Q}_y(\mathcal{D}_y\boldsymbol{U}) + (\mathcal{D}_y\boldsymbol{U})^T\mathcal{Q}_y^T\tilde{\boldsymbol{I}}\boldsymbol{U} \qquad (5.40)\\ &= \boldsymbol{U}^T\tilde{\boldsymbol{I}}(\mathcal{B}_y - \mathcal{Q}_y^T)(\mathcal{D}_y\boldsymbol{U}) + (\mathcal{D}_y\boldsymbol{U})^T(\mathcal{B}_y - \mathcal{Q}_y)\tilde{\boldsymbol{I}}\boldsymbol{U}\\ &= 2\boldsymbol{U}^T\tilde{\boldsymbol{I}}\mathcal{B}_y(\mathcal{D}_y\boldsymbol{U}) - \boldsymbol{U}^T\tilde{\boldsymbol{I}}\mathcal{Q}_y^T(\mathcal{D}_y\boldsymbol{U}) - (\mathcal{D}_y\boldsymbol{U})^T\mathcal{Q}_y\tilde{\boldsymbol{I}}\boldsymbol{U},\end{aligned}$$

and by inserting the unit matrices $\mathcal{P}^{-1}\mathcal{P}$ and $\mathcal{P}\mathcal{P}^{-1}$ on the indefinite terms in (5.40), **diff** becomes

$$\mathbf{diff} = 2\boldsymbol{U}^T\tilde{\boldsymbol{I}}\mathcal{B}_y(\mathcal{D}_y\boldsymbol{U}) - 2||\mathcal{D}_y\boldsymbol{U}||^2_{\tilde{\boldsymbol{I}}}. \tag{5.41}$$

The penalty term denoted by **pen** is

$$\mathbf{pen} = \boldsymbol{U}^T\mathcal{P}\mathcal{S} + (\boldsymbol{U}^T\mathcal{P}\mathcal{S})^T \tag{5.42}$$

where

$$\begin{aligned}\boldsymbol{U}^T\mathcal{P}\mathcal{S} = \boldsymbol{U}^T(I_3 \otimes (P_x \otimes E_M))\boldsymbol{\Sigma}_{_N}(\boldsymbol{L}_{_N}\boldsymbol{U} - G_{_N}) + \boldsymbol{U}^T(I_3 \otimes (E_N \otimes P_y))\boldsymbol{\Sigma}_{_E}(\boldsymbol{L}_{_E}\boldsymbol{U} - G_{_E}) \\ + \boldsymbol{U}^T(I_3 \otimes (P_x \otimes E_0))\boldsymbol{\Sigma}_{_S}(\boldsymbol{L}_{_S}\boldsymbol{U} - G_{_S}) + \boldsymbol{U}^T(I_3 \otimes (E_0 \otimes P_y))\boldsymbol{\Sigma}_{_W}(\boldsymbol{L}_{_W}\boldsymbol{U} - G_{_W}).\end{aligned}$$

The discrete line integrals operators acting on the lth ($l = N, E, S, W$) boundary mimicking the continuous $\oint_{\partial\Omega}(\cdot)ds$ are given by matrices

$$\mathcal{P}_{\partial\Omega_l} = \begin{cases} P_x \otimes E_M & \text{north boundary,} \\ E_N \otimes P_y & \text{east boundary,} \\ P_x \otimes E_0 & \text{south boundary,} \\ E_0 \otimes P_y & \text{west boundary.} \end{cases} \tag{5.43}$$

Next, we project the solution $\boldsymbol{U}$ onto the boundary and remove the contribution of the internal grid points using the projection matrix E_l

$$\boldsymbol{U}_l = E_l\,\boldsymbol{U}$$

where

$$E_l = \begin{cases} I_N \otimes E_M & \text{north boundary,} \\ E_N \otimes I_M & \text{east boundary,} \\ I_N \otimes E_0 & \text{south boundary,} \\ E_0 \otimes I_M & \text{west boundary.} \end{cases}$$

Remark 6. The size of $\boldsymbol{U}_l$ is $NM \times 1$ and contains nonzero elements only on the positions corresponding to the boundary points. Hence, the number of nonzero elements is M for the east, west boundary and N for the north, south boundary.

By substituting (5.39), (5.41) and (5.42) into (5.38), we obtain a discrete energy rate which is an analogue to (5.27)

$$\frac{d}{dt}||\,\boldsymbol{U}\,||^2_{\tilde{\mathcal{P}}} + 2||\mathcal{D}_y\,\boldsymbol{U}||^2_{\tilde{\mathcal{P}}} = \boldsymbol{BT}. \tag{5.44}$$

In the expression above, $\boldsymbol{BT} = BT_N + BT_E + BT_S + BT_W$ denotes the boundary terms, where

$$\begin{aligned}
BT_N =& - \boldsymbol{q}_N^T (I_4 \otimes \mathcal{P}_{\partial\Omega_N}) \boldsymbol{M}_N \boldsymbol{q}_N + \boldsymbol{U}_N^T (I_3 \otimes \mathcal{P}_{\partial\Omega_N}) \boldsymbol{\Sigma}_N (\boldsymbol{L}_N \boldsymbol{U}_N - G_N) \\
& + (\boldsymbol{U}_N^T (I_3 \otimes \mathcal{P}_{\partial\Omega_N}) \boldsymbol{\Sigma}_N (\boldsymbol{L}_N \boldsymbol{U}_N - G_N))^T, \\
BT_E =& - \boldsymbol{q}_E^T (I_4 \otimes \mathcal{P}_{\partial\Omega_E}) \boldsymbol{M}_E \boldsymbol{q}_E + \boldsymbol{U}_E^T (I_3 \otimes \mathcal{P}_{\partial\Omega_E}) \boldsymbol{\Sigma}_E (\boldsymbol{L}_E \boldsymbol{U}_E - G_E) \\
& + (\boldsymbol{U}_E^T (I_3 \otimes \mathcal{P}_{\partial\Omega_E}) \boldsymbol{\Sigma}_E (\boldsymbol{L}_E \boldsymbol{U}_E - G_E))^T, \\
BT_S =& - \boldsymbol{q}_S^T (I_4 \otimes \mathcal{P}_{\partial\Omega_S}) \boldsymbol{M}_S \boldsymbol{q}_S + \boldsymbol{U}_S^T (I_3 \otimes \mathcal{P}_{\partial\Omega_S}) \boldsymbol{\Sigma}_S (\boldsymbol{L}_S \boldsymbol{U}_S - G_S) \\
& + (\boldsymbol{U}_S^T (I_3 \otimes \mathcal{P}_{\partial\Omega_S}) \boldsymbol{\Sigma}_S (\boldsymbol{L}_S \boldsymbol{U}_S - G_S))^T, \\
BT_W =& - \boldsymbol{q}_W^T (I_4 \otimes \mathcal{P}_{\partial\Omega_W}) \boldsymbol{M}_W \boldsymbol{q}_W + \boldsymbol{U}_W^T (I_3 \otimes \mathcal{P}_{\partial\Omega_W}) \boldsymbol{\Sigma}_W (\boldsymbol{L}_W \boldsymbol{U}_W - G_W) \\
& + (\boldsymbol{U}_W^T (I_3 \otimes \mathcal{P}_{\partial\Omega_W}) \boldsymbol{\Sigma}_W (\boldsymbol{L}_W \boldsymbol{U}_W - G_W))^T.
\end{aligned}$$

Here, $\boldsymbol{q}_l$ and $\boldsymbol{M}_l$ ($l \in \{N, E, W, S\}$) are the discrete versions of q and M in (5.10) respectively and given by

$$\boldsymbol{q}_l = \begin{bmatrix} \mathbf{u}_l \\ \mathbf{v}_l \\ \mathbf{p}_l \\ \mathcal{D}_y \mathbf{u}_l \end{bmatrix}, \qquad \boldsymbol{M}_l = \begin{bmatrix} U_{\mathbf{n}_l} & \mathbf{0} & \alpha_l I_{NM} & -\beta_l I_{NM} \\ \mathbf{0} & \mathbf{0} & \beta_l I_{NM} & \mathbf{0} \\ \alpha_l I_{NM} & \beta_l I_{NM} & \mathbf{0} & \mathbf{0} \\ -\beta_l I_{NM} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}. \tag{5.45}$$

In (5.45), we introduced the normal velocity $U_{\mathbf{n}_l} = \alpha_l U_l + \beta_l V_l$, where $\mathbf{n}_l = (\alpha_l, \beta_l)$ is the outward point normal vector. Similar to remark 6, $\boldsymbol{BT}$ consists of NM decoupled terms, however, only the terms corresponding to the boundary points are nonzero.

Analogously to the continuous case (5.14), we decompose $\boldsymbol{M}_l$ in terms of the eigenvalue and associated eigenvector matrices.

Remark 7. The eigenvalues of $\boldsymbol{M}_l$ are given by the pointwise version of (5.13a) – (5.13d) for each l boundary. We denote them by $NM \times 1$ vectors and arrange them in an increase order as $\boldsymbol{\lambda}_1^l, \boldsymbol{\lambda}_2^l, \boldsymbol{\lambda}_3^l, \boldsymbol{\lambda}_4^l$.

Therefore, the discrete diagonalization of $\boldsymbol{M}_l$ using the eigenvalue matrix decomposition is given

$$\boldsymbol{M}_l = \boldsymbol{X}_l \boldsymbol{\Lambda}_l \boldsymbol{X}_l^T, \quad \boldsymbol{\Lambda}_l = \operatorname{diag}(\boldsymbol{\lambda}_1^l, \boldsymbol{\lambda}_2^l, \boldsymbol{\lambda}_3^l, \boldsymbol{\lambda}_4^l) \tag{5.46}$$

and $\boldsymbol{X}_l$ is the eigenvector matrix associated with the eigenvalue matrix $\boldsymbol{\Lambda}_l$. We organize $\boldsymbol{\Lambda}_l$ and $\boldsymbol{X}_l$ in terms of the positive, zero and negative components

$$\boldsymbol{\Lambda}_l = \operatorname{diag}(\boldsymbol{\Lambda}_l^+, \boldsymbol{\Lambda}_l^0, \boldsymbol{\Lambda}_l^-), \quad \boldsymbol{X}_l = [\boldsymbol{X}_l^+, \boldsymbol{X}_l^0, \boldsymbol{X}_l^-].$$

Here, $\boldsymbol{\Lambda}^+$ and $\boldsymbol{\Lambda}^-$ are $NM \times NM$ diagonal matrices containing non-negative and non-positive elements only on the respective boundary points. For each boundary, we have;

i. north boundary

$$\boldsymbol{\Lambda}_N^+ = \begin{bmatrix} I_{NM}/2 & \mathbf{0} \\ \mathbf{0} & \mathrm{diag}(\boldsymbol{\lambda}_4^N/(1+(\boldsymbol{\lambda}_4^N)^2) \end{bmatrix}, \quad \boldsymbol{X}_N^+ = \begin{bmatrix} \mathbf{0} & I_{NM} & I_{NM} & \mathbf{0} \\ \mathrm{diag}(\boldsymbol{\lambda}_4^N) & \mathbf{0} & \mathbf{0} & -I_{NM} \end{bmatrix}^T$$

$$\boldsymbol{\Lambda}_N^- = \begin{bmatrix} \mathrm{diag}(\boldsymbol{\lambda}_1^N/(1+(\boldsymbol{\lambda}_1^N)^2) & \mathbf{0} \\ \mathbf{0} & -I_{NM}/2 \end{bmatrix}, \quad \boldsymbol{X}_N^- = \begin{bmatrix} \mathrm{diag}(\boldsymbol{\lambda}_1^N) & \mathbf{0} & \mathbf{0} & -I_{NM} \\ \mathbf{0} & I_{NM} & -I_{NM} & \mathbf{0} \end{bmatrix}^T$$

ii. south boundary

$$\boldsymbol{\Lambda}_S^+ = \begin{bmatrix} I_{NM}/2 & \mathbf{0} \\ \mathbf{0} & \mathrm{diag}(\boldsymbol{\lambda}_4^S/(1+(\boldsymbol{\lambda}_4^S)^2) \end{bmatrix}, \quad \boldsymbol{X}_S^+ = \begin{bmatrix} \mathbf{0} & I_{NM} & I_{NM} & \mathbf{0} \\ \mathrm{diag}(\boldsymbol{\lambda}_4^S) & \mathbf{0} & \mathbf{0} & -I_{NM} \end{bmatrix}^T$$

$$\boldsymbol{\Lambda}_S^- = \begin{bmatrix} \mathrm{diag}(\boldsymbol{\lambda}_1^S/(1+(\boldsymbol{\lambda}_1^S)^2) & \mathbf{0} \\ \mathbf{0} & -I_{NM}/2 \end{bmatrix}, \quad \boldsymbol{X}_S^- = \begin{bmatrix} \mathrm{diag}(\boldsymbol{\lambda}_1^S) & \mathbf{0} & \mathbf{0} & I_{NM} \\ \mathbf{0} & I_{NM} & -I_{NM} & \mathbf{0} \end{bmatrix}^T$$

iii. east boundary

$$\boldsymbol{\Lambda}_E^+ = \mathrm{diag}(\boldsymbol{\lambda}_4^E/(1+(\boldsymbol{\lambda}_4^E)^2), \quad \boldsymbol{X}_E^+ = \begin{bmatrix} \mathrm{diag}(\boldsymbol{\lambda}_4^E) & \mathbf{0} & -I_{NM} & \mathbf{0} \end{bmatrix}^T$$

$$\boldsymbol{\Lambda}_E^- = \mathrm{diag}(\boldsymbol{\lambda}_1^E/(1+(\boldsymbol{\lambda}_1^E)^2), \quad \boldsymbol{X}_E^- = \begin{bmatrix} \mathrm{diag}(\boldsymbol{\lambda}_1^E) & \mathbf{0} & -I_{NM} & \mathbf{0} \end{bmatrix}^T$$

iv. west boundary

$$\boldsymbol{\Lambda}_W^+ = \mathrm{diag}(\boldsymbol{\lambda}_4^W/(1+(\boldsymbol{\lambda}_4^W)^2), \quad \boldsymbol{X}_W^+ = \begin{bmatrix} \mathrm{diag}(\boldsymbol{\lambda}_4^W) & \mathbf{0} & I_{NM} & \mathbf{0} \end{bmatrix}^T$$

$$\boldsymbol{\Lambda}_W^- = \mathrm{diag}(\boldsymbol{\lambda}_1^W/(1+(\boldsymbol{\lambda}_1^W)^2), \quad \boldsymbol{X}_W^- = \begin{bmatrix} \mathrm{diag}(\boldsymbol{\lambda}_1^W) & \mathbf{0} & I_{NM} & \mathbf{0} \end{bmatrix}^T.$$

Let $\boldsymbol{W} = \boldsymbol{X}^T \boldsymbol{q}$ be the discrete characteristic variables. We partition $\boldsymbol{W}$ into the positive and negative components

$$\boldsymbol{W} = \boldsymbol{W}^+ + \boldsymbol{W}^- = (\boldsymbol{X}^+)^T \boldsymbol{q} + (\boldsymbol{X}^-)^T \boldsymbol{q}, \quad \boldsymbol{W}^+ \cap \boldsymbol{W}^- = \emptyset$$

where $\boldsymbol{W}^+$ and $\boldsymbol{W}^-$ denotes the incoming and outgoing characteristics. The discrete version of the boundary conditions (5.18) can be written as

$$\boldsymbol{L}\boldsymbol{U} = \boldsymbol{W}_l^+ - \boldsymbol{R}_l \boldsymbol{W}_l^- = G_l \tag{5.47}$$

where $l \in \{N, E, S, W\}$ and $\boldsymbol{R} = R \otimes I_{NM}$ is a block-diagonal matrix mimicking R in (5.17). Moreover, the explicit form of $\boldsymbol{L}\boldsymbol{U}$ in terms of eigenvectors becomes

$$\begin{aligned} \boldsymbol{L}_l \boldsymbol{U}_l = (\boldsymbol{L}_l^+ - \boldsymbol{R}_l \boldsymbol{L}_l^-)\, \boldsymbol{U}_l &= \left[(\boldsymbol{X}_l^+)^T + \boldsymbol{R}_l (\boldsymbol{X}_l^-)^T \right] \boldsymbol{q}_l \\ &= \left[(\boldsymbol{X}_l^+)^T \boldsymbol{T}_l + \boldsymbol{R} (\boldsymbol{X}_l^-)^T \boldsymbol{T}_l \right] \boldsymbol{U}_l. \end{aligned} \tag{5.48}$$

Equation (5.48) implies that $\boldsymbol{L}_l^+ = (\boldsymbol{X}_l^+)^T \boldsymbol{T}_l$ and $\boldsymbol{L}_l^- = (\boldsymbol{X}_l^-)^T \boldsymbol{T}_l$ which corresponds to $L^\pm$ in the continuous setting (5.24). Here, $\boldsymbol{T}_l$ is a transformation relating $\boldsymbol{U}$ and $\boldsymbol{q}$

$$\boldsymbol{q}_l = \boldsymbol{T}_l \boldsymbol{U}_l, \quad \boldsymbol{T}_l = \begin{bmatrix} I_{NM} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & I_{NM} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & I_{NM} \\ \mathcal{D}_{y_l} & \mathbf{0} & \mathbf{0} \end{bmatrix}.$$

Remark 8. The size of $\boldsymbol{\Lambda}^+$ and $\boldsymbol{\Lambda}^-$ in (5.46) is $2NM \times 2NM$ for the north, south boundary and $NM \times NM$ for the east, west boundary. Similarly, the size of corresponding normalized eigenvector matrices $\boldsymbol{X}^+$ and $\boldsymbol{X}^-$ is $4NM \times 2NM$ for the north, south boundary and for the east, west boundary is $4NM \times NM$.

Remark 9. Consequently, using remark 8, the size of $\boldsymbol{W}^\pm$ for the north, south boundary is $2NM \times NM$ and $NM \times NM$ for the east, west boundary.

Next, we substitute (5.46) and (5.47) into (5.44). Therefore, $\boldsymbol{BT}$ becomes

$$\begin{aligned}
BT_N = &- \begin{bmatrix} \boldsymbol{W}_N^+ \\ \boldsymbol{W}_N^- \end{bmatrix}^T (I_4 \otimes \mathcal{P}_{\partial\Omega_N}) \begin{bmatrix} \boldsymbol{\Lambda}_N^+ & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_N^- \end{bmatrix} \begin{bmatrix} \boldsymbol{W}_N^+ \\ \boldsymbol{W}_N^- \end{bmatrix} \\
&+ \boldsymbol{U}_N^T (I_3 \otimes \mathcal{P}_{\partial\Omega_N}) \boldsymbol{\Sigma}_N (\boldsymbol{W}_N^- - \boldsymbol{R}_N \boldsymbol{W}_N^+ - G_N) \\
&+ \big(\boldsymbol{U}_N^T (I_3 \otimes \mathcal{P}_{\partial\Omega_N}) \boldsymbol{\Sigma}_N (\boldsymbol{W}_N^- - \boldsymbol{R}_N \boldsymbol{W}_N^+ - G_N)\big)^T, \\
BT_E = &- \begin{bmatrix} \boldsymbol{W}_E^+ \\ \boldsymbol{W}_E^- \end{bmatrix}^T (I_2 \otimes \mathcal{P}_{\partial\Omega_E}) \begin{bmatrix} \boldsymbol{\Lambda}_E^+ & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_E^- \end{bmatrix} \begin{bmatrix} \boldsymbol{W}_E^+ \\ \boldsymbol{W}_E^- \end{bmatrix} \\
&+ \boldsymbol{U}_E^T (I_3 \otimes \mathcal{P}_{\partial\Omega_E}) \boldsymbol{\Sigma}_E (\boldsymbol{W}_E^- - \boldsymbol{R}_E \boldsymbol{W}_E^+ - G_E) \\
&+ \big(\boldsymbol{U}_E^T (I_3 \otimes \mathcal{P}_{\partial\Omega_E}) \boldsymbol{\Sigma}_E (\boldsymbol{W}_E^- - \boldsymbol{R}_E \boldsymbol{W}_E^+ - G_E)\big)^T, \\
BT_S = &- \begin{bmatrix} \boldsymbol{W}_S^+ \\ \boldsymbol{W}_S^- \end{bmatrix}^T (I_4 \otimes \mathcal{P}_{\partial\Omega_S}) \begin{bmatrix} \boldsymbol{\Lambda}_S^+ & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_S^- \end{bmatrix} \begin{bmatrix} \boldsymbol{W}_S^+ \\ \boldsymbol{W}_S^- \end{bmatrix} \\
&+ \boldsymbol{U}_S^T (I_3 \otimes \mathcal{P}_{\partial\Omega_S}) \boldsymbol{\Sigma}_S (\boldsymbol{W}_S^- - \boldsymbol{R}_S \boldsymbol{W}_S^+ - G_S) \\
&+ \big(\boldsymbol{U}_S^T (I_3 \otimes \mathcal{P}_{\partial\Omega_S}) \boldsymbol{\Sigma}_S (\boldsymbol{W}_S^- - \boldsymbol{R}_S \boldsymbol{W}_S^+ - G_S)\big)^T, \\
BT_W = &- \begin{bmatrix} \boldsymbol{W}_W^+ \\ \boldsymbol{W}_W^- \end{bmatrix}^T (I_2 \otimes \mathcal{P}_{\partial\Omega_W}) \begin{bmatrix} \boldsymbol{\Lambda}_W^+ & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_E^- \end{bmatrix} \begin{bmatrix} \boldsymbol{W}_W^+ \\ \boldsymbol{W}_W^- \end{bmatrix} \\
&+ \boldsymbol{U}_W^T (I_3 \otimes \mathcal{P}_{\partial\Omega_E}) \boldsymbol{\Sigma}_W (\boldsymbol{W}_W^- - \boldsymbol{R}_W \boldsymbol{W}_W^+ - G_W) \\
&+ \big(\boldsymbol{U}_W^T (I_3 \otimes \mathcal{P}_{\partial\Omega_W}) \boldsymbol{\Sigma}_W (\boldsymbol{W}_W^- - \boldsymbol{R}_W \boldsymbol{W}_W^+ - G_W)\big)^T.
\end{aligned} \tag{5.49}$$

Remark 10. The boundary terms in (5.49) have a similar structure. Therefore, for simplicity of

presentation and minimizing the derivations, we limit our focus to the north boundary terms only. The other boundary terms can be treated in a similar way.

Let us consider the north boundary terms denoted by BT_N in (5.49). We seek the penalty coefficient matrix $\boldsymbol{\Sigma}_N$ and constraints on $\boldsymbol{R}_N$ such that BT_N bounds the energy rate (5.44). To achieve this, we repeat the continuous analysis and consider boundary conditions (5.47) with zero and nonzero data respectively.

In (5.29), we made the ansatz $U^T\Sigma = \mathbf{w}^{-T}\tilde{\Sigma}$ where $\tilde{\Sigma}$ could be chosen arbitrarily to rewrite the boundary terms in terms of the characteristics only. We proceeded to show (see Proposition 5.3) that the choice $\tilde{\Sigma} = \Lambda^-$ yields an estimate and subsequently determined Σ explicitly in (5.30). This choice also led to the estimate in the case of non-homogeneous boundary conditions (see Proposition 5.4). We imitate this ansatz discretely as

$$\begin{aligned}
\boldsymbol{U}^T(I_3 \otimes \mathcal{P}_{\partial\Omega_{l_1}})\boldsymbol{\Sigma}_{l_1} &= (\boldsymbol{W}_{l_1}^-)^T(I_2 \otimes \mathcal{P}_{\partial\Omega_{l_1}})\boldsymbol{\Lambda}_{l_1}^-, \quad l_1 \in \{N, S\} \\
\boldsymbol{U}^T(I_3 \otimes \mathcal{P}_{\partial\Omega_{l_2}})\boldsymbol{\Sigma}_{l_2} &= (\boldsymbol{W}_{l_2}^-)^T(I_1 \otimes \mathcal{P}_{\partial\Omega_{l_2}})\boldsymbol{\Lambda}_{l_2}^-, \quad l_2 \in \{E, W\}.
\end{aligned} \tag{5.50}$$

The BT_N term with ansatz (5.50) becomes

$$\begin{aligned}
BT_N = &- \begin{bmatrix} \boldsymbol{W}_N^+ \\ \boldsymbol{W}_N^- \end{bmatrix}^T (I_4 \otimes \mathcal{P}_{\partial\Omega_N}) \begin{bmatrix} \boldsymbol{\Lambda}_N^+ & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_N^- \end{bmatrix} \begin{bmatrix} \boldsymbol{W}_N^+ \\ \boldsymbol{W}_N^- \end{bmatrix} \\
&+ (\boldsymbol{W}_N^-)^T(I_2 \otimes \mathcal{P}_{\partial\Omega_N})\boldsymbol{\Lambda}_N^-(\boldsymbol{W}_N^- - \boldsymbol{R}_N\boldsymbol{W}_N^+ - G_N) \\
&+ \big((\boldsymbol{W}_N^-)^T(I_2 \otimes \mathcal{P}_{\partial\Omega_N})\boldsymbol{\Lambda}_N^-(\boldsymbol{W}_N^- - \boldsymbol{R}_N\boldsymbol{W}_N^+ - G_N)\big)^T.
\end{aligned} \tag{5.51}$$

Proposition 5.5. *The boundary term in (5.44) with zero data bounds the energy rate if $\boldsymbol{R}_l$ ($l \in \{N, E, S, W\}$) can be chosen such that the matrix*

$$\boldsymbol{\Lambda}_l^+ + \boldsymbol{R}_l^T\boldsymbol{\Lambda}_l^-\boldsymbol{R}_l \tag{5.52}$$

is positive semi-definite.

Proof. Consider (5.51) with $G_N = \mathbf{0}$

$$\begin{aligned}
BT_N = &- \begin{bmatrix} \boldsymbol{W}_N^+ \\ \boldsymbol{W}_N^- \end{bmatrix}^T (I_4 \otimes \mathcal{P}_{\partial\Omega_N}) \begin{bmatrix} \boldsymbol{\Lambda}_N^+ & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_N^- \end{bmatrix} \begin{bmatrix} \boldsymbol{W}_N^+ \\ \boldsymbol{W}_N^- \end{bmatrix} \\
&+ (\boldsymbol{W}_N^-)^T(I_2 \otimes \mathcal{P}_{\partial\Omega_N})\boldsymbol{\Lambda}_N^-(\boldsymbol{W}_N^- - \boldsymbol{R}_N\boldsymbol{W}_N^+) \\
&+ \big((\boldsymbol{W}_N^-)^T(I_2 \otimes \mathcal{P}_{\partial\Omega_N})\boldsymbol{\Lambda}_N^-(\boldsymbol{W}_N^- - \boldsymbol{R}_N\boldsymbol{W}_N^+)\big)^T. \\
= &- \begin{bmatrix} \boldsymbol{W}_N^+ \\ \boldsymbol{W}_N^- \end{bmatrix}^T (I_4 \otimes \mathcal{P}_{\partial\Omega_N}) \begin{bmatrix} \boldsymbol{\Lambda}_N^+ & \boldsymbol{R}_N^T\boldsymbol{\Lambda}_N^- \\ \boldsymbol{\Lambda}_N^-\boldsymbol{R}_N & -\boldsymbol{\Lambda}_N^- \end{bmatrix} \begin{bmatrix} \boldsymbol{W}_N^+ \\ \boldsymbol{W}_N^- \end{bmatrix}.
\end{aligned} \tag{5.53}$$

Next, we add and subtract $(\boldsymbol{W}_N^+)^T(I_2 \otimes \mathcal{P}_{\partial\Omega_N})\big[\boldsymbol{R}_N^T\boldsymbol{\Lambda}_N^-\boldsymbol{R}_N\big]\boldsymbol{W}_N^+$ on the RHS of (5.53) and

rearrange terms to obtain

$$
\begin{aligned}
BT_N = &\begin{bmatrix} \boldsymbol{W}_N^+ \\ \boldsymbol{W}_N^- \end{bmatrix}^T (I_4 \otimes \mathcal{P}_{\partial\Omega_N}) \begin{bmatrix} \boldsymbol{R}_N^T \boldsymbol{\Lambda}_N^- \boldsymbol{R}_N & -\boldsymbol{R}_N^T \boldsymbol{\Lambda}_N^- \\ -\boldsymbol{\Lambda}_N^- \boldsymbol{R}_N & \boldsymbol{\Lambda}_N^- \end{bmatrix} \begin{bmatrix} \boldsymbol{W}_N^+ \\ \boldsymbol{W}_N^- \end{bmatrix} \\
&- (\boldsymbol{W}_N^+)^T (I_2 \otimes \mathcal{P}_{\partial\Omega_N}) \big[\boldsymbol{\Lambda}_N^+ + \boldsymbol{R}_N^T \boldsymbol{\Lambda}_N^- \boldsymbol{R}_N\big] \boldsymbol{W}_N^+ \\
&= \big(\boldsymbol{W}_N^- - \boldsymbol{R}_N \boldsymbol{W}_N^+\big)^T (I_2 \otimes \mathcal{P}_{\partial\Omega_N}) \boldsymbol{\Lambda}_N^- \big(\boldsymbol{W}_N^- - \boldsymbol{R}_N \boldsymbol{W}_N^+\big) \\
&- (\boldsymbol{W}_N^+)^T (I_2 \otimes \mathcal{P}_{\partial\Omega_N}) \big[\boldsymbol{\Lambda}_N^+ + \boldsymbol{R}_N^T \boldsymbol{\Lambda}_N^- \boldsymbol{R}_N\big] \boldsymbol{W}_N^+.
\end{aligned} \tag{5.54}
$$

The first term on the RHS of (5.54) is negative since $\boldsymbol{\Lambda}_N^-$ contains pointwise negative eigenvalues , hence, the energy bound depends on the sign of the second term. Therefore, $BT_N \leq 0$ if (5.52) is positive semi-definite. Similarly, it can be shown that $BT_{E,S,W}$ bounds the energy rate (5.44). □

Remark 11. Temporal integration on the bounded time domain $[0, T]$ leads to an estimate

$$
||\boldsymbol{U}||_{\tilde{\mathcal{P}}}^2 \leq ||F||_{\tilde{\mathcal{P}}}^2,
$$

hence, the approximation (5.37) with zero data is stable.

Proposition 5.6. *The boundary term in* (5.44) *with nonzero data bounds the energy rate if*

i. $\boldsymbol{R}_l$ $(l \in \{N, E, S, W\})$ *is chosen such that* $\boldsymbol{\Lambda}_l^+ + \boldsymbol{R}_l^T \boldsymbol{\Lambda}_l^- \boldsymbol{R}_l$ *is positive definite,*

ii. there exist a positive semi-definite matrix $\mathcal{G}_l$ *such that* $(\boldsymbol{\Lambda}_l^- \boldsymbol{R}_l)(\boldsymbol{\Lambda}_l^+ + \boldsymbol{R}_l^T \boldsymbol{\Lambda}_l^- \boldsymbol{R}_l)^{-1} (\boldsymbol{\Lambda}_l^- \boldsymbol{R}_l)^T - \boldsymbol{\Lambda}_l^- \leq \mathcal{G}_l < \infty$ *holds.*

Proof. For brevity, we consider the north boundary terms BT_N only. Equation (5.51) with nonzero data can be written as

$$
BT_N = -\begin{bmatrix} \boldsymbol{W}_N^+ \\ \boldsymbol{W}_N^- \\ G_N \end{bmatrix}^T (I_6 \otimes \mathcal{P}_{\partial\Omega_N}) \begin{bmatrix} \boldsymbol{\Lambda}_N^+ & \boldsymbol{R}_N^T \boldsymbol{\Lambda}_N^- & \boldsymbol{0} \\ \boldsymbol{\Lambda}_N^- \boldsymbol{R}_N & -\boldsymbol{\Lambda}_N^- & \boldsymbol{\Lambda}_N^- \\ \boldsymbol{0} & \boldsymbol{\Lambda}_N^- \boldsymbol{0} & \end{bmatrix} \begin{bmatrix} \boldsymbol{W}_N^+ \\ \boldsymbol{W}_N^- \\ G_N \end{bmatrix} \tag{5.55}
$$

which is analogue to (5.31). Therefore, we employ a similar matrix splitting technique (5.32)

$$
\begin{aligned}
\begin{bmatrix} \boldsymbol{\Lambda}_N^+ & \boldsymbol{R}_N^T \boldsymbol{\Lambda}_N^- & \boldsymbol{0} \\ \boldsymbol{\Lambda}_N^- \boldsymbol{R}_N & -\boldsymbol{\Lambda}_N^- & \boldsymbol{\Lambda}_N^- \\ \boldsymbol{0} & \boldsymbol{\Lambda}_N^- & \boldsymbol{0} \end{bmatrix} &= \begin{bmatrix} \boldsymbol{\Lambda}_N^+ + \boldsymbol{R}_N^T \boldsymbol{\Lambda}_N^- \boldsymbol{R}_N & \boldsymbol{0} & \boldsymbol{R}_N^T \Lambda^- \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{\Lambda}_N^- \boldsymbol{R}_N & \boldsymbol{0} & \boldsymbol{\Lambda}_N^- + \mathcal{G}_N \end{bmatrix} + \begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & -\mathcal{G}_N \end{bmatrix} \\
&+ \begin{bmatrix} -\boldsymbol{R}_N^T \boldsymbol{\Lambda}_N^- \boldsymbol{R}_N & \boldsymbol{R}_N^T \boldsymbol{\Lambda}_N^- & -\boldsymbol{R}_N^T \boldsymbol{\Lambda}_N^- \\ \boldsymbol{\Lambda}_N^- \boldsymbol{R}_N & -\boldsymbol{\Lambda}_N^- & \boldsymbol{\Lambda}_N^- \\ -\boldsymbol{\Lambda}_N^- \boldsymbol{R}_N & \boldsymbol{\Lambda}_N^- & -\boldsymbol{\Lambda}_N^- \end{bmatrix}.
\end{aligned}
$$

By substituting the results above into (5.55), we obtain

$$
\begin{aligned}
BT_N = &- \begin{bmatrix} \boldsymbol{W}_N^+ \\ G_N \end{bmatrix}^T (I_4 \otimes \mathcal{P}_{\partial\Omega_N}) \underbrace{\begin{bmatrix} \boldsymbol{\Lambda}_N^+ + \boldsymbol{R}_N^T \boldsymbol{\Lambda}_N^- \boldsymbol{R}_N & \boldsymbol{R}_N^T \boldsymbol{\Lambda}_N^- \\ \boldsymbol{\Lambda}_N^- \boldsymbol{R}_N & \boldsymbol{\Lambda}_N^- + \mathcal{G}_N \end{bmatrix}}_{\mathcal{C}_1} \begin{bmatrix} \boldsymbol{W}_N^+ \\ G_N \end{bmatrix} \\
&+ G_N^T (I_2 \otimes \mathcal{P}_{\partial\Omega_N}) \mathcal{G}_N G_N \\
&- \begin{bmatrix} \boldsymbol{W}_N^+ \\ \boldsymbol{W}_N^- \\ G_N \end{bmatrix}^T (I_6 \otimes \mathcal{P}_{\partial\Omega_N}) \underbrace{\begin{bmatrix} -\boldsymbol{R}_N^T \boldsymbol{\Lambda}_N^- \boldsymbol{R}_N & \boldsymbol{R}_N^T \boldsymbol{\Lambda}_N^- & -\boldsymbol{R}_N^T \boldsymbol{\Lambda}_N^- \\ \boldsymbol{\Lambda}_N^- \boldsymbol{R}_N & -\boldsymbol{\Lambda}_N^- & \boldsymbol{\Lambda}_N^- \\ -\boldsymbol{\Lambda}_N^- \boldsymbol{R}_N & \boldsymbol{\Lambda}_N^- & -\boldsymbol{\Lambda}_N^- \end{bmatrix}}_{\mathcal{C}_2} \begin{bmatrix} \boldsymbol{W}_N^+ \\ \boldsymbol{W}_N^- \\ G_N \end{bmatrix}
\end{aligned} \tag{5.56}
$$

which mimics (5.33) discretely. Also here, in order to bound BT_N, we require the matrices $\mathcal{C}_1$ and $\mathcal{C}_2$ to be positive semi-definite and matrix $\mathcal{G}$ to be finite such that the data term $G_N^T(I_2 \otimes \mathcal{P}_{\partial\Omega_N})\mathcal{G}G_N$ is bounded.

Let's consider the matrix $\mathcal{C}_1$ in (5.56). To determine the signs of the eigenvalues of $\mathcal{C}_1$, we rotate it into a diagonal form using the rotation technique in the proof of Proposition 5.2. By multiplying $\mathcal{C}_1$ on the right with a non-singular upper triangular matrix and on the left with its transpose, we obtain a new matrix $\boldsymbol{X}$

$$
\boldsymbol{X} = \begin{bmatrix} I & \boldsymbol{\xi} \\ \boldsymbol{0} & I \end{bmatrix}^T \begin{bmatrix} \boldsymbol{\Lambda}_N^+ + \boldsymbol{R}_N^T \boldsymbol{\Lambda}_N^- \boldsymbol{R}_N & \boldsymbol{R}_N^T \Lambda_N^- \\ \boldsymbol{\Lambda}_N^- \boldsymbol{R}_N & \boldsymbol{\Lambda}_N^- + \mathcal{G}_N \end{bmatrix} \begin{bmatrix} I & \boldsymbol{\xi} \\ \boldsymbol{0} & I \end{bmatrix}.
$$

The choice $\boldsymbol{\xi} = -(\boldsymbol{\Lambda}_N^+ + \boldsymbol{R}_N^T \boldsymbol{\Lambda}_N^- \boldsymbol{R}_N)^{-1} (\boldsymbol{\Lambda}_N^- \boldsymbol{R}_N)^T$ transforms $\boldsymbol{X}$ into a block-diagonal matrix

$$
\boldsymbol{X} = \begin{bmatrix} \boldsymbol{\Lambda}_N^+ + \boldsymbol{R}_N^T \boldsymbol{\Lambda}_N^- \boldsymbol{R}_N & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Lambda}_N^- - (\boldsymbol{\Lambda}_N^- \boldsymbol{R}_N)(\boldsymbol{\Lambda}_N^+ + \boldsymbol{R}_N^T \boldsymbol{\Lambda}_N^- \boldsymbol{R}_N)^{-1} (\boldsymbol{\Lambda}_N^- \boldsymbol{R}_N)^T + \mathcal{G}_N \end{bmatrix}
$$

and it follows immediately that $\boldsymbol{X}_{11} = \boldsymbol{\Lambda}_N^+ + \boldsymbol{R}_N^T \boldsymbol{\Lambda}_N^- \boldsymbol{R}_N > 0$ is required in order for the inverse matrix in $\boldsymbol{\xi}$ to exist. The matrix $\boldsymbol{X}$ is positive semi-definite if the Schur complement $\boldsymbol{X}_{22}$ is positive semi-definite i.e.

$$
\boldsymbol{\Lambda}_N^- - (\boldsymbol{\Lambda}_N^- \boldsymbol{R}_N)(\boldsymbol{\Lambda}_N^+ + \boldsymbol{R}_N^T \boldsymbol{\Lambda}_N^- \boldsymbol{R}_N)^{-1} (\boldsymbol{\Lambda}_N^- \boldsymbol{R}_N)^T + \mathcal{G}_N \geq 0 \tag{5.57}
$$

which implies

$$
\mathcal{G}_N \geq (\boldsymbol{\Lambda}_N^- \boldsymbol{R}_N)(\boldsymbol{\Lambda}_N^+ + \boldsymbol{R}_N^T \boldsymbol{\Lambda}_N^- \boldsymbol{R}_N)^{-1} (\boldsymbol{\Lambda}_N^- \boldsymbol{R}_N)^T - \boldsymbol{\Lambda}_N^-.
$$

As before, the signs of the eigenvalues of $\boldsymbol{X}$ coincides with those of $\mathcal{C}_1$. Hence, $\mathcal{C}_1$ is positive semi-definite if the inequality (5.57) holds and lastly, we choose $\mathcal{G}_N < \infty$ in order to bound the second term in (5.56).

Next, we show that the matrix $\mathcal{C}_2$ in (5.56) is positive semi-definite. Since $\mathcal{C}_2$ is a discrete version of (5.34), we use a similar matrix decomposition. Consider the following factorization

$$\mathcal{C}_2 = \mathcal{A}^T(Z \otimes \mathbf{\Lambda}_N^-)\mathcal{A}, \quad \text{where} \quad \mathcal{A} = \text{diag}(\boldsymbol{R}_N, I_{2NM}, I_{2NM})$$

where the symmetric matrix Z is given in (5.34). By using the eigenvalue decomposition of Z derived in (5.35), the block-matrix $Z \otimes \mathbf{\Lambda}_N^-$ can be written as

$$Z \otimes \mathbf{\Lambda}_N^- = (W\Lambda_Z W^T) \otimes (I_{2NM}\mathbf{\Lambda}_N^- I_{2NM}) = (W \otimes I_{2NM})(\Lambda_Z \otimes \mathbf{\Lambda}_N^-)(W \otimes I_{2NM})^T$$

and notice that

$$\Lambda_Z \otimes \mathbf{\Lambda}_N^- = \text{diag}(-3, 0, 0) \otimes \mathbf{\Lambda}_N^- = \text{diag}(-3\mathbf{\Lambda}_N^-, \mathbf{0}, \mathbf{0})$$

is non-negative. Hence, matrix $\mathcal{C}_2$ is positive semi-definite and BT_N in (5.56) is bounded by data .i.e $BT_N \leq G_N^T(I_2 \otimes \mathcal{P}_{\partial\Omega_N})\mathcal{G}_N G_N < \infty$.

Similarly, the east, south and west boundary terms are bounded by data. Therefore, $\boldsymbol{BT}$ in (5.44) becomes

$$\begin{aligned}\boldsymbol{BT} \leq\ & G_N^T(I_2 \otimes \mathcal{P}_{\partial\Omega_N})\mathcal{G}_N G_N + G_E^T(\mathcal{P}_{\partial\Omega_E})\mathcal{G}_E G_E + G_S^T(I_2 \otimes \mathcal{P}_{\partial\Omega_S})\mathcal{G}_S G_S \\ & + G_W^T(\mathcal{P}_{\partial\Omega_W})\mathcal{G}_W G_W < \infty\end{aligned} \tag{5.58}$$

where $\mathcal{G}_l$ and $\boldsymbol{R}_l$ satisfies

$$(\mathbf{\Lambda}_l^- \boldsymbol{R}_l)(\mathbf{\Lambda}_l^+ + \boldsymbol{R}_l^T\mathbf{\Lambda}_l^-\boldsymbol{R}_l)^{-1}(\mathbf{\Lambda}_l^-\boldsymbol{R}_l)^T - \mathbf{\Lambda}_l^- \leq \mathcal{G}_l < \infty \quad \text{and} \quad \mathbf{\Lambda}_l^+ + \boldsymbol{R}_l^T\mathbf{\Lambda}_l^-\boldsymbol{R}_l > \mathbf{0}.$$

□

Proposition 5.7. *The approximation (5.37) is strongly stable.*

Proof. The temporal integration of (5.44) over finite time domain $[0, T]$ yields

$$||\boldsymbol{U}||_{\tilde{\mathcal{P}}}^2 \leq ||F||_{\tilde{\mathcal{P}}}^2 + \int_0^T \boldsymbol{BT}dt$$

where $\boldsymbol{BT}$ is given in (5.58). □

5.3 Numerical computations

The computational accuracy of the approximation (5.37) is evaluated in this section. To set up a test case, we consider problem (5.5a) with constant (linearized) coefficient matrices A and B posed on the domain $\Omega = [0, 1] \times [0, 1]$. Notice that this is such that the divergence relation in (5.4) is satisfied. The boundary operator L and the penalty coefficients for this test problem are linear. To validate the accuracy of the approximation, we employ the method

of manufactured solution (MMS) and second derivatives are computed using the wide-stencil SBP operators. The specific steady MMS considered is

$$\begin{aligned} u(x,y) &= x^2 cos(2\pi y), \\ v(x,y) &= -\frac{x}{\pi}\sin(2\pi y), \\ p(x,y) &= y^2 e^{-3x}. \end{aligned} \tag{5.59}$$

The forcing function and boundary data $\mathbf{g}$ are obtained by substituting solution (5.59) into (5.5a) and (5.5b) with the appropriate form of L given in (5.24) respectively. To compute the steady solution, an implicit formulation is employed.

Table 5.1: The convergence rates of the $u-$velocity are presented for different orders of accuracy of the SBP operators.

$N = M$	$q^{(2,1)}$	$q^{(4,2)}$	$q^{(6,3)}$	$q^{(8,4)}$
40	2.097	3.268	0.345	5.173
60	2.051	3.194	3.231	5.351
80	2.035	3.148	3.578	5.321
100	2.027	3.120	3.712	5.282
120	2.022	3.100	3.782	5.249

To compute the spatial derivatives in (5.5a), we use the $(2s, s)$ accurate finite difference operators on SBP form. Here, as before, $2s$ denotes the order of accuracy of the internal stencil and s denotes the accuracy near the boundaries. As outlined in Chapter 3, the expected global order of accuracy of the approximation (5.5a) is $s + 1$. The convergence rates $q^{(2s,s)}$ for the variables u, v and p are presented in Table. 5.1 - 5.3. Here, $q^{(2s,s)}$ is computed as in (4.11) where $s = 1, \ldots, 4$. The computational results shows that the convergence rates throughout corroborate the analysis.

Table 5.2: The convergence rates of the $v-$velocity are presented for different orders of accuracy of the SBP operators.

$N = M$	$q^{(2,1)}$	$q^{(4,2)}$	$q^{(6,3)}$	$q^{(8,4)}$
40	2.082	3.251	0.957	5.168
60	2.043	3.181	3.197	5.346
80	2.030	3.138	3.564	5.316
100	2.023	3.112	3.703	5.277
120	2.019	3.094	3.776	5.244

Table 5.3: The convergence rates of the pressure p are presented for different orders of accuracy of the SBP operators.

$N = M$	$q^{(2,1)}$	$q^{(4,2)}$	$q^{(6,3)}$	$q^{(8,4)}$
40	2.083	3.249	1.071	5.168
60	2.044	3.179	3.192	5.346
80	2.031	3.137	3.562	5.316
100	2.023	3.111	3.702	5.277
120	2.019	3.093	3.775	5.244

5.4 The Blasius solution

We consider the Blasius solution to the laminar boundary problem (5.4). The length of a flat plate is taken as $L = 10$ and the transverse length perpendicular to the flow is chosen such that the freestream boundary is outside the boundary layer. We further assume a uniform stream velocity U_∞ such that the pressure gradients vanishes. The Blasius solution is based on the *similarity* solution method which reduces PDE to ordinary differential equation (ODE). This is done by combining x and y to form a single dimensionless parameter called a similarity variable η. The steady solution of (5.4) with constant pressure field is

$$u = U_\infty f'(\eta), \quad v = \sqrt{\frac{U_\infty \nu}{2x}}\big(\eta f'(\eta) - f(\eta)\big) \tag{5.60}$$

where

$$\eta(x, y) = y\sqrt{\frac{U_\infty}{2\nu x}},$$

$f(\eta)$ and $f'(\eta)$ denotes stream function and its derivative respectively (see Appendix A).

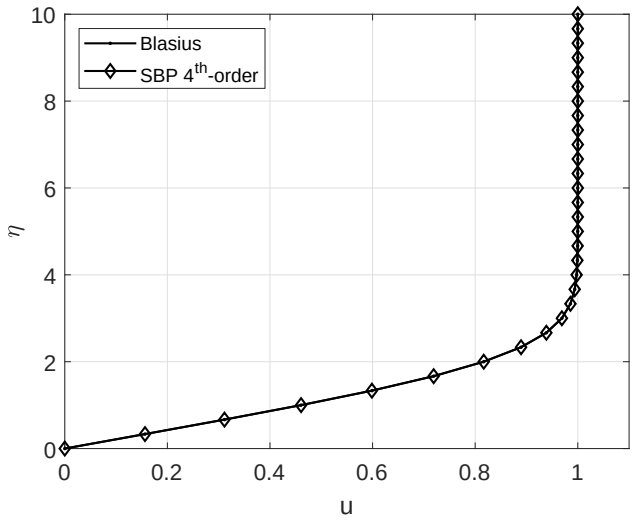

Figure 5.3: The streamwise velocity of the laminar flow past a flat-plate obtained by 4^{th} order accurate SBP-SAT scheme is compared with the Blasius similarity solution.

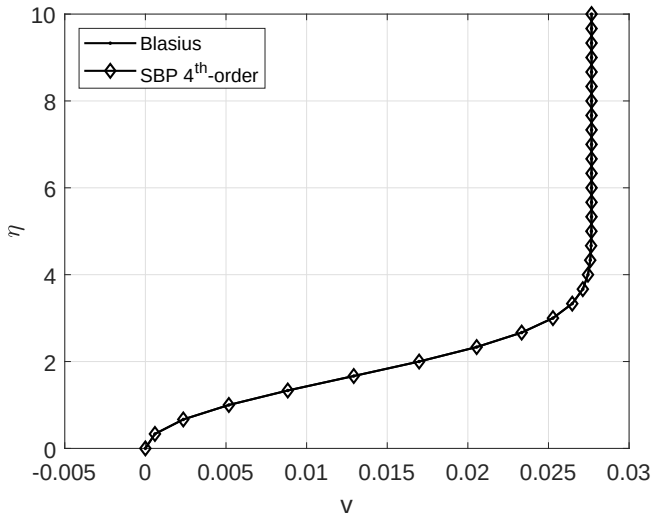

Figure 5.4: The transverse velocity of the laminar flow past a flat-plate obtained by 4^{th} order accurate SBP-SAT scheme is compared with the Blasius similarity solution.

To compute a numerical solution to (5.5a), we employ a regular quadrilateral mesh with the grids in $y-$direction stretched in order to capture the boundary layer. The approximation (5.37) is linearized by injecting Blasius solution (5.60) into the coefficient matrices $\boldsymbol{A}$, $\boldsymbol{B}$ as well as the penalty terms $\mathcal{S}$. The results are compared with the ODE solution presented in Table. 6.1. The comparative plots presented in Fig. 5.3 and Fig. 5.4 shows that the approximation scheme

(5.37) is stable. The accuracy of the high-order operators have been verified previously using the MMS (see Table. 5.1-5.3).

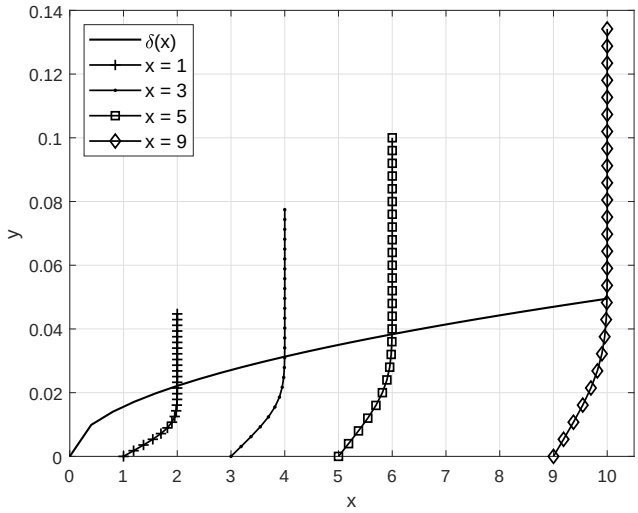

Figure 5.5: The streamwise velocity profile at different positions on the plate.

The boundary layer thickness $\delta(x)$ is defined as a distance from the solid wall boundary to a point in the flow where the flow's velocity reached approximately 99% of the stream velocity. By using the similarity variables η in (5.60), we obtain a relation

$$y = \eta\sqrt{\frac{2\nu x}{U_\infty}} \quad \Rightarrow \quad \delta(x) = \eta_{99\%}\sqrt{\frac{2\nu x}{U_\infty}}$$

and from Table. 6.1, we have that $u = 0.99$ when $\eta_{99\%} \approx 3.5$. Figure. 5.5 shows the evolution of the streamwise velocity profile along the plate. Notice that $\delta(x)$ is steep in the vicinity of the leading edge and grows asymptotically as the distance from the leading edge increases. The streamwise velocity profiles at each x position have similar shape. The approximation (5.37) works for the linearized case, the next step is to consider the nonlinear case and compute the approximate solution to the boundary layer equations.

Chapter 6

Conclusions and future work

This thesis was concerned with the development of a provably stable and high-order accurate numerical approximation for the nonlinear incompressible laminar boundary layer equations. By applying the energy method and Gauss's theorem to the continuous problem, a new set of energy stable boundary conditions were derived. These were obtained by diagonalizing the boundary matrix using eigenvalue-matrix decomposition. It has been shown that both the weak and strong implementation of these boundary conditions yields an energy estimate.

The corresponding discrete problem was formulated using the SBP-SAT technique in order to mimic the underlying properties of the PDEs. By mimicking the continuous analysis, the penalty coefficients were determined such that an energy estimate was obtained. In all the cases considered in this study, the discrete estimates mimicked precisely their continuous counterparts. The accuracy and stability of the SBP-SAT developed formulation was evaluated using the method of manufactured solutions and the design orders of accuracy coinciding with the theoretical ones were obtained. Hence, we conclude that the approximation is stable and high-accurate for the linearized problem.

The linear stability and accuracy of the developed scheme was also validated by solving the celebrated laminar flat plate flow problem. This was done by injecting the Blasius solution into the coefficient matrix as well as SAT boundary conditions and then solving for the primitives (velocity and pressure). The solved velocity and pressure matched the Blasius solution proving accuracy and linear stability.

The next step is to consider the fully nonlinear case and compute the numerical solution of the laminar boundary layer equations. The results will be compared with the Blasius similarity solution [10]. Once we can stably compute the solution to the laminar boundary layer equations to high order, the solution to the velocity-divergence formulation of the nonlinear incompressible Navier-Stokes equations will follow.

Bibliography

[1] L. Smith, O. Oxtoby, A. Malan, and J. Meyer, "An interactive boundary layer modelling methodology for aerodynamic flows," *International Journal of Numerical Methods for Heat and Fluid Flow*, vol. 23, no. 8, pp. 1373–1392, 2013.

[2] J. Marshall, A. Adcroft, C. Hill, L. Perelman, and C. Heisey, "A finite-volume, incompressible navier stokes model for, studies of the ocean on parallel computers," *Journal of Geophysical Research C: Oceans*, vol. 102, no. C3, pp. 5753–5766, 1997.

[3] A. Taneja and J. Higdon, "A fully-coupled discontinuous Galerkin spectral element method for two-phase flow in petroleum reservoirs," *Journal of Computational Physics*, vol. 352, pp. 341–372, 2018. [Online]. Available: https://doi.org/10.1016/j.jcp.2017.09.059

[4] A. W. Vreman, "The projection method for the incompressible Navier-Stokes equations: The pressure near a no-slip wall," *Journal of Computational Physics*, vol. 263, no. 2014, pp. 353–374, 2014. [Online]. Available: http://dx.doi.org/10.1016/j.jcp.2014.01.035

[5] J. Nordström, K. Mattsson, and C. Swanson, "Boundary conditions for a divergence free velocity-pressure formulation of the Navier-Stokes equations," *Journal of Computational Physics*, vol. 225, no. 1, pp. 874–890, 2007.

[6] R. Fernandez-Feria and E. Sanmiguel-Rojas, "An explicit projection method for solving incompressible flows driven by a pressure difference," *Computers and Fluids*, vol. 33, no. 3, pp. 463–483, 2004.

[7] A. J. Chorin, "A numerical method for solving incompressible viscous flow problems," *Journal of Computational Physics*, vol. 2, no. 1, pp. 12–26, 1967.

[8] A. G. Malan, R. W. Lewis, and P. Nithiarasu, "An improved unsteady, unstructured, artificial compressibility, finite volume scheme for viscous incompressible flows: Part I. Theory and implementation," *International Journal for Numerical Methods in Engineering*, vol. 54, no. 5, pp. 695–714, 2002.

[9] T. Ohwada and P. Asinari, "Artificial compressibility method revisited: Asymptotic numerical method for incompressible Navier-Stokes equations," *Journal of Computational Physics*, vol. 229, no. 5, pp. 1698–1723, 2010. [Online]. Available: http://dx.doi.org/10.1016/j.jcp.2009.11.003

[10] W. M. Frank, *Viscous Fluid Flow*. McGraw-Hill, 2006.

[11] H.-O. KREISS and G. SCHERER, "Finite Element and Finite Difference Methods for Hyperbolic Partial Differential Equations," in *Mathematical Aspects of Finite Elements in Partial Differential Equations*. Elsevier, 1974, pp. 195–212. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/B9780122083501500121

[12] B. Strand, "Summation by Parts for Finite Difference Approximations for d/dx," *Journal of Computational Physics*, vol. 110, no. 1, pp. 47–67, jan 1994.

[13] M. H. Carpenter, D. Gottlieb, and S. Abarbanel, "Time-Stable Boundary Conditions for Finite-Difference Schemes Solving Hyperbolic Systems: Methodology and Application to High-Order Compact Schemes," *Journal of Computational Physics*, vol. 111, no. 2, pp. 220–236, apr 1994. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0021999184710576

[14] J. Nordström, F. Ham, M. Shoeybi, E. van der Weide, M. Svärd, K. Mattsson, G. Iaccarino, and J. Gong, "A hybrid method for unsteady inviscid fluid flow," *Computers and Fluids*, vol. 38, no. 4, pp. 875–882, 2009. [Online]. Available: http://dx.doi.org/10.1016/j.compfluid.2008.09.010

[15] T. Lundquist, A. Malan, and J. Nordström, "A hybrid framework for coupling arbitrary summation-by-parts schemes on general meshes," *Journal of Computational Physics*, vol. 362, pp. 49–68, 2018. [Online]. Available: https://doi.org/10.1016/j.jcp.2018.02.018

[16] A. A. Ruggiu, P. Weinerfelt, and J. Nordström, "A new multigrid formulation for high order finite difference methods on summation-by-parts form," *Journal of Computational Physics*, vol. 359, pp. 216–238, 2018. [Online]. Available: https://doi.org/10.1016/j.jcp.2018.01.011

[17] J. E. Hicken and D. W. Zingg, "Parallel newton-krylov solver for the euler equations discretized using simultaneous-approximation terms," *AIAA Journal*, vol. 46, no. 11, pp. 2773–2786, 2008.

[18] J. Nordström and T. Lundquist, "Summation-by-parts in time," *Journal of Computational Physics*, vol. 251, pp. 487–499, oct 2013.

[19] B. Gustafsson, *High Order Difference Methods for Time Dependent PDE*, ser. Springer Series in Computational Mathematics. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, vol. 38. [Online]. Available: http://link.springer.com/10.1007/978-3-540-74993-6

[20] B. Gustafsson, H.-O. Kreiss, and J. Oliger, *Time-Dependent Problems and Difference Methods*. Hoboken, NJ, USA: John Wiley & Sons, Inc., sep 2013. [Online]. Available: http://doi.wiley.com/10.1002/9781118548448

[21] J. Sudirham, J. Van Der Vegt, and R. Van Damme, "A study on Discontinuous Gakerkin finite elements methods for elliptic problems," *Memorandum No. 1690, University of Twente, Faculty of EEMCS*, no. 1690, 2003.

[22] D. N. Arnold, F. Brezzi, B. Cockburn, and L. Donatella Marini, "Unified analysis of discontinuous Galerkin methods for elliptic problems," *SIAM Journal on Numerical Analysis*, vol. 39, no. 5, pp. 1749–1779, 2001.

[23] J. Nordström, "A Roadmap to Well Posed and Stable Problems in Computational Physics," *Journal of Scientific Computing*, vol. 71, no. 1, pp. 365–385, 2017.

[24] F. Ghasemi and J. Nordström, "An energy stable coupling procedure for the compressible and incompressible Navier-Stokes equations," *Journal of Computational Physics*, vol. 396, pp. 280–302, 2019. [Online]. Available: https://doi.org/10.1016/j.jcp.2019.07.022

[25] S. Ghader and J. Nordström, "Revisiting well-posed boundary conditions for the shallow water equations," *Dynamics of Atmospheres and Oceans*, vol. 66, pp. 1–9, 2014. [Online]. Available: http://dx.doi.org/10.1016/j.dynatmoce.2014.01.002

[26] R. Bodenmann, "Summation by Parts Formula for Noncentered Finite Differences Summation," *Serminar for applied mathematics, ETH Zürish, CH-8092 , Research report 95-07*, no. 95, 1995.

[27] K. Mattsson, "Diagonal-norm upwind SBP operators," *Journal of Computational Physics*, vol. 335, pp. 283–310, 2017. [Online]. Available: http://dx.doi.org/10.1016/j.jcp.2017.01.042

[28] M. Svärd and J. Nordström, "On the order of accuracy for difference approximations of initial-boundary value problems," *Journal of Computational Physics*, vol. 218, no. 1, pp. 333–352, 2006.

[29] J. Manzanero, G. Rubio, D. A. Kopriva, E. Ferrer, and E. Valero, "A free-energy stable nodal discontinuous Galerkin approximation with summation-by-parts property for the Cahn-Hilliard equation," *Journal of Computational Physics*, vol. 1, p. 109072, 2019. [Online]. Available: http://arxiv.org/abs/1902.08089

[30] J. Chan, "On discretely entropy conservative and entropy stable discontinuous Galerkin methods," *Journal of Computational Physics*, vol. 362, pp. 346–374, 2018. [Online]. Available: https://doi.org/10.1016/j.jcp.2018.02.033

[31] N. K. Yamaleev and M. H. Carpenter, "A family of fourth-order entropy stable nonoscillatory spectral collocation schemes for the 1-D Navier–Stokes equations," *Journal of Computational Physics*, vol. 331, pp. 90–107, 2017. [Online]. Available: http://dx.doi.org/10.1016/j.jcp.2016.11.039

[32] F. Ham, K. Mattsson, and G. Iaccarino, "Accurate and stable finite volume operators for unstructured flow solvers," *Center for Turbulence Research Annual Research Briefs*, pp. 243–261, 2006.

[33] J. Nordström, K. Forsberg, C. Adamsson, and P. Eliasson, "Finite volume methods, unstructured meshes and strict stability for hyperbolic problems," *Applied Numerical Mathematics*, vol. 45, no. 4, pp. 453–473, 2003.

[34] K. Mattsson and J. Nordström, "Summation by parts operators for finite difference approximations of second derivatives," *Journal of Computational Physics*, vol. 199, no. 2, pp. 503–540, 2004.

[35] K. Mattsson, "Diagonal-norm summation by parts operators for finite difference approximations of third and fourth derivatives," *Journal of Computational Physics*, vol. 274, pp. 432–454, oct 2014. [Online]. Available: http://dx.doi.org/10.1016/j.jcp.2014.06.027

[36] M. Svärd and J. Nordström, "Review of summation-by-parts schemes for initial-boundary-value problems," *Journal of Computational Physics*, vol. 268, pp. 17–38, 2014. [Online]. Available: http://dx.doi.org/10.1016/j.jcp.2014.02.031

[37] Q. Abbas and J. Nordström, "Weak Versus Strong No-Slip Boundary Conditions for the Navier-Stokes Equations," *Engineering Applications of Computational Fluid Mechanics*, vol. 4, no. 1, pp. 29–38, 2010.

[38] M. Svärd and J. Nordström, "A stable high-order finite difference scheme for the compressible Navier-Stokes equations. No-slip wall boundary conditions," *Journal of Computational Physics*, vol. 227, no. 10, pp. 4805–4824, 2008.

[39] J. Nordström, "Conservative finite difference formulations, variable coefficients, energy estimates and artificial dissipation," *Journal of Scientific Computing*, vol. 29, no. 3, pp. 375–404, 2006.

[40] C. H. Tai, J. C. Leong, and C. Y. Lin, "Effects of golf ball dimple configuration on aerodynamics, trajectory, and acoustics," *Journal of Flow Visualization and Image Processing*, vol. 14, no. 2, pp. 183–200, 2007.

[41] F. Alam, H. Chowdhury, H. Moria, R. L. Brooy, and A. Subic, "A comparative study of golf ball aerodynamics," *17th Australasian Fluid Mechanics Conference 2010*, no. December 2014, pp. 599–602, 2010.

[42] L. E. ERICSSON, "Effect of boundary-layer transition on vehicle dynamics," *Journal of Spacecraft and Rockets*, vol. 6, no. 12, pp. 1404–1409, dec 1969. [Online]. Available: https://arc.aiaa.org/doi/10.2514/3.29838

[43] D. Hlevca and M. Degeratu, "European Journal of Mechanics / B Fluids Atmospheric boundary layer modeling in a short wind tunnel," *European Journal of Mechanics / B Fluids*, vol. 79, pp. 367–375, 2020. [Online]. Available: https://doi.org/10.1016/j.euromechflu.2019.10.003

[44] A. Jafari, F. Ghanadi, M. Arjomandi, M. J. Emes, and B. S. Cazzolato, "Correlating turbulence intensity and length scale with the unsteady lift force on flat plates in an atmospheric boundary layer flow," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 189, no. April, pp. 218–230, 2019. [Online]. Available: https://doi.org/10.1016/j.jweia.2019.03.029

[45] J. Nordström and C. La Cognata, "Energy stable boundary conditions for the nonlinear incompressible Navier–Stokes equations," *Mathematics of Computation*, vol. 88, no. 316, pp. 665–690, aug 2018. [Online]. Available: http://www.ams.org/mcom/2019-88-316/S0025-5718-2018-03375-0/

[46] J. Nordström and M. Svärd, "Well-Posed Boundary Conditions for the Navier–Stokes Equations," *SIAM Journal on Numerical Analysis*, vol. 43, no. 3, pp. 1231–1255, jan 2005. [Online]. Available: http://epubs.siam.org/doi/10.1137/040604972

[47] S. N. Ha, "A nonlinear shooting method for two-point boundary value problems," *Computers and Mathematics with Applications*, vol. 42, no. 10-11, pp. 1411–1420, 2001.

Appendix A: Similarity solution for laminar boundary layer equations

As mentioned in Chapter 5, the pressure field is related to the freestream velocity U_∞ through the incompressible Bernoulli's theorem

$$p \approx -\frac{U_\infty^2}{2} \quad \Rightarrow \quad \frac{dp}{dx} \approx -U_\infty \frac{dU_\infty}{dx}.$$

Therefore, for constant U_∞, the pressure gradients in (5.4) vanishes and the steady flow problem reduces to

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0, \tag{6.1a}$$

$$u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} = \nu\frac{\partial^2 u}{\partial y^2} \tag{6.1b}$$

which satisfies the following boundary conditions

$$u = v = 0 \quad y = 0$$
$$u = U_\infty \quad y \to \infty.$$

The system above is a simplified form of the incompressible N-S equations specific to laminar flow past a flat plate Fig. 5.1. To date, explicit solutions $u(x, y)$ and $v(x, y)$ hasn't been found, however, Blasius (1904) proposed a method that reduces boundary layer equations (6.1a) – (6.1b) to ODE. This procedure is based on the *similarity* solution technique.

Let $\psi(x, y)$ be a stream function such that

$$u = \frac{\partial \psi}{\partial y}, \quad v = -\frac{\partial \psi}{\partial x} \tag{6.2}$$

holds. This choice automatically satisfies the divergence relation (6.1a) since $\psi_{xy} = \psi_{yx}$. The $x-$momentum equation becomes

$$\frac{\partial \psi}{\partial y}\frac{\partial^2 \psi}{\partial x \partial y} - \frac{\partial^2 \psi}{\partial y^2} = \nu\frac{\partial^3 \psi}{\partial y^3}. \tag{6.3}$$

The idea behind the similarity solution is to reduce the parameters in the problem by combining them. This approached is applicable to problems which lacks characteristic scales. Next, we consider similarity variable $\eta = \eta(x, y)$ and define ψ as

$$\eta = y\sqrt{\frac{U_\infty}{2\nu x}}, \quad \psi = \sqrt{2\nu U_\infty x} f(\eta) \tag{6.4}$$

where $f(\eta)$ is a function yet to be determined. The following derivatives are essential for the subsequent derivations

$$\begin{aligned} \frac{\partial \eta}{\partial x} &= -\frac{y}{2x}\sqrt{\frac{U_\infty}{2\nu x}} = -\frac{\eta y}{2x}, \\ \frac{\partial \eta}{\partial y} &= \sqrt{\frac{U_\infty}{2\nu x}}, \\ \frac{\partial \psi}{\partial \eta} &= \sqrt{2\nu U_\infty x} f'(\eta). \end{aligned} \tag{6.5}$$

Next, by substituting (6.4) into (6.2) and utilizing partial derivatives in (6.5), we rewrite u and v in terms of η

$$\begin{aligned} u &= \frac{\partial \psi}{\partial y} = \frac{\partial \psi}{\partial \eta}\frac{\partial \eta}{\partial y} = U_\infty f'(\eta), \\ v &= -\frac{\partial \psi}{\partial x} = -\Big[\frac{\partial}{\partial x}(\sqrt{2U_\infty \nu x}) f(\eta) + \sqrt{2U_\infty \nu x}\frac{\partial f}{\partial \eta}\frac{\partial \eta}{\partial x}\Big] \\ &= \sqrt{\frac{U_\infty \nu}{2x}}\big(\eta f'(\eta) - f(\eta)\big). \end{aligned} \tag{6.6}$$

Therefore from (6.6), the derivatives in (6.1b) are computed as

$$\begin{aligned} u_x &= \frac{\partial^2 \psi}{\partial x \partial y} = U_\infty \frac{\partial f'}{\partial \eta}\frac{\partial \eta}{\partial x} = -\frac{U_\infty \eta}{2x} f''(\eta), \\ u_y &= \frac{\partial^2 \psi}{\partial y^2} = U_\infty \frac{\partial f'}{\partial \eta}\frac{\partial \eta}{\partial y} = U_\infty \sqrt{\frac{U_\infty}{2\nu x}} f''(\eta), \\ u_{yy} &= \frac{\partial^3 \psi}{\partial y^3} = U_\infty \sqrt{\frac{U_\infty}{2\nu x}}\frac{\partial f'}{\partial \eta}\frac{\partial \eta}{\partial y} = \frac{U_\infty^2}{2\nu x} f'''(\eta). \end{aligned} \tag{6.7}$$

By substituting (6.6) and (6.7) into (6.3), we obtain third-order nonlinear ODE

$$f'''(\eta) + f(\eta) f''(\eta) = 0. \tag{6.8}$$

In order to solve (6.8), at least one boundary condition for each f, f' and f'' is required. Since $u = v = 0$ at $y = 0$, by using (6.6), we get

$$u(x, 0) = U_\infty f'(0) = 0 \quad \Rightarrow \quad f'(0) = 0,$$

similarly

$$u(x, y \to \infty) = U_\infty f'(\eta \to \infty) = U_\infty \quad \Rightarrow \quad f'(\eta \to \infty) = 1.$$

Subsequently, transverse velocity at $y = 0$ gives

$$v(x,0) = -\sqrt{\frac{U_\infty \nu}{2x}} f(0) = 0 \quad \Rightarrow \quad f(0) = 0.$$

To compute $f''(0)$, nonlinear *shooting* method [47] is employed. Let $r_1 = f$, $r_2 = f'$ and $r_3 = f''$, equation (6.8) can be written as a system of equations

$$\begin{aligned}
\frac{dr_1}{d\eta} &= f' = r_2, \\
\frac{dr_2}{d\eta} &= f'' = r_3 \\
\frac{dr_3}{d\eta} &= f''' = -ff'' = -r_1 r_3.
\end{aligned}$$

or compactly as

$$\frac{d\mathbf{r}}{d\eta} = g(\eta, r_1, r_2, r_3) \tag{6.9}$$

where $\mathbf{r} = (r_1, r_2, r_3)$ and $g = (r_2, r_3, -r_1 r_3)$. Therefore, by using forth-order explicit Runge-Kutta scheme and nonlinear shooting method to integrate (6.9), we obtain solution r_1, r_2 and r_3. We also find that $r_3(0) = f''(0) = 0.4696$. The computational results are presented in Fig. 6.1 and Table. 6.1.

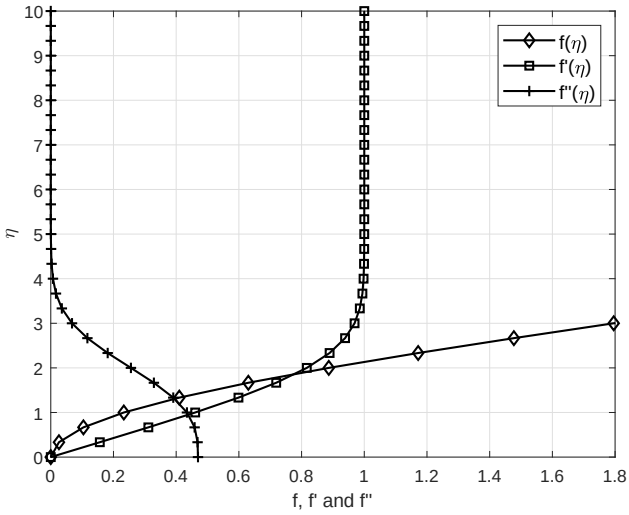

Figure 6.1: Blasius solution of the boundary layer equations.

Table 6.1: Blasius numerical solution of the flat-plate boundary layer equations.

$\eta = y\sqrt{\frac{U_\infty}{2\nu x}}$	$f(\eta)$	$f'(\eta) = \frac{u}{U_\infty}$	$f''(\eta)$
0	0.0000	0.0000	0.4696
0.2	0.0094	0.0939	0.4693
0.4	0.0376	0.1876	0.4673
0.6	0.0844	0.2806	0.4617
0.8	0.1497	0.3720	0.4512
1	0.2330	0.4606	0.4344
1.2	0.3337	0.5452	0.4106
1.4	0.4507	0.6244	0.3797
1.6	0.5830	0.6967	0.3425
1.8	0.7289	0.7611	0.3004
2	0.8868	0.8167	0.2557
2.2	1.0550	0.8633	0.2106
2.4	1.2315	0.9011	0.1676
2.6	1.4148	0.9306	0.1286
2.8	1.6033	0.9529	0.0951
3	1.7956	0.9691	0.0677
3.2	1.9906	0.9804	0.0464
3.4	2.1875	0.9880	0.0305
3.6	2.3856	0.9929	0.0193
3.8	2.5845	0.9959	0.0118
4	2.7839	0.9978	0.0069
4.2	2.9836	0.9988	0.0039
4.4	3.1834	0.9994	0.0021
4.6	3.3833	0.9997	0.0011
4.8	3.5833	0.9999	0.0005
5	3.7832	0.9999	0.0003
5.2	3.9832	1.0000	0.0001
5.4	4.1832	1.0000	0.0001