STABLE NUMERICAL METHODS FOR HYPERBOLIC PARTIAL DIFFERENTIAL EQUATIONS USING OVERLAPPING DOMAIN DECOMPOSITION

BY

ADAM HAROLD REICHERT

DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Computer Science in the Graduate College of the University of Illinois at Urbana-Champaign, 2011

Urbana, Illinois

Doctoral Committee:

Professor Michael T. Heath, Co-Chair and Co-Director of Research
Assistant Professor Daniel J. Bodony, Co-Chair and Co-Director of Research
Professor William Gropp
Doctor William Henshaw
Assistant Professor Luke Olson
Abstract

Overlapping domain decomposition methods, otherwise known as overset or chimera methods, are useful approaches for simplifying the discretizations of partial differential equations in or around complex geometries. While in wide use, the methods are prone to numerical instability unless numerical diffusion or some other form of regularization is used. This is especially true for higher-order methods. To address this, high-order, provably stable, overlapping domain decomposition methods are derived for hyperbolic initial-boundary-value problems. The overlap is treated by splitting the domain into pieces and using newly derived generalized summation-by-parts derivative operators and polynomial interpolation. Numerical regularization is not required for stability in the linear limit. Applications to linear and nonlinear problems in one and two dimensions are presented and new high-order generalized summation-by-parts derivative operators are derived.
To my wife Wen-Ting.
Acknowledgments

I would like to extend thanks to my co-advisers, Professor Daniel J. Bodony and Professor Michael T. Heath, for their time, valuable guidance, and financial support. I also want to thank the other members of my committee: Professor Luke Olson, Professor William Gropp, and Dr. William Henshaw.

I am who I am because of my family and their years of support. Thank you to my father, who always made sure that I had the best opportunities available, be it Stanford or a 4 × 5 camera. Thank you to my mother, who always pushed me to be the best and always made sure that my room was ready when I came home. Thank you to my brother Daniel, who read an early draft of the work and helped with the introduction. Finally, thank you to my beautiful wife Wen-Ting, whose love and encouragement made this possible.

This work was partially supported by the U.S. Department of Energy through the University of California under subcontract B523819.
Table of Contents

List of Figures ........................................... vii
List of Abbreviations ..................................... ix

Chapter 1 Introduction ..................................... 1

Chapter 2 Terminology and Problem Formulation ............ 5
  2.1 Overlapping Domains .................................. 5
  2.2 Composite Grids ...................................... 6
  2.3 Hyperbolic Equations .................................. 7
  2.4 Model Problems in One Dimension ..................... 8
    2.4.1 Right-Moving Advection .......................... 8
    2.4.2 Left-Moving Advection ............................ 9
    2.4.3 General Advection ................................ 9
    2.4.4 Hyperbolic System in One Dimension ............ 9
    2.4.5 System of Decoupled Advection Equations ...... 10
  2.5 Model Problems in Two Dimensions .................... 10
    2.5.1 Advection in Two Dimensions .................... 10
    2.5.2 Hyperbolic System in Two Dimensions ........... 11
  2.6 Approximate Solutions ................................ 11
  2.7 Error Measures ...................................... 12
  2.8 Convergence ......................................... 12
  2.9 Uniform Convergence ................................ 14
  2.10 Equivalent Norms ................................... 14
  2.11 Kronecker Product ................................. 14
  2.12 Reverse of a Matrix ............................... 16

Chapter 3 Consistency and Stability ...................... 18
  3.1 Consistency ......................................... 18
  3.2 Lax Stable ......................................... 18
  3.3 Energy Stability .................................... 19
  3.4 Strong Stability ................................... 20
  3.5 Strong Pointwise Stability ........................... 21

Chapter 4 Summation-by-Parts and Simultaneous Approximation Term .......... 23
  4.1 Summation-By-Parts ................................ 23
  4.2 Simultaneous Approximation Term ..................... 26
    4.2.1 SBP/SAT for Advection in One Dimension ....... 27
    4.2.2 SBP/SAT for Heat Equation in One Dimension ... 28
    4.2.3 GSBP/SAT for Advection in One Dimension .... 29
  4.3 GSBP Systems ....................................... 29
List of Figures

1.1 (a) Keyhole domain $\Omega$, with dark line around outside representing boundary $\partial \Omega$. (b) Overlapping domain decomposition comprising a square and a circle, with darker region representing overlap. (c) Non-overlapping decomposition. .......................................................... 1

2.1 (a) Original domain. (b) Domain decomposed into two smaller overlapping rectangles, with darker region representing overlap. (c) Overlapping composite grid associated with this decomposition. .......................................................... 5

2.2 Uniform grid in one dimension. ............................................................................ 6

2.3 Uniform grid in two dimensions. ............................................................................ 7

3.1 Relationships among stabilities. ............................................................................ 22

5.1 Interval $(a, b)$ discretized by two overlapping grids. ........................................... 32

5.2 Illustration of method (5.8) applied to right-moving advection problem (2.5). Downward pointing arrow indicates interpolation. .......................................................... 34

5.3 Composite grid used for testing scalar problems. Downward arrow indicates second-order interpolation $\mathcal{I}$. Upward arrow indicates injection $\tilde{\mathcal{I}}$. .......................................................... 51

5.4 (a) Approximate solution to problem (2.5) using method (5.8) at times $t = 0$, 0.5, and 1. At $t = 0.5$, blue squares give solution on left subdomain and red asterisks give solution on right. (b) Local convergence plot for problem (2.5) using method (5.8) with fourth-order Padé SBP pairs, showing errors $\|e^L\|_{R_L}$ and $\|e^R\|_{R_R}$ as functions of mesh spacings $h_L$ and $h_R$, respectively. (c) Same, for problem (2.6) using method (5.17) with same fourth-order Padé SBP pairs. (d) Spectrum of system matrix of method (5.8) ........................................................................ 54

5.5 (a) First component of approximate solution of method (5.31) when applied to problem (5.33). (b) Same, showing second component. (c) Same, showing third component. (d) Spectrum of system matrix of method (5.31). (e) Local convergence plot for problem (5.33) using method (5.31) with fourth-order Padé SBP pairs, showing errors $\|e^L\|_{R_L}$ and $\|e^R\|_{R_R}$ as functions of mesh spacings $h_{x,L}$ and $h_{x,R}$, respectively. ........................................................................ 55

5.6 (a) Right-moving approximate solution crossing interface. (b) Energy of simulation of right-moving advection problem (2.5) with method (5.8). (c) Energy of simulation of problem (5.33) with method (5.31). ........................................................................ 56

6.1 Illustration of method (6.1) applied to right-moving advection problem (2.5). Downward pointing arrow indicates interpolation. Leftmost node of right domain is always within fixed number of nodes of rightmost node of left domain. .......................................................... 57

6.2 Nonzero patterns of $Q_L + Q_L^T$ and $\mathcal{I}$ required by hypotheses of Theorem 6.1.1. Entries that may be nonzero are represented by colored blocks. ........................................................................ 59

6.3 (a) Local convergence plot for problem (2.5) using method (6.1) with new fourth-order Padé GSBP pairs, showing errors $\|e^L\|_{P_L}$ and $\|e^R\|_{P_R}$ as functions of mesh spacings $h_L$ and $h_R$, respectively. (b) Same, using new sixth-order Padé GSBP pairs. ........................................................................ 67
7.1 (a) Approximate solution to problem (7.1) at time \( t = 0.2 \) (b) Local convergence plot for problem (7.1) using nonlinear extension of method (6.1), showing errors \( \|e^L\|_{P_L} \) and \( \|e^R\|_{P_R} \) as functions of mesh spacings \( h_L \) and \( h_R \), respectively. .......................... 69

7.2 Numerical solution of linearized Euler equations (7.3) on overlapping domains. (a) Pressure at \( t = 0 \). (b) Pressure at \( t = 0.5 \). (c) Velocity in \( x \)-direction at \( t = 0.5 \). (d) Velocity in \( y \)-direction at \( t = 0.5 \). (e) Local convergence plot for problem (7.3) using method (5.31) with fourth-order Padé SBP pair, showing errors \( \|e^L\|_{R_L} \) and \( \|e^R\|_{R_R} \) as functions of mesh spacings \( h_{x,L} \) and \( h_{x,R} \), respectively. .......................... 73

7.3 Numerical solution of vortex problem (7.5) on overlapping domains. (a) Contour plot showing pressure before, during, and after vortex has crossed interface. (b) Same, showing vortex at interface. .......................... 74

7.4 Errors in numerical solution of vortex problem (7.5) on overlapping domains. (a) Scaled error \( \frac{p_{num} - p_{exact}}{s_p} \) vs. time at \( (-\frac{1}{2}, \frac{1}{2}) \). (b) Magnification of (a). (c) Scaled error vs. time at \( (\frac{1}{2}, \frac{1}{2}) \). (d) Scaled error vs. mesh spacing at \( (-\frac{1}{2}, \frac{1}{2}) \) and peak reflection time \( t \approx 0.8716 \). .......................... 75

8.1 (a) Approximate solution to problem (8.1) using method (8.3) at times \( t = 0, 0.5, \) and 1. (b) Local convergence plot for problem (8.1) using method (8.3) with fourth-order Padé SBP pairs, showing errors \( \|e^L\|_{R_L} \) and \( \|e^R\|_{R_R} \) as functions of mesh spacings \( h_L \) and \( h_R \), respectively. (c) Spectrum of system matrix of method (8.3) when diffusion constant \( \epsilon = 1 \). (d) Same, when \( \epsilon = 10^{-4} \) and \( \epsilon = 0 \). .......................... 78

8.2 (a) Approximate solution to problem (8.1) using method (8.8) at times \( t = 0, 0.5, \) and 1. (b) Local convergence plot for problem (8.1) using method (8.8) with GSBP matrices in Appendices B.3 and B.4, showing error \( \|e\|_{R} \) as function of mesh spacing \( h \). .......................... 80

8.3 (a) Approximate solution to problem (8.1) using method (8.9) at times \( t = 0, 0.5, \) and 1. (b) Local convergence plot for problem (8.1) using method (8.9) with second-order explicit GSBP pairs and new second-derivative operators, showing errors \( \|e^L\|_{R_L} \) and \( \|e^R\|_{R_R} \) as functions of mesh spacings \( h_L \) and \( h_R \), respectively. (c) Spectrum of system matrix of method (8.9) and diffusion constant \( \epsilon = 10^{-2} \). .......................... 82

B.1 (a) Modified wave number vs. wave number at boundary point \( x_{39} \). (b) Same at near-boundary point \( x_{39} \). .......................... 95

B.2 (a) Modified wave number vs. wave number at boundary point \( x_{39} \). (b) Same at near-boundary point \( x_{39} \). .......................... 97

B.3 (a) Modified wave number vs. wave number at boundary point \( x_{39} \). (b) Same at near-boundary point \( x_{39} \). .......................... 99

D.1 (a) Maximum eigenvalue of \( N_m \) with \( \alpha = \frac{1}{15} \) and \( \tau_R = 2 \) for GSBP4 matrices. (b) Upper boundaries for GSBP4 energy stability regions for five values of \( \theta \). .......................... 104
## List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSBP</td>
<td>Generalized Summation-by-Parts.</td>
</tr>
<tr>
<td>IBVP</td>
<td>Initial-Boundary-Value Problem.</td>
</tr>
<tr>
<td>SAT</td>
<td>Simultaneous Approximation Term.</td>
</tr>
<tr>
<td>SBP</td>
<td>Summation-by-Parts.</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

Domain decomposition methods offer an efficient way to solve initial-boundary value problems (IBVPs) for partial differential equations (PDEs) on general domains. In this approach, an IBVP is solved by splitting its domain into pieces, called subdomains, and then defining a new IBVP subproblem for each subdomain. The boundary conditions for each subproblem describe how that subproblem relates to its neighbors. An overlapping decomposition is one in which the subdomains have interior points in common. If none of the subdomains have overlapping interiors, then the decomposition is called non-overlapping. In this thesis, we describe new, provably stable, overlapping domain decomposition methods.

Introduced by Schwarz [77], the use of domain decomposition can be a powerful tool for solving IBVPs. Schwarz’s original use was to solve Laplace’s equation on a domain composed of an overlapping circle and square, like the keyhole domain in Figure 1.1. In order to construct a solution to this problem, Schwarz transformed Laplace’s equation on the keyhole domain into a sequence of problems that could be solved analytically, namely Laplace’s equation on the circle and on the square, in an iterative method now known as the alternating Schwarz method.

![Figure 1.1](image)

**Figure 1.1:** (a) Keyhole domain $\Omega$, with dark line around outside representing boundary $\partial \Omega$. (b) Overlapping domain decomposition comprising a square and a circle, with darker region representing overlap. (c) Non-overlapping decomposition.

Nearly a century later, domain decomposition was used to transform an IBVP into a set of interrelated IBVPs that could be stored and solved on a computer. Early non-overlapping domain decomposition methods, called substructuring methods, were used to solve elliptic problems arising from structural analysis...
Comprehensive theoretical results are available for overlapping domain decomposition methods for elliptic problems [13, 72, 91]. For example, there is a provably convergent method for numerically solving Laplace’s equation on polygons that have been decomposed into overlapping pieces [94, 95]. Others deal with elliptic equations on regions with curved boundaries [82] or consider finite element based approaches [44].

Overlapping domain decomposition methods have been applied to many time-dependent, nonlinear, fluid-dynamic problems (e.g., [14, 24, 47, 69, 73, 99]) and many aerodynamic and aeroacoustic problems (e.g., [5, 18, 38, 78, 79, 89, 100]). Their benefits include:

- **Efficient discretization and solution adaptation.** Due to the lack of a need to align internal boundaries, a global domain can be broken into pieces that can be discretized with simple grids. Generating grids efficiently is especially important in adaptive mesh refinement, where moving subdomains require that discretizations be updated frequently. Examples of adaptive overset methods include [32, 36, 56, 57, 76].

- **Ease of use.** There is a wealth of existing overlapping domain decomposition software. Software for generating decompositions includes CMPGRD [16, 67, 68], PEGASUS-5 [75] and Beggar [48]. Software for numerically simulating partial differential equations on overlapping subdomains include OVERFLOW [7] and INS3D [39]. The Overture framework [6, 31] combines both grid generation and numerical simulation in order to reduce development time. Many other techniques for grid and solution management have been developed (e.g., [8, 35, 43, 46, 65, 80, 97]).

- **Inherent parallelism.** Each subproblem can be assigned to a separate processor or processors. Parallelization has been investigated by many authors (e.g., [9, 33, 70, 98]).

Previous theoretical work on domain decomposition methods for time-dependent problems focused on enforcing two properties: stability and conservation. The amount of a conserved quantity modeled by a time dependent, nonlinear, fluid dynamic IBVP will increase or decrease as it is added, removed, or flows across boundaries. It is desirable that the numerical method behaves in a similar way. Such a method is called conservative. For example, if the original IBVP models a problem with a fixed amount of mass, then the “numerical mass” described by a conservative method will remain constant. Previous authors have successfully created conservative domain decomposition methods. Conservation can be enforced locally by balancing fluxes between non-overlapping subdomains. For example, in [73], the flow through a turbomachine comprising a rotor and stator is modeled. Conservation of mass is enforced between the non-overlapping rotor and the stator sections of the domain. Conservation inside the rotor and stator sections, each of which comprises overlapping subdomains, is not discussed. Later numerical methods enforced conservation...
of mass globally on overlapping domain decompositions. Global conservation can be enforced by solving a system of equations in order to determine interpolation coefficients [17] or by converting an overlapping decomposition into a non-overlapping one [96]. In [4], the authors develop a fully conservative scheme for hyperbolic problems with rectangular subdomains.

Unfortunately, conservative methods are not necessarily stable [66]. A method is stable if the growth of the approximate solution it describes is bounded. Stability of finite difference schemes has been heavily investigated, and many stable, non-overlapping domain decomposition methods have been developed. Most relevant to this thesis are the methods developed using the simultaneous approximation term (SAT) methodology in conjunction with summation-by-parts (SBP) derivative operators. Introduced in [42], this set of finite difference techniques and tools has been successfully used to create stable, non-overlapping domain decomposition methods for many problems, including the advection equation [40], constant-coefficient hyperbolic systems [49, 60], problems with diffusion [11, 12, 51], the second-order wave equation [50, 52], incompletely parabolic and parabolic systems [87], the Euler equations [53], and the Navier-Stokes equations on structured [58, 59, 61, 85, 88] and unstructured [23, 62] grids. SBP matrices were later extended to generalized summation-by-parts (GSBP) matrices [1, 2, 15], although these were not applied in the context of domain decomposition.

The non-overlapping proof techniques do not easily generalize to the overlapping case, and, even though overlapping methods are widely used, few proofs are available for their stability. Major theoretical results relating to the stability of finite difference methods are difficult to apply to overlapping domain decomposition methods. For example, the Courant-Friedrichs-Lewy condition [20] and the Godunov-Ryabenkii condition [26], both concerned with potential sources of instability, are necessary but not sufficient for stability. Necessary and sufficient conditions, such as Lax stability [37, 74] and Gustafsson-Kreiss-Sundström stability [28, 41, 92], are algebraically infeasible to apply to methods with complicated coefficients that can emerge in problems involving high-order interpolation.

The theoretical results describing the construction of stable overlapping domain decomposition methods for time dependent IBVPs are less complete. Overlapping numerical methods have typically been proven stable on a method-by-method basis. Stability can be enforced by the addition of an empirically determined amount of artificial dissipation [64]. An overlapping domain decomposition method for the specific case of the advection equation in one spatial dimension using the Lax-Wendroff method has been proven stable [83].

This thesis derives and proves the stability of new, overlapping domain decomposition methods for solving strictly hyperbolic IBVPs. Whereas other authors have enforced stability of overlapping domain decomposition methods by the addition of artificial dissipation or proven that specific, low-order schemes
are stable, we present stable methods for time-dependent hyperbolic IBVPs on overlapping domains that are both high-order and free of artificial dissipation. This is accomplished using the SAT methodology with SBP and GSBP matrices, previously proven stable only for single-domain and non-overlapping decompositions. High-order, provably stable methods without dissipation can be achieved by using high-order SBP and GSBP matrices. We proceed in the following way:

- *Chapter 2* defines our terminology.

- *Chapter 3* reviews the four forms of stability dealt with in this thesis: Lax, strong, strong pointwise, and energy-based.

- *Chapter 4* introduces summation-by-parts matrices, generalized summation-by-parts matrices, and the simultaneous approximation term methodology. These will be our primary tools in constructing stable methods.

- *Chapter 5* presents the first class of methods, Lax stable methods that are constructed from summation-by-parts parts operators. Under certain assumptions, these methods can also be shown to be strongly pointwise stable.

- *Chapter 6* presents the second class of methods, energy stable methods that are constructed from generalized summation-by-parts parts operators.

- *Chapter 7* presents numerical results utilizing both classes of methods.

- *Chapter 8* addresses problems with diffusion.
Chapter 2

Terminology and Problem Formulation

2.1 Overlapping Domains

A *domain* is an open and connected subset of $\mathbb{R}^d$, where $d$ is called the *spatial dimension*. A *decomposition* of a domain $\Omega$ is a finite set of domains $\{\Omega_i\}_{i=1}^{\ell}$ such that

$$\bar{\Omega} = \bigcup_{i=1}^{\ell} \bar{\Omega}_i,$$

where $\bar{\Omega}_i$ and $\bar{\Omega}$ are the closures of $\Omega_i$ and $\Omega$, respectively. The $\Omega_i$ are called *subdomains*. If there are $i$ and $j \neq i$ such that $\Omega_i \cap \Omega_j$ is non-empty, then the domain decomposition is described as *overlapping*. In this thesis, we frequently consider the case of two subdomains and adopt the convention that $\Omega_L = \Omega_1$ is on the left and the $\Omega_R = \Omega_2$ on the right, as in Figure 2.1.

![Figure 2.1](image)

*Figure 2.1:* (a) Original domain. (b) Domain decomposed into two smaller overlapping rectangles, with darker region representing overlap. (c) Overlapping composite grid associated with this decomposition.
2.2 Composite Grids

A composite grid $\Omega$ is a discrete representation of a domain decomposition, specifically a collection of grids $\{\hat{\Omega}_i\}_1^\ell$, one for each $\Omega_i$, such that $\hat{\Omega}_i$ is a set of points in $\bar{\Omega}_i$ and

$$\hat{\Omega} = \bigcup_{i=1}^\ell \hat{\Omega}_i.$$ 

Observe that $\hat{\Omega}_i$ can include points on the boundary of $\Omega_i$. If the domain decomposition is overlapping, we describe the composite grid as overlapping. Figure 2.1 shows a progression from a simple rectangular domain to an overlapping composite grid. We say that $\hat{\Omega}_i$ and $\hat{\Omega}$ are discretizations of $\Omega_i$ and $\Omega$, respectively.

We consider the case where $\hat{\Omega}_i$ are uniform grids. In one dimension, a uniform grid with $n$ points is the discretization of the open interval $(a, b) \subset \mathbb{R}$ comprising $n$ equally spaced nodes, where the first and last nodes are the end points of the interval,

$$h = (b - a) / (n - 1),$$
$$x_i = a + (i - 1)h, \quad i = 1, \ldots, n.$$ (2.1)

Figure 2.2: Uniform grid in one dimension.

A uniform grid in higher dimensions is defined as the Cartesian product of uniform grids in each dimension. The uniform grid discretizing $(a_x, b_x) \times (a_y, b_y)$ with $n_x$ nodes in the $x$-direction and $n_y$ nodes in the $y$-direction is defined by

$$h_x = (b_x - a_x) / (n_x - 1), \quad x_i = a_x + (i - 1)h_x,$$
$$h_y = (b_y - a_y) / (n_y - 1), \quad y_j = a_y + (j - 1)h_y.$$ (2.2)

In this case, the sequential order of the nodes is shown in Figure 2.3.
2.3 Hyperbolic Equations

Consider the following notation for a \( k \times k \), real symmetric matrix \( A \):

- \( \mathcal{N}_A \), the indicator of \( A \), is a \( k \times k \) matrix defined by
  \[
  \{\mathcal{N}_A\}_{i,j} = \begin{cases} 
  1 & \{A\}_{i,j} > 0 \\
  0 & \{A\}_{i,j} \leq 0
  \end{cases}.
  \]

- \( \Lambda_A = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_k) \), where \( \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_k \) are the (necessarily real) eigenvalues of \( A \).

- \( A^+ = TN\Lambda_A \Lambda_AT^T \) and \( A^- = A - A^+ \), where \( T \) is an orthogonal matrix that diagonalizes \( A \),
  \[
  A = T\Lambda_AT^T.
  \]

The matrices \( A^+ \) and \( A^- \) are uniquely defined: if \( T_1 \) and \( T_2 \) are two matrices that both satisfy (2.3), then the orthogonal matrix \((T_1^TT_2)\) commutes with \((\mathcal{N}_A, \Lambda_A)\).

The problems under consideration are constant-coefficient, hyperbolic IBVPs in \( d \) spatial dimensions with \( k \) unknowns. Let \( \Omega \subset \mathbb{R}^d \) and \( u : \Omega \times \mathbb{R} \to \mathbb{R}^k \). Let \( A_1, \ldots, A_d \) be \( k \times k \) real symmetric matrices. For a point \( z \) on \( \partial\Omega \), let \( n(z) \in \mathbb{R}^d \) be the outward pointing unit normal. Define \( B(z) \), the boundary matrix, as
  \[
  B(z) = n_1(z)A_1 + n_2(z)A_2 + \ldots + n_d(z)A_d,
  \]
where \( n(z) = [n_1(z), n_2(z), \ldots, n_d(z)] \).

The system of partial differential equations under consideration is

\[
\begin{align*}
  u_t &= \sum_{i=1}^{d} A_i u_{x_i}, & \forall t > 0, \ z \in \Omega, \\
  (B(z))^{+} u &= f(z, t) & \forall t > 0, \ z \in \partial \Omega, \\
  u(z, 0) &= g(z) & \forall z \in \Omega,
\end{align*}
\]

(2.4)

where \( u_{x_i} = \partial_{x_i} u, \ f(z) \) characterizes the boundary conditions, and \( g(z) \) gives the initial conditions. We always assume that \( \Omega, \ A_i, \ f, \) and \( g \) are such that the system is well-posed and has a suitably continuous solution, as discussed in [63].

IBVP (2.4) is called energy stable because a change in the norm of the solution can be bounded by an integral over the boundary,

\[
\begin{align*}
  \|u\|^2 &= \int_{\Omega} u(z, t)^{T} u(z, t) \, dz, \\
  \frac{d}{dt} \|u\|^2 &\leq \int_{\partial \Omega} f(z, t)^{T} C(z) f(z, t) \, dz,
\end{align*}
\]

where \( C(z) \) is the pseudo-inverse of \( B^{+}(z) \) defined by the Moore-Penrose conditions [22]. It follows that the existence of a solution implies the uniqueness that solution. Moreover, the distance between any two solutions with different initial values never increases.

## 2.4 Model Problems in One Dimension

We will consider several one-dimensional, constant-coefficient instances of problem (2.4) and list them here for future reference. In all cases, the domain \( \Omega = (a, b) \).

### 2.4.1 Right-Moving Advection

A unit-speed, right-moving wave is described by

\[
\begin{align*}
  u_t &= -u_x, \\
  u(x, 0) &= g(x), \\
  u(a, t) &= f(t).
\end{align*}
\]

(2.5)
2.4.2 Left-Moving Advection

A unit-speed, left-moving wave is described by

\[ u_t = u_x, \]
\[ u(x, 0) = g(x), \]
\[ u(b, t) = f(t). \]  (2.6)

2.4.3 General Advection

The general scalar advection problem in one dimension is

\[ u_t = \lambda u_x, \]
\[ u(x, 0) = g(x), \]
\[ |\lambda^-| u(a, t) = f_a(t), \]
\[ |\lambda^+| u(b, t) = f_b(t). \]  (2.7)

The values \(|\lambda^-|\) and \(|\lambda^+|\) “switch on” the appropriate boundary condition. The well-posedness of this problem is discussed in [21].

2.4.4 Hyperbolic System in One Dimension

The general form of (2.4) in one dimension is

\[ u_t = Au_x, \]
\[ u(x, 0) = g(x), \]
\[ -A^- u(a, t) = f_a(t), \]
\[ A^+ u(b, t) = f_b(t). \]  (2.8)

where \(A\) is a \(k \times k\) symmetric matrix.
2.4.5 System of Decoupled Advection Equations

A system of $k$ decoupled advection equations is described by

\begin{align*}
    u_t &= \Lambda u_x, \\
    u(x, 0) &= g(x), \\
    -\Lambda^- u(a, t) &= f_a(t), \\
    \Lambda^+ u(b, t) &= f_b(t),
\end{align*}

where $\Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_k)$.

2.5 Model Problems in Two Dimensions

We will consider two-dimensional, constant-coefficient instances of problem (2.4) and list them here for future reference. In all cases, the domain $\Omega = (a_x, b_x) \times (a_y, b_y)$.

2.5.1 Advection in Two Dimensions

A simple wave moving with velocity $(\lambda_x, \lambda_y)$ is described by

\begin{align*}
    u_t &= \lambda_x u_x + \lambda_y u_y, \\
    u(x, 0) &= g(x), \\
    |\lambda^-| u(a_x, t) &= f_{a_x}(t), \\
    |\lambda^+| u(b_x, t) &= f_{b_x}(t), \\
    |\lambda^-| u(a_y, t) &= f_{a_y}(t), \\
    |\lambda^+| u(b_y, t) &= f_{b_y}(t),
\end{align*}

where $u_x = \partial_x u$ and $u_y = \partial_y u$. 
2.5.2 Hyperbolic System in Two Dimensions

The statement of (2.4) in two dimensions for \( k \) unknowns is

\[
\begin{align*}
  u_t &= A_x u_x + A_y u_y, \\
  u(x,0) &= g(x), \\
  -A_x^+ u(a_x,t) &= f_{a_x}(y,t), \\
  A_y^+ u(b_x,t) &= f_{b_x}(y,t), \\
  -A_y^+ u(a_y,t) &= f_{a_y}(x,t), \\
  A_y^+ u(b_y,t) &= f_{b_y}(x,t),
\end{align*}
\]

(2.11)

where \( A_x \) and \( A_y \) are \( k \times k \) real symmetric matrices, \( f_{a_x}(t) = f(a_x,y) \), \( f_{b_x}(t) = f(b_x,y) \), \( f_{a_y}(x,t) = f(x,a_y,t) \), and \( f_{b_y}(x,t) = f(x,b_y,t) \).

2.6 Approximate Solutions

Let \( \hat{\Omega} = \{z_i\}_1^n \), where the \( z_i \) are points in \( \Omega \), be the set of all points in some grid, uniform or composite, discretizing \( \Omega \). An approximate solution of problem (2.4) is denoted by

\[
v(t) = \left[ v_1(t)^T, \ldots, v_n(t)^T \right]^T,
\]

where \( v_i(t) \approx u(z_i,t) \in \mathbb{R}^k \). Here the \( v_i \) are continuous functions of the continuous variable \( t \). Define the injection \( \hat{f} \) of a function \( f : \Omega \times \mathbb{R} \rightarrow \mathbb{R}^k \) by

\[
\hat{f}(t) = \left[ f(z_1,t)^T, \ldots, f(z_n,t)^T \right]^T.
\]

It is convenient to refer to approximate solution and injected true solution as

\[
\begin{bmatrix}
  v_L \\
  v_R
\end{bmatrix}
\]

\[
\begin{bmatrix}
  \hat{u}_L \\
  \hat{u}_R
\end{bmatrix},
\]

where the superscripts \( L \) and \( R \) refer to sets of values corresponding to left and right grids, respectively.

The time evolution of \( v \) is given by a system of linear ordinary differential equations obtained by the
method of lines [30],

\[ \dot{\mathbf{v}} = M \mathbf{v} + \mathbf{b}, \]
\[ \mathbf{v}(0) = \hat{g}. \] (2.12)

where \( \dot{\mathbf{v}} = \frac{d}{dt} \mathbf{v} \). The matrix \( M \), called the system matrix, represents a finite difference operator that approximates the differential operator \( \sum A_i \partial_{x_i} \) and the boundary operator \( B^+ \) in a form that will be made precise later. The vector \( \mathbf{b} \), called the inhomogeneous term, captures the effects of the boundary function \( f \) in (2.4).

### 2.7 Error Measures

We use two measures of error. The truncation error is given by

\[ t = \mathbf{u} - (M \mathbf{u} + \mathbf{b}). \] (2.13)

The cumulative error is defined by

\[ e = \mathbf{v} - \mathbf{u}. \]

Our goal is to make \( e \) small in norms described in the next section.

### 2.8 Convergence

Any real symmetric, positive definite matrix \( H \) induces an inner product and norm defined by \( \langle \mathbf{x}, \mathbf{y} \rangle_H = \mathbf{x}^T H \mathbf{y} \) and \( \| \mathbf{x} \|_H = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle_H} \), respectively, where \( \mathbf{x} \) and \( \mathbf{y} \) are vectors of the appropriate dimension. We measure the cumulative error in method (2.12) using the natural norm. Denoted by \( R \), this a discrete approximation to the \( L^2 \)-norm. Let \( I_m \) be the \( m \times m \) identity matrix. The uniform discretization (2.1) in one spatial dimension for a problem with \( k \) unknowns has the associated natural norm

\[ R = h I_{kn}. \]

The uniform discretization (2.2) in two spatial dimensions with \( k \) unknowns has the associated natural norm

\[ R = h_x h_y I_{kn_x n_y}. \]
Observe that, in both cases,
\[ \| \hat{u} \|_R^2 \approx \int_{\Omega} u^T u \, dx. \]

We solve problem (2.4) by generating a sequence of approximate solutions and show that the cumulative error goes to zero in the natural norm. Let \( \hat{\Omega}^1, \hat{\Omega}^2, \ldots \) be a sequence of grids discretizing \( \Omega \). Let \( n_i \) be the number of nodes in \( \hat{\Omega}^i \). Suppose that a numerical method produces a sequence of approximate solutions \( v^1, v^2, \ldots \), one for each discretization of the domain, then we say the method converges to the true solution \( u \) at fixed time \( t_f \) if
\[ \lim_{i \to \infty} \| e^i(t_f) \|_{R_i} = 0, \tag{2.14} \]
where \( R_i \in \mathbb{R}^{n_i \times n_i} \) is the natural norm associated with \( \hat{\Omega}^i \), and \( e^i(t_f) \) is the cumulative error on \( \hat{\Omega}^i \) at time \( t_f \). Henceforth, we cease using sub- and superscripts to specify a sequence of discretizations. We use an underset \( \hat{\Omega} \) to indicate a limit over discretizations. In this new notation, we write (2.14) as
\[ \lim_{\hat{\Omega}} \| e(t_f) \|_R = 0. \tag{2.15} \]

We say that a method converges with order \( p \) if
\[ \| e(t_f) \|_R = \mathcal{O}(h^p), \]
where \( h \) is the maximum spacing between adjacent nodes. This asymptotic bound should be interpreted in the limit that \( h \to 0 \).

Given a sequence of composite grids in which each comprises two uniform grids, as in Figure 2.1c, we say the method converges to the true solution \( u \) at time \( t_f \) if it converges in the sense of (2.15) on each uniform grid,
\[ \lim_{\hat{\Omega}} \| e^L(t_f) \|_{R_L} = 0, \quad \lim_{\hat{\Omega}} \| e^R(t_f) \|_{R_R} = 0, \tag{2.16} \]
where \( R_L \) is the natural norm for the left grid, and \( R_R \) is the natural norm for the right. We say that a method converges with order \( p \) on the left and order \( q \) on the right if
\[ \| e^L(t_f) \|_{R_L} = \mathcal{O}(h^p), \quad \| e^R(t_f) \|_{R_R} = \mathcal{O}(h^q), \]
where \( h \) the maximum spacing between adjacent points on either grid.
2.9 Uniform Convergence

A method of the form (2.12) is said to converge uniformly with order $p$ if

$$\|e\|_\infty = O(h^p),$$

where $\|x\|_\infty$ is the magnitude of the entry of $x$ of the largest magnitude and $h$ is the maximum spacing between adjacent nodes.

2.10 Equivalent Norms

Consider a sequence of uniform discretizations, each with corresponding natural norm $R$. A sequence of norms $H$ is said to be equivalent to the natural norm if there are $c_1 > 0$ and $c_2 > 0$, such that for all discretizations,

$$c_1 \|x\|_R^2 \leq \|x\|_H^2 \leq c_2 \|x\|_R^2 \quad \forall x.$$

(2.17)

We call $c_1$ and $c_2$ the lower and upper bound constants. Convergence in the natural norm and convergence in an equivalent norm have the same order of accuracy, as the next theorem states:

Theorem 2.10.1 Consider a sequence of uniform discretizations, each with a corresponding natural norm $R$. A method converges at time $t_f$ with order $p$ if and only if $\|e(t_f)\|_H = O(h^p)$, where $h$ is the maximum spacing between adjacent nodes.

The proof follows from norm equivalence. It can be difficult to show convergence in the natural norm directly. Rather, we will show that a method converges in some equivalent norm and apply Theorem 2.10.1.

2.11 Kronecker Product

The Kronecker product, denoted by $\otimes$, offers a convenient way to express finite difference operators for systems of PDEs in higher spatial dimensions in terms of operators for a scalar equation in one dimension. The Kronecker product of an $m \times n$ matrix $A$ and a $p \times q$ matrix $B$ is the $(mp) \times (nq)$ matrix defined by

$$A \otimes B = \begin{bmatrix} \{A\}_{1,1} B & \cdots & \{A\}_{1,n} B \\ \vdots & \ddots & \vdots \\ \{A\}_{m,1} B & \cdots & \{A\}_{m,n} B \end{bmatrix}.$$
More on the Kronecker product can be found in [22, p. 180]. Here we note that the Kronecker product satisfies the following properties, where $A$, $B$, $C$, and $D$ are any matrices of appropriate dimension:

- $(A \otimes B) \otimes C = A \otimes (B \otimes C)$. In this case, we write $A \otimes B \otimes C$.

- $A \otimes B$ is invertible if and only if $A$ and $B$ are invertible. In this case, $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$.

- $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$. This is often called the mixed-product property.

Let $\hat{\Omega}$ be the uniform discretization of $\Omega = (a, b) \times (a_y, b_y)$ defined by (2.1). Further, let $u : \Omega \times \mathbb{R} \rightarrow \mathbb{R}^k$, $A \in \mathbb{R}^{k \times k}$, and $D_x \in \mathbb{R}^{n \times n}$ be an finite difference approximation to the first spatial derivative $\partial_x$. A finite difference approximation to the operator $(A\partial_x)$ can be written concisely as

$$\hat{A}u_x \approx (D_x \otimes A) \hat{u}.$$  

If $\Omega = (a_x, b_x) \times (a_y, b_y)$ and $\hat{\Omega}$ is the uniform discretization (2.2), then

$$\hat{A}u_x \approx (D_x \otimes I_{n_y} \otimes A) \hat{u},$$

$$\hat{A}u_y \approx (I_{n_x} \otimes D_y \otimes A) \hat{u},$$

where $D_x \in \mathbb{R}^{n_x \times n_x}$ and $D_y \in \mathbb{R}^{n_y \times n_y}$ are finite difference approximations to $\partial_x$ and $\partial_y$, respectively.

Finally, we will implicitly use the following theorem:

**Theorem 2.11.1** Let $C$ be a matrix and $x$ be a vector such that the matrix-vector product $Cx$ makes sense. Let $y$ be a column vector of any dimension. Then the following equalities hold:

$$(C \otimes y)x = Cx \otimes y,$$

$$(y \otimes C)x = y \otimes Cx. \quad (2.18)$$

For any vector $x$,

$$x = x \otimes 1 = 1 \otimes x.$$  

The identities (2.18) follow from the mixed-product property,

$$(C \otimes y)x = (C \otimes y)(x \otimes 1) = Cx \otimes y,$$

$$(y \otimes C)x = (y \otimes C)(1 \otimes x) = y \otimes Cx.$$  

15
2.12 Reverse of a Matrix

For convenience, we introduce the following notation for matrices. Let $A$ be an $m \times n$ matrix. The matrix $A^\sharp$, called the reverse of $A$, is the $m \times n$ matrix containing the elements of $A$ permuted according to

$$\{A^\sharp\}_{i,j} = \{A\}_{m-i+1,n-j+1}.$$ 

For example,

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}^\sharp = \begin{bmatrix} 6 & 5 & 4 \\ 3 & 2 & 1 \end{bmatrix}.$$

The reversing operation satisfies the following properties, where $A$ and $B$ are any matrices:

- $(A^\sharp)^\sharp = A.$
- $(\alpha A + \beta B)^\sharp = \alpha A^\sharp + \beta B^\sharp.$
- $(AB)^\sharp = A^T B^\sharp.$
- If $A$ is invertible, then $A^\sharp$ is also invertible, with $(A^\sharp)^{-1} = (A^{-1})^\sharp.$ In this case we write $A^{-\sharp}$.

If $A = A^\sharp$, then $A$ is called reversible. If $A = -A^\sharp$, then $A$ is called antireversible. For the matrices in (4.4), $P$ is reversible and $Q$ is antireversible. Note that for any reversible square matrix $A$ and any vector of the same dimension $x$,

$$x^T A x = (x^\sharp)^T A (x^\sharp). \quad (2.19)$$

The reversing operation behaves nicely with respect to discrete derivative operators, as the next theorem shows:

**Theorem 2.12.1** If $D$ is a finite difference approximation to the derivative $\partial_x$ of approximation order $p$, then $-D^\sharp$ is an approximation to the derivative $\partial_x$ of the same approximation order.

Let $f(x)$ be a differentiable function. Define $g(x) = f(-x)$. Observe that

$$\hat{f} = \hat{g}^\sharp, \quad \hat{f}_x = -\hat{g}_x^\sharp.$$

Computing the truncation error associated with $-D^\sharp$ reveals that

$$-D^\sharp \hat{f} - \hat{f}_x = -D^\sharp \hat{g}^\sharp + \hat{g}_x^\sharp = - (D\hat{g} - \hat{g}_x)^\sharp.$$
Because the natural norm $R$ is reversible, it follows from (2.19) that

$$\| -D\hat{f} - \hat{f}_x \|_R = \| D\hat{g} - \hat{g}_x \|_R = O(h^\nu).$$

In Chapter 5, when using generalized summation-by-parts operators, we will apply the corollary:

**Corollary 2.12.2** Let $\{P, Q\}$ be a pair of matrices such that $P^{-1}Q$ is a finite difference approximation to the first spatial derivative $\partial_x$. Then $-P^{-2}Q^2$ is an approximation to $\partial_x$ of the same approximation order.
Chapter 3

Consistency and Stability

This section presents a review of stability assuming, for the moment, that $\Omega$ is discretized as a single domain. Again, consider a sequence of uniform grids.

3.1 Consistency

A sequence of ODEs (2.12) defined by system matrices $M$ and inhomogeneous terms $b$ is **consistent** if

$$\lim_{\hat{\Omega}} \max_{t \in [0, t_f]} \| t(t) \|_R = 0,$$

where $t$ is the truncation error (2.13). Further, if $h$ is the maximum distance between any two adjacent grid points, then the method is consistent with order $p$ if

$$\sup_{t \in [0, t_f]} \| t(t) \|_R = \mathcal{O}(h^p)$$

as $h \to 0$.

3.2 Lax Stable

The sequence of ODEs defined by system matrices $M$ is said to be **Lax stable** at fixed time $t_f$ if

$$\lim_{\hat{\Omega}} \max_{t \in [0, t_f]} \| e^{tM} \|_R < c(t_f) < \infty,$$

where $\| e^{tM} \|_R$ is the matrix norm induced by the natural norm applied to the matrix exponential. The **Lax equivalence theorem** relates consistency, stability, and convergence:

**Theorem 3.2.1 (Lax equivalence)** Method (2.12) converges to the solution of a well-posed IBVP if and only if it is consistent.
Demonstrating Lax stability by directly proving bound (3.1) is challenging. Instead, we will show stability by first proving convergence and then applying Lax equivalence.

3.3 Energy Stability

A sequence of ODEs (2.12) defined by system matrices $M$ is energy stable in the $H$-norm if there is a sequence of norms $H$, each corresponding to a discretization, such that the following properties hold:

(a) $H$ is equivalent to the natural norm: there are $c_1 > 0$ and $c_2 > 0$, such that for each discretization,

$$c_1 \|x\|_R^2 \leq \|x\|_H^2 \leq c_2 \|x\|_R^2 \quad \forall x,$$

where $R$ is the natural norm.

(b) $M$ is negative semidefinite in the $H$-norm,

$$x^T H M x \leq 0 \quad \forall x. \tag{3.3}$$

Note that (3.3) is equivalent to all of the eigenvalues of the real symmetric matrix $(HM + M^T H)$ being less than or equal to zero and implies that the eigenvalues of $M$ have non-positive real parts.

Energy stability is stronger than Lax stability, as the next theorem states:

**Theorem 3.3.1** An energy stable sequence of matrices $M$ is Lax stable for all $t_f > 0$.

To prove Theorem 3.3.1, assume that $M$ is energy stable in the $H$-norm and define

$$\Sigma = HM + M^T H.$$

Bound (3.3) implies that $\lambda_{\text{max}}(\Sigma)$, the maximum eigenvalue of $\Sigma$, is nonpositive. It is shown in [19] that the natural norm of the matrix exponential obeys

$$\|e^{tM}\|_R \leq \sqrt{\frac{c_2}{c_1}} e^{-\alpha t},$$

where $c_1$ and $c_2$ are the constants from bound (3.2) and $\alpha = \frac{\lambda_{\text{min}}(-H^{-1} \Sigma)}{2}$. Note that if $x$ is any vector of appropriate dimension, then

$$x^T \left( -H^{-\frac{1}{2}} \Sigma H^{-\frac{1}{2}} \right) x = - \left( H^{-\frac{1}{2}} x \right)^T \Sigma \left( H^{-\frac{1}{2}} x \right) \geq 0,$$
where the last inequality follows from Σ being a negative semidefinite matrix. Therefore, the spectrum of the symmetric matrix \((-H^{-\frac{1}{2}} \Sigma H^{-\frac{1}{2}})\) is nonnegative. Because \((H^{-1} \Omega)\) and \((-H^{-\frac{1}{2}} \Sigma H^{-\frac{1}{2}})\) are similar, they share a common spectrum and

\[
\alpha = \lambda_{\min} (-H^{-1} \Sigma) = \lambda_{\min} (-H^{-\frac{1}{2}} \Sigma H^{-\frac{1}{2}}) \geq 0.
\]

Therefore,

\[
\lim_{t \to \infty} \max_{\bar{\Omega}} \sup_{t \in [0, tf]} \| t(t) \|_H \leq \sqrt{\frac{c_2}{c_1}}.
\]

Because all energy stable approximations are Lax stable, it follows from the Lax Equivalence Theorem that a consistent and energy stable method converges. An important theorem relates energy stability, the order of the truncation error, and the rate of convergence:

**Theorem 3.3.2** A consistent and energy stable method converges with bound

\[
\| \hat{\mathbf{u}}(t_f) - \mathbf{v}(t_f) \|_R \leq \frac{t_f}{c_1} \left( \sup_{t \in [0, t_f]} \| \mathbf{t}(t) \|_H \right),
\]

where \(c_1\) is the constant from bound (3.2).

A proof can be found in [10]. Note that if \(M\) is consistent with order \(p\), then

\[
\| \hat{\mathbf{u}}(t_f) - \mathbf{v}(t_f) \|_R = \mathcal{O}(h^p).
\]

The bound in (3.4) is often not tight. Making stronger assumptions on the structure of the product of \(H\) and \(M\) allows one to derive a slightly tighter bound [40]. For some specific cases, one can derive asymptotically tighter bounds. For example, [51] proves a tighter bound for the advection-diffusion equation with particular approximations to the first and second derivatives.

### 3.4 Strong Stability

Consider the inhomogeneous right-moving advection problem

\[
\begin{align*}
\hat{u}_t &= -u_x + F, \\
u(x, 0) &= g(x), \\
u(a, t) &= f(t).
\end{align*}
\]
Applying the method of lines yields
\[ \begin{align*}
\dot{v} &= Mv + b + \hat{F}, \\
v(0) &= \hat{g},
\end{align*} \quad (3.6) \]

where the vector \( b \) captures the effects of the boundary function \( f \), and vector \( \hat{F} \) is the injection of \( F \). A sequence of ODEs (3.6) defined by system matrices \( M \), inhomogeneous terms \( b \) approximating the right-moving advection problem (3.5) is said to be **strongly stable in the \( H \)-norm** if there exists a \( K \) depending on \( t \) but not on the mesh spacing \( h \) such the approximate solution \( v \) obeys the bound
\[ \| v(t) \|_H^2 \leq K(t) \left( \| \hat{g} \|_H^2 + \| \hat{F} \|_H^2 + \max_{t' \in [0,t]} f^2 \right). \quad (3.7) \]

A strongly stable scheme is Lax stable, as the next theorem states:

**Theorem 3.4.1** If a method of form (2.12) that approximates problem (2.5) is strongly stable in a norm equivalent to the natural norm, then the method is also Lax stable.

Let \( H \) be any norm equivalent to the natural norm, and consider a problem of form (2.5) in which \( f = F = 0 \). Then
\[ \| e^{tM} \hat{g} \|_R \leq \frac{1}{\sqrt{c_1}} \| v(t) \|_H \leq \sqrt{\frac{K(t)}{c_1}} \| \hat{g} \|_H \leq \sqrt{\frac{K(t)c_2}{c_1}} \| \hat{g} \|_R, \quad (3.8) \]

where \( c_1 \) and \( c_2 \) are the upper and lower bound constants. Because bound (3.8) holds for any \( \hat{g} \),
\[ \| e^{tM} \|_R \leq \sqrt{\frac{c_2K(t)}{c_1}}. \]

### 3.5 Strong Pointwise Stability

A sequence of ODEs (3.6) defined by system matrices \( M \) and inhomogeneous terms \( b \) approximating the right-moving advection problem (3.5) is said to be **strongly pointwise stable in the \( H \)-norm** if there exists a \( K \) depending on \( t \) but not on the mesh spacing \( h \) such the approximate solution \( v \) obeys the bound
\[ \| v(t) \|_\infty^2 \leq K(t) \left( \| \hat{g} \|_H^2 + \| \hat{F} \|_H^2 + \max_{t' \in [0,t]} f^2 \right). \quad (3.9) \]

Strong pointwise stability is stricter than strong stability, as the next theorem states:

**Theorem 3.5.1** If a method of form (2.12) that approximates problem (2.5) on domain \( \Omega \) is strongly pointwise stable in a norm equivalent to the natural norm, then the method is also strongly stable.
Let $H$ be any norm equivalent to the natural norm with upper bound constant $c_2$. The proof follows from the fact that, for all $x$,

$$\|x\|_H^2 \leq \frac{1}{c_2} \|x\|_R^2 \leq \frac{2 \text{Area}(\Omega)}{c_2} \|x\|_\infty^2,$$

where $\text{Area}(\Omega)$ is the length of $\Omega$ in one dimension or its area in two dimensions. Figure 3.1 shows the relationships among the four stabilities.

![Diagram showing the relationships among stabilities](image)

**Figure 3.1:** Relationships among stabilities.
Chapter 4

Summation-by-Parts and Simultaneous Approximation Term

This chapter introduces summation-by-part matrices and the simultaneous approximation term methodology. These are the building blocks with which we will construct provably stable, overlapping domain decomposition methods.

4.1 Summation-By-Parts

Let \( \hat{\Omega} \) be a uniform discretization of \( \Omega = (a, b) \) with \( n \) nodes and spacing \( h \). The pair of matrices \( \{P, Q\} \) is a pair of summation-by-parts (SBP) matrices of approximation order \( p \) if the following conditions hold:

(a) The matrix \( P \) is real symmetric, positive definite, and equivalent to the natural norm. That is, there are positive constants \( \mu_1 \) and \( \mu_2 \) such that for all \( x \),

\[
\mu_1 \|x\|^2_R \leq \|x\|^2_P \leq \mu_2 \|x\|^2_R,
\]

(4.1)

where \( R = hI_n \), the natural norm.

(b) If \( f : [a, b] \to \mathbb{R} \) is sufficiently differentiable then

\[
\left\| \hat{f}_x - P^{-1}Q\hat{f} \right\|_P = \mathcal{O}(h^p).
\]

(4.2)

That is, \( P^{-1}Q \) is an order \( p \) approximation to \( \partial_x \).

(c) \( Q \) satisfies

\[
Q + Q^T = \text{diag}(-1, 0, 0, \ldots, 0, 1).
\]

(4.3)
An example pair of SBP matrices is

\[
P = h \begin{bmatrix} \frac{1}{2} & 1 & \cdots & 1 \\ 1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 \\ 1 & \cdots & 1 & \frac{1}{2} \end{bmatrix}, \quad Q = \begin{bmatrix} -\frac{1}{2} & \frac{1}{2} & \cdots & \cdots \\ -\frac{1}{2} & 0 & \frac{1}{2} & \cdots \\ \cdots & \cdots & \ddots & \ddots \\ -\frac{1}{2} & 0 & \frac{1}{2} & \cdots \end{bmatrix}.
\]

The three SBP properties imply that a discrete analog of the integration-by-parts property holds, as given in the following theorem:

**Theorem 4.1.1** Let \( \{P, Q\} \) be a pair of \( n \times n \) SBP matrices of any approximation order. For any \( x \) and \( y \) in \( \mathbb{R}^n \),

\[
\langle P^{-1}Qx, y \rangle_P = x_n y_n - x_1 y_1 - \langle x, P^{-1}Qy \rangle_P.
\]

A useful corollary to Theorem 4.1.1, which is used in the next section, can be derived by setting \( x = y \):

\[
2x^TQx = x_n^2 - x_1^2.
\]

There are many ways to construct pairs of SBP matrices \( \{P, Q\} \). One can construct pairs for which \( P^{-1}Q \) is up to a locally eighth-order approximation to \( \partial_x \) [10, 12, 27, 84]. If \( P \) is diagonal, then the operator \( P^{-1}Q \) is termed *explicit*. Otherwise, it is termed *implicit*. One can show that assuming a certain structure for \( Q \) precludes \( P \) from being proportional to the identity (see Appendix C). In what follows, we shall state any assumptions made on \( P \) and \( Q \).

We can extend SBP matrices to higher dimensions using the Kronecker product, as introduced in Section 2.11. Let \( \Omega \) be a uniform discretization of \( \Omega = (a_x, b_x) \times (a_y, b_y) \) with \( n_x \) nodes in the \( x \)-direction and \( n_y \) nodes in the \( y \)-direction. Let \( \{P_x, Q_x\} \) be a pair of \( n_x \times n_x \) SBP matrices of approximation order \( p \). Let \( \{P_y, Q_y\} \) be a pair of \( n_y \times n_y \) SBP matrices of approximation order \( q \). Define the following,

\[
H = P_x \otimes P_y,
\]

\[
G_x = Q_x \otimes P_y,
\]

\[
G_y = P_x \otimes Q_y.
\]

**Theorem 4.1.2** Let \( \{P_x, Q_x\} \) be a pair of SBP matrices of approximation order \( p \) and \( \{P_y, Q_y\} \) be a pair
of SBP matrices of approximation order $q$. If $f$ is sufficiently differentiable in $x$, then $H^{-1}G_x$ is an order $p$ approximation to $\partial_x$ in the $H$-norm,

$$\left\| \hat{f}_x - H^{-1}G_x \hat{f} \right\|_H = O(h_x^p).$$

If $f$ is sufficiently differentiable in $y$, then $H^{-1}G_y$ is an order $q$ approximation to $\partial_y$ in the $H$-norm,

$$\left\| \hat{f}_y - H^{-1}G_y \hat{f} \right\|_H = O(h_y^q).$$

An important analog of Theorem 4.1.1 relates inner products to boundary terms:

**Theorem 4.1.3** Let $\{P_x, Q_x\}$ and $\{P_y, Q_y\}$ be two pairs of SBP matrices of any approximation orders. Let $x$ and $y$ be any vectors in $\mathbb{R}^{n_x \times n_y}$. Then the following identities hold,

$$\langle x, H^{-1}G_x y \rangle_H = \langle x_E, y_E \rangle_{P_y} - \langle x_W, y_W \rangle_{P_y} - \langle H^{-1}G_x x, y \rangle_H,$$

$$\langle x, H^{-1}G_y y \rangle_H = \langle x_N, y_N \rangle_{P_x} - \langle x_S, y_S \rangle_{P_x} - \langle H^{-1}G_y x, y \rangle_H,$$

where $x_W$ is the vector composed of the elements of $x$ on the left boundary with order persevered, $x_E$ on the right boundary, $x_N$ on the top boundary, and $x_S$ in the lower boundary. The vectors $y_{E/W/N/S}$ are defined similarly.

The proof of Theorem 4.1.3 follows from the definition of the Kronecker product and the SBP properties. Let $\Delta = \text{diag}(-1, 0, \ldots, 0, 1)$. To show the first identity, note that

$$\langle x, H^{-1}G_x y \rangle_H = x^T (Q_x \otimes P_y) y$$

$$= x^T \left( (\Delta - Q_x^T) \otimes P_y \right) y$$

$$= x^T (\Delta \otimes P_y) y - x^T (Q_x^T \otimes P_y) y$$

$$= x^T (\text{block,diag}(-P_y, 0, \ldots, 0, P_y)) y - \langle H^{-1}G_x x, y \rangle_H$$

$$= \langle x_E, y_E \rangle_{P_y} - \langle x_W, y_W \rangle_{P_y} - \langle H^{-1}G_x x, y \rangle_H.$$

The proof of the second identity is analogous.

Again, setting $x = y$, we derive two useful identities. Let $x$ be any vector in $\mathbb{R}^{n_x \times n_y}$, then

$$2xG_x x = ||x_E||_{P_y}^2 - ||x_W||_{P_y}^2,$$

$$2xG_y x = ||x_N||_{P_x}^2 - ||x_S||_{P_x}^2.$$ (4.6)
The following theorem bounds the $H$-norm by the natural norm for a uniform discretization in two dimensions:

**Theorem 4.1.4** Let $R_x = h_x I_{n_x}$ and $R_y = h_y I_{n_y}$ be the natural norms associated with the uniform discretizations in each direction. By (4.1) there are $\mu_1^L$, $\mu_2^L$, $\mu_1^R$, and $\mu_2^R$ such that for any $x \in \mathbb{R}^{n_x}$ and $y \in \mathbb{R}^{n_y}$,

\[
\begin{align*}
\mu_1^L \|x\|^2_{R_x} &\leq \|x\|^2_{P_x} \leq \mu_2^L \|x\|^2_{R_x}, \\
\mu_1^R \|y\|^2_{R_y} &\leq \|y\|^2_{P_y} \leq \mu_2^R \|y\|^2_{R_y}.
\end{align*}
\]

Let $R = h_x h_n I_{n_x n_y}$, the natural norm for the two-dimensional discretization. Then

\[
\eta_1 \|w\|^2_R \leq \|w\|^2_H \leq \eta_2 \|w\|^2_R \quad \forall w \in \mathbb{R}^{n_x n_y},
\]

(4.7) where $\eta_1 = \mu_1^L \mu_1^R$ and $\eta_2 = \mu_2^L \mu_2^R$.

We say that a pair of matrices $\{P_c, Q_c\}$ satisfies the generalized summation-by-parts (GSBP) property if it satisfies the first two SBP properties (a) and (b) and a modified third property

\[(c') \quad Q_c \text{ satisfies } Q_c + Q_c^T = \text{block-diag}(\tilde{Q}_u, 0, 0, \ldots, 0, \tilde{Q}_l).
\]

(4.8) Matrix $\tilde{Q}_u$ is called the upper symmetric block and $\tilde{Q}_l$ the lower symmetric block.

Clearly any pair of matrices satisfying the SBP property also satisfies the GSBP property.

### 4.2 Simultaneous Approximation Term

The simultaneous approximation term (SAT) methodology is a process for using SBP and GSBP finite difference operators to enforce boundary conditions such that the resulting method is energy stable. It is based on three principles [10, 42, 51]:

(a) A PDE is discretized using finite difference operators constructed using SBP or GSBP matrices.

(b) Boundary conditions are enforced weakly through the use of penalty terms.

(c) Each penalty term contains penalty parameters, values used to enforce energy stability that may or may not directly correspond to quantities in the original PDE. These parameters are here denoted by the letters $\tau$ and $\sigma$. 

26
The power of the SAT methodology is in reducing the problem of finding energy stable methods to enforcing that a small number of penalty parameters are set to the correct values. In order to illustrate the general approach, we present a short summary of SAT methods for several problems.

### 4.2.1 SBP/SAT for Advection in One Dimension

Consider the advection problem (2.5) describing right-moving flow on the interval domain $\Omega = (a, b)$. Let $\{P, Q\}$ be a pair of SBP matrices. One possible SAT formulation of this problem using these matrices is

$$
\begin{align*}
\dot{v} &= -P^{-1}Qv - \tau P^{-1}s(v_1(t) - f(t)), \\
v(0) &= \hat{g}, \\
s &= [1, 0, \ldots, 0, 0]^T.
\end{align*}
$$

(4.9)

Observe that the three SAT properties are met:

(a) The continuous derivative operator $\partial_x$ has been replaced by the finite difference matrix $P^{-1}Q$.

(b) The boundary condition at the left point $x = a$ is enforced weakly through the penalty term $\tau P^{-1}s(v_1 - f)$. The method does not require $v_1$, the value that approximates $u(a, t)$, to be equal to $f(t)$. Strictly enforcing boundary conditions can result in an unstable system [10].

(c) The penalty term contains the penalty parameter $\tau$, a value that does not correspond to any quantity in the original PDE (2.5).

The purpose of the penalty parameter is to enforce energy stability and convergence by being “large enough,” as the next theorem shows:

**Theorem 4.2.1** Let $\{P, Q\}$ be a pair of SBP matrices of approximation order $p$. If $\tau \geq \frac{1}{2}$, then method (4.9) is energy stable in the $P$-norm and converges to the true solution $u$ at fixed time $t_f$ with order $p$,

$$
\|e(t_f)\|_p = \mathcal{O}(h^p).
$$

The first step in the proof of Theorem 4.2.1 is establishing that the system matrix of method (4.9) is energy stable in the $P$-norm. Observe that $P$ is equivalent to the natural norm. By construction, bounds (3.2) are satisfied with lower and upper bound constants $\mu_1$ and $\mu_2$, respectively, where $\mu_1$ and $\mu_2$ are defined by (4.1). For method (4.9), the system matrix is

$$
M = -P^{-1}Q - \tau P^{-1}ss^T.
$$
If $\tau \geq \frac{1}{2}$, then $M$ is negative semidefinite in the $P$-norm,

$$
x^T P M x = -x^T Q x - \tau x^T s s^T x
= \frac{(1 - 2\tau)x_1^2 - x_n^2}{2} \leq 0.
$$

Therefore, the system matrix $M$ is energy stable in the $P$-norm. By construction, $M$ is also consistent with approximation order $p$. Method (4.9) is consistent and energy stable, and therefore by Theorem 3.3.2, it converges with order $p$.

### 4.2.2 SBP/SAT for Heat Equation in One Dimension

SAT methods have been extended to problems with second derivatives, but more complicated penalty terms may be required. Consider the heat equation on the interval domain $\Omega = (a, b)$,

$$
\begin{align*}
  u_t & = \epsilon u_{xx}, \\
  u(x, 0) & = g(x), \\
  u(a, t) & = f_l(t), \\
  u(b, t) & = f_r(t),
\end{align*}
$$

where $\epsilon > 0$. A SAT method for this problem is

$$
\begin{align*}
  \dot{v} & = \epsilon D^2 v + \epsilon \sigma_1 P^{-1} (D^T s) (v_1 - f_l) + \epsilon P^{-1} (D^T r)(v_n - f_r), \\
  v(0) & = \hat{g}, \\
  s & = [1, 0, \ldots, 0, 0]^T, \\
  r & = [0, 0, \ldots, 0, 1]^T,
\end{align*}
$$

where $\{P, Q\}$ is a pair of SBP matrices of any approximation order, and $D = P^{-1}Q$. In this case, the penalty parameters must assume specific values, as the next theorem demonstrates:

**Theorem 4.2.2** Let $\{P, Q\}$ be a pair of SBP matrices of approximation order $p$. If $\sigma_1 = +1$ and $\sigma_2 = -1$, then method (4.11) is energy stable in the $P$-norm and converges with order $p$ to the true solution of (4.10) for any fixed time $t_f > 0$.

It suffices to show negative semidefiniteness. If $\sigma_1 = +1$ and $\sigma_2 = -1$, then for all $x$

$$
x^T P M x = \epsilon \left[ (\sigma_1 - 1)x_1 (Dx)_1 + (\sigma_2 + 1)x_n (Dx)_n - \|Dx\|_{P}^2 \right] \leq 0.
$$
4.2.3 GSBP/SAT for Advection in One Dimension

An example GSBP method for the right-moving advection problem (2.5) is

\[
\begin{align*}
\dot{v} &= -P^{-1}Qv + \tau P^{-1}w(v_1(t) - f(t)), \\
v(0) &= \hat{g}, \\
w &= \left[\tau \{Q\}_{1,1}, \{Q\}_{1,2} + \{Q\}_{2,1}, \ldots, 0, 0\right]^\top,
\end{align*}
\]

(4.12)

where \(\{P, Q\}\) is \(GSBP_4\), the GSBP pair given in Appendix A. We must, for the moment, restrict ourselves to GSBP4 because this particular pair of GSBP matrices and the above penalty vector \(w\) are, in a sense, matched. A different GSBP pair might require the choice of a different penalty vector.

**Theorem 4.2.3** Let \(\{P, Q\}\) be \(GSBP_4\). Let \(\hat{\Omega}\) be a uniform discretization of \((a, b)\). If \(\tau \geq 1\), then (4.12) is energy stable in the \(P\)-norm and converges with fourth-order to the solution of (2.5) for any \(t_f > 0\).

The proof proceeds as before. Note that

\[
x^\top PMx \leq x^\top Gx^\top,
\]

where

\[
G = \begin{bmatrix}
(\tau - 1)\frac{5}{8} & 0 \\
0 & -\frac{1}{8} \\
& & \ddots \\
& & \frac{1}{8} & -\frac{1}{8}
\end{bmatrix}.
\]

If \(\tau \geq 1\) then the maximum eigenvalue of \(G\) is zero, \(G\) is negative semidefinite, and system matrix \(M\) is negative semidefinite in the \(P\)-norm.

4.3 GSBP Systems

Proofs about SAT methods constructed from GSBP operators tend to be specific to a certain set of matrices with certain penalty terms and parameters. To abstract many of these details, we introduce the following notation. The triplet \(\{P, Q, w\}\) is called a **GSBP system** of approximation order \(p\) and size \(n\) if the following conditions hold:

(a) \(\{P, Q\}\) is a GSBP pair of dimension \(n \times n\) and approximation order \(p\).
(b) \( w \in \mathbb{R}^n \). Vector \( w \) is called the \textit{penalty vector}.

(c) The matrix \( G = -Q + ws^T \), where \( s = [1, 0, \ldots, 0]^T \in \mathbb{R}^n \), is negative semidefinite. \( G \) is called the \textit{energy matrix}.

The entries of \( w \) will depend on penalty parameters and the GSBP system being used. For example, if \( \{P, Q\} \) is a pair of SBP matrices, \( w = -\tau s \), and \( \tau \geq \frac{1}{2} \), then \( \{P, Q, w\} \) form a GSBP system, which follows from the proof of Theorem 4.2.1. Using GSBP systems, an energy stable method for solving (2.5) on uniform discretization (2.1) can be stated succinctly:

**Theorem 4.3.1** Let \( \{P, Q, w\} \) be a GSBP system of approximation order \( p \). Then method

\[
\dot{v} = -P^{-1}Qv + P^{-1}w (v_1(t) - f(t)) ,
\]

\[
v(0) = \hat{g}
\]

is energy stable in the \( P \)-norm and converges to the true solution of (2.5) at time \( t_f \) with order \( p \),

\[
\|e(t_f)\|_R = \mathcal{O}(h^p).
\]

Let \( G \) be the energy matrix associated with \( \{P, Q, w\} \) and \( M \) be the system matrix for (4.13). The proof follows from the fact that \( PM = G \).
Chapter 5

SBP-Based Methods

In Chapters 5 and 6, we present new, provably stable overlapping domain decomposition methods for the constant-coefficient hyperbolic system (2.8) in one dimension and the advection equation (2.10) in two dimensions. Many domain decomposition methods transfer information between grids by interpolating the approximate solution on one grid and overwriting the approximate solution on another (e.g., [3]). The methods in this chapter and the next are different. Much as the single domain method (4.9) enforced boundary conditions weakly, our overlapping decomposition methods weakly enforce continuity at the grid interfaces and do not use overwriting.

We will start with the scalar case \((k = 1\) in (2.4)) and build to systems \((k > 1)\). While these methods are interesting in their own right, we analyze them in order to motivate their generalizations for nonlinear problems in higher dimensions. It is often impossible to prove the stability or convergence of a method for a nonlinear problem. We shall follow the path set by others and prove that our methods are stable and convergent for linear problems and demonstrate empirically that they are for nonlinear ones.

5.1 Advection Systems in One Dimension

This chapter will describe provably Lax stable and convergent methods constructed from SBP matrices. We will show in Section 5.1.3 that, under certain assumptions, these methods are also strongly pointwise stable. The domain \(\Omega\) is decomposed into \(\Omega_1 = (a = a_L, b_L)\) and \(\Omega_2 = (a_R, b_R = b)\). The sets \(\hat{\Omega}_1\) and \(\hat{\Omega}_2\) are uniform discretizations defined by

\[
\begin{align*}
x^L_i &= a_L + (i - 1)h_L \quad i = 1, \ldots, n, \\
h_L &= (b_L - a_L) / (n - 1), \\
x^R_i &= a_R + (i - 1)h_R \quad i = 1, \ldots, m, \\
h_R &= (b_R - a_R) / (m - 1).
\end{align*}
\]

This layout is illustrated in Figure 5.1. The left grid, represented by the upper line, discretizes \((a = a_L, b_L)\). The right grid, represented by the lower line, discretizes \((a_R, b_R = b)\). Numbers above grid points give the
ordering. As the left and right discretizations are refined, the values \( b_L \) and \( a_R \) are constant, but the number of nodes that are contained in the overlap interval \((a_R, b_L)\) increases. In Chapter 6, we will consider the case when the number of points in the overlap region is held fixed.

Finally, let \( \{P, Q\} \) be a pair of SBP matrices. For convenience, define

\[
D = P^{-1}Q + \tau P^{-1}s_m s_m^T, \\
c = \tau P^{-1}s, \\
\tilde{D} = P^{-1}Q - \tau P^{-1}r_m r_m^T, \\
\tilde{c} = \tau P^{-1}r \\
s = [1, 0, \ldots, 0]^T, \\
r = [0, \ldots, 0, 1]^T. 
\] (5.2)

Figure 5.1: Interval \((a, b)\) discretized by two overlapping grids.

### 5.1.1 Single Domain Bounds

Our first step in proving the stability of SBP-based SAT methods on overlapping domains is to bound the single-domain method (4.9):

**Lemma 5.1.1** If \( \tau > \frac{1}{2} \), then method (4.9) obeys the bound

\[
\|v(t)\|_P^2 \leq \|v(0)\|_P^2 + \frac{\tau^2}{2\tau - 1} \int_0^t f^2 dt'. 
\] (5.3)

**Theorem 5.1.2** Assume that \( \tau > \frac{1}{2} \) and \( f \) is bounded. Then method (4.9) is strongly stable and obeys the bound

\[
\|v(t)\|_P \leq \|v(0)\|_P + \sqrt{\frac{\tau^2}{2\tau - 1}} \max_{t' \in [0, t]} (|f(t')|). 
\] (5.4)
The key to proving Theorem 5.1.2 is to bound the derivative of the square of the $P$-norm,

$$
\frac{d\|v\|^2_P}{dt} = \langle \frac{dv}{dt}, v \rangle_P + \langle v, \frac{dv}{dt} \rangle_P = 2 \langle v, \frac{dv}{dt} \rangle_P = 2 \left( v^T Q v - \tau v^T s_n s_n^T v + \tau f(t) v^T s_n \right) = \left( -v_N^2 + (1 - 2\tau)v_1^2 + 2\tau f(t)v_1 \right) .
$$

We follow [51] and rewrite (5.5) in terms of values with definite sign,

$$
-v_N^2 + (1 - 2\tau)v_1^2 + 2\tau f(t)v_1 = \frac{\tau^2}{2\tau - 1} f(t)^2 - v_N^2 - (2\tau - 1) \left( v_1 - \frac{\tau}{2\tau - 1} f(t) \right)^2 .
$$

Therefore,

$$
\frac{d \|v\|^2_P}{dt} \leq \frac{\tau^2}{2\tau - 1} f^2.
$$

(5.6)

Integrating (5.6) proves Theorem 5.1.2. The integral in (5.3) can then be bounded, yielding,

$$
\|v(t)\|^2_P \leq \|v(0)\|^2_P + t \frac{\tau^2}{2\tau - 1} \max_{t' \in [0,t]} (f^2) .
$$

(5.7)

Therefore, method (4.9) obeys the bound

$$
\|v(t)\|^2_P \leq K(t) \left( \|\hat{g}\|^2_P + \max_{t' \in [0,t]} f^2 \right) ,
$$

$$
K(t) = \max \left( 1, \frac{\tau^2}{2\tau - 1} \right)
$$

and is therefore strongly stable. Finally, applying square roots to both sides of (5.7) proves bound (5.4).

### 5.1.2 Overlapping Domains

In this section, we present a new, stable overlapping domain decomposition method that solves problem (2.5). We will consider the case of two overlapping domains, but the proof techniques, which are based on propagation of discretization and interpolation errors from upwind to downwind domains, can be generalized to cases with more than two overlapping domains.

In order to solve the advection problem (2.5) on two overlapping, uniform discretizations, we make the following ansatz:

- Each individual discretization should be energy stable in some norm. In this case, we will approximate the first derivative on the left and right subdomains using SBP derivative operators.
Problem (2.5) describes an advective flow from left to right. The initial condition restricted to the left subdomain $g|_{\Omega_1}$ and the boundary condition $u(a, t) = f(a, t)$ fully specify the continuous solution to problem (2.5) on the left subdomain $u|_{\Omega_1}$. Our numerical method should respect this same pattern: $\hat{g}^L$ and $f(a, t)$ will fully specify the approximate solution on the left uniform grid $\hat{v}^L$. The approximate solution on the right uniform grid $v^R$ will be specified by $v^L$ and $\hat{g}^R$.

Let (5.1) define two overlapping uniform grids, as in Figure 5.1. Let $\{P_L, Q_L\}$ be a pair of $n \times n$ SBP matrices of approximation order $p$. Let $\tau_L$ be the left penalty parameter. Using $P_L$, $Q_L$, and $\tau_L$, let $D_L$ and $c_L$ be the derivative operator and penalty vector defined by (5.2) when $P = P_L$, $Q = Q_L$ and $\tau = \tau_L$. The right operators are defined similarly. Let $\{P_R, Q_R\}$ be a pair of $m \times m$ SBP matrices of approximation order $q$. Let $\tau_R$ be the right penalty parameter. Define $D_R$ and $c_R$ by letting $P = P_R$, $Q = Q_R$ and $\tau = \tau_R$ in (5.2). Let $I$ be an $n \times 1$ matrix describing the interpolation from the left grid to the point $a_R$ on the right grid,

$$I \hat{u}^L(t) \equiv u(a_R, t) + e_{\text{interp}},$$

where $e_{\text{interp}}$ is the interpolation error. Our new method for the overlapping discretizations is given by

$$\begin{align*}
\dot{v} &= -Dv + f(t)d, \\
D &= \begin{bmatrix} D_L & 0_{n,m} \\
-\mathbf{c}_R I & D_R \end{bmatrix}, \\
d &= \begin{bmatrix} c_L \\
0_m \end{bmatrix}, \\
v(0) &= \hat{g}.
\end{align*}$$

Informally, the values of $\hat{v}^L$ surrounding the leftmost node of the right domain are interpolated in order to construct an approximation to $u(a_R, t)$. If $\hat{v}^R_i$ does not equal this value, then the equation governing the
evolution of $v^R$ is penalized. Figure 5.2 illustrates the case when three values of $v^L$ are used to approximate $u(a_R, t)$. This would be the case if $I$ were third-order polynomial interpolation.

With these particular choices, the eigenvalues of the system matrix of method (5.8) have nonpositive real parts. Because the system matrix $M = (-D)$ is block lower triangular, its spectrum is the union of the spectra of the two diagonal blocks. Therefore, if $\tau_L \geq \frac{1}{2}$ and $\tau_R \geq \frac{1}{2}$, then all eigenvalues of the system matrix will be in the left half plane. Proving convergence requires the following slightly stronger assumptions. The main result of this section is:

**Theorem 5.1.3** Assume the following:

- $\tau_L \geq \frac{1}{2}$ and $\tau_R > \frac{1}{2}$.
- The pair of SBP matrices $\{P_L, Q_L\}$ approximate $\partial_x$ to order $p$,
  \[ \|\hat{u}_x - P^{-1}_LQ_L\hat{u}\|_{P_L} \leq C_1h_P^p. \]
- The pair of SBP matrices $\{P_R, Q_R\}$ approximate $\partial_x$ to order $q$,
  \[ \|\hat{u}_x - P^{-1}_RQ_R\hat{u}\|_{P_R} \leq C_2h_R^q. \]
- The interpolation $I$ has the following properties:
  - It is accurate to order $r$,
    \[ |u(a_R, t) - I\hat{u}^L| \leq C_3h_L^r. \] \[(5.9)\]
  - There is a constant $C$ such that for vectors $x \in \mathbb{R}^n$
    \[ x^T I^T I x < C x^T x. \] \[(5.10)\]

Let $\mu$ be the lower bound constant from (4.1) relating $P_L$ and $R_L = h_L I_n$,

\[ \mu \|x\|_{R_L}^2 \leq \|x\|_{P_L}^2. \]
Then

\[
\|e^L(t_f)\|_{P_L} \leq t_f C_1 h_L^p, \\
\|e^R(t_f)\|_{P_R} \leq t_f C_3 h_R^q + S \sqrt{t_f} \left( C_3 h_R^r + \sqrt{\frac{C}{\mu} C_1 t_f h_L^{p - \frac{1}{2}}} \right),
\]

where \( S = \sqrt{\frac{\tau^2}{2 \tau h - 1}}. \)

The following is an important corollary:

**Corollary 5.1.4** Let \( h = \max(h_L, h_R) \) and \( s = \min(p - \frac{1}{2}, q, r) \). Then

\[
\|e^L(t_f)\|_{P_L} = O(h^p), \\
\|e^R(t_f)\|_{P_R} = O(h^s).
\]

The proof of Theorem 5.1.3 proceeds in the following way:

- Due to the lower block triangular structure of the system matrix, we use the energy stability of the left domain to show that \( v^L \) converges to \( \hat{u}^L \).

- The approximate solution on the right domain \( v^R \) is the sum of three parts:
  
  (a) a component corresponding to exact boundary conditions
  
  (b) two components corresponding to errors

- Using strong stability, we show that (a) converges to \( \hat{u}^R \) and (b) goes to zero.

The rest of this section will concern the details. The proofs of Theorem 5.1.3 and Corollary 5.1.4 will use the following lemma:

**Lemma 5.1.5** Let \( e^L \) be the cumulative error associated with the left discrete domain. Then

\[
\max_{t \in [0, t_f]} |\mathcal{I} e^L| \leq \sqrt{\frac{C}{\mu h_L}} \max_{t \in [0, t_f]} \|e^L\|_{P_L}.
\]

The proof follows from bound (5.10),

\[
|\mathcal{I} e^L| = \sqrt{e^L^T \mathcal{I}^T \mathcal{I} e^L} \leq \sqrt{C e^L^T e^L} \\
= \sqrt{\frac{C}{\mu h_L}} \|e^L\|_{R_L} \leq \sqrt{\frac{C}{\mu h_L}} \|e^L\|_{P_L}.
\]
To show convergence, consider the left and right discrete domains separately. The left approximate solution $v^L$ is independent of the right approximate solution $v^R$. Thus, by Theorem 4.2.1,

$$\|e^L(t_f)\|_{P_L} \leq t_f C_1 h^p_L.$$

Now, decompose $Iv^L$ into several pieces,

$$Iv^L = I (\hat{u}^L + e^L)$$

$$= u(a_R, t) + e_{\text{interp}} + I e^L,$$

where $e_{\text{interp}}$ is the interpolation error. Consider three subproblems,

$$\begin{align*}
\dot{w} &= -D_R w + u(a_R, t) c_R \\
w(0) &= \hat{g}^R \\
\dot{y} &= -D_R y + (e_{\text{interp}}) c_R \\
y(0) &= 0 \\
\dot{z} &= -D_R z + (I e^L) c_R \\
v(0) &= 0
\end{align*}$$

By linearity, $v^R = w + y + z$. The error $e^R$ is equal to the error in $w$ plus the solutions of the second and third ODEs,

$$e^R = v^R - \hat{u}^R = w + y + z - \hat{u}^R \equiv e_w + y + z.$$

Theorem 4.2.1 bounds $\|e_w\|_{P_R}$, and Theorem 5.1.2 bounds $\|y\|_{P_R}$ and $\|z\|_{P_R}$:

$$\begin{align*}
\|e_w(t_f)\|_{P_R} &\leq t_f C_2 h^9_R, \\
\|y(t_f)\|_{P_R} &\leq \sqrt{t_f} \frac{\mu}{(2\tau - 1)} \max_{t \in [0, t_f]} (|e_{\text{interp}}|) = S \sqrt{\tau^2} (C_3 h_L^p), \\
\|z(t_f)\|_{P_R} &\leq \sqrt{t_f} \frac{\mu}{(2\tau - 1)} \max_{t \in [0, t_f]} (|I e^L|) \leq S \sqrt{\frac{C_L}{\mu h_L}} \max_{t \in [0, t_f]} \|e^L\|_{P_L} \\
&\leq S \sqrt{\frac{C_L}{\mu h_L}} C_1 t_f h^p_L \leq S \sqrt{\tau^2} \left( \sqrt{\frac{C_L}{\mu}} C_1 t_f h^p_L \right). 
\end{align*}$$

(5.12)
Putting these together yields the desired result,

\[
\| \mathbf{e}^R \|_{PR} = \| \mathbf{e}_w + y + z \|_{PR} \\
\leq \| \mathbf{e}_w \|_{PR} + \| y \|_{PR} + \| z \|_{PR} \\
\leq t_f C_3 h R^q + S \sqrt{t_f} \left( C_3 h R^r + \sqrt{C \mu C_1 t_f h R^{p-\frac{1}{2}}} \right).
\]

Finally, we show that the requirements on the interpolation vector are satisfied by polynomial interpolation:

**Theorem 5.1.6** Assume that the solution \( u(x, t) \) is \( r \) times continuously differentiable in \( x \). For each discretization, let \( q \) be an integer such that

\[
x^L_1 \leq x^L_q \leq a_R \leq x^L_{q+r} \leq x^L_n.
\]

Then the following choice of \( \mathcal{I} \) satisfies (5.9) and (5.10) with \( C = \sqrt{r(r-1)^{2r-2}} \):

\[
\{\mathcal{I}\}_{1,i} = \begin{cases} 
0 & i < q \\
\ell_i(a_R) & i = q, \ldots, q + r \\
0 & i > q + r
\end{cases}
\]

where \( \ell_i \) is the \( i \)th Lagrange polynomial,

\[
\ell_i(a_R) = \prod_{j=q,j \neq i}^{q+r} \frac{a_R - x^L_j}{x^L_i - x^L_j}.
\]

The proof of (5.9) can be found in [30]. To show (5.10), we use the fact that the grid is equally spaced. For any \( i \),

\[
|\ell_i(a_R)| \leq \prod_{j=q,j \neq i}^{q+r} \left| \frac{a_R - x^L_j}{x^L_i - x^L_j} \right| \leq \prod_{j=q,j \neq i}^{q+r} \frac{h(r-1)}{h} = (r-1)^{r-1}.
\]

Finally, applying the Cauchy-Schwarz inequality yields

\[
x^T \mathcal{I}^T \mathcal{I} x \leq (\mathcal{I}^T \mathcal{I}) (\mathbf{x}^T) \leq r(r-1)^{2r-2} (\mathbf{x}^T \mathbf{x}).
\]

Note that the convergence rate is bounded by the order of the interpolation less one half. For example, if we used fourth-order SBP pairs in conjunction with fourth-order interpolation, we would find that the downwind domain converged with order 3.5. If there were more than two domains, all downwind domains
would also converge with order 3.5. We address this loss of one-half order in the next section.

5.1.3 Pointwise Stability

Placing stronger assumptions on the SBP pairs \( \{P_L, Q_L\} \) and \( \{P_R, Q_R\} \) allows stronger asymptotic bounds on convergence, as the next theorem states:

**Theorem 5.1.7** Assume all conditions of Theorem 5.1.3 hold. Additionally, assume that

- A method of form (4.9) constructed from \( \{P_L, Q_L\} \) is strongly pointwise stable and uniformly convergent with order \( p \).
- A method of form (4.9) constructed from \( \{P_R, Q_R\} \) is strongly pointwise stable and uniformly convergent with order \( q \).
- The matrix \( I \) contains at most \( r \) nonzero entries, each of magnitude bounded by \( B \). That is, for all \( i \),

\[
|\{I\}_{1,i}| \leq B.
\]

Then method (5.8) is strongly pointwise stable in the \( H \)-norm, where \( H = \text{block diag}(P_L, P_R) \). Further, the errors on the left and right domains obey

\[
\|e^L(t_f)\|_\infty = O(h^p),
\]

\[
\|e^R(t_f)\|_\infty = O(h^s),
\]

where \( h = \max(h_L, h_R) \) and \( s = \min(p, q, r) \).

If \( \{P_L, Q_L\} \) and \( \{P_L, Q_L\} \) are strongly pointwise stable, then there are \( K_1(t) \) and \( K_2(t) \) such that

\[
\|v^L\|_\infty^2 \leq K_1 \left( \|\hat{g}^L\|_{P_L}^2 + \max f^2 \right),
\]

\[
\|v^R\|_\infty^2 \leq K_2 \left( \|\hat{g}^R\|_{P_R}^2 + \max (Iv^L)^2 \right),
\]

where \( \max = \max_{t' \in [0,t]} \). Observe that

\[
(Iv^L)^2 = \left( \sum_{i=1}^n \{I\}_{1,i} v^L_i \right)^2 \leq r B^2 \|v\|_\infty^2.
\]

It follows immediately that

\[
\max (Iv^L)^2 \leq r B^2 \max \|v\|_\infty^2.
\]
Assume, without loss of generality, that $K_1$ is monotonically nondecreasing and bound $\|v^R\|_\infty^2$ by

$$
\|v^R\|_\infty^2 \leq K_2 \left( \|\hat{g}^R\|_{P_R}^2 + \max (Jv^L)^2 \right) \\
\leq K_2 \left( \|\hat{g}^R\|_{P_R}^2 + rB^2 \max \|v\|_\infty^2 \right) \\
\leq K_2 \left( \|\hat{g}^R\|_{P_R}^2 + K_1 \left( \|\hat{g}^L\|_{P_L}^2 + \max f^2 \right) \right) \\
\leq K_2 \max (1, rB^2K_1) \left( \|\hat{g}^R\|_{P_R}^2 + \|\hat{g}^L\|_{P_L}^2 + \max f^2 \right).
$$

(5.15)

Next, sum bounds (5.14) and combine with bound (5.15), yielding

$$
\|v^L\|_\infty^2 + \|v^R\|_\infty^2 \leq K_1 \left( \|\hat{g}^L\|_{P_L}^2 + \max f^2 \right) + K_2 \max (1, rB^2K_1) \left( \|\hat{g}^R\|_{P_R}^2 + \|\hat{g}^L\|_{P_L}^2 + \max f^2 \right) \\
\leq 2 \max (K_1, K_2 \max (1, rB^2K_1)) \left( \|\hat{g}^R\|_{P_R}^2 + \|\hat{g}^L\|_{P_L}^2 + \max f^2 \right).
$$

Noting that

$$
\|v\|_\infty^2 \leq \|v^R\|_\infty^2 + \|v^R\|_\infty^2, \\
\|\hat{g}\|_H^2 = \|\hat{g}^L\|_{P_L}^2 + \|\hat{g}^L\|_{P_L}^2
$$

yields the desired result.

To show the two convergence bounds (5.13), we follow a path similar that of Corollary 5.1.4. Let $h_L$ and $h_R$ be the spacing on the left and right grids. Note that $\|e^L\|_\infty = O(h_L^p)$. We again consider three problems and use each domain’s individual strong pointwise stability. Using the same notation as that in the proof of Theorem 5.1.3, observe that

$$
\|e_w\|_\infty = O(h_R^p), \\
\|y\|_\infty = O(\max (|e_{\text{interp}}|)) = O(h_L^r), \\
\|z\|_\infty = O(\max |J e^L|) \leq O(\max t \in [0, t_f] \|e^L\|_\infty) = O(h_L^p).
$$

(5.16)

The primary difference between (5.12) and (5.16) is that the uniform convergence assumption removes the loss of a half-order of accuracy.

We conclude by noting that all strong pointwise stability and uniform convergence bounds hold for certain low-order SBP pairs [87] and that the requirements on the interpolation operator hold if one uses polynomial interpolation, which follows from Theorem 5.1.6. Strong pointwise stability is conjectured to hold for all central difference based SBP pairs. Pointwise stability can be enforced by using dissipative SBP operators, which we have not used in our work and are discussed in [54].
5.1.4 Non-Unit-Velocity Waves

Consider the left-moving advection problem (2.6). The SBP/SAT formulation for overlapping grids is defined by considering a left-moving wave as a right-moving wave under a change of variables. From this it can be shown that

\[
\begin{align*}
\dot{v} &= \bar{D} v + f(t) \tilde{d}, \\
\bar{D} &= \begin{bmatrix}
\bar{D}_L & \tilde{c}_L \tilde{I} \\
0_{m,n} & \bar{D}_R
\end{bmatrix}, \\
\tilde{d} &= \begin{bmatrix}
0_n \\
\tilde{c}_R
\end{bmatrix}, \\
v(0) &= \hat{g},
\end{align*}
\]

(5.17)

where \(\tilde{I}\) is a \(1 \times m\) matrix that characterizes the interpolation,

\[
\tilde{I} u^R(t) \approx u(b_L, t),
\]

and is consistent and bounded in the sense of (5.9) and (5.10). We now have all the tools to consider the general advection problem (2.7):

**Theorem 5.1.8** Let \(D, \bar{D}, d, \) and \(\tilde{d}\) denote the matrices and inhomogeneous terms, respectively, defined by (5.8) and (5.17). If \(\tau_L > \frac{1}{2}\) and \(\tau_R > \frac{1}{2}\), then the method

\[
\begin{align*}
\dot{v} &= (\lambda^- D + \lambda^+ \bar{D}) v + f_a(t) d + f_b(t) \tilde{d}, \\
v(0) &= \hat{g},
\end{align*}
\]

(5.18)

converges for any finite time \(t_f\) to the solution of (2.7).

The key step in this proof is to perform a change of variables by rescaling time \(t' = |\lambda| t\).

5.1.5 Systems of Advection Equations in One Dimension

First, consider a system of decoupled advection equations of form (2.9):

**Theorem 5.1.9** Let \(\tilde{D}\) and \(D\) be defined by (5.8) and (5.17). Assume \(\tau_L > \frac{1}{2}\) and \(\tau_R > \frac{1}{2}\). Then the following method converges to the solution of (2.9):

\[
\begin{align*}
\dot{v} &= \left( (\tilde{D} \otimes \Lambda^+) + (D \otimes \Lambda^-) \right) v + (d \otimes f_a(t)) + (\tilde{d} \otimes f_b(t)), \\
v(0) &= \hat{g}.
\end{align*}
\]

(5.19)
The eigenvalues of the new system are the union of the eigenvalues of the scalar systems. Therefore, all eigenvalues of the scalar systems lie in the left half plane. The entire system converges because $v$ is a linear combination of $k$ independent components, each of which converges.

Next consider problem (2.8), a system that is not in characteristic form. Let $T$ be any orthogonal matrix that diagonalizes $A$,
$$T^TAT = \Lambda_A.$$  

Let $w(x,t) = T^Tu(x,t)$ be characteristic variables. Note that $w$ solves
$$
\begin{align*}
  w_t &= \Lambda_Aw_x, \\
  w(x,0) &= T^Tg(x), \\
  -\Lambda_A^-w(a,t) &= T^Tf_a(t), \\
  \Lambda_A^+w(b,t) &= T^Tf_b(t),
\end{align*}
$$

a system in the form of (2.9). Applying the diagonal method (5.19) to this new system and then transforming back gives the following result:

**Theorem 5.1.10** Assume $\tau_L > \frac{1}{2}$ and $\tau_R > \frac{1}{2}$. Then the following method converges to the solution of (2.8):
$$
\dot{v} = \left(\begin{array}{c}
\tilde{D} \otimes A^+ \end{array}\right) v + \left(\begin{array}{c}
D \otimes f_a(t) \end{array}\right) + \left(\begin{array}{c}
\tilde{d} \otimes f_b(t) \end{array}\right),
$$
$$v(0) = \hat{g}. \tag{5.20}$$

We have applied a similarity transformation to the system matrix from (5.19). Because similarity transformations preserve eigenvalues, all eigenvalues of (5.20) also lie in the left half plane.

**5.1.6 Brief Note on When $\tau_R = \frac{1}{2}$**

Theorem 5.1.3 requires that $\tau_R > \frac{1}{2}$, whereas $\tau_L \geq \frac{1}{2}$. The case $\tau_R = \frac{1}{2}$ is not covered by our proof. This follows from a lack of a bound equivalent to (5.6). Observe that there is no $C$ such that
$$\frac{d\|v\|^2_{\mu}}{dt} = -\nabla_N^2 - f(t)v_1 \leq Cf(t)^2$$

for all $v$ and $f$. In the proof of Theorem 5.1.3, because there is no proven strong stability, one cannot use the bounds on $e_{\text{interp}}$ and $e_L^L$ to bound $e^R$. We have observed empirically, however, that the convergence rates when $\tau_R = \frac{1}{2}$ and $\tau_R > \frac{1}{2}$ are the same. In the next section we will construct a new method for overlapping grids in two dimensions. The proof of stability of this method also relies on strong stability. In this context,
strong stability will again require that some of the penalty parameters be larger than $\frac{1}{2}$.

### 5.2 Advection Systems in Two Dimensions

In this section, we will numerically solve (2.10) by extending the results of the previous sections. At a high level, the construction and proof of convergence of the two-dimensional method for overlapping domains will be the same as that of the previous section: first show strong stability of the single domain problem and use this to bound the influence of the truncation and interpolation errors on the downwind domain.

Let $\Omega = (a_x, b_x) \times (a_y, b_y)$ and consider problem (2.10). Initially, let $\lambda_x < 0$ and $\lambda_y < 0$. The statement corresponding to (2.4) is

$$
\begin{align*}
  u_t &= \lambda_x u_x + \lambda_y u_y, \\
  u(x, y, 0) &= g(x), \\
  |\lambda_x| u(a_x, y, t) &= f(a_x, y, t), \\
  |\lambda_y| u(x, a_y, t) &= f(x, a_y, t).
\end{align*}
$$

We review an SBP/SAT method for solving problem (5.21) on a single subdomain, and then extend this method to overlapping subdomains. In the second case, $\Omega$ is decomposed into left subdomain $\Omega_1 = (a = a^L_x, b_x) \times (a_y, b_y)$ and right subdomain $\Omega_2 = (a^R_x, b_x = b_x) \times (a_y, b_y)$. The set $\hat{\Omega}_1$ is defined by

$$
\begin{align*}
  h_{x,L} &= (b_x^L - a_x^L) / (n_x - 1), \\
  x_i^L &= a_x + (i - 1)h_{x,L}, \\
  h_{y,L} &= (b_y - a_y) / (n_y - 1), \\
  y_i^L &= a_y + (i - 1)h_{y,L}, \\
  n &= n_xn_y.
\end{align*}
$$

The set $\hat{\Omega}_2$ is defined by

$$
\begin{align*}
  h_{x,R} &= (b_x^R - a_x^R) / (m_x - 1), \\
  x_i^L &= a + (i - 1)h_{x,R}, \\
  h_{y,R} &= (b_y - a_y) / (m_y - 1), \\
  y_i^L &= a_y + (i - 1)h_{y,R}, \\
  m &= m_xm_y.
\end{align*}
$$

This layout is illustrated in Figure 2.1. As in the previous section, when the left and right discretizations are refined, the area of the overlap region is fixed and the number of nodes that are contained in the overlap region increases.
5.2.1 Single Domain Bounds

Let \( \tilde{\Omega} \) be the uniform discretization of \( \Omega \) defined by (2.2). As in the one-dimensional case, we define positive and negative SAT derivative operators. Let \( \{P_x, Q_x\} \) and \( \{P_y, Q_y\} \) be two pairs of SBP matrices. Define the positive (marked by an overset tilde) and negative (no tilde) SAT derivative operators and penalty vectors as

\[
D_x = \left( (P_x^{-1}Q_x + \tau_x P_x^{-1}s_{n_x}s_{n_x}^T) \otimes I_{n_y} \right),
\]

\[
c_x = \tau_x P_x^{-1}s_{n_x},
\]

\[
\hat{D}_x = \left( (P_x^{-1}Q_x - \tau_x P_x^{-1}r_{n_x}r_{n_x}^T) \otimes I_{n_y} \right),
\]

\[
\hat{c}_x = \tau_x P_x^{-1}r_{n_x},
\]

\[
D_y = \left( I_{n_x} \otimes \left( P_y^{-1}Q_y + \tau_y P_y^{-1}s_{n_y}s_{n_y}^T \right) \right),
\]

\[
c_y = \tau_y P_y^{-1}s_{n_y},
\]

\[
\hat{D}_y = \left( I_{n_x} \otimes \left( P_y^{-1}Q_y - \tau_y P_y^{-1}r_{n_y}r_{n_y}^T \right) \right),
\]

\[
\hat{c}_y = \tau_y P_y^{-1}r_{n_y}.
\]

Also define \( H = P_x \otimes P_y \).

Let \( f_{a_x} = f(a_x, y, t) \) and \( f_{a_y} = f(x, a_y, t) \). The SBP/SAT formulation for this problem is

\[
\dot{v} = (\lambda_x D_x + \lambda_y D_y) v + \left( c_x \otimes f_{a_x}(t) \right) + \left( \hat{f}_{a_x}(t) \otimes c_y \right),
\]

\[
v(0) = g,
\]

\[
\dot{f}_{a_x}(t) = \left[ f(a_x, y_1, t), \ldots, f(a_x, y_{n_y}, t) \right]^T,
\]

\[
\dot{f}_{a_y}(t) = \left[ f(x_1, a_y, t), \ldots, f(x_{n_x}, a_y, t) \right]^T.
\]

**Theorem 5.2.1** Let \( \{P_x, Q_x\} \) and \( \{P_y, Q_y\} \) be two pairs of SBP matrices of approximation order \( p \). If \( \tau_x \geq \frac{1}{2} \) and \( \tau_y \geq \frac{1}{2} \), then method (5.23) is energy stable in the \( H \)-norm and converges to the solution of (5.21) for any \( t_f > 0 \). Moreover,

\[
\|e(t_f)\|_H = O \left( h_x^p + h_y^p \right).
\]

The system is strongly stable under stricter assumptions:

**Theorem 5.2.2** Assume \( \tau_x > \frac{1}{2} \) and \( \tau_y > \frac{1}{2} \). Then method (5.23) is strongly stable and

\[
\|v(t)\|_H^2 \leq \|v(0)\|_H^2 + \int_0^t \frac{\tau_x^2}{2|\lambda_x|} f_{a_x}^2 + \frac{\tau_y^2}{|\lambda_y|^2} \|\hat{f}_{a_y}\|_{P_y}^2 \, dt'.
\]

We will use the following corollary:

**Corollary 5.2.3** Let \( R_x = h_x I_{n_y} \) and \( R_y = h_y I_{n_y} \). Let \( \mu_1 \) and \( \mu_2 \) be grid-independent upper bound constants
Again, we rewrite (5.24) in terms of values with definite sign, defined by

\[ \| \hat{f} \|_{P_y} \leq \mu_1 \| x \|_{P_x}^2 \quad \forall x \in \mathbb{R}^{n_x}, \]
\[ \| \hat{y} \|_{P_y} \leq \mu_2 \| y \|_{P_y}^2 \quad \forall y \in \mathbb{R}^{n_y}. \]

Let \( \mu = \max(\mu_1, \mu_2) \). Assume that \( n_x \) and \( n_y \) are both larger than 2. Then

\[
\| v(t) \|_{H} \leq \| v(0) \|_{H} + \sqrt{2\mu} \left( \frac{\tau_x^2}{|\lambda_x| (2\tau_x - 1)} \left( b_y - a_y \right) \max_{\tau \in [0, t]} \| \hat{f}_y \|_{P_y} \right) + \sqrt{\frac{\tau_y^2}{|\lambda_y| (2\tau_y - 1)}} \left( b_x - a_x \right) \max_{\tau \in [0, t]} \| \hat{f}_x \|_{P_x} \),
\]

where \( \| x \|_{\infty} = \max_i |x_i| \).

Define \( \hat{f}_1 \) and \( \hat{f}_2 \) by

\[
\hat{f}_a = |\lambda_y| \hat{f}_1, \quad \hat{f}_x = |\lambda_x| \hat{f}_2.
\]

Using Theorem 4.1.3, the time derivative of \( \| v(t) \|_{H}^2 \) can be related to boundary terms,

\[
\frac{d}{dt} \| v \|_{H}^2 = 2\lambda_y v^T \left( ((Q_x + \tau_x s_{n_x} s_{n_x}^T) \otimes P_y) v - (\tau_x s_{n_y} \otimes P_y \hat{f}_2) \right) + 2\lambda_y v^T \left( (P_x \otimes (Q_y + \tau_y s_{n_y} s_{n_y}^T)) v - (P_x \hat{f}_1 \otimes \tau_y s_{n_y}) \right) = \lambda_y \left( (2\tau_x - 1) \| v_E \|_{P_y}^2 + \| v_W \|_{P_y}^2 - 2\tau_x \langle v_E, \hat{f}_2 \rangle_{P_y} \right) + \lambda_y \left( (2\tau_y - 1) \| v_S \|_{P_x}^2 + \| v_N \|_{P_x}^2 - 2\tau_y \langle v_S, \hat{f}_1 \rangle_{P_x} \right).
\]

Again, we rewrite (5.24) in terms of values with definite sign,

\[
\frac{d}{dt} \| v \|_{H}^2 = \lambda_y \left( \| v_W \|_{P_y}^2 + \| v_E \|_{P_y} \right) + \lambda_y \left( \| v_N \|_{P_x}^2 + \| v_S \|_{P_x} \right) \leq \lambda_y \left( \frac{\tau_x^2}{2\tau_x - 1} \| \hat{f}_2 \|_{P_y}^2 + \frac{\tau_y^2}{2\tau_y - 1} \| \hat{f}_1 \|_{P_x}^2 \right) \leq \frac{\tau_x^2}{|\lambda_x| (2\tau_x - 1)} \| \hat{f}_x \|_{P_x}^2 + \frac{\tau_y^2}{|\lambda_y| (2\tau_y - 1)} \| \hat{f}_y \|_{P_x}^2.
\]

Integrating the bound (5.25) proves Theorem 5.2.2. Let \( R_x = h_x I_{n_x} \) and \( R_y = h_y I_{n_y} \). Let \( B_1 \) be any bound on \( |f_{a_y}(x, t)| \). Then we have the bound

\[
\| \hat{f}_y \|_{P_y}^2 \leq \mu \| \hat{f}_x \|_{P_x}^2 \leq \mu h_x n_x B_1^2 = \mu B_1^2 ((b_x - a_x) + h_x) \leq 2\mu B_1^2 (b_x - a_x).
\]
Deriving the analogous result for the $y$-direction and taking square roots proves Corollary 5.2.3.

5.2.2 Overlapping Domains

We next extend this result to a domain that has been decomposed into two subdomains, each with its own uniform discretization. Define the following four pairs of SBP matrices:

<table>
<thead>
<tr>
<th></th>
<th>Size</th>
<th>Approx. Order</th>
</tr>
</thead>
<tbody>
<tr>
<td>${P_{x,L}, Q_{x,L}}$</td>
<td>$n_x \times n_x$</td>
<td>$p$</td>
</tr>
<tr>
<td>${P_{y,L}, Q_{y,L}}$</td>
<td>$n_y \times n_y$</td>
<td>$p$</td>
</tr>
<tr>
<td>${P_{x,R}, Q_{x,R}}$</td>
<td>$m_x \times m_x$</td>
<td>$q$</td>
</tr>
<tr>
<td>${P_{y,R}, Q_{y,R}}$</td>
<td>$m_y \times m_y$</td>
<td>$q$</td>
</tr>
</tbody>
</table>

For simplicity, in the following definitions we will use a single penalty parameter denoted by $\tau$. Using $\{P_{x,L}, Q_{x,L}\}$, $\{P_{y,L}, Q_{y,L}\}$, and $\tau$, define the four derivative operators, four penalty vectors, and a norm for the left domain by setting $P_x = P_{x,L}$, $P_y = P_{y,L}$, $Q_x = Q_{x,L}$, $Q_y = Q_{y,L}$, $\tau_x = \tau$ and $\tau_y = \tau$ in (5.22), and similarly for the right domain using $\{P_{x,R}, Q_{x,R}\}$, $\{P_{y,R}, Q_{y,R}\}$, and $\tau$. For example, the norm for the right domain and the negative derivative operator for the left are

$$
H_R = P_{x,R} \otimes P_{y,R},
$$

$$
D_{y,L} = I_{n_x} \otimes \left( P_{y,L}^{-1} Q_{y,L} + \tau P_{y,L}^{-1} s_n s_n^T \right).
$$

The method for solving (5.21) on two overlapping uniform discretizations is

$$
\dot{v} = (\lambda_x D_x + \lambda_y D_y) v + \begin{bmatrix} c_{x,L} \otimes \hat{f}_{L,a_x}(t) & 0_n \end{bmatrix} + \begin{bmatrix} \hat{f}_{L,a_y}(t) \otimes c_{y,L} & \hat{f}_{R,a_y}(t) \otimes c_{y,R} \end{bmatrix},
$$

$$
D_x = \begin{bmatrix} D_{x,L} & 0_{n,m} \\ - (c_{x,R} \otimes I) & D_{x,R} \end{bmatrix},
$$

$$
D_y = \begin{bmatrix} D_{y,L} & 0_{n,m} \\ 0_{m,n} & D_{y,R} \end{bmatrix},
$$

$$
v(0) = \hat{g}.
$$

The following are the requirements for convergence:

**Theorem 5.2.4** Assume the following:

- The mesh is refined in such a way that the ratio of any two step sizes is a fixed constant. Let $h = \max(h_{x,L}, h_{y,L}, h_{x,R}, h_{y,R})$. 

46
• \( \tau > \frac{1}{2} \).

• The interpolation matrix satisfies the following properties:
  
  – The interpolation is consistent,
    \[
    \| \mathcal{I} \hat{u} - \hat{u}_E^R \|_{P_y,R} = \mathcal{O}(h^r). \tag{5.27}
    \]
  
  – For all \( x \in \mathbb{R}^{n_x} \),
    \[
    x^T \mathcal{I}^T \mathcal{I} x \leq C x^T x. \tag{5.28}
    \]

Then method (5.26) converges to the solution of (5.21) for any \( t_f > 0 \). Moreover,

\[
\| e^L(t_f) \|_{H_L} = \mathcal{O}(h^p),
\]

\[
\| e^R(t_f) \|_{H_R} = \mathcal{O}(h^s),
\]

where \( s = \min(p - \frac{1}{2}, q, r) \).

The proof of Theorem 5.2.4 is analogous to that of Theorem 5.1.3. The analog to Lemma 5.1.5 is:

**Lemma 5.2.5** Let \( R_y,R = h_y,R I_{m_y} \), the natural norm for the \( y \)-discretization of the right domain. Let \( \mu \) be the upper bound constant from (4.1), so that for all \( y \in \mathbb{R}^{m_y} \),

\[
\| y \|_{P_y,R}^2 \leq \mu \| y \|_{R_y,R}^2.
\]

Let \( \eta \) be the lower bound constant from (4.7), so that for all \( w \in \mathbb{R}^{n_x} \),

\[
\eta \| y \|_{R_L}^2 \leq \| y \|_{H_L}^2.
\]

Then

\[
\max_{t \in [0,t_f]} \| \mathcal{I} \! e^L \|_{P_y,R} \leq \mathcal{O}(h^{-\frac{1}{2}}) \max_{t \in [0,t_f]} \| e^L \|_{H_L}. \]

The proof of Lemma 5.2.5 follows from the norm equivalences. For any \( t \in [0, t_f] \),

\[
\| \mathcal{I} e^L \|_{P_y,R} \leq \sqrt{\mu} \| \mathcal{I} e^L \|_{R_y,R} \leq \sqrt{\frac{C \mu h_y,R}{\eta h_x,L h_y,L}} \| e^L \|_{H_L} = \mathcal{O}(h^{p-\frac{1}{2}}). \]
The left domain is independent of the right and converges with order $p$. We analyze the right domain by considering three subproblems. The first corresponds to the exact boundary conditions. This contributes an error of order $q$. The second corresponds to the error introduced by interpolation. This contributes an error of order $r$. Finally, the third problem corresponds to the interpolation of $e^L$. By Lemma 5.2.5, this contributes an error of order $(p - \frac{1}{2})$. The asymptotically largest of these three errors determines the asymptotic error behavior.

The requirements on interpolation are very general. The following gives a set of sufficient conditions:

**Theorem 5.2.6** Let $I$ satisfy the following properties:

- The interpolation defined by $I$ is locally $O(h^r)$,

$$|u(a_R, y^L_i) - (I v^L)_i| = O(h^r) \quad i = 1, \ldots, m_y. \quad (5.29)$$

- There is a $B_r$ such that for all discretizations, each row of $I$ contains at most $B_r$ non-zeros.

- There is a $B_c$ such that for all discretizations, each column of $I$ contains at most $B_c$ non-zeros.

- There is a $B$ such that for all discretizations, for all entries,

$$|\{I\}_{i,j}| \leq B.$$

Then $I$ satisfies (5.27) and (5.28) with

$$C = B^2 B_c B_r.$$

The proof that (5.29) implies (5.27) follows from the norm-equivalence between the $P_y,R$-norm and $R_y,R$-norm. To show (5.28), we will use facts about matrix norms, all of which can be found in [22]. Observe that

$$\|I\|_2^2 \leq \|I\|_1 \|I\|_\infty,$$

where $\|I\|_2$ is the matrix 2-norm of $I$, $\|I\|_1$ is the matrix 1-norm, and $\|I\|_\infty$ is the matrix infinity norm. The matrix 1-norm is equal to the maximum absolute column sum,

$$\|I\|_1 = \max_j \sum_{i=1}^{m_y} |\{I\}_{i,j}| \leq B_c B.$$
The matrix $\infty$-norm is the maximum absolute row sum,

$$
\|\mathcal{I}\|_\infty = \max_i \sum_{j=1}^{n_x n_y} |(\mathcal{I})_{i,j}| \leq B_r B.
$$

Bound (5.28) is derived from

$$
x^T \mathcal{I} x \leq \|\mathcal{I}\|_2^2 (x^T x) \leq B^2 B c B r x^T x.
$$

### 5.2.3 General Advection

Thus far we have required $\lambda_x < 0$ and $\lambda_y < 0$. The remaining cases can be considered by applying a change of variables. For example, if $\lambda_x > 0$ and $\lambda_y > 0$, let $x' = -x$ and $y' = -y$. Summarizing the results,

$$
\dot{v} = \left( \lambda_x^+ \mathcal{D}_x + A_1^+ \mathcal{D}_x + \lambda_y^+ \mathcal{D}_y + A_2^+ \mathcal{D}_y \right) v
+ \mathcal{C}_x \dot{f}_{a_x} + \mathcal{C}_x \dot{f}_{b_x} + \mathcal{C}_y \dot{f}_{a_y} + \mathcal{C}_y \dot{f}_{b_y},
$$

$$v(0) = \dot{g},
$$

(5.30)

where

$$
\mathcal{D}_x = \begin{bmatrix}
D_{x,L} & 0_{n,m} \\
-(c_{x,R} \otimes I) & D_{x,R}
\end{bmatrix},
\mathcal{D}_y = \begin{bmatrix}
D_{y,L} & 0_{n,m} \\
0_{m,n} & D_{y,R}
\end{bmatrix},
$$

$$
\mathcal{D}_x^+ = \begin{bmatrix}
\tilde{D}_{x,L} & 0_{n,m} \\
c_{x,L} \otimes I_{n_y} & \tilde{D}_{x,R}
\end{bmatrix},
\mathcal{D}_y^+ = \begin{bmatrix}
\tilde{D}_{y,L} & 0_{n,m} \\
0_{m,n} & \tilde{D}_{y,R}
\end{bmatrix},
$$

$$
\mathcal{C}_x = \begin{bmatrix}
c_{x,L} \otimes I_{n_y} \\
0_{m,n_y}
\end{bmatrix},
\mathcal{C}_y = \begin{bmatrix}
I_{n_x} \otimes c_{y,L} & 0_{n,m_x} \\
0_{m,n_x} & I_{m_x} \otimes c_{y,R}
\end{bmatrix},
$$

$$
\mathcal{C}_x^+ = \begin{bmatrix}
0_{n,m_y} \\
c_{x,R} \otimes I_{n_y}
\end{bmatrix},
\mathcal{C}_y^+ = \begin{bmatrix}
I_{n_x} \otimes \tilde{c}_{y,L} & 0_{n,m_x} \\
0_{m,n_x} & I_{m_x} \otimes \tilde{c}_{y,R}
\end{bmatrix},
$$

Note that we are applying + and − operators to the scalars $\lambda_x$ and $\lambda_y$. For example, if $\lambda_x = -1$, then $\lambda_x^+ = 0$ and $\lambda_x^- = -1$. This has the effect of switching components on or off.

49
5.2.4 Systems of Equations in Two Dimensions

We next consider systems of advection equations in two dimensions of the form (2.11). The generalization of method (5.30) is

$$\frac{dv}{dt} = \left( (D_x \otimes A_x) + (\tilde{D}_x \otimes A_x^+) + (D_y \otimes A_y^-) + (\tilde{D}_y \otimes A_y^+) \right) v$$

$$+ (C_x \otimes I_k) \hat{f}_{a_x} + (\tilde{C}_x \otimes I_k) \hat{f}_{b_x} + (C_y \otimes I_k) \hat{f}_{a_y} + (\tilde{C}_y \otimes I_k) \hat{f}_{b_y},$$

$$v(0) = \hat{g}.$$ (5.31)

If $A_x$ and $A_y$ are simultaneously diagonalizable, then (5.31) describes a transformed system of advection equations in two dimensions. Therefore, by a proof similar to that of Theorem 5.1.8, method (5.31) converges to the solution of problem (2.11). We are unable to prove convergence and stability in the case that $A_x$ and $A_y$ are not simultaneously diagonalizable. However, Section 7.2 demonstrates empirically that method (5.31) is stable and convergent in the specific case of the linearized Euler equations.

5.3 Numerical Results

In this section, we present numerical results for the methods presented in this chapter.

5.3.1 Methods for problems in One Dimension

We will solve the two model IBVPs. The domain $\Omega = (-1, 1)$ is decomposed into overlapping subdomains: $\Omega_1 = (-1, 0)$ and $\Omega_2 = (0, 1)$. Uniform grids are constructed that meet the following requirements:

- The leftmost node of the right domain is halfway between two nodes of the left domain,

$$a_R = 0.1 = x_i^R = \frac{x_{i+1}^L + x_i^L}{2}.$$

- The rightmost node of the left domain is aligned with a node on the right domain,

$$b_L = 0 = x_n^L = x_j^R.$$

- $h_L = 2h_R$. 

50
\( \tilde{I} \) and \( I \) are defined by second-order linear interpolation,

\[
\tilde{I}u^L(t) = \frac{1}{2} (u(x_i^L, t) + u(x_{i+1}^L, t)) = u(a_R, t) + \mathcal{O}(h_R^2),
\]

\[
\tilde{I}u^R(t) = u(x_j^R, t) = u(b_L, t).
\]

The layout is illustrated in Figure 5.3. These uniform grids were chosen in order to highlight the details of Theorem 5.1.3 and show how interpolation error affects the approximate solutions. In the following examples, \( \{P_L, Q_L\} \) and \( \{P_R, Q_R\} \) are both of approximation order 3.5, \( \tau_L = 1 \), and \( \tau_R = 1 \). Time integration was performed using the standard fourth-order Runge-Kutta method. All simulations used the same time step, which was sufficiently small to ensure stability. Results are shown in Figure 5.4.

Figure 5.3: Composite grid used for testing scalar problems. Downward arrow indicates second-order interpolation \( I \). Upward arrow indicates injection \( \tilde{I} \).

Right-Moving Wave

We solve problem (2.5) with a Gaussian initial pulse using method (5.8). The results are shown in Figure 5.4. Theorem 5.1.3 predicts that the convergence rate of the left domain will be 3.5 and that of the right will be 2. Note that the left domain converges at fourth-order. Reasons to expect that, although the SBP pairs are of approximation order 3.5, the left domain still converges with fourth-order can be found in [25, 86]. The right domain receives the second-order interpolation error and converges with second-order, the expected rate.

Left-Moving Wave

We solve problem (2.6) using method (5.17). Observe in Figure 5.4c that both domains converge with fourth-order because the left domain receives no interpolation error from the right domain.
5.3.2 Methods for Problems in Two Dimensions

We next consider a problem in two spatial dimensions. The domain $\Omega = (-1, 1) \times (0, 1)$ is decomposed into overlapping subdomains: $\Omega_1 = (-1, 0) \times (0, 1)$ and $\Omega_2 = (-0.1, 1) \times (0, 1)$. Consider the hyperbolic system

$$u_t = A_x u_x + A_y u_y,$$

$$A_x = \text{diag} \left( -1, 1, \frac{1}{\sqrt{2}} \right),$$

$$A_y = \text{diag} \left( 0, 0, -\frac{1}{\sqrt{2}} \right).$$

The initial and boundary data are provided by

$$u = [u_1, u_2, u_3]^T,$$

$$u_1 = e^{-50 (x-t+\frac{1}{2})^2 + (y-\frac{1}{2})^2},$$

$$u_2 = e^{-50 (x+t-\frac{1}{2})^2 + (y-\frac{1}{2})^2},$$

$$u_3 = e^{-50 \left( x+\frac{1}{2} - \frac{1}{\sqrt{2}} y \right)^2 + \left( y - \frac{1}{\sqrt{2}} + \frac{x}{2} \right)^2}.$$

Note that each component of the solution of problem (5.33) models a Gaussian pulse that moves with unit speed. The first component $u_1$ describes a pulse that moves to the left, the second component $u_2$ describes a pulse that moves to the right, and the third component $u_3$ describes a pulse that moves up and to the left.

We apply method (5.31) to problem (5.33). Because $A_x$ and $A_y$ are diagonal, they are simultaneously diagonalizable, and method (5.31) is provably Lax stable and convergent. Both subdomains are discretized by two-dimensional uniform grids. We use the same fourth-order SBP pair as that in Section 5.3.1 and fourth-order interpolation at the boundaries. Time integration is performed using the standard fourth-order Runge-Kutta method. We measure fourth-order convergence. The results are shown in Figure 5.5.

5.4 Toward Energy Stable Methods

We have been unable to prove that the methods in this chapter are energy stable. In this section, we illustrate the challenge of satisfying both requirements (3.2) and (3.3). We use method (5.8), which numerically solves the right-moving problem (2.5), as an example.

Consider the norm defined by $H = \text{block_diag}(H_L, H_R)$, where $H_L$- and $H_R$-norms approximate the $L^2$-norm. Method (5.8) is not energy stable in the $H$-norm. Consider the right-moving pulse shown in Figure 5.6a. As the pulse enters the interface region, the global energy that is measured by the $H$-norm will increase as the energy of the pulse is counted by both the left $H_L$-norm and the right $H_R$-norm, as shown
in Figure 5.6b. Because the energy is not non-increasing, method (5.8) is not energy stable in the $H$-norm.

Another approach is to enforce explicitly that the system matrix of method (5.8) is negative semidefinite in the $H$-norm. Method (5.8) is negative semidefinite in some norm $H$ if and only if there is a positive semidefinite matrix $\Sigma$ such that

\[ HM + M^T H = -\Sigma, \]

where $M$ is the system matrix of method (5.8). If all eigenvalues of $M$ have negative real parts, then

\[ H = \int_0^\infty e^{tM} \Sigma e^{tM} \, dt. \]

We have been unable to prove that the spectrum of $M$ lies in the strict left-half plane, but it is true in all cases we have tested. Figure 5.4d shows the spectrum of one simulation. Assuming that the spectrum lies in the strict left-half plane leaves the theoretical hurdle of choosing $\Sigma$ in such a way that the resulting $H$-norm is equivalent to the natural norm.

Our solution is to construct provably energy stable overlapping domain decomposition methods using generalized summation-by-parts operators, which we present in the next chapter.
Figure 5.4: (a) Approximate solution to problem (2.5) using method (5.8) at times $t = 0, 0.5,$ and 1. At $t = 0.5$, blue squares give solution on left subdomain and red asterisks give solution on right. (b) Local convergence plot for problem (2.5) using method (5.8) with fourth-order Padé SBP pairs, showing errors $\|e_L\|_{R_L}$ and $\|e_R\|_{R_R}$ as functions of mesh spacings $h_L$ and $h_R$, respectively. (c) Same, for problem (2.6) using method (5.17) with same fourth-order Padé SBP pairs. (d) Spectrum of system matrix of method (5.8)
Figure 5.5: (a) First component of approximate solution of method (5.31) when applied to problem (5.33). (b) Same, showing second component. (c) Same, showing third component. (d) Spectrum of system matrix of method (5.31). (e) Local convergence plot for problem (5.33) using method (5.31) with fourth-order Padé SBP pairs, showing errors $\|e^L\|_{RL}$ and $\|e^R\|_{RR}$ as functions of mesh spacings $h_{x,L}$ and $h_{x,R}$, respectively.
Figure 5.6: (a) Right-moving approximate solution crossing interface. (b) Energy of simulation of right-moving advection problem (2.5) with method (5.8). (c) Energy of simulation of problem (5.33) with method (5.31).
Chapter 6

GSBP-Based Methods

Whereas the previous chapter presented Lax-stable methods, this chapter describes new, energy stable, overlapping domain decomposition methods. The assumptions about the ways in which domains are decomposed and discretized are different than those in Chapter 5. In the last chapter, we assumed that the length of the overlap region remained constant as the mesh was refined. In this chapter, we assume that the length of the overlap region decreases as the mesh is refined, but the number of nodes inside this region remains fixed, as illustrated in Figure 6.1.

![Figure 6.1: Illustration of method (6.1) applied to right-moving advection problem (2.5). Downward pointing arrow indicates interpolation. Leftmost node of right domain is always within fixed number of nodes of rightmost node of left domain.](image)

MATLAB-style notation is used to denote subsets of vectors,

\[ \{x\}_{i:j} = [x_i, \ldots, x_j]^T. \]

We use an unsubscripted norm \( \|x\| = \sqrt{x^T x}. \)

6.1 Energy Stable Overlapping Method for Advection Equation

Again consider the overlapping domain decomposition (5.1). The main result of this chapter is:

**Theorem 6.1.1** Assume that \( \Omega = (a, b) \) is discretized according to (5.1). Let \( \{P_L, Q_L, w_L\} \) be a GSBP system of approximation order \( p \) and size \( n \), with energy matrix \( G_L \). Let \( \{P_R, Q_R, w_R\} \) be a GSBP system...
of approximation order $q$ and size and $m$, with energy matrix $G_R$. Let $I \in \mathbb{R}^n$ be an interpolation vector satisfying

$$I^T \hat{u}^L(t) = u(a_R, t) + \mathcal{O}(h^r).$$

Consider the method

$$\begin{align*}
\dot{v}^L &= -P_L^{-1}Q_L v^L + P_L^{-1} w_L (v^L_I - f(t)), \\
v^L(0) &= \hat{g}^L, \\
\dot{v}^R &= -P_R^{-1}Q_R v^R + P_R^{-1} w_R (v^R_I - I^T v^L), \\
v^R(0) &= \hat{g}^R.
\end{align*}$$

(6.1)

Assume the following

- $Q_L + Q_L^T = \text{block}_d \left( \ldots, 0, \tilde{Q}_L^I \right)$, where $\tilde{Q}_L^I$ is $s^L \times s^L$ and $s^L \geq 1$. Further, $\tilde{Q}_L^I$ is positive definite.
- $G_R + G_R^T = \text{block}_d \left( \tilde{G}_R^u, 0, \ldots \right)$, where $\tilde{G}_R^u$ is $s^R \times s^R$ and $s^R \geq 1$. Further, $\tilde{G}_R^u$ is negative definite.
- The right penalty vector has a fixed number of entries of fixed size:

$$\begin{align*}
\{w_R\}_j &= 0, \quad j = s^R + 1, \ldots, m, \\
\|\bar{w}_R\| &= C_1,
\end{align*}$$

where $\bar{w}_R = \{w_R\}_{[1:s^R]}$.

- Most entries of $I$ are zero, and the rest have bounded size:

$$\begin{align*}
I_j &= 0, \quad j = 1, \ldots, n - s^L, \\
\|\bar{I}\| &< C_2,
\end{align*}$$

(6.2)

where $\bar{I} = \{I\}_{[n-s^L+1:n]}$.

- Define

$$\alpha^* = \frac{-\lambda_{\min}(\tilde{Q}_L^I) \lambda_{\max}(\tilde{G}_R^u)}{C_1^2 C_2^2},$$

(6.3)

where $\lambda_{\min}(\tilde{Q}_L^I)$ and $\lambda_{\max}(\tilde{G}_R^u)$ denote the minimum eigenvalue of $\tilde{Q}_L^I$ and the maximum eigenvalue of $\tilde{G}_R^u$, respectively.

- Define $s = \min(p, q, r)$. 

58
Figure 6.2: Nonzero patterns of $Q_L + Q^T_L$ and $I$ required by hypotheses of Theorem 6.1.1. Entries that may be nonzero are represented by colored blocks.

Then method (6.1) is energy stable in the $H$-norm, where

$$H = \begin{bmatrix} P_L & \alpha P_R \end{bmatrix}$$

(6.4)

and $\alpha$ is any scalar satisfying $0 < \alpha \leq \alpha^\ast$, and convergent in the sense of (2.16) with bounds

$$\|e^L(t_f)\|_{R_L} = O(h^s), \quad \|e^R(t_f)\|_{R_R} = O(h^s).$$

(6.5)

Figure 6.2 illustrates the pattern of non-zero entries specified by Theorem 6.1.1. Our proof of stability requires that $Q_L + Q^T_L$, represented by the square in the lower right corner, is strictly positive definite and has dimension commensurate with that of the nonzeros of $I$.

Both SBP-based method (5.8) and GSBP-based method (6.1) solve problem (2.5) in similar ways: values of $v^L$ surrounding the leftmost node of the right domain are interpolated in order to construct a consistent approximation to $a(a_R, t)$. However, the stability proof of method (6.1) will require stronger assumptions about the discretization of the underlying domain. Proving stability of method (5.8) required that the interpolation operator be consistent and bounded. To prove the stability of method (6.1), the interpolation must not only be consistent and bounded, but must also have the specific nonzero structure given by (6.2).

We will make two assumptions in order to satisfy these three requirements:

- Operator $I$ interpolates the last $s^L$ entries of $v^L$ using polynomial interpolation.
• The length of the overlap region, $b_L - a_R$, is chosen such that

$$b_L - a_R < (s^L - 1)h_L. \quad (6.6)$$

Bound (6.6) implies that the length of the overlap region decreases as the mesh is refined.

These two assumptions are illustrated in Figure 6.1 for the case when $s^L$ and the interpolation order are both equal to 3.

The parameter $\alpha$ is neither related to the original PDE (2.5), nor does it appear in method (6.1). This reflects that the norm matrix $H$ is a theoretical tool used to show several properties. Its existence proves that method (6.1) is energy stable, and by Theorem 3.3.2, it follows that method (6.1) is also convergent. By the Lyapunov theorem [19], one can prove that all eigenvalues of the system matrix have non-positive real parts.

We prove Theorem 6.1.1 by first proving an auxiliary lemma and then applying that lemma to method (6.1):

**Lemma 6.1.2** Let $a$, $b$, $c$, and $\beta$ be real numbers. Assume $a > 0$ and $c > 0$. Define

$$f_\beta(x, y) = -ax^2 + \beta bxy - \beta cy^2.$$ 

Then there exists $\beta^* > 0$ such that if $0 \leq \beta \leq \beta^*$, then for all $x, y \in \mathbb{R}$, $f_\beta(x, y) \leq 0$.

The statement holds if $b = 0$ and $\beta^* = 1$. Assume $b \neq 0$. Observe that

$$f_\beta(x, y) = [x, y] F \begin{bmatrix} x \\ y \end{bmatrix}, \quad F = \begin{bmatrix} -a & \beta b \\ \beta b & -\beta c \end{bmatrix}.$$ 

Let $\beta^* = \frac{4ac}{b^2}$. If $0 \leq \beta \leq \beta^*$, then $\det F \geq 0$. By Sylvester’s criterion, matrix $F$ is negative semidefinite.

Next, we apply Lemma 6.1.2 to method (6.1). We define the following vectors that represent the approximate solution on the rightmost nodes of the left domain and the leftmost nodes of the right domain,

$$l = \{v^L\}_{[n-s^L+1:n]}, \quad r = \{v^R\}_{[1:s^R]}.$$ 

Observe that

$$(v^R)^T w_R \bar{T} v^L = r^T \bar{w}_R \bar{T} l.$$
If $\alpha \leq \alpha^*$, then
\[
\frac{d}{dt} \|v\|_H^2 = 2v^THMv \\
\leq -1^T\tilde{Q}_L^t1 + 2\alpha r^T\tilde{w}_R^t\tilde{T}1 + \alpha r^T\tilde{G}_R^u r \\
\leq -\lambda_{\text{min}}(\tilde{Q}_L^t) \|l\|^2 + 2\alpha \|\tilde{w}_R\| \|\tilde{T}\| \|l\| \|r\| + \alpha \lambda_{\text{max}}(\tilde{G}_R^u) \|r\|^2 \\
\leq 0,
\]
where the last inequality follows from Lemma 6.1.2. The proof of bounds (6.5) follows from the block structure of $H$ and norm equivalence.

### 6.2 Example

We illustrate Theorem 6.1.1 by using GSBP4, given in Appendix A. These GSBP matrices were chosen because they have a simple structure with $2 \times 2$ upper and lower symmetric blocks. Assume that $\Omega$ has been discretized by two overlapping uniform grids, as in Figure 5.1. Assume that the leftmost node of the right domain lies between the two rightmost nodes of the left domain,

$$x_{n-1}^L \leq a_R \leq x_n^L. \quad (6.7)$$

Let $\theta = 1 - \frac{a_R - b_L}{h_r}$. For this choice of GSBP pair and using second-order linear interpolation, we have

$$s^L = 2, \quad s^R = 2,$$

$$\tilde{w}_R = [-\tau \frac{5}{8}, -\frac{1}{4}], \quad \tilde{Q}_L^t = \begin{bmatrix}
\frac{1}{4} & \frac{1}{4} \\
\frac{1}{4} & \frac{3}{4}
\end{bmatrix},$$

$$\tilde{T} = [\theta, 1-\theta], \quad \tilde{G}_R^u = \text{diag} \left((1-\tau) \frac{5}{4}, -\frac{1}{4}\right),$$

where $\tau$ is a penalty parameter. Therefore,

$$\|\tilde{w}_R\|^2 = \tau^2 \frac{25}{64} + \frac{1}{16}, \quad \|\tilde{T}\|^2 \leq 1,$$

$$\lambda_{\text{max}}(\tilde{G}_R^u) = \max ((1-\tau) \frac{5}{4}, -\frac{1}{4}), \quad \lambda_{\text{min}}(\tilde{Q}_L^t) \approx 0.146.$$ 

Letting $\tau = 2$, then $\alpha^* \approx 0.0225$. This case is further investigated in Appendix D.
6.3 New GSBP Pairs

We have constructed two new GSBP pairs \( \{ P, Q \} \). The first is a fourth-order approximation to \( \partial_x \) on the interior and third-order at the boundary. In this case, the upper symmetric block is \( 4 \times 4 \), which admits the use of fourth-order interpolation. The second pair of matrices is sixth-order on the interior and fifth-order at the boundary. The lower symmetric block is \( 6 \times 6 \) and admits sixth-order interpolation. These pairs of matrices are constructed by simultaneously enforcing linear constraints ensuring that \( P^{-1}Q \) is a suitable first derivative operator and nonlinear constraints ensuring that \( P \) is positive definite and that the upper and lower symmetric blocks are indefinite and positive definite respectively. We present the construction of the matrices and a Fourier analysis in Appendix B.

6.4 Brief Note on Permissible Interpolation

The requirement that \( I_j = 0 \) for \( j = 1, \ldots, n - s^L \) is strict. The next theorem states that if it is not satisfied, then method (6.1) is not energy stable in the \( H \)-norm:

**Theorem 6.4.1** Assume that the hypotheses of Theorem 6.1.1 hold true, except that there is at least one \( j \leq (n - s^L) \) such that

\[
I_j \neq 0.
\]

Then there is no \( \alpha > 0 \) such that method (6.1) is energy stable in the \( H \)-norm.

Let \( I_j \neq 0 \), where \( j \leq (n - s^L) \), select an \( \ell \) such that \( \{ w_R \}_\ell \neq 0 \). Let \( M \) be the system matrix associated with (6.1). Define

\[
\Sigma = HM + M^\top H.
\]

Through explicit computation, one finds

\[
\{ \Sigma \}_{[j,n+\ell],[j,n+\ell]} = \alpha \begin{bmatrix}
0 & I_j \{ w_R \}_\ell \\
I_{1,j} \{ w_R \}_\ell & G_{R,\ell,\ell}^n
\end{bmatrix}, \tag{6.8}
\]

where \( \{ \Sigma \}_{[j,n+\ell],[j,n+\ell]} \) is the principal minor of \( \Omega \) composed of the \( j \)th and \( (n + \ell) \)th elements in the \( j \)th and \( (n + \ell) \)th rows. Taking the determinant of (6.8),

\[
\det \left( \{ \Sigma \}_{[j,n+\ell],[j,n+\ell]} \right) = -\alpha^2 (I_j \{ w_R \}_\ell)^2 < 0.
\]
Therefore $\Sigma$ is indefinite because it contains an indefinite principle minor, and system matrix $M$ is not energy stable in the $H$-norm.

## 6.5 Energy Stable Method for Hyperbolic Systems in One Dimension

In this section, we show how to extend the method presented in the previous section to an energy stable method for the hyperbolic systems. We show how to extend the GSBP method from the previous section to scalar advection problems with arbitrary wave velocities. Finally, we solve the hyperbolic problem (2.8).

### 6.6 Non-Unit-Velocity Waves

The first step in extending the energy stable method (6.1) to problem (2.8) is to consider the left-moving wave (2.6).

**Theorem 6.6.1** Let $\{P_L, Q_L, w_L\}$ and $\{P_R, Q_R, w_R\}$ be GSBP systems of order $p$ and $q$, respectively. Let $\tilde{I}$ describe an order $r$ interpolation,

$$
\tilde{I}^T \tilde{u}^R(t) = u(b_L, t) + O(h_R^r).
$$

Consider

$$
\begin{align*}
\dot{v}^L &= -P_L^{-1}Q_L^t v^L + P_L^{-1} w_L^t (v_n^L - \tilde{I}^T v^R), \\
v^L(0) &= \tilde{g}^L, \\
\dot{v}^R &= -P_R^{-1}Q_R^t v^R + P_R^{-1} w_R^t (v_m^R - f(t)), \\
v^R(0) &= \tilde{g}^R.
\end{align*}
$$

(6.9)

Under the same assumptions as Theorem 6.1.1 with $I = \tilde{I}^\sharp$, method (6.9) is energy stable in the $\tilde{H}$-norm, where

$$
\tilde{H} = \begin{bmatrix}
\alpha P_L^t \\
P_R^t
\end{bmatrix}
$$

(6.10)

and $\alpha$ is any scalar satisfying $0 < \alpha \leq \alpha^*$, and converges to the solution of problem (2.6) with order $\min (p, q, r)$.

Note that $-P^{-1}Q$ approximates the derivative $+\partial_x$ on both the left and right domains, which follows from Corollary 2.12.2. The proof follows from considering a left-moving wave as a right-moving wave under a
change of variables.

We now proceed to systems with arbitrary wave velocity. Let \( \{P, Q, w\} \) be a GSBP system and define

\[
\begin{align*}
D &= P^{-1}Q - P^{-1}ws^T, \\
c &= P^{-1}w, \\
\tilde{D} &= -P^{-1}Q^T + P^{-1}w^sr^T, \\
\tilde{c} &= -P^{-1}w^s, \\
s &= [1, 0, \ldots, 0]^T, \\
r &= [0, \ldots, 0, 1]^T.
\end{align*}
\] (6.11)

Consider the general advection equation in one dimension (2.7). Let \( \{P_L, Q_L, w_L\} \) and \( \{P_R, Q_R, w_R\} \) be two GSBP systems. Using \( P_L, Q_L, \) and \( w_L \), let \( D_L, c_L, \tilde{D}_L, \tilde{c}_L \) be the derivative operators and penalty vectors defined by (6.11) when \( P = P_L, Q = Q_L, w = w_L \). The right operators are defined similarly. Using \( P_R, Q_R, \) and \( w_R \), let \( D_R, c_R, \tilde{D}_R, \tilde{c}_R \) be the derivative operators and penalty vectors defined by (6.11) when \( P = P_R, Q = Q_R, w = w_R \). Further, assume \( \{P_L, Q_L, w_L\}, \{P_R, Q_R, w_R\}, I, \) and \( \tilde{I} \) meet all the criteria of Theorems 6.1.1 and 6.6.1. Let

\[
\begin{align*}
\dot{v} &= \left(\lambda^+\tilde{D} + \lambda^-D\right)v - f_b(t)\tilde{d} - f_a(t)d, \\
v(0) &= \hat{g}.
\end{align*}
\] (6.12)

where

\[

\begin{align*}
D &= \begin{bmatrix} D_L & 0_{n,m} \\ c_RI & D_R \end{bmatrix}, \quad d = \begin{bmatrix} c_L \\ 0_m \end{bmatrix}, \\
\tilde{D} &= \begin{bmatrix} \tilde{D}_L & -c_L\tilde{I} \\ 0_{m,n} & \tilde{D}_R \end{bmatrix}, \quad \tilde{d} = \begin{bmatrix} 0_n \\ \tilde{c}_R \end{bmatrix}.
\end{align*}
\]

**Theorem 6.6.2** Let

\[
H_\lambda = N_\lambda + \tilde{H} + (1 - N_\lambda)H,
\]

with \( H \) and \( \tilde{H} \) defined by (6.4) and (6.10), respectively. Then method (6.12) is energy stable in the \( H_\lambda \)-norm and converges to the solution of (2.7) for any \( t_f > 0 \).

The key step in this proof is to perform a change of variables by rescaling time \( t' = |\lambda|t \). The factor \( N_\lambda \) “switches on” the appropriate norm.


6.7 Systems of Advection Equations in One Dimension

A system of hyperbolic equations can be solved by decomposing it into characteristic variables and using the scalar method (6.12). We begin with system (2.9), which is already in characteristic form.

**Theorem 6.7.1** Let $\tilde{D}$ and $D$ and all penalty parameters be defined as in Theorem 6.6.2. Further, let

$$
\mathcal{H} = \left( \tilde{H} \otimes N_{\lambda^+} \right) + \left( H \otimes (I - N_{\lambda^+}) \right),
$$

with $H$ and $\tilde{H}$ defined by (6.4) and (6.10), respectively. Then method

$$
\begin{align*}
\mathbf{v} &= \left( \left( \tilde{D} \otimes \Lambda^+ \right) + \left( D \otimes \Lambda^- \right) \right) \mathbf{v} + \left( d \otimes f_a(t) \right) + \left( \tilde{d} \otimes f_b(t) \right), \\
\mathbf{v}(0) &= \hat{g}
\end{align*}
$$

is energy stable in the $\mathcal{H}$-norm and converges to the solution of (2.9).

Method (5.19) is energy stable and thus convergent because it is the sum of $k$ energy stable, convergent, and independent components. Observe that $\mathbf{v}$ is a linear combination of $k$ components, one for each wave velocity $\lambda_i$. The time evolution of the component corresponding to $\lambda_i$ is described by method (6.12) and is energy stable in the $\mathcal{H}_{\lambda_i}$-norm. The $\mathcal{H}$-norm is the sum of the $\mathcal{H}_{\lambda_1}, \ldots, \mathcal{H}_{\lambda_k}$ norms, each applied to the corresponding component. Because the $\mathcal{H}_{\lambda_i}$-norm of each component is non-increasing, their sum is also non-increasing. Because $\mathcal{H}$ is a symmetric permutation of a block diagonal matrix and each block is either $H$ or $\tilde{H}$, $\mathcal{H}$ is positive definite and satisfies bound (3.2). Note also that the spectrum of the system matrix of (6.14) is the union of the spectra of the scalar systems (6.1) and (6.9). Because both are energy stable, all eigenvalues of the system matrix (6.14) have non-positive real parts.

Next consider a general system (2.8) that is not in characteristic form. Let $T$ be any orthogonal matrix that diagonalizes $A$,

$$
T^T A T = \Lambda_A.
$$

Let $w(x, t) = T^T u(x, t)$ be characteristic variables. Note that $w$ solves

$$
\begin{align*}
w_t &= \Lambda_A w_x, \\
w(x, 0) &= T^T g(x), \\
-\Lambda^- w(a, t) &= T^T f_a(t), \\
\Lambda^+_A w(b, t) &= T^T f_b(t),
\end{align*}
$$

65
a system in the form of (2.9). Applying the diagonal method (6.14) to this new system and then transforming back gives the following result:

**Theorem 6.7.2** Let $\tilde{D}$ and $D$ and all penalty parameters be defined as in Theorem 6.6.2. Further, let $T$ be any orthogonal matrix that diagonalizes $A$ and define

$$\mathcal{H} = \left( \tilde{H} \otimes [T^T N_\alpha T^T] \right) + \left( H \otimes [T^T (I - N_\alpha) T^T] \right),$$

with $H$ and $\tilde{H}$ defined by (6.4) and (6.10), respectively. Then method

$$\dot{v} = \left( \left( \tilde{D} \otimes A^+ \right) + \left( D \otimes A^- \right) \right) v + \left( d \otimes f_a(t) \right) + \left( \tilde{d} \otimes f_b(t) \right), \quad v(0) = \tilde{g} \quad (6.15)$$

is energy stable in the $\mathcal{H}$-norm and converges to the solution of (2.4).

We have applied a similarity transformation to the system matrix from (6.14). Because similarity transformations preserve spectra, all eigenvalues of (6.15) also have non-positive real parts.

### 6.8 Numerical Results

We test method (6.1) solving the right-moving wave problem (2.5), where $g(x) = \sin(2\pi x)$ and $f(t) = \sin(2\pi(1 - t))$. The exact solution is $u(x, t) = \sin(2\pi(x - t))$. The original domain $\Omega$ is decomposed into overlapping subdomains: $\Omega_1 = (-1, 0)$ and $\Omega_2 = (-\frac{L}{2}, 1)$. In order to achieve an energy stable method that converges with fourth-order on both domains, we must use high-order interpolation. By Theorem 6.1.1, energy stability requires the GSBP pair to have large lower symmetric blocks. We construct a new GSBP pair meeting this criterion in Appendix B.1. We have used this new GSBP pair in conjunction with fourth-order interpolation to achieve fourth-order convergence on both domains. Time integration was performed using the standard fourth-order Runge-Kutta method. The results are shown in Figure 6.3a. Appendix B.2 presents a new high-order GSBP pair with $6 \times 6$ lower symmetric blocks. We use these matrices and sixth-order polynomial interpolation to achieve sixth-order convergence on both domains. The results are shown in Figure 6.3b.
Figure 6.3: (a) Local convergence plot for problem (2.5) using method (6.1) with new fourth-order Padé GSBP pairs, showing errors $\|e^L\|_{P^L}$ and $\|e^R\|_{P^R}$ as functions of mesh spacings $h_L$ and $h_R$, respectively. (b) Same, using new sixth-order Padé GSBP pairs.
Chapter 7

Numerical Results

In this chapter, we will apply the methods from Chapters 5 and 6 to linear and nonlinear problems for which we cannot prove stability. In the one-dimensional problem, the computational domain is $\Omega = (−1, 1)$. In the two dimensional problems, the computational domain is $\Omega = (−1, 1) \times (0, 1)$. In all cases, initial and boundary data were provided by the analytic solution to the Cauchy problem. Time integration was performed using the standard fourth-order Runge-Kutta method.

7.1 Burgers’ Equation

Nonlinear problems of the form (2.8), where $A = A(u)$, can be solved using either explicit or implicit GSBP pairs. However, the use of explicit GSBP pairs creates an opportunity for efficiency. For many explicit GSBP pairs $\{P, Q\}$, the row of the discrete derivative operator $P^{-1}Q$ that approximates the derivative at the interior point $x_i$ and the corresponding row of $-P^{-\sharp}Q^{\sharp}$ are the same. Such an operator allows a discrete approximation to $Au_x|_{x_i}$ to be calculated without diagonalizing $A$. We have constructed such an explicit GSBP pair, which is shown in Appendix B.3. The operator $P^{-1}Q$ is a fourth-order approximation to $\partial_x$ in the interior and second-order at the boundary. We have applied this explicit discrete derivative operator to the inviscid Burgers’ equation with a Gaussian initial pulse,

$$
\begin{align*}
  u_t &= -u \, u_x, \\
  u(x, 0) &= e^{-10x^2}.
\end{align*}
$$

The solution, defined implicitly by

$$
  u = e^{-10(x-ut)^2},
$$

is smooth for all $x$ and all $t < \frac{\sqrt{5}}{10} \approx 0.2236$ [55]. Characteristics from the interior specify the solution at $x = 1$. Our numerical method reflects this property by not enforcing boundary conditions at this endpoint. The analytic solution to the Cauchy problem (7.2) specifies the solution at $x = -1$. As shown in Figure 7.1b, we achieve greater than third-order convergence on both the left and right subdomains.
7.2 Linearized Euler Equations

To demonstrate the efficacy of the SBP-based approach given in Chapter 5, we solve the non-dimensional linearized Euler equations in two dimensions in the form the First CAA Benchmark, Category 3 [29]. This problem models a Gaussian perturbation of amplitude $G$, width $\alpha$, and center $(c_x, c_y)$ to a fluid of otherwise constant density and constant flow in the $(M_x, M_y)$-direction. Let $\rho$ be the perturbation to the density, and $(v_x, v_y)$ be the perturbation to the velocity. The linearized Euler equations are

$$ u_t = A_x u_x + A_x u_y, $$

$$ A_x = \begin{bmatrix} -M_x & -1 & 0 \\ -1 & -M_x & 0 \\ 0 & 0 & -M_x \end{bmatrix}, $$

$$ A_y = \begin{bmatrix} -M_y & 0 & -1 \\ 0 & -M_y & 0 \\ -1 & 0 & -M_y \end{bmatrix}. $$

(7.3)
where
\[
\begin{align*}
\mathbf{u} &= [\rho, v_x, v_y]^T, \\
\rho(x, y, 0) &= Ge^{-\alpha r^2}, \\
v_x(x, y, 0) &= 0, \\
v_y(x, y, 0) &= 0, \\
r^2 &= (x - c_x)^2 + (y - c_y)^2.
\end{align*}
\]

Problem (7.3) has the analytic solution
\[
\begin{align*}
\rho &= \frac{G}{2\pi} \int_0^\infty ze^{-\frac{z^2}{4\alpha}\sin(tz)J_0(z\eta)} \, dz, \\
v_x &= \frac{G}{2\pi\eta^2} \int_0^\infty ze^{-\frac{z^2}{4\alpha}\cos(tz)J_0(z\eta)} \, dz, \\
v_y &= \frac{G}{2\pi\eta^2} \int_0^\infty ze^{-\frac{z^2}{4\alpha}\cos(tz)J_0(z\eta)} \, dz, \\
x &= x - M_x t - c_x, \\
\tilde{y} &= y - M_y t - c_y, \\
\eta &= \sqrt{x^2 + \tilde{y}^2},
\end{align*}
\] (7.4)

where \(J_i\) is the \(i\text{th}\) Bessel function of the first kind \([90]\). In this normalization, the perturbation to density \(\rho\) has the same value as the perturbation to pressure

For these simulations, \((c_x, c_y) = (-3.10, 1.2), \alpha = 100, \text{ and } (M_x, M_y) = (1, 0)\). Derivatives are approximated using fourth-order Padé SBP matrices, with fourth-order Lagrangian interpolation. We refine the mesh in such a way that all step-sizes are fixed ratios of the \(x\)-spacing on the left grid \(h_{x,L}\). Figures 7.2a and 7.2b present the results of one simulation by plotting the contour lines both before and after the pulse has crossed the interfaces, which are represented by the hashed lines. Some distortion can be seen. Figures 7.2c and 7.2d show velocities \(v_x\) and \(v_y\), respectively, after the pulse has crossed the interface. We measure fourth-order convergence in the natural norm \(R_L\) on the left domain and in the natural norm \(R_R\) on the right, which is demonstrated in Figure 7.2e.

### 7.3 Euler Equations

We apply the methods presented in Chapter 6 to a two-dimensional vortex governed by the Euler equations. In the reference frame of the vortex, fluid rotates around a stationary local pressure minimum. In the reference frame used for this computation, the vortex advects with the mean flow to the right. After normalization, let \((x, y), t, \rho, v_x, v_y, E, \text{ and } p\) be the non-dimensional coordinates, time, density, \(x\)-velocity,
y-velocity, energy density, and pressure, respectively. With these choices, the Euler equations are

\[
A_x = \begin{bmatrix}
0 & 1 & 0 & 0 \\
\hat{\gamma}(v_x^2 + v_y^2) - v_x^2 & (3 - \gamma)v_x^2 & -\hat{\gamma}v_y & \hat{\gamma} \\
-v_xv_y & v_y & v_x & 0 \\
-\gamma \frac{\rho E}{\rho} + \hat{\gamma}v_x(v_x^2 + v_y^2) & -\gamma \frac{E}{\rho} - \frac{\hat{\gamma}}{2}(3v_x + v_y^2) & -\hat{\gamma}v_xv_y & \gamma v_x
\end{bmatrix},
\]

\[
A_y = \begin{bmatrix}
0 & 0 & 1 & 0 \\
0 & v_y & v_x & 0 \\
\hat{\gamma}(v_x^2 + v_y^2) - v_y^2 & (3 - \gamma)v_y^2 & \hat{\gamma} \\
-\gamma \frac{\rho E}{\rho} + \hat{\gamma}v_y(v_x^2 + v_y^2) & -\gamma \frac{E}{\rho} - \frac{\hat{\gamma}}{2}(v_x^2 + 3v_y^2) & \gamma v_y
\end{bmatrix},
\]

\[
u = \begin{bmatrix}
\rho, \rho v_x, \rho v_y, E
\end{bmatrix}^T,
\]

\[
\hat{\gamma} = \gamma - 1.
\]

The vortex is described by

\[
\rho = \left(1 - \frac{\epsilon^2}{8\pi \gamma} e^{-\beta^2 \tilde{r}^2}\right)^{\frac{1}{\gamma - 1}},
\]

\[
v_x = 1 - \frac{\epsilon}{2\pi} \tilde{y} e^{-\beta^2 \tilde{r}^2},
\]

\[
v_y = \frac{\epsilon}{2\pi} \tilde{x} e^{-\beta^2 \tilde{r}^2},
\]

\[
p = \rho \hat{\gamma},
\]

\[
E = \frac{\rho v_x^2}{\gamma - 1} + \frac{1}{2} \rho \left(v_x^2 + v_y^2\right),
\]

\[
\tilde{x} = x - x_0 - t,
\]

\[
\tilde{y} = y - y_0,
\]

\[
\tilde{r}^2 = \tilde{x}^2 + \tilde{y}^2,
\]

where the ratio of specific heats \(\gamma = 1.4\), the scaling factor \(\beta = 11\), and \(\epsilon = 1\).

We discretize the left and right subdomains with overlapping, uniform composite grids. For each simulation, the discretizations of both the left and right domains are defined by equations (2.2) with node counts given by the following table:

<table>
<thead>
<tr>
<th>Left</th>
<th>Right</th>
</tr>
</thead>
<tbody>
<tr>
<td>(n_x)</td>
<td>(n_y)</td>
</tr>
<tr>
<td>43</td>
<td>45</td>
</tr>
<tr>
<td>83</td>
<td>85</td>
</tr>
<tr>
<td>163</td>
<td>165</td>
</tr>
</tbody>
</table>

The leftmost column of the right discretization always lies halfway between the rightmost two columns of
the left discretization. The fourth-order GSBP pair presented in Appendix B.1 is used to approximate the
derivatives $\partial_x$ and $\partial_y$ on both the left and right domains. Fourth-order polynomial interpolation is used
at the interfaces of the two regions. Figure 7.3 presents the results of the third simulation by plotting the
pressure contours both before and after the vortex has crossed the interfaces, which are represented by the
hashed lines. Observe that circular contour lines are retained.

We next quantify the error in our method by measuring the errors at points away from boundaries and
interfaces. These three discretizations were chosen because each has nodes placed at $c_L = (-\frac{1}{2}, \frac{1}{2})$ and
$c_R = (\frac{1}{2}, \frac{1}{2})$. We investigate the error in the computed pressure,

$$\frac{p_{\text{num}} - p_{\text{exact}}}{\delta p},$$

where $p_{\text{num}}$ is the computed pressure, $p_{\text{exact}}$ is the pressure given by the analytical solution, and $1 - \delta p$ is
pressure at the vortex center. Figures 7.4a, 7.4b, and 7.4c show the scaled errors at $c_L$ and $c_R$ as functions
of time for each discretization. In Figure 7.4a, the increase in the magnitude of the error near time $t = \frac{1}{4}$
corresponds to discretization error as the vortex moves over $c_L$. Similarly in Figure 7.4c, the larger error
near $t = \frac{5}{4}$ corresponds to the discretization error as the vortex moves over $c_R$. All simulations show a peak
in the scaled error at $c_L$ and $t \approx 0.8716$. Measuring this peak shows the error reflected back into the domain
from the interface. Figure 7.4d shows that this error is $O \left( h_x^{3.9} \right)$, where $h_x$ is the $x$-spacing between adjacent
nodes in the left domain.
Figure 7.2: Numerical solution of linearized Euler equations (7.3) on overlapping domains. (a) Pressure at $t = 0$. (b) Pressure at $t = 0.5$. (c) Velocity in $x$-direction at $t = 0.5$. (d) Velocity in $y$-direction at $t = 0.5$. (e) Local convergence plot for problem (7.3) using method (5.31) with fourth-order Padé SBP pair, showing errors $\|e^L\|_{RL}$ and $\|e^R\|_{RR}$ as functions of mesh spacings $h_{x,L}$ and $h_{x,R}$, respectively.
Figure 7.3: Numerical solution of vortex problem (7.5) on overlapping domains. (a) Contour plot showing pressure before, during, and after vortex has crossed interface. (b) Same, showing vortex at interface.
Figure 7.4: Errors in numerical solution of vortex problem (7.5) on overlapping domains. (a) Scaled error \( \frac{p_{\text{num}} - p_{\text{exact}}}{\delta p} \) vs. time at \((-\frac{1}{2}, \frac{1}{2})\). (b) Magnification of (a). (c) Scaled error vs. time at \((\frac{1}{2}, \frac{1}{2})\). (d) Scaled error vs. mesh spacing at \((-\frac{1}{2}, \frac{1}{2})\) and peak reflection time \(t \approx 0.8716\).
Chapter 8

Toward Provably Stable Methods for Diffusion

This chapter begins laying the groundwork for stable methods for problems with diffusion. We will restrict our focus to a right-moving, diffusive wave on the interval \( \Omega = (a, b) \) that is described by

\[
\begin{align*}
\frac{du}{dt} &= \epsilon u_{xx} + \lambda u_x, \\
u(x, 0) &= g(x), \\
u(a, t) &= f_l(t), \\
u(b, t) &= f_r(t),
\end{align*}
\]

where \( \epsilon > 0 \) and \( \lambda < 0 \). The well-posedness and energy stability of this problem are discussed in [21] and [34], respectively.

8.1 Single Domain SBP-SAT Method

We begin with a method for problem (8.1) on the single uniform discretization (2.1). Let \( \{P, Q\} \) be a pair of SBP matrices. An SBP-based SAT method for problem (8.1) is

\[
\begin{align*}
\dot{v} &= \epsilon D^2 v + \lambda D v \\
&\quad + P^{-1} (\epsilon \sigma_1 D^T s + \lambda \tau s) (v_1 - f_l) + P^{-1} (\epsilon \sigma_2 D^T r) (v_n - f_r), \\
v(0) &= \hat{g}, \\
D &= P^{-1} Q, \\
s &= [1, 0, \ldots, 0, 0]^T, \\
r &= [0, 0, \ldots, 0, 1]^T.
\end{align*}
\]

If penalty parameters \( \sigma_1 \) and \( \sigma_2 \) assume certain values, then method (8.2) is energy stable, as the next theorem states:

**Theorem 8.1.1** Let \( \{P, Q\} \) be a pair of SBP matrices of approximation order \( p \). If \( \sigma_1 = +1, \sigma_2 = -1, \) and \( \tau \geq \frac{1}{2} \), then method (8.2) is energy stable in the \( P \)-norm and converges with order \( p \) to the true solution of
for any fixed time \( t_f > 0 \).

The proof follows from the fact that the system matrix of method (8.2) is the sum of the system matrices of methods (4.9) and (4.11).

## 8.2 Overlapping GSBP-SAT Method

We next move to overlapping subdomains. Consider the overlapping discretization defined by (5.1). Let \( \{P_L, Q_L\} \) be a pair of \( n \times n \) SBP matrices of approximation order \( p \) and \( \{P_R, Q_R\} \) be a pair of \( m \times m \) SBP matrices of approximation order \( q \). One possible SAT-SBP method for problem (8.1) on overlapping domains, which results from joining two copies method (8.2), is

\[
\dot{v}_L = \epsilon D^2_L v_L + \lambda D_L v_L + P^{-1}_L (\epsilon \sigma^2_L D^T_s L + \lambda \tau_L s_L)(v_L - f_l) + P^{-1}_L \epsilon \sigma_2^2 D^T r_L (v_n^L - \tilde{I} v_R),
\]

\[
\dot{v}_R = \epsilon D^2_R v_R + \lambda D_R v_R + P^{-1}_R (\epsilon \sigma^2_R D^T s_R + \lambda \tau_R s_R)(v_R - I v_L) + P^{-1}_R \epsilon \sigma_2^2 D^T r_R (v_m^R - f_r),
\]

\[
\dot{v}(0) = \hat{g},
\]

\[
D_{L/R} = P_{L/R}^{-1} Q_{L/R},
\]

\[
s_{L/R} = [1, 0, \ldots, 0, 0]^T,
\]

\[
r_{L/R} = [0, 0, \ldots, 0, 1]^T,
\]

where \( I \) and \( \tilde{I} \) are the interpolation matrices satisfying

\[
I \hat{u}_L(t) = u(a_R, t) + O(h_L^r),
\]

\[
\tilde{I} \hat{u}_R(t) = u(b_L, t) + O(h_R^r).
\]

We have been unable to prove that method (8.3) converges. Because the system matrix of method (8.3) lacks the block-triangular structure of the system matrix of method (5.8), we have been unable to prove that all eigenvalues of the system matrix of (8.3) lie in the left-half of the complex plane or construct a convergence proof similar to that of Theorem 5.1.3.

Lacking provable convergence and stability, we test method (8.3) numerically by applying it to IBVP (8.1). Initial and boundary data are defined by the solution to the Cauchy problem

\[
u(x, t) = \sqrt{\frac{\delta}{t+\delta}} e^{-\frac{(x+\lambda t+\frac{1}{2})^2}{4t(t+\delta)}},
\]

\[
\delta = (4\epsilon \sigma)^{-1},
\]

(8.5)
where diffusion constant $\epsilon = \frac{1}{100}$, wave velocity $\lambda = -1$, and scale factor $\sigma = 50$. Both $\{P_L, Q_L\}$ and $\{P_R, Q_R\}$ are fourth-order. Penalty parameters $\tau_L = \tau_R = 2$, $\sigma_L^I = \sigma_R^I = 1$, $\sigma_L^f = \sigma_R^f = -1$, and both $I$ and $\tilde{I}$ utilize fourth-order Lagrangian interpolation. Time integration was performed using the standard fourth-order Runge-Kutta method. The results are shown in Figure 8.1. Both domains show convergence with fourth-order.

![Figure 8.1: (a) Approximate solution to problem (8.1) using method (8.3) at times $t = 0, 0.5,$ and $1$. At $t = 0.5$, blue asterisks give solution on left subdomain and red squares give solution on right. (b) Local convergence plot for problem (8.1) using method (8.3) with fourth-order Padé SBP pairs, showing errors $\|e^L\|_{R_L}$ and $\|e^R\|_{R_R}$ as functions of mesh spacings $h_L$ and $h_R$, respectively. (c) Spectrum of system matrix of method (8.3) when diffusion constant $\epsilon = 1$. (d) Same, when $\epsilon = 10^{-4}$ and $\epsilon = 0$.](image)
8.3 Single Domain GSBP-SAT Method

Let \( \{P, Q, w\} \) be a GSBP system of approximation order \( p \). Assume that there is a matrix \( N \) such that if \( f : [a, b] \rightarrow \mathbb{R} \) is sufficiently differentiable then

\[
\left\| \hat{f}_{xx} - P^{-1}N\hat{f} \right\|_p = O(h^q). \tag{8.6}
\]

That is, \( P^{-1}N \) is an order \( q \) approximation to \( \partial_{xx} \). Further, assume that this second derivative operator obeys

\[
\langle v, P^{-1}Nv \rangle_p \leq v_n(\hat{D}v)_n - v_1(\hat{D}v)_1 \tag{8.7}
\]

for some \( \hat{D} \). Then a GSBP-SAT method for problem (8.1) on overlapping domains is

\[
\dot{v} = \epsilon D_2v + \lambda Dv + P^{-1}(\epsilon \sigma_1 \hat{D}^T s - \lambda w)(v_1 - f_i) + P^{-1}(\epsilon \sigma_2 \hat{D}^T r)(v_n - f_i)
\]

\[
v(0) = \hat{g},
\]

\[
D = P^{-1}Q, \quad D_2 = P^{-1}N,
\]

\[
s = [1, 0, \ldots, 0, 0]^T, \quad r = [0, 0, \ldots, 0, 1]^T.
\]

This method is energy stable and convergent:

**Theorem 8.3.1** Let \( \{P, Q, w\} \) be a GSBP system of order \( p \). Assume that \( P^{-1}N \) is an order \( q \) approximation and \( \hat{D} \) obeys (8.7). If \( \sigma_1 = +1 \) and \( \sigma_2 = -1 \), then method (8.8) is energy stable in the \( P \)-norm and converges with order \( p \) to the true solution of (8.1) for any fixed time \( t_f > 0 \).

Let \( G \) be the energy matrix associated with the GSBP system \( \{P, Q, w\} \). The proof follows from the fact that

\[
x^T P M x = \epsilon \left[ (\sigma_1 - 1)x_1(Dx)_1 + (\sigma_2 + 1)x_n(Dx)_n - \|Dx\|_p^2 \right] + |\lambda| x^T G x \leq 0.
\]

Appendix B.4 gives new matrices \( N \) and \( \hat{D} \) that obey (8.6) and (8.7) in the case that \( \{P, Q, w\} \) is given by Appendix B.3. Figure 8.2 shows the results when these matrices are used in conjunction with method (8.8) to solve problem (8.1). The system matrices were too stiff to permit the use of the standard fourth-order Runge-Kutta method. Instead, time integration was performed by MATLAB’s \texttt{ode23s}. While we
measure second-order convergence, the system matrix of method (8.8) is very stiff. In one simulation, shown in Figure 8.2a, in which the domain is discretized by a uniform discretization with 50 nodes, the eigenvalue with the most negative real part is about $-1.5 \times 10^8$. Moreover, Figure 8.2a shows that this operator can introduce distortion at the right boundary.

![Figure 8.2: Approximate solution to problem (8.1) using method (8.8) at times $t = 0, 0.5,$ and 1.](a)

![Figure 8.2: Local convergence plot for problem (8.1) using method (8.8) with GSBP matrices in Appendices B.3 and B.4, showing error $\|e\|_R$ as function of mesh spacing $h$.](b)

8.4 Overlapping GSBP-SAT Method

Again consider the overlapping discretization defined by (5.1). Let \{P_L, Q_L, w_L\} and \{P_R, Q_R, w_R\} be the GSBP systems of sizes $n$ and $m$, respectively. Let $\hat{D}_L$ and $N_L$ be the matrices of dimension $n$ obeying (8.6) and (8.7) with $P = P_L$. Similarly, let $\hat{D}_R$ and $N_R$ be the matrices of dimension $m$ obeying (8.6) and (8.7)
with \( P = P_R \). Joining two copies method (8.8) results in

\[
\begin{align*}
\dot{v}^L &= \epsilon D_{2,L}v^L + \lambda D_L v^L \\
&\quad + P_{L}^{-1} \left( \epsilon \sigma_1^L \hat{D}^T_L s_L - \lambda w_L \right) (v^L_1 - f_l) + P_{L}^{-1} (\epsilon \sigma_2^L \hat{D}^T_L r_L)(v^L_m - \tilde{I}v^R), \\
\dot{v}^R &= \epsilon D_{2,R}v^R + \lambda D_R v^R \\
&\quad + P_{R}^{-1} \left( \epsilon \sigma_1^R \hat{D}^T_R s_R - \lambda w_R \right) (v^R_1 - \tilde{I}v^L) + P_{R}^{-1} (\epsilon \sigma_2^R \hat{D}^T_R r_R)(v^R_m - f_r),
\end{align*}
\]

\( \dot{v}(0) = \hat{g} \),

\[
\begin{align*}
D_{L/R} &= P_{L/R}^{-1} Q_{L/R}, \\
D_{2,L/R} &= P_{L/R}^{-1} N_{L/R}, \\
s_{L/R} &= [1, 0, \ldots, 0, 0]^T, \\
r_{L/R} &= [0, 0, \ldots, 0, 1]^T,
\end{align*}
\]  

(8.9)

and \( I \) and \( \tilde{I} \) are the interpolation matrices satisfying (8.4).

We test method (8.3) numerically by applying it to IBVP (8.1) when initial and boundary data are defined by the solution to the Cauchy solution (8.5). GSBP systems \( \{P_L, Q_L, w_L\} \) and \( \{P_R, Q_R, w_R\} \) are given in Appendix B.3. Matrices \( N_L, \hat{D}_L, N_R, \) and \( \hat{D}_R \) are given in Appendix B.4. As in the previous section, \( \epsilon = \frac{1}{100}, \) wave velocity \( \lambda = -1, \) and scale factor \( \sigma = 50. \) Penalty parameters \( \sigma_1^L = \sigma_1^R = 1 \) and \( \sigma_2^L = \sigma_2^R = -1. \) Both \( I \) and \( \tilde{I} \) utilize second-order Lagrangian interpolation. The system matrix of method (8.9) is very stiff, as seen in Figure 8.3c. Time integration was performed by MATLAB’s ode23s. The results are shown in Figure 8.3. Figure 8.3a shows that the strong distortion occurs on left side of the boundary. The left domain shows convergence with order 2.5. The order of convergence on the right domain tends to 1 as interface terms begin to dominate.
Figure 8.3: (a) Approximate solution to problem (8.1) using method (8.9) at times $t = 0$, $0.5$, and $1$. At $t = 0.5$, blue asterisks give solution on left subdomain and red squares give solution on right. (b) Local convergence plot for problem (8.1) using method (8.9) with second-order explicit GSBP pairs and new second-derivative operators, showing errors $\| e^L \|_{RL}$ and $\| e^R \|_{RR}$ as functions of mesh spacings $h_L$ and $h_R$, respectively. (c) Spectrum of system matrix of method (8.9) and diffusion constant $\epsilon = 10^{-2}$. 
Chapter 9

Conclusions and Future Work

Overlapping domain decomposition methods are applicable to a wide range of problems but are prone to instability unless numerical diffusion or some other form of regularization is used. To address this, we derived high-order, stable, overlapping domain decomposition methods for hyperbolic initial-boundary value problems that do not require artificial dissipation. We derived two classes of methods. In Chapter 5, we constructed Lax stable methods, which are under certain assumptions also strongly pointwise stable. The key attributes of these methods are that interpolation transferred information from the upwind domain to the downwind domain, but not from the downwind to the upwind, and that the downwind domain is strongly stable. We used this technique to create methods in both one and two dimensions, whose stability requirements are stated in Theorems 5.1.3 and 5.2.4, respectively. We proved that the eigenvalues of the system matrices had nonpositive real parts and empirically found that they had negative real parts, but we were unable to find a norm in which the methods were energy stable.

Next, in Chapter 6, we constructed provably energy stable overlapping domain decomposition methods by utilizing GSBP derivative operators. Theorem 6.1.1 shows that high-order methods can be constructed by using high-order GSBP derivative operators and states the conditions that these operators must satisfy. Appendix B presents new Padé GSBP derivative operators that meet these criteria and that can be used to construct fourth-order and sixth-order methods, respectively. Because the methods are energy stable, the spectra of the system matrices lie in the left half of the complex plane.

Future work includes the following tasks, which are presented here in ascending order of perceived difficulty:

- **Find a GSBP system \( \{ P, Q, w \} \) for which the lower symmetric block of \( Q \) is positive definite and larger than \( 1 \times 1 \) and for which \( P = P^\dagger \).** Currently, any theoretical results describing GSBP-methods in two dimensions, even on a single domain, are limited to the case where \( A_x \) and \( A_y \) are simultaneously diagonalizable. Let \( \hat{\Omega} \) be a uniform discretization of \( \Omega = (a_x, b_x) \times (a_y, b_y) \) with \( n_x \) nodes in the \( x \)-direction and \( n_y \) nodes in the \( y \)-direction. Let \( \{ P_x, Q_x, w_x \} \) and \( \{ P_y, Q_y \} \) \( w_y \) be GSBP systems of size \( n_x \) and \( n_y \), respectively. Consider problem (2.11). Extending GSBP matrices on a single domain...
discretization gives rise to the system matrix

\[ M = (D_x \otimes I_{n_y} \otimes A_x^-) + (\tilde{D}_x \otimes I_{n_y} \otimes A_x^+) + (I_{n_x} \otimes D_y \otimes A_y^-) + (I_{n_x} \otimes \tilde{D}_y \otimes A_y^+), \]

where \( D_x \) and \( \tilde{D}_x \) are defined by (6.11) with \( P = P_x \) and \( Q = Q_x \), and \( D_y \) and \( \tilde{D}_y \) are defined by (6.11) with \( P = P_y \) and \( Q = Q_y \). If \( A_x \) and \( A_y \) are real symmetric and simultaneously diagonalizable, we can show energy stability. Let \( T \) be any orthogonal matrix that diagonalizes both \( A_x = T \Lambda_x T^T \) and \( A_y = T \Lambda_y T^T \). In this case, \( M \) is negative definite in the \( H \)-norm, where

\[ H = \left( P_x \otimes P_y \otimes T \Lambda_x \right) \left( N_{\Lambda_x} \right)^{-1} \left( N_{\Lambda_y} \right)^{-1} \left( P_x \otimes P_y \otimes T \Lambda_y \right)^T + \left( P_x \otimes P_y \otimes T \Lambda_x \right) \left( N_{\Lambda_x} \right)^{-1} \left( N_{\Lambda_y} \right)^{-1} \left( P_x \otimes P_y \otimes T \Lambda_y \right)^T. \]

If \( A_x \) and \( A_y \) are not simultaneously diagonalizable, then \( M \) is not negative semidefinite in the above norm. However, if there exists a GSBP pair for which \( P = P^t \), then \( M \) is negative semidefinite in the \( \tilde{H} \)-norm, where \( \tilde{H} = (P_x \otimes P_y \otimes I_k) \).

- **Show method** (5.31) **is stable when** \( A_y \) **and** \( A_y \) **are not simultaneously diagonalizable.** We believe this to be true based on promising numerical results given in Section 7.2 and the fact that (5.31) is a stable and convergent method when \( A_x \) and \( A_y \) are simultaneously diagonalizable. However, none of our proof techniques extend to the case.

- **Prove or disprove Conjecture 3.6 in [87].** Pointwise stability of the single domain SAT-SBP method (4.9) is often assumed, but it has so far resisted proof.

- **Find a local stability that is sufficient for global stability.** All domain decomposition methods constructed using the SAT methodology are, at their core, a collection of interrelated single domain, local SAT methods. As such, the stability of these local SAT methods is necessary, but not sufficient, for global stability. The proofs of global stability can be complicated even for simple problems. For example, in [11], a stable method for the one-dimensional advection-diffusion equation on two non-overlapping subdomains containing eighteen penalty parameters is considered. Is there a form of local stability such that an overlapping domain decomposition method that comprises methods with this form of local stability is globally stable?
Appendix A

GSBP4

GSBP4 is a pair of GSBP matrices that are a locally fourth-order approximation to the first derivative on the interior and third-order at the boundary [15]. Note the factor of $h$ multiplying $P$:

$$P = h \begin{bmatrix}
\frac{79}{288} & \frac{11}{36} & -\frac{25}{288} & \frac{5}{288} \\
\frac{11}{36} & \frac{13}{9} & \frac{35}{288} & \frac{5}{144} \\
-\frac{25}{288} & \frac{35}{288} & \frac{387}{288} & \frac{1}{4} \\
\frac{5}{288} & \frac{5}{144} & \frac{1}{3} & 1 & \frac{1}{4} \\
& \frac{1}{4} & 1 & \frac{1}{4} & \ddots \ddots \\
& & \frac{1}{4} & 1 & \frac{1}{4} \\
& & & \frac{1}{4} & 1 & \frac{1}{4} & \frac{1}{36} & \frac{1}{96} \\
& & & & \frac{1}{4} & 287 & \frac{53}{288} & -\frac{13}{288} \\
& & & & & \frac{1}{36} & 53 & \frac{10}{9} & \frac{7}{36} \\
& & & & & & \frac{1}{36} & -\frac{13}{288} & \frac{7}{96} & \frac{47}{288}
\end{bmatrix}$$
\[
Q = \begin{bmatrix}
-\frac{5}{8} & 451 & -\frac{29}{144} & 25 & \frac{5}{576} \\
-\frac{595}{576} & \frac{1}{8} & \frac{181}{192} & -\frac{5}{144} \\
\frac{29}{144} & -\frac{181}{192} & 0 & \frac{427}{576} \\
-\frac{25}{576} & \frac{5}{144} & -\frac{427}{576} & 0 & \frac{3}{4} \\
-\frac{3}{4} & 0 & \frac{3}{4} \\
\ldots & \ldots & \ldots & \ldots \\
-\frac{3}{4} & 0 & \frac{3}{4} \\
-\frac{3}{4} & 0 & \frac{143}{192} & -\frac{1}{48} & \frac{5}{48} \\
-\frac{143}{192} & 0 & \frac{163}{192} & -\frac{5}{48} \\
\frac{1}{48} & -\frac{163}{192} & \frac{1}{8} & \frac{45}{64} \\
-\frac{5}{192} & \frac{5}{48} & -\frac{29}{64} & 3 & \frac{3}{8} \\
\end{bmatrix}
\]

Explicit calculation reveals that \( Q + Q^T = \text{block diag} \left( \tilde{Q}_u, 0, \ldots, 0, \tilde{Q}_l \right) \), where

\[
\tilde{Q}_u = \begin{bmatrix}
-\frac{5}{4} & -\frac{1}{4} \\
-\frac{1}{4} & \frac{1}{4} \\
\end{bmatrix},
\]

\[
\tilde{Q}_l = \begin{bmatrix}
\frac{1}{4} & \frac{1}{4} \\
\frac{1}{4} & \frac{3}{4} \\
\end{bmatrix}.
\]

Define the penalty vector

\[
w = \left[ -\frac{5}{8}, -\frac{1}{4}, 0, \ldots, 0 \right]^T.
\]

If \( \tau \geq 1 \), then \( \{P, Q, w\} \) is a GSBP system. All conditions in this thesis are satisfied if \( \tau > 1 \).
Appendix B

Matrices for Overlapping Domain Decomposition

B.1 Low-Order Implicit GSBP

We present one possible procedure for constructing a pair of GSBP matrices \( \{P, Q\} \) that is a locally third-order approximation to \( \partial_x \) at the boundaries and fourth-order on the interior. The algorithm has three high-level steps:

(a) Choose a Padé central difference approximation.

(b) Select the sizes of the left and right boundaries.

(c) Construct the boundary blocks of \( P \) and \( Q \).

The problem of determining a GSBP pair for which \( P^{-1}Q \) approximates the derivatives with a given order of accuracy does not have a unique solution. We construct a solution by restricting ourselves to GSBP pairs in which both matrices are banded with corrections to the upper and lower diagonal blocks. The coefficients of the Padé scheme chosen in (a) determine the entries in the bands, and the sizes of the boundary blocks selected in (b) determine the sizes of the corrections. This three-step procedure is similar to that followed by other authors [10, 12, 15, 84], but we complete step (c) in a new way. Whereas other authors analytically solve large systems, we use numerical optimization to improve an initial guess iteratively, in conjunction with a perturbation step that ensures that the pair \( \{P, Q\} \) satisfy in exact arithmetic a set of consistency constraints, which are defined later.

The first step is to select an appropriate Padé approximation. We desire \( P^{-1}Q \) to be a fourth-order approximation on the interior. In order to satisfy this requirement, we select the fourth-order Padé approximation

\[
\frac{h}{4} f'(x_{i+1}) + hf'(x_i) + \frac{h}{4} f'(x_{i-1}) \approx \frac{3}{4} f(x_{i+1}) - \frac{3}{4} f(x_{i-1}),
\]

where \( x_i \) and \( h \) are defined by (2.1). For simplicity, we have used the same Padé approximation used in the construction of GSBP4, but we emphasize that this is just one of many possible fourth-order choices [45].
Next, the sizes of the boundary blocks must be chosen. We choose all boundary blocks to be $4 \times 4$. Again, one could use differently sized blocks.

The choices of Padé approximation and sizes of the boundary blocks assign the following structures to $P$ and $Q$,

\[
P = h \begin{bmatrix}
p_1 & p_2 & p_3 & p_4 \\
p_5 & p_6 & p_7 \\
p_3 & p_6 & p_8 & p_9 \\
p_4 & p_7 & p_9 & p_{10} & \frac{1}{4} \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
\frac{1}{4} & 1 & \frac{1}{4} & \frac{1}{4} & p_{11} & p_{12} & p_{13} & p_{14} \\
p_{12} & p_{15} & p_{16} & p_{17} \\
p_{13} & p_{16} & p_{18} & p_{19} \\
p_{14} & p_{17} & p_{19} & p_{20}
\end{bmatrix}
\]
$Q = \begin{bmatrix} q_1 & q_2 & q_3 & q_4 \\ q_5 & q_6 & q_7 & q_8 \\ q_9 & q_{10} & q_{11} & q_{12} \\ q_{13} & q_{14} & q_{15} & \frac{3}{4} \\ \frac{3}{4} & 0 & \frac{3}{4} \\ \vdots & \vdots & \vdots \\ -\frac{3}{4} & 0 & \frac{3}{4} \\ -\frac{3}{4} & q_{17} & q_{18} & q_{19} & q_{20} \\ q_{21} & q_{22} & q_{23} & q_{24} \\ q_{25} & q_{26} & q_{27} & q_{28} \\ q_{29} & q_{30} & q_{31} & q_{32} \end{bmatrix},$

where the $p_i$ are unknowns in $P$ and $q_i$ in $Q$. Observe that the coefficients of the Padé approximation (B.1) have become the values in the bands of $P$ and $Q$. Also note that the sizes of the boundary blocks determine the number of unknowns.

We now solve for these unknowns. Define

$$
\hat{P} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{12} & p_{15} & p_{16} & p_{17} \\ p_{13} & p_{16} & p_{18} & p_{19} \\ p_{14} & p_{17} & p_{19} & p_{20} \end{bmatrix}, \quad \hat{Q} = \begin{bmatrix} q_{17} & q_{18} & q_{19} & q_{20} \\ q_{21} & q_{22} & q_{23} & q_{24} \\ q_{25} & q_{26} & q_{27} & q_{28} \\ q_{29} & q_{30} & q_{31} & q_{32} \end{bmatrix}.
$$
\( P^{-1}Q \) being a third-order approximation to the derivative at the right boundary is equivalent to enforcing that

\[
\begin{bmatrix}
0 & 1 & 0 & 0 \\
0 & 1 & 2 & 3 \\
0 & 1 & 4 & 12 \\
0 & 1 & 6 & 27
\end{bmatrix}
+ \frac{1}{4}
\begin{bmatrix}
0 & 1 & -2 & 3 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0
\end{bmatrix}
= \begin{bmatrix}
1 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 \\
1 & 2 & 4 & 8 \\
1 & 3 & 9 & 27
\end{bmatrix}
+ \frac{3}{4}
\begin{bmatrix}
1 & -1 & 1 & -1 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0
\end{bmatrix},
\]

which can be rewritten

\[
\hat{Q} = \hat{P}
\begin{bmatrix}
-\frac{11}{6} & 3 & -\frac{3}{2} & \frac{1}{3} \\
-\frac{1}{3} & -\frac{1}{2} & 1 & -\frac{1}{6} \\
\frac{1}{6} & -1 & 1 & \frac{1}{3} \\
-\frac{1}{3} & \frac{3}{2} & -3 & \frac{11}{9}
\end{bmatrix}
+ \begin{bmatrix}
\frac{24}{12} & -\frac{17}{5} & \frac{5}{4} & -\frac{7}{24} \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0
\end{bmatrix}.
\]

Note that the choice of \( \hat{P} \) in equation (B.2) uniquely determines \( \hat{Q} \).

So far, we have proceeded as in [15]. Next we obtain a \( \hat{P} \) such that both \( \hat{P} \) and \( (\hat{Q} + \hat{Q}^T) \) are symmetric positive definite. Work by other authors places more restrictions on system (B.2), uses symbolic math software to determine the number of degrees of freedom, and then gives choices for these degrees of freedom that result in \( \hat{P} \) and \( \hat{Q} \) satisfying the desired properties. Instead, we use a numerical optimizer to improve iteratively a random initial guess \( \hat{P}_0 \). To do this, we create the formal maximization problem

\[
\hat{P}_{\text{trial}} = \arg \max_{\text{s.p.d. } C} \left[ \min \lambda(D + D^T) \right],
\]

where \( C \) is a 4 \( \times \) 4 real symmetric, positive definite matrix, \( D \) is defined by letting \( \hat{Q} \to D \) and \( \hat{P} \to C \) in equation (B.2), and \( \lambda(X) \) is the spectrum of \( X \). After solving (B.3) with MATLAB’s optimizer using the random initial guess \( \hat{P}_0 \), we obtain the solution

\[
\hat{P}_{\text{trial}} \approx
\begin{bmatrix}
0.7181 & -0.1581 & 0.7142 & -0.2963 \\
-0.1581 & 4.4364 & -5.2697 & 1.8557 \\
0.7142 & -5.2697 & 9.3293 & -3.7795 \\
-0.2963 & 1.8557 & -3.7795 & 2.623
\end{bmatrix}.
\]
The elements of $\hat{P}$ are defined to be rational approximations of the corresponding elements of $\hat{P}_{\text{trial}},$

\[
\hat{P} = \begin{bmatrix}
163 & -46 & 927 & -8 & 27 \\
227 & 291 & 1298 & 8 & 27 \\
-46 & 244 & -469 & 180 & 97 \\
291 & 55 & 89 & 85 & 68 \\
927 & -469 & 793 & 257 & 160 \\
1298 & 89 & 85 & 257 & 68 \\
-8 & 180 & -257 & 160 & 61 \\
27 & 97 & 68 & 61 & \\
\end{bmatrix}.
\]

Matrix $\hat{Q}$ is calculated according to equation (B.2), yielding

\[
\hat{Q} = \begin{bmatrix}
1007857804 & 1080757783 & 324312701 & 3066198755 \\
1157516811 & 1029903832 & 257225958 & 32590836 \\
-2904587 & 2545369 & -3350641 & 3616697 \\
8735503 & 474819 & 749830 & 117832440 \\
2678010 & 3927480 & 4969685 & 318475391 \\
-102958735 & 21354653 & 54017993 & 102364499 \\
65181672 & 3621204 & 7242408 & 32908536
\end{bmatrix}.
\]

Note that $\hat{P}$ and $(\hat{Q} + \hat{Q}^T)$ are positive definite,

\[
\lambda(\hat{P}) \approx \{0.500, 0.781, 1.578, 14.248\},
\]

\[
\lambda(\hat{Q} + \hat{Q}^T) \approx \{0.258, 0.258, 1.784, 35.7689\}.
\]

It is not necessary perform an analogous procedure for the upper blocks. Because the upper blocks of GSBP4 satisfy the necessary requirements, we assign the upper blocks of $P$ and $Q$ to be the corresponding values.
of the GSBP4 matrices. The final matrices are

\[ P = h \]

\[ \begin{bmatrix}
    \frac{79}{288} & \frac{11}{36} & -\frac{25}{288} & \frac{5}{288} \\
    \frac{11}{36} & \frac{13}{9} & \frac{35}{288} & \frac{5}{144} \\
    -\frac{25}{288} & \frac{35}{288} & \frac{287}{288} & \frac{1}{4} \\
    \frac{5}{288} & \frac{5}{144} & \frac{1}{4} & \frac{1}{4} \\
    \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\
    \frac{1}{4} & \frac{163}{227} & -\frac{46}{291} & \frac{927}{1298} & -\frac{8}{27} \\
    -\frac{46}{291} & \frac{244}{55} & -\frac{469}{89} & \frac{180}{97} \\
    \frac{927}{1298} & -\frac{469}{89} & \frac{793}{85} & -\frac{257}{68} \\
    -\frac{8}{27} & \frac{180}{97} & -\frac{257}{68} & \frac{160}{61}
\end{bmatrix} \]
Define the penalty vector
\[ \mathbf{w} = \begin{bmatrix} -\tau \frac{5}{8}, -\frac{1}{4}, 0, \ldots, 0 \end{bmatrix}^T. \]

If \( \tau \geq 1 \), then \( \{ P, Q, \mathbf{w} \} \) is a GSBP system. All conditions in this thesis are satisfied if \( \tau > 1 \).

We next present a Fourier analysis of the discrete derivative operator \( P^{-1}Q \) defined by the above matrices. This type of analysis provides an effective way to quantify the resolving characteristics of a finite difference operator. Introductions to this topic as well as further reading can be found in [27, 93]. We follow [45] in the details of our analysis.

A uniform discretization of the domain \( \Omega = [0, 2\pi) \) is the set \( n \) equally spaced points
\[ h = \frac{2\pi}{n}, \quad x_i = (i-1)h, \quad i = 1, \ldots, n. \] (B.4)

Let \( f : \Omega \to \mathbb{R} \). The vector \( \hat{f} \), called the injection of \( f \), is defined by
\[ \hat{f} = [f(x_1), \ldots, f(x_n)]^T. \]
The Fourier basis is the set of functions \( \{ g_k(x) \}_{k \in \mathbb{Z}} \), where

\[ g_k(x) = \exp(i k x), \]

and \( i = \sqrt{-1} \). The elements of \( \hat{f} \) can be written in terms of the Fourier basis according to

\[ \hat{f}_j = \sum_{k=L}^{U} \tilde{f}_k g_k(x_j), \]

where, if \( n \) is odd then \( L = -\frac{(n-1)}{2} \) and \( U = -L \), and if \( n \) is even then \( L = -\frac{n}{2} + 1 \) and \( U = -L + 1 \). The wavenumber \( w \in [\pi, \pi] \) is a scaled index defined by \( w = \frac{2\pi k}{n} \). For convenience, we define \( g_w = g_{k(w)} \). Note that

\[ \frac{\partial g_w}{\partial x} \bigg|_{x=x_j} = i w h g_w(x_j). \]

An approximate first derivative \( D \in \mathbb{R}^{n \times n} \) defines a modified wavenumber \( w'_j(w) \) by

\[ \{ D \tilde{g}_w \}_j = i \frac{w'_j}{h} g_w(x_j). \]

The exact differentiation operator \( \partial_x \) corresponds to \( w'_j(w) = w \) for all \( j \). The value \( \text{Re}(w') - w \) measures *dispersive error*, and \( \text{Im}(w') \) measures *dissipative error* [45]. Figure B.1 plots the modified wavenumber \( w'_j(w) \) vs. the wavenumber \( w \) for the discrete derivative operator \( P^{-1}Q \) at the boundary point \( x_{49} \) and the near-boundary point \( x_{39} \). Figure B.1a shows that the GSBP modifications introduce dissipative error, while Figure B.1b shows that this error has diminished in the interior points.

In the next sections, we present two new pairs of GSBP matrices that can also be used to create energy stable methods for overlapping subdomains. The same procedure is followed except that different Padé approximations and sizes of boundary blocks are used.
B.2 High-Order Implicit GSBP

The following pair of GSBP matrices can be used to construct an approximation to the first derivative that is locally sixth-order on the interior, fifth-order at the boundaries:

\[
\begin{bmatrix}
6965509 & 30563 & 281 & 51007 & 19075 & 19273 \\
27993600 & 77760 & 1296 & 349920 & 373248 & 2342800 \\
30563 & 10875041 & 4159 & 171491 & 209207 & 45253 \\
77760 & 5598720 & 13552 & 466560 & 1399680 & 1866240 \\
281 & 4159 & 1786169 & 10219 & 27791 & 7471 \\
1296 & 13552 & 1399680 & 466560 & 1399680 & 699840 \\
51007 & 171491 & 10219 & 1382789 & 1 & 0 \\
349920 & 466560 & 466560 & 1399680 & 3 & 0 \\
-19075 & 209207 & 27791 & 5603021 & 3 & 1 \\
373248 & 1399680 & 466560 & 1399680 & 3 & 0 \\
19273 & 45253 & 7471 & 27992569 & 1 & 3 \\
2342800 & 1866240 & 699840 & 27993600 & 3 & 1 \\
\end{bmatrix}
\]

\[
P = h
\]
\[ Q = \begin{bmatrix}
\frac{1}{3} & 1 & \frac{1}{3} \\
\frac{1}{3} & 298 & 34 & 27 & \frac{1}{3} & 13 & 2 & \frac{1}{3} \\
\frac{1}{3} & 34 & 91 & 28 & \frac{1}{3} & 137 & 2 & \frac{1}{3} \\
\frac{1}{3} & 91 & 28 & 29 & \frac{1}{3} & 103 & 3 & \frac{1}{3} \\
\frac{1}{3} & 298 & 34 & 13 & \frac{1}{3} & 259 & 2 & \frac{1}{3} \\
\frac{1}{3} & 34 & 91 & \frac{1}{3} & 13 & 2 & \frac{1}{3} \\
\frac{1}{3} & 28 & \frac{1}{3} & 103 & 2 & \frac{1}{3} \\
\frac{1}{3} & 137 & 2 & \frac{1}{3} & 137 & 2 & \frac{1}{3} \\
\end{bmatrix} \]

Define the penalty vector
\[ w = \left[ -\frac{2}{3}, -\frac{1}{3}, 0, \ldots, 0 \right]^T. \]

If \( \tau \geq 1 \), then \( \{P, Q, w\} \) is a GSBP system. All conditions in this thesis are satisfied if \( \tau > 1 \).

Figure B.2 plots the modified wavenumber vs. the wavenumber for the discrete derivative operator defined
by $P^{-1}Q$. Figure B.2a shows that this operator introduces dissipative error near the boundary, while Figure B.2b shows that the dissipative error is much smaller in the interior.

Figure B.2: (a) Modified wave number vs. wave number at boundary point $x_{49}$. (b) Same at near-boundary point $x_{39}$.
B.3 Explicit GSBP

This final pair of GSBP matrices can be used to construct an approximation to the first derivative that is locally fourth-order on the interior, second-order at the boundaries:

\[
P = h \text{ diag} \left( \frac{17}{48}, \frac{59}{48}, \frac{43}{48}, \frac{49}{48}, 1, \ldots, 1, \frac{7}{8}, \frac{1}{2}, \frac{1}{8} \right),
\]

\[
Q = \begin{bmatrix}
-\frac{1}{2} & \frac{59}{96} & -\frac{1}{12} & -\frac{1}{32} \\
-\frac{59}{96} & 0 & \frac{59}{96} & 0 \\
\frac{1}{12} & -\frac{59}{96} & 0 & -\frac{1}{12} \\
\frac{1}{32} & 0 & -\frac{59}{96} & 0 & \frac{2}{3} & -\frac{1}{12} \\
& & \frac{1}{12} & -\frac{2}{3} & 0 & \frac{2}{3} & -\frac{1}{12} \\
& & & \ddots & \ddots & \ddots & \ddots & \ddots \\
& & & & \frac{1}{12} & -\frac{2}{3} & 0 & \frac{2}{3} & -\frac{1}{12} \\
& & & & & \frac{1}{12} & -\frac{2}{3} & \frac{3}{16} & \frac{5}{12} & -\frac{1}{48} \\
& & & & & & \frac{1}{12} & -\frac{1}{2} & \frac{1}{4} & \frac{1}{8} \\
& & & & & & & \frac{1}{16} & -\frac{1}{4} & \frac{3}{16}
\end{bmatrix}.
\]

Define the penalty vector

\[
w = \left[ -\frac{1}{2}, 0, \ldots, 0 \right]^T.
\]

If \(\tau \geq 1\), then \(\{P, Q, w\}\) is a GSBP system. All conditions in this thesis are satisfied if \(\tau > 1\).

Figure B.3 plots the modified wavenumber vs. the wavenumber for the discrete derivative operator defined by \(P^{-1}Q\). Figure B.3a shows that this operator introduces dissipative error near the boundary. The derivative operator \(P^{-1}Q\) approximates the first derivative \(\partial_x\) in the interior using a fourth-order centered finite-difference approximation. Such approximations do not introduce dissipative error at the interior points. Figure B.3b confirms this. Note that \(\text{Im}(w_3'(w)) = 0\) for all \(w\).
**Figure B.3:** (a) Modified wave number vs. wave number at boundary point $x_{49}$. (b) Same at near-boundary point $x_{39}$.

### B.4 Second Derivative

Let $\{P, Q, w\}$ be the GSBP system from Appendix B.3. In Section 8.3, first and second spatial derivatives are approximated by finite difference operators $P^{-1}Q$ and $P^{-1}N$, respectively, where

$$ R = \frac{1}{\pi} \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & -\frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & -\frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & -\frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix} $$

For all vectors $v$ of appropriate dimension,

$$ \left\langle v, P^{-1}Nv \right\rangle_P \leq v_n(\hat{D}v)_n - v_1(\hat{D}v)_1, $$
where

\[ \hat{D} = \frac{1}{h} \begin{bmatrix} -1 & 1 \\ \frac{1}{2} & -2 & \frac{3}{2} \end{bmatrix}. \]
Appendix C

SBP Pairs and Natural Norms

Consider the open interval $\Omega = (a, b)$, and let $\hat{\Omega}$ be a uniform discretization with $n$ nodes and associated natural norm $R$. Let $p_i = x^i$ be the $i$th monomial basis function. Let $\hat{p}_i$ be the injection of $p_i$. A finite difference operator is said to be exact for all polynomials of degree $r$ if $D\hat{p}_0 = 0$ and $D\hat{p}_i = i\hat{p}_{i-1}$ for all $i \in [1, \ldots, r]$. If a finite difference operator is exact for all polynomials of degree $r$ and $f$ is a sufficiently differentiable function, then

$$\|\hat{f}_x - Df\|_R = O(h^r).$$

The easiest way to ensure that a pair of $\{P, Q\}$ is of approximation order $r$ is to require $P^{-1}Q$ to be exact for all polynomials of degree $r$. In [42], the original paper introducing SBP matrices, the authors show that there exists no SBP pair $\{P, Q\}$ such that $P^{-1}Q$ is exact for polynomials of degree 2 and $P$ is the natural norm. We offer a shorter proof of a slightly stronger theorem:

**Theorem C.1.1** There exists no pair of SBP matrices $\{P, Q\}$ that is exact for polynomials of degree 1 and for which $P = hI$.

Let $\Omega = (0, 1)$ and $\hat{\Omega}$ be a uniform discretization of $\Omega$ with $n$ nodes and spacing $h$ defined by (2.1). Note that

$$hn = \frac{n}{n-1} > 1.$$

Assume that there exists an SBP pair $\{P, Q\}$ that is exact for all polynomials of degree 1 and for which $P = hI$. We will evaluate the same inner product in two different ways and show a contradiction. Observe that

$$\langle \hat{p}_1, P^{-1}Q\hat{p}_0 \rangle_P = 0.$$
On the other hand,

\[
\langle \hat{p}_1, P^{-1} Q \hat{p}_0 \rangle_p = (\hat{p}_0)_n (\hat{p}_1)_n - (\hat{p}_0)_1 (\hat{p}_1)_1 - \langle P^{-1} Q \hat{p}_1, \hat{p}_0 \rangle_p \\
= 1 - \langle \hat{p}_0, \hat{p}_0 \rangle_p \\
= 1 - n h \\
< 0.
\]

To our knowledge, all authors construct high order SBP pairs \( \{ P, Q \} \) by enforcing that \( P^{-1} Q \) is exact for polynomials of a high degree and must, as a consequence, consider complicated \( P \)-norms. It is possible that simpler norms could be derived by considering other forms of differentiation.
Appendix D

Better Bound

In Theorem 6.1.1, we showed that if $H = \text{block}_\text{diag}(P_L, \alpha P_R)$ and $\alpha \leq \alpha^*$, where $\alpha^*$ is defined by (6.3), then method (6.1) is energy stable in the $H$-norm. In this section, we show that the hypotheses of Theorem 6.1.1 are sufficient but not necessary. Specifically, $\alpha$ may be larger than $\alpha^*$ and the method can still be energy stable. Consider again the example GSBP pair and interpolation operator from Section 6.2.

**Lemma D.1.1** If $\{P_R, Q_R\}$ and $\{P_L, Q_L\}$ are both GSBP4, $\tau_L \geq 1$, $\alpha = \frac{1}{15} > \alpha^*$, and $\tau_R = 2$, then (6.1) is energy stable in the $H$-norm for all overlaps satisfying (6.7).

Let $M$ be the system matrix associated with method (6.1). In order to show that $M$ is negative semidefinite in the $H$-norm, we will examine the eigenvalues of $(HM + M^T H)$. By explicit computation, observe that

\[
HM + M^T H = \text{block}_\text{diag} (N_u, 0, \ldots, 0, N_m, 0, \ldots, 0, N_l),
\]

where

\[
N_u = \begin{bmatrix}
2(1 - \tau_L) \{Q_L\}_{1,1} & 0 \\
0 & -2 \{Q_L\}_{2,2}
\end{bmatrix},
\]

\[
N_m = \begin{bmatrix}
-2 \{Q_L\}_{n-1,n-1} & -\delta^L_{n-1} & -\alpha(1 - \theta)\tau_R \{Q_R\}_{1,1} & -\alpha(1 - \theta)\delta^R_l \\
-\delta^L_{n-1} & -2 \{Q_L\}_{n,n} & -\alpha\theta\tau_R \{Q_R\}_{1,1} & -\alpha\theta\delta^R_l \\
-\alpha(1 - \theta)\tau_R \{Q_R\}_{1,1} & \theta\tau_R \{Q_R\}_{1,1} & 2\alpha(1 - \tau_R) \{Q_R\}_{1,1} & 0 \\
-\alpha(1 - \theta)\delta^R_l & -\alpha\theta\delta^R_l & 0 & -2\alpha \{Q_R\}_{2,2}
\end{bmatrix},
\]

\[
N_l = \alpha \begin{bmatrix}
-2 \{Q_R\}_{m-1,m-1} & -\delta^R_{m-1} \\
-\delta^R_{m-1} & -2 \{Q_R\}_{m,m}
\end{bmatrix},
\]

\[
\delta^L_i/R = \{Q_L/R\}_{i,i+1} + \{Q_L/R\}_{i+1,i},
\]

\[
\theta = 1 - \frac{a_R - b_L}{h_L}.
\]

Note that (6.7) implies that $\theta \in [0, 1]$. Observe that the spectrum of $(HM + M^T H)$ is equal to the union of the spectra of $N_l$, $N_m$, and $N_u$. That all eigenvalues of $N_l$ and $N_u$ are non-positive is shown in [15]. To
show that the spectrum of $N_m$ is negative definite for all overlaps satisfying (6.7), we compute the maximum eigenvalue of $N_m$ and show that it is negative. The result is shown in Figure D.1a. Note that the largest eigenvalue is always negative.

We stress that the choice $\alpha = \frac{1}{15}$ and $\tau_R = 2$ is somewhat arbitrary. There is a region of possible choices that has been illustrated by plotting the stability regions of different choices of $\theta$ in Figure D.1b. If a point $(\tau_R^*, \alpha^*)$ is below the curve marked $\theta^*$, then $N_m$ in (D.1) with $\alpha = \alpha^*$, $\tau_R = \tau_R^*$, and $\theta = \theta^*$ is negative semidefinite. The black diamond shows that the particular choice made in Lemma D.1.1 and that this choice is inside the stability region for the five values of $\theta$ shown. Figure D.1a shows that this point is inside the stability region for all $\theta \in [0, 1]$.

Figure D.1: (a) Maximum eigenvalue of $N_m$ with $\alpha = \frac{1}{15}$ and $\tau_R = 2$ for GSBP4 matrices. (b) Upper boundaries for GSBP4 energy stability regions for five values of $\theta$. 

Figure D.1: (a) Maximum eigenvalue of $N_m$ with $\alpha = \frac{1}{15}$ and $\tau_R = 2$ for GSBP4 matrices. (b) Upper boundaries for GSBP4 energy stability regions for five values of $\theta$. 

104
Appendix E

Sobolev Bound

Consider the following Sobolev inequality theorems, both of which are proven in [27]:

**Theorem E.1.1** If \( u \) is a continuously differentiable function on the closed interval \([0, l]\), then for every \( \epsilon > 0 \)
\[
\| u \|_\infty^2 \leq (\epsilon^{-1} + l^{-1}) \| u \|_2^2 + \epsilon \| u_x \|_2^2,
\]  
where \( \| f \|_2^2 = \int_0^l f(x)^2 \, dx \) and \( \| f \|_\infty^2 = \max_{x \in [0, l]} f(x)^2 \).

**Theorem E.1.2** Let \( \hat{\Omega} \) be the uniform discretization of \( \Omega = (0, l) \) defined by (2.1) with \( n \) nodes, step spacing \( h = \frac{l}{n-1} \), and natural norm \( R = h I_n \). Define the norm \( \hat{R} = h I_{n-1} \) and the \((n-1) \times n \) derivative operator
\[
D_1 = \frac{1}{h} \begin{bmatrix}
-1 & 1 \\
-1 & 1 \\
\vdots & \vdots \\
-1 & 1 \\
-1 & 1
\end{bmatrix}.
\]
If \( v \in \mathbb{R}^n \), then for any \( \epsilon > 0 \)
\[
\| v \|_\infty^2 \leq (\epsilon^{-1} + l^{-1}) \| v \|_R^2 + \epsilon \| D_1 v \|_{\hat{R}}^2.
\]  

Theorem E.1.2 provides an important link between pointwise and global behavior. Our goal is to mimic (E.2) using SBP operators:

**Conjecture E.1.3** Let \( \hat{\Omega} \) be a sequence of uniform discretizations of \( \Omega = (0, l) \) in which each has an associated SBP pair \( \{ P, Q \} \). For all discretizations, assume that there are \( \mu_1, \mu_2, \) and \( B \) such that
\[
\mu_1 \| x \|_R^2 \leq \| x \|_P^2 \leq \mu_2 \| x \|_R^2,
\]
\[
\| Q \|_R < B.
\]  

105
In this case, there are $c_1 > 0$ and $c_2 > 0$, which possibly depend on $\mu_1$, $\mu_2$ and $B$ but not on mesh spacing $h$, such that

$$\|v\|_\infty^2 \leq c_1 \|v\|_P^2 + c_2 \|P^{-1}Qv\|_P^2.$$  

The requirements on $\{P, Q\}$ stress that bound constants $c_1$ and $c_2$ may depend on the choice of SBP pair. A bound of this form would provide an important link between boundedness of $v$ and $Dv$ and pointwise boundedness. It is quite difficult, however, to find a suitable $c_1$ and $c_2$. In this section, we survey the difficulties in this search.

A first natural step is to combine the discrete Sobolev bound (E.2) and assumption (E.3), which results in

$$\|v\|_\infty^2 \leq \left(\epsilon^{-1} + l^{-1}\right) \mu_1 \|v\|_P^2 + \epsilon \|D_1v\|_P^2. \quad (E.4)$$

Handling the derivative term is more difficult. In Appendix D of [11], the authors attempt to deal with the derivative term using norm equivalence. We will outline their approach and then show where their search for bound constants that are independent of $h$ breaks down. Let $D$ be any derivative operator. If the null space of $D$ contains the vector $[1, 1, 1, \ldots, 1]^T$, then the rows of $D$ can be written as a linear combination of the rows of $D_1$, and, therefore, there is an $n \times (n - 1)$ matrix $N$ such that $D = ND_1$. Further, if $\text{rank}(D) = (n - 1)$, then $N^T P N$ is an $(n - 1) \times (n - 1)$ symmetric positive definite matrix and

$$\|D_1v\|_{N^T P N}^2 = \|Dv\|_P^2.$$  

The next step is to bound the $N^T P N$-norm by the $\hat{R}$-norm. That is, determine $b_1$ and $b_2$ such that for all $x \in \mathbb{R}^{n-1}$,

$$b_1 \|x\|_\hat{R}^2 \leq \|x\|_{N^T P N}^2 \leq b_2 \|x\|_\hat{R}^2. \quad (E.5)$$

Combining (E.4) and (E.5) yields that

$$\|v\|_\infty^2 \leq \left(\epsilon^{-1} + l^{-1}\right) \mu_1 \|v\|_P^2 + \frac{\epsilon}{b_1} \|D_1v\|_\hat{R}^2. \quad (E.6)$$
The challenge is that \( b_1 \) and \( b_2 \) can depend on \( h \). Consider the SBP pair defined by (4.4). For these particular matrices, one can show by explicit computation that

\[
N = \frac{1}{2} \begin{bmatrix}
2 \\
1 & 1 \\
& 
\cdot & 
& 
\cdot \\
& 
& 
\cdot \\
1 & 1 \\
& 
& 
2 \\
\end{bmatrix}, \quad N^T P N = \frac{h}{4} \begin{bmatrix}
3 & 1 \\
1 & 2 & 1 \\
& 
\cdot & 
& 
\cdot \\
& 
& 
\cdot \\
1 & 2 & 1 \\
& 
& 
1 & 3 \\
\end{bmatrix}.
\]

In [101], it is shown that

\[
\lambda \left( N^T P N \right) = \left\{ \frac{h}{2} + \frac{h}{2} \cos \left( \frac{k\pi}{n-1} \right) \bigg| k = 0, \ldots, n-2 \right\},
\]

where \( \lambda \left( N^T P N \right) \) is the spectrum of \( N^T P N \). It follows that, for this particular \( \{P, Q\} \),

\[
b_1(h) \|x\|^2_R \leq \|x\|^2_{N^T P N} \leq b_2 \|x\|^2_R,
\]

where

\[
b_1(h) = \frac{1+\cos\left(\frac{\pi(1-h)}{2}\right)}{2} = \frac{1+\cos(1-h)}{2}, \quad b_2 = 2.
\]

A Taylor series analysis shows that \( b_1 = \mathcal{O}(h^2) \) as \( h \) goes to 0. Because of this, it is not possible to chose \( \epsilon \) in (E.6) to create norm independent bounds. If \( \epsilon \) is selected so that the constant multiplying the derivative term in (E.6) is bounded, then the \( \left( \epsilon^{-1} + l^{-1} \right) \) grows without bound. Likewise, if \( \left( \epsilon^{-1} + l^{-1} \right) \) is bounded, then the coefficient multiplying the derivative term grows without bound. Finally, this growth is not due to our choice of norms. In any bound of the form

\[
\|D_1 x\|^2_R \leq c(h) \|D x\|^2_P
\]

that holds for all \( x \), the bound scaler \( c(h) \) must be be larger than \( h^{-1} \). Let \( z = [+1, -1, +1, \ldots]^T \). For this vector

\[
\|Dz\|^2_P = \frac{4}{\pi}, \quad \|D_1 z\|^2_R = \frac{4}{\pi^2},
\]

and \( \|D_1 z\|^2_R = h^{-1} \|Dz\|^2_P \).
References


