VSB TECHNICAL UNIVERSITY OF OSTRAVA | FACULTY OF ELECTRICAL ENGINEERING AND COMPUTER SCIENCE

Parallel Space-Time Discretisation Methods

Paralelní metody diskretizace v časo-prostorové oblasti

Ing. Ladislav Foltyn

Supervisor: doc. Ing. Dalibor Lukáš, Ph.D.

Ostrava, 2023

Abstrakt

Hlavní náplní disertační práce je aplikace paralelních metod pro řešení parabolických parciálních diferenciálních rovnic. Konkrétně, pro řešení úloh vedoucí na rovnici vedení tepla (difúze). Jednou z možností, jak řešit dané úlohy, je použití konečně prvkové semi-diskretizační metody, kdy se nejprve diskretizuje prostorová oblast za pomocí metody konečných prvků. Získáme tak soustavu obyčejných diferenciálních rovnic, která může být následně diskretizována za použití krokové metody, jako je například Eulerova nebo Crank-Nicolsonova. Na vzniklé časoprostorové semi-diskretizační schéma lze aplikovat paralelní metodu známou pod názvem Parareal (Lions et al. 2001), která poskytuje poměrně jednoduché schéma pro paralelizaci v čase. Pro zvýšení efektivity paralelního řešení, je v rámci disertační práce navržena kombinace Parareal metody s doménovou dekompozicí založenou na pricipu aproximace Schurova doplňku (Bramble et al. 1986). Doménová dekompozice navyšuje míru paralelismu v čase o paralelismus v prostoru, kdy v každém časovém řezu je prostorová oblast rozdělena na dílčí samostatné podoblasti.

Další možností, jak řešit parabolické úlohy, je přímé použití konečně prvkové metody na celou časoprostorovou oblast (Steinbach 2015). Tímto přístupem nám odpadá nutnost použití tenzorové struktury, která se skrývá v pozadí předchozího přístupu, a můžeme tak použít obecnější nestrukturované konečné prvky. Nicméně, poslední část disertační práce pojednává o tzv. Rychlé Diagonalizační Metodě (Langer et al. 2021), která využívá principu tenzorového součinu, aby byla oddělena soustava lineárních rovnic v čase od soustavy rovnic příslušící prostorové oblasti. Tím je umožněno řešit nezávislé úlohy v dílčích časových krocích na základě vlastních čísel matic příslušící diskretizaci úlohy v časovém intervalu. Jelikož výsledné vlastní čísla jsou komplexní, je v rámci této části navržena kombinace Rychlé Diagonalizační Metody s metodou PRESB – Preconditioning for REal matrices with Square Blocks (Axelsson; Lukáš 2019; Axelsson; Neytcheva 2018). PRESB využívá komplexní struktury vzniklých prostorových úloh pro sestavení efektivního předpomíňovače.

Hlavní přínosy disertační práce jsou následující:

1. Detailnější zpracování teorie existence a jednoznačnosti slabé formulace parabolické úlohy od autora Eberharda H. E. Zeidlera.

2. Shrnutí metody Parareal, možností její praktické implementace a provedení kombinace Parareal s doménovou dekompozicí založenou na aproximaci Schurova doplňku.

3. Shrnutí teorie časoprostorových konečných prvků a použití metody PRESB (spolu s FGMRES metodou a multigridem), která demonstruje potenciál úplné paralelizace časosoprostorové úlohy v kombinaci s Rychlou diagonalizační metodou.

Klíčová slova

paralelismus, doménova dekompozice, časoprostorové konečné prvky, konečně prvková semi-diskretizační metoda, Parareal, parabolická úloha, rovnice vedení tepla

Abstract

The main aim of this doctoral thesis is to employ parallel methods to solve parabolic partial differential equations, specifically those leading to transient heat (diffusion) equations. One method for solving these problems is through the use of the semi-discrete finite element method. This scheme involves first discretising the spatial domain using the finite element method, which results in a system of ordinary differential equations. Next, this system is discretised over time using a time-stepping scheme, such as the Euler or Crank-Nicolson method. The resulting semi-discrete problem can then be solved in parallel using the Parareal method (Lions et al. 2001), which offers a relatively straightforward approach. The author of this thesis proposes a novel combination of the Parareal algorithm and a domain decomposition method based on the Schur complement approximation (Bramble et al. 1986) to increase the parallelism. This domain decomposition allows for the concurrent solutions of the spatial subproblems within each time slice.

Another method for solving parabolic partial differential equations is to use the finite element method directly on the space-time domain (Steinbach 2015). This approach does not require any underlying tensor structure as in the previous case and allows for the use of general unstructured finite elements. However, in the final part of the doctoral thesis, the so-called Fast Diagonalisation Method (Langer et al. 2021) is discussed, which uses the tensor-product technique to divide the space-time domain into a system of linear equations along the time interval and the system associated with the spatial domain. This method provides a parallel scheme along the time interval using the eigenvalues of the linear system in time. As the eigenvalues are complex, the author proposes a novel combination of the Fast Diagonalisation Method with the Preconditioning for REal matrices with Square Blocks (PRESB) method (Axelsson; Lukáš 2019; Axelsson; Neytcheva 2018). The PRESB method utilises the complex structure of the obtained spatial systems to construct an efficient preconditioner.

The main contributions of this doctoral thesis are as follows:

1. A more in-depth examination of the Main Theorem of the well-posedness of a weak formulation for a parabolic problem, which Eberhard H. E. Zeidler has established in his work.

2. An overview of the Parareal method, including a discussion of potential implementation options and the proposal of a novel combination of the Parareal with the DDM based on the Schur complement approximation.

3. A summary of the space-time finite element method and the proposal of a novel combination of the Fast Diagonalisation Method and the PRESB algorithm (along with FGMRES method and multigrid), which demonstrates a potential of a full parallelisation of the space-time problem.

Keywords

parallelism, domain decomposition, space-time FEM, FE semi-discrete method, Parareal, parabolic problem, heat equation

Acknowledgement

Firstly, I would like, in memoriam, to thank prof. RNDr. Radim Blaheta, Csc., for his thorough review of my thesis proposal. His insights guided me throughout the completion of this work. I would also like to thank my supervisor, doc. Ing. Dalibor Lukáš, Ph.D., for his patient guidance and prof. RNDr. Jaroslav Haslinger, DrSc. for his expert notes during his additional lectures at VSB-TUO. Lastly, I would like to thank my love Petra, my family, my friends Michal Běloch, Michal Béreš, and Lukáš Zbijovský, as well as my fluffy dogs, Dream Dust and Cidaris, for their unwavering support.

This work was supported by the Ministry of Education, Youth and Sports of the Czech Republic through the e-INFRA CZ (ID:90140).

Contents

List of symbols and abbreviations

$\mathbb{R}$	–	Real numbers
$\mathbb{C}$	–	Complex numbers
V^*	–	Dual space to vector space V
$(\cdot,\cdot)_V$	–	Inner product on vector space V
$\langle\cdot,\cdot\rangle_V$	–	Duality pairing on vector space V
$\|\cdot\|_V$	–	Norm on normed vector space V
$\|\cdot\|_V$	–	Seminorm on normed vector space V
$\nabla u(\mathbf{x},t)$	–	Laplacian applied to function u
$\Delta u(\mathbf{x},t)$	–	Gradient of function u
$\nabla_{\mathbf{x}} u(\mathbf{x},t)$	–	Gradient of function u along the spatial domain
$\Delta_{\mathbf{x}} u(\mathbf{x},t)$	–	Laplace of function u along the spatial domain
A, B, ...	–	Matrices
$\mathbf{u}$, $\mathbf{v}$, ...	–	Vectors
$\otimes$	–	Kronecker product
a.e.	–	almost every
CPU	–	Central Processing Unit
ODE	–	Ordinary Differential Equation
PDE	–	Partial Differential Equation
DDM	–	Domain Decomposition Method
SPD	–	Symmetric positive-definite
FEM	–	Finite Element Method
FE	–	Finite Element

DGM	–	Discontinuous Galerkin Method
FDM	–	Fast Diagonalisation Method
PCG	–	Preconditioned Conjugate Gradient method
FGMRES	–	Flexible Inner-Outer Preconditioned Generalised Minimal Residual method
RHS	–	Right-Hand Side
PRESB	–	Preconditioning for REal matrices with Square Blocks

List of Figures

List of Tables

Chapter 1

Introduction

Solving boundary value problems for elliptic partial differential equations in parallel is a well-established practice today. The most commonly used class of methods are Domain Decomposition Methods (DDMs). There are two main types of DDMs. The first type decomposes the spatial domain into overlapping subdomains (Smith et al. 1996), while the second into the non-overlapping ones (Toselli et al. 2004). Both result in local spatial subproblems that are solved in parallel. The subproblems are linked to the global problem by a coarse solution having a much smaller size than the original problem. This thesis utilises the non-overlapping DDM based on the Schur complement (Bramble et al. 1986). Possible alternatives include the Balancing Domain Decomposition (Mandel et al. 1996) or the finite element tearing and interconnecting method (Farhat et al. 1991). These DDMs generally combine direct methods for subdomains and coarse problems to provide a preconditioner for iterative methods to solve the original system. The resulting condition number is poly-logarithmic in terms of H/h, where H is the diameter of the subdomain (a coarse step), and h is a discretisation step. Such methods offer strong scalability while the computational time and memory consumption are inversely proportional to the number of used CPUs.

While focusing on the parabolic partial differential equations, the situation is more complicated. As the nature of evolution problems is sequential, meaning each time step depends on the previous one, it was long believed that it was not possible to break this connection and develop a parallel algorithm. One possible way to solve parabolic PDEs is to use the semi-discretisation method, such as the method of lines (Thomée 2006). This technique treats the time variable differently than the spatial variable. Firstly, it discretises the spatial domain using a finite element method, resulting in a system of ordinary differential equations. Then, this system of ODEs is solved using a time-stepping method like the Euler or Crank-Nicolson method. In 2001, the Parareal algorithm was introduced (Lions et al. 2001), providing a parallel scheme along the time interval. It is based on the predictor-corrector technique, where a coarse solution corrects solutions obtained by fine solvers in the subdomains. The convergence of the Parareal is proven in (M. Gander et al. 2007a,b). If the time interval is bounded, the order of convergence is super-linear. Otherwise, the convergence is linear. The connection of the Parareal method to the multiple shooting method and the multigrid method is provided in (M. Gander et al. 2007b). Some engineering applications of the Parareal can be found in (Mercerat et al. 2009; Schöps et al. 2017). Three possible implementation options, along with theoretical speedups,

are discussed in (Aubanel 2011).

An alternative to the semi-discrete method is the Discontinuous Galerkin Method (DGM) (Dolejsi et al. 2015). This approach is based on piecewise discontinuous polynomial approximations and thus does not require inter-element continuity. This characteristic makes it well-suited to solving problems with solutions comprising discontinuities and steep gradients, such as compressible flow problems. Another advantage of DGM is its higher-order accuracy on unstructured meshes and the capability of utilising hanging nodes. However, a significant disadvantage is the larger number of Degrees of Freedom (DOFs) compared to the semi-discrete method or the continuous Galerkin Method. Consequently, the resulting system is less sparse than what is achieved through the standard finite element method, and specific parameters have to be selected to ensure stability. To reduce the increased number of DOFs, one option is to use hybrid techniques, as described in (Lehrenfeld 2010). Another hybridised space-time method is discussed in (Neumüller 2013). These methods divide the space-time domain into non-overlapping subdomains that can be additionally solved in parallel. The space-time multigrid method is also proposed and analysed using a two-grid cycle in (Neumüller 2013). A comparison of DGM to the hybridised DGM can be found in (Fidkowski 2019; Woopen et al. 2014).

Instead of the discontinuous Galerkin method, a continuous variant can also be used to solve the parabolic partial differential equations (PDEs). In (Steinbach 2015), the standard finite element method for solving evolutionary problems, known as the space-time finite element method (FEM), is described. The idea is based on the increasing computing capacities which allow for considering an overall solution of a space-time domain. The whole space-time domain is discretised into finite elements. The space-time FEM assumes that the time variable is like an additional spatial variable. This approach requires no underlying tensor structure, so it can handle general finite element meshes and make adaptive refinements simultaneously in space and time. However, finding an efficient preconditioner for unstructured meshes has been challenging in recent years.

For a more detailed insight into existing parallel-in-time methods, from the multiple-shooting method to the direct space-time solvers and their history, see (M. J. Gander 2015). A history of continuous space-time finite element methods for the parabolic evolution equations is summarised in (Steinbach; Yang 2019), which also contains the development of the *a posteriori* error estimates.

1.1 Main objectives

The main objectives of this doctoral thesis are as follows:

1. The first objective is to provide a more in-depth examination of the Main Theorem of the well-posedness of a weak formulation for a parabolic problem, as Eberhard H. E. Zeidler has established. In other words, a proof of the well-posedness of the weak formulation is provided, including clear and detailed steps.

2. The second objective is to deliver an overview of the Parareal method, including a discussion of potential implementation options. Furthermore, a novel combination of the Parareal

with the DDM based on the Schur complement approximation is proposed. Given combination allows us to increase the parallelism in time by the parallelism in space.

3. The third and last objective is to summarise the space-time finite element method and propose a novel combination of the Fast Diagonalisation Method (FDM) with the Preconditioning for REal matrices with Square Blocks (PRESB) method (Axelsson; Neytcheva 2018). Together with the PRESB method, the Flexible Inner-Outer Preconditioned GMRES (FGMRES) method with the underlying multigrid algorithm is utilised. The PRESB method leverages the complex structure of the obtained spatial systems from the FDM to construct an efficient preconditioner.

1.2 Outline

In Chapter 2, the author recalls some definitions and propositions necessary to prove the well-posedness of the weak formulation of the parabolic partial differential equation. The Main Theorem established by Eberhard H. E. Zeidler, which is processed in more detail, can be found in Chapter 3. In Chapter 4, the semi-discrete method using the finite element method (FEM) in the spatial domain and the Euler and Crank-Nicolson methods in the time interval is discussed. Chapter 5 encases the scheme of the Parareal method with an example of a possible practical implementation. The author also proposes a combination of the Parareal algorithm with the DDM based on the Schur complement approximation. Finally, the theory of the space-time FEM is summarised in Chapter 6, where the author describes a combination of the FDM with the PRESB method.

Chapter 2

Preliminaries

In this chapter, we recall the spaces used throughout the thesis, including an important proposition with proof. The mentioned proposition is essential in a specific section of the thesis. Summarised knowledge from (Zeidler 1990a) is being included here to maintain the structure of the following chapters.

2.1 Lebesgue spaces

Let Ω be a nonempty bounded open set in $\mathbb{R}^n$, $n \geq 1$ and let $1 \leq p < \infty$. By

$$L^p(\Omega) := \left\{ v \colon \Omega \to \mathbb{R} \colon \|v\|_{L^p(\Omega)} < \infty \right\},$$

where

$$\|v\|_{L^p(\Omega)} := \left(\int_\Omega |v|^p \, \mathrm{d}\mathbf{x} \right)^{1/p}$$

is the corresponding norm, and v is a measurable function, we note the set of all measurable functions with the given property. The elements of $L^p(\Omega)$ are all the equivalence classes of measurable functions $u \colon \Omega \to \mathbb{R}$ where two measurable functions u, $v \colon \Omega \to \mathbb{R}$ are considered equivalent if they are equal for almost all (a.e.) $\mathbf{x} \in \Omega$. The space $L^p(\Omega)$ is known as the Lebesgue space and forms a Banach space. Furthermore, the following holds true:

(i) $L^p(\Omega)$ is separable.

(ii) The space $C_0^\infty(\Omega)$ is dense in $L^p(\Omega)$.

(iii) The Banach space $L^p(\Omega)$ is reflexive iff $1 < p < \infty$.

(iv) The space $L^2(\Omega)$ with the scalar product

$$(u, v) := \int_\Omega u(\mathbf{x})\, v(\mathbf{x}) \, \mathrm{d}\mathbf{x}$$

is a Hilbert space.

Assume $1 < p,\, q < \infty$, $p^{-1} + q^{-1} = 1$, then

$$\left(L^p(\Omega)\right)^* = L^q(\Omega),$$

where $\left(L^p(\Omega)\right)^*$ is the dual space to $L^p(\Omega)$. In other words, the space $\left(L^p(\Omega)\right)^*$ is the space of all linear continuous functionals $U\colon L^p(\Omega) \to \mathbb{R}$

$$U(v) := \int_\Omega u\, v \,\mathrm{d}\mathbf{x} \quad \forall v \in L^p(\Omega),$$

where $u \in L^q(\Omega)$. We can identify U with u and use a notation

$$\langle u, v \rangle := \int_\Omega u\, v \,\mathrm{d}\mathbf{x} \quad \forall u \in L^q(\Omega), \forall v \in L^p(\Omega)$$

as the duality pairing.

2.2 Sobolev spaces

Let Ω be a nonempty bounded domain in $\mathbb{R}^n$, $n \in \mathbb{N}$, $k \in \mathbb{N} \cup \{0\}$, and let $1 \le p < \infty$. The Sobolev space

$$W^{k,p}(\Omega)$$

is defined as a completion of the vector space $C^\infty(\overline{\Omega})$ using the norm

$$\|u\|_{W^{k,p}(\Omega)} := \left(\sum_{|\alpha| \le k} \int_\Omega |D^\alpha u|^p \,\mathrm{d}\mathbf{x} \right)^{1/p},$$

where

- $\alpha := (\alpha_1, \alpha_2, \ldots, \alpha_n) \in \mathbb{R}^n$ is a multi-index with $\alpha_i \in \mathbb{N} \cup \{0\}$,
- $|\alpha| := \alpha_1 + \alpha_2 + \cdots + \alpha_n$ is a length of the multi-index,
- and

  $$D^\alpha u := \frac{\partial^{|\alpha|} u}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \ldots \partial x_n^{\alpha_n}}$$

  is a generalised (weak) derivative.

It holds

$$W^{k,p}(\Omega) \subset \widetilde{W}^{k,p}(\Omega) := \left\{ u \in L^p(\Omega) \colon D^\alpha u \in L^p(\Omega) \text{ for each multi-index } \alpha, |\alpha| \le k \right\}.$$

If Ω is the domain with the Lipschitz boundary, then

$$W^{k,p}(\Omega) = \widetilde{W}^{k,p}(\Omega).$$

Further, by

$$W_0^{k,p}(\Omega),$$

we define the Sobolev space as a completion of $C_0^\infty(\Omega)$ in $W^{k,p}(\Omega)$

$$W_0^{k,p}(\Omega) := \left\{ u \in W^{k,p}(\Omega) \colon \exists (u_n) \subset C_0^\infty(\Omega) \text{ such that } u_n \to u \text{ in } W^{k,p}(\Omega) \right\}.$$

The space

$$W^{k,2}(\Omega) = \widetilde{W}^{k,2}(\Omega), \tag{2.1}$$

where Ω is the domain with Lipschitz boundary, with the inner product

$$(u, v) := \sum_{|\alpha| \le k} \int_\Omega D^\alpha u D^\alpha v \, \mathrm{d}\mathbf{x}$$

forms the separable Hilbert space. Thus, the notations

$$H^k(\Omega) := W^{k,2}(\Omega), \quad H_0^k(\Omega) := W_0^{k,2}(\Omega)$$

are used to emphasise that we talk about Hilbert spaces. Additionally, the space $W^{k,p}(\Omega)$ is reflexive iff $1 < p < \infty$.

2.3 Evolution triples

Definition 2.1 *Let the following be valid.*

(i) V *is a real, separable, and reflexive Banach space.*

(ii) H *is a real, separable Hilbert space.*

(iii) The embedding $V \subseteq H$ *is continuous and* V *is dense in* H*.*

Then

$$\text{“}V \subseteq H \subseteq V^* \text{”}$$

is an evolution triple.

With the use of the evolution triple, we can identify every $h \in H$ with a corresponding $\overline{h} \in V^*$, as stated in (Zeidler 1990a, Proposition 23.13). This allows us to write

$$\langle h, v \rangle_V = (h, v)_H \quad \forall h \in H, \forall v \in V. \tag{2.2}$$

In particular, we get

$$\langle u, v \rangle_V = \langle v, u \rangle_V \quad \forall u, v \in V.$$

Moreover, it holds that

$$\|h\|_{V^*} \le C \|h\|_H \quad \forall h \in H,$$

where $C > 0$.

2.4 Bochner spaces

Let X be a Banach space, $0 < T < \infty$, and $1 \leq p < \infty$.

By

$$C^m([0,T];X)$$

with norm

$$\|u\|_{C^m([0,T];X)} := \sum_{i=0}^{m} \max_{0\leq t\leq T} \|u^{(i)}(t)\|_X, \tag{2.3}$$

where $m \in \mathbb{N} \cup \{0\}$, we denote the space consisting of all continuous functions $u\colon [0,T] \to X$ that possess continuous derivatives up to the order m on $[0,T]$. It is worth noting that we use the notation $C([0,T];X)$ instead of $C^0([0,T];X)$.

The space of all measurable functions $u\colon (0,T) \to X$ satisfying

$$\|u\|_{L^p((0,T);X)} := \left(\int_0^T \|u(t)\|_X^p \,\mathrm{d}t \right)^{1/p} < \infty,$$

where $\|u\|_{L^p((0,T);X)}$ is the corresponding norm, is denoted by

$$L^p((0,T);X).$$

The set of all polynomials of degree n, where $n \in \mathbb{N} \cup \{0\}$, $w\colon [0,T] \to X$

$$w(t) = a_0 + a_1 t + \cdots + a_n t^n = \sum_{i=0}^{n} a_i t^i$$

with coefficients $a_i \in X$ for all $i = 0, 1, \ldots, n$, we denote as

$$\mathcal{P}_n([0,T];X) := \left\{ w\colon [0,T] \to X \colon w(t) = \sum_{i=0}^{n} a_i t^i;\ \ a_i \in X\ \forall i = 0,1,\ldots,n \right\}.$$

The following holds true:

(i) $C^m([0,T];X)$ with the norm (2.3) is a Banach space over $\mathbb{R}$ (over $\mathbb{C}$).

(ii) $L^p((0,T);X)$ with the norm $\|u\|_{L^p((0,T);X)}$ is a Banach space over $\mathbb{R}$ (over $\mathbb{C}$) in the case where one identifies functions that are equal almost everywhere on $(0,T)$.

(iii) $C([0,T];X)$ is dense in $L^p((0,T);X)$, and the embedding

$$C([0,T];X) \subseteq L^p((0,T);X)$$

is continuous.

(iv) As demonstrated in the proof of Proposition 2.2, the set $\mathcal{P}_n([0,T];X)$ is dense in $C([0,T];X)$. Additionally, $\mathcal{P}_n([0,T];X)$ is also dense in $L^p((0,T);X)$.

(v) If X is a Hilbert space with the inner product $(\cdot,\cdot)_X$, then $L^2((0,T);X)$ is also a Hilbert space with the inner product

$$(u,v) = \int_0^T (u,v)_X \,\mathrm{d}t.$$

(vi) $L^p((0,T);X)$ is separable iff X is separable and $1 \le p < \infty$.

The proposition with proof can be found in (Zeidler 1990a, Proposition 23.2).

Let V be a reflexive and separable Banach space and let $1 < p < \infty$, $p^{-1}+q^{-1} = 1$. According to (Zeidler 1990a, Proposition 23.7), we can identify a real Banach space $\Big(L^p((0,T);V)\Big)^*$ with a real Banach space $L^q((0,T);V^*)$, i.e.,

$$\Big(L^p((0,T);V)\Big)^* = L^q((0,T);V^*), \tag{2.4}$$

and we write

$$\langle u,v\rangle_{L^p((0,T);V)} = \int_0^T \langle u(t),v(t)\rangle_V \,\mathrm{d}t \quad \forall u \in L^q((0,T);V^*),\ \forall v \in L^p((0,T);V),$$

$$\|u\|_{L^q((0,T);V^*)} = \left(\int_0^T \|u(t)\|_{V^*}^q \,\mathrm{d}t\right)^{1/q} \qquad \forall u \in L^q((0,T);V^*).$$

Here, we understand the space $\Big(L^p((0,T);V)\Big)^*$ as the dual space to $L^p((0,T);V)$.

Let "$V \subseteq H \subseteq V^*$" be an evolution triple, $1 \le p,\ q \le \infty$, $0 < T < \infty$, and let $u \in L^p((0,T);V)$. Then by (Zeidler 1990a, Proposition 23.20, (b)) there exists the generalised derivative

$$u^{(n)} \in L^q((0,T);V^*)$$

iff there is a function $w \in L^q((0,T);V^*)$ such that

$$\int_0^T (u(t),v)_H \varphi^{(n)}(t)\,\mathrm{d}t = (-1)^n \int_0^T \langle w(t),v\rangle_V \varphi(t)\,\mathrm{d}t \quad \forall v \in V,\ \forall \varphi \in C_0^\infty((0,T)). \tag{2.5}$$

Then $u^{(n)} = w$ and

$$\frac{\partial^n}{\partial t^n}(u(t),v)_H = \langle u^{(n)}(t),v\rangle_V \quad \forall v \in V \text{ and a.e. } t \in (0,T), \tag{2.6}$$

where $\frac{\partial^n}{\partial t^n}$ means the n-th generalised derivative of real functions on $(0,T)$.

2.5 Polynomials as dense subset

The following proposition is a part of (Zeidler 1990a, Proposition 23.23). This result is essential in proving the convergence of Galerkin approximations in $C([0,T];H)$, where H is a real separable Hilbert space. In practice, $H := L^2(\Omega)$.

Proposition 2.2 *Let "$V \subseteq H \subseteq V^*$" be an evolution triple, and let $1 < p < \infty$, $p^{-1} + q^{-1} = 1$, $0 < T < \infty$. Then it holds that $\mathcal{P}_n([0,T];V)$ is dense in the space $W^{1,p}((0,T);V,H) := \left\{v \in L^p((0,T);V) \colon \frac{\partial v}{\partial t} \in L^q((0,T);V^*)\right\}$, where $\frac{\partial v}{\partial t}$ is a generalised derivative of v.*

The exact definition of the space $W^{1,p}((0,T);V,H)$ is in (Zeidler 1990a, Section 23.6).

Proof: The first part of the proof is based on the properties of Berstein polynomials, which are utilised to prove the Weierstrass theorem. This elaboration provides a detailed explanation of the proof from (Zeidler 1990a).

(i) *Generalised approximation theorem of Weierstrass*: Let X be a Banach space. We show that $\mathcal{P}_n([0,T];X)$ is dense in $C([0,T];X)$.

Consider the Bernstein polynomials

$$b_k(t) = \binom{n}{k} t^k (1-t)^{n-k},$$

where $t \in \mathbb{R}$, $k \in \mathbb{N} \cup \{0\}$, $n \in \mathbb{N}$, $k \leq n$. Then, the following holds

$$\sum_{k=0}^{n} b_k(t) = 1 \tag{2.7}$$

$$\sum_{k=0}^{n} b_k(t)(nt-k)^2 = nt(1-t). \tag{2.8}$$

The first identity results from the Binomial theorem, which reads

$$(r+s)^n = \sum_{k=0}^{n} \binom{n}{k} r^k s^{n-k} \quad \forall r, s \in \mathbb{R},\ n \in \mathbb{N}, \tag{2.9}$$

so by $r := t$, $s := 1-t$,

$$\sum_{k=0}^{n} b_k(t) = [t + (1-t)]^n = 1,$$

which proves (2.7).

To obtain the second identity, we have to derive a few terms. Assume $(r+s) > 0$. By differentiating (2.9) by r, we obtain

$$\underbrace{\binom{n}{0} r^0 s^n}_{=0} + \sum_{k=1}^{n} \binom{n}{k} k r^{k-1} s^{n-k} = n(r+s)^{n-1}.$$

Since

$$\binom{n}{k} k r^{k-1} s^{n-k} = 0 \text{ for } k = 0,$$

we can write

$$\sum_{k=0}^{n} \binom{n}{k} k r^{k-1} s^{n-k} = n(r+s)^{n-1}.$$

By multiplying the latter by r, we arrive at

$$\sum_{k=0}^{n} \binom{n}{k} k r^k s^{n-k} = nr(r+s)^{n-1}. \tag{2.10}$$

Further, we have to differentiate (2.10) by r once again. The right-hand side of the equation is split into two cases:

(a) $n(r+s)^0 + 0$ for $n = 1$,

(b) $n(r+s)^{n-1} + n(n-1)r(r+s)^{n-2}$ for $n > 1$.

Since

$$n(n-1)r(r+s)^{n-2} = 0 \text{ for } n = 1,$$

we can write

$$n(r+s)^{n-1} + n(n-1)r(r+s)^{n-2} \text{ for } n \in \mathbb{N}.$$

Thus, we obtain

$$\sum_{k=0}^{n} \binom{n}{k} k^2 r^{k-1} s^{n-k} = n(r+s)^{n-1} + n(n-1)r(r+s)^{n-2}$$

and we multiply the latter by r

$$\sum_{k=0}^{n} \binom{n}{k} k^2 r^k s^{n-k} = nr(r+s)^{n-1} + n(n-1)r^2(r+s)^{n-2}. \tag{2.11}$$

Once more, denote $r := t$, $s := 1 - t$, then (2.10) and (2.11) yield

$$\sum_{k=0}^{n} \binom{n}{k} k t^k (1-t)^{n-k} = \sum_{k=0}^{n} b_k(t) k = nt,$$
$$\sum_{k=0}^{n} \binom{n}{k} k^2 t^k (1-t)^{n-k} = \sum_{k=0}^{n} b_k(t) k^2 = nt + n(n-1)t^2.$$

Finally, combining (2.7), (2.10), and (2.11),

$$\sum_{k=0}^{n} b_k(t)(nt-k)^2 = n^2t^2 \sum_{k=0}^{n} b_k(t) + 2nt \sum_{k=0}^{n} b_k(t) k + \sum_{k=0}^{n} b_k(t) k^2 = nt(1-t).$$

Let $t \in [0,1]$ and $u \in C([0,1]; X)$ with the norm $\|u\|_C = \max\limits_{0 \le t \le 1} \|u(t)\|_X$. Set

$$B_n(t) = \sum_{k=0}^{n} u\left(\frac{k}{n}\right) b_k(t).$$

With the use of (2.7) and the triangle inequality, we get

$$\begin{aligned}\|u(t)-B_n(t)\|_X &= \left\|u(t)\sum_{k=0}^{n} b_k(t) - \sum_{k=0}^{n} u\Big(\frac{k}{n}\Big)\, b_k(t)\right\|_X \\ &= \left\|\sum_{k=0}^{n}\left[u(t)-u\Big(\frac{k}{n}\Big)\right] b_k(t)\right\|_X \\ &\le \sum_{k=0}^{n}\Big\|\underbrace{u(t)-u\Big(\frac{k}{n}\Big)}_{\in X}\Big\|_X |\underbrace{b_k(t)}_{\in\mathbb{R}_0^+}| = \sum_{k=0}^{n}\left\|u(t)-u\Big(\frac{k}{n}\Big)\right\|_X b_k(t).\end{aligned} \tag{2.12}$$

We divide the given sum into two parts. Let us fix u, n and t. Given $\varepsilon>0$, by continuity of u

$$(\exists\delta>0)\colon \left|t-\frac{k}{n}\right|<\delta \quad\Rightarrow\quad \left\|u(t)-u\Big(\frac{k}{n}\Big)\right\|_X \le \varepsilon. \tag{2.13}$$

We shall denote the related part of the sum, i.e., $\left\{k\colon |t-\frac{k}{n}|<\delta\right\}$, by $\sum_1$.

The second part, denoted as $\sum_2$, contains those elements satisfying

$$\left|t-\frac{k}{n}\right|\ge\delta \quad\Rightarrow \left(\frac{nt-k}{\delta n}\right)^2 \ge 1.$$

Together with the triangle inequality

$$\left\|u(t)-u\Big(\frac{k}{n}\Big)\right\|_X \le \|u(t)\|_X + \left\|u\Big(\frac{k}{n}\Big)\right\|_X \le 2\|u\|_C\cdot 1 \le 2\|u\|_C\,\left(\frac{nt-k}{\delta n}\right)^2. \tag{2.14}$$

Using (2.7), (2.8), (2.13), and (2.14), from (2.12) we get

$$\begin{aligned}\sum_{k=0}^{n}\left\|u(t)-u\Big(\frac{k}{n}\Big)\right\|_X b_k(t) &\le \sum\nolimits_1 \varepsilon b_k(t) + \sum\nolimits_2 2\|u\|_C\left(\frac{nt-k}{\delta n}\right)^2 b_k(t) \\ &= \varepsilon\sum\nolimits_1 b_k(t) + 2\|u\|_C\,\frac{1}{\delta^2 n^2}\sum\nolimits_2 (nt-k)^2\, b_k(t) \\ &\le \varepsilon\sum_{k=0}^{n} b_k(t) + 2\|u\|_C\,\frac{1}{\delta^2 n^2}\sum_{k=0}^{n}(nt-k)^2\, b_k(t) \\ &= \varepsilon + 2\|u\|_C\,\frac{\overbrace{t(1-t)}^{\le 1}}{\delta^2 n} \le \varepsilon + \frac{2\|u\|_C}{\delta^2 n}.\end{aligned}$$

For chosen sufficiently big n_0, it holds

$$n\ge n_0\colon \frac{2\|u\|_C}{\delta^2 n}\le\varepsilon.$$

We have proven that $\forall u\in C\big([0,1];X\big)$, $\forall t\in[0,1]$

$$(\forall\varepsilon>0)(\exists\delta>0)(\exists n_0\in\mathbb{N}\colon \forall n\ge n_0)\colon \|u(t)-B_n(t)\|_X\le 2\varepsilon,$$

i.e., $\mathcal{P}_n\big([0,1];X\big)$ is dense in $C\big([0,1];X\big)$, where X is the Banach space. With the use of similarity transformation, we can prolong $T=1$ to a general T.

(ii) *Dense subset of the space* $C^1([0,T];X)$, (Zeidler 1990a, Problem 23.3). Let X be a Banach space. Show that $\mathcal{P}_n([0,T];X)$ is dense in $C^1([0,T];X)$.

We already know that $\mathcal{P}_n([0,T];X)$ is dense in $C([0,T];X)$. Thus, for each $u \in C^1([0,T];X)$, there exists a sequence of the polynomials $q_n\colon [0,T] \to X$ such that

$$q_n \to u' \quad \text{in } C \quad \text{as } n \to \infty. \tag{2.15}$$

As $u \in C^1([0,T];X)$, by integrating u', we arrive at

$$u(t) = u(0) + \int_0^t u'(s)\,\mathrm{d}s.$$

Set

$$p_n(t) = u(0) + \int_0^t q_n(s)\,\mathrm{d}s.$$

Then we get $p_n' = q_n$. Therefore, using Majorant criterion (Zeidler 1990a, Appendix, (17)), for each $t \in [0,T]$

$$\begin{aligned} 0 \le \|u(t) - p_n(t)\|_X &= \left\| \int_0^t u'(s) - q_n(s)\,\mathrm{d}s \right\|_X \le \int_0^t \|u'(s) - q_n(s)\|_X \,\mathrm{d}s \\ &\le \|u' - q_n\|_{C([0,T];X)} \int_0^T 1\,\mathrm{d}s \\ &= T\,\|u' - q_n\|_{C([0,T];X)} \to 0 \text{ as } n \to \infty. \end{aligned}$$

Since the estimate is independent of t, we obtain $p_n \to u$ in $C([0,T];X)$ as $n \to \infty$, and together with (2.15) we have $p_n \to u$ in $C^1([0,T];X)$ as $n \to \infty$.

(iii) Consider $X = V$, where V is the Banach space from the evolution triple. It can be shown that $C^1([0,T];V)$ is dense in $W^{1,p}((0,T);V,H)$. The proof is outlined in (Zeidler 1990a, Problem 23.10b).

□

Chapter 3

Weak formulation of parabolic problem

Consider the initial-boundary value problem for the heat equation with constant material coefficients

$$\begin{cases} c_H \dfrac{\partial u}{\partial t}(\mathbf{x},t) - \Delta_{\mathbf{x}} u(\mathbf{x},t) = f(\mathbf{x},t) & \forall (\mathbf{x},t) \in Q_T, \\ u(\mathbf{x},t) = 0 & \forall (\mathbf{x},t) \in \Gamma \times [0,T], \\ u(\mathbf{x},0) = u_0(\mathbf{x}) & \forall \mathbf{x} \in \Omega, \end{cases} \tag{3.1}$$

where $Q_T = \Omega \times (0,T)$, $T > 0$, $\Omega \subset \mathbb{R}^n$, $n = 1, 2, 3$ is a bounded domain with Lipschitz boundary, $\Gamma = \partial\Omega$, $f(\mathbf{x},t)$ is a given source term, $u_0(\mathbf{x})$ is a given initial condition, $c_H > 0$ is a heat capacity and $u(\mathbf{x},t)$ has to be found. The variable $\mathbf{x} = (x_1, \ldots, x_n)$ represents a vector of spatial variables.

Note that the entire chapter is based on the work of Eberhard H. E. Zeidler, specifically focusing on the main theorem on first-order linear evolution equations and the Galerkin method. Therefore, we refer to his work (Zeidler 1990a) for additional information. Furthermore, for the purposes of this chapter, we consider $c_H = 1$ without loss of generality.

3.1 Weak formulation

We assume that a classical solution $u(\mathbf{x},t)$ of the problem (3.1) exists. Therefore, the partial derivatives of $u(\mathbf{x},t)$ must exist and be continuous at every point within Q_T. These requirements on the smoothness of $u(\mathbf{x},t)$ can be relaxed, leading to a weak formulation of the problem (3.1). This alternative formulation is represented by an integral equation, which allows us to use a more comprehensive set of functions and apply numerical methods to solve (3.1). It is worth noting that along with the relaxation of the requirements on $u(\mathbf{x},t)$, the requirements on the input data $f(\mathbf{x},t)$ and $u_0(\mathbf{x})$ are also weakened. Here, the idea of obtaining the weak formulation, as presented in (Zeidler 1990a, Section 23.1), along with a few comments, is discussed.

Initially, we shall focus on the spatial domain Ω. We multiply the first equation of (3.1) by a test function $v(\mathbf{x}) \in C_0^\infty(\Omega)$ and integrate it over the spatial domain Ω. We obtain the first

equation of (3.1) in the variational sense

$$\begin{aligned}\int_\Omega \frac{\partial u}{\partial t}(\mathbf{x},t)\, v(\mathbf{x})\,\mathrm{d}\mathbf{x} - \int_\Omega \Delta_{\mathbf{x}} u(\mathbf{x},t)\, v(\mathbf{x})\,\mathrm{d}\mathbf{x} \\ = \int_\Omega f(\mathbf{x},t)\, v(\mathbf{x})\,\mathrm{d}\mathbf{x} \quad \forall v(\mathbf{x}) \in C_0^\infty(\Omega),\ \forall t \in (0,T).\end{aligned} \tag{3.2}$$

Using Green's theorem, the second term of (3.2) is rewritten to

$$-\int_\Omega \Delta_{\mathbf{x}} u(\mathbf{x},t)\, v(\mathbf{x})\,\mathrm{d}\mathbf{x} = \int_\Omega \nabla_{\mathbf{x}} u(\mathbf{x},t)\, \nabla_{\mathbf{x}} v(\mathbf{x}) \quad \forall v \in C_0^\infty(\Omega),\ \forall t \in (0,T) \tag{3.3}$$

since the functions $v(\mathbf{x})$ vanish on Γ. Combining (3.3) and (3.2), we arrive at

$$\begin{aligned}\int_\Omega \frac{\partial u}{\partial t}(\mathbf{x},t)\, v(\mathbf{x})\,\mathrm{d}\mathbf{x} + \int_\Omega \nabla_{\mathbf{x}} u(\mathbf{x},t)\, \nabla_{\mathbf{x}} v(\mathbf{x})\,\mathrm{d}\mathbf{x} \\ = \int_\Omega f(\mathbf{x},t)\, v(\mathbf{x})\,\mathrm{d}\mathbf{x} \quad \forall v \in C_0^\infty(\Omega),\ \forall t \in (0,T).\end{aligned} \tag{3.4}$$

By (Zeidler 1990b, Appendix (25)), we can interchange the partial derivative $\frac{\partial}{\partial t}$ with the integral $\int_\Omega$ in the first term of (3.4)

$$\begin{aligned}\frac{\partial}{\partial t}\int_\Omega u(\mathbf{x},t)\, v(\mathbf{x})\,\mathrm{d}\mathbf{x} + \int_\Omega \nabla_{\mathbf{x}} u(\mathbf{x},t)\, \nabla_{\mathbf{x}} v(\mathbf{x})\,\mathrm{d}\mathbf{x} \\ = \int_\Omega f(\mathbf{x},t)\, v(\mathbf{x})\,\mathrm{d}\mathbf{x} \quad \forall v \in C_0^\infty(\Omega),\ \forall t \in (0,T).\end{aligned} \tag{3.5}$$

Let

$$V := H_0^1(\Omega), \quad H := L^2(\Omega).$$

We shall generalise (3.5) for

$$u(t) \in V \ \forall t \in (0,T),$$

and $\forall v \in V$. For brevity, we use the shorthand notation $u(t)$ to represent an element of V, which is defined as the function $\mathbf{x} \mapsto u(\mathbf{x},t)$, where $\mathbf{x}$ is the spatial variable and $t \in (0,T)$ is fixed. Therefore, if we vary the time t in the interval $(0,T)$, we get the function $t \mapsto u(t)$ with values in the Banach space V. With this at hand, we have to find $u(t) \in V$ such that

$$\begin{cases} \dfrac{\partial}{\partial t}\big(u(t),v\big)_H + a(u(t),v) = (f(t),v)_H \quad \forall v \in V,\ \forall t \in (0,T), \\ u(0) = u_0 \in H, \end{cases} \tag{3.6}$$

where

$$\frac{\partial}{\partial t}\big(u(t),v\big)_H := \frac{\partial}{\partial t}\int_\Omega u(\mathbf{x},t)\, v(\mathbf{x})\,\mathrm{d}\mathbf{x},$$

$$a(u(t), v) := \int_\Omega \nabla_{\mathbf{x}} u(\mathbf{x}, t)\, \nabla_{\mathbf{x}} v(\mathbf{x})\, \mathrm{d}\mathbf{x}, \tag{3.7}$$

$$(f(t), v)_H := \int_\Omega f(\mathbf{x}, t)\, v(\mathbf{x})\, \mathrm{d}\mathbf{x}. \tag{3.8}$$

The choice $u_0 \in H$ will be discussed later.

At this point, we have already weakened the requirements on solution u in the spatial domain Ω. The next step is to weaken the requirements on $u(\mathbf{x}, t)$ along the time interval $(0, T)$. Since we consider that the mapping $t \mapsto u(t)$ is continuous $\forall t \in (0, T)$, the solution u belongs to $L^2\big((0, T); V\big)$ by (Zeidler 1990a, Example 23.3). Note that V is the separable Hilbert space. Hence, (3.6) shall be further generalised for $u \in L^2\big((0, T); V\big)$.

Hereafter, we shall assume that $f \in L^2(Q_T)$. By (Zeidler 1990a, Example 23.4), if $f \in L^2(Q_T)$, then $\forall t \in [0, T]$ there exists $b(t) \in V^*$ such that

$$\langle b(t), v\rangle_V := \int_\Omega f(t)\, v\, \mathrm{d}\mathbf{x} \quad \forall v \in V$$

holds for almost every $t \in (0, T)$ and $t \mapsto b(t)$ belongs to $L^2\big((0, T); V^*\big)$. Moreover, we get

$$\|b\|^2_{L^2((0,T);V^*)} \le \int_{Q_T} |f(\mathbf{x}, t)|^2\, \mathrm{d}\mathbf{x}\, \mathrm{d}t.$$

Furthermore, we shall understand the time derivative $\frac{\partial}{\partial t}$ as a generalised derivative, i.e.,

$$\int_0^T \frac{\partial}{\partial t}\big(u(t), v\big)_H\, \varphi(t)\, \mathrm{d}t = -\int_0^T \big(u(t), v\big)_H\, \varphi'(t)\, \mathrm{d}t \quad \forall v \in V,\ \forall \varphi \in C_0^\infty(0, T). \tag{3.9}$$

Then, using the Variational Lemma (Zeidler 1990a, Proposition 23.10), (3.6) is equivalent to

$$\int_0^T \frac{\partial}{\partial t}\big(u(t), v\big)_H\, \varphi(t)\, \mathrm{d}t + \int_0^T a(u(t), v)\, \varphi(t)\, \mathrm{d}t = \int_0^T (f(t), v)_H\, \varphi(t)\, \mathrm{d}t \quad \forall v \in V$$

and $\forall \varphi(t) \in C_0^\infty\big((0, T)\big)$. In other words, (3.6) has to be satisfied only for almost every $t \in (0, T)$. Moreover, since V and H form the evolution triple "$V \subseteq H \subseteq V^*$", one can define the generalised derivative $u' := \frac{\partial u}{\partial t}$ of u by (2.6), i.e.,

$$\frac{\partial}{\partial t}\big(u(t), v\big)_H = \big\langle u'(t), v\big\rangle_V \quad \forall v \in V \text{ and a.e. } t \in (0, T), \tag{3.10}$$

and the solution space

$$W = W^{1,2}\big((0, T); V, H\big) := \Big\{u \in L^2\big((0, T); V\big) \colon u' \in L^2\big((0, T); V^*\big)\Big\}. \tag{3.11}$$

The norm of W is given as

$$\|u\|_W := \left(\int_0^T \|u(t)\|_V^2 \, \mathrm{d}t \right)^{1/2} + \left(\int_0^T \|u'(t)\|_{V^*}^2 \, \mathrm{d}t \right)^{1/2}. \tag{3.12}$$

The precise definition of the space W is in (Zeidler 1990a, Section 23.6). We shall note that the space W with the norm $\|\cdot\|_W$, which is not a standard norm, forms a real Banach space. Hence, we arrive at the variational formulation to find $u \in W$ such that

$$\begin{cases} \dfrac{\partial}{\partial t}(u(t), v)_H + a(u(t), v) = \langle b(t), v \rangle_V \quad \forall v \in V \text{ and a.e. } t \in (0, T), \\ u(0) = u_0 \in H. \end{cases} \tag{3.13}$$

The space H was introduced according to the inner product in H, to which the generalised derivative $u'(t)$ was defined. i.e., $u_0 \in H$ follows from (2.2) and (3.10). Furthermore, the proof of the well-posedness of the weak formulation highlights that the continuous dependency on the initial condition u_0 is only defined in relation to the inner product in H.

Remark 3.1 There exists a continuous embedding

$$W \hookrightarrow C([0, T], H). \tag{3.14}$$

by (Zeidler 1990a, Proposition 23.23, ii).

Theorem 3.2 (Well-posedness of weak formulation) *Let the following hold.*

(i) *"$V \subseteq H \subseteq V^*$" is an evolution triple with* $\dim V = \infty$, $0 < T < \infty$, *where the spaces V and H are real Hilbert spaces.*

(ii) *Mapping $a : V \times V \to \mathbb{R}$ is bilinear, bounded*

$$(\exists C > 0) \colon a(u, v) \leq C \, \|u\|_V \, \|v\|_V, \, \forall u, v \in V,$$

and coercive

$$(\exists c > 0) \colon c \, \|v\|_V^2 \leq a(v, v), \, \forall v \in V.$$

Moreover, we are given $u_0 \in H$ and $b \in L^2((0, T); V^*)$.

(iii) *$\{w_1, w_2, \dots\}$ is a basis in V, which is also dense in V, and (u_{n0}) is a sequence in H with*

$$u_{n0} \to u_0 \text{ in } H \quad \text{as } n \to \infty,$$

where

$$u_{n0} \in \operatorname{span}\{w_1, w_2, \dots, w_n\} \quad \forall n.$$

Then:

(a) **Existence and uniqueness**. *The weak formulation* (3.13) *has exactly one solution u.*

(b) **Continuous dependence on the data**. *The map*

$$(u_0, b) \mapsto u$$

is linear and continuous from $H \times L^2\big((0,T);V^*\big)$ *to* $W^{1,2}\big((0,T);V,H\big)$, *i.e., there is a constant* $C > 0$ *such that*

$$\|u\|_W \leq C\,\Big(\|u_0\|_H + \|b\|_{L^2((0,T);V^*)}\Big)\,, \tag{3.15}$$

for all $u_0 \in H$ *and* $b \in L^2\big((0,T);V^*\big)$.

(c) **Convergence of Galerkin method**. *For all* $n \in \mathbb{N}$, *the Galerkin approximations*

$$\begin{cases} \big(u_n'(t), w_i\big)_H + a\big(u_n(t), w_i\big) = \big\langle b(t), w_i\big\rangle_V \quad \forall t \in (0,T),\ i = 1, 2, \ldots, n, \\ u_n(0) = u_{n0} \end{cases} \tag{3.16}$$

have unique solutions

$$u_n \in W,$$

where $u_n(t) = \sum\limits_{j=1}^{n} c_{jn}(t)\, w_j$ *is an approximate solution for fixed* n *with* $c_{jn}(t) \in C^1([0,T])$, *and* $u_{n0} = \sum\limits_{j=1}^{n} \alpha_{jn}\, w_j$ *is some approximation of the initial condition* $u_0 \in H$. *The sequence* (u_n) *converges as* $n \to \infty$ *to the solution* u *of* (3.13) *in the following sense*

$$\begin{aligned} &u_n \to u \quad \text{in } L^2\big((0,T);V\big), \\ &\max_{0\leq t\leq T} \|u_n(t) - u(t)\|_H \to 0. \end{aligned}$$

Remark 3.3 The equations (3.16) result in a linear system of ordinary differential equations. As stated in (Zeidler 1986, Corollary 3.8), these equations possess a unique classical solution on the interval $[0,T]$ for each $n \in \mathbb{N}$.

Corollary 3.4 The original problem (3.13) is equivalent to the following operator problem. Find $u \in W$ such that

$$\begin{cases} u'(t) + Au(t) = b(t) \quad \text{in } V^* \text{ for a.e. } t \in (0,T), \\ u(0) = u_0 \in H. \end{cases} \tag{3.17}$$

Here, the operator $A\colon V \to V^*$ defined by

$$\big\langle Au, v\big\rangle_V = a(u,v) \quad \forall u, v \in V, \tag{3.18}$$

is a linear, continuous

$$(\exists C > 0)\colon \big\langle Au, v\big\rangle_V \leq C\,\|u\|_V \|v\|_V \quad \forall u, v \in V,$$

and coercive

$$(\exists c > 0)\colon c\,\|v\|_V^2 \leq \big\langle Av, v\big\rangle_V \quad \forall v \in V.$$

3.2 Proof of well-posedness

The proof consists of the following parts:

1. Equivalence between the weak formulation (3.13) and the operator equation (3.17) is shown.

2. Using the operator equation (3.17), uniqueness of the solution is proved.

3. To prove the existence of the solution, the subsequent steps are done.

   3.1 Boundeness of the Galerkin solutions in $L^2\big((0,T);V\big)$ is proved, i.e.,

   $$\int_0^T \|u_n(t)\|_V^2 \,\mathrm{d}t \le K\left(\|u_n(0)\|_H^2 + \int_0^T \|b(t)\|_{V^*}^2 \,\mathrm{d}t\right) \tag{3.19}$$

   for all n, where the constant $K > 0$ is independent of n.

   3.2 With the bound (3.19) at hand, the existence of the weak limit

   $$u_n \rightharpoonup \widetilde{u} \quad \text{in } L^2\big((0,T);V\big) \text{ as } n \to \infty. \tag{3.20}$$

   is discussed.

   3.3 Integral identity

   $$\begin{aligned} -(u_0,v)_H\,\varphi(0) - \int_0^T \langle \widetilde{u}(t),v\rangle_V\,\varphi'(t)\,\mathrm{d}t + \\ + \int_0^T \langle A\widetilde{u}(t),v\rangle_V\,\varphi(t)\,\mathrm{d}t = \int_0^T \langle b(t),v\rangle_V\,\varphi(t)\,\mathrm{d}t \end{aligned} \tag{3.21}$$

   for all $v \in V$ and all real functions

   $$\varphi \in C^1\big([0,T]\big) \quad \text{with} \quad \varphi(T) = 0$$

   is proved.

   3.4 At last, the proof of the existence relies on a convergence of the Galerkin approximations (3.16) to the operator equation (3.17). This convergence is proved using the integral identity (3.21). In other words, $\widetilde{u}$ solve (3.17), thus $\widetilde{u} = u$.

4. Continuous dependence on the data is shown.

5. Convergence of the Galerkin method in $C\big([0,T];H\big)$ is shown.

6. Finally, strong convergence of the Galerkin method in the space $L^2\big((0,T);V\big)$ is proved.

3.2.1 Operator equation as equivalent equation

We show that the first equation of (3.13) is equivalent to the operator equation (3.17).

Let $u \in W$ and $A\colon V \to V^*$ be linear, continuous and coercive defined by (3.18). With this at hand and using (3.10), we rewrite (3.13) to

$$\langle u'(t), v\rangle_V + \langle Au(t), v\rangle_V = \langle b(t), v\rangle_V \quad \forall v \in V \text{ and a.e. } t \in (0,T).$$

The latter yields

$$\langle u'(t) + Au(t) - b(t), v\rangle_V = 0 \quad \forall v \in V \text{ and a.e. } t \in (0,T),$$

because duality pairing is bilinear. By the definition of V^*, we get

$$u'(t) + Au(t) - b(t) = 0 \quad \text{in } V^* \text{ and a.e. } t \in (0,T),$$

and so

$$u'(t) + Au(t) = b(t) \quad \text{in } V^* \text{ and a.e. } t \in (0,T).$$

3.2.2 Uniqueness

We show that the operator (3.17) gives the unique solution.

Let $u_1 \in W$ solves

$$\begin{aligned} u_1'(t) + Au_1(t) &= b(t) \quad \text{in } V^* \text{ for a.e. } t \in (0,T),\\ u_1(0) &= u_0, \end{aligned}$$

and $u_2 \in W$ solves

$$\begin{aligned} u_2'(t) + Au_2(t) &= b(t) \quad \text{in } V^* \text{ for a.e. } t \in (0,T),\\ u_2(0) &= u_0. \end{aligned}$$

As A is the linear mapping, we get

$$\begin{aligned} \big(u_1'(t) - u_2'(t)\big) + A\big(u_1(t) - u_2(t)\big) &= 0 \quad \text{in } V^* \text{ for a.e. } t \in (0,T),\\ \big(u_1(0) - u_2(0)\big) &= 0. \end{aligned}$$

Let $u(t) := u_1(t) - u_2(t) \in W$, we obtain

$$\begin{aligned} u'(t) + A\left(u(t)\right) &= 0 \quad \text{in } V^* \text{ for a.e. } t \in (0,T),\\ u(0) &= 0, \end{aligned}$$

so we have to show that

$$u(t) = 0 \quad \text{in } V^* \text{ for a.e. } t \in (0,T).$$

The equation

$$u'(t) + Au(t) = 0$$

is equal to

$$u'(t) = -Au(t).$$

We make duality pairing $\langle \cdot, u(t) \rangle_V$, $u(t) \in V$, and integrate it over $\langle 0, T \rangle$

$$\int_0^T \langle u'(t), u(t) \rangle_V \, \mathrm{d}t = -\int_0^T \langle Au(t), u(t) \rangle_V \, \mathrm{d}t \tag{3.22}$$

By (Zeidler 1990a, Proposition 23.23, (iv)), we get following integration by parts formula

$$\int_0^T \langle u'(t), u(t) \rangle_V \, \mathrm{d}t = \frac{1}{2} \Big(\big(u(T), u(T)\big)_H - \big(u(0), u(0)\big)_H \Big),$$

where $\big(u(0), u(0)\big)_H = 0$ ($u(0) = 0$), so

$$\int_0^T \langle u'(t), u(t) \rangle_V \, \mathrm{d}t = \frac{1}{2} \|u(T)\|_H^2.$$

Further, we have

$$\langle Au(t), v \rangle_V = a\big(u(t), v\big) \quad \forall v \in V.$$

Thus, we rewrite (3.22) to

$$\frac{1}{2} \|u(T)\|_H^2 = \underbrace{-\int_0^T a\big(u(t), u(t)\big) \, \mathrm{d}t \le -c \int_0^T \|u(t)\|_V^2 \, \mathrm{d}t}_{\text{as } a(\cdot,\cdot) \text{ is coercive}},$$

and arrive at

$$0 \le c \int_0^T \|u(t)\|_V^2 \, \mathrm{d}t \le \frac{1}{2} \, \|u(T)\|_H^2 + c \int_0^T \|u(t)\|_V^2 \, \mathrm{d}t \le 0.$$

Finally,

$$\int_0^T \|u(t)\|_V^2 \, \mathrm{d}t = \|u\|^2_{L^2((0,T);V)} = 0 \quad \Rightarrow \quad u = u_1 - u_2 = 0 \quad \Rightarrow \quad u_1 = u_2.$$

3.2.3 Existence proof via Galerkin method

For the sake of simplicity, let us assume that the function $t \mapsto \langle b(t), v \rangle_V$ is continuous on $[0, T]$ for all $v \in V$. For a more general scenario, where $b \in L^2\big((0,T); V^*\big)$, we refer to (Zeidler 1990a, Proof of Corollary 23.26).

3.2.3.1 Boundeness of Galerkin solutions

We prove (3.19).

We shall recall the Galerkin approximations

$$\begin{aligned} \left(u_n'(t), w_i\right)_H + a(u_n(t), w_i) &= \langle b(t), w_i\rangle_V \quad i = 1, \dots, n \\ u_n(0) &= u_{n0}, \end{aligned} \tag{3.23}$$

where

$$u_n(t) = \sum_{j=1}^{n} c_{jn}(t)\, w_j,$$

and w_j are basis functions in V $\forall j = 1, \dots, n$. Multiplying (3.23) by $c_{in}(t)$ and summing over i

$$\sum_{i=1}^{n} c_{in}(t)\, \left(u_n'(t), w_i\right)_H + \sum_{i=1}^{n} c_{in}(t)\, a\big(u_n(t), w_i\big) = \sum_{j=1}^{n} c_{in}(t)\, \langle b(t), w_i\rangle_V$$

yields

$$\left(u_n'(t), u_n(t)\right)_H + a\big(u_n(t), u_n(t)\big) = \big\langle b(t), u_n(t)\big\rangle_V. \tag{3.24}$$

By (Zeidler 1990b, Appendix, (25)), the following formula holds true

$$\frac{\partial}{\partial t}\big(u_n(t), u_n(t)\big)_H = \left(u_n'(t), u_n(t)\right)_H + \left(u_n(t), u_n'(t)\right)_H = 2\left(u_n'(t), u_n(t)\right)_H. \tag{3.25}$$

Combining the first term of (3.24) with (3.25), we arrive at

$$\begin{aligned} \frac{1}{2}\frac{\partial}{\partial t}\big(u_n(t), u_n(t)\big)_H + a\big(u_n(t), u_n(t)\big) &= \big\langle b(t), u_n(t)\big\rangle_V, \\ \frac{\partial}{\partial t}\left\|u_n(t)\right\|_H^2 + 2\, a\big(u_n(t), u_n(t)\big) &= 2\,\big\langle b(t), u_n(t)\big\rangle_V. \end{aligned}$$

Further, we integrate the latter over $[0, T]$

$$\int_0^T \frac{\partial}{\partial t}\left\|u_n(t)\right\|_H^2 \,\mathrm{d}t + 2\int_0^T a\big(u_n(t), u_n(t)\big)\,\mathrm{d}t = 2\int_0^T \big\langle b(t), u_n(t)\big\rangle_V \,\mathrm{d}t.$$

Since $a(\cdot,\cdot)$ is coercive, we get

$$\begin{aligned} \left\|u_n(T)\right\|_H^2 - \left\|u_n(0)\right\|_H^2 + 2\,c\int_0^T \left\|u_n(t)\right\|_V^2 \,\mathrm{d}t &\le 2\int_0^T \big\langle b(t), u_n(t)\big\rangle_V \,\mathrm{d}t, \\ \left\|u_n(T)\right\|_H^2 + 2\,c\int_0^T \left\|u_n(t)\right\|_V^2 \,\mathrm{d}t &\le \left\|u_n(0)\right\|_H^2 + 2\int_0^T \big\langle b(t), u_n(t)\big\rangle_V \,\mathrm{d}t. \end{aligned} \tag{3.26}$$

For arbitrary $\hat{v} \in V$ it holds that

$$\|b\|_{V^*}\, \|\hat{v}\|_V \ge \langle b, \hat{v}\rangle_V.$$

In our case $\hat{v} = u_n$. Therefore, we rewrite (3.26) to

$$\|u_n(T)\|_H^2 + 2\,c \int_0^T \|u_n(t)\|_V^2 \,\mathrm{d}t \leq \|u_n(0)\|_H^2 + 2 \int_0^T \|b(t)\|_{V^*} \,\|u_n(t)\|_V \,\mathrm{d}t. \tag{3.27}$$

Now, we focus on the second term of the right-hand side

$$\begin{aligned} 2 \int_0^T \|b(t)\|_{V^*} \,\|u_n(t)\|_V \,\mathrm{d}t &= \int_0^T 2\,\sqrt{c^{-1}} \,\|b(t)\|_{V^*} \,\sqrt{c}\,\|u_n(t)\|_V \,\mathrm{d}t \\ &\leq \int_0^T c^{-1} \,\|b(t)\|_{V^*}^2 + c\,\|u_n(t)\|_V^2 \,\mathrm{d}t \\ &= c^{-1} \int_0^T \|b(t)\|_{V^*}^2 \,\mathrm{d}t + c \int_0^T \|u_n(t)\|_V^2 \,\mathrm{d}t, \end{aligned} \tag{3.28}$$

where $c > 0$ is the constant of coercive property of the bilinear form $a(\cdot,\cdot)$. Combining (3.27) with (3.28), we arrive at

$$\|u_n(T)\|_H^2 + c \int_0^T \|u_n(t)\|_V^2 \,\mathrm{d}t \leq \|u_n(0)\|_H^2 + c^{-1} \int_0^T \|b(t)\|_{V^*}^2 \,\mathrm{d}t.$$

As $\|u_n(T)\|_H^2 \geq 0$, we obtain

$$c \int_0^T \|u_n(t)\|_V^2 \,\mathrm{d}t \leq \|u_n(T)\|_H^2 + c \int_0^T \|u_n(t)\|_V^2 \,\mathrm{d}t \leq \|u_n(0)\|_H^2 + c^{-1} \int_0^T \|b(t)\|_{V^*}^2 \,\mathrm{d}t.$$

Finally,

$$\begin{aligned} \int_0^T \|u_n(t)\|_V^2 \,\mathrm{d}t &\leq \frac{1}{c^2} \left(c\,\|u_n(0)\|_H^2 + 1 \int_0^T \|b(t)\|_{V^*}^2 \,\mathrm{d}t \right) \\ &\leq \frac{\max\{c,1\}}{c^2} \left(\|u_n(0)\|_H^2 + \int_0^T \|b(t)\|_{V^*}^2 \,\mathrm{d}t \right) \\ &= K \left(\|u_n(0)\|_H^2 + \int_0^T \|b(t)\|_{V^*}^2 \,\mathrm{d}t \right), \end{aligned}$$

where K is independent of n. Hence, according to (3.19), the sequence of the Galerkin solutions (u_n) is bounded in Hilbert space $L^2((0,T);V)$, since

$$u_n(0) \to u_0 \quad \text{in } H \text{ as } n \to \infty.$$

3.2.3.2 Weak convergence of Galerkin method in $L^2\big((0,T);V\big)$

As the space $L^2\big((0,T);V\big)$ is reflexive, there exists a weakly convergent subsequence $(\hat{u}_n)$

$$\hat{u}_n \rightharpoonup \tilde{u} \quad \text{in } L^2\big((0,T);V\big) \text{ as } n \to \infty$$

as stated by *Eberlein-Šmuljan theorem*. With this weak convergence, we shall demonstrate below that the Galerkin approximations tend towards the operator equation (3.17). Furthermore, as previously established in subsection 3.2.2, the equation (3.17) has a unique solution. Therefore, each weakly convergent subsequence converges to the same limit, as stated in (Zeidler 1990a, Proposition 21.23(i)). Hence, we obtain the existence of u, as can be seen below.

3.2.3.3 Integral identity

In order to prove the existence of the weak solution and to demonstrate that the weak limit (3.20) converges to solution u, we shall justify (3.21).

Choose $\varphi \in C^1\big([0,T]\big)$, $\varphi(T) = 0$. Multiplying the Galerkin equations (3.23) by chosen φ

$$\big(u_n'(t), w_i\big)_H \varphi(t) + a\big(u_n(t), w_i\big)\, \varphi(t) = \big\langle b(t), w_i \big\rangle_V \varphi(t),$$

and integrating it over $[0,T]$

$$\int_0^T \big(u_n'(t), w_i\big)_H \varphi(t)\,\mathrm{d}t + \int_0^T a\big(u_n(t), w_i\big)\, \varphi(t)\,\mathrm{d}t = \int_0^T \big\langle b(t), w_i \big\rangle_V \varphi(t)\,\mathrm{d}t. \tag{3.29}$$

Let us focus on the first term of the (3.29). By Fubini's theorem

$$\int_0^T \big(u_n'(t), w_i\big)_H \varphi(t)\,\mathrm{d}t = \int_\Omega \int_0^T u_n'(t)\, \varphi(t)\,\mathrm{d}t\, w_i\,\mathrm{d}\mathbf{x}$$

along with integration by parts

$$\int_0^T u_n'(t)\, \varphi(t)\,\mathrm{d}t = \underbrace{u_n(T)\, \varphi(T)}_{=0} - u_n(0)\, \varphi(0) - \int_0^T u_n(t)\, \varphi'(t)\,\mathrm{d}t,$$

we obtain

$$\begin{aligned} -\big(u_n(0), w_i\big)_H \varphi(0) - \int_0^T \big(u_n(t), w_i\big)_H \varphi'(t)\,\mathrm{d}t + \\ + \int_0^T a\big(u_n(t), w_i\big)\, \varphi(t)\,\mathrm{d}t = \int_0^T \big\langle b(t), w_i \big\rangle_V \varphi(t)\,\mathrm{d}t \end{aligned}$$

for all $i = 1, 2, \ldots, n$. Furthermore, by (2.2), we can identify

$$\big(u_n(t), w_i\big)_H = \big\langle u_n(t), w_i \big\rangle_V.$$

Hence,

$$
\begin{aligned}
-\left(u_n(0), w_i\right)_H \varphi(0) - \int_0^T \left\langle u_n(t), w_i \right\rangle_V \varphi'(t)\,\mathrm{d}t + \\
+ \int_0^T a(u_n(t), w_i)\,\varphi(t)\,\mathrm{d}t = \int_0^T \left\langle b(t), w_i \right\rangle_V \varphi(t)\,\mathrm{d}t.
\end{aligned}
\tag{3.30}
$$

Along with the weak limit (3.20), we get

$$
\begin{aligned}
-\left(u_0, w_i\right)_H \varphi(0) - \int_0^T \left\langle \widetilde{u}(t), w_i \right\rangle_V \varphi'(t)\,\mathrm{d}t + \\
+ \int_0^T a(\widetilde{u}(t), w_i)\,\varphi(t)\,\mathrm{d}t = \int_0^T \left\langle b(t), w_i \right\rangle_V \varphi(t)\,\mathrm{d}t.
\end{aligned}
\tag{3.31}
$$

We need to justify the weak limit (3.20), which means that the second and the third terms in the previous equation (3.30) have to be linear continuous functionals on $L^2\big((0,T);V\big)$.

The convergence of the first term follows from (3.30) as $u_n(0) \to u_0$ in H when $n \to \infty$. The continuity of the second term is established by the inequalities

$$
\left|\left\langle u_n(t), w_i \right\rangle_V\right| \le \|u_n(t)\|_{V^*}\ \|w_i\|_V\,,
$$

and

$$
(\exists C_1 > 0)\colon\ \|v\|_{V^*} \le C_1\ \|v\|_V\ \ \forall v \in V.
$$

Thus, the following inequalities are obtained

$$
\begin{aligned}
\left|\int_0^T \left\langle u_n(t), w_i \right\rangle_V \varphi'(t)\,\mathrm{d}t\right| &\le \int_0^T \left|\left\langle u_n(t), w_i \right\rangle_V\right|\ |\varphi'(t)|\ \mathrm{d}t \le \\
&\le \int_0^T \|u_n(t)\|_{V^*}\ \|w_i\|_V\ |\varphi'(t)|\ \mathrm{d}t \le \\
&\le C_1\ \|w_i\|_V \int_0^T \|u_n(t)\|_V\ |\varphi'(t)|\ \mathrm{d}t.
\end{aligned}
$$

Finally, by the Hölder inequality, we get for the second term of (3.30)

$$
\left|\int_0^T \left\langle u_n(t), w_i \right\rangle_V \varphi'(t)\,\mathrm{d}t\right| \le C_1\,\|w_i\|_V\ \left(\int_0^T \|u_n(t)\|_V^2\ \mathrm{d}t\right)^{1/2} \underbrace{\left(\int_0^T |\varphi'(t)|^2\ \mathrm{d}t\right)^{1/2}}_{=k_1} \le
$$

$$
\le \underbrace{C_1\,k_1\,\|w_i\|_V}_{=C_2}\ \|u_n(t)\|_{L^2((0,T);V)}\,.
$$

Analogously, for the third term of (3.30)

$$\left| \int_0^T a(u_n(t), w_i)\varphi(t)\,\mathrm{d}t \right| \le \int_0^T \left| \langle u_n(t), w_i \rangle_V \right| \, |\varphi(t)| \,\mathrm{d}t \le$$
$$\le \underbrace{C_1\, k_2\, \|w_i\|_V}_{=C_3} \, \|u_n(t)\|_{L^2((0,T);V)} \,,$$

where we have used (3.18) while identifying $Au_n(t)$ with $u_n(t)$.

At last, we have to replace w_i in (3.31) by $v \in V$ since we need to obtain (3.17). By the assumption (iii) in Theorem 3.2, it follows from the density of basis $\{w_1, w_2, \dots\}$ in V that there exists a sequence (v_n) with

$$v_n \to v \quad \text{in } V \text{ as } n \to \infty, \tag{3.32}$$

where each v_n is a finite linear combination of certain basis elements w_i. Then, the equation (3.31) holds true if we replace w_i with v for $n \to \infty$. The limit (3.32) can be applied if all the terms in (3.31) are linear continuous functionals on V, with respect to w_i. This is true for the terms on the left-hand side, as we can see above. The last one can be shown in a similar way

$$\left| \int_0^T \langle b(t), w_i \rangle_V \, \varphi(t)\,\mathrm{d}t \right| \le \int_0^T \|b(t)\|_{V^*} \, \|w_i\|_V \, |\varphi(t)| \,\mathrm{d}t \le$$
$$\le C_4 \, \|b\|_{L_2((0,T);V^*)} \, \|w_i\|_V \,.$$

Therefore, (3.30) tends to (3.21) for $u_n \rightharpoonup \widetilde{u}$ in $L^2((0,T);V)$ and $v_n \to v$ in V as $n \to \infty$.

3.2.3.4 Galerkin approximations as operator equation

(i) We show that the Galerkin equations (3.16) tend to the first equation of (3.17).

From the integral identity (3.21), we obtain

$$-\int_0^T \langle \widetilde{u}(t), v \rangle_V \varphi'(t)\,\mathrm{d}t + \int_0^T \langle A\widetilde{u}(t), v \rangle_V \varphi(t)\,\mathrm{d}t - \int_0^T \langle b(t), v \rangle_V \varphi(t)\,\mathrm{d}t = 0$$

$\forall v \in V$, $\forall \varphi \in C_0^\infty((0,T))$. By (Zeidler 1990a, Proposition 23.9, (b)) along with the bilinearity of duality pairing and the definition of V^*, we get

$$\int_0^T \langle \widetilde{u}(t), v \rangle_V \varphi'(t)\,\mathrm{d}t = -\int_0^T \Big\langle \underbrace{b(t) - A\widetilde{u}(t)}_{=w(t)}, v \Big\rangle \varphi(t)\,\mathrm{d}t \quad \forall v \in V,\, \forall \varphi \in C_0^\infty((0,T)).$$

Then $w(t) = \widetilde{u}'(t)$ is a generalised derivative of $\widetilde{u}(t)$ on $(0,T)$, i.e., $b(t) - A\widetilde{u}(t) = \widetilde{u}'(t)$.

From

$$\int_0^T \langle \widetilde{u}'(t) + A\widetilde{u}(t) - b(t), v\rangle\, \varphi(t)\,\mathrm{d}t = 0 \quad \forall v \in V,\, \forall \varphi \in C_0^\infty((0,T)),$$

we arrive at

$$\widetilde{u}'(t) + A\widetilde{u}(t) = b(t) \quad \text{in } V^* \text{ for a.e. } t \in (0,T),$$

by definition of V^*.

(ii) Now, we show that $\widetilde{u} \in W$.

Since $\widetilde{u} \in L^2((0,T);V)$ and

$$(\exists C > 0)\colon\ \|Av\|_{V^*} \le C\,\|v\|_V \quad \forall v \in V, \tag{3.33}$$

we get

$$\|A\widetilde{u}\|_{L^2((0,T);V^*)} = \int_0^T \|A\widetilde{u}(t)\|_{V^*}^2\,\mathrm{d}t \le C^2 \int_0^T \|\widetilde{u}(t)\|_V^2\,\mathrm{d}t = C^2\,\|\widetilde{u}\|_{L^2((0,T);V)}^2 < \infty. \tag{3.34}$$

Hence, $A\widetilde{u} \in L^2((0,T);V^*)$. Furthermore, $b \in L^2((0,T);V^*)$. Therefore,

$$\widetilde{u}' = b - A\widetilde{u} \in L^2((0,T);V^*).$$

(iii) At last, we shall show that $\widetilde{u}(0) = u_0$.

Consider, once again, the integral identity (3.21), which could be rewritten to

$$-(u_0, v)_H \varphi(0) = \int_0^T \langle \widetilde{u}(t), v\rangle_V \varphi'(t)\,\mathrm{d}t + \int_0^T \langle \underbrace{b(t) - A\widetilde{u}(t)}_{=\widetilde{u}'(t)}, v\rangle_V \varphi(t)\,\mathrm{d}t$$

$\forall v \in V$, $\forall \varphi \in C^1([0,T])$, $\varphi(T) = 0$. As $\varphi(t)$ is a scalar constant in V, we can write

$$-(u_0, \varphi(0)v)_H = \int_0^T \langle \widetilde{u}(t), \varphi'(t)v\rangle_V + \langle \widetilde{u}'(t), \varphi(t)v\rangle_V\,\mathrm{d}t \tag{3.35}$$

$\forall v \in V$, $\forall \varphi \in C^1([0,T])$, $\varphi(T) = 0$. Furthermore, integration by parts formula (Zeidler 1990a, Proposition 23.23, (iv)) yields

$$\underbrace{(\widetilde{u}(T), \varphi(T)v)_H}_{=0} - (\widetilde{u}(0), \varphi(0)v)_H = \int_0^T \langle \widetilde{u}(t), \varphi'(t)v\rangle_V + \langle \widetilde{u}'(t), \varphi(t)v\rangle_V\,\mathrm{d}t.$$

We subtract the last equation from (3.35)

$$(\widetilde{u}(0), \varphi(0)v)_H - (u_0, \varphi(0)v)_H = 0.$$

In particular, for $\varphi(0) = 1$, we get

$$\left(\widetilde{u}(0) - u_0, v\right)_H = 0 \quad \forall v \in V.$$

Since V is dense in H, there exists a sequence (v_n) in V such that $v_n \to \widetilde{u}(0) - u_0$ in H as $n \to \infty$. Therefore,

$$\left|\left(\widetilde{u}(0) - u_0, v_n\right)_H\right| \to \|\widetilde{u}(0) - u_0\|_H = 0 \quad \Rightarrow \quad \widetilde{u}(0) = u_0.$$

The existence proof of the Theorem 3.2 is complete. We have proved that the weak limit (3.30) is the unique solution to the operator equation (3.17). Thus, $\widetilde{u} = u$.

3.2.4 Continuous dependence on input data

Denote

$$X := L^2((0,T);V),\ X^* := L^2((0,T);V^*). \tag{3.36}$$

The latter space is indeed the dual space to X since (2.4) holds. Combining

$$u_n \rightharpoonup u \quad \text{in } X \quad \text{as} \quad n \to \infty,$$

the estimate (3.19) and (Zeidler 1990a, Proposition 21.23, (c)), it follows that

$$\begin{aligned} \|u\|_X &\le \liminf_{n\to\infty} \|u_n\|_X \le \liminf_{n\to\infty} \left[K^{1/2} \left(\|u_n(0)\|_H^2 + \int_0^T \|b(t)\|_{V^*}^2 \, \mathrm{d}t \right)^{1/2} \right] \\ &= K^{1/2} \left(\|u_0\|_H^2 + \int_0^T \|b(t)\|_{V^*}^2 \, \mathrm{d}t \right)^{1/2}. \end{aligned} \tag{3.37}$$

Furthermore, using (3.34), we get

$$\begin{aligned} \|u\|_W^2 = \|u\|_X^2 + \|u'\|_{X^*}^2 &= \|u\|_X^2 + \|b - Au\|_{X^*}^2 \\ &\le \|u\|_X^2 + \left(\|b\|_{X^*} + \|Au\|_{X^*}\right)^2 \\ &\le \|u\|_X^2 + 2\,\|b\|_{X^*}^2 + 2\,\|Au\|_{X^*}^2 \\ &\le (1 + 2\,C^2)\,\|u\|_X^2 + 2\,\|b\|_{X^*}^2. \end{aligned}$$

By (3.37), we arrive at

$$\begin{aligned} \|u\|_W^2 &\le (1 + 2\,C^2) K \left(\|u_0\|_H^2 + \|b\|_{X^*}^2 \right) + 2\,\|b\|_{X^*}^2 \\ &= (1 + 2\,C^2) K\,\|u_0\|_H^2 + \left[(1 + 2\,C^2) K + 2 \right] \|b\|_{X^*}^2 \\ &\le \left[(1 + 2C^2) K + 2 \right] \left(\|u_0\|_H^2 + \int_0^T \|b(t)\|_{V^*}^2 \, \mathrm{d}t \right), \end{aligned}$$

so

$$\|u\|_W \le D\left(\|u_0\|_H^2 + \int_0^T \|b(t)\|_{V^*}^2 \,\mathrm{d}t\right)^{1/2}. \tag{3.38}$$

3.2.5 Convergence of Galerkin method in $C\big([0,T];H\big)$

We shall show that

$$\max_{0\le t\le T} \|u_n(t) - u(t)\|_H \to 0 \quad \text{as } n\to\infty. \tag{3.39}$$

We have $u \in W$, and $u_n \in W$, $\forall n$. Consequently,

$$u, u_n \in C([0,T];H) \quad \forall n,$$

by (3.14). Further, by (Zeidler 1990a, Proposition 23.23, (iii)), for each $u \in W$ and $\varepsilon > 0$ there exists a polynomial $p \in \mathcal{P}_n\big([0,T];V\big)$ such that

$$\|u-p\|_W = \left(\int_0^T \|u(t)-p(t)\|_V^2 \,\mathrm{d}t\right)^{1/2} + \left(\int_0^T \|u'(t)-p'(t)\|_{V^*}^2 \,\mathrm{d}t\right)^{1/2} < \varepsilon.$$

Set

$$V_n = \operatorname{span}\{w_1, w_2, \ldots, w_n\},$$

so there holds a relation

$$V_n \subseteq V \subseteq H.$$

The set $\bigcup\limits_n V_n$ is dense in V. Thus the set of all polynomials with coefficients in $\bigcup\limits_n V_n$ is dense in W. In other words, there exists a sequence of polynomials (p_n) in $\mathcal{P}_n\big([0,T];V_n\big)$ with

$$p_n \to u \quad \text{in } W \quad \text{as } n\to\infty. \tag{3.40}$$

From the embedding (3.14) it follows that

$$(\exists C>0)\colon \max_{0\le t\le T} \|u(t)-p_n(t)\|_H \le \mathrm{C}\,\|u-p_n\|_W \to 0 \quad \text{as } n\to\infty. \tag{3.41}$$

We need to prove that

$$\max_{0\le t\le T} \|u_n(t)-p_n(t)\|_H \to 0 \quad \text{as } n\to\infty. \tag{3.42}$$

(i) From (3.41) it follows that

$$\|u(0)-p_n(0)\|_H \le \max_{0\le t\le T} \|u(t)-p_n(t)\|_H \le \mathrm{C}\,\|u-p_n\|_W \to 0, \tag{3.43}$$

i.e.,

$$\|u(0)-p_n(0)\|_H \to 0.$$

Combining the latter with $u_n(0) \to u(0)$ in H as $n\to\infty$, we have

$$\|u_n(0)-p_n(0)\|_H \le \|u_n(0)-u(0)\|_H + \|u(0)-p_n(0)\|_H \to 0 \quad \text{as } n\to\infty.$$

(ii) The Galerkin approximations in the operator form with $v_n = u_n(t) - p_n(t)$ yield

$$\begin{aligned}\langle u_n'(t), u_n(t) - p_n(t)\rangle_V &= \langle b(t), u_n(t) - p_n(t)\rangle_V - \langle Au_n(t), u_n(t) - p_n(t)\rangle_V \\ &= \langle \underbrace{b(t)}_{=u'(t)+Au(t)} - Au_n(t), u_n(t) - p_n(t)\rangle_V \\ &= \langle u'(t) + A(u(t) - u_n(t)), u_n(t) - p_n(t)\rangle_V.\end{aligned} \tag{3.44}$$

(iii) Consider the notations (3.36), again. By integration by parts (Zeidler 1990a, Proposition 23.23, (iv)), we have

$$\frac{1}{2}\|u_n(t) - p_n(t)\|_H^2 - \frac{1}{2}\|u_n(0) - p_n(0)\|_H^2 = \int_0^t \langle u_n'(s) - p_n'(s), u_n(s) - p_n(s)\rangle_V \,\mathrm{d}s. \tag{3.45}$$

Further, using (3.44), we rewrite the right-hand side of the last equation

$$\begin{aligned}&\int_0^t \langle u'(s) + A(u(s) - u_n(s)) - p_n'(s), u_n(s) - p_n(s)\rangle_V \,\mathrm{d}s \\ &\quad = \int_0^t \langle u'(s) - p_n'(s), u_n(s) - p_n(s)\rangle_V \,\mathrm{d}s \\ &\qquad + \int_0^t \langle A(u(s) - u_n(s)), u_n(s) \underbrace{-u(s) + u(s)}_{=0} - p_n(s)\rangle_V \,\mathrm{d}s \\ &\quad = \int_0^t \langle u'(s) - p_n'(s), u_n(s) - p_n(s)\rangle_V \,\mathrm{d}s - \int_0^t \langle A(u(s) - u_n(s)), u(s) - u_n(s)\rangle_V \,\mathrm{d}s \\ &\qquad + \int_0^t \langle A(u(s) - u_n(s)), u(s) - p_n(s)\rangle_V \,\mathrm{d}s.\end{aligned}$$

By the positivity of $a(\cdot,\cdot)$, i.e., $-\langle A(u - u_n), u - u_n\rangle_V \le 0$, we obtain

$$\begin{aligned}&\int_0^t \langle u'(s) + A(u(s) - u_n(s)) - p_n'(s), u_n(s) - p_n(s)\rangle_V \,\mathrm{d}s \\ &\quad \le \int_0^t \langle u'(s) - p_n'(s), u_n(s) - p_n(s)\rangle_V \,\mathrm{d}s + \int_0^t \langle A(u(s) - u_n(s)), u(s) - p_n(s)\rangle_V \,\mathrm{d}s.\end{aligned}$$

Therefore, (3.45) yields

$$\begin{aligned}\frac{1}{2}\|u_n(t)-p_n(t)\|_H^2-\frac{1}{2}\|u_n(0)-p_n(0)\|_H^2 &\\ \le\int_0^t\langle u'(s)-p_n'(s),u_n(s)-p_n(s)\rangle_V\,\mathrm{d}s &\\ +\int_0^t\langle A(u(s)-u_n(s)),u(s)-p_n(s)\rangle_V\,\mathrm{d}s &\\ \le\int_0^t\|u'(s)-p_n'(s)\|_{V^*}\|u_n(s)-p_n(s)\|_V\,\mathrm{d}s &\\ +\int_0^t\|A(u(s)-u_n(s))\|_{V^*}\|u(s)-p_n(s)\|_V\,\mathrm{d}s.&\end{aligned}$$

Furthermore, by Hölder inequality with $t=T$ within integrals, we get

$$\begin{aligned}\frac{1}{2}\|u_n(t)-p_n(t)\|_H^2-\frac{1}{2}\|u_n(0)-p_n(0)\|_H^2 &\\ \le\|u'-p_n'\|_{X^*}\|u_n-p_n\|_X+\|Au-Au_n\|_{X^*}\|u-p_n\|_X&\end{aligned}$$

The sequences (u_n), (p_n) are bounded in X (weak convergence in X), and (Au_n) is bounded in V^* by (3.33), so

$$\begin{aligned}\frac{1}{2}\|u_n(t)-p_n(t)\|_H^2-\frac{1}{2}\|u_n(0)-p_n(0)\|_H^2 &\\ \le\max\{\|u_n-p_n\|_X,\|Au-Au_n\|_{X^*}\} &\\ \times(\|u-p_n\|_X+\|u'-p_n'\|_{X^*}) &\\ \le\text{const}\,\|u-p_n\|_W.&\end{aligned}$$

Hence,

$$\|u_n(t)-p_n(t)\|_H^2\to 0\quad\text{as } n\to\infty,$$

by (3.40) and (3.43). The limit (3.42) is proved.

Finally,

$$\begin{aligned}\max_{0\le t\le T}\|u_n(t)-u(t)\|_H&=\max_{0\le t\le T}\|u_n(t)\underbrace{-p_n(t)+p_n(t)}_{=0}-u(t)\|_H\\ &\le\max_{0\le t\le T}\|u_n(t)-p_n(t)\|_H+\max_{0\le t\le T}\|p_n(t)-u(t)\|_H\to 0\end{aligned}$$

as $n\to\infty$.

3.2.6 Strong convergence of Galerkin method in $L^2\Big((0,T);V\Big)$

Before we prove the strong convergence, we shall prepare some useful relations.

(i) We have $u_n \rightharpoonup u$ in X as $n \to \infty$ holds. This implies

$$Au_n \to Au \quad \text{in } X^*,$$

since A is linear and continuous mapping by (3.34). Therefore,

$$\int_0^T \langle Au_n(t), u(t)\rangle_V \,\mathrm{d}t \to \int_0^T \langle Au(t), u(t)\rangle_V \,\mathrm{d}t \quad \text{as } n \to \infty$$

and

$$\int_0^T \langle b(t), u_n(t)\rangle_V \,\mathrm{d}t \to \int_0^T \langle b(t), u(t)\rangle_V \,\mathrm{d}t \quad \text{as } n \to \infty.$$

We note that the variable t is omitted in the following text for the sake of simplicity. The first useful formula is obtained by imitating integration by parts formula (Zeidler 1990a, Proposition 23.9, (iv)) for u_n, u

$$\begin{aligned}\int_0^T \langle b - Au_n, u\rangle_V + \langle b - Au, u_n\rangle_V \,\mathrm{d}t \; - \big(u_n(T), u(T)\big)_H + \big(u_n(0), u(0)\big)_H \\ \to 2\int_0^T \langle \underbrace{b - Au}_{=u'}, u\rangle_V \,\mathrm{d}t - \|u(T)\|_H^2 + \|u(0)\|_H^2 = 0\end{aligned} \tag{3.46}$$

(ii) The next two relations are derived with the use of integration by parts formula, again.

$$\frac{1}{2}\|u(T) - u_n(T)\|_H^2 = \frac{1}{2}\|u(0) - u_n(0)\|_H^2 + \int_0^T \langle b - Au - u_n', u - u_n\rangle_V \,\mathrm{d}t \tag{3.47}$$

and

$$\big(u_n(T), u(T)\big)_H - \big(u_n(0), u(0)\big)_H = \int_0^T \langle u_n', u\rangle_V + \langle b - Au, u_n\rangle \,\mathrm{d}t, \tag{3.48}$$

where $u' = b - Au$. The Galerkin approximations (3.16) yields

$$\begin{aligned}\langle u_n', u_n\rangle_V + \langle Au_n, u_n\rangle_V &= \langle b, u_n\rangle_V \\ \langle b - Au_n - u_n', u_n\rangle_V &= 0.\end{aligned} \tag{3.49}$$

Since $a(\cdot,\cdot)$ is coercive, it follows that

$$\begin{aligned} c\,\|u-u_n\|_X^2 &\equiv \int_0^T c\,\|u(t)-u_n(t)\|^2\,\mathrm{d}t \le \\ &\le \int_0^T \langle A(u-u_n), u-u_n\rangle\,\mathrm{d}t \le \\ &\le \int_0^T \langle A(u-u_n), u-u_n\rangle\,\mathrm{d}t + \frac{1}{2}\,\|u(T)-u_n(T)\|_H^2\,. \end{aligned} \tag{3.50}$$

Finally, we can show that $u_n \to u$ in X as $n \to \infty$. Combining (3.50) with (3.47) yields

$$\begin{aligned} c\,\|u-u_n\|_X^2 \le \int_0^T &\langle Au, u-u_n\rangle_V - \langle Au, u-u_n\rangle_V\,\mathrm{d}t \\ &+ \langle b - Au_n - u_n', u-u_n\rangle_V\,\mathrm{d}t + \frac{1}{2}\,\|u(0)-u_n(0)\|_H^2 \\ = \int_0^T &\langle b - Au_n - u_n', u\rangle_V - \underbrace{\langle b - Au_n - u_n', u_n\rangle_V}_{=0 \text{ by } (3.49)}\,\mathrm{d}t \\ &+ \frac{1}{2}\,\|u(0)-u_n(0)\|_H^2 \end{aligned}$$

Furthermore, by (3.46) and (3.48), we get

$$\begin{aligned} c\,\|u-u_n\|_X^2 &\le \frac{1}{2}\,\|u(0)-u_n(0)\|_H^2 + \int_0^T \langle b - Au_n, u\rangle_V\,\mathrm{d}t - \int_0^T \langle u_n', u\rangle_V\,\mathrm{d}t = \\ &= \frac{1}{2}\,\|u(0)-u_n(0)\|_H^2 + \int_0^T \langle b - Au_n, u\rangle_V + \langle b - Au, u_n\rangle_V\,\mathrm{d}t - \\ &\quad - \big(u_n(T), u(T)\big)_H + \big(u_n(0), u(0)\big)_H \to 0 \end{aligned}$$

as $n \to \infty$. Hence,

$$\|u-u_n\|_X^2 \to 0 \quad \Rightarrow \quad u_n \to u \text{ in } X.$$

□

Chapter 4

Finite element semi-discrete method

4.1 Finite element scheme

Consider the equation (3.1). We assume that the domain Ω is an interval for the 1d spatial domain, polygonal for the 2d spatial domain, or a polyhedral for the 3d spatial domain. A semi-discrete finite element method is employed to solve (3.1). This method divides the spatial domain into smaller manageable sections known as finite elements. Once complete, a time-stepping scheme is used to solve the resulting system of ordinary differential equations. Therefore, we assume that the domain Ω is divided into finite elements

$$\mathcal{T}_{h_x} := \{\omega_{x,i} \subset \mathbb{R}^d\}_{i=1}^{M_x}; \quad \Omega := \bigcup_i \{\omega_{x,i} \colon \omega_{x,i} \in \mathcal{T}_{h_x}\},$$

where M_x is a number of elements $\omega_{x,i} \subset \mathbb{R}^d$ with mesh sizes

$$h_{x,i} := \operatorname{diam} \omega_{x,i} = \max_{x,y \in \omega_{x,i}} \|x - y\|$$

and a maximal mesh size

$$h_x := \max_{\omega_{x,i} \in \mathcal{T}_{h_x}} h_{x,i}.$$

The spatial elements $\omega_{x,i}$ represent intervals in 1d, triangles in 2d, and tetrahedral elements in 3d. Furthermore, let

$$\rho_{x,i} := \max_{S_i} \operatorname{diam} S_i,$$

where S_i is a ball inscribed in element $\omega_{x,i}$. We assume that the mesh $\mathcal{T}_{h_x}$ is a shape-regular

$$(\exists \sigma \geq 1)(\forall h_x > 0) \colon \max_{\omega_{x,i} \in \mathcal{T}_{h_x}} \frac{h_{x,i}}{\rho_{x,i}} \leq \sigma \tag{4.1}$$

and quasi-uniform

$$(\exists \tau > 0)(\forall h_x > 0) \colon \min_{\omega_{x,i} \in \mathcal{T}_{h_x}} h_{x,i} \geq \tau h_x. \tag{4.2}$$

For the sake of brevity, assume a finite element space

$$V_h := \operatorname{span} \{\varphi_1(\mathbf{x}), \varphi_2(\mathbf{x}), \ldots, \varphi_N(\mathbf{x})\} \subset V, \tag{4.3}$$

where $\varphi_i(\mathbf{x})$ are piece-wise linear nodal FEM basis functions on $\omega_{x,i}$. Thus, we have to find $u_h \in W_h \subset W$ such that

$$\begin{cases} c_H \dfrac{\partial}{\partial t}(u_h(t), v_h)_H + a(u_h(t), v_h) = \langle b(t), v_h \rangle_V \quad \forall v_h \in V_h \subset V \ \text{ and a.e. } t \in (0,T) \ , \\ u_h(0) = u_{0,h}. \end{cases} \tag{4.4}$$

In general, the initial condition $u_{0,h}$ represents some approximation of u_0 in V_h. For example, $u_{0,h}$ could be the L^2-projection of u_0 onto V_h.

Consider the discrete solution $u_h(\mathbf{x}, t)$, the discrete test function $v_h(\mathbf{x})$, and the discrete initial condition $u_{0,h}(\mathbf{x})$ defined by

$$u_h(\mathbf{x}, t) := \sum_{j=1}^{N} \big(\mathbf{u}(t)\big)_j \, \varphi_j(\mathbf{x}), \quad v_h(\mathbf{x}) := \sum_{i=1}^{N} v_i \, \varphi_i(\mathbf{x}), \quad u_{0,h}(\mathbf{x}) := \sum_{j=1}^{N} \big(\mathbf{u}_0\big)_j \, \varphi_j(\mathbf{x}), \tag{4.5}$$

where $\big(\mathbf{u}(t)\big)_j \in C^1((0,1))$ for all j. By replacing (4.5) in (4.4), we arrive at the Galerkin approximations (3.16) in a form

$$\begin{aligned} c_H \sum_{j=1}^{N} \big(\mathbf{u}'(t)\big)_j \, (\varphi_j(\mathbf{x}), \varphi_i(\mathbf{x}))_H + \sum_{j=1}^{N} \big(\mathbf{u}(t)\big)_j \, a\,(\varphi_j(\mathbf{x}), \varphi_i(\mathbf{x})) = \langle b(t), \varphi_i(\mathbf{x}) \rangle_V, \\ \big(\mathbf{u}(0)\big)_i = \big(\mathbf{u}_0\big)_i. \end{aligned} \tag{4.6}$$

$\forall i = 1, \ldots, N$. The equations (4.6) represent the system of the ordinary differential equations

$$\begin{cases} \mathrm{M}\,\mathbf{u}'(t) + \mathrm{A}\,\mathbf{u}(t) = \mathbf{b}(t) \quad \forall t \in (0,T), \\ \mathbf{u}(0) = \mathbf{u}_0, \end{cases} \tag{4.7}$$

where $\forall i, j = 1, 2, \ldots, N$

$$\begin{aligned} (\mathrm{M})_{i,j} &:= c_H \int_\Omega \varphi_i(\mathbf{x}) \, \varphi_j(\mathbf{x}) \, \mathrm{d}\mathbf{x}, \\ (\mathrm{A})_{i,j} &:= \int_\Omega \nabla_{\mathbf{x}} \varphi_i(\mathbf{x}) \, \nabla_{\mathbf{x}} \varphi_j(\mathbf{x}) \, \mathrm{d}\mathbf{x}, \\ (\mathbf{b}(t))_i &:= \int_\Omega f(t, \mathbf{x}) \, \varphi_i(\mathbf{x}) \, \mathrm{d}\mathbf{x}. \end{aligned}$$

The problem (4.7) is uniquely solvable since matrices A and M are symmetric positive definite.

Finally, we need to use a time-stepping method to solve the system of ordinary differential equations (4.7). To make it simple, we divide the time interval of $[0, T]$ into m equally spaced subintervals, using a time step of $\Delta t > 0$, i.e.,

$$[0, T] := \bigcup_{k=0}^{m-1} [t_k, t_{k+1}], \tag{4.8}$$

where $t_k = k \cdot \Delta t$, $k = 0, 1, \ldots, m$. Further, we approximate the time derivative of $\mathbf{u}$ by

$$\mathbf{u}'(t) \approx \frac{\mathbf{u}^{k+1} - \mathbf{u}^k}{\Delta t}, \tag{4.9}$$

where $\mathbf{u}^k := \mathbf{u}(t_k)$. Combining (4.9) with (4.7), we obtain the backward Euler scheme

$$\begin{cases} \mathrm{M}\dfrac{\mathbf{u}^{k+1} - \mathbf{u}^k}{\Delta t} + \mathrm{A}\,\mathbf{u}^{k+1} = \mathbf{b}^{k+1}, \quad k = 0, 1, \ldots, m, \\ \mathbf{u}(0) = \mathbf{u}_0, \end{cases} \tag{4.10}$$

and the Crank-Nicolson method

$$\begin{cases} \mathrm{M}\dfrac{\mathbf{u}^{k+1} - \mathbf{u}^k}{\Delta t} + \dfrac{1}{2}\mathrm{A}\left(\mathbf{u}^{k+1} + \mathbf{u}^k\right) = \dfrac{1}{2}\left(\mathbf{b}^{k+1} + \mathbf{b}^k\right), \quad k = 0, 1, \ldots, m, \\ \mathbf{u}(0) = \mathbf{u}_0, \end{cases} \tag{4.11}$$

where $\mathbf{b}^k := \mathbf{b}(t_k)$. The approximate solution takes the form given by (4.5),

$$u_h^k(\mathbf{x}) := \sum_{j=1}^{N} \left(\mathbf{u}^k\right)_j \varphi_j(\mathbf{x}). \tag{4.12}$$

4.2 Convergence results of semi-discrete method

Convergence results for the specified time-stepping methods have already been proposed in (Thomée 2006). In this instance, we shall adopt the presentation from (Foltyn et al. 2020) which shall be essential in studying the behaviour of the proposed schemes in combination with the Parareal method.

Consider the problem (3.1), where $u_0 \in L^2(\Omega)$ and $f \in L^2(Q_T)$, once more. We shall remind that according to Theorem 3.2, we possess the continuous dependence on the data (3.15), that is,

$$(\exists C > 0)\colon \ \|u\|_{L^2((0,T);H_0^1(\Omega))} + \|u'\|_{L^2((0,T);H^{-1}(\Omega))} \leq C\left(\|u_0\|_{L^2(\Omega)} + \|f\|_{L^2(Q_T)}\right). \tag{4.13}$$

Let us suppose that $u_h(\mathbf{x}, t)$, as defined by (4.5), is the solution to (4.7). Furthermore, let us assume that $u(\mathbf{x}, t)$ is the solution to (3.13), and $u_{0,h}(\mathbf{x}) = 0$ on Γ. Then, as stated in (Foltyn et al. 2020, Theorem 2.2), there exists a constant $C > 0$, independent of h, such that for $r \in [1, 2]$ and $t \geq 0$

$$\begin{aligned} \|u_h(\mathbf{x}, t) - u(\mathbf{x}, t)\|_{L^2(\Omega)} \leq{}& \|u_{0,h}(\mathbf{x}) - u_0(\mathbf{x})\|_{L^2(\Omega)} \\ &+ C\,h^r \left(\|u_0(\mathbf{x})\|_{H^r(\Omega)} + \int_0^t \left\| \frac{\partial u}{\partial s}(\mathbf{x}, s) \right\|_{H^r(\Omega)} \mathrm{d}s \right). \end{aligned}$$

and

$$\|\nabla_{\mathbf{x}} u_h(\mathbf{x},t) - \nabla_{\mathbf{x}} u(\mathbf{x},t)\|_{L^2(\Omega)} \leq \|\nabla_{\mathbf{x}} u_{0,h}(\mathbf{x}) - \nabla_{\mathbf{x}} u_0(\mathbf{x})\|_{L^2(\Omega)} + C\, h^{r-1} \left\{ \|u_0(\mathbf{x})\|_{H^r(\Omega)} + \|u(\mathbf{x},t)\|_{H^r(\Omega)} + \left(\int_0^t \left\| \frac{\partial u}{\partial s}(\mathbf{x},s) \right\|^2_{H^{r-1}(\Omega)} \mathrm{d}s \right)^{1/2} \right\} .$$

Consider the Euler method, where $u_h^k(\mathbf{x})$, $k \geq 0$, is the solution to (4.10), (4.12). Let there exists $K > 0$ that is independent of h, such that for all $r \in [1,2]$ the following holds:

$$\|u_{0,h}(\mathbf{x}) - u_0(\mathbf{x})\|_{L^2(\Omega)} \leq K\, h^r \|u_0(\mathbf{x})\|_{H^r(\Omega)} \tag{4.14}$$

and

$$u_0(\mathbf{x}) = 0 \text{ on } \Gamma. \tag{4.15}$$

Then by (Foltyn et al. 2020, Theorem 2.3), there exists $C > 0$ such that for $k \geq 0$ and $r \in [1;2]$, it holds that

$$\left\| u_h^k(\mathbf{x}) - u(\mathbf{x},t_k) \right\|_{L^2(\Omega)} \leq C\, h^r \left(\|u_0(\mathbf{x})\|_{H^r(\Omega)} + \int_0^{t_k} \left\| \frac{\partial u}{\partial s}(\mathbf{x},s) \right\|_{H^r(\Omega)} \mathrm{d}s \right) + \Delta t \int_0^{t_k} \left\| \frac{\partial^2 u}{\partial s^2}(\mathbf{x},s) \right\|_{L^2(\Omega)} \mathrm{d}s \ .$$

In similar way, consider the Crank-Nicolson method, where $u_h^k(\mathbf{x})$, $k \geq 0$, is the solution to (4.11), (4.12) and that exists $K > 0$ such that (4.14) and (4.15) hold true. Then by (Foltyn et al. 2020, Theorem 2.4), there exists $C > 0$ such that for $k \geq 0$ and $r \in [1,2]$ it holds that

$$\left\| u_h^k(\mathbf{x}) - u(\mathbf{x},t_k) \right\|_{L^2(\Omega)} \leq C\, h^r \left(\|u_0(\mathbf{x})\|_{H^r(\Omega)} + \int_0^{t_k} \left\| \frac{\partial u}{\partial s}(\mathbf{x},s) \right\|_{H^r(\Omega)} \mathrm{d}s \right) + C\, (\Delta t)^2 \int_0^{t_k} \left(\left\| \frac{\partial^3 u}{\partial s^3}(\mathbf{x},s) \right\|_{L^2(\Omega)} + \left\| \Delta_{\mathbf{x}} \left(\frac{\partial^2 u}{\partial s^2}(\mathbf{x},s) \right) \right\|_{L^2(\Omega)} \right) \mathrm{d}s.$$

To verify the estimates, we present numerical examples that are similar to those found in (Foltyn et al. 2020, Section 4.1). Within this thesis, we also extend the numerical evidence to 2d and 3d spatial problems.

4.3 Numerical experiments

The following numerical experiments were performed in MATLAB (with an Academic license obtained via VSB-TUO) on an HP-Spectre x360 Convertible 13-ap0xxx, using an Intel(R) Core(TM) i7-8565U processor with a clock speed of 1.80 GHz and 16GB of RAM. Two cases were examined. In the first scenario, we assumed a zero source term $f(\mathbf{x},t)$ but a non-zero initial condition $u_0(\mathbf{x})$. In contrast, the second scenario featured a non-zero source term $f(\mathbf{x},t)$ but a zero initial condition $u_0(\mathbf{x})$.

Thus, consider the system of ODEs (4.7) with $\Omega := (0,1)^d$ and $T := 2$.

1. **Example 1**: Let $c_H := 25$, $f(\mathbf{x},t) := 0$ and

$$u_0(\mathbf{x}) := \prod_{i=1}^{d} \sin(\pi x_i).$$

The exact solution to (3.1) for $d = 1, 2, 3$ is

$$u(\mathbf{x},t) := \mathrm{e}^{-d\cdot\frac{\pi^2}{c_H}t} \prod_{i=1}^{d} \sin(\pi x_i). \tag{4.16}$$

We study the errors of the Euler and Crank-Nicolson methods in the $L^2(\Omega)$-norm and the $H^1(\Omega)$-seminorm, that is,

$$\|u_h(\mathbf{x},T) - u(\mathbf{x},T)\|_{L^2(\Omega)} \quad \text{and} \quad \|\nabla_{\mathbf{x}} u_h(\mathbf{x},T) - \nabla_{\mathbf{x}} u(\mathbf{x},T)\|_{L^2(\Omega)}\,, \tag{4.17}$$

where $u_h(\mathbf{x},T) := u_h^m(\mathbf{x})$ represents the approximate solution (4.12) of the Euler method (4.10) and the Crank-Nicolson method (4.11) at the end time $T := m \cdot \Delta t$. The exact solution is denoted as $u(\mathbf{x},t)$. The spatial step h_x is the same for all d and is equal to the time step Δt, i.e., $h_x = \Delta t$. In Table 4.1, we present the errors of the Euler method in the $L^2(\Omega)$-norm and the $H^1(\Omega)$-seminorm. As noted in (Foltyn et al. 2020, Section 4.1), we can observe linear convergence of the Euler method for both cases through all spatial dimensions. In the $L^2(\Omega)$-norm, we observe some instabilities for coarser steps, but it was verified by further tests that the eoc tends to be 1 for finer steps. In Table 4.2, we provide the errors of the Crank-Nicolson scheme in the $L^2(\Omega)$-norm and the $H^1(\Omega)$-seminorm. The Crank-Nicolson method demonstrates quadratic convergence in the $L^2(\Omega)$-norm, and linear convergence in the $H^1(\Omega)$-seminorm for all d. Given results confirm ones from (Foltyn et al. 2020, Section 4.1). To achieve quadratic convergence for the Crank-Nicolson scheme in the $H^1(\Omega)$-seminorm, we have already suggested in (Foltyn et al. 2020, Section 4.1) to use of a higher-order finite element approximation in the spatial domain.

2. **Example 2**: Let $c_H := 1$, $u_0(\mathbf{x}) := 0$ and

$$f(\mathbf{x},t) := -(1-t)\cdot \mathrm{e}^{-t} \cdot \left[\prod_{i=1}^{d} x_i(x_i-1)\right] + 2\cdot t \cdot \mathrm{e}^{-t} \cdot \left\{\sum_{i=1}^{d}\left[\prod_{j=1;i\neq j}^{d} x_j(x_j-1)\right]\right\}, \tag{4.18}$$

where

$$\sum_{i=1}^{d}\left[\prod_{j=1;i\neq j}^{d} x_j(x_j-1)\right] = 1 \quad \text{for } d = 1.$$

The exact solution to (3.1) for $d = 1, 2, 3$ is

$$u(\mathbf{x},t) := -t\cdot \mathrm{e}^{-t} \cdot \left[\prod_{i=1}^{d} x_i(x_i-1)\right].$$

In Table 4.3, we display the errors in the $L^2(\Omega)$-norm and the $H^1(\Omega)$-seminorm of the Eu-

Table 4.1: Error of the backward Euler scheme – example 1.

$h_x = \Delta t$		1/8	1/16	1/32	1/64	1/128
$\|\cdot\|_{L^2(\Omega)}$	$d=1$	2.22e−3	1.34e−3	1.10e−3	6.65e−4	3.62e−4
	eoc		0.73	0.28	0.73	0.88
	$d=2$	2.10e−3	1.82e−3	1.44e−3	8.62e−4	4.66e−4
	eoc		0.21	0.34	0.74	0.89
	$d=3$	1.10e−3	1.67e−3	1.14e−3	6.49e−4	3.44e−4
	eoc		−0.60	0.55	0.81	0.92
$\|\cdot\|_{H^1(\Omega)}$	$d=1$	1.14e−1	5.76e−2	2.89e−2	1.45e−2	7.24e−3
	eoc		0.98	1.00	1.00	1.00
	$d=2$	9.02e−2	4.65e−2	2.37e−2	1.20e−2	6.01e−3
	eoc		0.96	0.77	1.19	1.00
	$d=3$	4.01e−2	2.21e−2	1.17e−2	6.01e−3	3.05e−3
	eoc		0.86	0.92	0.96	0.98

Table 4.2: Error of the Crank-Nicholson scheme – example 1.

$h_x = \Delta t$		1/8	1/16	1/32	1/64	1/128
$\|\cdot\|_{L^2(\Omega)}$	$d=1$	7.60e−3	1.91e−3	4.79e−4	1.20e−4	3.00e−5
	eoc		1.99	2.00	2.00	2.00
	$d=2$	8.91e−3	2.30e−3	5.78e−4	1.45e−4	3.62e−5
	eoc		1.95	1.99	2.00	2.00
	$d=3$	4.68e−3	1.22e−3	3.10e−4	7.76e−5	1.94e−5
	eoc		1.94	1.98	2.00	2.00
$\|\cdot\|_{H^1(\Omega)}$	$d=1$	1.15e−1	5.72e−2	2.86e−2	1.43e−2	7.15e−3
	eoc		1.01	1.00	1.00	1.00
	$d=2$	9.18e−2	4.52e−2	2.25e−2	1.12e−2	5.62e−3
	eoc		1.02	1.01	1.01	0.99
	$d=3$	4.19e−2	1.97e−2	9.65e−3	4.80e−3	2.40e−3
	eoc		1.09	1.03	1.01	1.00

ler scheme. It shows quadratic convergence in the L^2-norm for all spatial dimensions, as the Crank-Nicolson method does. This can be attributed to the sufficiently smooth behaviour of the source term in the spatial domain. In the $H^1(\Omega)$-seminorm, the Euler method behaves similarly to the previous example. The Crank-Nicolson method yields similar results as before, as seen in Table 4.4.

Table 4.3: Error of the backward Euler scheme – example 2.

$h_x = \Delta t$		1/8	1/16	1/32	1/64	1/128
$\|\cdot\|_{L^2(\Omega)}$	$d=1$	8.23e−4	2.10e−4	5.47e−5	1.47e−5	4.24e−6
	eoc		1.97	1.94	1.90	1.79
	$d=2$	3.99e−4	1.01e−4	2.55e−5	6.43e−6	1.63e−6
	eoc		1.98	1.99	1.99	1.98
	$d=3$	8.13e−5	2.06e−5	5.16e−6	1.29e−6	3.27e−7
	eoc		1.98	2.00	2.00	1.98
$\|\cdot\|_{H^1(\Omega)}$	$d=1$	1.95e−2	9.77e−3	4.88e−3	2.44e−3	1.22e−3
	eoc		1.00	1.00	1.00	1.00
	$d=2$	8.16e−3	4.11e−3	2.06e−3	1.03e−3	5.15e−4
	eoc		0.99	1.00	1.00	1.00
	$d=3$	1.96e−3	9.84e−4	4.93e−4	2.46e−4	1.23e−4
	eoc		0.99	1.00	1.00	1.00

Table 4.4: Error of the Crank-Nicholson scheme – example 2.

$h_x = \Delta t$		1/8	1/16	1/32	1/64	1/128
$\|\cdot\|_{L^2(\Omega)}$	$d=1$	7.95e−4	1.99e−4	4.97e−5	1.24e−5	3.11e−6
	eoc		2.00	2.00	2.00	2.00
	$d=2$	3.97e−3	1.01e−4	2.53e−5	6.33e−6	1.58e−6
	eoc		1.97	2.00	2.00	2.00
	$d=3$	8.12e−5	2.05e−5	5.14e−6	1.28e−6	3.21e−7
	eoc		1.99	2.00	2.00	2.00
$\|\cdot\|_{H^1(\Omega)}$	$d=1$	1.95e−2	9.77e−3	4.88e−3	2.44e−3	1.22e−3
	eoc		1.00	1.00	1.00	1.00
	$d=2$	8.16e−3	4.11e−3	2.06e−3	1.03e−3	5.15e−4
	eoc		1.00	1.00	1.00	1.00
	$d=3$	1.96e−3	9.84e−4	4.93e−4	2.46e−4	1.23e−4
	eoc		0.99	1.00	1.00	1.00

Chapter 5

Application of Parareal to solve partial differential equations

5.1 Parareal

The Parareal method is a parallel-in-time technique for solving ordinary differential equations (ODEs) that was first proposed in (Lions et al. 2001) along with a convergence estimate. This method is based on a predictor-corrector scheme that employs time-stepping techniques such as Euler, Crank-Nicolson, or Runge-Kutta. The Parareal method divides the given time interval into n independent subproblems using a coarse time step Δt, which are then solved in parallel using a finer step δt, $\delta t \ll \Delta t$. The resulting predicted values are corrected by the propagated error (the difference between the coarse and fine solutions at a given time) over the coarse problem, as can be seen in Section 5.1.1.

In this section, we recall the Parareal scheme presented in (Lions et al. 2001) to clarify its definition. Then, we update the scheme to incorporate the use of the Euler and Crank-Nicolson methods for solving partial differential equations (PDEs). In Section 5.1.3, we further update the Parareal scheme to reduce communication between processes. Additionally, the concept of a distributed program, which was introduced in (Aubanel 2011), is discussed. Lastly, we present an application of the Parareal method combined with a spatial domain decomposition method for 2d spatial problem + time.

In Section 5.2.1, we present an illustrative numerical experiment of an ordinary differential equation to demonstrate the application of the Parareal method. In Section 5.2.2, we describe the application of the Parareal method to the semi-discrete finite element method for solving PDEs. The efficiency of the distributed Parareal program is evaluated in Section 5.2.3. At last, in Section 5.2.4, we present a numerical experiment demonstrating the combination of the Parareal and spatial domain decomposition methods.

5.1.1 Scheme of Parareal method: ODE

Consider a linear ordinary differential equation

$$\begin{cases} \dfrac{\mathrm{d}y(t)}{\mathrm{d}t} = -a \cdot y(t),\ t \in [0;T], \\ y(0) = y_0, \end{cases} \tag{5.1}$$

where a represents a real constant, y_0 initial condition, and $T > 0$ time horizon. To solve the problem (5.1) by the Parareal, we use the implicit Euler scheme for simplicity.

First, we have to define the approximate coarse solutions $Y_0^k := y(T_k)$, $k = 0, 1, 2, \ldots, n-1$, $n \in \mathbb{N}$, with a step size $\Delta t > 0$, as follows:

$$\begin{cases} \dfrac{Y_0^{k+1} - Y_0^k}{\Delta t} + a \cdot Y_0^{k+1} = 0, \\ Y_0^0 = y_0. \end{cases} \tag{5.2}$$

The resulting solutions Y_0^k represent initial conditions for each time subinterval $[T_k, T_{k+1}]$, where $T_k = k \cdot \Delta t$, giving us equally spaced subintervals. As a result, we have defined n independent subproblems

$$\begin{cases} \dfrac{\mathrm{d}y_{k,0}}{\mathrm{d}t}(t) = -a \cdot y_{k,0}(t),\ t \in [T_k, T_{k+1}], \\ y_{k,0}(T_k) = Y_0^k, \end{cases} \tag{5.3}$$

which can be solved in parallel. These problems are solved with the fine step $\delta t > 0$, $\delta t \ll \Delta t$, that is,

$$\begin{cases} \dfrac{y_{k,0}^{l+1} - y_{k,0}^l}{\delta t} + a \cdot y_{k,0}^{l+1} = 0 \text{ in } [T_k, T_{k+1}], \\ y_{k,0}^0 = Y_0^k, \end{cases} \tag{5.4}$$

where $l = 0, 1, 2, \ldots, m-1$, $m \in \mathbb{N}$, $y_{k,0}^l := y_{k,0}(t_l)$, $t_l \in [T_k, T_{k+1}]$, and $y_{k,0}^0 := y_{k,0}(T^k)$. We assume that m is common for all n subintervals in this case. After the fine solutions are computed, we define a loop of the method for $i = 1, 2, \ldots, n$, as follows:

1. Compute the jumps (the difference between the fine and the coarse solution at time T_k):

$$\begin{aligned} &s_i^k = y_{k-1,i-1}(T_k) - Y_{i-1}^k, \quad k = 1, 2, \ldots, n-1, \\ &s_i^0 = 0. \end{aligned} \tag{5.5}$$

2. Propagate these jumps with the coarse step

$$\begin{cases} \dfrac{q_i^{k+1} - q_i^k}{\Delta t} + a \cdot q_i^{k+1} = \dfrac{s_i^k}{\Delta t}, \quad k = 0, 1, \ldots, n-1, \\ q_i^0 = 0. \end{cases} \tag{5.6}$$

3. Update the coarse solution:

$$\begin{aligned} &Y_i^k = y_{k-1,i-1}(T_k) + q_i^k, \quad k = 1, 2, \ldots, n-1, \\ &Y_i^0 = Y_0^0. \end{aligned} \tag{5.7}$$

4. Solve the fine problems again

$$\begin{cases} \dfrac{y_{k,i}^{l+1} - y_{k,i}^{l}}{\delta t} + a \cdot y_{k,i}^{l+1} = 0 \text{ in } [T_k, T_{k+1}],\ l = 0, 1, \ldots, m-1,\ k = 0, 1, \ldots, n-1, \\ y_{k,i}^{0} = \mathrm{Y}_i^k. \end{cases} \tag{5.8}$$

Since we have defined n independent subproblems, the algorithm converges at least for $i = n$. Moreover, the first and the second step of the loop could be rewritten into

$$\begin{cases} e'_{k,i}(t) + a \cdot e_{k,i}(t) = 0 \quad \forall t \in [T_k, T_{k+1}],\ k = 0, 1, \ldots, n-1, \\ e_{k,i}(T_k) - e_{k-1,i}(T_k) = \underbrace{y_{k-1,i-1}(T_k) - Y_{i-1}^k}_{=s_i^k},\ k = 1, 2, \ldots, n, \\ e_{0,i}(0) = 0, \end{cases} \tag{5.9}$$

$\forall i = 1, 2, \ldots, n$, as it was proposed in (Foltyn et al. 2020), where $q_i^k := e_{k,i}(T_k)$. The propagation of the error in the form (5.9) explains how to propagate jumps while using the Parareal to solve PDEs. It is worth noting that the described scheme can also be extended to nonlinear equations as described (Lions et al. 2001).

5.1.2 Scheme of Parareal method: PDE

Consider the system of ODEs (4.7), which is obtained through the finite element semi-discretisation method. Utilising the Euler method to solve such a problem, the coarse solution (5.2) and the fine solutions (5.4) and (5.8) are computed by (4.10). As a result, the propagation of the jumps (5.6) can be defined as

$$\begin{cases} \left(\dfrac{1}{\Delta t} \cdot \mathrm{M} + \mathrm{A}\right) \cdot \mathbf{q}_i^{k+1} = \dfrac{1}{\Delta t} \cdot \mathrm{M} \cdot \left(\mathbf{q}_i^k + \mathbf{s}_i^k\right), \quad k = 0, 1, \ldots, n-1, \\ \mathbf{q}_i^0(0) = \mathbf{0}, \end{cases} \tag{5.10}$$

$\forall i = 1, \ldots, n$, where n is the number of time subintervals. To better understand (5.10), we refer to (5.9).

Proceeding to the Crank-Nicolson method, the coarse solution (5.2) and the fine solutions (5.4) and (5.8) are calculated by (4.11). To ensure consistency, the propagation of jumps, as outlined in (5.6), is resolved by

$$\begin{cases} \left(\dfrac{1}{\Delta t} \cdot \mathrm{M} + \dfrac{1}{2}\mathrm{A}\right) \cdot \mathbf{q}_i^{k+1} = \left(\dfrac{1}{\Delta t} \cdot \mathrm{M} - \dfrac{1}{2}\mathrm{A}\right) \cdot \left(\mathbf{q}_i^k + \mathbf{s}_i^k\right), \quad k = 0, 1, \ldots, n-1, \\ \mathbf{q}_i^0(0) = \mathbf{0}, \end{cases} \tag{5.11}$$

$\forall i = 1, \ldots, n$. Using the Euler method to propagate the error is also feasible, even if the Crank-Nicolson scheme is employed to solve the coarse and the fine solutions. As previously stated, we present the error propagation by the Crank-Nicolson method to ensure consistency for subsequent experiments.

5.1.3 Parareal method as MPI distributed program

We have adopted an idea presented in (Aubanel 2011). The author proposed the following three implementation options for the Parareal algorithm:

1. **Manager/Workers with overlap**: Consider equally spaced time subintervals for the fine solutions in this option. Initially, the Manager (main processor) calculates the coarse solution and subsequently dispatches the corresponding time slices to the Workers (other processors) as the initial conditions. The Workers compute their fine solutions and transmit the result back to the Manager. The Manager returns the updated coarse solution to the individual Workers and the loop restarts. As the Manager and the Workers have the same length of time subintervals for the fine solution, a gap exists between the Parareal loop's steps. The Workers have to wait until the Manager calculates its fine solution, collects the other fine solutions from the Workers, and sends the updated coarse solutions back to the Workers.

2. **Manager/Workers with improved overlap**: The proposed option is akin to the preceding one, except that in this scenario, we utilise a shorter time interval for the fine solution on the Manager, thereby reducing the gap between iterations. The rest of the overall time interval $[0, T]$ is divided evenly among the Workers.

3. **Distributed program**: The last option of the algorithm employs the concept of distributed programming, wherein the initial processor evaluates the coarse propagator only once. Subsequently, it transmits the outcome to the next processor. The initial processor then calculates its fine solution, updates the corresponding coarse solution and forwards it to the next processor again. Then the initial processor is over. After the next processor receives the coarse solution from the preceding process, it concurrently evaluates the next coarse propagator and transfers the resulting coarse solution to the following processor in a sequential manner. The processor then calculates the fine solution, awaits receipt of the corrected coarse solution from the preceding processor, then updates its coarse solution and transmits the result to the next processor. This procedure repeats until the specified precision is obtained or the n iterations are reached. The distinction between the distributed algorithm and the previous two options is that it requires less memory at the expense of slightly reduced efficiency, as stated in (Aubanel 2011). The distributed algorithm was subject to testing in Section 5.2.3.

As described above, there is minimal communication overhead within the iterations. The algorithms only transmit and receive the corresponding time slice of the coarse solution. To accomplish this, the Parareal scheme outlined in Section 5.1.1 must be updated. One potential solution is to reduce steps (5.5), (5.6) and (5.7) into a single step, as represented by equation

$$U_{k+1}^{i+1} := \mathcal{G}(U_k^{i+1}) + \mathcal{F}(U_k^i) - \mathcal{G}(U_k^i) \tag{5.12}$$

$k = 0, 1, \ldots, n-1$, $i = 1, 2, \ldots, n$, where $\mathcal{G}$ is the coarse propagator with time step Δt, $\mathcal{F}$ the fine propagator with time step δt, and U_k^i is the coarse solution at the time T_k in the ith iteration of the Parareal. Here, the term $\mathcal{G}(U_k^{i+1}) - \mathcal{G}(U_k^i)$ represents the propagated error (5.9)

by which the predicted solution $\mathcal{F}(U_k^i)$ is corrected. This concept has already been proposed in (Aubanel 2011; Maday 2008). If steps (5.5), (5.6) and (5.7) are retained in their original form, the corresponding jumps s_k^i and propagated jumps q_k^i have to be also transmitted and received by the processors in the distributed algorithm, resulting in increased communication overhead. In contrast, in the Manager/Workers version, the update of the coarse solution is performed on the Manager side. Thereby, the same level of communication overhead for the algorithm can be maintained.

The author of (Aubanel 2011) also introduces theoretical speedups for each version of the Parareal method. The speedup for the distributed algorithm, which was tested here, is

$$\psi := \frac{N}{Nr + i(1+r)}, \tag{5.13}$$

where r is the ratio of the time taken by the coarse propagator $\mathcal{G}$ to the time of the fine propagator $\mathcal{F}$, N is the number of processors (CPUs), and i is the ith iteration of the Parareal loop.

5.1.4 Combining Parareal and spatial DDM to solve PDE

In order to enhance the parallelism of the Parareal method, the spatial domain can be resolved by utilising a domain decomposition method (DDM). DDMs are well-established techniques for solving elliptic partial differential equations in parallel. In this particular case, we shall focus on the non-overlapping variant, based on the Schur complement's approximation, as described in (Bramble et al. 1986; Lukáš et al. 2015; Toselli et al. 2004).

Consider the parabolic problem (3.1) in 2d. Furthermore, assume that the spatial domain Ω is discretised into the non-overlapping right-angled isosceles triangles using piece-wise basis functions (4.3), resulting in the system of ordinary differential equations (4.7). Additionally, the time interval $[0, T]$ is divided into m equidistant subintervals through a time step $\Delta t > 0$, allowing for the use of, for example, the Euler method (4.10). Therefore, at each iteration of the Euler scheme (4.10), symmetric positive-definite (SPD) system, denoted as

$$\mathrm{A}\,\mathbf{u} = \mathbf{b}, \tag{5.14}$$

is solved. The SPD system can subsequently be decomposed into N independent subproblems. To achieve this, the spatial domain Ω has to be discretised using two discretisation steps: a fine step $h_x > 0$ resulting in FEM elements (triangles), and a coarse step $H_x > 0$, where $h_x \ll H_x$, resulting in independent subdomains Ω_i, $i = 1, 2, \ldots, N$ (blue dots in Figure 5.1). The boundaries of the subdomains located within the domain Ω define the so-called skeleton (red crosses in Figure 5.1), which serves as the interface between adjacent subdomains Ω_i.

After the discretisation mentioned above, the FEM basis functions (4.3) are regrouped into $N + 1$ sets. The N sets, denoted as $\mathcal{I}_i$, where $i = 1, 2, \ldots, N$, are defined by the interior nodes belonging to each individual subdomain Ω_i, i.e., the sets are formed by indices whose basis functions have support in $\overline{\Omega_i}$. The last $(N + 1)$-th set, denoted as $\mathcal{S}$, includes the nodes corresponding to the skeleton or, in general, to the Neumann part of the boundary $\partial\Omega$. Nodes associated with the Dirichlet boundary condition are not considered within the decomposition process.

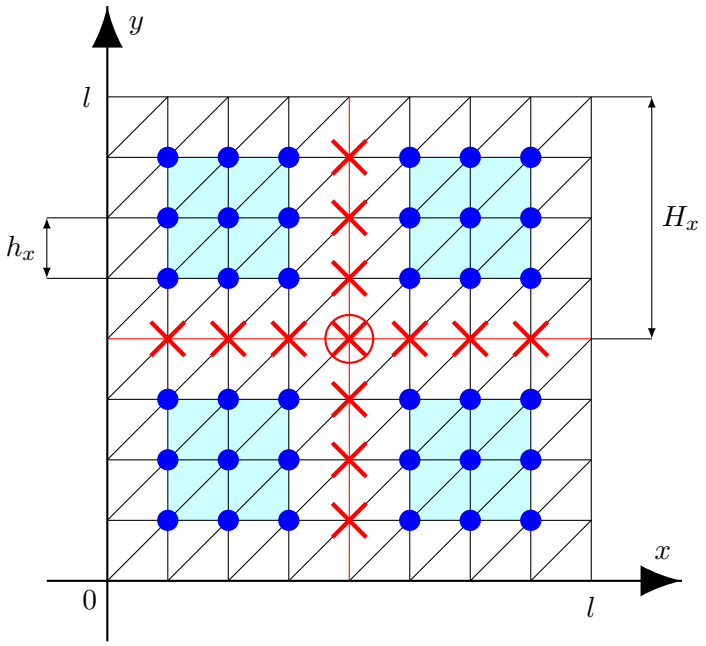


Figure 5.1: Subdomains and skeleton of used DDM.

The resulting matrix of the system (5.14) is reordered into a form

$$\mathrm{A} := \begin{pmatrix} \mathrm{A}_{\mathcal{I}\mathcal{I}} & \mathrm{A}_{\mathcal{I}\mathcal{S}} \\ \mathrm{A}_{\mathcal{S}\mathcal{I}} & \mathrm{A}_{\mathcal{S}\mathcal{S}} \end{pmatrix}, \tag{5.15}$$

where $\mathrm{A}_{\mathcal{I}\mathcal{I}}$ is a block diagonal matrix (each block represents a single inner subdomain), $\mathrm{A}_{\mathcal{S}\mathcal{S}}$ consists of the skeleton's indices only and matrices $\mathrm{A}_{\mathcal{I}\mathcal{S}}$, $\mathrm{A}_{\mathcal{S}\mathcal{I}}$ describe the contributions of inner subdomains to the skeleton. The Schur complement is represented by the equation

$$\mathrm{S} := \mathrm{A}_{\mathcal{S}\mathcal{S}} - \mathrm{A}_{\mathcal{S}\mathcal{I}}\,\mathrm{A}_{\mathcal{I}\mathcal{I}}^{-1}\,\mathrm{A}_{\mathcal{I}\mathcal{S}}. \tag{5.16}$$

Using notations (5.15), (5.16) and a particular-solution approach, the origin system (5.14) is solved within three steps:

1. $\mathrm{A}_{\mathcal{I}_i\mathcal{I}_i}\,\mathbf{u}_{\mathcal{I}_i}^P = \mathbf{b}_{\mathcal{I}_i}$, $i = 1, 2, \ldots, N$,
2. $\mathrm{S}\,\mathbf{u}_{\mathcal{S}}^H = \mathbf{b}_{\mathcal{S}} - \sum_{i=1}^{N} \mathrm{A}_{\mathcal{S}\mathcal{I}_i}\,\mathbf{u}_{\mathcal{I}_i}^P$,
3. $\mathrm{A}_{\mathcal{I}_i\mathcal{I}_i}\,\mathbf{u}_{\mathcal{I}_i}^H = -\mathrm{A}_{\mathcal{I}_i\mathcal{S}}\,\mathbf{u}_{\mathcal{S}}^H$, $i = 1, 2, \ldots, N$.

The solution is given by $\mathbf{u} := \mathbf{u}^H + \mathbf{u}^P$. Steps 1 and 3 can be solved concurrently in parallel. The critical step is the second step, in which the expensive linear system of the Schur complement S must be evaluated. The idea behind the given DDM, referred to as the vertex-based method, is to replace the Schur complement S with its approximation $\hat{\mathrm{S}}$, for which the inversion is less costly to compute.

Initially, the nodes are regrouped in a similar manner as it was done for the matrix A. By $\mathcal{E}_i$, where $i = 1, 2, \ldots, N_{\mathcal{E}}$ and $N_{\mathcal{E}}$ is the number of all edges in the skeleton, we denote a set

of interior nodes forming a single edge of the skeleton (red crosses in Figure 5.1). The set of vertices joining adjacent edges of the skeleton (red cross in a circle in Figure 5.1) is denoted by $\mathcal{V}$. The original Schur complement S is subsequently rearranged into a form

$$\mathrm{S} := \begin{pmatrix} \mathrm{S}_{\mathcal{E}\mathcal{E}} & \mathrm{S}_{\mathcal{E}\mathcal{V}} \\ \mathrm{S}_{\mathcal{V}\mathcal{E}} & \mathrm{S}_{\mathcal{V}\mathcal{V}} \end{pmatrix}. \tag{5.17}$$

Furthermore, we introduce a matrix $\mathrm{R}_{\mathcal{E}}$ which represents an extension of the basis functions $\varphi(\mathbf{x})$ covering the adjacent finite elements to the basis functions $\hat{\varphi}(\mathbf{x})$ over the skeleton edges, as shown in Figure 5.2. Since we are considering triangular or rectangular subdomains Ω_i, the interpolation of the function values $\hat{w}(\mathbf{x})$ into the interior nodes of the corresponding edge is linear.

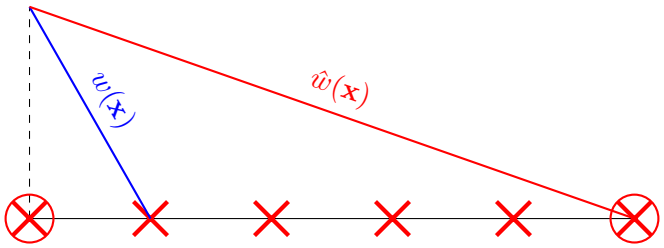


Figure 5.2: Extension of a basis function.

With $\mathrm{R}_{\mathcal{E}}$ matrix at hand, the Schur complement (5.17) is decomposed into

$$\mathrm{S} := \begin{pmatrix} \mathrm{I}_{\mathcal{E}} & \mathrm{O} \\ -\mathrm{R}_{\mathcal{E}} & \mathrm{I}_{\mathcal{V}} \end{pmatrix} \begin{pmatrix} \mathrm{S}_{\mathcal{E}\mathcal{E}} & \tilde{\mathrm{S}}_{\mathcal{E}\mathcal{V}} \\ \tilde{\mathrm{S}}_{\mathcal{V}\mathcal{E}} & \tilde{\mathrm{S}}_{\mathcal{V}\mathcal{V}} \end{pmatrix} \begin{pmatrix} \mathrm{I}_{\mathcal{E}} & -\mathrm{R}_{\mathcal{E}}^{\mathrm{T}} \\ \mathrm{O} & \mathrm{I}_{\mathcal{V}} \end{pmatrix}, \tag{5.18}$$

where $\mathrm{I}_{\mathcal{E}}$, $\mathrm{I}_{\mathcal{V}}$ are identity matrices and O are zero matrices having appropriate size. In the approximation of the Schur complement $\hat{\mathrm{S}}$, the matrix $\mathrm{S}_{\mathcal{E}\mathcal{E}}$ is replaced by its block-diagonal part $\hat{\mathrm{S}}_{\mathcal{E}\mathcal{E}} := \mathrm{diag}\left(\mathrm{S}_{\mathcal{E}1,\mathcal{E}1}, \mathrm{S}_{\mathcal{E}2,\mathcal{E}2}, \ldots, \mathrm{S}_{\mathcal{E}_{N\mathcal{E}},\mathcal{E}_{N_{\mathcal{E}}}}\right)$. Furthermore, the matrices $\tilde{\mathrm{S}}_{\mathcal{E}\mathcal{V}}$ and $\tilde{\mathrm{S}}_{\mathcal{V}\mathcal{E}}$ are omitted. The matrix $\tilde{\mathrm{S}}_{\mathcal{V}\mathcal{V}}$ represents the contribution of the basis functions $\hat{w}(\mathbf{x})$ over the skeleton. In other words, it represents the bilinear form A over the coarse mesh, denoted as A_H, i.e, $\mathrm{A}_H := \tilde{\mathrm{S}}_{\mathcal{V}\mathcal{V}}$.

Finally, the Schur complement approximation $\hat{\mathrm{S}}$ reads

$$\hat{\mathrm{S}} = \begin{pmatrix} \mathrm{I}_{\mathcal{E}} & \mathrm{O} \\ -\mathrm{R}_{\mathcal{E}} & \mathrm{I}_{\mathcal{V}} \end{pmatrix} \begin{pmatrix} \hat{\mathrm{S}}_{\mathcal{E}\mathcal{E}} & \mathrm{O} \\ \mathrm{O} & \mathrm{A}_H \end{pmatrix} \begin{pmatrix} \mathrm{I}_{\mathcal{E}} & -\mathrm{R}_{\mathcal{E}}^{\mathrm{T}} \\ \mathrm{O} & \mathrm{I}_{\mathcal{V}} \end{pmatrix}. \tag{5.19}$$

As all resulting linear systems (5.14) can be solved by, e.g., preconditioned conjugate gradient method, the inversion of $\hat{\mathrm{S}}$ is needed. The formula is

$$\hat{\mathrm{S}}^{-1} = \sum_{i=1}^{N_{\mathcal{E}}} \left[\begin{pmatrix} \mathrm{I}_{\mathcal{E}_i} \\ \mathrm{O} \end{pmatrix} \left(\hat{\mathrm{S}}_{\mathcal{E}_i\,\mathcal{E}_i}\right)^{-1} \begin{pmatrix} \mathrm{I}_{\mathcal{E}_i} & \mathrm{O} \end{pmatrix} \right] + \begin{pmatrix} (\mathrm{R}_{\mathcal{E}})^T \\ \mathrm{I}_{\mathcal{V}} \end{pmatrix} (\mathrm{A}_H)^{-1} \begin{pmatrix} \mathrm{R}_{\mathcal{E}} & \mathrm{I}_{\mathcal{V}} \end{pmatrix}. \tag{5.20}$$

This gives us a modification of the second step of the particular-solution approach, as defined above, which can take advantage of solving $N_{\mathcal{E}}$ independent subproblems over individual edges in parallel. The modification is as follows

2a. Set

$$\mathbf{c}_{\mathcal{S}} := \begin{pmatrix} \mathbf{c}_{\mathcal{E}} \\ \\ \mathbf{c}_{\mathcal{V}} \end{pmatrix} = \mathbf{b}_{\mathcal{S}} - \mathrm{A}_{\mathcal{S}\mathcal{I}}\, \mathbf{u}_{\mathcal{I}}^{P}.$$

2b. Solve $N_{\mathcal{E}}$ independent local systems $\mathrm{S}_{\mathcal{E}_i \mathcal{E}_i}\, \mathbf{w}_{\mathcal{E}_i} = \mathbf{c}_{\mathcal{E}}$.

2c. Solve the global coarse system $\mathrm{A}_H\, \mathbf{w}_H = \mathbf{c}_{\mathcal{V}} + \mathrm{R}_{\mathcal{E}}\, \mathbf{c}_{\mathcal{E}}$.

2d. Set

$$\hat{\mathbf{u}}_{\mathcal{S}}^{H} := \begin{pmatrix} \mathbf{w}_{\mathcal{E}} + (\mathrm{R}_{\mathcal{E}})^{\mathrm{T}}\, \mathbf{w}_H \\ \\ \mathbf{w}_H \end{pmatrix}.$$

Given modification is proposed in (Lukáš et al. 2015) along with the resulting condition number

$$\kappa\left(\hat{\mathrm{S}}^{-1}\, \mathrm{S}\right) \leq C \left(1 + \log \frac{H}{h}\right)^2, \tag{5.21}$$

where $C > 0$ depends only on the shape of Ω.

5.2 Numerical experiments

We have previously defined the solution of the system of ODEs (4.7) using the Euler and Crank-Nicolson methods in Section 4.3. An example of solving such a system of ODEs using the Parareal method is provided in Section 5.2.1. The application of the Parareal method to solve PDEs is discussed in Section 5.2.2. In Section 5.2.3, the Parareal method as a distributed algorithm is tested. Finally, in Section 5.2.4, a numerical experiment combining the Parareal method and the spatial domain decomposition method is presented.

5.2.1 Solving ODE by Parareal

Consider a linear ODE problem

$$\begin{cases} \dfrac{\mathrm{d}y}{\mathrm{d}t}(t) = -y(t),\ t \in [0, 1], \\ \ \ y(0) = 1. \end{cases} \tag{5.22}$$

The exact solution of the (5.22) is known, that is,

$$y(t) := \mathrm{e}^{-t}. \tag{5.23}$$

In order to demonstrate the convergence behaviour of the Parareal method, the problems outlined in equations (5.4) and (5.8) are solved using the analytic solution, i.e., we solve the problem (5.3). No fine solutions are used in this example. The coarse time step Δt is set to $\frac{1}{4}$. Table

5.1 shows the computed values of the jumps (5.5) for each iteration of the algorithm. As can be observed, the last non-zero value occurs in the fourth iteration, i.e., $i = 4$. Therefore, the exact solution is obtained once the problem (5.8) is solved for $i = n = 4$. This experiment serves as an example of how the Parareal algorithm converges after n steps, at the very least. Of course, the algorithm may converge sooner by using certain stopping criteria.

Table 5.1: The jumps of the ith iteration of the loop using the analytic solution.

	s_i^0	s_i^1	s_i^2	s_i^3	s_i^4
$i = 1$	0	$-2.12e{-2}$	$-1.70e{-2}$	$-1.36e{-2}$	$-1.09e{-2}$
$i = 2$	0	0	$4.49e{-4}$	$7.19e{-4}$	$8.63e{-4}$
$i = 3$	0	0	0	$-9.53e{-6}$	$-2.29e{-5}$
$i = 4$	0	0	0	0	$2.02e{-7}$

Table 5.2 displays the computed error values, which are calculated as

$$e^i := \max_k |y_{k-1,i-1}(T_k) - y_{exact}(T_k)|,$$

where $y_{k-1,i-1}(T_k)$ is the analytic solution at the time T_k of the $(i-1)$th Parareal iteration and $y_{exact}(T_k)$ is the exact solution (5.23) at the time T_k. The decimal logarithm of this error is then plotted in Figure 5.3. The $i = 0$ column in the table represents the state before the loop is initiated, that is, after the first analytic solutions (5.3) have been solved. The zero values in the columns $i = 3$ and $i = 4$ indicate that the computational arithmetic precision has been achieved.

Table 5.2: Error in the maximum norm using the analytic solution.

	$i = 0$	$i = 1$	$i = 2$	$i = 3$	$i = 4$
e^i	$3.09e{-2}$	$8.33e{-4}$	$7.42e{-6}$	0	0

The convergence is similar to the previous case while solving (5.4) and (5.8) approximately, using the backward Euler method with fine step δt. However, it is important to note that the precision of the Parareal method is determined by the time-stepping method used with the step δt over the entire interval $[0, T]$, i.e., the sequential Euler method. Therefore, an appropriate δt has to be selected to achieve sufficient precision.

In this example, the system of equations (5.4) and (5.8) are being solved with a fine step $\delta t = \frac{1}{2^{20}}$. The coarse time step $\Delta t = \frac{1}{4}$ remains the same. The computed jumps are presented in Table 5.3. As in the previous example, the convergence of the method is demonstrated using the maximum norm

$$e^i := \max_k |y_{k-1,i-1}(T_k) - y_{seq}(T_k)|, \tag{5.24}$$

where $y_{k-1,i-1}(T_k)$ is the fine solution at the time T_k of the $(i-1)$th Parareal iteration and $y_{seq}(T_k)$ is the sequential solution at the time T_k. Table 5.4 presents the resulting error values. The decimal logarithm of the error values is plotted in Figure 5.4. It can be seen that the al-

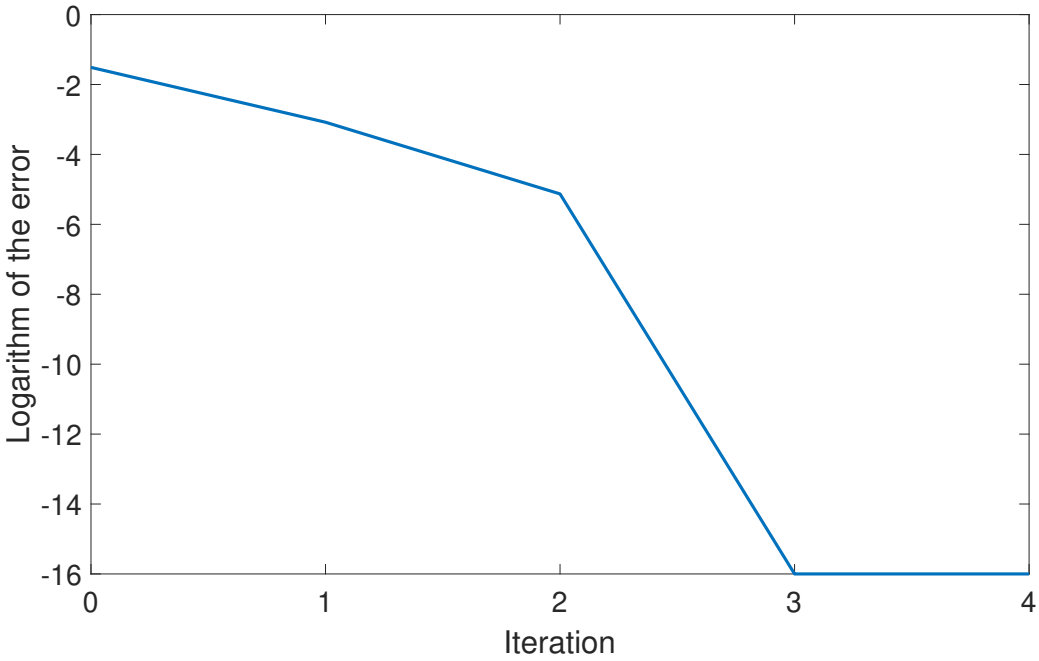


Figure 5.3: Logarithm of the error using the analytic solution (Table 5.2).

gorithm has the same convergence rate as in the case where the fine solutions were computed using the analytic equation.

Table 5.3: The jumps of the ith iteration of the loop using the approximate solution.

	s_i^0	s_i^1	s_i^2	s_i^3	s_i^4
$i=1$	0	$-2.12e{-2}$	$-1.70e{-2}$	$-1.36e{-2}$	$-1.09e{-2}$
$i=2$	0	0	$4.49e{-4}$	$7.19e{-4}$	$8.63e{-4}$
$i=3$	0	0	0	$-9.53e{-6}$	$-2.29e{-5}$
$i=4$	0	0	0	0	$2.02e{-7}$

Table 5.4: Error in the maximum norm using the approximate solution.

	$i=0$	$i=1$	$i=2$	$i=3$	$i=4$
e^i	$3.09e{-2}$	$8.33e{-4}$	$7.42e{-6}$	0	0

In the final ODE experiment, we examine the use of processors with between 4 and 64 cores. Our objective is to determine if the utilisation of a greater number of available cores, through the increase in the number of independent subintervals, results in a faster convergence rate for the Parareal algorithm. It is assumed that the coarse step Δt and the fine step δt are defined as follows:

$$\Delta t := \frac{1}{2^{j+1}} \quad j = 1, 2, \ldots, 5; \quad \delta t := \frac{1}{2^{20}}.$$

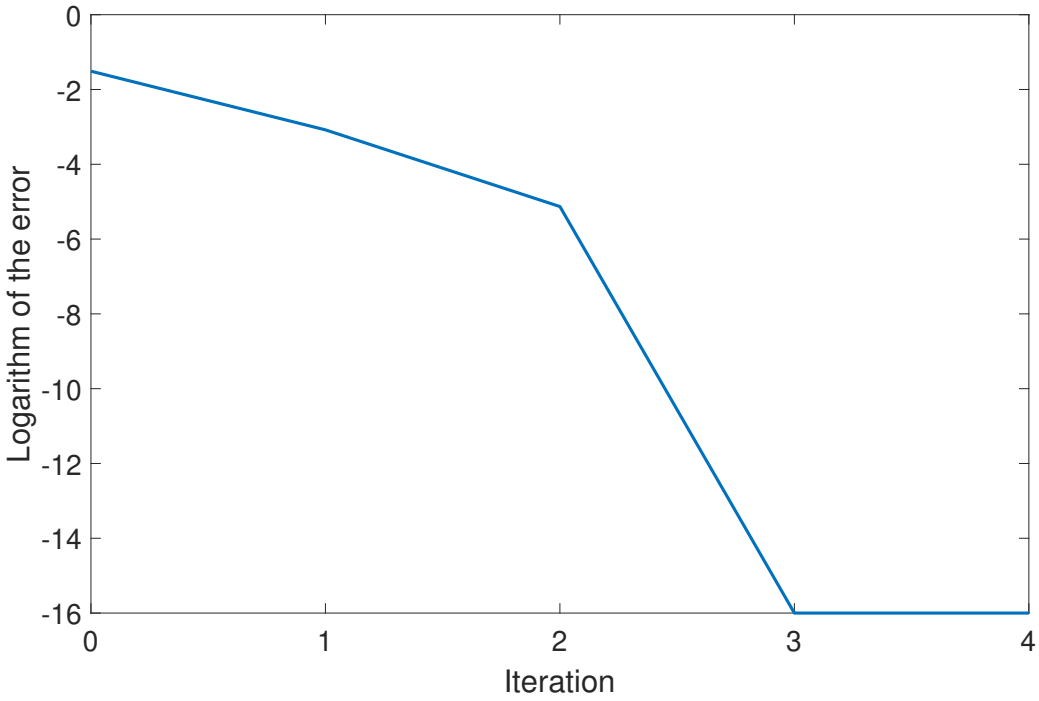


Figure 5.4: Logarithm of the error using the approximate solution (Table 5.4).

In other words, the fine step of the backward Euler method over the interval $[0, 1]$ (in sequential meaning) remains constant for all cases. The smaller problems are solved within the subintervals $[T^k; T^{k+1}]$, utilising as many cores as possible. The logarithm of the error (5.24) is presented in Table 5.5. Furthermore, the decimal logarithm of the error values is plotted in Figure 5.5.

Table 5.5: Logarithm of the error using using 4 up to 64 cores.

Δt	$i = 0$	$i = 1$	$i = 2$	$i = 3$	$i = 4$	$i = 5$	$i = 6$
$1/4$	$3.09e{-2}$	$8.33e{-4}$	$7.42e{-6}$	0	0	0	0
$1/8$	$1.91e{-2}$	$4.15e{-4}$	$5.00e{-6}$	$3.61e{-8}$	$1.56e{-10}$	$3.71e{-13}$	$4.44e{-16}$
$1/16$	$1.05e{-2}$	$1.38e{-4}$	$1.13e{-6}$	$6.34e{-9}$	$2.61e{-11}$	$9.17e{-14}$	$3.89e{-16}$
$1/32$	$5.50e{-3}$	$3.95e{-5}$	$1.83e{-7}$	$6.13e{-10}$	$1.58e{-12}$	$7.38e{-15}$	$1.67e{-16}$
$1/64$	$2.81e{-3}$	$1.05e{-5}$	$2.59e{-8}$	$4.70e{-11}$	$8.38e{-14}$	$1.22e{-15}$	$1.67e{-16}$

As can be observed in Table 5.5, the convergence rate is seen to improve slightly when utilising a greater number of cores. It is important to note two key points. Firstly, iterations $i = 5$ and $i = 6$ are not necessary for the first row of Table 5.5 and are included only for comparative purposes. Secondly, it can be observed that the algorithm converges after three iterations for the case where $\Delta t := 1/4$. However, it should be noted that for $\Delta t := 1/4$, larger problems are being solved within the independent subdomains compared to the case where $\Delta t := 1/64$.

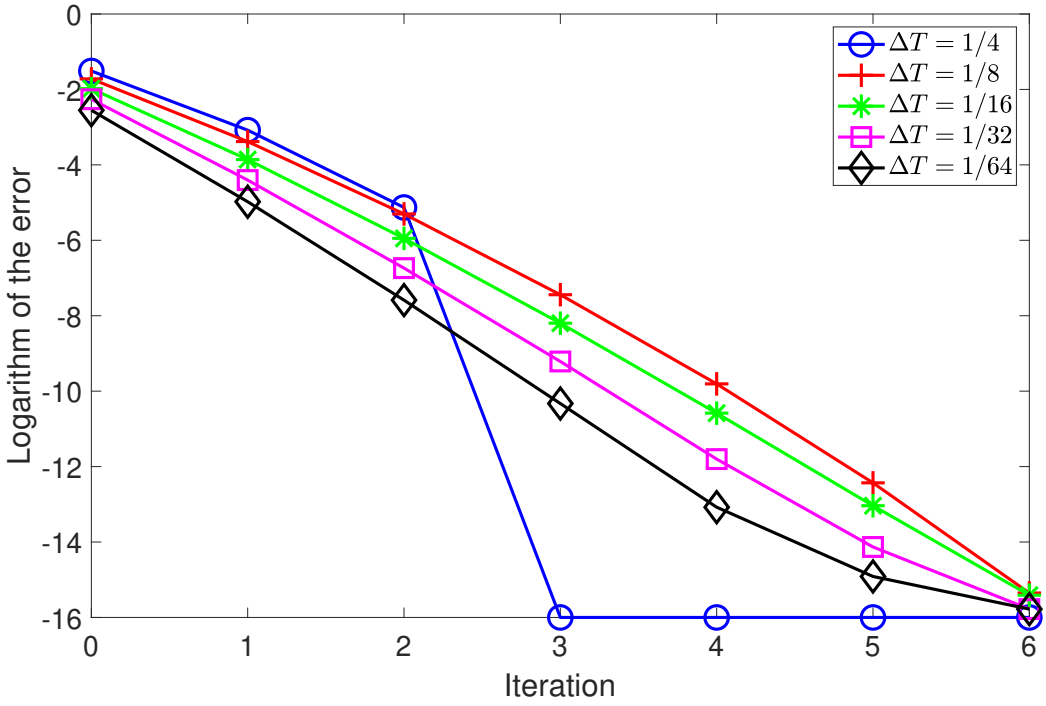


Figure 5.5: Logarithm of the error using using 4 up to 64 cores (Table 5.5).

5.2.2 Solving PDE by Parareal

Consider the problem outlined in the first example of Section 4.3, precisely, the heat equation with a non-zero initial condition and a zero source term. To facilitate comparison with the results presented in Table 4.1 and Table 4.2, we again assume a coarse time step of $\Delta t := 1/4$. The fine steps are identical to those used in Section 4.3, specifically, $\delta t := 1/16,\ 1/32,\ 1/64,\ 1/128$. The spatial step is set to $h_x := \delta t$. This setup corresponds to columns 3-6 of given tables. Thus, as the first example, the Euler method is employed as outlined in Section 5.1.2. In this case, the number of independent subintervals is $n = 8$ as $T = 2$ and $\Delta t := 1/4$. We assume that the spatial dimension is $d = 1$.

To observe the efficiency of the Parareal algorithm, we use the error defined as

$$\mathrm{e}^{i}_{seq} := \|u^{\Delta t,\delta t}_{h,i}(\mathbf{x},T) - u_h(\mathbf{x},T)\|_{L^2(\Omega)}, \tag{5.25}$$

where $u^{\Delta t,\delta t}_{h,i}(\mathbf{x},T)$ represents the solution of the Parareal method of the ith Parareal iteration at time T for the given steps Δt and δt, and $u_h(\mathbf{x},T)$ is the sequential solution at the time T as outlined in Section 4.3. The resulting errors of the backward Euler method are presented in Table 5.6. With a precision of $1e{-}10$, it can be seen that the Parareal method converges to the sequential solution at the fourth iteration out of a total of 8.

To ensure that the Parareal algorithm converges to a reasonable solution, we also computed the error with respect to the exact solution

$$\mathrm{e}^{i}_{exact} := \|u^{\Delta t,\delta t}_{h,i}(\mathbf{x},T) - u(\mathbf{x},T)\|_{L^2(\Omega)}, \tag{5.26}$$

where $u(\mathbf{x}, T)$ is the exact solution as defined in equation (4.16). The error values can be found in Table 5.6. For example, looking at the $i = 6$ column, it can be observed that the error value is the same as that obtained in the problem mentioned in the first example of Section 4.3. Precisely, by comparing the $i = 6$ column with the first line of Table 4.1. This suggests that the Parareal algorithm does not alter the convergence behaviour of the original sequential method when applied to PDEs.

Table 5.6: The error of the Euler method in $L_2(\Omega)$-norm for 1d + time problem.

	δt	$i = 1$	$i = 2$	$i = 3$	$i = 4$	$i = 5$	$i = 6$
e^i_{seq}	1/16	$7.92e{-5}$	$4.48e{-7}$	$1.52e{-9}$	$3.10e{-12}$	$3.55e{-15}$	$8.55e{-17}$
	1/32	$1.09e{-4}$	$7.21e{-7}$	$2.86e{-9}$	$6.82e{-12}$	$9.10e{-15}$	$1.41e{-16}$
	1/64	$1.26e{-4}$	$8.94e{-7}$	$3.82e{-9}$	$9.77e{-12}$	$1.40e{-14}$	$8.60e{-17}$
	1/128	$1.34e{-4}$	$9.91e{-7}$	$4.38e{-9}$	$1.16e{-11}$	$1.72e{-14}$	$2.71e{-16}$
e^i_{exact}	1/16	$1.26e{-3}$	$1.34e{-3}$	$1.34e{-3}$	$1.34e{-3}$	$1.34e{-3}$	$1.34e{-3}$
	1/32	$9.92e{-4}$	$1.10e{-3}$	$1.10e{-3}$	$1.10e{-3}$	$1.10e{-3}$	$1.10e{-3}$
	1/64	$5.40e{-4}$	$6.66e{-4}$	$6.65e{-4}$	$6.65e{-4}$	$6.65e{-4}$	$6.65e{-4}$
	1/128	$2.27e{-4}$	$3.63e{-4}$	$3.62e{-4}$	$3.62e{-4}$	$3.62e{-4}$	$3.62e{-4}$

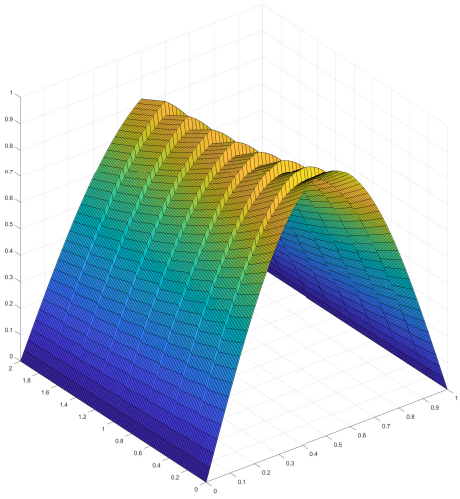
As previously noted in Section 5.2.2, the Crank-Nicolson method can be used as an alternative to the Euler method within the Parareal algorithm. In Table 5.7, we present the error with respect to both the exact and sequential solutions. The table shows that the minimum reachable error of the sequential method is preserved as in the previous case. By examining Table 5.7 and assuming a precision of $1e{-10}$, it can be seen that the Parareal method with the Crank-Nicolson scheme converges to the sequential solution in the sixth iteration. This represents an increase of 2 iterations compared to the Euler method. However, Table 5.7 also shows that in the fourth iteration, the Crank-Nicolson method has slightly better precision than the Euler scheme in relation to the exact solution.

Table 5.7: The error of the Crank-Nicolson method in $L_2(\Omega)$-norm for 1d + time problem.

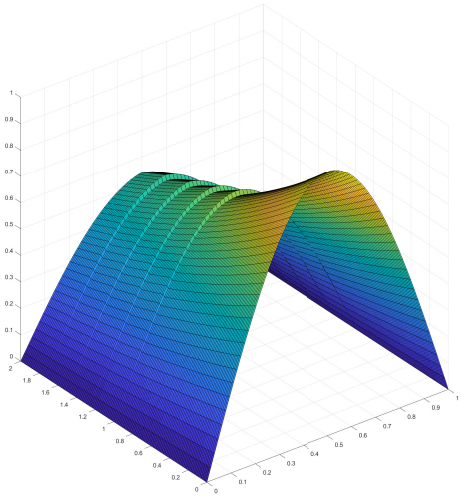
	δt	$i = 1$	$i = 2$	$i = 3$	$i = 4$	$i = 5$	$i = 6$
e^i_{seq}	1/16	$4.47e{-5}$	$5.88e{-6}$	$4.46e{-7}$	$1.99e{-8}$	$4.88e{-10}$	$5.09e{-12}$
	1/32	$5.60e{-5}$	$8.66e{-6}$	$7.66e{-7}$	$3.98e{-8}$	$1.13e{-9}$	$1.37e{-11}$
	1/64	$6.16e{-5}$	$1.02e{-5}$	$9.71e{-7}$	$5.40e{-8}$	$1.65e{-9}$	$2.13e{-11}$
	1/128	$6.43e{-5}$	$1.11e{-5}$	$1.09e{-6}$	$6.24e{-8}$	$1.96e{-9}$	$2.63e{-11}$
e^i_{exact}	1/16	$1.87e{-3}$	$1.92e{-3}$	$1.91e{-3}$	$1.91e{-3}$	$1.91e{-3}$	$1.91e{-3}$
	1/32	$4.24e{-4}$	$4.87e{-4}$	$4.78e{-4}$	$4.79e{-4}$	$4.79e{-4}$	$4.79e{-4}$
	1/64	$6.17e{-5}$	$1.30e{-4}$	$1.19e{-4}$	$1.20e{-4}$	$1.20e{-4}$	$1.20e{-4}$
	1/128	$3.59e{-5}$	$4.07e{-5}$	$2.89e{-5}$	$3.00e{-5}$	$3.00e{-5}$	$3.00e{-5}$

Similar results are obtained while considering a spatial dimension $d = 2, 3$. Additionally, if we were to consider a non-zero source term, as in the second example in Section 4.3, the results would remain consistent.

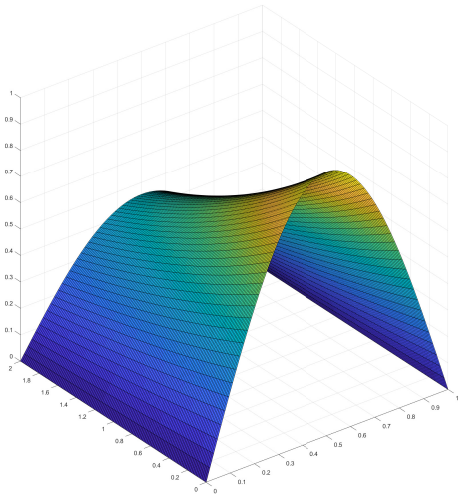
In Figure 5.6, the Parareal method is illustrated for the case in which the Euler method is used. The figure illustrates how the values of each subdomain are corrected until the end of the loop. The jumps between subintervals are highlighted by using a suitable constant for better visibility.



(a) Step 0 (before the loop).



(b) Step 4.



(c) Step 8.

Figure 5.6: The Euler method for 1d + time problem.

5.2.3 Parareal as distributed program

In the study presented in (Aubanel 2011), the Parareal algorithm was also implemented as a distributed program. Here, the program was implemented in Python 3 using the PETSc interface

petsc4py. The PETSc library offers an efficient method for the assembly of FEM matrices and the usage of subsequent solvers. In this case, the conjugate gradient solver with an incomplete LU preconditioner was employed at each step of the Euler method. The parallelisation was achieved through the use of the MPI package mpi4py. The program was tested on the Karolina cluster at IT4Innovations, VSB – Technical University of Ostrava, Ostrava, Czech Republic.

We considered the system of ODEs as defined by (4.7) in 2d, with $\Omega := (0,1)^2$, $T := 1$, $u_0(\mathbf{x}) := 0$, and the source term as defined by (4.18). The coarse time step was set to $\Delta t = 1/n$, where n represents the number of cores (MPI processes), and the spatial step was fixed at $h_x := 1/256$.

1. In the first example, the fine time step was set to $\delta t = 1/256$. The sequential solution was computed in 48.45 seconds, and the error in the $L^2(\Omega)$-norm (4.17) was $1.8276e{-}6$. As the stopping criteria for the Parareal algorithm, we defined $|\mathcal{G}(U_k^{i+1}) - \mathcal{G}(U_k^i)| < \varepsilon$, where $\varepsilon := 1e{-}8$. Table 5.8 presents the resulting times in seconds for the iterations $i = 1$ and $i = 2$ of the Parareal method (the first and the second row), as well as the time taken until the last processor receives the initial coarse solution (the third row), and the overall computational time (the fourth row). Table 5.8 also presents the iteration when the algorithm converges (the fifth row) and the theoretical speedup ψ (the last row). The times for $i = 1$ and $i = 2$ do not include the time taken to send the initial coarse solution from the first processor to the last one. As can be seen, the time required to send the initial coarse solution increases by a factor of 2 with the increasing number of used CPUs. The theoretical speedup ψ was equal for 8, 16 and 32 CPUs. However, the fastest solution was obtained when using 8 cores. In other words, the ratio r in equation (5.13) is minimal for 8 cores. As the number of cores increases, the ratio r in equation (5.13) tends towards 1. This causes the parallel algorithm to behave in a similar manner to a sequential one, resulting in a degradation of efficiency. Additionally, the delay caused by communication between processes must also be taken into account.

Table 5.8: Times in seconds for $\delta t = 1/256$ with speedup ψ.

	n				
	2	4	8	16	32
$i = 1$	27.76	12.32	6.35	3.96	2.77
$i = 2$	51.64	24.61	12.69	7.93	5.70
coarse (s)	0.65	1.37	2.28	4.52	8.84
overall (s)	52.29	38.28	33.92	38.37	40.77
iteration	2	3	5	9	14
ψ	0.98	1.27	1.42	1.42	1.42

The resulting overall times are plotted in Figure 5.7 for improved readability.

2. In the second example, the fine time step was set to $\delta t = 1/1024$. The sequential solution was computed in 148.7917 seconds, and the error in the $L^2(\Omega)$-norm (4.17) was $8.7065e{-}7$.

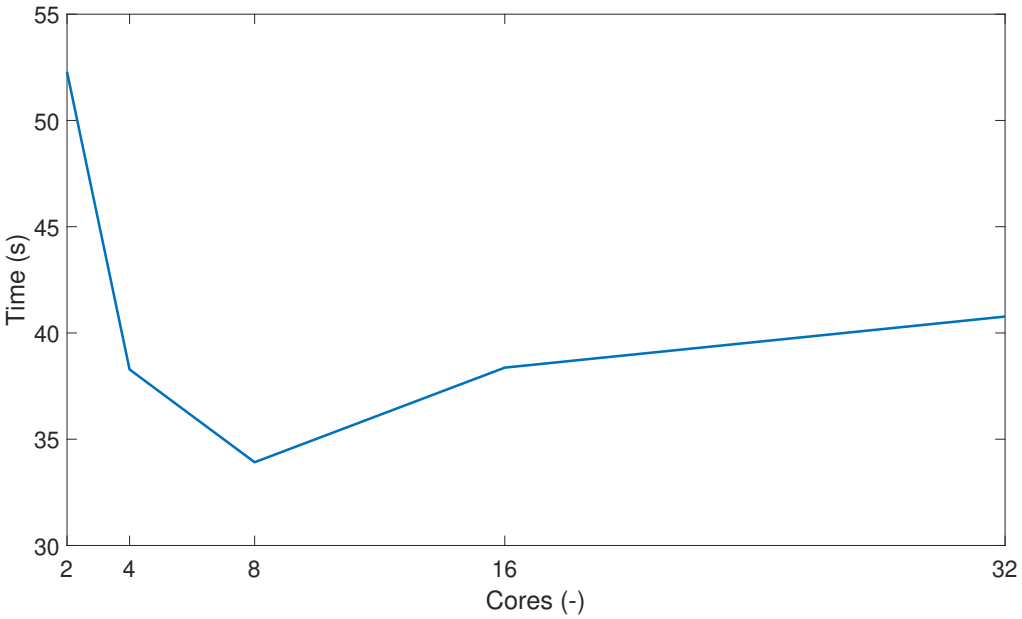


Figure 5.7: Solve time of distributed Parareal for $\delta t := 1/256$.

The precision remained the same, i.e., $\varepsilon := 1e{-}8$. As in the previous example, Table 5.9 presents the resulting times in seconds for the iterations $i = 1$ and $i = 2$ of the Parareal method (the first and the second row), as well as the time taken until the last processor receives the initial coarse solution (the third row), and the overall computational time (the fourth row). Table 5.9 also presents iteration when the algorithm converges (the fifth row) and the theoretical speedup ψ (the last row). Again, the times for $i = 1$ and $i = 2$ do not include the time taken to send the initial coarse solution from the first processor to the last one. Similar to the prior example, this time increases by a factor of 2. In contrast to the previous case, the most efficient solution was achieved when utilising 16 cores.

Table 5.9: Times in seconds for $\delta t = 1/1024$ with speedup ψ.

	n				
	2	4	8	16	32
$i = 1$	90.86	37.11	18.84	11.25	8.53
$i = 2$	164.33	74.18	37.61	21.99	17.31
coarse (s)	0.58	1.18	1.69	4.77	9.34
times (s)	164.92	112.39	95.63	91.08	102.63
iteration	2	3	5	8	14
ψ	0.99	1.31	1.54	1.84	1.97

Again, the resulting overall times are plotted in Figure 5.8 for improved readability.

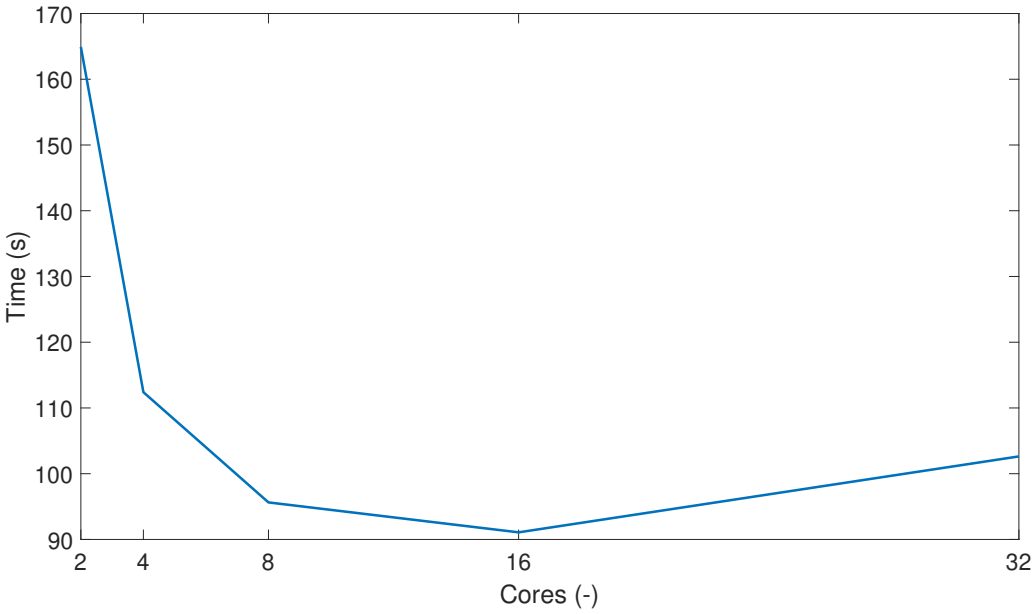


Figure 5.8: Solve time of distributed Parareal for $\delta t := 1/1024$.

5.2.4 Combining Parareal and spatial DDM to solve PDE

In this section, the results previously proposed in (Foltyn et al. 2020) are presented. The 2-dimensional problem is considered, as defined in the first part of Section 4.3. To examine the combination, a relative error is studied, given by:

$$\mathrm{e}_{rel}^{i} = \frac{\|u_{h,i}^{\Delta t,\delta t}(\mathbf{x},T) - u_h(\mathbf{x},T)\|_{L^2(\Omega)}}{\|u_h(\mathbf{x},T)\|_{L^2(\Omega)}}, \tag{5.27}$$

where $u_{h,i}^{\Delta t,\delta t}(\mathbf{x},T)$ is the solution of the Parareal method of ith iteration at time T for given steps Δt and δt, and $u_h(\mathbf{x},T)$ is the sequential solution at time T. The spatial and temporal discretisation steps are fixed, i.e., $h = 1/32$ and $\delta t = 1/512$. The backward Euler method (4.10) is used as a time-stepping scheme. Before examining the robustness of the DDM coupled with the Parareal, the relative error (5.27) for different coarse time steps Δt without using the DDM is provided.

The resulting errors are shown in Table 5.10. It can be observed that, in order to achieve a given precision (e.g., $1e{-}8$), the number of iterations decreases (i.e., $i = 6,\ 5,\ 5$) with an increasing parallelism in time (i.e., $\Delta t = 1/4,\ 1/8,\ 1/16$) as was demonstrated in the previous experiments. This implies that the overall complexity of the predictor steps demonstrates optimal parallel scalability. However, in practice, the parallel speedup is partially degraded by the sequential corrector steps, as can be seen in Section 5.2.3.

Concerning the combination of the Parareal algorithm and the Domain Decomposition Method, it is noted that the spatial subproblems within each iteration of equation (4.10) are solved using the preconditioned conjugate gradients (PCG) method to a relative precision of $1e{-}8$. The DDM preconditioner is being utilised as described in Section 5.1.4. The results presented in Table 5.11 demonstrate the error after three iterations of the Parareal method.

Table 5.10: Relative error after three parareal iterations in 2d.

Δt	$i=1$	$i=2$	$i=3$	$i=4$	$i=5$	$i=6$
1/2	2.04e−1	1.28e−2	2.62e−4	0	0	0
1/4	1.28e−1	6.72e−3	1.93e−4	3.33e−6	3.86e−8	1.87e−9
1/8	7.08e−2	2.28e−3	4.52e−5	6.20e−7	8.10e−9	1.29e−9
1/16	3.70e−2	6.53e−4	7.41e−6	6.09e−8	5.18e−10	5.78e−11

It is observed that the convergence is not dependent on the level of spatial parallelism and, in fact, improves as the level of parallelism in time increases. Additionally, it is worth noting that the convergence in column $i = 3$ of Table 5.10 remains unaffected. It is also worth noting that the maximum number of PCG iterations required were 6, 15, 27, and 21, respectively, for the DDM coarse steps of $H = 1/2$, $1/4$, $1/8$, and $1/16$.

Table 5.11: Relative error after three parareal-DDM iterations in 2d.

Δt	$H=1/2$	$H=1/4$	$H=1/8$	$H=1/16$
1/2	2.62e−4	2.62e−4	2.62e−4	2.62e−4
1/4	1.92e−4	1.92e−4	1.92e−4	1.92e−4
1/8	4.40e−5	4.40e−5	4.40e−5	4.40e−5
1/16	6.95e−6	6.95e−6	6.95e−6	6.95e−6

A drawback of the combination above is that it wastes memory, as each process is required to solve the same DDM for each time slice. A possible solution to this issue would be to employ a combination of the Parareal with a single DDM that is capable of handling multiple right-hand sides.

Chapter 6

Space-time finite element method

This chapter is based on the work presented in (Steinbach 2015), where stability and a priori error analysis for the space-time finite element method (FEM) are proposed. The author of the given work has derived quasi-optimal error estimates which do not depend on the spatial dimension but only on the discretisation step. Furthermore, there is a more detailed analysis of the piecewise linear interpolation error when using right-angled isosceles triangular elements for the case of a 1d + time. The author also recommends the use of multigrid approaches such as geometric multilevel techniques or domain decomposition methods to solve space-time problems but does not focus on their practical implementation in a parallel environment. Due to the nature of time, there remains a significant amount of work to be done in terms of parallel implementation of this approach. As stated in the introduction chapter, dealing with time is challenging as every subsequent subproblem depends on the subproblem from the previous time slice. Thus, building an efficient preconditioner for the space-time FEM is a big challenge nowadays. In this chapter, we summarise the space-time FEM theory in the Bochner-Sobolev space (Steinbach 2015) and in the anisotropic Sobolev space (Langer et al. 2021). For further details and proofs, please refer to the mentioned references.

6.1 Bochner-Sobolev space, existence and uniqueness

Once more, let us consider the heat equation with constant material coefficients (3.1). In this case, the initial condition

$$u_0(\mathbf{x}) \in H_0^1(\Omega), \tag{6.1}$$

and the source term

$$f(\mathbf{x},t) \in L^2\Big((0,T);H^{-1}(\Omega)\Big). \tag{6.2}$$

Furthermore, assume the following spaces:

$$X := L^2\Big((0,T);H_0^1(\Omega)\Big) \cap H^1\Big((0,T);H^{-1}(\Omega)\Big), \tag{6.3}$$

$$\widetilde{X} := \Big\{v \in L^2\Big((0,T);H_0^1(\Omega)\Big) \cap H^1\Big((0,T);H^{-1}(\Omega)\Big) : v(\mathbf{x},0) = 0 \text{ for } \mathbf{x} \in \Omega\Big\}, \tag{6.4}$$

$$Y := L^2\Big((0,T);H_0^1(\Omega)\Big), \tag{6.5}$$

$$Y^* := L^2\Big((0,T);H^{-1}(\Omega)\Big). \tag{6.6}$$

In contrast to the weak formulation outlined in Chapter 3, a test function $v(\mathbf{x},t) \in Y$ is assumed. The test function is defined in space and time simultaneously. Thus, the variational formulation corresponding to (3.1) along with (6.1) and (6.2) is to find $u \in X$, $u(\mathbf{x},0) = u_0(\mathbf{x})$ for $\mathbf{x} \in \Omega$, such that

$$c_H \int_0^T \int_\Omega \partial_t u(\mathbf{x},t) v(\mathbf{x},t) \,\mathrm{d}\mathbf{x}\,\mathrm{d}t + \int_0^T \int_\Omega \nabla_{\mathbf{x}} u(\mathbf{x},t) \nabla_{\mathbf{x}} v(\mathbf{x},t) \,\mathrm{d}\mathbf{x}\,\mathrm{d}t = \int_0^T \int_\Omega f(\mathbf{x},t) v(\mathbf{x},t) \,\mathrm{d}\mathbf{x}\,\mathrm{d}t \quad (6.7)$$

for all $v(\mathbf{x},t) \in Y$, $v(\mathbf{x},0) = 0$ for $\mathbf{x} \in \Omega$, where $\partial_t u := \frac{\partial u}{\partial t}$ for the sake of simplicity. Additionally, we treat the initial condition u_0 as a Dirichlet condition. In doing so, the solution u can be represented in the form $u(\mathbf{x},t) = \widetilde{u}(\mathbf{x},t) + \widetilde{u}_0(\mathbf{x},t)$ for $(\mathbf{x},t) \in Q_T$, where $\widetilde{u}_0(\mathbf{x},t) \in X$ is some extension of $u_0 \in H_0^1(\Omega)$. Hence, we arrive at the variational formulation to find $\widetilde{u} \in \widetilde{X}$ such that

$$a(\widetilde{u}, v) = \langle f, v \rangle_{Q_T} - a(\widetilde{u}_0, v) \quad \forall v \in Y, \quad (6.8)$$

where

$$a(u,v) := c_H \int_0^T \int_\Omega \partial_t u(\mathbf{x},t) v(\mathbf{x},t) \,\mathrm{d}\mathbf{x}\,\mathrm{d}t + \int_0^T \int_\Omega \nabla_{\mathbf{x}} u(\mathbf{x},t) \nabla_{\mathbf{x}} v(\mathbf{x},t) \,\mathrm{d}\mathbf{x}\,\mathrm{d}t, \quad (6.9)$$

$$\langle f, v \rangle_{Q_T} := \int_0^T \int_\Omega f(\mathbf{x},t) v(\mathbf{x},t) \,\mathrm{d}\mathbf{x}\,\mathrm{d}t. \quad (6.10)$$

Within the given variational formulation, two different spaces are utilised. The first space is defined in relation to the solution u, and the second is defined as the space of test functions. This formulation is known as a Petrov-Galerkin variational formulation. Consequently, the idea of ellipticity cannot be applied to establish unique solvability in this scenario. Additional tools have to be provided to ensure the unique solvability of the variational formulation. The first of them is the quasi-static elliptic Dirichlet boundary value problem

$$\begin{aligned} -\Delta_{\mathbf{x}} w(\mathbf{x},t) &= c_H \Phi(\mathbf{x},t) \quad && \forall (\mathbf{x},t) \in Q_T, \\ w(\mathbf{x},t) &= 0 && \forall (\mathbf{x},t) \in \Gamma \times (0,T), \end{aligned} \quad (6.11)$$

The variational formulation to (6.11) is to find $w(\mathbf{x},t) \in Y$ such that

$$\int_0^T \int_\Omega \nabla_{\mathbf{x}} w(\mathbf{x},t) \nabla_{\mathbf{x}} v(\mathbf{x},t) \,\mathrm{d}\mathbf{x}\,\mathrm{d}t = c_H \int_0^T \int_\Omega \Phi(\mathbf{x},t) v(\mathbf{x},t) \,\mathrm{d}\mathbf{x}\,\mathrm{d}t \quad (6.12)$$

for all $v(\mathbf{x},t) \in Y$, and any given $\Phi(\mathbf{x},t) \in Y^*$. The solution of (6.12) implies that for any given $\Phi(\mathbf{x},t) \in Y^*$, there exists a unique $w(\mathbf{x},t) \in Y$ that defines the Newton potential

$$N\Phi := w, \quad N \colon Y^* \to Y. \quad (6.13)$$

With these at hand, the norm equivalence

$$\|w\|_Y = \|N\Phi\|_Y = \|\Phi\|_{Y^*}.$$

holds true and results in

$$\|\Phi\|_{Y^*}^2 = c_H \langle \Phi, N\Phi \rangle_{Q_T} \quad \forall \Phi \in Y^*.$$

Furthermore, a stability condition is established by utilising the inclusion $X \subset Y$.

Theorem 6.1 (Stability condition) *For all $u \in \widetilde{X}$ there holds the stability condition*

$$\frac{1}{2\sqrt{2}} \|u\|_X \leq \sup_{v \neq 0,\ v \in Y} \frac{a(u,v)}{\|v\|_Y}. \tag{6.14}$$

Finally, a corollary of the existence and uniqueness can be stated, as it is done in (Steinbach 2015), since $\widetilde{X} \subset Y$.

Corollary 6.2 (Existence and uniqueness) Let $\widetilde{u}_0 \in X$ be some extension of the given initial datum $u_0 \in H_0^1(\Omega)$, and assume $f \in Y^*$. Since the bilinear form $a(\cdot,\cdot)$ as given in (6.9) is bounded satisfying

$$a(u,v) \leq \sqrt{2} \|u\|_X \|v\|_Y \quad \forall u \in X,\, \forall v \in Y,$$

and satisfies the stability condition (6.14), there exists a unique solution $\widetilde{u} \in \widetilde{X}$ of the variational formulation (6.8) satisfying

$$\|\widetilde{u}\|_X \leq 2\sqrt{2} \left[\|f\|_{Y^*} + \sqrt{2} \|\widetilde{u}_0\|_X \right].$$

6.1.1 Petrov-Galerkin discretisation

Let us consider the finite-dimensional spaces $X_h \subset \widetilde{X}$ and $Y_h \subset Y$. The inclusion $X_h \subset Y_h$ is assumed, as in the continuous case. The discrete variational formulation is to find $\widetilde{u}_h \in X_h$ such that

$$a(\widetilde{u}_h, v_h) = \langle f, v_h \rangle_{Q_T} - a(\widetilde{u}_0, v_h) \quad \forall v_h \in Y_h. \tag{6.15}$$

Again, as in the continuous case, we suppose the discrete quasi-static variational formulation to find $w_h(\mathbf{x},t) \in Y_h$ such that

$$\int_0^T \int_\Omega \nabla_{\mathbf{x}} w_h(\mathbf{x},t) \nabla_{\mathbf{x}} v_h(\mathbf{x},t) \,\mathrm{d}\mathbf{x}\,\mathrm{d}t = c_H \int_0^T \int_\Omega \Phi(\mathbf{x},t) v_h(\mathbf{x},t) \,\mathrm{d}\mathbf{x}\,\mathrm{d}t \tag{6.16}$$

for all $v_h(\mathbf{x},t) \in Y_h$, and we define the approximate Newton potential

$$N_h \Phi := w_h, \quad N_h \colon Y^* \to Y_h \subset Y. \tag{6.17}$$

Then a discrete stability condition can be established.

Theorem 6.3 (Discrete stability condition) *Assume $X_h \subset \widetilde{X}$, $Y_h \subset Y$, and $X_h \subset Y_h$. Then there holds the discrete stability condition*

$$\frac{1}{\sqrt{2}} \|u_h\|_{X_h} \leq \sup_{v_h \neq 0,\ v_h \in Y_h} \frac{a(u_h, v_h)}{\|v_h\|_Y} \quad \forall u_h \in X_h. \tag{6.18}$$

The well-known Galerkin orthogonality can be defined by utilising the inclusion $Y_h \subset Y$ and combining (6.8) with (6.15). That is,

$$a(\widetilde{u} - \widetilde{u}_h, v_h) = 0 \quad \forall v_h \in Y_h. \tag{6.19}$$

In this point, the quasi-optimal error estimate can be proved.

Theorem 6.4 *Let $\widetilde{u} \in \widetilde{X}$ and $\widetilde{u}_h \in X_h$ be the unique solutions of the variational formulations* (6.8) *and* (6.15)*, respectively. Then there holds the a priori error estimate*

$$\|\widetilde{u} - \widetilde{u}_h\|_{X_h} \leq 5 \inf_{z_h \in X_h} \|\widetilde{u} - z_h\|_{\widetilde{X}}.$$

6.1.2 Finite element spaces and error estimates

We shall consider the discretisation of the space-time domain $Q_T := (0,T) \times \Omega \subset \mathbb{R}^{d+1}$, where $d = 1, 2, 3$, into shape-regular (4.1) and quasi-uniform (4.2) finite elements. The spatial domain Ω is assumed to be an interval in 1d, or polygonal in 2d, or polyhedral in 3d, and can be written as

$$\mathcal{T}_h := \{\omega_i\}_{i=1}^M; \quad Q_T := \bigcup_i \{\omega_i \colon \omega_i \in \mathcal{T}_h\},$$

where M is a number of elements ω_i with mesh sizes h_i and the maximal mesh size $h := \max_i h_i$. Here, we denote both the temporal steps $h_{t,i}$ and the spatial steps $h_{x,i}$ in a unified way as h_i for simplicity. Let us suppose the space-time finite element space

$$S_h^1(Q_T) = \operatorname{span}\{\varphi_1, \ldots, \varphi_N\}, \tag{6.20}$$

where φ_k are piece-wise linear and continuous nodal FEM basis functions over ω_i. The elements ω_i represent triangles in the case of the 1d + time, tetrahedra in the 2d + time, and pentatops (Neumüller; Steinbach 2011) elements in the 3d + time domain. Alternatively, for example, multilinear continuous basis functions over quadrilaterals using 1d + time, and hexahedra using 2d + time, can also be employed. The resulting finite element spaces are defined as follows

$$X_h := S_h^1(Q_T) \cap \widetilde{X}, \quad Y_h := S_h^1(Q_T) \cap Y. \tag{6.21}$$

Finally, by utilising the finite-dimensional spaces (6.21) which satisfy Theorem 6.3, the following a priori error estimate can be established.

Theorem 6.5 (Energy error estimate) *Let $\widetilde{u} \in \widetilde{X}$ and $\widetilde{u}_h \in \widetilde{X}_h = S_h^1(Q_{T,h}) \cap \widetilde{X}$ be the unique solutions of the variational formulations* (6.8) *and* (6.15)*, respectively. Assume $\widetilde{u} \in H^2(Q_T)$. Then there holds the energy error estimate*

$$\|\widetilde{u} - \widetilde{u}_h\|_Y \leq c\, h\, |\widetilde{u}|_{H^2(Q_T)}. \tag{6.22}$$

Corollary 6.6 Let $\widetilde{u} \in \widetilde{X}$ and $\widetilde{u}_h \in \widetilde{X}_h = S_h^1(Q_{T,h}) \cap \widetilde{X}$ be the unique solutions of the variational formulations (6.8) and (6.15), respectively. Assume $\widetilde{u} \in H^s(Q_T)$ for some $s \in [1,2]$. Then there holds the energy error estimate

$$\|\widetilde{u} - \widetilde{u}_h\|_Y \leq c\, h^{s-1}\, |\widetilde{u}|_{H^s(Q_T)}. \tag{6.23}$$

Unfortunately, the estimate (6.22) is not optimal in general, for example, in cases of solutions with some singular behaviour. The reason can be found in the proof of the given estimate, where the spatial and temporal derivatives were treated in a unified way. For further information, please refer to (Steinbach 2015).

6.2 Anisotropic Sobolev Spaces, existence and uniqueness

An alternative to the Bochner-Sobolev formulation is the use of so-called anisotropic Sobolev spaces, as defined in (Langer et al. 2021). Here, a brief summary of the existing results from (Langer et al. 2021; Zank 2019) is provided with no further investigation. In Section 6.3, which follows along with subsequent numerical experiments, the Bochner-Sobolev space is assumed.

The anisotropic Sobolev spaces are formulated as follows

$$\begin{aligned}
H^{1,1/2}_{0;0,}(Q_T) &:= H^{1/2}_{0,}\left((0,T);L^2(\Omega)\right) \cap L^2\left((0,T);H^1_0(\Omega)\right),\\
H^{1,1/2}_{0;,0}(Q_T) &:= H^{1/2}_{,0}\left((0,T);L^2(\Omega)\right) \cap L^2\left((0,T);H^1_0(\Omega)\right),
\end{aligned}$$

along with Hilbertian norms

$$\begin{aligned}
\|v\|_{H^{1,1/2}_{0;0,}(Q_T)} &:= \sqrt{\|v\|^2_{H^{1/2}_{0,}((0,T);L^2(\Omega))} + \|v\|^2_{L^2((0,T);H^1_0(\Omega))}},\\
\|v\|_{H^{1,1/2}_{0;,0}(Q_T)} &:= \sqrt{\|v\|^2_{H^{1/2}_{,0}((0,T);L^2(\Omega))} + \|v\|^2_{L^2((0,T);H^1_0(\Omega))}}.
\end{aligned}$$

The definition of the Bochner spaces on the right-hand side is

$$\begin{aligned}
H^{1/2}\left((0,T);L^2(\Omega)\right) &:= \left\{v \in L^2(Q_T)\colon \|v\|_{H^{1/2}((0,T);L^2(\Omega))} < \infty\right\},\\
H^{1/2}_{0,}\left((0,T);L^2(\Omega)\right) &:= \left\{v \in H^{1/2}\left((0,T);L^2(\Omega)\right) : \|v\|_{H^{1/2}_{0,}((0,T);L^2(\Omega))} < \infty\right\},\\
H^{1/2}_{,0}\left((0,T);L^2(\Omega)\right) &:= \left\{v \in H^{1/2}\left((0,T);L^2(\Omega)\right) : \|v\|_{H^{1/2}_{,0}((0,T);L^2(\Omega))} < \infty\right\}.
\end{aligned}$$

The corresponding norms are defined by

$$\begin{aligned}
\|v\|_{H^{1/2}((0,T);L^2(\Omega))} &:= \sqrt{\|v\|^2_{L^2(Q_T)} + \int_0^T \int_0^T \frac{\|v(\cdot,t) - v(\cdot,\xi)\|^2_{L^2(\Omega)}}{|t-\xi|^2}\,\mathrm{d}t\,\mathrm{d}\xi},\\
\|v\|_{H^{1/2}_{0,}((0,T);L^2(\Omega))} &:= \sqrt{\|v\|^2_{H^{1/2}((0,T);L^2(\Omega))} + \int_0^T \frac{\|v(\cdot,t)\|^2_{L^2(\Omega)}}{t}\,\mathrm{d}t},
\end{aligned}$$

$$\|v\|_{H^{1/2}_{,0}((0,T);L^2(\Omega))} := \sqrt{\|v\|^2_{H^{1/2}((0,T);L^2(\Omega))} + \int_0^T \frac{\|v(\cdot,t)\|^2_{L^2(\Omega)}}{T-t}\,\mathrm{d}t}.$$

The meaning of the subscript "0," (",0") is that the function is 0 for $t = 0$ ($t = T$) in the sense of the norm $\|\cdot\|_{H^{1/2}_{0,}((0,T);L^2(\Omega))}$ ($\|\cdot\|_{H^{1/2}_{0,}((0,T);L^2(\Omega))}$). The dual space $\left(H^{1,1/2}_{0;,0}(Q_T)\right)^*$ is defined as completition of $L^2(Q_T)$ according to the Hilbertian norm

$$\|f\|_{(H^{1,1/2}_{0;,0}(Q_T))^*} := \sup_{v\neq 0, v\in H^{1,1/2}_{0;,0}(Q_T)} \frac{|\langle f, v\rangle_{Q_T}|}{\|v\|_{H^{1,1/2}_{0;,0}(Q_T)}}.$$

At this point, all necessary preparations have been made to establish the variational formulation. Let us consider the parabolic problem (3.1). For simplicity, the initial condition $u_0(\mathbf{x}) = 0$ is assumed. The variational formulation is to find $u \in H^{1,1/2}_{0;0,}(Q_T)$ such that

$$a(u,v) = \langle f, v\rangle_{Q_T} \quad \forall v \in H^{1,1/2}_{0;,0}(Q_T), \tag{6.24}$$

where $f \in \left(H^{1,1/2}_{0;,0}(Q_T)\right)^*$ is given and

$$a(u,v) := \langle \partial_t u, v\rangle_{L_2(Q_T)} + \left(\nabla_x u, \nabla_x v\right)_{L_2(Q_T)}$$

is continuous bilinear form for $u \in H^{1,1/2}_{0;0,}(Q_T)$ and $v \in H^{1,1/2}_{0;,0}(Q_T)$.

Theorem 6.7 (Unique solution) *Let the right hand side $f \in \left(H^{1,1/2}_{0;,0}(Q_T)\right)^*$ be given. Then the variational formulation* (6.24) *has a unique solution $u \in H^{1,1/2}_{0;0,}(Q_T)$, satisfying*

$$\|u\|_{H^{1,1/2}_{0;0,}(Q_T)} \leq C\,\|f\|_{\left(H^{1,1/2}_{0;,0}(Q_T)\right)^*}$$

with a constant $C > 0$. Furthermore, the solution operator

$$\mathcal{L}\colon \left(H^{1,1/2}_{0;,0}(Q_T)\right)^* \to H^{1,1/2}_{0;,0}(Q_T), \quad \mathcal{L}f := u$$

is an isomorphism. In addition, the bilinear form

$$a(\cdot,\cdot)\colon H^{1,1/2}_{0;0,}(Q_T) \times H^{1,1/2}_{0;,0}(Q_T) \to \mathbb{R}, \quad a(u,v) := \langle \partial_t u, v\rangle_{L_2(Q_T)} + \left(\nabla_x u, \nabla_x v\right)_{L_2(Q_T)}$$

is continuous and fulfils an inf-sup condition and the surjectivity condition.

The theorem with proof can be found in (Zank 2019, Section 3.3). As highlighted in the remark in (Zank 2019, Section 3.3), it is important to note that the use of different solution space $H^{1,1/2}_{0;0,}(Q_T)$ and test space $H^{1,1/2}_{0;,0}(Q_T)$ is crucial for the bilinear form $a(u,v)$, as there is no continuous extension of $\langle \partial_t u, v\rangle_{(0,T)}$ for u, $v \in C_0^\infty((0,T))$ to

$$H^{1/2}_{0,}((0,T)) \times H^{1/2}_{0,}((0,T)) \quad \text{or} \quad H^{1/2}((0,T)) \times H^{1/2}((0,T)).$$

Additionally, the discrete formulation of (6.24) using a conforming solution space $S_h^1(Q_T) \cap H_{0;0,}^{1/2}(Q_T)$ and conforming test space $S_h^1(Q_T) \cap H_{0;,0}^{1/2}(Q_T)$, where $S_h^1(Q_T)$ is defined by (6.20), leads to an unstable method as stated in (Zank 2019, Section 3.3). To overcome this drawback, a modified Hilbert transformation $\mathcal{H}_T$ introduced in (Zank 2019) can be utilised. The modified Hilbert transformation $\mathcal{H}_T$ is defined as

$$(\mathcal{H}_T u)(\mathbf{x}, t) := \sum_{i=1}^{\infty} \sum_{k=0}^{\infty} u_{i,k} \cos\left(\left(\frac{\pi}{2} + k\pi\right) \frac{t}{T}\right) \phi_i(\mathbf{x}), \quad (\mathbf{x}, t) \in Q_T, \tag{6.25}$$

where the given $u \in L^2(Q_T)$ is expressed by

$$u(\mathbf{x}, t) := \sum_{i=1}^{\infty} \sum_{k=0}^{\infty} u_{i,k} \sin\left(\left(\frac{\pi}{2} + k\pi\right) \frac{t}{T}\right) \phi_i(\mathbf{x}), \quad (\mathbf{x}, t) \in Q_T. \tag{6.26}$$

The functions $\phi_i \in H_0^1(\Omega)$ represent the eigenfunctions, and along with the eigenvalues $\mu_i \in \mathbb{R}$, satisfy

$$-\Delta_{\mathbf{x}} \phi_i = \mu_i \phi_i \quad \text{in } \Omega, \quad \phi_i = 0 \quad \text{on } \partial\Omega, \quad \|\Phi_i\|_{L^2(\Omega)} = 1, \quad i \in \mathbb{N}.$$

Using the Hilbert transformation $\mathcal{H}_T$ the variational formulation (6.24) is rewritten to find $u \in H_{0;0,}^{1,1/2}(Q_T)$ such that

$$a(u, \mathcal{H}_T v) = \langle f, \mathcal{H}_T v \rangle_{Q_T} \quad \forall v \in H_{0;0,}^{1,1/2}(Q_T). \tag{6.27}$$

One can observe that the solution space and the space of the test functions are equal. Additionally, the discretisation of (6.27) using any conforming finite element space $V_h \subset H_{0;0,}^{1,1/2}(Q_T)$ gives an unconditionally stable method and leads to the discrete variational formulation to find $u_h \in V_h$ such that

$$a(u_h, \mathcal{H}_T v_h) = \langle f, \mathcal{H}_T v_h \rangle_{Q_T} \quad \forall v_H \in V_h. \tag{6.28}$$

The following theorem can be proved.

Theorem 6.8 (Unique solution of Hilbert transformation) *Let $V_h \subset H_{0;0,}^{1,1/2}(Q_T)$ be a conforming space-time finite element space, and let $f \in \left(H_{0;,0}^{1,1/2}(Q_T)\right)^*$ be a given right-hand side. Then a unique solution $u_h \in V_h$ of the Galerkin variational formulation* (6.28) *exists. If, in addition, the right-hand side fulfills $f \in \left(H_{,0}^{1/2}((0,T); L^2(\Omega))\right)^* \subset \left(H_{0;,0}^{1,1/2}(Q_T)\right)^*$, then the stability estimate*

$$\|u_h\|_{H_{0,}^{1/2}((0,T);L^2(\Omega))} \leq c \|f\|_{\left(H_{,0}^{1/2}((0,T);L^2(\Omega))\right)^*}$$

is true with a constant $c > 0$.

The proof is given by (Zank 2019, Theorem 3.4.20).

As it is stated in (Langer et al. 2021, Section 2), the space-time error estimates are derived when the tensor-product space-time finite element space

$$Q_h^p(Q_T) := V_{h_x}^p(\Omega) \otimes S_{h_t}^p((0,T)), \tag{6.29}$$

where $p \in \mathbb{N}$ is a fixed polynomial degree of the piece-wise polynomial continuous functions, is considered. For example, if $p = 1$, then $V_{h_x}^1(\Omega)$ is either the space $S_{h_x}^1(\Omega) \cap H_0^1(\Omega)$ of piece-wise linear continuous functions over intervals for 1d spatial domain, triangles for 2d spatial domain, and tetrahedra for 3d spatial domain, or it is the space $Q_{h_x}^1(\Omega) \cap H_0^1(\Omega)$ of piece-wise multilinear continuous functions over intervals for 1d spatial domain, quadrilaterals for 2d spatial domain, and hexahedra for 3d spatial domain. Then $V_h = Q_h^p(Q_T) \cap H_{0;0,}^{1,1/2}(Q_T)$ in (6.28), and the following estimates hold true

$$\begin{aligned} \|u - u_h\|_{H_{0,}^{1/2}((0,T);L^2(\Omega))} &\leq c\, h^{p+1/2}, \\ \|u - u_h\|_{L^2(Q_T)} &\leq c\, h^{p+1}, \\ |u - u_h|_{H^1(Q_T)} &\leq c\, h^p, \end{aligned}$$

where h is the maximal mesh size and $c > 0$ for a sufficiently smooth solution $u \in H_{0;0,}^{1,1/2}(\Omega)$ and a sufficiently regular domain Ω. For the error in the space $H^1(Q_T)$, it is additionally assumed that the sequence of decompositions of Ω is globally quasi-uniform (4.2).

6.3 Combining Fast Diagonalisation Method and PRESB

The section is based on the Fast diagonalisation method (FDM), as it is described in (Langer et al. 2021), and the PRESB method presented in (Axelsson; Lukáš 2019; Axelsson; Neytcheva 2018). Consider the heat equation (3.1) with the initial condition $u_0(\mathbf{x}) = 0$, and the tensor-product space-time finite element space (6.29) for $p = 1$. Thus, the space

$$S_{h_x}^1(\Omega) = \operatorname{span}\{\psi_i, \ldots, \psi_{N_x}\}$$

is the finite dimensional space of piece-wise linear continuous functions over intervals for 1d spatial domain, triangles for 2d spatial domain, and tetrahedra for 3d spatial domain, and

$$S_{h_t}^1(\Omega) = \operatorname{span}\{\varphi_i, \ldots, \varphi_{N_t}\}$$

is the finite dimensional space of piece-wise linear continuous functions over intervals. Then the discrete space-time variational formulation (6.15) is obtained leading to the following system of linear equations

$$(\mathrm{A}_{h_t} \otimes \mathrm{M}_{h_x} + \mathrm{M}_{h_t} \otimes \mathrm{A}_{h_x})\, \mathbf{u} = \mathbf{b}. \tag{6.30}$$

The FDM allows us to construct n independent spatial subproblems of the tensor-product system (6.30). The temporal information is conveyed to the spatial subdomains through the eigenvalues of a matrix pencil $(\mathrm{M}_{h_t}, \mathrm{A}_{h_t})$. Thus, it is necessary to solve the generalised eigenvalue problem

$$\mathrm{M}_{h_t}\, \mathbf{z} = \lambda\, \mathrm{A}_{h_t}\, \mathbf{z}, \tag{6.31}$$

where $\lambda = \alpha + \beta i \in \mathbb{C}$ are complex generalised eigenvalues and $\mathbf{z} = \mathbf{x} + \mathbf{y}i \in \mathbb{C}^{N_t}$ are complex generalised eigenvectors. Specifically, the FDM uses an eigenvalue decomposition given by

$$\mathrm{A}_{h_t}^{-1}\, \mathrm{M}_{h_t} = \mathrm{X}_t\, \mathrm{D}_t\, \mathrm{X}_t, \tag{6.32}$$

where each column of $\mathrm{X}_t \in \mathbb{C}^{N_t \times N_t}$ represents the complex generalised eigenvector $\mathbf{z}_k$ for the corresponding complex eigenvalue λ_k, $k = 1, \ldots, N_t$, and $\mathrm{D}_t \in \mathbb{C}^{N_t \times N_t}$ is a diagonal matrix of the complex eigenvalues λ_k, i.e., $\mathrm{D_t} = \mathrm{diag}\{\lambda_1, \ldots, \lambda_{N_t}\}$. It has to be ensured that

$$\Re(\lambda_k) = \alpha_k > 0, \quad k = 1, \ldots, N_t,$$

to secure the convergence of the FDM. Since the matrix A_{h_t} is not symmetric, the resulting eigenvalues forms conjugate pairs $\lambda = \alpha \pm \beta i$. Moreover, due to lack of symmetry of A_{h_t}, the condition number of X_t is not equal to 1 and may be large, i.e., X_t is not unitary. Therefore, it is recommended to provide an additional singular value decomposition

$$\mathrm{X}_t = \mathrm{U}_t \, \Sigma_t \, \mathrm{V}_t^*, \tag{6.33}$$

where U_t, $\mathrm{V}_t \in \mathbb{C}^{N_t \times N_t}$ are unitary matrices and $\Sigma_t \in \mathbb{C}^{N_t \times N_t}$ is a diagonal matrix. This is done in order to dampen the numerical instabilities that may occur when computing the inverse of X_t. Furthermore, by defining

$$\mathrm{Y}_t := (\mathrm{A}_{h_t} \, \mathrm{X}_t) = \mathrm{V}_t \, \Sigma_t^{-1} \, \mathrm{U}_t^* \, \mathrm{A}_{h_t}^{-1}$$

we obtain the representations

$$\mathrm{A}_{h_t} = \mathrm{Y}_t^{-1} \, \mathrm{X}_t^{-1} \quad \text{and} \quad \mathrm{M}_{h_t} = \mathrm{Y}_t^{-1} \, \mathrm{D}_t \, \mathrm{X}_t^{-1}.$$

Then, using the diagonalisation (6.32), the solution $\mathbf{u}$ of (6.30) can be rewritten to the form

$$\mathbf{u} = (\mathrm{X}_t \otimes \mathrm{I}_{N_x}) \, (\mathrm{I}_{N_t} \otimes \mathrm{M}_{h_x} + \mathrm{D}_t \otimes \mathrm{A}_{h_x})^{-1} \, (\mathrm{Y}_t \otimes \mathrm{I}_{N_x}) \, \mathbf{b} \tag{6.34}$$

as described in (Langer et al. 2021). Here, $\mathrm{I}_{N_t} \in \mathbb{R}^{N_t \times N_t}$ and $\mathrm{I}_{N_x} \in \mathbb{R}^{N_x \times N_x}$ are identity matrices. The FDM algorithm can be summarised as follows:

1. Compute the eigenvalue decomposition (6.32).
2. Compute the singular value decomposition (6.33).
3. Transform the right-hand side

$$\mathbf{g} = (\mathbf{g}_1, \ldots, \mathbf{g}_{N_t})^T := (\mathrm{Y}_t \otimes \mathrm{I}_{N_x}) \, \mathbf{b} \in \mathbb{C}^{N_t \cdot N_x}. \tag{6.35}$$

4. Solve in parallel for $k = 1, \ldots, N_t$

$$(\mathrm{M}_{h_x} + \lambda_k \mathrm{A}_{h_x}) \, \mathbf{z}_k = \mathbf{g}_k, \tag{6.36}$$

   where $\mathbf{z} = (\mathbf{z}_1, \ldots, \mathbf{z}_{N_t})^T \in \mathbb{C}^{N_t \cdot N_x}$.

5. Compute solution $\mathbf{u}$ using the following transformation

$$\mathbf{u} = (\mathrm{X}_t \otimes \mathrm{I}_{N_x}) \, \mathbf{z} = (\mathrm{U}_t \, \Sigma_t \, \mathrm{V}_t^* \otimes \mathrm{I}_{N_x}) \, \mathbf{z} \in \mathbb{C}^{N_t \cdot N_x}. \tag{6.37}$$

Assuming the right-hand side vectors $\mathbf{b}$ and $\mathbf{z}$ are written in matrix form as B and $\mathrm{Z} \in \mathbb{C}^{N_x \times N_t}$, where each column is associated with a spatial subproblem at time t_k, $k = 1, \ldots, N_t$, the transformation steps (6.35) and (6.37) can be computed by

$$\begin{aligned} \mathrm{G} &:= \left(\mathrm{Y}_t \, \mathrm{B}^T\right)^T \in \mathbb{C}^{N_x \times N_t}, \\ \mathrm{U}_s &:= \left[\left(\mathrm{U}_t \, \Sigma_t \, \mathrm{V}_t^*\right) \, \mathrm{Z}^T\right]^T \in \mathbb{C}^{N_x \times N_t}. \end{aligned}$$

Then the resulting RHS vector $\mathbf{g}$ and the solution vector $\mathbf{u}$ are obtained by

$$\begin{aligned} \mathrm{g} &:= \left(G_1^c, \ldots, G_{N_t}^c\right)^T \in \mathbb{C}^{N_t \cdot N_x}, \\ \mathrm{u} &:= \left(\mathrm{U}_{1,s}^c, \ldots, \mathrm{U}_{N_t,s}^c\right)^T \in \mathbb{C}^{N_t \cdot N_x}, \end{aligned}$$

where superscript "c" represents the kth column of corresponding matrix. As stated in (Langer et al. 2021, Section 4.4.1), the overall computational cost is $\mathcal{O}\left(N_t^3 + N_x N_t^2 + C_C(N_x) \cdot N_t\right)$, where $C_C(\cdot)$ is some cost function, and memory consumption is $\mathcal{O}\left(N_t^2 + N_x N_t + C_S(N_x) \cdot N_t\right)$, where $C_S(\cdot)$ is some storage function. Given bounds are valid when a Cholesky factorisation of A_{h_t}, along with sparse direct solver described in (Langer et al. 2021, Section 3.2), is used.

Since the system (6.36) is complex, i.e.,

$$\left(\mathrm{M}_{h_x} + \overbrace{(\alpha_k + \beta_k i)}^{=\lambda_k} \mathrm{A}_{h_x}\right) \overbrace{(\mathbf{u}_k + \mathbf{v}_k i)}^{=\mathbf{z}_k} = \overbrace{(\mathbf{b}_k + \mathbf{c}_k i)}^{=\mathbf{g}_k},$$

it can be rewritten into a two-by-two block matrix form

$$\begin{pmatrix} \mathrm{M}_{h_x} + \alpha_k \mathrm{A}_{h_x} & -\beta_k \mathrm{A}_{h_x} \\ \beta_k \mathrm{A}_{h_x} & \mathrm{M}_{h_x} + \alpha_k \mathrm{A}_{h_x} \end{pmatrix} \cdot \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} = \begin{pmatrix} \mathbf{b} \\ \mathbf{c} \end{pmatrix}, \tag{6.38}$$

which can be preconditioned using the PRESB method. As mentioned above, the resulting eigenvalues λ form conjugate pairs. Thus two cases of the two-by-two block matrix have to be considered. Let $k = 1, \ldots, N_t$.

1. If $\beta_k < 0$, then (6.38) leads to

$$\begin{pmatrix} \mathcal{A}_k & \mathcal{B}_k^I \\ -\mathcal{B}_k^I & \mathcal{A}_k \end{pmatrix} \cdot \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} = \begin{pmatrix} \mathbf{b} \\ \mathbf{c} \end{pmatrix}, \tag{6.39}$$

   where $\mathcal{A}_k := \mathrm{M}_{h_x} + \alpha_k \mathrm{A}_{h_x}$ and $\mathcal{B}_k^I := -\beta_k \mathrm{A}_{h_x}$. With this at hand, the resulting matrix $\mathcal{A}_k + \mathcal{B}_k = \mathrm{M}_{h_x} + (\alpha_k - \beta_k)\mathrm{A}_{h_x}$ which is required within the preconditioning, is symmetric positive definite.

2. If $\beta_k > 0$, the following trick is done. Assume

$$\widetilde{\mathbf{v}} := -\mathbf{v} \tag{6.40}$$

and multiply the second row of (6.38) by -1, then subsequent two-by-two block matrix is obtained

$$\begin{pmatrix} \mathcal{A}_k & \mathcal{B}_k^{II} \\ -\mathcal{B}_k^{II} & \mathcal{A}_k \end{pmatrix} \cdot \begin{pmatrix} \mathbf{u} \\ \widetilde{\mathbf{v}} \end{pmatrix} = \begin{pmatrix} \mathbf{b} \\ -\mathbf{c} \end{pmatrix}, \tag{6.41}$$

where $\mathcal{B}_k^{II} := \beta_k \mathrm{A}_{h_x}$. As in the previous case, the resulting matrix $\mathcal{A}_k + \mathcal{B}_k = \mathrm{M}_{h_x} + (\alpha_k + \beta_k)\mathrm{A}_{h_x}$ is symmetric positive definite. Since it is assumed (6.40), the backward substitution have to be provided after the imaginary part of the solution $\widetilde{\mathbf{v}}$ is computed.

The preconditioner, denoted as $\mathcal{C}$, to the system (6.39) and (6.41) is defined as

$$\mathcal{C} := \begin{pmatrix} \mathcal{A}_k + 2\mathcal{B}_k^j & \mathcal{B}_k^j \\ -\mathcal{B}_k^j & \mathcal{A}_k \end{pmatrix}, \tag{6.42}$$

where $j = \{I, II\}$. The preconditioner $\mathcal{C}$ can be used in a Krylov subspace type of iteration method. Given that the resulting system is not symmetric, the Flexible Inner-Outer Preconditioned GMRES method (FGMRES), as described in (Saad 1993), is utilised. A linear matrix preconditioning equation is defined as follows:

$$\mathcal{C} \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = \begin{pmatrix} \mathbf{r} \\ \mathbf{s} \end{pmatrix}. \tag{6.43}$$

Precisly, the preconditioning equation (6.43) is equal to

$$\begin{aligned} \left(\mathcal{A}_k + 2\mathcal{B}_k^j\right)\mathbf{x} + \mathcal{B}_k^j\,\mathbf{y} &= \mathbf{r}, \\ -\mathcal{B}_k^j\,\mathbf{x} + \mathcal{A}_k\,\mathbf{y} &= \mathbf{s}, \end{aligned} \tag{6.44}$$

$j = \{I, II\}$. By adding the first equation of (6.44) to the second one, the equivalent system of equations is obtained

$$\begin{aligned} \left(\mathcal{A}_k + 2\mathcal{B}_k^j\right)\mathbf{x} + \mathcal{B}_k^j\,\mathbf{y} &= \mathbf{r}, \\ \left(\mathcal{A}_k + \mathcal{B}_k^j\right)\mathbf{x} + \left(\mathcal{A}_k + \mathcal{B}_k^j\right)\mathbf{y} &= \mathbf{r} + \mathbf{s}. \end{aligned} \tag{6.45}$$

Using a substitution $\mathbf{z} := \mathbf{x} + \mathbf{y}$, the resulting system is in the form

$$\begin{pmatrix} \mathcal{A}_k + \mathcal{B}_k^j & \mathcal{B}_k^j \\ \mathrm{O} & \mathcal{A}_k + \mathcal{B}_k^j \end{pmatrix} \cdot \begin{pmatrix} \mathbf{x} \\ \mathbf{z} \end{pmatrix} = \begin{pmatrix} \mathbf{r} \\ \mathbf{r} + \mathbf{s} \end{pmatrix}.$$

Thus the preconditioning algorithm can be summarised as follows.

1. Solve $\left(\mathcal{A}_k + \mathcal{B}_k^j\right)\mathbf{z} = \mathbf{r} + \mathbf{s}$,
2. Solve $\left(\mathcal{A}_k + \mathcal{B}_k^j\right)\mathbf{x} = \mathbf{r} + \mathcal{B}_k^j\mathbf{z}$,
3. Compute $\mathbf{y} = \mathbf{z} - \mathbf{x}$.

As mentioned above, the matrix $\left(\mathcal{A}_k + \mathcal{B}_k^j\right)$, $j = \{I, II\}$, which is applied in the first and the second step, is symmetric positive definite.

6.4 Numerical experiments

In this section, numerical experiments of the space-time FEM in 1d + time and 2d + time are presented in order to confirm the theoretical estimation (6.23) in the Bochner-Sobolev space. Furthermore, the iteration numbers of the coupled Fast Diagonalisation Method with the PRESB method for a case of 3d + time domain are presented to demonstrate the scalability of the proposed combination.

6.4.1 Space-time FEM

The experiments were performed in MATLAB on the HP-Spectre laptop again. The space-time problems in 1d + time and 2d + time using the space-time FEM were solved. For the case of 1d + time, the space-time domain were decomposed into right-angled isosceles triangles. In the case of 2d + time, uniform tetrahedrons were used. Thus the mesh sizes h_x and h_t, where h_x is the discretisation step in the spatial domain and h_t in the time interval, were equal. We observed three errors. The first one was given by

$$e_1 := \|u_h(\mathbf{x},t) - u(\mathbf{x},t)\|_{L^2(Q_T)} = \left(\int_0^T \int_\Omega \left(u_h(\mathbf{x},t) - u(\mathbf{x},t)\right)^2 \, \mathrm{d}\mathbf{x}\, \mathrm{d}t \right)^{1/2}, \tag{6.46}$$

the second by

$$e_2 := \left\|\nabla_{\mathbf{x}}(u_h(\mathbf{x},t) - u(\mathbf{x},t))\right\|_{L^2(Q_T)} = \left(\int_0^T \int_\Omega \left[\nabla_{\mathbf{x}}(u_h(\mathbf{x},t) - u(\mathbf{x},t))\right]^2 \, \mathrm{d}\mathbf{x}\, \mathrm{d}t \right)^{1/2}, \tag{6.47}$$

and the last one by

$$\begin{aligned} e_3 &:= \|u_h(\mathbf{x},t) - u(\mathbf{x},t)\|_{L^2((0,T);H_0^1(\Omega))} \\ &= \left(\int_0^T \int_\Omega \left(u_h(\mathbf{x},t) - u(\mathbf{x},t)\right)^2 + \left[\nabla_{\mathbf{x}}(u_h(\mathbf{x},t) - u(\mathbf{x},t))\right]^2 \, \mathrm{d}\mathbf{x}\, \mathrm{d}t \right)^{1/2}, \end{aligned} \tag{6.48}$$

where $u(\mathbf{x},t)$ is the exact solution and $u_h(\mathbf{x},t)$ is the approximate solution obtained by the space-time FEM.

1. Consider the first example as in 4.3. The resulting errors are presented in Table 6.1. For both cases, 1d + time and 2d + time, the eoc of the e_1 was 2 and the eoc of the e_2 was 1. As it was proposed by the theory, i.e., by the estimate (6.23), the eoc of the e_3 was 1.

2. Consider the second example as in 4.3. The resulting errors are shown in Table 6.2. As in the previous example, the eoc of the e_1 was 2, the eoc of the e_2 was 1, and the eoc of the e_3 was 1.

Table 6.1: Errors of the space-time FEM – example 1.

$h_x = h_t$		1/8	1/16	1/32	1/64	1/128
e_1	$d = 1$	$9.73e{-3}$	$2.44e{-3}$	$6.10e{-4}$	$1.52e{-4}$	$3.81e{-5}$
	eoc		2.00	2.00	2.00	2.00
	$d = 2$	$1.22e{-2}$	$3.10e{-3}$	$7.80e{-4}$	$1.95e{-4}$	$4.89e{-5}$
	eoc		1.98	1.99	2.00	2.00
e_2	$d = 1$	$2.12e{-1}$	$1.06e{-1}$	$5.32e{-2}$	$2.66e{-2}$	$1.33e{-2}$
	eoc		1.00	1.00	1.00	1.00
	$d = 2$	$2.32e{-1}$	$1.15e{-1}$	$5.72e{-2}$	$2.85e{-2}$	$1.42e{-2}$
	eoc		1.01	1.01	1.00	1.00
e_3	$d = 1$	$2.12e{-1}$	$1.06e{-1}$	$5.32e{-2}$	$2.66e{-2}$	$1.33e{-2}$
	eoc		1.00	1.00	1.00	1.00
	$d = 2$	$2.32e{-1}$	$1.15e{-1}$	$5.72e{-2}$	$2.85e{-2}$	$1.42e{-2}$
	eoc		1.01	1.01	1.00	1.00

Table 6.2: Errors of the space-time FEM – example 2.

$h_x = h_t$		1/8	1/16	1/32	1/64	1/128
e_1	$d = 1$	$1.22e{-3}$	$3.08e{-4}$	$7.73e{-5}$	$1.93e{-5}$	$4.83e{-6}$
	eoc		1.99	2.00	2.00	2.00
	$d = 2$	$3.02e{-4}$	$7.59e{-5}$	$1.90e{-5}$	$4.75e{-6}$	$1.20e{-6}$
	eoc		1.99	2.00	2.00	1.98
e_2	$d = 1$	$2.46e{-2}$	$1.23e{-2}$	$6.17e{-3}$	$3.08e{-3}$	$1.54e{-3}$
	eoc		1.00	1.00	1.00	1.00
	$d = 2$	$7.14e{-3}$	$3.58e{-3}$	$1.79e{-3}$	$8.94e{-4}$	$4.47e{-4}$
	eoc		1.00	1.00	1.00	1.00
e_3	$d = 1$	$2.46e{-2}$	$1.23e{-2}$	$6.17e{-3}$	$3.08e{-3}$	$1.54e{-3}$
	eoc		1.00	1.00	1.00	1.00
	$d = 2$	$7.15e{-3}$	$3.58e{-3}$	$1.79e{-3}$	$8.94e{-4}$	$4.47e{-4}$
	eoc		1.00	1.00	1.00	1.00

6.4.2 Combination of the FDM and PRESB method

The Fast Diagonalisation Method, in combination with the PRESB technique 6.3, was tested on the Karolina cluster at the IT4Innovations, VSB – Technical University of Ostrava, Ostrava, Czech Republic, and was implemented in C++. The study focused on the heat equation on a cube $(-1,1)^3$ over the time interval $(0,1)$. In particular, the equation (3.1) in three spatial dimensions in time was analysed, with an initial condition of $u_0(\mathbf{x}) = 0$. As a case study, the spatial subproblems (6.36) were assumed to be in the form

$$(\mathrm{M}_{h_x} + \lambda_k \mathrm{A}_{h_x})\, \mathbf{z}_k = \mathbf{1} \tag{6.49}$$

for all $k = 1, 2, \ldots, N_t$. The given subproblems, which were further considered in the form (6.39) if $\beta_k < 0$ and (6.41) if $\beta_k > 0$, were solved using the FGMRES method in which the PRESB was implemented. Within the PRESB method, the multigrid algorithm with PCG was utilised. The temporal steps were $h_t := 1/32,\ 1/64,\ 1/128,\ 1/256$. The precision of the FGMRES was set to $1e{-}8$, and the precision of the PCG to $1e{-}2$. The precision $1e{-}2$ is sufficient for the purpose of the preconditioning.

The study focused on the number of iterations of the FGMRES method and the inner PCG method for different multigrid levels. Initially, the FGMRES for a 0-level multigrid, i.e., the FGMRES without the multigrid, was carried out. The number of spatial DOFs was 2 395. In the second experiment, the proposed method using a 1-level multigrid was tested. The number of refined spatial DOFs was 16 433. In the final experiment, a 2-level multigrid was employed. The number of refined spatial DOFs for the 2-level multigrid was 121 265. The results for the number of iterations for the outer FGMRES method and the inner PCG are presented in Table 6.3. The lowest number of iterations, i.e., 7 for GMRES and 12 for PCG in the first column of Table 6.3 for the 0-level multigrid, correspond to the complex number λ_k with the largest real part α_k. On the contrary, the largest number of iterations, i.e., 12 for FGMRES and 25 for PCG, correspond to the complex numbers with the lowest real part α_k. Additionally, it can be observed that the multigrid refinement did not affect the number of iterations. We note that each iteration number represents the total iteration number which is needed to solve the spatial subproblem (6.49) for the corresponding λ_k, $k = 1, 2, \ldots, N_t$.

Table 6.3: Iterations of FGMRES and underlying PCG using PRESB preconditioning.

multigrid	spatial DOFs		$h_t = 1/32$	$h_t = 1/64$	$h_t = 1/128$	$h_t = 1/256$
0-level	2 395	FGMRES	7 – 12	7 – 12	7 – 12	7 – 12
		PCG	12 – 23	13 – 23	13 – 23	13 – 23
1-level	16 433	FGMRES	8 – 13	8 – 13	8 – 13	8 – 13
		PCG	15 – 25	15 – 25	15 – 25	15 – 25
2-level	121 265	FGMRES	8 – 13	8 – 13	8 – 13	8 – 13
		PCG	17 – 25	17 – 25	17 – 25	17 – 25

Chapter 7

Conclusion

In the first part of this doctoral thesis, the well-posedness of the weak formulation of the parabolic equation via the Galerkin method was proposed, as stated in reference (Zeidler 1990a). The theorem was followed by the proof, which was examined in greater detail than provided in the mentioned reference. The section on the weak formulation of the parabolic equation shall serve as a theoretical background for future work to solve more comprehensive parabolic problems. Another potential topic for future work is to establish a theoretical basis for the well-posedness of the weak formulation of linear partial differential equations of the second order, commonly known as the wave equation.

In Section 5, the Parareal algorithm was briefly outlined. It offers a straightforward parallel scheme for solving space-time problems by utilising the semi-discrete method. Additionally, three potential implementation options of the Parareal method were described. The distributed version of the algorithm was tested within the thesis providing promising results. As an alternative to the distributed program, the version with improved overlap, in which all used CPUs are well-balanced, could be employed. However, the improved overlap algorithm has a higher memory demand than the distributed algorithm but offers superior efficiency. At the end of this section, a combination of the Parareal with the DDM based on the Schur complement approximation was proposed. This combination allows us to increase the parallelism in time through the parallelism in the spatial subproblems. The results are also promising, but they require validation through large-scale testing. A potential drawback of the proposed method is that it wastes memory, as the same DDM is employed at each time slice. Therefore, an efficient combination of the Parareal with a single DDM that can handle multiple right-hand sides should be developed in future research.

In the last section, the space-time FEM theory was briefly summarised. Two cases of variational formulations were provided: the formulation in the Bochner-Sobolev space and the formulation in the anisotropic Sobolev space. Subsequently, numerical examples utilising conforming space-time FEM in the Bochner-Sobolev space were presented. The results confirm the theoretical estimates outlined in (Steinbach 2015). Furthermore, a combination of the Fast Diagonalisation Method with the PRESB method was also proposed. This combination yielded promising results, as the number of iterations appeared to be stable when refining the problem using the underlying multigrid method. However, as this combination was tested while assuming a regular solution, testing with solutions exhibiting some singular behaviour should be consid-

ered in future research. Furthermore, large-scale tests have to also be performed. Additionally, future work should investigate the application of various DDM to space-time FEM, such as the finite element tearing and interconnecting method – FETI (Farhat et al. 1991).

Bibliography

AUBANEL, Eric, 2011. Scheduling of tasks in the parareal algorithm. *Parallel Computing*. Vol. 37, no. 3, pp. 172–182. ISSN 0167-8191. Available from DOI: 10.1016/j.parco.2010.10.004.

AXELSSON, Owe; LUKÁŠ, Dalibor, 2019. Preconditioning methods for eddy-current optimally controlled time-harmonic electromagnetic problems. *Journal of Numerical Mathematics*. Vol. 27, no. 1, pp. 1–21. Available from DOI: 10.1515/jnma-2017-0064.

AXELSSON, Owe; NEYTCHEVA, Maya, 2018. *Preconditioners for two-by-two block matrices with square blocks*. Department of Information Technology, Uppsala Universitet.

BRAMBLE, James H.; PASCIAK, Joseph E.; SCHATZ, Alfred H., 1986. The Construction of Preconditioners for Elliptic Problems by Substructuring. I. *Mathematics of Computation*. Vol. 47, no. 175, pp. 103–134.

DOLEJSI, Vit; FEISTAUER, Miloslav, 2015. *Discontinuous Galerkin Method*. ISBN 978-3-319-19266-6. Available from DOI: 10.1007/978-3-319-19267-3.

FARHAT, Charbel; ROUX, Francois-Xavier, 1991. A method of finite element tearing and interconnecting and its parallel solution algorithm. *International Journal for Numerical Methods in Engineering*. Vol. 32, no. 6, pp. 1205–1227. Available from DOI: 10.1002/nme.1620320604.

FIDKOWSKI, Krzysztof J, 2019. Comparison of hybrid and standard discontinuous Galerkin methods in a mesh-optimisation setting. *International Journal of Computational Fluid Dynamics*. Vol. 33, no. 1-2, pp. 34–42. Available from DOI: 10.1080/10618562.2019.1588962.

FOLTYN, Ladislav; LUKÁŠ, Dalibor; PETEREK, Ivo, 2020. Domain Decomposition Methods Coupled with Parareal for the Transient Heat Equation in 1 and 2 Spatial Dimension. *Applications of Mathematics*. Vol. 65, pp. 173–190. Available from DOI: 10.21136/AM.2020.0219-19.

GANDER, Martin J., 2015. 50 Years of Time Parallel Time Integration. In: CARRARO, Thomas; GEIGER, Michael; KÖRKEL, Stefan; RANNACHER, Rolf (eds.). *Multiple Shooting and Time Domain Decomposition Methods*. Cham: Springer International Publishing, pp. 69–113. ISBN 978-3-319-23321-5.

GANDER, Martin J.; VANDEWALLE, Stefan, 2007a. Analysis of the parareal time-parallel time-integration method. *SIAM Journal on Scientific Computing*. Vol. 29, no. 2, pp. 556–578. Available from DOI: 10.1137/05064607X.

GANDER, Martin J.; VANDEWALLE, Stefan, 2007b. On the superlinear and linear convergence of the parareal algorithm. In: WIDLUND, Olof B.; KEYES, David E. (eds.). *Domain decomposition methods in science and engineering XVI*. Springer Berlin Heidelberg, pp. 291–298. ISBN 978-3-540-34469-8.

LANGER, Ulrich; ZANK, Marco, 2021. Efficient Direct Space-Time Finite Element Solvers for Parabolic Initial-Boundary Value Problems in Anisotropic Sobolev Spaces. *SIAM Journal on Scientific Computing*. Vol. 43, no. 4, pp. A2714–A2736. Available from DOI: 10.1137/20M1358128.

LEHRENFELD, Christoph, 2010. *Hybrid Discontinuous Galerkin methods for solving incompressible flow problems*. Computational Engineering Science, Rheinisch-Westfalischen Technischen Hochschule Aachen.

LIONS, Jacques-Louis; MADAY, Yvon; TURINICI, Gabriel, 2001. Résolution d'EDP par un schéma en temps ≪ pararéel ≫. *Comptes Rendus de l'Académie des Sciences*. Vol. 332, no. 7, pp. 661–668. ISSN 0764-4442. Available from DOI: 10.1016/S0764-4442(00)01793-6.

LUKÁŠ, Dalibor; BOUCHALA, Jiří.; VODSTRČIL, Petr; MALÝ, Lukáš, 2015. 2-Dimensional primal domain decomposition theory in detail. *Applications of Mathematics*. Vol. 60, pp. 265–283. Available from DOI: 10.1007/s10492-015-0095-5.

MADAY, Yvon, 2008. The Parareal in time algorithm. In: MAGOULÈS, Frédéric (ed.). *Substructuring Techniques and Domain Decomposition Methods*. Stirlingshire, UK: Saxe-Coburg Publications, chap. 2, pp. 19–44. Available from DOI: 10.4203/csets.24.2.

MANDEL, Jan; BREZINA, Marian, 1996. Balancing Domain Decomposition for Problems with Large Jumps in Coefficients. *Mathematics of Computation*. Vol. 65, no. 216, pp. 1387–1401. Available from DOI: 10.1090/S0025-5718-96-00757-0.

MERCERAT, Diego; GUILLOT, Laurent; VILOTTE, Jean-Pierre, 2009. Application of the parareal algorithm for acoustic wave propagation. Vol. 1168, no. 1, pp. 1521–1524. Available from DOI: 10.1063/1.3241388.

NEUMÜLLER, Martin, 2013. *Space-Time Methods: Fast Solvers and Applications*. PhD thesis. Technische Universität Graz.

NEUMÜLLER, Martin; STEINBACH, Olaf, 2011. Refinement of flexible space–time finite element meshes and discontinuous Galerkin methods. *Computing and Visualization in Science*. Vol. 14, no. 5, pp. 189–205. Available from DOI: 10.1007/s00791-012-0174-z.

SAAD, Youcef, 1993. A flexible inner-outer preconditioned GMRES algorithm. *SIAM Journal on Scientific Computing*. Vol. 14, no. 2, pp. 461–469. Available from DOI: 10.1137/0914028.

SCHÖPS, Sebastian; NIYONZIMA, Innocent; CLEMENS, Markus, 2017. Parallel-in-time simulation of eddy current problems using parareal. *IEEE Transactions on Magnetics*. Vol. 54, no. 3, pp. 1–4. Available from DOI: 10.1109/TMAG.2017.2763090.

SMITH, Barry F.; BJØRSTAD, Petter E.; GROPP, William D., 1996. *Domain Decomposition: Parallel Multilevel Methods for Elliptic Partial Differential Equations*. Cambridge University Press. ISBN 978-0-521-60286-0.

STEINBACH, Olaf, 2015. Space-Time Finite Element Methods for Parabolic Problems. *Computational Methods in Applied Mathematics*. Vol. 15, no. 4, pp. 551–566. Available from DOI: 10.1515/cmam-2015-0026.

STEINBACH, Olaf; YANG, Huidong, 2019. Space-time finite element methods for parabolic evolution equations: discretization, a posteriori error estimation, adaptivity and solution. In: *Applications to Partial Differential Equations*. Ed. by LANGER, Ulrich; STEINBACH, Olaf. Berlin, Boston: De Gruyter, pp. 207–248. ISBN 978-3-110-54848-8. Available from DOI: 10.1515/9783110548488-007.

THOMÉE, Vidar, 2006. *Galerkin Finite Element Methods for Parabolic Problems*. Springer. ISBN 978-3-540-33121-6. Available from DOI: doi.org/10.1007/3-540-33122-0.

TOSELLI, Andrea; WIDLUND, Olof B., 2004. *Domain Decomposition Methods - Algorithms and Theory*. Springer Science & Business Media. ISBN 978-3-540-20696-5. Available from DOI: doi.org/10.1007/b137868.

WOOPEN, Michael; BALAN, Aravind; MAY, Georg; SCHÜTZ, Jochen, 2014. A comparison of hybridized and standard DG methods for target-based hp-adaptive simulation of compressible flow. *Computers & Fluids*. Vol. 98, pp. 3–16. Available from DOI: 10.1016/j.compfluid.2014.03.023.

ZANK, Marco, 2019. *Inf-Sup Stable Space-Time Methods for Time-Dependent Partial Differential Equations*. PhD thesis. Technischen Universität Graz.

ZEIDLER, Eberhard, 1986. *Nonlinear Functional Analysis and its Applications I: Fixed-point Theorems*. Springer. ISBN 978-0-387-90914-1.

ZEIDLER, Eberhard, 1990a. *Nonlinear Functional Analysis and its Applications II/A: Linear Monotone Operators*. Springer. ISBN 978-0-387-96802-5. Available from DOI: 10.1007/978-1-4612-0985-0.

ZEIDLER, Eberhard, 1990b. *Nonlinear Functional Analysis and its Applications II/B: Nonlinear Monotone Operators*. Springer. ISBN 978-0-387-97167-4. Available from DOI: doi.org/10.1007/978-1-4612-0981-2.

Appendix A

Articles and projects

A.1 Articles

A.1.1 Thesis related articles

- Foltyn, L., Lukáš, D., and Peterek, I. (2020). Domain decomposition methods coupled with parareal for the transient heat equation in 1 and 2 spatial dimensions. Applications of Mathematics (IF 0.537), 65(2), 173-190.

A.1.2 Thesis unrelated articles

- Foltyn, Ladislav; Vlach, Oldřich. Implementation of full linearization in semismooth Newton method for 2D contact problem. Programs and Algorithms of Numerical Mathematics, 2017, 30-36.

- Foltyn, L., Vysocký, J., Prettico, G., Běloch, M., Praks, P., and Fulli, G. (2021). OPF solution for a real Czech urban meshed distribution network using a genetic algorithm. Sustainable Energy, Grids and Networks, 26, 100437.

- Vysocký, J., Foltyn, L., Brkić, D., Praksová, R., and Praks, P. (2022). Steady-State Analysis of Electrical Networks in Pandapower Software: Computational Performances of Newton–Raphson, Newton–Raphson with Iwamoto Multiplier, and Gauss–Seidel Methods. Sustainability, 14(4), 2002.

A.1.3 Thesis related projects

- Matematické modelování a vývoj algoritmů pro výpočetně náročné inženýrské úlohy (SP2020/114), role: student/co-worker, The Ministry of Education, Youth and Sports.

- Matematické modelování a vývoj algoritmů pro výpočetně náročné inženýrské úlohy (SP2019/84), role: student/co-worker, The Ministry of Education, Youth and Sports.

- Matematické modelování a vývoj algoritmů pro výpočetně náročné inženýrské úlohy (SP2018/165), role: student/co-worker, The Ministry of Education, Youth and Sports.

- Inovace kurzů numerických metod (RPP2017/197), role: Co-worker, The Ministry of Education, Youth and Sports.
- Matematické modelování a vývoj algoritmů pro výpočetně náročné inženýrské úlohy (SP2017/122), role: student/co-worker, The Ministry of Education, Youth and Sports.
- Matematické modelování a vývoj algoritmů pro výpočetně náročné inženýrské úlohy (SP2016/108), role: student/co-worker, The Ministry of Education, Youth and Sports.
- Matematické modelování a vývoj algoritmů pro výpočetně náročné inženýrské úlohy (SP2015/100), role: student/co-worker, The Ministry of Education, Youth and Sports.

A.1.4 Thesis unrelated projects

- Project of Technology Agency of the Czech Republic (TA CR) Energy System for Grids (ES4G) (TK02030039), role: co-worker in WP 5 Mining, Modelling, Forecasting responsible for grids modelling and optimization of processes.
  https://starfos.tacr.cz/cs/project/TK02030039
- EuroCC, National Competence Centres in the framework of EuroHPC, HORIZON 2020 (951732), role: co-worker, INDUSTRIAL LEADERSHIP - Leadership in enabling and industrial technologies - Information and Communication Technologies (ICT).
  https://cordis.europa.eu/project/id/951732
- Optimalizace provozních parametrů elektrické distribuční soustavy s využitím umělé inteligence (TJ02000157), role: co-worker, TA ČR - Zéta
- Interaktivní úlohy pro podporu výuky matematických předmětů (RPP2019/61), role: leader, The Ministry of Education, Youth and Sports.
- Interaktivní 3D grafika pro podporu výuky diferenciálního počtu funkcí více proměnných (RPP2016/126), role: leader, The Ministry of Education, Youth and Sports.

A.1.5 Thesis unrelated application results

- Vysocký, Jan, Lukáš Prokop, Stanislav Mišák, Pavel Praks a Ladislav Foltyn. System for optimizing the electrical distribution network operation. 2020. Utility model - software.
- Foltyn, Ladislav, Marek Lampart a Topolánek David. Software for verification and validation optimised models. Vysoká škola báňská - Technická univerzita Ostrava, 2021. Software.
- Foltyn, Ladislav, Renáta Praksová a Pavel Praks. Stochastic model for indentification of critical components in energy grid (VP5). Vysoká škola báňská - Technická univerzita Ostrava, 2021. Software.
- Martinovič, Tomáš, Judita Buchlovská Nagyová, Ladislav Foltyn a Radek Halfar. NTS - Network Traversal Simulator. VŠB - TU Ostrava, IT4Innovations, 2021. Software.

- Vysocký, Jan, Pavel Praks a Ladislav Foltyn. Software for optimizing the electrical distribution network operation under abnormal operating conditions of the network and in the event of network failures, 2020. Software.
- Foltyn, Ladislav, Jan Vysocký a Pavel Praks. Optimisation software - Optimisation Algorithm. Vysoká škola báňská - Technická univerzita Ostrava, 2020. Software.
- Vysocký, Jan, Michal Běloch, Pavel Praks a Ladislav Foltyn. Optimization Software - Mathematical model of a controlled distribution network. Vysoká škola báňská - Technická univerzita Ostrava, 2020. Software.
- Foltyn, Ladislav. Developing Shiny application. 2022. Best practise guide.