Nonconforming space–time methods for evolution PDEs

Joint PhD program in Computational Mathematics and Decision Sciences

Università degli studi di Pavia

Università della Svizzera Italiana

Presented by:

Sergio Alejandro Gómez Macías
(XXXVI cycle)

Accepted on the recommendation of:

Dr. Andrea Moiola (advisor)

Dr. Jay Gopalakrishnan

Dr. Karsten Urban

February 2024

*"... sabiendo que la tribulación
produce paciencia;
y la paciencia, prueba;
y la prueba, esperanza;
y la esperanza no avergüenza. . . "*
Romanos 5:3-5.

Acknowledgements

There are many people and institutions that I would like to thank:

- To Andrea Moiola, for all his guidance, patience, and advice during these years. Doubtlessly, I received his unconditional support even before the beginning of my PhD.

- To Lorenzo Mascotto, for the many laughs, the great discussions that taught me so much, and his friendship.

- To Ilaria Perugia, for the many enjoyable hours of work (and tea) together. I consider meeting her as one of the most wonderful experiences during my PhD.

- To Paul Castillo, for his lasting friendship (despite the distance) and the many hours of conversation that encouraged me during the hardest times. I owe him a big part of what I have achieved in my academic career.

- To Karsten Urban and Jay Gopalakrishnan for the time they dedicated to read and evaluate this thesis despite their busy schedules.

- To my PhD colleagues at the University of Pavia, for the interesting conversations, exchange of ideas, and relaxing hours we spent together.

- To Kai Hormann, for receiving me at Università della Svizzera Italiana during my Erasmus traineeship there, and exposing me to the world of Generalized Barycentric Coordinates.

- To the Erwin Schrödinger International Institute for Mathematics and Physics, for hosting me in Vienna during half a year as a Junior Research Fellow.

- To the Department of Mathematics of the University of Pavia, for the opportunity of carrying out my doctoral studies there.

- To my parents, for all their love and sacrifices, without which it would have been impossible to get where I am.

- To my wife, that is the main source of my inspiration and motivation. For all her support, love, and patience, specially during the time we were far from each other.

During my PhD I have also had the opportunity of meeting some friends from other institutions: Matteo Ferrari, Paul Stocker, Sara Fraschini, and many other nice people.

iv

Abstract

Even though the foundation of space–time Galerkin methods dates back to the 70s of the last century [57], and a few contributions were developed in the twenty years to follow [5, 6, 31, 53, 34], in the last two decades there has been a growing attention on this topic.

In this dissertation, space–time Galerkin methods are designed for the discretization of the Schrödinger and heat equations.

For the Schrödinger equation, we design an ultra-weak space–time discontinuous Galerkin variational formulation on general prismatic meshes. The method is well-posed, stable and quasi-optimal for very general discrete spaces. The approximation properties of the method are studied for four choices of discrete spaces: *i)* the polynomial Trefftz space for problems with zero potential; *ii)* a pseudo-plane wave Trefftz space for problems with piecewise-constant potentials; *iii)* the full polynomial space, and *iv)* a quasi-Trefftz polynomial space for piecewise-smooth potentials.

For the heat equation, we design a nonconforming space–time virtual element method, which is also the first space–time virtual element discretization of a time-dependent PDE in the literature. The method is proven to be well-posed, based on a discrete inf-sup condition. An *a priori* error analysis show that optimal h-convergence rates are obtained for sufficiently smooth solutions. The method performs well also for singular solutions such as those arising from the incompatibility of initial and boundary conditions. Since the method allows for nonmatching space-like and time-like facets, it is naturally suitable for h- and hp-versions.

Contents

List of figures

List of tables

Some useful notation

We denote the first partial derivative with respect to the time variable t by ∂_t, and the spatial gradient and Laplacian operators by $\nabla_{\mathbf{x}}$, $\Delta_{\mathbf{x}}$, respectively.

Standard notation for Sobolev spaces is employed, see e.g., [78]. For a given domain $D \subset \mathbb{R}^d$, $d \in \mathbb{N}$, $H^s(D)$ represents the standard Sobolev space of order $s \in \mathbb{N}$ endowed with the standard inner product $(\cdot, \cdot)_{H^s(D)}$, the seminorm $|\cdot|_{H^s(D)}$, and the norm $\|\cdot\|_{H^s(D)}$. In particular, we let $H^0(D)$ be the space $L^2(D)$ of Lebesgue square integrable functions over D and $H_0^1(D)$ be the subspace of functions in $H^1(D)$ with zero trace on ∂D. If s is a non-integer or negative number, then the Sobolev space $H^s(D)$ is defined by means of interpolation and duality, respectively. We denote the duality product between $H^{-1}(D)$ and $H_0^1(D)$ by $\langle \cdot, \cdot \rangle$. The Sobolev spaces on ∂D are defined analogously and denoted by $H^s(\partial D)$.

Given s in $\mathbb{R}$, a time interval (a, b), and a Banach space $(Z, \|\cdot\|_Z)$, we introduce the Bochner space $H^s(a, b; Z)$.

We denote the space of polynomials in d variables of degree at most p on a domain $D \subset \mathbb{R}^d$ by $\mathbb{P}^p(D)$, with dimension $r_{d,p} = \binom{p+d}{d}$.

We use the standard multi-index notation for partial derivatives and monomials, adapted to the space–time setting: for $\boldsymbol{j} = (\boldsymbol{j}_{\mathbf{x}}, j_t) = (j_{x_1}, \ldots, j_{x_d}, j_t) \in \mathbb{N}^{d+1}$, and $\mathbf{z} = (\mathbf{z}_{\mathbf{x}}, z_t) = (z_{x_1}, \ldots, z_{x_d}, z_t) \in \mathbb{N}^{d+1}$,

$$\boldsymbol{j}! := j_{x_1}! \cdots j_{x_d}! j_t!, \qquad |\boldsymbol{j}| := |\boldsymbol{j}_{\mathbf{x}}| + j_t := j_{x_1} + \cdots + j_{x_d} + j_t,$$

$$\binom{\boldsymbol{j}}{\mathbf{z}} := \frac{\boldsymbol{j}!}{\mathbf{z}!(\boldsymbol{j} - \mathbf{z})!}, \qquad \boldsymbol{j} \le \mathbf{z} \Leftrightarrow j_{x_i} \le z_{x_i} \ (1 \le i \le d) \text{ and } j_t \le z_t,$$

$$D^{\boldsymbol{j}} f := \partial_{x_1}^{j_{x_1}} \cdots \partial_{x_d}^{j_{x_d}} \partial_t^{j_t} f, \qquad \mathbf{x}^{\boldsymbol{j}_{\mathbf{x}}} t^{j_t} := x_1^{j_{x_1}} \cdots x_d^{j_{x_d}} t^{j_t}.$$

We write $a \lesssim b$ meaning that there exists a positive constant c independent of the meshsize, such that $a \le cb$. We also write $a \simeq b$ if $a \lesssim b$ and $b \lesssim a$ at once.

List of publications

The results presented in this dissertation are based on the following peer-reviewed papers and preprints [41, 42, 43, 44, 45]

Paper/preprint	Corresponding chapter(s)
[44] Gómez, S. and Moiola, A. (2023). Space–time DG method for the Schrödinger equation with variable potential. arXiv:2306.05780.	Chapters 2 and 5
[45] Gómez, S., Moiola, A., Perugia, I., and Stocker, P. (2023). On polynomial Trefftz spaces for the linear time-dependent Schrödinger equation. Appl. Math. Lett. 146(C):108824.	Chapter 3
[43] Gómez, S. and Moiola, A. (2022). A space–time Trefftz discontinuous Galerkin method for the linear Schrödinger equation. SIAM J. Num. Anal., 60(2):688–714.	Chapter 4
[41] Gómez, S., Mascotto, L., Moiola, A., and Perugia, I. (2023). Space–time virtual elements for the heat equation. To appear in SIAM J. Numer. Anal.	Chapters 6 and 7
[42] Gómez, S., Mascotto, L., and Perugia, I. (2024). Design and performance of a space–time virtual element method for the heat equation on prismatic meshes. Comput. Methods Appl. Mech. Eng., 418(A):116491.	Chapters 8 and 9

Chapter 1

Introduction

Contents

We introduce the general classes of problems and methods that will be analysed in the rest of this dissertation.

1.1 The time-dependent Schrödinger equation

The first part of this dissertation concerns the approximation of the solution to the time-dependent Schrödinger equation on a space–time cylinder $Q_T = \Omega \times I$, where Ω is an open, bounded polytopic domain in $\mathbb{R}^d$ ($d \in \mathbb{N}$) with Lipschitz boundary $\partial\Omega$, and $I = (0, T)$ for some final time $T > 0$:

$$\mathcal{S}_\epsilon \psi := i\epsilon \partial_t \psi + \frac{\epsilon^2}{2} \Delta_\mathbf{x} \psi - V\psi = 0 \qquad \text{in } Q_T, \tag{1.1a}$$

$$\psi = g_\mathrm{D} \qquad \text{on } \Gamma_\mathrm{D} \times I, \tag{1.1b}$$

$$\partial_{\mathbf{n_x}} \psi = g_\mathrm{N} \qquad \text{on } \Gamma_\mathrm{N} \times I, \tag{1.1c}$$

$$\partial_{\mathbf{n_x}} \psi - i\vartheta\psi = g_\mathrm{R} \qquad \text{on } \Gamma_\mathrm{R} \times I, \tag{1.1d}$$

$$\psi(\mathbf{x}, 0) = \psi_0(\mathbf{x}) \qquad \text{on } \Omega. \tag{1.1e}$$

Here i is the imaginary unit; ϵ is a positive parameter; $\partial_{\mathbf{n_x}}(\cdot)$ is the normal derivative-in-space operator; $V : Q_T \to \mathbb{R}$ is the potential energy function; $\vartheta \in L^\infty(\Gamma_\mathrm{R} \times I)$ is a positive "impedance" function; the Dirichlet (g_D), Neumann (g_N), Robin (g_R) and initial condition (ψ_0) data are given functions; $\Gamma_\mathrm{D}, \Gamma_\mathrm{N}, \Gamma_\mathrm{R}$ are a polytopic partition of $\partial\Omega$. The standard dimensionless Schrödinger equation corresponds to $\epsilon = 1$. Problem 1.1 is well-posed if, e.g., $\psi_0 \in H_0^1(\Omega)$, $\Gamma_\mathrm{R} = \Gamma_\mathrm{N} = \emptyset$, and $g_\mathrm{D} = 0$, by [74, Ch. 3, Thm. 10.1, Rem. 10.2]; in this case $\psi \in C^0(0, T; H_0^1(\Omega)) \cap C^1(0, T; H^{-1}(\Omega))$.

The model problem (1.1) has a wide range of applications. In quantum physics [63], the solution ψ is a quantum-mechanical wave function determining the dynamics of one or multiple particles in a potential V. In electromagnetic wave propagation [73], it is called "paraxial

wave equation" and ψ is a function associated with the field component in a two-dimensional electromagnetic problem where the energy propagates at small angles from a preferred direction. In such problems, the function V depends on the refractive index and the wave number. In underwater sound propagation [61], it is referred to as "parabolic equation" and ψ describes a time harmonic wave propagating primarily in one direction. In molecular dynamics [10], by neglecting the motion of the atomic nuclei, the Born-Oppenheimer approximation leads to a Schrödinger equation in the *semi-classical* regime ($0 < \epsilon \ll 1$), where ϵ^2 is proportional to the mass ratio of electrons and nuclei, and the solution ψ describes the dynamics of the electrons in the molecules. The approximation of the solution to the Schrödinger equation in the semi-classical regime is very challenging from the computational point of view, due to the highly oscillatory behaviour of its solution, see the monographs [58, 71]. Problems of such type require a very delicate treatment, which is beyond the scope of this dissertation, but it is an interesting topic of research to be addressed in the future.

1.2 Heat equation

The second part of this dissertation concerns the approximation of the solution to the heat equation on a space–time cylinder $Q_T := \Omega \times (0, T)$, where $\Omega \subset \mathbb{R}^d$ ($d = 1,\ 2,\ 3$) is a bounded Lipschitz space domain and $T > 0$,

$$
\begin{aligned}
c_H \partial_t u - \lambda \Delta_{\mathbf{x}} u &= f && \text{in } Q_T, \\
u &= g && \text{on } \partial\Omega \times (0, T), \\
u(\mathbf{x}, 0) &= u_0(\mathbf{x}) && \text{on } \Omega.
\end{aligned}
\tag{1.2}
$$

Here, $c_H > 0$ and $\lambda > 0$ are given positive constant volumetric heat capacity and thermal conductivity, respectively; the source term $f : Q_T \to \mathbb{R}$, the initial condition $u_0 : \Omega \times \{0\} \to \mathbb{R}$, and the Dirichlet boundary condition $g : \partial\Omega \times (0, T) \to \mathbb{R}$ are given functions.

Introduce the functional spaces

$$
Y := L^2\left(0, T; H_0^1(\Omega)\right), \quad X := \left\{ v \in Y \cap H^1\left(0, T; H^{-1}(\Omega)\right) \mid v = 0 \text{ in } \Omega \times \{0\} \right\},
\tag{1.3}
$$

endowed with the norms

$$
\|v\|_Y^2 := \left\| \lambda^{\frac{1}{2}} \nabla_{\mathbf{x}} v \right\|_{L^2(Q_T)^d}^2, \qquad \|v\|_X^2 := \|c_H \partial_t v\|_{L^2(0,T;H^{-1}(\Omega))}^2 + \|v\|_Y^2,
$$

respectively. We have used the following definition:

$$
\text{for any } \phi \text{ in } L^2(0, T; H^{-1}(\Omega)), \qquad \|\phi\|_{L^2(0,T;H^{-1}(\Omega))} := \sup_{0 \neq v \in Y} \frac{\int_0^T \langle \phi, v \rangle \, \mathrm{dt}}{\|v\|_Y},
$$

where $\langle \cdot, \cdot \rangle$ denotes the duality between $H_0^1(\Omega)$ and $H^{-1}(\Omega)$.

Next, we define the space–time bilinear form $b(\cdot, \cdot) : X \times Y \to \mathbb{R}$ as

$$
b(u, v) := \int_0^T \left(\langle c_H \partial_t u, v \rangle + \int_\Omega \lambda \nabla_{\mathbf{x}} u \cdot \nabla_{\mathbf{x}} v \, \mathrm{d}\mathbf{x} \right) \mathrm{dt}.
\tag{1.4}
$$

Several space–time discretizations of the heat equation (1.2) are based on the following standard Petrov-Galerkin weak formulation for homogeneous initial and Dirichlet boundary conditions (see [19, Ch. XVIII, §4.1]):

$$\begin{cases} \text{find } u \in X \text{ such that} \\ b(u, v) = \int_0^T \langle f, v \rangle \mathrm{dt} \qquad \forall v \in Y. \end{cases} \tag{1.5}$$

Problem (1.5) is well posed; see, e.g., [93, Corollary 2.3].

Remark 1.1 (Inhomogeneous initial and boundary conditions). Given (f, u_0) in $L^2(0, T; H^{-1}(\Omega)) \times L^2(\Omega)$, consider the following problem: find $u \in Y \cap H^1(0, T; H^{-1}(\Omega))$ such that

$$\begin{cases} \int_0^T \left(\langle c_H \partial_t u, v \rangle + \int_\Omega \nu \nabla_{\mathbf{x}} u \cdot \nabla_{\mathbf{x}} v \, \mathrm{d}\mathbf{x} \right) \mathrm{dt} = \int_0^T \langle f, v \rangle \mathrm{dt} & \forall v \in L^2(0, T; H^{-1}(\Omega)) \\ \int_\Omega u(\cdot, 0) w \, dx = \int_\Omega u_0 w \, dx & \forall w \in L^2(\Omega). \end{cases} \tag{1.6}$$

The well-posedness of problem (1.6) is discussed, e.g., in [92, §5].

The case of inhomogeneous Dirichlet boundary conditions $u = g$ on $\partial\Omega \times (0, T)$ can be dealt with assuming g in $H^1(0, T; H^{1/2}(\partial\Omega))$. Denote by $G : Q_T \to \mathbb{R}$ the solution to the family of elliptic problems $-\nu\Delta_{\mathbf{x}} G(\cdot, t) = 0$ in Ω with $G(\cdot, t) = g(\cdot, t)$ on $\partial\Omega$ for all $0 \leq t \leq T$. The function G belongs to $H^1(0, T; H^1(\Omega))$, since $\partial_t G$ solves a similar family of elliptic problems with boundary data $\partial_t g$ in $L^2(0, T; H^{1/2}(\partial\Omega))$).

For the case of inhomogeneous initial and boundary conditions, denote by w the solution to problem (1.6) with source term $f - c_H \partial_t G$ and initial condition $u_0 - g(\cdot, 0)$. Then, $u = G + w$ solves the inhomogeneous initial-boundary value problem with data (f, u_0, g). In particular, u belongs to $L^2(0, T; H^1(\Omega)) \cap H^1(0, T; H^{-1}(\Omega))$. ∎

1.3 Space–time methods for evolution PDEs

Space–time Galerkin methods discretize all the variables in a time-dependent PDE at once; this is in contrast with the method of lines, which combines a spatial discretization and a time-stepping scheme. Space–time methods have attracted a lot of attention in the last decades, also due to availability of modern computer resources. For recent surveys of space–time discretizations of time-dependent PDEs, we refer to [69, 70, 83].

We list some of the main features and advantages of space–time methods.

- Space–time methods can achieve high order convergence rates in space and time.

- They provide discrete solutions that are available on the whole space–time domain, see Figure 1.1.

- They may allow for local refinements/coarsening of the space–time mesh, see Figure 1.2.

- They may be used to solve problems on evolving domains.

- They can be combined with space–time parallel solvers.

(a) Finite differences in space + time-stepping

(b) Finite elements in space + time-stepping

(c) Space-time methods
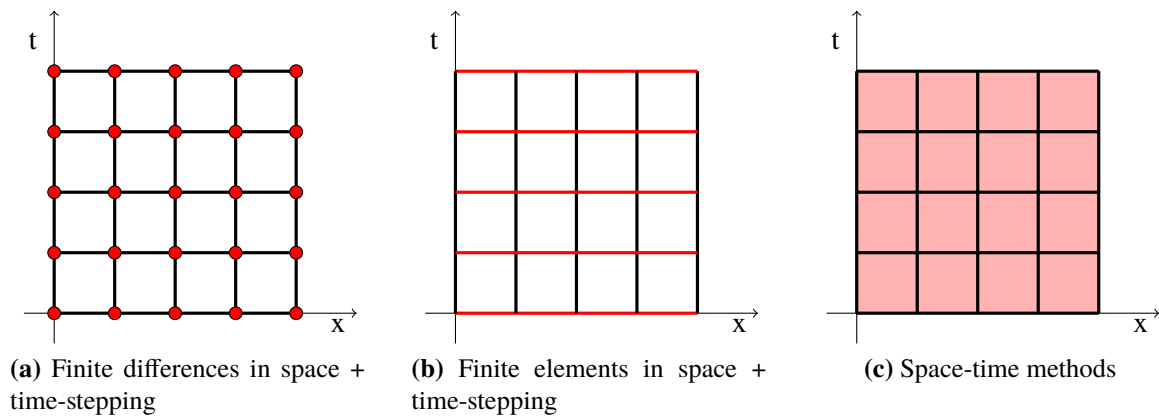
Figure 1.1 Schematic representation of the approximations provided by the method of lines and space–time methods. The **red** color represents the domain of the discrete solution of each method: (a) grid nodes on each time-step; (b) spatial mesh on each time-step; (c) the whole space–time domain.

(a) Tensor-product mesh

(b) Local time-stepping

(c) Local space–time refinements
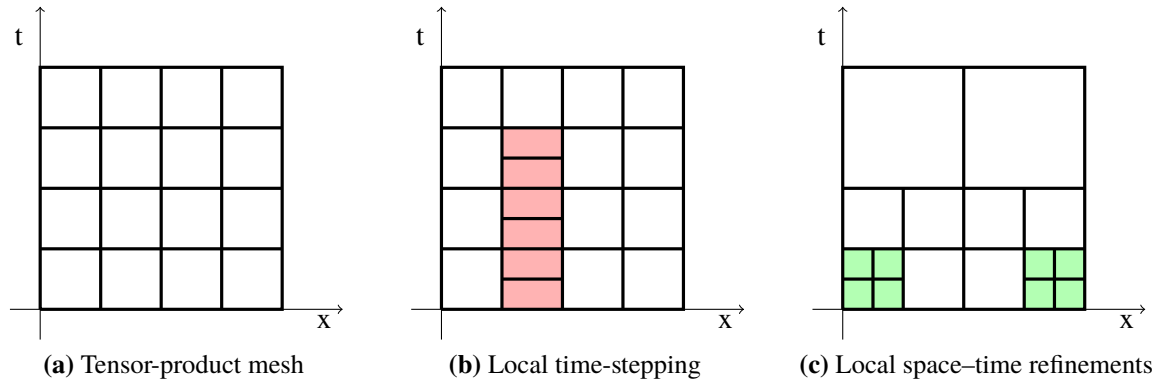
Figure 1.2 Examples of $(1 + 1)$-dimensional space–time meshes.

1.4 Trefftz discontinuous Galerkin methods

Trefftz methods are Galerkin discretizations characterized by using test and trial approximation spaces that lie in the kernel of the target differential operator, i.e., spaces spanned by functions that are local solutions to the considered PDE. Trefftz schemes are mainly motivated by their significant reduction in the computational cost and the number of degrees of freedom with respect to traditional polynomial approximations, and by their effectiveness in dealing with the intrinsic highly oscillatory behavior in the solution of certain problems. Trefftz methods with lower-dimensional spaces than standard finite element spaces, but similar approximation properties, have been designed for many problems, e.g., Laplace and solid-mechanics problems [87]; the Helmholtz equation [52]; the time-harmonic [51], and time-dependent [28] Maxwell's equations; the acoustic wave equation in second-order [7] and first-order [80] form; the Schrödinger equation [43]; among others.

Discontinuous Galerkin (DG) methods [24] do not impose continuity on the approximation in a strong sense, so they are specially suitable to be combined with Trefftz bases, which are naturally discontinuous.

In many cases, well-posedness and quasi-optimality are proven for ultra-weak DG variational formulations on very general discrete Trefftz spaces, as the analysis for such formulations do not require the use of inverse estimates, see e.g., [40]. In fact, showing inverse estimates for non-polynomial Trefftz spaces is a very difficult task. On the other hand, polynomial Trefftz spaces can be easily combined with any DG variational formulation. *A priori* error estimates depend on the specific choice of the discrete Trefftz space, which is desired to possess good approximation properties.

1.5   Virtual element method

The virtual element method (VEM) was introduced in [9] as an extension of the finite element method to general polytopic meshes for the approximation of solutions to the Poisson equation. Trial and test spaces consist of functions that are solutions to local problems related to the PDE problem to be approximated. Moreover, they typically contain polynomials of a given maximum degree, together with nonpolynomial functions allowing for the enforcement of the desired type of conformity in the global spaces. Such functions are not required to be explicitly known. Suitable sets of degrees of freedom (DoFs) are chosen so that projections from local virtual element spaces onto polynomial spaces can be computed out of them. Such polynomial projectors and certain stabilizing bilinear forms are used to define the discrete bilinear forms. A nonconforming version of the VEM for the Poisson equation was proposed in [3]. Unlike its conforming counterpart, the nonconforming VEM can be presented in a unified framework for any dimension, which significantly simplifies its analysis and implementation.

1.6   General outline of the dissertation

In the present dissertation, we propose and analyze some novel space–time methods for the discretization of the time-dependent Schrödinger equation and the heat equation.

A space–time ultra-weak DG variational formulation for the Schödinger equation (1.1) is introduced in Chapter 2. The method is proven to be well-posed, stable and quasi-optimal for very general discrete spaces. We study the approximation properties of different discrete spaces: the polynomial Trefftz space for problems with zero potential $V = 0$ in Chapter 3; a Trefftz space consisting of complex pseudo-plane wave functions for piecewise-constant potentials in Chapter 4; the full polynomial space and a quasi-Trefftz subspace for piecewise-smooth potentials in Chapter 5.

A nonconforming space–time virtual element method for the discretization of the heat equation (6.19) and its well-posedness is presented in Chapter 6. An *a priori* error analysis of the method is carried out in Chapter 7, where optimal convergence rates are proven for sufficiently smooth solutions. An extension of the method to general prismatic mehes and variable degrees of accuracy is presented in Chapter 8. In Chapter 9, we discuss several numerical aspects of the method: some flagging strategies to speed up the assembling and solution of the linear system stemming from the proposed method; the hp-convergence of the method for singular solutions; an h-adaptive procedure driven by a residual-type error indicator.

Part I: Time-dependent Schrödinger equation

Chapter 2 We propose a space–time ultra-weak DG variational formulation for the Schrödinger equation on very general prismatic space–time meshes. By choosing upwind numerical fluxes on the space-like facets, the method allows for the decomposition of the global linear system into a sequence of smaller linear systems on time slabs. Relying on the self-adjointness property of the Schrödinger operator $\mathcal{S}_\epsilon(\cdot)$, a consistent volume penalty parameter is added order to guarantee the coercivity of the sesquilinear form of the method for very general discrete spaces. We show that *a priori h*-convergence estimates in a mesh-dependent norm can be derived by simply studying the approximation properties of the local discrete spaces, which will be studied in the next chapters. We refer to Table 2.1 for an overview of the main features of each discrete space.

Chapter 3 We analyze the approximation properties of the polynomial Trefftz space for the Schrödinger equation with zero potential ($V = 0$). To the best of our knowledge, this is the first study of a polynomial Trefftz-DG space for a PDE with derivatives of different orders. We show that optimal *h*-convergence is obtained if Trefftz polynomials of degree $2p$ in $\mathbb{R}^{d+1}$ are used as discrete spaces, by proving that such a space contains some extended Taylor polynomials of any sufficiently regular solution to the Schrödinger equation (1.1a). We provide a practical way to construct a basis for the space on any space dimension $d \in \mathbb{N}$, and prove that its dimension is equal to that of the space of polynomials of degree $2p$ in $\mathbb{R}^d$.

Chapter 4 We introduce a pseudo-plane wave Trefftz space for the Schrödinger equation (1.1) with piecewise-constant potential V. The definition of the basis functions depends on the choice of some real parameters $\{k_\ell\}$, and some unitary directions $\{\mathbf{d}_\ell\} \subset \mathbb{R}^d$. Based on the ideas presented by O. Cessenat and B. Després in [15], we show that optimal h-convergence can be obtained, if for any sufficiently regular solution ψ to the Schrödinger equation (1.1a), the local Trefftz space contains an element with the same degree-p Taylor polynomial as that of ψ. For $d = 1$ and $d = 2$, we provide some sufficient conditions on the parameters $\{k_\ell\}$ and the directions $\{\mathbf{d}_\ell\}$ that guarantee that such a condition is satisfied for the pseudo-plane wave Trefftz space.

Chapter 5 In this chapter, we consider the Schrödinger equation (1.1) with piecewise-smooth potential V. On each element K of the space–time mesh, we construct a quasi-Trefftz polynomial space that is tailored to contain the degree-p Taylor polynomial centered at some point $(\mathbf{x}_K, t_K) \in K$ of any sufficiently regular solution to the Schrödinger equation (1.1a). We show optimal h-convergence estimates for the space of polynomials of degree p in $\mathbb{R}^{d+1}$ dimensions, and for the quasi-Trefftz space.

<center>Part II: Heat equation</center>

Chapter 6 We introduce a space–time virtual element method for the discretization of the heat equation (1.2) on tensor-product-in-time meshes. We design local virtual element spaces consisting of functions that are solutions to local heat problems with polynomial data, and find a set of unisolvent degrees of freedom. We introduce some projections onto the space of polynomials that are computable for all the functions in the local virtual element space via the aforementioned set of degrees of freedom. We construct global virtual element spaces that are nonconforming across time-like facets, and discontinuous across space-like facets. Continuity in time is weakly imposed through upwind fluxes that are defined in terms of the traces of a

<center>6</center>

polynomial projection. We prove the well-posedness of the method, relying on a discrete inf-sup condition.

Chapter 7 We present an *a priori* error analysis for the method introduced in Chapter 6. We show some identities related to the nonconformity across time-like facets and the polynomial inconsistency of the method. A Strang-type error bound is proven by using the discrete inf-sup condition. An optimal *h*-convergence error estimate is then obtained by estimating each term of the error bound, namely, *i)* a virtual element interpolation error, *ii)* the variational crime on the computation of the right-hand side, *iii)* the nonconformity across time-like facets, and *iv)* polynomial error estimates. We present some numerical experiments that show optimal *h*-convergence of the method for smooth and singular solutions, and exponential convergence of the *p*-version for smooth solutions.

Chapter 8 We extend the virtual element method (6.24) in Chapter 6 to the case of general prismatic space–time meshes and variable degrees of accuracy. The definitions of the local virtual element spaces, the polynomial projections, the discrete bilinear forms, and the discrete norms are adapted to such a type of meshes. For the case of variable degrees, a maximum strategy is described to fix the degrees of freedom associated with each element of the prismatic space–time mesh.

Chapter 9 We discuss some numerical aspects of an efficient implementation of the virtual element method (8.10). More precisely, we present some element flagging strategies to handle efficiently the data structure of the space–time prismatic meshes, namely, a time-slab flagging and an element-topology flagging. The former allows us to identify a time-slab decomposition of the space–time mesh, which is then used to decompose the global linear system as a sequence of smaller problems associated with each time-slab. The latter allows for a reduction in the computational cost of assembling the linear system, by identifying elements with the same topology and using a set of reference elements to compute the projections and evaluate the discrete operators. We numerically assess the convergence of the *hp*-version of the method for singular solutions on geometrically refined space–time meshes, which shows the expected exponential convergence in terms of suitable roots of the total number of degrees of freedom. We also present an *h*-adaptive procedure driven by a residual-type error indicator, and compare its results with a continuous finite element discretization of (1.2).

1.7 Open problems and future work

We list here some possible future research directions.

Time-dependent Schrödinger equation:

- The proof of optimal error estimates in mesh-independent norms such as $L^2(Q_T)$ (cf. the duality approach used in [80, §5.4] for the wave equation).
- The *p*-convergence analysis for the quasi-Trefftz and Trefftz versions of the method, i.e. the proof of convergence rates for sequences of discrete spaces obtained by local enrichment on a fixed mesh. Only the best-approximation bounds are missing, as the quasi-optimality bound (2.8) is independent of the discrete space. This is a very challenging task, which was accomplished for the Helmholtz equation [79] but not yet for the wave equation.

- To carry out a careful analysis for the proposed method applied to the Schrödinger equation (1.1) in the semiclassical regime ($0 < \epsilon \ll 1$).

- The "sparsification" of the scheme, i.e. the combination with sparse-grid techniques in space–time to improve its efficiency, e.g. along the lines of [8].

- The extension of the h-convergence estimates for the pseudo-plane wave Trefftz spaces introduced in Chapter 4 to space dimensions higher than 2.

- A more accurate analysis of the pseudo-plane wave Trefftz spaces in order to optimize the choice of the parameters (k_ℓ and $\mathbf{d}_\ell$ in (4.1)), improving conditioning, robustness and accuracy.

- The analysis of the quasi-Trefftz version of the method with Sobolev regularity assumptions for the exact solution.

- The extension to initial boundary value problems with non-reflecting boundary conditions, which are often used to truncate unbounded domains (e.g. [2]). Under this respect the Trefftz approach is promising as it allows for the selection of outward-propagating basis functions on boundary cells, as in [27].

Heat equation:

- To prove explicit bounds for the stabilization term in the proposed virtual element method. This is an important aspect, as the stabilization term in the method play a key role in the conditioning of the linear system and the accuracy of the method, see the recent survey [75].

- To derive error estimates in stronger norms such as in the $L^2(Q_T)$ norm or $H^1(0, T; L^2(\Omega))$-type norms as it has been done in [14] for a DG method.

- The analysis of the hp version of the method. In fact, polytopic methods allow for extremely flexible geometries when refining towards the corners of the domain, and for an easy handling of refinement-coarsening strategies in adaptive algorithms.

- To carry out an error analysis under lower-regularity conditions on the exact solution and the data of the problem. This would explain the optimal convergence rates obtained also for singular solutions.

- To carry out an *a posteriori* error analysis of the method, based on an efficient and effective error estimator.

- The design of a Trefftz variant of the method, as those designed for the Poisson equation [76] and the Helmholtz equation [77].

- The design of space–time virtual element methods for other time-dependent PDEs (e.g., the Schrödinger equation, the wave equation, Maxwell's equations).

Part I

Time-dependent Schrödinger equation

Chapter 2

Space–time discontinuous Galerkin method for the Schrödinger equation

Contents

2.1 Introduction

The literature on space–time Galerkin methods for the Schrödinger equation is very scarce. In fact, the standard Petrov-Galerkin formulation for the Schrödinger equation, i.e., the analogous formulation to that proposed in [93] for the heat equation, is not inf-sup stable, see [49, §2.2]. In [59], Karakashian and Makridakis proposed a space–time method for the Schrödinger equation with nonlinear potential, combining a conforming Galerkin discretization in space and an upwind DG time-stepping. This method reduces to a Radau-IIA Runge-Kutta time discretization in the case of constant potentials. Moreover, under some restrictions on the mesh, that are necessary to preserve the accuracy of the method, it allows for changing the spatial mesh on each time-slab, but not for local time-stepping. A second version of the method, obtained by enforcing the transmission of information from the past through a projection, was proposed in [60]. This version reduces to a Legendre Runge-Kutta time discretization in the case of constant potentials. Recently, some space–time methods based on ultra-weak formulations of the Schrödinger equation have been designed. The well-posedness of such formulations requires weaker assumptions on the mesh. In [21], Demkowicz et al., proposed a discontinuous Petrov-Galerkin (DPG) formulation for the linear Schrödinger equation. The method is a conforming discretization of an ultra-weak formulation of the Schrödinger equation in graph spaces. Well-posedness and quasi-optimality of the method follow directly from the inf-sup stability (in a graph norm) of the continuous Petrov-Galerkin formulation. In [49], Hain and Urban proposed a space–time ultra-weak variational formulation for the Schrödinger equation with optimal inf-sup constant. The formulation in [49] is closely related to the DPG method in [21], but differs in the choice of the test and trial spaces. While in the latter one first fixes a trial space and then construct a suitable test space, the former requires the choice of a conforming test space and then the trial space is defined accordingly. We are not aware of publications proposing space–time DG methods for the Schrödinger equation other than [21, 49, 59, 60] and [43].

In this chapter, we introduce a space–time ultra-weak DG method for the discretization of the Schrödinger equation. The main advantages of the method are the following:

- The proposed ultra-weak DG variational formulation of (1.1) is well-posed, stable, and quasi-optimal in any space dimension for an almost arbitrary choice of piecewise-defined discrete spaces and variable potentials.

- *A priori* error estimates in a mesh-dependent norm can be obtained by simply analyzing the approximation properties of the local spaces.

- The method naturally allows for non-matching space-like and time-like facets and all our theoretical results hold under standard assumptions on the space–time mesh, which make the method suitable for adaptive versions and local time-stepping.

In Table 2.1, we summarize the features of the method with four different choices of discrete spaces that will be discussed in the next chapters:

- the polynomial Trefftz space for problems with zero potential in Chapter 3;

- a non-polynomial Trefftz space of complex pseudo-plane wave functions for problems with piecewise-constant potentials in Chapter 4;

- the full polynomial space and a quasi-Trefftz polynomial subspace for problems with piecewise-smooth potentials in Chapter 5.

Discrete space	Potential V	Local Dim.	Reg. Sol.	Conditioning
Polynomial Trefftz space $\mathbb{P}_{\mathbf{T}}^{2p}(K)$ in (3.3)	$= 0$	$\binom{d+2p}{2p} = O((2p)^d)$	$\psi \in H^{p+1}(K_t; H^{2p}(K_{\mathbf{x}}))$	$O(h^{-1})$
Pseudo-plane wave Trefftz space $\mathbb{T}^p(K)$ in (4.1)	Piecewise-constant	$\binom{d+p-1}{d}\frac{2p+d}{p} = O(p^d)$	$\psi \in H^{p+1}(K)$	$O(h^{-(2p+1)})$
Full polynomial space $\mathbb{P}^p(K)$	Piecewise-smooth	$\binom{d+p+1}{p} = O(p^{d+1})$	$\psi \in H^{p+1}(K)$	$O(h^{-1})$
Quasi-Trefftz polynomial space $\mathbb{Q}\mathbb{T}^p(K)$ in (5.2)	Piecewise-smooth	$\binom{d+p-1}{d}\frac{2p+d}{p} = O(p^d)$	$\psi \in C^{p+1}(K)$	$O(h^{-1})$

Table 2.1 Comparison of theoretical and numerical properties of different discrete spaces for the variational formulation (2.3). **Second column:** potentials V that are allowed. **Third column:** dimension of the local discrete space on a prismatic element $K = K_{\mathbf{x}} \times K_t$ of the space–time mesh $\mathcal{T}_h$ that guarantees convergence of order $O(h^p)$ for the error of the method in a mesh-dependent norm. **Fourth column:** Local regularity required to obtain optimal convergence. **Fifth column:** Behaviour of the condition number $\kappa_2(\cdot)$ of the stiffness matrix numerically observed in Section 5.5.5.

From Table 2.1, we conclude that the quasi-Trefftz and full polynomial versions of the method are the most robust, as they allow for more general problems and produce better conditioned matrices. Among them, the quasi-Trefftz version of the method requires much fewer degrees of freedom ($O(p^d)$ per element), than the full polynomial version which requires $O(p^{d+1})$ degrees of freedom per element, and achieves the same convergence rates under slightly stronger local regularity assumptions of the exact solution.

Structure of the chapter: In Section 2.2, we introduce some notation on the space–time meshes that we use. In Section 2.3, we propose an ultra-weak DG variational formulation on abstract spaces. Section 2.4 is devoted to the analysis of well-posedness, stability and quasi-optimality of the method. In Section 2.5 we introduce some preliminary results that will be used to derive *a priori* error estimates in the subsequent chapters.

2.2 Space–time mesh and DG notation

Let $\mathcal{T}_h$ be a non-overlapping prismatic partition of Q_T, i.e., each element $K \in \mathcal{T}_h$ can be written as $K = K_{\mathbf{x}} \times K_t$ for a d-dimensional polytope $K_{\mathbf{x}} \subset \Omega$ and a time interval $K_t \subset I$. We use the notation $h_{K_{\mathbf{x}}} = \mathrm{diam}(K_{\mathbf{x}})$, $h_{K_t} = |K_t|$ and $h_K = \mathrm{diam}(K) = (h_{K_{\mathbf{x}}}^2 + h_{K_t}^2)^{1/2}$. We call "mesh facet" any intersection $F = \partial K_1 \cap \partial K_2$ or $F = \partial K_1 \cap \partial Q_T$, for $K_1, K_2 \in \mathcal{T}_h$, that has positive d-dimensional measure and is contained in a d-dimensional hyperplane. We denote by $\vec{\mathbf{n}}_F = (\vec{\mathbf{n}}_F^{\mathbf{x}}, n_F^t) \in \mathbb{R}^{d+1}$ one of the two unit normal vectors orthogonal to F with either $n_F^t = 0$ or $n_F^t = 1$. We assume that each internal mesh facet F is either

$$\text{a space-like facet} \quad \text{if } \vec{\mathbf{n}}_F^{\mathbf{x}} = 0, \quad \text{or} \quad \text{a time-like facet} \quad \text{if } n_F^t = 0.$$

We further denote the mesh skeleton and its parts as

$$\mathcal{F}_h := \bigcup_{K \in \mathcal{T}_h} \partial K, \qquad \mathcal{F}_h^0 := \Omega \times \{0\}, \quad \mathcal{F}_h^T := \Omega \times \{T\},$$

$$\mathcal{F}_h^{\mathrm{D}} := \Gamma_{\mathrm{D}} \times (0, T), \quad \mathcal{F}_h^{\mathrm{N}} := \Gamma_{\mathrm{N}} \times (0, T), \quad \mathcal{F}_h^{\mathrm{R}} := \Gamma_{\mathrm{R}} \times (0, T),$$

$$\mathcal{F}_h^{\mathrm{time}} := \text{the union of all the internal time-like facets,}$$

$$\mathcal{F}_h^{\mathrm{space}} := \text{the union of all the internal space-like facets.}$$

We employ the standard DG notation for the averages $\{\!\!\{\cdot\}\!\!\}$ and space $[\![\cdot]\!]_{\mathbf{N}}$ and time $[\![\cdot]\!]_t$ jumps for piecewise complex scalar w and vector $\boldsymbol{\tau}$ fields:

$$\begin{cases} \{\!\!\{w\}\!\!\} := \frac{1}{2}\left(w|_{K_1} + w|_{K_2}\right) \\ \{\!\!\{\boldsymbol{\tau}\}\!\!\} := \frac{1}{2}\left(\boldsymbol{\tau}|_{K_1} + \boldsymbol{\tau}|_{K_2}\right) \end{cases} \qquad \text{on } \partial K_1 \cap \partial K_2 \subset \mathcal{F}_h^{\mathrm{time}},$$

$$\begin{cases} [\![w]\!]_{\mathbf{N}} := w|_{K_1}\vec{\mathbf{n}}_{K_1}^{\mathbf{x}} + w|_{K_2}\vec{\mathbf{n}}_{K_2}^{\mathbf{x}} \\ [\![\boldsymbol{\tau}]\!]_{\mathbf{N}} := \boldsymbol{\tau}|_{K_1} \cdot \vec{\mathbf{n}}_{K_1}^{\mathbf{x}} + \boldsymbol{\tau}|_{K_2} \cdot \vec{\mathbf{n}}_{K_2}^{\mathbf{x}} \end{cases} \qquad \text{on } \partial K_1 \cap \partial K_2 \subset \mathcal{F}_h^{\mathrm{time}},$$

$$[\![w]\!]_t := w|_{K_1}n_{K_1}^t + w|_{K_2}n_{K_2}^t = w^- - w^+ \qquad \text{on } \partial K_1 \cap \partial K_2 \subset \mathcal{F}_h^{\mathrm{space}},$$

where $\vec{\mathbf{n}}_K^{\mathbf{x}} \in \mathbb{R}^d$ and $n_K^t \in \mathbb{R}$ are the space and time components of the outward-pointing unit normal vectors on $\partial K \cap \mathcal{F}_h^{\mathrm{time}}$ and $\partial K \cap \mathcal{F}_h^{\mathrm{space}}$, respectively. The superscripts "−" and "+" are used to denote the traces of a function on a space-like facet from the elements "before" (−) and "after" (+) the facet.

We denote the space–time broken function spaces as

$$H^s(\mathcal{T}_h) := \{v \in L^2(Q_T), \ v|_K \in H^s(K) \ \forall K \in \mathcal{T}_h\} \qquad s \in \mathbb{R}^+,$$

$$C^s(\mathcal{T}_h) := \{v : Q_T \to \mathbb{C}, \ v|_K \in C^s(K) \ \forall K \in \mathcal{T}_h\} \qquad s \in \mathbb{N}.$$

2.3 Ultra-weak discontinuous Galerkin formulation

For any finite-dimensional subspace $\mathbb{V}_{hp}\,(\mathcal{T}_h)$ of the broken Bochner–Sobolev space

$$\mathbf{V}(\mathcal{T}_h) := \prod_{K \in \mathcal{T}_h} H^1\left(K_t; L^2(K_{\mathbf{x}})\right) \cap L^2\left(K_t; H^2(K_{\mathbf{x}})\right),$$

the proposed ultra-weak DG variational formulation for the Schrödinger equation (1.1) is:

Seek $\psi_{hp} \in \mathbb{V}_{hp}(\mathcal{T}_h)$ such that the following equation is satisfied for all $K \in \mathcal{T}_h$:

$$
\begin{aligned}
\int_K \psi_{hp}\overline{\mathcal{S}_\epsilon s_{hp}}\,\mathrm{d}V &+ \int_{\partial K}\left[i\epsilon\widehat{\psi}_{hp}\overline{s_{hp}}n_K^t + \frac{\epsilon^2}{2}\left(\widehat{\overline{\nabla_{\mathbf{x}}\psi}}_{hp}\overline{s_{hp}} - \widehat{\psi}_{hp}\nabla\overline{s_{hp}}\right)\cdot\vec{\mathbf{n}}_K^{\mathbf{x}}\right]\mathrm{d}S \\
&+ i\int_K \mu\mathcal{S}_\epsilon\psi_{hp}\overline{\mathcal{S}_\epsilon s_{hp}}\,\mathrm{d}V = 0 \qquad\qquad \forall s_{hp}\in\mathbb{V}_{hp}(\mathcal{T}_h),
\end{aligned}
\tag{2.1}
$$

where $\overline{\cdot}$ denotes the complex conjugate. The first two terms in Equation (2.1) are obtained integrating by parts the product of Equation (1.1a) and $\overline{s_{hp}}$ twice in space and once in time. The last term in Equation (2.1) is a local least-squares volume penalty term that is added to guarantee the well-posedness of the variational formulation (2.1) for an almost arbitrary choice of discrete spaces, see Proposition 2.5 below.

The so-called *numerical fluxes* $\widehat{\psi}_{hp}$ and $\widehat{\overline{\nabla_{\mathbf{x}}\psi}}_{hp}$ are approximations of the traces of ψ_{hp} and $\nabla_{\mathbf{x}}\psi_{hp}$ on $\mathcal{F}_h$. We choose them as:

$$
\widehat{\psi}_{hp} := \begin{cases}
\psi_{hp}^- & \text{on } \mathcal{F}_h^{\text{space}}, \\
\psi_{hp} & \text{on } \mathcal{F}_h^T, \\
\psi_0 & \text{on } \mathcal{F}_h^0, \\
\{\!\{\psi_{hp}\}\!\} - i\beta\,[\![\nabla_{\mathbf{x}}\psi_{hp}]\!]_{\mathbf{N}} & \text{on } \mathcal{F}_h^{\text{time}}, \\
g_{\mathrm{D}} & \text{on } \mathcal{F}_h^{\mathrm{D}}, \\
\psi_{hp} - i\beta\left(\partial_{\mathbf{n}_{\mathbf{x}}}\psi_{hp} - g_{\mathrm{N}}\right) & \text{on } \mathcal{F}_h^{\mathrm{N}}, \\
\psi_{hp} - \delta(i\vartheta)^{-1}\left(\partial_{\mathbf{n}_{\mathbf{x}}}\psi_{hp} - i\vartheta\psi_{hp} - g_{\mathrm{R}}\right) & \text{on } \mathcal{F}_h^{\mathrm{R}},
\end{cases}
$$

$$
\widehat{\overline{\nabla_{\mathbf{x}}\psi}}_{hp} := \begin{cases}
\{\!\{\nabla_{\mathbf{x}}\psi_{hp}\}\!\} + i\alpha\,[\![\psi_{hp}]\!]_{\mathbf{N}} & \text{on } \mathcal{F}_h^{\text{time}}, \\
\nabla_{\mathbf{x}}\psi_{hp} + i\alpha\left(\psi_{hp} - g_{\mathrm{D}}\right)\vec{\mathbf{n}}_\Omega^{\mathbf{x}} & \text{on } \mathcal{F}_h^{\mathrm{D}}, \\
g_{\mathrm{N}}\vec{\mathbf{n}}_\Omega^{\mathbf{x}}, & \text{on } \mathcal{F}_h^{\mathrm{N}}, \\
\nabla_{\mathbf{x}}\psi_{hp} - (1-\delta)\left(\nabla_{\mathbf{x}}\psi_{hp} - i\vartheta\psi_{hp} - g_{\mathrm{R}}\right)\vec{\mathbf{n}}_\Omega^{\mathbf{x}} & \text{on } \mathcal{F}_h^{\mathrm{R}}.
\end{cases}
$$

The mesh-dependent stabilization functions

$$
\begin{aligned}
\alpha &\in L^\infty(\mathcal{F}_h^{\text{time}}\cup\mathcal{F}_h^{\mathrm{D}}), & \operatorname{ess\,inf}_{\mathcal{F}_h^{\text{time}}\cup\mathcal{F}_h^{\mathrm{D}}}\alpha &> 0, \\
\beta &\in L^\infty(\mathcal{F}_h^{\text{time}}\cup\mathcal{F}_h^{\mathrm{N}}), & \operatorname{ess\,inf}_{\mathcal{F}_h^{\text{time}}\cup\mathcal{F}_h^{\mathrm{N}}}\beta &> 0, \\
\delta &\in L^\infty(\mathcal{F}_h^{\mathrm{R}}), & 0 < \delta &\leq \frac{1}{2},
\end{aligned}
$$

$$\mu \in L^\infty(Q_T), \qquad \text{ess inf}_{Q_T} \, \mu > 0,$$

play an important role in the convergence of the method. In order to guarantee optimal convergence in a mesh-dependent norm, we set them as

$$\min\left\{h_{K_t}^2, h_{K_x}^2\right\} \le \epsilon^2 \mu|_K \le \max\left\{h_{K_t}^2, h_{K_x}^2\right\}, \quad \delta|_K = \min\left(\vartheta h_{K_x}, \frac{1}{2}\right),$$

$$\alpha|_F = \frac{1}{h_{F_x}} \quad \forall F \subset \mathcal{F}_h^{\text{time}} \cup \mathcal{F}_h^{\text{D}}, \qquad \beta|_F = h_{F_x} \quad \forall F \subset \mathcal{F}_h^{\text{time}} \cup \mathcal{F}_h^{\text{N}}, \tag{2.2}$$

where

$$\begin{cases} h_{F_x} = h_{K_x} & \text{if } F \subset \partial K \cap \left(\mathcal{F}_h^{\text{D}} \cup \mathcal{F}_h^{\text{N}}\right), \\ \min\{h_{K_x^1}, h_{K_x^2}\} \le h_{F_x} \le \max\{h_{K_x^1}, h_{K_x^2}\} & \text{if } F = K^1 \cap K^2 \subset \mathcal{F}_h^{\text{time}}. \end{cases}$$

After summing Equation (2.1) over all the elements $K \in \mathcal{T}_h$ and substituting the definition of the numerical fluxes, the following ultra-weak DG variational formulation is obtained:

$$\text{Seek } \psi_{hp} \in \mathbb{V}_{hp}(\mathcal{T}_h) \text{ such that: } \mathcal{A}\left(\psi_{hp}; s_{hp}\right) = \ell(s_{hp}) \quad \forall s_{hp} \in \mathbb{V}_{hp}(\mathcal{T}_h), \tag{2.3}$$

where

$$\mathcal{A}\left(\psi_{hp}; s_{hp}\right) := \sum_{K \in \mathcal{T}_h} \int_K \psi_{hp} \overline{\mathcal{S}_\epsilon s_{hp}} \, dV + i\epsilon \left(\int_{\mathcal{F}_h^{\text{space}}} \psi_{hp}^- \llbracket \overline{s_{hp}} \rrbracket_t \, d\mathbf{x} + \int_{\mathcal{F}_h^T} \psi_{hp} \overline{s_{hp}} d\mathbf{x} \right)$$

$$+ \frac{\epsilon^2}{2} \int_{\mathcal{F}_h^{\text{time}}} \left(\{\!\!\{ \nabla_\mathbf{x} \psi_{hp} \}\!\!\} \cdot \llbracket \overline{s_{hp}} \rrbracket_\mathbf{N} + i\alpha \llbracket \psi_{hp} \rrbracket_\mathbf{N} \cdot \llbracket \overline{s_{hp}} \rrbracket_\mathbf{N} \right.$$

$$\left. - \{\!\!\{ \psi_{hp} \}\!\!\} \llbracket \nabla_\mathbf{x} \overline{s_{hp}} \rrbracket_\mathbf{N} + i\beta \llbracket \nabla_\mathbf{x} \psi_{hp} \rrbracket_\mathbf{N} \llbracket \nabla_\mathbf{x} \overline{s_{hp}} \rrbracket_\mathbf{N} \right) dS$$

$$+ \frac{\epsilon^2}{2} \int_{\mathcal{F}_h^{\text{D}}} \left(\partial_{\mathbf{n}_\mathbf{x}} \psi_{hp} + i\alpha \psi_{hp} \right) \overline{s_{hp}} dS$$

$$+ \frac{\epsilon^2}{2} \int_{\mathcal{F}_h^{\text{N}}} \left(-\psi_{hp} \partial_{\mathbf{n}_\mathbf{x}} \overline{s_{hp}} + i\beta \left(\partial_{\mathbf{n}_\mathbf{x}} \psi_{hp} \right) \left(\partial_{\mathbf{n}_\mathbf{x}} \overline{s_{hp}} \right) \right) dS$$

$$+ \frac{\epsilon^2}{2} \int_{\mathcal{F}_h^{\text{R}}} \left(\delta \partial_{\mathbf{n}_\mathbf{x}} \psi_{hp} + (1-\delta) i\vartheta \psi_{hp} \right) \left(\overline{s_{hp}} + \frac{i}{\vartheta} \partial_{\mathbf{n}_\mathbf{x}} \overline{s_{hp}} \right) dS$$

$$+ i \sum_{K \in \mathcal{T}_h} \int_K \mu \mathcal{S}_\epsilon \psi_{hp} \overline{\mathcal{S}_\epsilon s_{hp}} \, dV,$$

$$\ell(s_{hp}) := i\epsilon \int_{\mathcal{F}_h^0} \psi_0 \overline{s_{hp}} d\mathbf{x} + \frac{\epsilon^2}{2} \int_{\mathcal{F}_h^{\text{D}}} g_{\text{D}} \left(\partial_{\mathbf{n}_\mathbf{x}} \overline{s_{hp}} + i\alpha \overline{s_{hp}} \right) dS$$

$$+ \frac{\epsilon^2}{2} \int_{\mathcal{F}_h^{\text{N}}} g_{\text{N}} \left(-\overline{s_{hp}} + i\beta \partial_{\mathbf{n}_\mathbf{x}} \overline{s_{hp}} \right) dS + \frac{\epsilon^2}{2} \int_{\mathcal{F}_h^{\text{R}}} g_{\text{R}} \left((\delta - 1)\overline{s_{hp}} + \frac{i\delta}{\vartheta} \partial_{\mathbf{n}_\mathbf{x}} \overline{s_{hp}} \right) dS.$$

Remark 2.1 (Trefftz spaces). If we require the discrete space $\mathbb{V}_{hp}(\mathcal{T}_h)$ to satisfy the Trefftz property ($\mathcal{S}_\epsilon \psi_{hp}|_K = 0, \; \forall K \in \mathcal{T}_h$), all the volume terms in the variational formulation (2.3) vanish, thus recovering the Trefftz-DG formulation in [43]. In such a case, the definitions of $\mathcal{A}(\cdot; \cdot)$ and $\ell(\cdot)$ are independent of the potential V, which has an effect only on the discrete space. ∎

Remark 2.2 (Implicit time-stepping through slabs)**.** The variational problem (2.3) is a global problem involving all the degrees of freedom of the discrete solution for the whole space–time cylinder Q_T. However, as upwind numerical fluxes are taken on the space-like facets, if the space–time prismatic mesh $\mathcal{T}_h$ can be decomposed into time slabs (i.e., if the mesh elements can be grouped in sets of the form $\Omega \times (t_{n-1}, t_n)$ for a partition of the time interval of the form $0 = t_0 < t_1 < \ldots < t_N = T$, the global linear system stemming from (2.3) can be solved as a sequence of N smaller systems of the form

$$\mathbf{K}_n \Psi_h^{(n)} = b_n \quad 1 \le n \le N,$$

where $b_n = \mathbf{R}_n \Psi_h^{(n-1)}$ for $n = 2, \ldots, N$. This is comparable to an implicit time-stepping, and it naturally allows for local mesh refinement in different regions of the space–time cylinder Q_T. Moreover, when $\mathcal{T}_h$ is a tensor-product space–time mesh, the potential V does not vary in time, and the partition of the time interval is uniform, the matrices $\mathbf{K}_n$ and $\mathbf{R}_n$ are the same for every time slab. ∎

Remark 2.3 (Self-adjointness and volume penalty term)**.** The well-posedness of the variational formulation (2.3) strongly relies on the $L^2(K)$-self-adjointness of the Schrödinger operator $\mathcal{S}_\epsilon(\cdot)$ on each $K \in \mathcal{T}_h$ (in the sense that $\int_K \mathcal{S}_\epsilon \psi \, \overline{\varphi} \, dV = \int_K \psi \, \overline{\mathcal{S}_\epsilon \varphi} \, dV$ for all $\psi \in \mathbf{V}(\mathcal{T}_h)$, $\varphi \in C_0^\infty(K)$, thanks to the fact that the only odd derivative in $\mathcal{S}_\epsilon$ is multiplied by the imaginary unit), which makes the local Galerkin-least squares correction term consistent. On the one hand, such a term is essential in the proof of coercivity of the sesquilinear form $\mathcal{A}(\cdot\,;\cdot)$ (see Proposition 2.5 below). On the other hand, numerical experiments suggest that it can be neglected without losing accuracy and stability, see Section 5.5.2. This is also the case for the quasi-Trefftz DG method for the Helmholtz equation [56, §5.1.3] and for the wave equation [55, §5.1], where a similar correction term was used. Nonetheless, in the design of an ultra-weak DG discretization for a PDE with a non-self-adjoint differential operator $\mathcal{L}(\cdot)$ (e.g., the heat operator $\mathcal{L}(\cdot) = (\partial_t - \Delta_\mathbf{x})(\cdot)$), the corresponding local least-squares correction term $\sum_{K \in \mathcal{T}_h} \int_K \mu \mathcal{L}\psi_{hp} \overline{\mathcal{L}s_{hp}} \, dV$ would not control the consistency term $\sum_{K \in \mathcal{T}_h} \int_K \psi_{hp} \overline{\mathcal{L}^* s_{hp}} \, dV$ arising from the integration by parts. ∎

Remark 2.4 (Time-dependent potentials)**.** The variational problem (2.3) allows for time-dependent potentials V. This is an important feature as, in such a case, the method of separation of variables cannot be used to reduce the time-dependent problem (1.1) to the time-independent Schrödinger equation. ∎

2.4 Well-posedness, stability and quasi-optimality of the DG method

The theoretical results in this section are derived for any space dimension $d \in \mathbb{N}$, and are independent of the specific choice of the discrete space $\mathbb{V}_{hp}(\mathcal{T}_h)$. Moreover, inverse estimates are not needed in the analysis. This is an important advantage in comparison to other DG-type formulations (such as interior-penalty, see [7] for a Trefftz example) as inverse estimates for non-polynomial discrete spaces are in general hard to obtain (see [40, §3.2]).

Recalling that the volume penalty function μ, the stabilization functions α, β and the impedance function ϑ are positive, and that $\delta \in (0, \frac{1}{2}]$, we define the following mesh-dependent

norms on $\mathbf{V}(\mathcal{T}_h)$:

$$
\begin{aligned}
|||w|||_{\text{DG}}^2 &:= \sum_{K \in \mathcal{T}_h} \left\| \mu^{\frac{1}{2}} \left(i\epsilon \partial_t w + \frac{\epsilon^2}{2} \Delta_{\mathbf{x}} w - Vw \right) \right\|_{L^2(K)}^2 + \frac{\epsilon}{2} \left(\left\| [\![w]\!]_t \right\|_{L^2(\mathcal{F}_h^{\text{space}})}^2 + \|w\|_{L^2(\mathcal{F}_h^T \cup \mathcal{F}_h^0)}^2 \right) \\
&\quad + \frac{\epsilon^2}{2} \left(\left\| \alpha^{\frac{1}{2}} [\![w]\!]_{\mathbf{N}} \right\|_{L^2(\mathcal{F}_h^{\text{time}})^d}^2 + \left\| \beta^{\frac{1}{2}} [\![\nabla_{\mathbf{x}} w]\!]_{\mathbf{N}} \right\|_{L^2(\mathcal{F}_h^{\text{time}})}^2 + \left\| \alpha^{\frac{1}{2}} w \right\|_{L^2(\mathcal{F}_h^D)}^2 \right. \hspace{1cm} \text{(2.4a)} \\
&\quad \left. + \left\| \beta^{\frac{1}{2}} \partial_{\mathbf{n}_{\mathbf{x}}} w \right\|_{L^2(\mathcal{F}_h^N)}^2 + \left\| (\vartheta(1-\delta))^{\frac{1}{2}} w \right\|_{L^2(\mathcal{F}_h^R)}^2 + \left\| (\delta \vartheta^{-1})^{\frac{1}{2}} \partial_{\mathbf{n}_{\mathbf{x}}} w \right\|_{L^2(\mathcal{F}_h^R)}^2 \right), \\
|||w|||_{\text{DG}^+}^2 &:= |||w|||_{\text{DG}}^2 + \sum_{K \in \mathcal{T}_h} \left\| \mu^{-\frac{1}{2}} w \right\|_{L^2(K)}^2 + \frac{\epsilon}{2} \|w^-\|_{L^2(\mathcal{F}_h^{\text{space}})}^2 + \frac{\epsilon^2}{2} \left(\left\| \alpha^{-\frac{1}{2}} \{\!\!\{ \nabla_{\mathbf{x}} w \}\!\!\} \right\|_{L^2(\mathcal{F}_h^{\text{time}})^d}^2 \right. \\
&\quad \left. + \left\| \alpha^{-\frac{1}{2}} \partial_{\mathbf{n}_{\mathbf{x}}} w \right\|_{L^2(\mathcal{F}_h^D)}^2 + \left\| \beta^{-\frac{1}{2}} \{\!\!\{ w \}\!\!\} \right\|_{L^2(\mathcal{F}_h^{\text{time}})}^2 + \left\| \beta^{-\frac{1}{2}} w \right\|_{L^2(\mathcal{F}_h^N)}^2 + \left\| (\delta^{-1} \vartheta)^{\frac{1}{2}} w \right\|_{L^2(\mathcal{F}_h^R)}^2 \right).
\end{aligned}
$$

$$\hspace{14cm} \text{(2.4b)}$$

The sum of the $L^2(K)$-type terms ensures that $||| \cdot |||_{\text{DG}^+}$ is a norm. That $||| \cdot |||_{\text{DG}}$ is a norm on $\mathbf{V}(\mathcal{T}_h)$ follows from the following reasoning: if $w \in \mathbf{V}(\mathcal{T}_h)$ and $|||w|||_{\text{DG}} = 0$, then $w \in H^1(Q_T)$ and satisfies the Schrödinger equation (1.1) with homogeneous initial and boundary conditions. Moreover, integration by parts gives the following identity for all $t \in (0, T]$

$$
\begin{aligned}
0 &= \mathfrak{Im} \left(\int_0^t \int_\Omega \left(i\epsilon \partial_t w + \frac{\epsilon^2}{2} \Delta_{\mathbf{x}} w - Vw \right) \overline{w} \, d\mathbf{x} \, ds \right) \\
&= \mathfrak{Re} \left(\epsilon \int_0^t \int_\Omega \partial_t w \overline{w} \, d\mathbf{x} \, ds \right) + \mathfrak{Im} \left(\frac{\epsilon^2}{2} \int_0^t \|\nabla_{\mathbf{x}} w\|_{L^2(\Omega)^d}^2 \, ds \right. \\
&\quad \left. + \frac{i\epsilon^2}{2} \int_0^t \left\| \vartheta^{\frac{1}{2}} w \right\|_{L^2(\Gamma_R)}^2 \, ds - \int_0^t \int_\Omega V |w|^2 \, d\mathbf{x} \, ds \right) \\
&= \frac{\epsilon^2}{2} \left(\|w(\cdot, t)\|_{L^2(\Omega)}^2 - \|w(\cdot, 0)\|_{L^2(\Omega)}^2 + \int_0^t \left\| \vartheta^{\frac{1}{2}} w \right\|_{L^2(\Gamma_R)}^2 \right).
\end{aligned}
$$

Then, $\|w(\cdot, t)\|_{L^2(\Omega)}^2 \leq \|w(\cdot, 0)\|_{L^2(\Omega)}^2 = 0$, for all $t \in [0, T]$; therefore, $w = 0$.

The DG norms in (2.4a)–(2.4b) are chosen in order to ensure some properties of the sesquilinear form $\mathcal{A}(\cdot; \cdot)$ and the antilinear functional $\ell(\cdot)$, from which the well-posedness and quasi-optimality of the method (2.3) follow.

We recall some standard DG flux-jump identities that are instrumental in the proof of Proposition 2.5 below. For all functions $v \in H^1(\mathcal{T}_h)$ and $\boldsymbol{\tau} \in H^1(\mathcal{T}_h)^d$, it holds

$$
\int_{\mathcal{F}_h^{\text{space}}} \left(\mathfrak{Re}\left(v^- [\![\overline{v}]\!]_t \right) - \frac{1}{2} [\![|v|^2]\!]_t \right) d\mathbf{x} = \frac{1}{2} \int_{\mathcal{F}_h^{\text{space}}} \left| [\![v]\!]_t \right|^2 d\mathbf{x}, \hspace{1cm} \text{(2.5a)}
$$

$$
\int_{\mathcal{F}_h^{\text{time}}} \left(\{\!\!\{ v \}\!\!\} [\![\boldsymbol{\tau}]\!]_{\mathbf{N}} + \{\!\!\{ \boldsymbol{\tau} \}\!\!\} \cdot [\![v]\!]_{\mathbf{N}} \right) dS = \int_{\mathcal{F}_h^{\text{time}}} [\![v \boldsymbol{\tau}]\!]_{\mathbf{N}} \, dS. \hspace{1cm} \text{(2.5b)}
$$

Proposition 2.5 (Coercivity). *For all $w \in \mathbf{V}(\mathcal{T}_h)$ the following identity holds*

$$
\mathfrak{Im}\left(\mathcal{A}(w; w) \right) = |||w|||_{\text{DG}}^2.
$$

Proof. Elementwise integration by parts for $w \in \mathbb{V}_{hp}(\mathcal{T}_h)$ gives the identity:

$$
\mathfrak{Im}\left(\sum_{K \in \mathcal{T}_h} \int_K w \overline{\left(i\epsilon \partial_t w + \frac{\epsilon^2}{2} \Delta_\mathbf{x} w - V w \right)} \, dV \right)
$$

$$
= -\frac{\epsilon}{2} \sum_{K \in \mathcal{T}_h} \int_K \partial_t |w|^2 \, dV + \frac{\epsilon^2}{2} \mathfrak{Im}\left(\int_{\mathcal{F}_h^{\text{time}}} [\![w \nabla_\mathbf{x} \overline{w}]\!]_\mathbf{N} \, dS + \int_{\partial \Omega \times I} w \partial_{\mathbf{n_x}} \overline{w} \, dS \right)
$$

$$
= -\frac{\epsilon}{2}\left(\int_{\mathcal{F}_h^{\text{space}}} [\![|w|^2]\!]_t \, d\mathbf{x} + \int_{\mathcal{F}_h^T} |w|^2 \, d\mathbf{x} - \int_{\mathcal{F}_h^0} |w|^2 \, d\mathbf{x} \right)
$$

$$
+ \frac{\epsilon^2}{2} \mathfrak{Im}\left(\int_{\mathcal{F}_h^{\text{time}}} [\![w \nabla_\mathbf{x} \overline{w}]\!]_\mathbf{N} \, dS + \int_{\partial \Omega \times I} w \partial_{\mathbf{n_x}} \overline{w} \, dS \right). \tag{2.6}
$$

Together with the DG flux-jump identities in (2.5), this gives

$$
\mathfrak{Im}\left(\mathcal{A}(w; w) \right) \overset{(2.6)}{=} \int_{\mathcal{F}_h^{\text{space}}} \left(\mathfrak{Re}\left(\epsilon w^- [\![\overline{w}]\!]_t \right) - \frac{\epsilon}{2} [\![|w|^2]\!]_t \right) d\mathbf{x} + \frac{\epsilon}{2}\left(\int_{\mathcal{F}_h^T} |w|^2 \, d\mathbf{x} + \int_{\mathcal{F}_h^0} |w|^2 \, d\mathbf{x} \right)
$$

$$
+ \frac{\epsilon^2}{2}\left(\int_{\mathcal{F}_h^{\text{time}}} \mathfrak{Im}\left(\{\!\{ \nabla_\mathbf{x} w \}\!\} \cdot [\![\overline{w}]\!]_\mathbf{N} - \{\!\{ w \}\!\} [\![\nabla_\mathbf{x} \overline{w}]\!]_\mathbf{N} + [\![w \nabla_\mathbf{x} \overline{w}]\!]_\mathbf{N} \right) dS \right.
$$

$$
+ \int_{\mathcal{F}_h^{\text{time}}} \left(\alpha \left| [\![w]\!]_\mathbf{N} \right|^2 + \beta \left| [\![\nabla_\mathbf{x} w]\!]_\mathbf{N} \right|^2 \right) dS + \int_{\mathcal{F}_h^D} \alpha |w|^2 \, dS
$$

$$
\left. + \int_{\mathcal{F}_h^N} \beta \left| \partial_{\mathbf{n_x}} w \right|^2 dS + \int_{\mathcal{F}_h^R} \left(\vartheta(1 - \delta) |w|^2 + (\delta \vartheta^{-1}) \left| \partial_{\mathbf{n_x}} w \right|^2 \right) dS \right)
$$

$$
+ \sum_{K \in \mathcal{T}_h} \int_K \mu |\mathcal{S}_\epsilon w|^2 \, dV \overset{(2.5)}{=} |||w|||^2_{\text{DG}}. \qquad \square
$$

Proposition 2.6 (Continuity). *The sesquilinear form $\mathcal{A}(\cdot; \cdot)$ and the antilinear functional $\ell(\cdot)$ are continuous in the following sense: $\forall v, w \in \mathbf{V}(\mathcal{T}_h)$*

$$
|\mathcal{A}(v; w)| \leq 2 |||v|||_{\text{DG}^+} |||w|||_{\text{DG}}, \tag{2.7a}
$$

$$
|\ell(v)| \leq \left(2\epsilon \|\psi_0\|^2_{L^2(\mathcal{F}_h^0)} + \epsilon^2 \left\| \alpha^{\frac{1}{2}} g_D \right\|^2_{L^2(\mathcal{F}_h^D)} \right.
$$

$$
\left. + \epsilon^2 \left\| \beta^{\frac{1}{2}} g_N \right\|^2_{L^2(\mathcal{F}_h^N)} + \epsilon^2 \left\| \vartheta^{-\frac{1}{2}} g_R \right\|^2_{L^2(\mathcal{F}_h^R)} \right)^{\frac{1}{2}} |||w|||_{\text{DG}^+}. \tag{2.7b}
$$

Proof. Let $v, w \in \mathbf{V}(\mathcal{T}_h)$. After applying the Cauchy–Schwarz inequality several times and using the inequality $\delta \leq 1 - \delta < 1$, we obtain

$$
|\mathcal{A}(v; w)| \leq \sum_{K \in \mathcal{T}_h} \left(\left\| \mu^{-\frac{1}{2}} v \right\|_{L^2(K)} + \left\| \mu^{\frac{1}{2}} \mathcal{S}_\epsilon v \right\|_{L^2(K)} \right) \left\| \mu^{\frac{1}{2}} \mathcal{S}_\epsilon w \right\|_{L^2(K)}
$$

$$
+ \epsilon \left(\|v^-\|_{L^2(\mathcal{F}_h^{\text{space}})} \left\| [\![w]\!]_t \right\|_{L^2(\mathcal{F}_h^{\text{space}})} + \|v\|_{L^2(\mathcal{F}_h^T)} \|w\|_{L^2(\mathcal{F}_h^T)} \right)
$$

$$
+ \frac{\epsilon^2}{2}\left(\left\| \alpha^{-\frac{1}{2}} \partial_{\mathbf{n_x}} v \right\|_{L^2(\mathcal{F}_h^D)} + \left\| \alpha^{\frac{1}{2}} v \right\|_{L^2(\mathcal{F}_h^D)} \right) \left\| \alpha^{\frac{1}{2}} w \right\|_{L^2(\mathcal{F}_h^D)}
$$

$$+ \frac{\epsilon^2}{2} \left(\left\| \beta^{-\frac{1}{2}} v \right\|_{L^2(\mathcal{F}_h^{\mathrm{N}})} + \left\| \beta^{\frac{1}{2}} \partial_{\mathbf{n}_{\mathbf{x}}} v \right\|_{L^2(\mathcal{F}_h^{\mathrm{N}})} \right) \left\| \beta^{\frac{1}{2}} \partial_{\mathbf{n}_{\mathbf{x}}} w \right\|_{L^2(\mathcal{F}_h^{\mathrm{N}})}$$

$$+ \frac{\epsilon^2}{2} \left(\left\| (\delta \vartheta^{-1})^{\frac{1}{2}} \partial_{\mathbf{n}_{\mathbf{x}}} v \right\|_{L^2(\mathcal{F}_h^{\mathrm{R}})} + \left\| ((1-\delta)\vartheta)^{\frac{1}{2}} v \right\|_{L^2(\mathcal{F}_h^{\mathrm{R}})} \right) \left\| ((1-\delta)\vartheta)^{\frac{1}{2}} w \right\|_{L^2(\mathcal{F}_h^{\mathrm{R}})}$$

$$+ \frac{\epsilon^2}{2} \left(\left\| (\delta \vartheta^{-1})^{\frac{1}{2}} \partial_{\mathbf{n}_{\mathbf{x}}} v \right\|_{L^2(\mathcal{F}_h^{\mathrm{R}})} + \left\| (\delta^{-1}\vartheta)^{\frac{1}{2}} v \right\|_{L^2(\mathcal{F}_h^{\mathrm{R}})} \right) \left\| (\delta \vartheta^{-1})^{\frac{1}{2}} \partial_{\mathbf{n}_{\mathbf{x}}} w \right\|_{L^2(\mathcal{F}_h^{\mathrm{R}})}$$

$$+ \frac{\epsilon^2}{2} \left(\left\| \alpha^{-\frac{1}{2}} \{\!\{ \nabla_{\mathbf{x}} v \}\!\} \right\|_{L^2(\mathcal{F}_h^{\mathrm{time}})^d} + \left\| \alpha^{\frac{1}{2}} [\![v]\!]_{\mathbf{N}} \right\|_{L^2(\mathcal{F}_h^{\mathrm{time}})^d} \right) \left\| \alpha^{\frac{1}{2}} [\![w]\!]_{\mathbf{N}} \right\|_{L^2(\mathcal{F}_h^{\mathrm{time}})^d}$$

$$+ \frac{\epsilon^2}{2} \left(\left\| \beta^{-\frac{1}{2}} \{\!\{ v \}\!\} \right\|_{L^2(\mathcal{F}_h^{\mathrm{time}})} + \left\| \beta^{\frac{1}{2}} [\![\nabla_{\mathbf{x}} v]\!]_{\mathbf{N}} \right\|_{L^2(\mathcal{F}_h^{\mathrm{time}})} \right) \left\| \beta^{\frac{1}{2}} [\![\nabla w]\!]_{\mathbf{N}} \right\|_{L^2(\mathcal{F}_h^{\mathrm{time}})}.$$

Bound (2.7a) is then obtained by using the Cauchy–Schwarz inequality once again. The bound for the linear operator $\ell(\cdot)$ can be obtained in a similar way. $\qquad\square$

Theorem 2.7 (Quasi-optimality). *For any finite-dimensional subspace $\mathbb{V}_{hp}(\mathcal{T}_h)$ of $\mathbf{V}(\mathcal{T}_h)$, there exists a unique solution $\psi_{hp} \in \mathbb{V}_{hp}(\mathcal{T}_h)$ satisfying the variational formulation* (2.3). *Additionally, the following quasi-optimality bound holds:*

$$||| \psi - \psi_{hp} |||_{\mathrm{DG}} \le 3 \inf_{s_{hp} \in \mathbb{V}_{hp}(\mathcal{T}_h)} ||| \psi - s_{hp} |||_{\mathrm{DG}^+}. \tag{2.8}$$

Moreover, if $g_{\mathrm{D}} = 0$ and $g_{\mathrm{N}} = 0$ (or $\Gamma_{\mathrm{D}} = \emptyset$ and $\Gamma_{\mathrm{N}} = \emptyset$), then

$$||| \psi_{hp} |||_{\mathrm{DG}} \le \left(2\epsilon \, \|\psi_0\|_{L^2(\mathcal{F}_h^0)}^2 + \epsilon^2 \left\| \vartheta^{-1/2} g_{\mathrm{R}} \right\|_{L^2(\mathcal{F}_h^{\mathrm{R}})}^2 \right)^{1/2}. \tag{2.9}$$

Proof. Existence and uniqueness of the discrete solution $\psi_{hp} \in \mathbb{V}_{hp}(\mathcal{T}_h)$ of the variational formulation (2.3), and the quasi-optimality bound (2.8) follow directly from Propositions 2.5 and 2.6, the consistency of the variational formulation (2.3) and Lax–Milgram theorem. The continuous dependence on the data (2.9) follows from Proposition 2.5, and the fact that if $g_{\mathrm{D}} = 0$ and $g_{\mathrm{N}} = 0$ (or $\Gamma_{\mathrm{D}} = \emptyset$ and $\Gamma_{\mathrm{N}} = \emptyset$), the term $||| w |||_{\mathrm{DG}^+}$ on the right-hand side of (2.7b) can be replaced by $||| w |||_{\mathrm{DG}}$. $\qquad\square$

Theorem 2.7 implies that it is possible to obtain error estimates in the mesh-dependent norm $||| \cdot |||_{\mathrm{DG}}$ by studying the best approximation in $\mathbb{V}_{hp}(\mathcal{T}_h)$ of the exact solution in the $||| \cdot |||_{\mathrm{DG}^+}$ norm. Moreover, according to Proposition 2.8 below, *a priori* error estimates can be deduced from the local approximation properties of the space $\mathbb{V}_{hp}(\mathcal{T}_h)$ only, as the $||| \cdot |||_{\mathrm{DG}^+}$ norm can be bounded in terms of local volume Sobolev seminorms and norms.

So far, we have not imposed any restriction on the space–time mesh $\mathcal{T}_h$. Henceforth, in our analysis we assume:

- **Uniform star-shapedness**: There exists $0 < \rho \le \frac{1}{2}$ such that, each element $K \in \mathcal{T}_h$ is star-shaped with respect to the ball $B := B_{\rho h_K}(\mathbf{z}_K, s_K)$ centered at $(\mathbf{z}_K, s_K) \in K$ and with radius ρh_K.

- **Local quasi-uniformity in space**: there exists a number $\mathsf{lqu}(\mathcal{T}_h) > 0$ such that $h_{K_{1,\mathbf{x}}} \le h_{K_{2,\mathbf{x}}} \mathsf{lqu}(\mathcal{T}_h)$ for all $K_1 = K_{1,\mathbf{x}} \times K_{1,t}, K_2 = K_{2,\mathbf{x}} \times K_{2,t} \in \mathcal{T}_h$, whose intersection $K_1 \cap K_2$ has positive d-dimensional measure.

The proof of Proposition 2.8 is a direct consequence of a collection of trace inequalities (see [12, Thm. 1.6.6] and [80, Lemma 2]), which in our space–time setting can be written for any element $K = K_{\mathbf{x}} \times K_t \in \mathcal{T}_h$ as

$$\|\varphi\|^2_{L^2(K_{\mathbf{x}} \times \partial K_t)} \leq C_{\text{tr}} \left(h_{K_t}^{-1} \|\varphi\|^2_{L^2(K)} + h_{K_t} \|\partial_t \varphi\|^2_{L^2(K)} \right) \qquad \forall \varphi \in H^1 \left(K_t; L^2(K_{\mathbf{x}}) \right),$$

$$\|\varphi\|^2_{L^2(\partial K_{\mathbf{x}} \times K_t)} \leq C_{\text{tr}} \left(h_{K_{\mathbf{x}}}^{-1} \|\varphi\|^2_{L^2(K)} + h_{K_{\mathbf{x}}} \|\nabla_{\mathbf{x}} \varphi\|^2_{L^2(K)^d} \right) \qquad \forall \varphi \in L^2 \left(K_t; H^1 \left(K_{\mathbf{x}} \right) \right),$$

$$\|\nabla_{\mathbf{x}} \varphi\|^2_{L^2(\partial K_{\mathbf{x}} \times K_t)^d} \leq C_{\text{tr}} \left(h_{K_{\mathbf{x}}}^{-1} \|\nabla_{\mathbf{x}} \varphi\|^2_{L^2(K)^d} + h_{K_{\mathbf{x}}} \left\|D_{\mathbf{x}}^2 \varphi\right\|^2_{L^2(K)^{d \times d}} \right) \quad \forall \varphi \in L^2 \left(K_t; H^2 \left(K_{\mathbf{x}} \right) \right),$$
(2.10)

where $D_{\mathbf{x}}^2 \varphi$ is the spatial Hessian of φ, and $C_{\text{tr}} \geq 1$ only depends on the star-shapedness parameter ρ.

Proposition 2.8. *Assume that $V \in L^\infty(K)$, $\forall K \in \mathcal{T}_h$, and fix α, β, δ and μ as in (2.2). For all $\varphi \in \mathbf{V}(\mathcal{T}_h)$, the following bound holds*

$$\begin{aligned}
|||\varphi|||^2_{\text{DG}^+} \leq & 3C_{\text{tr}} \sum_{K = K_{\mathbf{x}} \times K_t \in \mathcal{T}_h} \left[\epsilon h_{K_t}^{-1} \|\varphi\|^2_{L^2(K)} + \epsilon h_{K_t} \|\partial_t \varphi\|^2_{L^2(K)} + \epsilon^2 \mathrm{a}_K^2 h_{K_{\mathbf{x}}}^{-1} \|\varphi\|^2_{L^2(K)} \right. \\
& + \epsilon^2 \left(\mathrm{a}_K^2 h_{K_{\mathbf{x}}} + \mathrm{b}_K^2 h_{K_{\mathbf{x}}}^{-1} \right) \|\nabla_{\mathbf{x}} \varphi\|^2_{L^2(K)^d} + \epsilon^2 \mathrm{b}_K^2 h_{K_{\mathbf{x}}} \left\|D_{\mathbf{x}}^2 \varphi\right\|^2_{L^2(K)^{d \times d}} \qquad (2.11) \\
& \left. + \epsilon^2 \left\|\mu^{\frac{1}{2}} \partial_t \varphi\right\|^2_{L^2(K)} + \epsilon^4 \left\|\mu^{\frac{1}{2}} \Delta_{\mathbf{x}} \varphi\right\|^2_{L^2(K)} + \|V\|^2_{L^\infty(K)} \left\|\mu^{\frac{1}{2}} \varphi\right\|^2_{L^2(K)} + \left\|\mu^{-\frac{1}{2}} \varphi\right\|^2_{L^2(K)} \right],
\end{aligned}$$

where

$$\mathrm{a}_K^2 := \max \left\{ \operatorname*{ess\,sup}_{\partial K \cap \left(\mathcal{F}_h^{\text{time}} \cup \mathcal{F}_h^{\text{D}}\right)} \alpha, \ \left(\operatorname*{ess\,inf}_{\partial K \cap \left(\mathcal{F}_h^{\text{time}} \cup \mathcal{F}_h^{\text{N}}\right)} \beta \right)^{-1}, \ \operatorname*{ess\,sup}_{\partial K \cap \mathcal{F}_h^{\text{R}}} \vartheta \right\},$$

$$\mathrm{b}_K^2 := \max \left\{ \left(\operatorname*{ess\,inf}_{\partial K \cap \left(\mathcal{F}_h^{\text{time}} \cup \mathcal{F}_h^{\text{D}}\right)} \alpha \right)^{-1}, \ \operatorname*{ess\,sup}_{\partial K \cap \left(\mathcal{F}_h^{\text{time}} \cup \mathcal{F}_h^{\text{N}}\right)} \beta, \ h_{K_{\mathbf{x}}} \right\}.$$

The volume term $\left\|\mu^{\frac{1}{2}} \left(i\epsilon \partial_t w + \frac{\epsilon^2}{2} \Delta_{\mathbf{x}} w - V w \right)\right\|^2_{L^2(K)}$ is controlled by the inequality $|\mathbf{y}|_1 \leq \sqrt{n} \, |\mathbf{y}|_2$, $\forall \mathbf{y} \in \mathbb{C}^n$. If φ is a Trefftz function, the last four terms in the right hand side of (2.11) can be dropped.

Remark 2.9 (Energy dissipation). It is well known that the Schrödinger equation (1.1) with homogeneous Dirichlet and/or Neumann boundary conditions and $\Gamma_{\text{R}} = \emptyset$ preserves the energy (or probability) functional $\mathcal{E}(t; \psi) := \frac{1}{2} \int_\Omega |\psi(\mathbf{x}, t)|^2 d\mathbf{x}$, i.e. $\frac{d}{dt} \mathcal{E}(t; \psi) = 0$.

The proposed DG method is dissipative, but the energy loss can be quantified in terms of the local least-squares error, the initial condition error, the jumps of the solution on the mesh skeleton, and the error on $\mathcal{F}_h^{\text{D}} \cup \mathcal{F}_h^{\text{N}}$ due to the weak imposition of the boundary conditions. More precisely, for $g_{\text{D}} = 0$, $g_{\text{N}} = 0$ and $\mathcal{F}_h^{\text{R}} = \emptyset$, the discrete solution to (2.3) satisfies

$$\mathcal{E}(0; \psi_0) - \mathcal{E}(T; \psi_{hp}) = \mathcal{E}_{loss} := \delta_{\mathcal{E}} + \frac{\epsilon}{2} \left\|\psi_0 - \psi_{hp}\right\|^2_{\mathcal{F}_h^0},$$

where $\quad \delta_{\mathcal{E}} := \sum_{K \in \mathcal{T}_h} \left\| \mu^{\frac{1}{2}} \mathcal{S}_\epsilon \psi_{hp} \right\|^2_{L^2(K)} + \frac{\epsilon}{2} \left\| [\![\psi_{hp}]\!]_t \right\|^2_{L^2(\mathcal{F}_h^{\text{space}})} + \frac{\epsilon^2}{2} \Big(\left\| \alpha^{\frac{1}{2}} \psi_{hp} \right\|^2_{L^2(\mathcal{F}_h^{\text{D}})}$

$$+ \left\| \beta^{\frac{1}{2}} \partial_{\mathbf{n_x}} \psi_{hp} \right\|^2_{L^2(\mathcal{F}_h^{\text{N}})} + \left\| \alpha^{\frac{1}{2}} [\![\psi_{hp}]\!]_{\mathbf{N}} \right\|^2_{L^2(\mathcal{F}_h^{\text{time}})^d} + \left\| \beta^{\frac{1}{2}} [\![\nabla_{\mathbf{x}} \psi_{hp}]\!]_{\mathbf{N}} \right\|^2_{L^2(\mathcal{F}_h^{\text{time}})} \Big).$$

This follows from the definition of the $||| \cdot |||_{\text{DG}}$ norm of the solution ψ_{hp}, the coercivity of the sesquilinear form $\mathcal{A}(\cdot\,;\cdot)$, the definition of the antilinear functional $\ell(\cdot)$, and simple algebraic manipulations

$$\epsilon \Big(\mathcal{E}(0; \psi_{hp}) + \mathcal{E}(T; \psi_{hp}) \Big) + \delta_{\mathcal{E}} \overset{(2.4a)}{=} |||\psi_{hp}|||^2_{\text{DG}} \overset{(2.5)}{=} \mathfrak{Im}\left(\mathcal{A}\left(\psi_{hp};\, \psi_h \right) \right) \overset{(2.3)}{=} \mathfrak{Im}\left(\ell(\psi_{hp}) \right)$$

$$= \mathfrak{Re}\left(\epsilon \int_{\mathcal{F}_h^0} \psi_0 \overline{\psi_{hp}} \mathrm{d}\mathbf{x} \right) = \epsilon \left(\mathcal{E}(0; \psi_0) + \mathcal{E}(0; \psi_{hp}) - \frac{1}{2} \left\| \psi_0 - \psi_{hp} \right\|^2_{\mathcal{F}_h^0} \right).$$

Further manipulations give the identity $\mathcal{E}_{loss} = |||\psi - \psi_{hp}|||^2_{\text{DG}} - \mathcal{E}(T, \psi - \psi_{hp})$. ∎

2.5 Approximation by Taylor polynomials

We recall the definition and approximation properties of multivariate Taylor polynomials, which constitute the basis of our error analysis.

Definition 2.10 (Taylor polynomial). On an open and bounded set $\Upsilon \subset \mathbb{R}^{d+1}$, the Taylor polynomial of order $m \in \mathbb{N}$ (and degree $m-1$), centered at $(\mathbf{z}, s) \in \Upsilon$, of a function $\varphi \in C^{m-1}(\Upsilon)$ is defined as

$$T^m_{(\mathbf{z},s)}[\varphi](\mathbf{x}, t) := \sum_{|\mathbf{j}| < m} \frac{1}{\mathbf{j}!} D^{\mathbf{j}} \varphi(\mathbf{z}, s)(\mathbf{x} - \mathbf{z})^{j_{\mathbf{x}}} (t - s)^{j_t}.$$

If $\varphi \in C^m(\Upsilon)$ and the segment $[(\mathbf{z}, s), (\mathbf{x}, t)] \subset \Upsilon$, the Lagrange's form of the Taylor remainder (see [13, Corollary 3.19]) is bounded as follows:

$$\left| \varphi(\mathbf{x}, t) - T^m_{(\mathbf{z},s)}[\varphi](\mathbf{x}, t) \right| \leq |\varphi|_{C^m(\Upsilon)} \sum_{|\mathbf{j}|=m} \frac{1}{\mathbf{j}!} \left| (\mathbf{x} - \mathbf{z})^{j_{\mathbf{x}}} (t - s)^{j_t} \right|$$

$$\leq \frac{(d+1)^{\frac{m}{2}}}{m!} h^m_\Upsilon |\varphi|_{C^m(\Upsilon)}, \tag{2.12}$$

where h_Υ is the diameter of Υ. In particular, if Υ is star-shaped with respect to $(\mathbf{z}, s)$, then the following estimate is obtained

$$\left\| \varphi(\mathbf{x}, t) - T^m_{(\mathbf{z},s)}[\varphi](\mathbf{x}, t) \right\|_{L^2(\Upsilon)} \leq \frac{(d+1)^{\frac{m}{2}} |\Upsilon|^{\frac{1}{2}}}{m!} h^m_\Upsilon |\varphi|_{C^m(\Upsilon)}, \tag{2.13}$$

which, together with the well-known identity (see [12, Prop. (4.1.17)])

$$D^{\mathbf{j}} T^m_{(\mathbf{z},s)}[\varphi] = T^{m-|\mathbf{j}|}_{(\mathbf{z},s)}[D^{\mathbf{j}} \varphi], \qquad |\mathbf{j}| < m, \tag{2.14}$$

gives the estimate

$$\left| \varphi - T^m_{(\mathbf{z},s)}[\varphi] \right|_{H^r(\Upsilon)} \leq \binom{d+r}{d}^{\frac{1}{2}} \frac{(d+1)^{\frac{m-r}{2}} |\Upsilon|^{\frac{1}{2}}}{(m-r)!} h^{m-r}_\Upsilon |\varphi|_{C^m(\Upsilon)} \quad r < m, \; \forall \varphi \in C^m(\Upsilon). \tag{2.15}$$

The Bramble–Hilbert lemma provides an estimate for the error of the averaged Taylor polynomial, see [26] and [12, Thm. 4.3.8].

Lemma 2.11 (Bramble–Hilbert)**.** *Let $\Upsilon \subset \mathbb{R}^{d+1}$, $1 \le d \in \mathbb{N}$, be an open and bounded set with diameter h_Υ, star-shaped with respect to the ball $B := B_{\rho h_\Upsilon}(\mathbf{z}, s)$ centered at $(\mathbf{z}, s) \in \Upsilon$ and with radius ρh_Υ, for some $0 < \rho \le \frac{1}{2}$. If $\varphi \in H^m(\Upsilon)$, the averaged Taylor polynomial of order m (and degree $m - 1$) defined as*

$$Q^m [\varphi] (\mathbf{x}, t) := \frac{1}{|B|} \int_B T^m_{(\mathbf{z},s)} [\varphi] (\mathbf{x}, t) \, \mathrm{d}V(\mathbf{z}, s), \tag{2.16}$$

satisfies the error bound

$$|\varphi - Q^m [\varphi]|_{H^s(\Upsilon)} \le C_{d,m,\rho} h_\Upsilon^{m-s} |\varphi|_{H^m(\Upsilon)} \le 2 \binom{d+s}{d} \frac{(d+1)^{m-s}}{(m-s-1)!} \frac{h_\Upsilon^{m-s}}{\rho^{\frac{d+1}{2}}} |\varphi|_{H^m(\Upsilon)} \quad \text{for } s < m. \tag{2.17}$$

A sharp bound on $C_{d,m,\rho} > 0$ is given in [26, p. 986] in dependence of d, s, m and ρ, and the second bound is proven in [80, Lemma 1].

From [12, Prop. (4.1.17)], we also have the identity

$$D^j Q^m [\varphi] = Q^{m-|j|} \left[D^j \varphi \right], \qquad |j| < m. \tag{2.18}$$

Chapter 3

A polynomial Trefftz space for zero potential $(V = 0)$

Contents

3.1   Introduction

When all the derivatives in a constant-coefficients differential operator are of the same order (e.g., the Laplace's equation, the wave equation, the time-dependent Maxwell's equations), the space of Trefftz polynomials of a certain maximum degree delivers the same convergence rates as the full polynomial space of the same degree, see [80, Lemma 1]. This is due to the fact that, by the identity (2.18), the averaged Taylor polynomial of the exact solution belongs to the polynomial Trefftz space, thus ensuring good approximation properties. On the contrary, when the differential operator includes derivatives of different orders (e.g., the Helmholtz's equation, the time-harmonic Maxwell's equation, the Schrödinger equation), non-polynomial spaces have been used in the literature, as no Trefftz subspace of polynomials of a given degree delivers the same accuracy as the full polynomial space of the same degree. For their simple shape and closed-form integration formulas, the preferred type of non-polynomial Trefftz spaces for time-harmonic problems are plane waves, see [51, 43]. Unfortunately, the use of plane waves Trefftz spaces brings some consequences: ill-conditioned matrices in the resulting linear system; lack of general tools for their analysis, and therefore need of developing novel tools for each problem; dependence of the design of such spaces on the physical dimension of the problem; dimension-dependent conditions on the parameters defining their basis functions in order to preserve accuracy. The use of polynomial Trefftz spaces would mitigate or overcome these issues.

In this chapter, we show that optimal *a priori* error estimates with polynomial Trefftz spaces can be obtained for a Trefftz-DG discretization of a PDE that includes derivatives of different orders, namely the time-dependent Schrödinger equation (1.1) in d space dimensions with zero potential, i.e., $V = 0$. For the variational formulation (2.3), we prove that the same asymptotic accuracy as for full polynomials of maximum degree p in $\mathbb{R}^{d+1}$, is obtained if Trefftz polynomials of degree $2p$ in $\mathbb{R}^{d+1}$ are used. We also propose a practical way to construct an explicit basis

for the polynomial Trefftz space in arbitrary dimensions and prove that its dimension is equal to that of the space of polynomials of degree $2p$ in $\mathbb{R}^d$.

Structure of the chapter: In Section 3.2, we present the reduced version of the space–time ultra-weak DG variational formulation (2.3) for discrete Trefftz spaces. In Section 3.3, we prove optimal h-convergence error estimates of order $O(h^p)$ in a mesh-dependent norm, for the space of Trefftz polynomials of degree $2p$ in $\mathbb{R}^d$. The error analysis is based on the approximation properties of some extended Taylor polynomials. A simple construction of a basis for the polynomial Trefftz space is presented in Section 5.3.1. In Section 3.5, we validate the error estimate, and assess numerically some additional features of the method.

3.2 Trefftz-DG method

We define the local and global Trefftz spaces:

$$\mathbf{T}(K) := \left\{ w \in H^1\left(K_t; L^2(K_\mathbf{x})\right) \cap L^2\left(K_t; H^2(K_\mathbf{x})\right) \text{ such that } \right.$$

$$\left. i\epsilon \partial_t w + \frac{\epsilon^2}{2}\Delta_\mathbf{x} w - V|_K w = 0 \text{ on } K = K_\mathbf{x} \times K_t \right\}, \tag{3.1}$$

$$\mathbf{T}(\mathcal{T}_h) := \left\{ w \in L^2\left(Q_T\right)^{d+1} \,\middle|\, w|_K \in \mathbf{T}(K), \; \forall K \in \mathcal{T}_h(Q_T) \right\}.$$

For any finite-dimensional subspace $\mathbb{T}_{hp}(\mathcal{T}_h) \subset \mathbf{T}(\mathcal{T}_h)$, the space–time ultra-weak DG variational formulation (2.3) reduces to

$$\text{Seek } \psi_{hp} \in \mathbb{T}_{hp}(\mathcal{T}_h) \text{ such that: } \mathcal{A}\left(\psi_{hp}; s_{hp}\right) = \ell(s_{hp}) \quad \forall s_{hp} \in \mathbb{T}_{hp}(\mathcal{T}_h), \tag{3.2}$$

where

$$\mathcal{A}\left(\psi_{hp}; s_{hp}\right) := i\epsilon \left(\int_{\mathcal{F}_h^{\text{space}}} \psi_{hp}^- \llbracket \overline{s_{hp}} \rrbracket_t \, d\mathbf{x} + \int_{\mathcal{F}_h^T} \psi_{hp} \overline{s_{hp}} d\mathbf{x} \right) + \frac{\epsilon^2}{2} \int_{\mathcal{F}_h^{\text{time}}} \left(\{\!\!\{ \nabla_\mathbf{x} \psi_{hp} \}\!\!\} \cdot \llbracket \overline{s_{hp}} \rrbracket_\mathbf{N} \right.$$

$$+ i\alpha \llbracket \psi_{hp} \rrbracket_\mathbf{N} \cdot \llbracket \overline{s_{hp}} \rrbracket_\mathbf{N} - \{\!\!\{ \psi_{hp} \}\!\!\} \llbracket \nabla_\mathbf{x} \overline{s_{hp}} \rrbracket_\mathbf{N} + i\beta \llbracket \nabla_\mathbf{x} \psi_{hp} \rrbracket_\mathbf{N} \llbracket \nabla_\mathbf{x} \overline{s_{hp}} \rrbracket_\mathbf{N} \left. \right) dS$$

$$+ \frac{\epsilon^2}{2} \int_{\mathcal{F}_h^{\text{D}}} \left(\partial_{\mathbf{n}_\mathbf{x}} \psi_{hp} + i\alpha \psi_{hp} \right) \overline{s_{hp}} dS$$

$$+ \frac{\epsilon^2}{2} \int_{\mathcal{F}_h^{\text{N}}} \left(-\psi_{hp} \partial_{\mathbf{n}_\mathbf{x}} \overline{s_{hp}} + i\beta \left(\partial_{\mathbf{n}_\mathbf{x}} \psi_{hp} \right) \left(\partial_{\mathbf{n}_\mathbf{x}} \overline{s_{hp}} \right) \right) dS$$

$$+ \frac{\epsilon^2}{2} \int_{\mathcal{F}_h^{\text{R}}} \left(\delta \partial_{\mathbf{n}_\mathbf{x}} \psi_{hp} + (1-\delta) i\vartheta \psi_{hp} \right) \left(\overline{s_{hp}} + \frac{i}{\vartheta} \partial_{\mathbf{n}_\mathbf{x}} \overline{s_{hp}} \right) dS,$$

$$\ell(s_{hp}) := i\epsilon \int_{\mathcal{F}_h^0} \psi_0 \overline{s_{hp}} d\mathbf{x} + \frac{\epsilon^2}{2} \int_{\mathcal{F}_h^{\text{D}}} g_{\text{D}} \left(\partial_{\mathbf{n}_\mathbf{x}} \overline{s_{hp}} + i\alpha \overline{s_{hp}} \right) dS$$

$$+ \frac{\epsilon^2}{2} \int_{\mathcal{F}_h^{\text{N}}} g_{\text{N}} \left(-\overline{s_{hp}} + i\beta \partial_{\mathbf{n}_\mathbf{x}} \overline{s_{hp}} \right) dS + \frac{\epsilon^2}{2} \int_{\mathcal{F}_h^{\text{R}}} g_{\text{R}} \left((\delta - 1) \overline{s_{hp}} + \frac{i\delta}{\vartheta} \partial_{\mathbf{n}_\mathbf{x}} \overline{s_{hp}} \right) dS.$$

Due to the Trefftz property of the space $\mathbb{T}_{hp}(\mathcal{T}_h)$, the volume integrals in the variational formulation (2.3) vanish, and thus the Trefftz-DG method (3.2) only involves integrals on the mesh skeleton.

All the abstract theory in Section 2.4 follows after removing the volume terms in the definition of the $||| \cdot |||_{\text{DG}}$ and $||| \cdot |||_{\text{DG}^+}$ norms. More precisely,

$$
|||w|||^2_{\text{DG}} := \frac{\epsilon}{2} \left(\|[\![w]\!]_t\|^2_{L^2(\mathcal{F}_h^{\text{space}})} + \|w\|^2_{L^2(\mathcal{F}_h^T \cup \mathcal{F}_h^0)} \right) + \frac{\epsilon^2}{2} \left(\left\| \alpha^{\frac{1}{2}} [\![w]\!]_{\mathbf{N}} \right\|^2_{L^2(\mathcal{F}_h^{\text{time}})^d} \right.
$$

$$
+ \left\| \beta^{\frac{1}{2}} [\![\nabla_{\mathbf{x}} w]\!]_{\mathbf{N}} \right\|^2_{L^2(\mathcal{F}_h^{\text{time}})} + \left\| \alpha^{\frac{1}{2}} w \right\|^2_{L^2(\mathcal{F}_h^{\text{D}})} + \left\| \beta^{\frac{1}{2}} \partial_{\mathbf{n}_{\mathbf{x}}} w \right\|^2_{L^2(\mathcal{F}_h^{\text{N}})}
$$

$$
\left. + \left\| (\vartheta(1-\delta))^{\frac{1}{2}} w \right\|^2_{L^2(\mathcal{F}_h^{\text{R}})} + \left\| (\delta \vartheta^{-1})^{\frac{1}{2}} \partial_{\mathbf{n}_{\mathbf{x}}} w \right\|^2_{L^2(\mathcal{F}_h^{\text{R}})} \right),
$$

$$
|||w|||^2_{\text{DG}^+} := |||w|||^2_{\text{DG}} + \frac{\epsilon}{2} \|w^-\|^2_{L^2(\mathcal{F}_h^{\text{space}})} + \frac{\epsilon^2}{2} \left(\left\| \alpha^{-\frac{1}{2}} \{\!\!\{ \nabla_{\mathbf{x}} w \}\!\!\} \right\|^2_{L^2(\mathcal{F}_h^{\text{time}})^d} \right.
$$

$$
\left. + \left\| \alpha^{-\frac{1}{2}} \partial_{\mathbf{n}_{\mathbf{x}}} w \right\|^2_{L^2(\mathcal{F}_h^{\text{D}})} + \left\| \beta^{-\frac{1}{2}} \{\!\!\{ w \}\!\!\} \right\|^2_{L^2(\mathcal{F}_h^{\text{time}})} + \left\| \beta^{-\frac{1}{2}} w \right\|^2_{L^2(\mathcal{F}_h^{\text{N}})} + \left\| (\delta^{-1} \vartheta)^{\frac{1}{2}} w \right\|^2_{L^2(\mathcal{F}_h^{\text{R}})} \right).
$$

3.3 *A priori* error estimates

We recall that, combining Proposition 2.8 and Theorem 2.7 in Chapter 2, it is possible to derive error estimates for the Trefftz-DG method (3.2) by simply studying the approximation properties of the local discrete spaces $\mathbb{T}_{hp}(K)$ for functions in $\mathbf{T}(\mathcal{T}_h)$. In particular, our analysis for the polynomial Trefftz space relies on the approximation properties of some Taylor-type polynomials.

Definition 3.1 (Polynomial Trefftz space). On each $K \in \mathcal{T}_h$, the space of Trefftz polynomials of maximum degree p is defined as

$$
\mathbb{P}^p_{\mathbf{T}}(K) := \mathbf{T}(K) \cap \mathbb{P}^p(K) = \left\{ q_p \in \mathbb{P}^p(K) \ : \ i\epsilon \partial_t q_p + \frac{\epsilon^2}{2} \Delta_{\mathbf{x}} q_p = 0 \right\}. \tag{3.3}
$$

It would be desirable to prove that, for all $\psi \in \mathbf{T}(K)$, either $T^{p+1}_{(\mathbf{z},s)}[\psi] \in \mathbb{P}^p_{\mathbf{T}}(K)$ for some $(\mathbf{z}, s) \in K$ or $Q^{p+1}[\psi] \in \mathbb{P}^p_{\mathbf{T}}(K)$, as this would be enough to achieve optimal h-convergence rates for sufficiently smooth solutions. Unfortunately, this is not true; e.g., for $d = 1$, $\epsilon = 1$ and $p = 1$ the function $\psi(x, t) = \exp\left(x + \frac{it}{2}\right)$ satisfies the Schrödinger equation (1.1); however, $T^2_{(0,0)}[\psi] = 1 + x + \frac{it}{2}$ does not belong to $\mathbb{P}^1_{\mathbf{T}}(K)$. Instead, in Proposition 3.5 below, we show that the following extended Taylor polynomials belong to $\mathbb{P}^{2p}_{\mathbf{T}}(K)$.

Definition 3.2 (Extended Taylor polynomial). Given $p \in \mathbb{N}$, and an open and bounded set $\Upsilon = \Upsilon_{\mathbf{x}} \times \Upsilon_t \subset \mathbb{R}^{d+1}$, for each complex-valued $\varphi \in C^p(\Upsilon_t; C^{2p}(\Upsilon_{\mathbf{x}}))$ we define its extended Taylor polynomial of degree $2p$ centered at some $(\mathbf{z}, t) \in \Upsilon$ as

$$
\widetilde{T}^{2p}_{(\mathbf{z},t)}[\varphi] = T^{p+1}_{(\mathbf{z},t)}[\varphi] + \sum_{\substack{2j_t + |\mathbf{j}_{\mathbf{x}}| \leq 2p \\ p+1 \leq j_t + |\mathbf{j}_{\mathbf{x}}|}} \frac{1}{\mathbf{j}!} D^{\mathbf{j}} \varphi(\mathbf{z}, s)(\mathbf{x} - \mathbf{z})^{\mathbf{j}_{\mathbf{x}}} (t - s)^{j_t} \qquad \mathbf{j} = (\mathbf{j}_{\mathbf{x}}, j_t) \in \mathbb{N}^{d+1}.
$$

Definition 3.3 (Extended averaged Taylor polynomial). Let $\Upsilon = \Upsilon_{\mathbf{x}} \times \Upsilon_t \subset \mathbb{R}^{d+1}$ be an open and bounded set with diameter h_Υ, star-shaped with respect to the ball $B := B_{\rho h_\Upsilon}(\mathbf{z}, s)$ centered at $(\mathbf{z}, s) \in \Upsilon$ and with radius ρh_Υ, for some $\rho \in (0, \frac{1}{2})$. For each complex-valued $\varphi \in H^p\left(\Upsilon_t; H^{2p}(\Upsilon_{\mathbf{x}})\right)$, we define its extended averaged Taylor polynomial of degree $2p$ as

$$
\widetilde{Q}^{2p}[\varphi](\mathbf{x}, t) := \frac{1}{|B|} \int_B \widetilde{T}^{2p}_{(\mathbf{z},s)}[\varphi](\mathbf{x}, t)\, \mathrm{d}V(\mathbf{z}, s)
$$

$$
= Q^{p+1}[\varphi] + \sum_{\substack{2j_t + |\boldsymbol{j}_{\mathbf{x}}| \leq 2p \\ p+1 \leq j_t + |\boldsymbol{j}_{\mathbf{x}}|}} \frac{1}{\boldsymbol{j}! |B|} \int_B D^{\boldsymbol{j}} \varphi(\mathbf{z}, s)(\mathbf{x} - \mathbf{z})^{\boldsymbol{j}_{\mathbf{x}}} (t - s)^{j_t}\, \mathrm{d}V(\mathbf{z}, s),
$$

where $Q^{p+1}[\cdot]$ is the standard averaged Taylor polynomial defined in (2.16).

Remark 3.4 (Heat polynomials). In the real-valued case, these extended Taylor polynomials can be used to analyze the approximation properties of the so-called heat polynomials, i.e., the polynomial solutions to the heat equation [89]. In fact, Proposition 3.5 below can be easily extended to the heat equation. However, as the heat operator $\mathcal{L}(\cdot) := (\partial_t - \Delta_{\mathbf{x}})(\cdot)$ is not self-adjoint, the setting of Theorem 2.7 cannot be used to analyze the ultra-weak DG formulation of the heat equation, see Remark 2.3 in Chapter 2. This is due to the fact that the corresponding bilinear form is not coercive. ∎

Proposition 3.5. *Let Υ be an open and bounded set in $\mathbb{R}^{d+1}$ satisfying the assumptions in Definition 3.3. For all $\psi \in \mathbf{T}(\Upsilon)$ sufficiently regular, both $\widetilde{T}^{2p}_{(\mathbf{z},s)}[\psi]$ and $\widetilde{Q}^{2p}[\psi]$ belong to $\mathbb{P}^{2p}_{\mathbf{T}}(\Upsilon)$.*

Proof. Denote by $\{\boldsymbol{e}_\ell\}_{\ell=1}^d$ the canonical basis of $\mathbb{R}^d$. Since $\psi \in \mathbf{T}(\Upsilon)$, we have that

$$
\epsilon \sum_{\ell=1}^d D^{(\boldsymbol{j}_{\mathbf{x}}+2\boldsymbol{e}_\ell, j_t)} \psi(\mathbf{z}, s) = -2i D^{(\boldsymbol{j}_{\mathbf{x}}, j_t+1)} \psi(\mathbf{z}, s) \qquad \forall \boldsymbol{j} = (\boldsymbol{j}_{\mathbf{x}}, j_t) \in \mathbb{N}^{d+1},\ |\boldsymbol{j}| \geq 0. \quad (3.4)
$$

Applying this to the Taylor expansion of ψ and using straightforward multi-index manipulations, we get

$$
-2i\partial_t T^{p+1}_{(\mathbf{z},s)}[\psi](\mathbf{x}, t) = \sum_{|\boldsymbol{j}| \leq p-1} \frac{-2i}{\boldsymbol{j}!} D^{(\boldsymbol{j}_{\mathbf{x}}, j_t+1)} \psi(\mathbf{z}, s)(\mathbf{x} - \mathbf{z})^{\boldsymbol{j}_{\mathbf{x}}} (t - s)^{j_t}
$$

$$
\overset{(3.4)}{=} \sum_{|\boldsymbol{j}| \leq p-1} \sum_{\ell=1}^d \frac{\epsilon}{\boldsymbol{j}!} D^{(\boldsymbol{j}_{\mathbf{x}}+2\boldsymbol{e}_\ell, j_t)} \psi(\mathbf{z}, s)(\mathbf{x} - \mathbf{z})^{\boldsymbol{j}_{\mathbf{x}}} (t - s)^{j_t}
$$

$$
= \epsilon \Delta_{\mathbf{x}} T^{p+1}_{(\mathbf{z},s)}[\psi](\mathbf{x}, t) + \sum_{|\boldsymbol{j}|=p-1} \sum_{\ell=1}^d \frac{\epsilon}{\boldsymbol{j}!} D^{(\boldsymbol{j}_{\mathbf{x}}+2\boldsymbol{e}_\ell, j_t)} \psi(\mathbf{z}, s)(\mathbf{x} - \mathbf{z})^{\boldsymbol{j}_{\mathbf{x}}} (t - s)^{j_t}.
$$

$$(3.5)$$

For the extended Taylor polynomial we have that

$$
\epsilon \Delta_{\mathbf{x}} \widetilde{T}^{2p}_{(\mathbf{z},s)}[\psi](\mathbf{x}, t) = \epsilon \Delta_{\mathbf{x}} T^{p+1}_{(\mathbf{z},s)}[\psi](\mathbf{x}, t) + \sum_{\substack{2j_t + |\boldsymbol{j}_{\mathbf{x}}| \leq 2p-2 \\ p-1 \leq |\boldsymbol{j}|}} \sum_{\ell=1}^d \frac{\epsilon}{\boldsymbol{j}!} D^{(\boldsymbol{j}_{\mathbf{x}}+2\boldsymbol{e}_\ell, j_t)} \psi(\mathbf{z}, s)(\mathbf{x} - \mathbf{z})^{\boldsymbol{j}_{\mathbf{x}}} (t - s)^{j_t}
$$

$$= \epsilon \Delta_{\mathbf{x}} T^{p+1}_{(\mathbf{z},s)}[\psi](\mathbf{x},t) + \sum_{|j|=p-1} \sum_{\ell=1}^{d} \frac{\epsilon}{j!} D^{(j_{\mathbf{x}}+2e_\ell, j_t)} \psi(\mathbf{z},s)(\mathbf{x}-\mathbf{z})^{j_{\mathbf{x}}}(t-s)^{j_t}$$

$$+ \sum_{\substack{2j_t+|j_{\mathbf{x}}|\leq 2p-2 \\ p\leq |j|}} \sum_{\ell=1}^{d} \frac{\epsilon}{j!} D^{(j_{\mathbf{x}}+2e_\ell, j_t)} \psi(\mathbf{z},s)(\mathbf{x}-\mathbf{z})^{j_{\mathbf{x}}}(t-s)^{j_t}$$

$$\overset{(3.4)}{\underset{(3.5)}{=}} -2i\partial_t T^{p+1}_{(\mathbf{z},s)}[\psi](\mathbf{x},t) - \sum_{\substack{2j_t+|j_{\mathbf{x}}|\leq 2p-2 \\ p\leq |j|}} \frac{2i}{j!} D^{(j_{\mathbf{x}}, j_t+1)} \psi(\mathbf{z},s)(\mathbf{x}-\mathbf{z})^{j_{\mathbf{x}}}(t-s)^{j_t}$$

$$= -2i\partial_t \widetilde{T}^{2p}_{(\mathbf{z},s)}[\psi](\mathbf{x},t),$$

and therefore $\widetilde{T}^{2p}_{(\mathbf{z},s)}[\psi] \in \mathbf{T}(\Upsilon)$. Since $\widetilde{T}^{2p}_{(\mathbf{z},s)}[\psi]$ also belongs to $\mathbb{P}^{2p}(\Upsilon)$, then $\widetilde{T}^{2p}_{(\mathbf{z},s)}[\psi] \in \mathbb{P}^{2p}_{\mathbf{T}}(\Upsilon)$. As the above identity is independent of the center point $(\mathbf{z},s)$, the fact that $\widetilde{Q}^{2p}[\psi] \in \mathbb{P}^{2p}_{\mathbf{T}}(\Upsilon)$ can be proven in a similar way. $\qquad\square$

Combining Theorem 2.7, Propositions 2.8 and 3.5, together with error bounds for extended averaged Taylor polynomials, which can be proven using that $\widetilde{Q}^{2p}[\psi] = Q^{p+1}[\psi] + O(h_K^{p+1})$, the following error estimate is obtained.

Theorem 3.6. *Let $p \in \mathbb{N}$, and fix α, β, δ and μ as in (2.2). Let ψ be the exact solution to (1.1) with $V = 0$, and $\psi_{hp} \in \mathbb{T}_{hp}(\mathcal{T}_h)$ be the solution to the Trefftz-DG method (3.2) with $\mathbb{T}_{hp}(\mathcal{T}_h) := \prod_{K\in\mathcal{T}_h} \mathbb{P}^{2p}_{\mathbf{T}}(K)$. If $\psi|_K \in H^{p+1}(K_t; H^{2p}(K_{\mathbf{x}}))$ and $h_{K_{\mathbf{x}}} \simeq h_{K_t}$ for all $K = K_{\mathbf{x}} \times K_t \in \mathcal{T}_h$, then there exists a positive constant C independent of the mesh size h, but depending on the degree p, the parameter ϵ, the local quasi-uniformity parameter $\mathsf{lqu}(\mathcal{T}_h)$, and the measure of the space–time domain Q_T such that*

$$|||\psi - \psi_{hp}|||_{\mathrm{DG}} \leq C \sum_{K\in\mathcal{T}_h} \max\{h_{K_{\mathbf{x}}}, h_{K_t}\}^p \|\psi\|_{H^{p+1}(K_t; H^{2p}(K_{\mathbf{x}}))}.$$

Remark 3.7 (Regularity assumptions)**.** Theorem 3.6 guarantees optimal h-convergence rates, provided that, for $p > 1$, the solution satisfy stronger regularity assumptions than those required for the complex pseudo-plane wave functions in Chapter 4, see Theorem 4.4. $\qquad\blacksquare$

Remark 3.8 (Zero potential)**.** There are no nontrivial polynomial solutions to (1.1a) for $V \neq 0$. Indeed, if $0 \neq q_p \in \mathbb{P}^p_{\mathbf{T}}(K)$ and $V \in \mathbb{R}\setminus\{0\}$, then $\deg(i\epsilon\partial_t q_p + \frac{\epsilon^2}{2}\Delta_{\mathbf{x}} q_p) < \deg(V q_p)$. Therefore, $\mathcal{S}_\epsilon q_p = 0$ if and only if $q_p = 0$. $\qquad\blacksquare$

3.4 Basis and dimension of the local polynomial Trefftz space

We study the dimension of the space $\mathbb{P}^{2p}_{\mathbf{T}}(K)$ and provide a practical way to construct a basis, which is valid for any space dimension. Given $p \in \mathbb{N}$ and $K = K_{\mathbf{x}} \times K_t \in \mathcal{T}_h$, let $\{m_J\}_{J=1}^{r_{d,2p}}$ be a basis for the space $\mathbb{P}^{2p}(K_{\mathbf{x}})$, where $r_{d,2p} = \dim(\mathbb{P}^{2p}(\mathbb{R}^d)) = \binom{2p+d}{d}$. For a fixed $t_K \in K_t$, we consider the following set of functions in $\mathbb{P}^{2p}_{\mathbf{T}}(K)$

$$\mathcal{B}^{\mathbf{T}}_{2p} := \left\{ b^{\mathbf{T}}_J \in \mathbb{P}^{2p}_{\mathbf{T}}(K) : b_J(\mathbf{x}, t_K) = m_J(\mathbf{x}), \ J = 1, \ldots, r_{d,2p} \right\}. \tag{3.6}$$

Proposition 3.9. *For any $p \in \mathbb{N}$ and $K \in \mathcal{T}_h$, the set $\mathcal{B}_{2p}^{\mathbf{T}}$ constitutes a basis for the space $\mathbb{P}_{\mathbf{T}}^{2p}(K)$.*

Proof. Let $q_{2p}^{\mathbf{T}} \in \mathbb{P}_{\mathbf{T}}^{2p}(K)$. Then, $q_{2p}^{\mathbf{T}}$ can be expressed in the scaled monomial basis as

$$q_{2p}^{\mathbf{T}}(\mathbf{x}, t) = \sum_{|\boldsymbol{j}| \leq 2p} C_{\boldsymbol{j}} \left(\frac{\mathbf{x} - \mathbf{x}_K}{h_{K_{\mathbf{x}}}} \right)^{\boldsymbol{j}_{\mathbf{x}}} \left(\frac{t - t_K}{h_{K_t}} \right)^{j_t}.$$

Since $q_{2p}^{\mathbf{T}} \in \mathbf{T}(K)$, the complex coefficients $\{C_{\boldsymbol{j}}\}_{|\boldsymbol{j}| \leq 2p}$ must satisfy the following recurrence relations

$$C_{\boldsymbol{j}_{\mathbf{x}}, j_t + 1} = \begin{cases} \frac{i \epsilon h_{K_t}}{2(j_t + 1) h_{K_{\mathbf{x}}}^2} \sum_{\ell=1}^{d} (\boldsymbol{j}_{\mathbf{x}_\ell} + 1)(\boldsymbol{j}_{\mathbf{x}_\ell} + 2) C_{\boldsymbol{j}_{\mathbf{x}} + 2e_\ell, j_t} & \text{if } |\boldsymbol{j}_{\mathbf{x}}| \leq 2p - 2, \\ 0 & \text{otherwise.} \end{cases} \tag{3.7}$$

As a consequence, once the coefficients $C_{\boldsymbol{j}_{\mathbf{x}}, 0}$ are fixed, the remaining coefficients are determined by (3.7). This implies that each element in $\mathbb{P}_{\mathbf{T}}^{2p}(K)$ is uniquely identified by its restriction to $t = t_K$.

On the other hand, by the definition of $\mathcal{B}_{2p}^{\mathbf{T}}$, there exist some coefficients $\{\gamma_J\}_{J=1}^{r_{d,2p}} \subset \mathbb{C}$ such that

$$q_{2p}^{\mathbf{T}}(\mathbf{x}, t_K) = \sum_{J=1}^{r_{d,2p}} \gamma_J m_J(\mathbf{x}) = \sum_{J=1}^{r_{d,2p}} \gamma_J b_J(\mathbf{x}, t_K) \implies q_{2p}^{\mathbf{T}} = \sum_{J=1}^{r_{d,2p}} \gamma_J b_J.$$

Finally, the linear independence of the set $\mathcal{B}_{2p}^{\mathbf{T}}$ follows from the restriction of its elements to $t = t_K$, and the linear independence of the basis $\{m_J\}_{J=1}^{r_{d,2p}}$. $\qquad \square$

To construct a basis for the local polynomial Trefftz space $\mathbb{P}_{\mathbf{T}}^{p}(K)$ one can now employ the recurrence relation (3.7). To initialize the recursion, one chooses a basis $\{m_J\}_{J=1}^{r_{d,2p}}$ of $\mathbb{P}^{2p}(K_{\mathbf{x}})$. Natural choices are (scaled and/or translated) monomials, Legendre, or Chebyshev polynomials. For instance, if $d = 1$, $\epsilon = 1$, and we set $\{m_J\}_{J=1}^{r_{1,2p}}$ as monomials, then $\mathbb{P}_{\mathbf{T}}^{2} = \text{span}\{1, x, it + x^2\}$ and $\mathbb{P}_{\mathbf{T}}^{4} = \text{span}\{1, x, it + x^2, x^3 + 3ixt, x^4 + 6ix^2t - 3t^2\}$.

Remark 3.10 (Dimension of $\mathbb{P}_{\mathbf{T}}^{2p}(K)$). For $d = 1$, the dimension of the polynomial Trefftz space of degree $2p$ is always smaller than that of the full polynomial space of degree p. However, for $d > 1$, the reduction in the total number of degrees of freedom takes place only for large values of p, e.g., $p > 7$ for $d = 2$ and $p > 24$ for $d = 3$. $\qquad \blacksquare$

3.5   Numerical results

We validate the error estimate of Theorem 3.6, and assess numerically some additional features of the method.

3.5.1   Smooth solution

On the space–time domain $Q_T = (0, 1) \times (0, 1)$ we consider a manufactured $(1+1)$-dimensional Schrödinger equation with initial and Dirichlet boundary conditions so that the exact solution is given by the complex wave function $\psi(x, t) = \exp\left(\frac{i}{\epsilon}(x - \frac{t}{2})\right)$, which is highly oscillatory

in space and time for $\epsilon \ll 1$. The Trefftz basis functions $\{b_J^{\mathbf{T}}\}_{J=1}^{r_{d,2p}}$ are constructed by choosing $\{m_J\}_{J=1}^{r_{2p,d}}$ in (3.6) as scaled monomials and by computing the remaining coefficients C_j with the relations (3.7).

In Figure 3.1, we show the errors of method (3.2) in the DG norm obtained for $\epsilon = 10^{-1}$, 10^{-2} and a sequence of meshes with $h_t = h_x = 0.1 \times 2^{-j}$, $j = 1, \ldots, 5$. Optimal rates of convergence of order $O(h^p)$ are obtained as predicted by Theorem 3.6. Observe that, for small values of ϵ, there is a preasymptotic behaviour due to the oscillations of the exact solution with frequency $1/\epsilon$ in space and time. In Figure 3.2, we observe exponential convergence of order $O\left(e^{-bN_{DoFs}}\right)$ for the p-version of the method, i.e., by fixing the space–time mesh and increasing the degree of accuracy p.

(a) $\epsilon = 10^{-1}$ **(b)** $\epsilon = 10^{-2}$
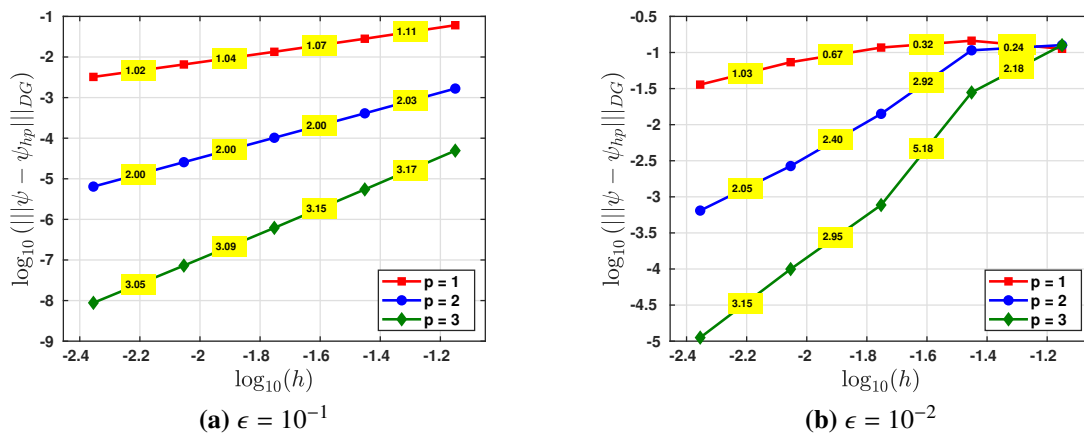
Figure 3.1 h-convergence of the Trefftz-DG method (3.2) for the $(1+1)$-dimensional problem with exact solution $\psi(x,t) = \exp\left(\frac{i}{\epsilon}(x - \frac{t}{2})\right)$.

(a) $\epsilon = 10^{-1}$ **(b)** $\epsilon = 10^{-2}$
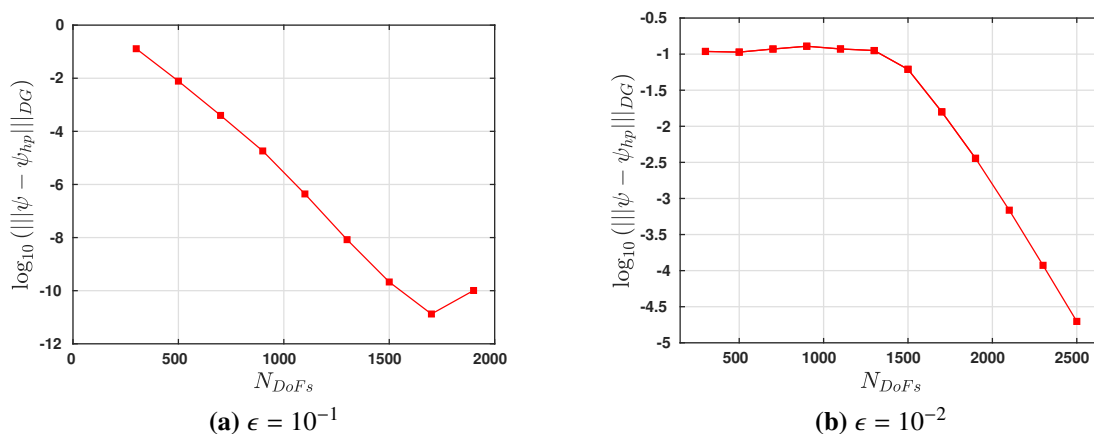
Figure 3.2 p-convergence of the Trefftz-DG method (3.2) for the $(1+1)$-dimensional problem with exact solution $\psi(x,t) = \exp\left(\frac{i}{\epsilon}(x - \frac{t}{2})\right)$ for a Cartesian space–time mesh with $h_x = h_t = 0.1$.

The κ_2-condition number of the stiffness matrix is studied numerically in Section 5.5.5.

Chapter 4

A pseudo-plane wave Trefftz space for piecewise-constant potentials

Contents

4.1   Introduction

In this chapter, we focus on the time-dependent Schrödinger equation (1.1) with piecewise-constant potential V. We construct Trefftz spaces consisting of simple complex pseudo-plane wave functions, which we use as discrete spaces in the Trefftz-DG variational formulation (3.2). More precisely, assuming that the space–time mesh $\mathcal{T}_h$ is aligned to the discontinuities in the potential V, the local discrete subspace $\mathbb{T}^p(K)$ of the local Trefftz space $\mathbf{T}(K)$ in (3.1) is defined for each $K = K_{\mathbf{x}} \times K_t \in \mathcal{T}_h$ and for $p \in \mathbb{N}$, as the following set of complex exponentials:

$$\mathbb{T}^p(K) := \operatorname{span}\left\{\phi_\ell(\mathbf{x}, t), \ \ell = 1, \ldots, n_{d+1,p}\right\}, \tag{4.1}$$

$$\text{where} \qquad \phi_\ell(\mathbf{x}, t) := \exp\left[\frac{i}{\epsilon}\left(k_\ell \mathbf{d}_\ell^\top \mathbf{x} - \left(\frac{k_\ell^2}{2} + V|_K\right)t\right)\right] \quad \text{for } \ell = 1, \ldots, n_{d+1,p},$$

for some parameters $\{k_\ell\} \subset \mathbb{R}$ and directions $\{\mathbf{d}_\ell\} \subset \mathcal{S}_1^d := \{\mathbf{v} \in \mathbb{R}^d, |\mathbf{d}| = 1\}$, which can be chosen differently in each cell K. These exponential functions are called "pseudo-plane waves" in the context of Fresnel optics in [46, Eqn. (11)].

We choose the local dimension $n_{d+1,p}$ in dependence of the space dimension d of the problem and a "degree" parameter p. The parameter p is to be understood as the degree of the polynomial space $\mathbb{P}^p(K)$ providing the same h-convergence rates as $\mathbb{T}^p(K)$. The Trefftz property allows us to construct $\mathbb{T}^p(K)$ with dimension $n_{d+1,p} \ll r_{d+1,p}$, where $r_{d,p}$ denotes the dimension of the space of polynomials of degree at most p in $\mathbb{R}^d$.

The key idea to derive optimal *a priori* error estimates for the considered Trefftz space was introduced by O. Cessenat and B. Després in the proof of [15, Thm. 3.7] (in the case of the ultra-weak variational formulation (UWVF) applied to the Helmholtz equation): given

any smooth PDE solution ψ, if the local discrete space contains an element with the same degree-p Taylor polynomial of ψ, then the space enjoys the same h-approximation properties as the space $\mathbb{P}^p(\mathbb{R}^{d+1})$. We prove that this condition is satisfied by the space $\mathbb{T}^p$ in low space dimensions $d = 1$ and $d = 2$ (we comment on the case $d \geq 3$ in Remark 4.8). This approach to the Trefftz approximation theory is completely different from that used for the Helmholtz equation in [79], which is based on the use of an integral (Vekua) transform and circular wave expansions. A shortcoming of the Taylor-polynomial approach of [15] and here, in contrast to the Vekua-transform technique in [79], is that the approximation analysis does not extend to the p-convergence case. This is a difficult task that, in the context of Trefftz schemes, has been achieved for time-harmonic equations but not yet for the wave equation.

Structure of the chapter: Section 4.2 is devoted to the error analysis for the complex pseudo-plane wave Trefftz spaces $\mathbb{T}^p$ in (4.1). We present a condition that guarantees optimal local approximation of the exact solution in a general discrete Trefftz space. Assuming that such condition is satisfied we prove an h-estimate in a mesh-skeleton norm. In Sections 4.3 and 4.4 we prove that the assumed condition holds for the $(1 + 1)$- and $(2 + 1)$-dimensional cases under some restrictions of the tuning parameters for our basis choice. Some numerical experiments validating our theoretical results are presented in Section 4.5.

4.2 *A priori* error estimates

We now give a condition on discrete Trefftz spaces $\mathbb{T}^p(K)$, and show how it entails optimal h-approximation estimates. In the next sections we prove that such a condition is satisfied by the Trefftz space (4.1).

Condition 4.1 states that, for any sufficiently smooth Schrödinger solution ψ, for each element $K \in \mathcal{T}_h$, and almost every point $(\mathbf{z}, s)$ in a ball $B \subset K$, the discrete Trefftz space contains an element whose Taylor polynomial centered at $(\mathbf{z}, s)$ matches that of ψ. We also require the coefficients $(a_1, \ldots, a_{n_{d+1,p}})$ of this approximant to be uniformly bounded with respect to the points $(\mathbf{z}, s)$ in the aforementioned ball B, so this is a condition on the basis rather than on the discrete space itself. To allow for general $\psi \in H^{p+1}(K)$, whose Taylor polynomial might not be defined everywhere and does not guarantee enough approximation, we match $T^{p+1}_{(\mathbf{z},s)}[\psi]$ in almost every point in a ball B (see Remarks 4.2 and 4.5).

Condition 4.1. Let $B \subset K$ be a $(d + 1)$-dimensional ball such that K is star-shaped with respect to B. Let $\{\phi_1, \ldots, \phi_{n_{d+1,p}}\} \subset C^\infty(K)$ be a basis of $\mathbb{T}^p(K)$. For every $\psi \in \mathbf{T}(K) \cap H^{p+1}(K)$, there exists a vector-valued function $\mathbf{a} \in L^1(B)^{n_{d+1,p}}$ satisfying the following two conditions

$$D^{\boldsymbol{j}}\psi(\mathbf{z}, s) = \sum_{\ell=1}^{n_{d+1,p}} a_\ell(\mathbf{z}, s) D^{\boldsymbol{j}}\phi_\ell(\mathbf{z}, s) \quad \text{for all } |\boldsymbol{j}| \leq p \quad \text{and a.e. } (\mathbf{z}, s) \in B, \tag{4.2a}$$

$$\||\mathbf{a}|_1\|_{L^1(B)} = \int_B \sum_{\ell=1}^{n_{d+1,p}} |a_\ell(\mathbf{z}, s)| \, dV(\mathbf{z}, s) \leq C_\star |K|^{1/2} \|\psi\|_{H^{p+1}(K)}, \tag{4.2b}$$

where $C_\star > 0$ might depend on d, p, and $\{\phi_\ell\}$ but is independent of K and ψ.

Multiplying (4.2a) by $\frac{(\mathbf{x}-\mathbf{z})^{\boldsymbol{j}_x}(t-s)^{j_t}}{\boldsymbol{j}!}$ and summing over $\boldsymbol{j}$, we observe that Condition 4.1 implies that, for every smooth Schrödinger solution ψ, there is an element of the discrete space

$\mathbb{T}^p(K)$ whose Taylor polynomial at $(\mathbf{z}, s)$ coincides with that of ψ:

$$T^{p+1}_{(\mathbf{z},s)}[\psi](\mathbf{x},t) = T^{p+1}_{(\mathbf{z},s)}\left[\sum_{\ell=1}^{n_{d+1,p}} a_\ell(\mathbf{z},s)\phi_\ell\right](\mathbf{x},t) = \sum_{\ell=1}^{n_{d+1,p}} a_\ell(\mathbf{z},s)T^{p+1}_{(\mathbf{z},s)}[\phi_\ell](\mathbf{x},t)$$

$$\text{for a.e. } (\mathbf{z},s) \in B, \quad \forall(\mathbf{x},t) \in K. \qquad (4.3)$$

Remark 4.2. Condition 4.1 requires the point value of the partial derivatives of a Sobolev function in almost every point of B. This is to be understood as follows:

Let $\psi \in H^{p+1}(K)$. By Calderón's extension theorem [78, Thm. A.4], ψ can be extended to a $\widetilde{\psi} \in H^{p+1}(\mathbb{R}^{d+1})$. This allows us to apply [1, Thm. 10.1.4], which implies that there exists a zero-measure set $\Upsilon \subset \mathbb{R}^{d+1}$ such that $\widetilde{\psi}$ admits a "differential of order p" at every point $(\mathbf{z}, s)$ of the complement of Υ. The differential of order p, as defined in [1, Def. 10.1.3], is a polynomial that coincides with the Taylor polynomial $T^{p+1}_{(\mathbf{z},s)}[\widetilde{\psi}]$ whenever this is defined. With a small abuse of notation, this is the polynomial appearing on the left-hand side of (4.3) for all $(\mathbf{z}, s) \in B \setminus \Upsilon$ (while $\phi_\ell \in C^\infty(K)$ so their Taylor polynomials are the classical ones). The partial derivatives $D^j\psi(\mathbf{z}, s)$ in (4.2a) are equal to the partial derivatives of $T^{p+1}_{(\mathbf{z},s)}[\psi]$ evaluated in $(\mathbf{z}, s)$. ∎

Proposition 4.3. *Let $K \in \mathcal{T}_h$ and $\psi \in \mathbf{T}(K) \cap H^{p+1}(K)$, with K satisfying the conditions in Lemma 2.11. Assume that the local space $\mathbb{T}^p(K)$ satisfies Condition 4.1. Then, there exists an element $\Phi \in \mathbb{T}^p(K)$, and $C > 0$, which depends on the constant $C_\star$ in (4.2b) but is independent of h_K and ψ, such that:*

$$\left\|D^j(\psi - \Phi)\right\|_{L^2(K)} \le Ch_K^{p+1-|j|}\|\psi\|_{H^{p+1}(K)} \qquad 0 \le |j| \le p. \qquad (4.4)$$

Proof. Let $\mathbf{a} \in L^1(B)$ be the coefficient function defined by Condition 4.1. We define the discrete Trefftz function $\Phi \in \mathbb{T}^p(K)$ as

$$\Phi(\mathbf{x},t) := \frac{1}{|B|}\int_B \sum_{\ell=1}^{n_{d+1,p}} a_\ell(\mathbf{z},s)\phi_\ell(\mathbf{x},t)\,dV(\mathbf{z},s) = \frac{1}{|B|}\sum_{\ell=1}^{n_{d+1,p}}\left(\int_B a_\ell(\mathbf{z},s)\,dV(\mathbf{z},s)\right)\phi_\ell(\mathbf{x},t).$$

$$(4.5)$$

Since $\Phi \in C^\infty(\overline{K})$, by the triangle inequality, for $|j| \le p$ we have

$$\left\|D^j(\psi - \Phi)\right\|_{L^2(K)} \le \left\|D^j\left(\psi - Q^{p+1}[\psi]\right)\right\|_{L^2(K)} + \left\|D^j\left(Q^{p+1}[\psi] - \Phi\right)\right\|_{L^2(K)}, \qquad (4.6)$$

where $Q^{p+1}[\cdot]$ is the averaged Taylor polynomial of order $p + 1$ defined in (2.16).

The first term is bounded by the Bramble–Hilbert lemma 2.11, while for the second term we first derive a pointwise estimate for all $(\mathbf{x}, t) \in K$ and $|j| \le p$ using the multivariate Taylor's theorem and the fact that $\phi_\ell \in C^\infty(K)$:

$$\left|D^j\left(Q^{p+1}[\psi](\mathbf{x},t) - \Phi(\mathbf{x},t)\right)\right|$$

$$\overset{(2.16),(4.5)}{=} \left|\frac{1}{|B|}\int_B D^j\left(T^{p+1}_{(\mathbf{z},s)}[\psi](\mathbf{x},t) - \sum_{\ell=1}^{n_{d+1,p}} a_\ell(\mathbf{z},s)\phi_\ell(\mathbf{x},t)\right)dV(\mathbf{z},s)\right|$$

$$\overset{(4.3)}{=} \left|\frac{1}{|B|}\int_B \sum_{\ell=1}^{n_{d+1,p}} a_\ell(\mathbf{z},s)D^j\left(T^{p+1}_{(\mathbf{z},s)}[\phi_\ell](\mathbf{x},t) - \phi_\ell(\mathbf{x},t)\right)dV(\mathbf{z},s)\right|$$

32

$$
\overset{(2.14)}{=} \left| \frac{1}{|B|} \int_B \sum_{\ell=1}^{n_{d+1,p}} a_\ell(\mathbf{z}, s) \left(T_{(\mathbf{z},s)}^{p+1-|\boldsymbol{j}|} \left[D^{\boldsymbol{j}} \phi_\ell \right](\mathbf{x}, t) - D^{\boldsymbol{j}} \phi_\ell(\mathbf{x}, t) \right) \mathrm{d}V(\mathbf{z}, s) \right|
$$

$$
\overset{(2.12)}{\le} \frac{(d+1)^{(p+1-|\boldsymbol{j}|)/2}}{|B|\,(p+1-|\boldsymbol{j}|)!} h_K^{p+1-|\boldsymbol{j}|} \max_{\ell=1,\dots,n_{d+1,p}} \left\{ |\phi_\ell|_{C^{p+1}(K)} \right\} \int_B |\mathbf{a}(\mathbf{z}, s)|_1 \, \mathrm{d}V(\mathbf{z}, s)
$$

$$
\overset{(4.2b)}{\le} \frac{|K|^{1/2}(d+1)^{(p+1-|\boldsymbol{j}|)/2}}{|B|(p+1-|\boldsymbol{j}|)!} h_K^{p+1-|\boldsymbol{j}|} \max_{\ell=1,\dots,n_{d+1,p}} \left\{ |\phi_\ell|_{C^{p+1}(K)} \right\} C_\star \, \|\psi\|_{H^{p+1}(K)}.
$$

Then,

$$
\left\| D^{\boldsymbol{j}} \left(Q^{p+1}\left[\psi\right] - \Phi \right) \right\|_{L^2(K)} \le C_\star \frac{|K|}{|B|} \frac{(d+1)^{(p+1-|\boldsymbol{j}|)/2}}{(p+1-|\boldsymbol{j}|)!} h_K^{p+1-|\boldsymbol{j}|} \max_{\ell=1,\dots,n_{d+1,p}} \left\{ |\phi_\ell|_{C^{p+1}(K)} \right\} \|\psi\|_{H^{p+1}(K)},
$$

which, combined with $|K| \le |B_{h_K}| \le \rho^{d+1}|B_{\rho h_K}| = \rho^{d+1}|B|$, the triangle inequality (4.6) and Bramble–Hilbert lemma estimate (2.17), completes the proof. $\qquad\square$

Theorem 4.4 provides the error estimate for the Trefftz-DG approximation of (1.1) in the $|||\cdot|||_{\mathrm{DG}}$-norm assuming that Condition 4.1 holds true. It is a consequence of Theorem 2.7 and Propositions 2.8 and 4.3.

Theorem 4.4. *Let $p \in \mathbb{N}$, and fix α, β, δ and μ as in (2.2). Let $\psi \in \mathbf{T}(\mathcal{T}_h) \cap H^{p+1}(\mathcal{T}_h)$ be the exact solution of (1.1) and $\psi_{hp} \in \mathbb{T}^p(\mathcal{T}_h)$ be the Trefftz-DG approximation solving (3.2) with $\mathbb{T}^p(\mathcal{T}_h)$ satisfying Condition 4.1 for all $K \in \mathcal{T}_h$. If $h_{K_\mathbf{x}} \simeq h_{K_t}$ for all $K \in \mathcal{T}_h$, there exists a constant C independent on the mesh size such that*

$$
|||\psi - \psi_{hp}|||_{\mathrm{DG}} \le C \sum_{K=K_\mathbf{x} \times K_t \in \mathcal{T}_h} \max\{h_{K_\mathbf{x}}, h_{K_t}\}^p \, \|\psi\|_{H^{p+1}(K)}.
$$

No special property of the Schrödinger equation has been used in the proof of Proposition 4.3, so the same result extends to any linear PDE for which one has at hand a discrete Trefftz space that can reproduce the Taylor polynomials of any PDE solution.

Remark 4.5. The use of averaged Taylor polynomials $Q^m[\cdot]$ (2.16) allows us to treat any $\psi \in H^{p+1}(K)$. Under the stronger regularity assumption $\psi \in C^{p+1}(K)$ the argument of Proposition 4.3 simplifies considerably: one could use the standard Taylor polynomial $T_{(\mathbf{z},s)}^m[\cdot]$, assume that K is star-shaped with respect to $(\mathbf{z}, s) \in K$, and require the identity (4.2a) for a single point $(\mathbf{z}, s)$ only. $\qquad\blacksquare$

4.3 Best approximation in the (1+1)-dimensional case

In this section, we show that in one space dimension the choice of exponential basis functions (4.1) with different values of k_ℓ (and equal direction $d_\ell = 1$) is enough to ensure Condition 4.1 and the approximation properties of $\mathbb{T}^p(K)$. Only $2p+1$ degrees of freedom per element are needed to obtain $O(h^p)$ convergence. An example of the $(1+1)$-dimensional basis (4.7) for $p = 3$ and $\epsilon = 10^{-1}$ is plotted in Figure 4.1.
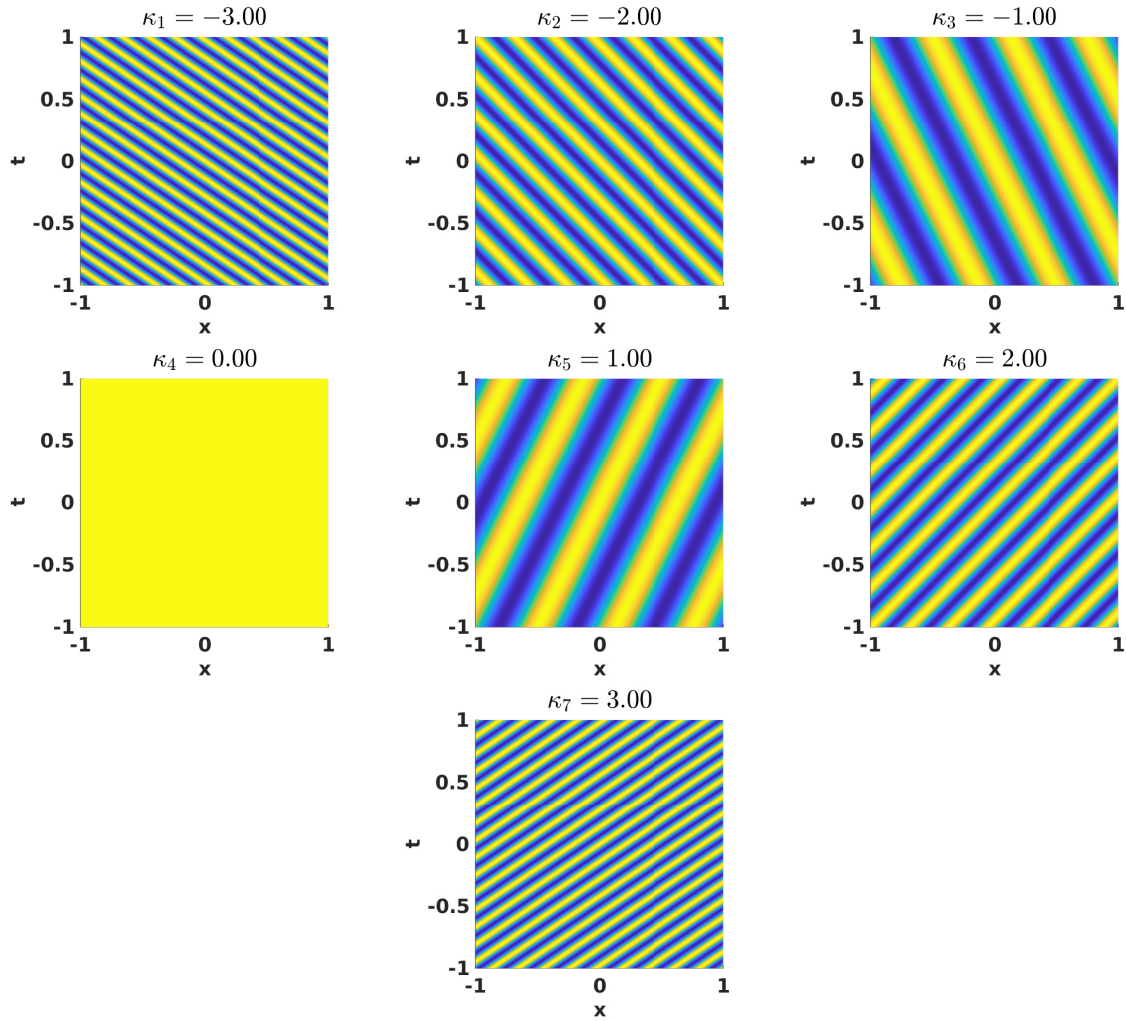
Figure 4.1 The real parts ($\epsilon^{-1}\cos(k_\ell x - k_\ell^2 t/2)$) of the Trefftz basis functions ϕ_ℓ of (4.7) for the potential $V = 0$, $\epsilon = 10^{-1}$, and $p = 3$, plotted on the space–time square $(x,t) \in (-1,1) \times (-1,1)$. The parameters k_ℓ are chosen as $k_\ell = -3, -2, \ldots, 3$. The space- and time-frequencies increase linearly and quadratically with $|k_\ell|$, respectively. The space spanned by these $n_{2,3} = 7$ basis functions approximates Schrödinger solutions with the same convergence rates of cubic polynomials.

Proposition 4.6. *Let $d = 1$, $p \in \mathbb{N}$, $n_{2,p} = 2p + 1$ and the parameters $\{k_\ell\}_{\ell=1}^{2p+1} \subset \mathbb{R}$ be all different from one another. Let*

$$\phi_\ell(x,t) = \exp\left[\frac{i}{\epsilon}\left(k_\ell x - \left(\frac{k_\ell^2}{2} + V|_K\right)t\right)\right], \qquad \ell = 1, \ldots, 2p + 1, \tag{4.7}$$

be the basis of the discrete Trefftz space $\mathbb{T}^p(K)$. Then, Condition 4.1 is satisfied.

Proof. Let $\psi \in \mathbf{T}(K) \cap H^{p+1}(K)$ and $(\mathbf{z}, s) \in B \setminus \Upsilon$, where Υ is the zero-measure set introduced in Remark 4.2. Since all the derivatives of ψ of order at most p are defined at (z, s) and each basis function $\phi_\ell \in C^\infty(K)$, the Taylor polynomials centered at (z, s) of order $p + 1$ of ψ and

each ϕ_ℓ can be written as:

$$T_{(z,s)}^{p+1}[\psi](x,t) = \sum_{j_x=0}^{p} \sum_{j_t=0}^{p-j_x} b_{\boldsymbol{j}}(x-z)^{j_x}(t-s)^{j_t}, \qquad \text{where } \boldsymbol{j} = (j_x, j_t), \tag{4.8a}$$

$$T_{(z,s)}^{p+1}[\phi_\ell](x,t) = \sum_{j_x=0}^{p} \sum_{j_t=0}^{p-j_x} M_{\boldsymbol{j},\ell}(x-z)^{j_x}(t-s)^{j_t} \quad \text{for } \ell = 1, \ldots, n_{2,p}, \tag{4.8b}$$

where $\{b_{\boldsymbol{j}}\}$ and $\{M_{\boldsymbol{j},\ell}\}$ denote the Taylor polynomial coefficients for ψ and ϕ_ℓ. We aim to prove that there exists $\mathbf{a}(z,s) \in \mathbb{C}^{2p+1}$ that satisfies (4.2a), i.e.,

$$\sum_{\ell=1}^{n_{2,p}} a_\ell(z,s) M_{\boldsymbol{j},\ell} = b_{\boldsymbol{j}} \quad \text{for } |\boldsymbol{j}| \le p, \quad \boldsymbol{j} = (j_x, j_t).$$

This can be arranged as a rectangular linear system of the form $\mathbf{M}\,\mathbf{a}(z,s) = \mathbf{b}$, with $\mathbf{b} \in \mathbb{C}^{r_{2,p}}$ and $\mathbf{M} \in \mathbb{C}^{r_{2,p} \times n_{2,p}}$, $r_{2,p} = (p+1)(p+2)/2 \ge n_{2,p}$.

The case $p = 1$ is rather special, since $n_{2,1} = r_{2,1} = 3$ and the matrix $\mathbf{M}$ is square; hence we just need to show that $\mathbf{M}$ is not singular. For $p \ge 2$, in order to prove the existence of such an $\mathbf{a}(z,s)$, we need to show that $\mathbf{b} \in \mathrm{Im}\,(\mathbf{M})$. To this purpose, after equation (4.10) below, we define a set $\mathcal{D} \subset \mathbb{C}^{r_{2,p}}$ that contains $\{\mathbf{b}\} \cup \mathrm{Im}\,(\mathbf{M})$. Afterwards, it only remains to show that $\dim \mathcal{D} = \mathrm{rank}\,(\mathbf{M})$.

For all smooth $\varphi \in \mathbf{T}(K)$ and each multi-index $\boldsymbol{j}$, with $|\boldsymbol{j}| \le p$, $D^{\boldsymbol{j}}\varphi(z,s) = D^{\boldsymbol{j}}T_{(z,s)}^{p+1}[\varphi](z,s)$. Since φ and its derivatives satisfy the Schrödinger equation, this implies that $D^{\boldsymbol{j}}T_{(z,s)}^{p+1}[\varphi]$ satisfies the same equation at the single point (z,s):

$$\left(\left(i\epsilon \frac{\partial}{\partial t} + \frac{\epsilon^2}{2}\Delta_{\mathbf{x}} - V \right) D^{\boldsymbol{j}}T_{(z,s)}^{p+1}[\varphi] \right)(z,s) = 0 \quad \text{for } p \ge 2 \text{ and } |\boldsymbol{j}| \le p-2. \tag{4.9}$$

We want to express these equations in terms of the coefficients of $T_{(z,s)}^{p+1}[\varphi](x,t)$. Setting $\mathbf{C} \in \mathbb{C}^{r_{2,p}}$ to be the vector with components $C_{j_x,j_t} := \frac{1}{\boldsymbol{j}!}D^{\boldsymbol{j}}\varphi(z,s)$ with $\boldsymbol{j} = (j_x, j_t)$, so that $T_{z,s}^{p+1}[\varphi](x,t) = \sum_{|\boldsymbol{j}|\le p} C_{j_x,j_t}(x-z)^{j_x}(t-s)^{j_t}$, we have

$$\partial_t T_{z,s}^{p+1}[\varphi](x,t) = \sum_{j_x=0}^{p} \sum_{j_t=0}^{p-1-j_x} C_{j_x,j_t+1}(j_t+1)(x-z)^{j_x}(t-s)^{j_t},$$

$$\frac{\partial^2}{\partial x^2}T_{z,s}^{p+1}[\varphi](x,t) = \sum_{j_x=0}^{p-2} \sum_{j_t=0}^{p-2-j_x} C_{j_x+2,j_t}(j_x+1)(j_x+2)(x-z)^{j_x}(t-s)^{j_t}.$$

Expanding (4.9), we get the following $p(p-1)/2$ relations between the coefficients of the Taylor polynomial of φ:

$$i\epsilon(j_t+1)C_{j_x,j_t+1} + \frac{\epsilon^2}{2}(j_x+1)(j_x+2)C_{j_x+2,j_t} = V C_{j_x,j_t} \tag{4.10}$$

$$\text{for } |\boldsymbol{j}| \le p-2, \quad \boldsymbol{j} = (j_x, j_t).$$

We define $\mathcal{D} := \left\{ \mathbf{C} \in \mathbb{C}^{r_{2,p}} \mid \mathbf{C} \text{ satisfies (4.10)} \right\}$. Since $\psi, \phi_\ell \in \mathbf{T}(K)$, it is evident that $\mathbf{b} \in \mathcal{D}$ and $\mathrm{Im}\,(\mathbf{M}) \subset \mathcal{D}$.

The key point of the proof is to show that $\mathrm{rank}\,(\mathbf{M}) = \dim(\mathcal{D}) = 2p + 1$, which guarantees that the overdetermined linear system $\mathbf{M}\mathbf{a}(z, s) = \mathbf{b}$ has a unique solution.

Figure 4.2 illustrates the relations that define $\mathcal{D}$ for both cases $V = 0$ and $V \neq 0$. It shows that given the $2p+1$ entries C_{j_x, j_t} of $\mathbf{C} \in \mathcal{D}$ corresponding to $j_x \in \{0, 1\}$, all other entries are uniquely determined by the conditions (4.10). The dots in the (j_x, j_t) plane represent the coefficients C_{j_x, j_t} of $\mathbf{C} \in \mathcal{D} \subset \mathbb{C}^{r_{2,p}}$, $r_{2,p} = \frac{(p+1)(p+2)}{2}$. In the case $V \neq 0$ (right panel), each shape $\overset{\bullet}{\underset{\bullet}{\diagdown}}\!\!\bullet$ connects three dots located at the points $(j_x, j_t + 1)$, (j_x, j_t) and $(j_x + 2, j_t)$: this shape represents one of the equations (4.10) which, given C_{j_x, j_t+1} and C_{j_x, j_t}, allows us to compute C_{j_x+2, j_t}. If the $2p + 1$ values with $j_x \in \{0, 1\}$ (corresponding to the nodes in the shaded region) are given, then these relations uniquely determine all the other coefficients, which can be computed sequentially by proceeding left to right in the diagram. In the figure $p = 7$, the number of nodes is $r_{2,p} = 36$, the number of nodes in the shaded region is $n_{2,p} = 15$, the number of relations is $r_{2,p} - n_{2,p} = 21$. The case $V = 0$ (left panel) is slightly simpler: the coefficient C_{j_x+2, j_t} is determined by C_{j_x, j_t+1} only, and each relation in (4.10) is depicted as a segment $\overset{\bullet}{\diagdown}\!\bullet$.

(a) Case $V = 0$ (b) Case $V \neq 0$
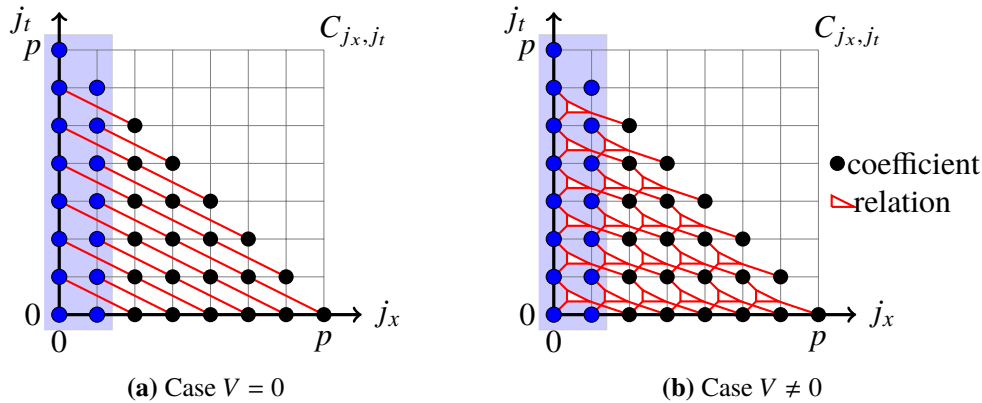
Figure 4.2 Illustration of the relations defining the set $\mathcal{D}$ in the $(1 + 1)$-dimensional case.

Therefore,
$$\mathrm{rank}\,(\mathbf{M}) \leq \dim\,(\mathcal{D}) \leq 2p + 1.$$

The rest of the proof consists in proving that $\mathrm{rank}\,(\mathbf{M}) \geq 2p + 1$, i.e. that $\mathbf{M}$ is full-rank. To do so, we first recall that $\phi_\ell(x, t) = \exp\left[\frac{i}{\epsilon} \left(k_\ell x - \left(\frac{k_\ell^2}{2} + V \right) t \right) \right]$ from the basis definition (4.7), and observe that the entries of matrix $\mathbf{M}$ are given by

$$M_{\mathbf{j}, \ell} = \frac{1}{j_x! j_t!} D^{\mathbf{j}} \phi_\ell(z, s) = q_{j_x, j_t}(k_\ell) \phi_\ell(z, s) \quad \mathbf{j} = (j_x, j_t), \quad \ell = 1, \dots, 2p + 1,$$

where $q_{j_x, j_t}(k) = \frac{1}{j_x! j_t!} \left(i\epsilon^{-1} k \right)^{j_x} \left(-i\epsilon^{-1} \left(\frac{k^2}{2} + V \right) \right)^{j_t}$ is a complex-valued polynomial of degree exactly $j_x + 2j_t$. We define a square submatrix $\mathbf{M}_\square$ of the matrix $\mathbf{M}$ by taking the rows corresponding to $j_x \in \{0, 1\}$. This, in turn, can be decomposed as $\mathbf{M}_\square = \mathbf{V} \mathbf{D}_{z,s}$, where $\mathbf{D}_{z,s} = \mathrm{diag}(\phi_1(z, s), \dots, \phi_{2p+1}(z, s))$ and the Vandermonde-like matrix $\mathbf{V} \in \mathbb{C}^{(2p+1) \times (2p+1)}$ is given by

$$
\mathbf{V} = \begin{pmatrix}
q_{0,0}(k_1) & q_{0,0}(k_2) & \dots & q_{0,0}(k_{2p+1}) \\
q_{1,0}(k_1) & q_{1,0}(k_2) & \dots & q_{1,0}(k_{2p+1}) \\
\vdots & \vdots & & \vdots \\
q_{0,j_t}(k_1) & q_{0,j_t}(k_2) & \dots & q_{0,j_t}(k_{2p+1}) \\
q_{1,j_t}(k_1) & q_{1,j_t}(k_2) & \dots & q_{1,j_t}(k_{2p+1}) \\
\vdots & \vdots & & \vdots \\
q_{0,p-1}(k_1) & q_{0,p-1}(k_2) & \dots & q_{0,p1}(k_{2p+1}) \\
q_{1,p-1}(k_1) & q_{1,p-1}(k_2) & \dots & q_{1,p-1}(k_{2p+1}) \\
q_{0,p}(k_1) & q_{0,p}(k_2) & \dots & q_{0,p}(k_{2p+1})
\end{pmatrix} . \tag{4.11}
$$

We observe that $V_{m,\ell} = p_{m-1}(k_\ell)$, $m = 1, \dots, 2p + 1$, for some polynomials $p_m(\cdot)$ of degree m. Therefore, there exists a lower triangular matrix $\mathbf{P}$ such that $\mathbf{PV} = \widetilde{\mathbf{V}}$, where $\widetilde{V}_{m,\ell} = k_\ell^{m-1}$ for $m, \ell = 1, \dots, 2p + 1$. The entries of the m-th row of the inverse of $\mathbf{P}$ are the coefficients of the monomial expansion of the polynomial p_{m-1}. This means that $\widetilde{\mathbf{V}}$ is a Vandermonde matrix, so it is invertible [50, §22.1] (recall that all k_ℓ are different from one another). We conclude that the matrix $\mathbf{V}$ is invertible, independent of (z, s) and $\mathbf{M}$ is full rank (namely, rank $(\mathbf{M}) = 2p + 1$).

Denoting by $\mathbf{b}_\square$ the subvector of $\mathbf{b}$ corresponding to the indices $j_x \in \{0, 1\}$, the unique solution of the linear system $\mathbf{Ma}(z, s) = \mathbf{b}$ is $\mathbf{a}(z, s) = \mathbf{M}_\square^{-1} \mathbf{b}_\square$ and satisfies condition (4.2a). Moreover, the following bound holds

$$
|\mathbf{a}(z, s)|_1 \le |||\mathbf{V}^{-1}|||_1 |||\mathbf{D}_{z,s}^{-1}|||_1 |\mathbf{b}_\square|_1 , \tag{4.12}
$$

where $|||\mathbf{D}_{z,s}^{-1}|||_1 = 1$ for all $(z, s) \in B$. We now recall that (z, s) was chosen arbitrarily in $B \setminus \Upsilon$. Setting $\mathbf{a}(z, s) = \mathbf{0}$ and $\mathbf{b}_\square(z, s) = \mathbf{0}$ for $(z, s) \in \Upsilon$ (which has zero measure), recalling from (4.8a) that $b_j = \frac{1}{j!} D^j \psi$, and integrating (4.12) over B, we obtain

$$
\||\mathbf{a}|_1\|_{L^1(B)} \le |||\mathbf{V}^{-1}|||_1 \left\||\mathbf{b}_\square|_1\right\|_{L^1(B)} \le |||\mathbf{V}^{-1}|||_1 \sqrt{2p + 1} \left\||\mathbf{b}_\square|_2\right\|_{L^1(B)}
$$
$$
\le |||\mathbf{V}^{-1}|||_1 \sqrt{2p + 1} \sqrt{|B|} \|\psi\|_{H^{p+1}(B)} .
$$

This implies the assertion (4.2b) with $C_\star = \sqrt{2p + 1} |||\mathbf{V}^{-1}|||_1$, since $|B| \le |K|$. $\qquad\square$

Since the constant $C_\star$ depends on the norm of the inverse of the Vandermonde matrix $\mathbf{V}$, which in turn depends on the parameters $\{k_\ell\}_{\ell=1}^{2p+1}$ and ϵ, the choice of the parameters defining the Trefftz basis functions has an important influence on the accuracy of the method; this is illustrated in the numerical experiments in Section 4.5.1.

4.4 Best approximation in the (2+1)-dimensional case

Following the same strategy of the previous section, in Proposition 4.7 we show that, for a sensible choice of the parameters $\{k_\ell\}$ and the directions $\{\mathbf{d}_\ell\}$ in (4.1), Condition 4.1 is true for $\mathbb{T}^p(\mathcal{T}_h)$ in the $(2 + 1)$-dimensional case.

The basis functions $\phi_{m,\lambda}$ are plane waves in space indexed by two parameters: m identifying the wavenumber k_m, and λ identifying the propagation direction $\theta_{m,\lambda}$. For every wavenumber we take a different number of directions following a strategy similar to that used for the construction of plane-wave Trefftz bases for the 3D Helmholtz equation in [79, Lemma 4.2]. The time-dependence of each basis element is harmonic with frequency $\epsilon^{-1}(k_m^2/2 + V|_K)$. A sample basis for $p = 2$ and $\epsilon = 10^{-1}$ is shown in Figure 4.3.

Proposition 4.7. *Let $d = 2$ and $n_{3,p} = (p + 1)^2$. Let the parameters k_m and $\theta_{m,\lambda}$ satisfy the following conditions:*

$$k_m \in \mathbb{R} \setminus \{0\} \;\; for \; m = 0, \dots, p, \;\; with \;\; k_{m_1}^2 \neq k_{m_2}^2 \; for \; m_1 \neq m_2,$$
$$\theta_{m,\lambda} \in [0, 2\pi) \;\; for \; m = 0, \dots, p, \; \lambda = 1, \dots, 2m + 1, \;\; with \;\; \theta_{m,\lambda_1} \neq \theta_{m,\lambda_2} \; for \; \lambda_1 \neq \lambda_2.$$

Define the directions $\mathbf{d}_{m,\lambda} = (\cos(\theta_{m,\lambda}), \sin(\theta_{m,\lambda}))$ and the basis functions

$$\phi_{m,\lambda}(\mathbf{x}, t) = \exp\left[\frac{i}{\epsilon}\left(k_m \mathbf{d}_{m,\lambda}^\top \mathbf{x} - \left(\frac{k_m^2}{2} + V|_K\right) t\right)\right] \;\; for \; m = 0, \dots, p, \; \lambda = 1, \dots, 2m + 1.$$

Then, Condition 4.1 holds true.
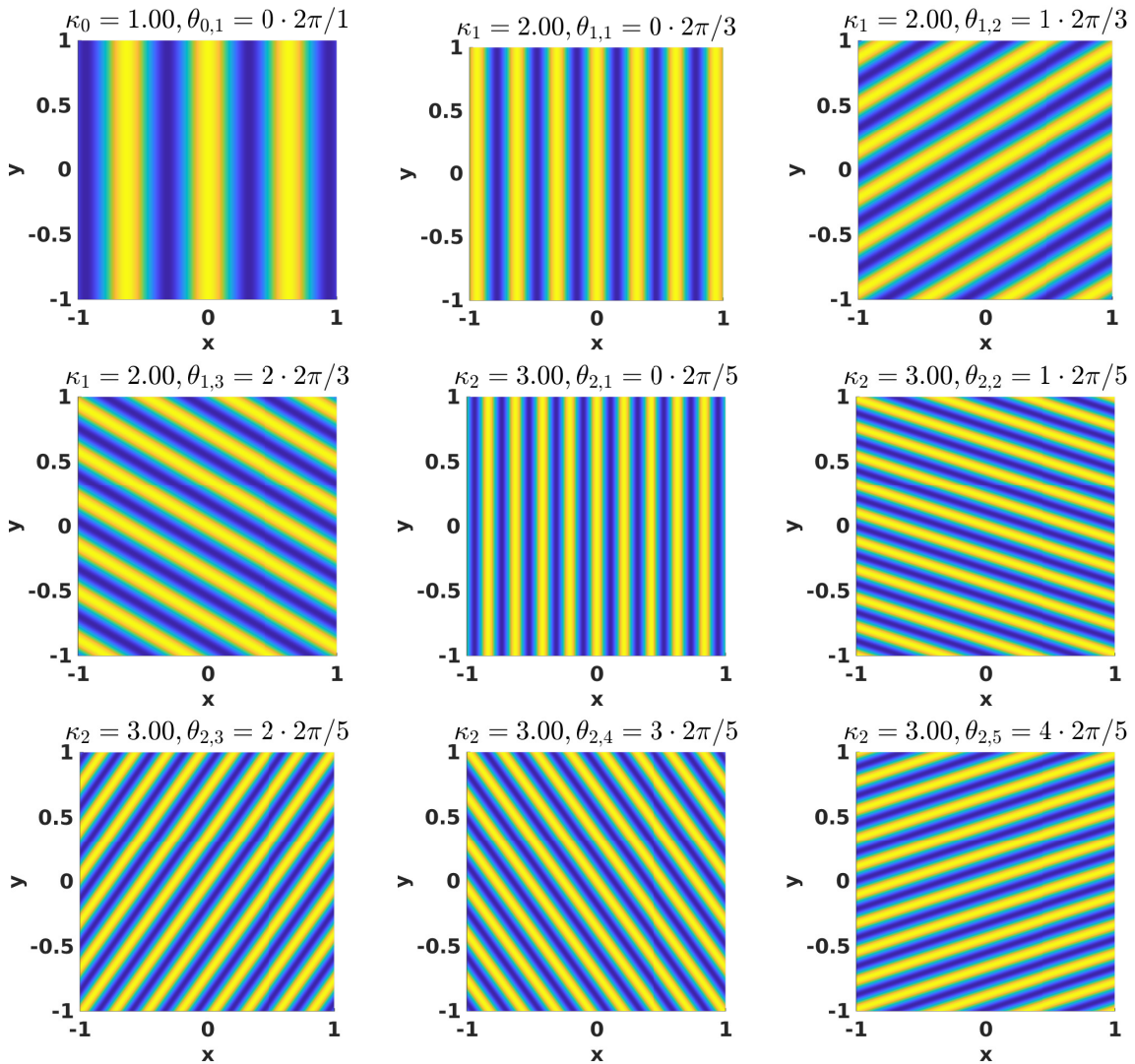
Figure 4.3 The real parts of the Trefftz basis functions defined in (4.7) for $V = 0$, $\epsilon = 10^{-1}$, and $p = 2$ at time $t = 0$ on the space domain $(-1, 1)^2$. Here $k_m = m + 1$ for $m = 0, 1, 2$ and $\theta_{m,\lambda} = \frac{2\pi(\lambda-1)}{2m+1}$ for $\lambda = 1, \dots, 2m + 1$. Note that, differently from Figure 4.1, here we only plot the space dependence of ϕ_ℓ.

Proof. Let $\psi \in \mathbf{T}(K) \cap H^{p+1}(K)$ and $(\mathbf{z}, s) \subset B \setminus \Upsilon$, with the set Υ as in Remark 4.2. As in Proposition 4.6, we can arrange the coefficients of the Taylor polynomials of ψ and each basis function ϕ_ℓ in a vector $\mathbf{b} \in \mathbb{C}^{r_{3,p}}$ and a matrix $\mathbf{M} \in \mathbb{C}^{r_{3,p} \times n_{3,p}}$ with $r_{3,p} = (p+1)(p+2)(p+3)/6$. In order to prove (4.2b), we look for a vector $\mathbf{a}(\mathbf{z}, s) \in \mathbb{C}^{n_{3,p}}$ such that $\mathbf{Ma}(\mathbf{z}, s) = \mathbf{b}$.

Similarly to the $(1+1)$-dimensional case, $\mathbf{b} \in \mathcal{D}$ and $\mathrm{Im}(\mathbf{M}) \subset \mathcal{D}$ for $\mathcal{D}$ the set of all the vectors $\mathbf{C} \in \mathbb{C}^{r_{3,p}}$ satisfying the following relations

$$
i\epsilon(j_t + 1)C_{j_x, j_y, j_t+1} + \frac{\epsilon^2}{2}(j_x + 1)(j_x + 2)C_{j_x+2, j_y, j_t}
$$
$$
+ \frac{\epsilon^2}{2}(j_y + 1)(j_y + 2)C_{j_x, j_y+2, j_t} = VC_{j_x, j_y, j_t} \tag{4.13}
$$
$$
\text{for } |\mathbf{j}| \le p - 2, \ \mathbf{j} = (j_x, j_y, j_t).
$$

Figure 4.4 depicts the equations (4.13) as relations between the coefficients of the vector $\mathbf{C}$, which are represented as points in the (j_x, j_y, j_t) space. In particular, it shows that the $(p + 1)^2$ entries of any $\mathbf{C} \in \mathcal{D}$ with $j_x \in \{0, 1\}$ determine all the other entries of $\mathbf{C}$, thus $\mathrm{rank}(\mathbf{M}) \le \dim(\mathcal{D}) \le (p + 1)^2$. The colored dots in position $\mathbf{j} = (j_x, j_y, j_t)$, $|\mathbf{j}| \le p$, correspond to the entries C_{j_x, j_y, j_t} of the vector $\mathbf{C} \in \mathcal{D} \subset \mathbb{C}^{r_{3,p}}$ (here $p = 5$ and $r_{3,p} = 56$). Each white circle is connected by the segments to four nodes and represents one of the equations in (4.13): given C_{j_x, j_y, j_t}, C_{j_x, j_y, j_t+1} and C_{j_x, j_y+2, j_t}, it allows to compute C_{j_x+2, j_y, j_t} (the leftmost of the four nodes connected to a given white circle). The red dot exemplifies one of these relations, for $\mathbf{j} = (0, 1, 2)$. Given the $(p+1)^2$ coefficients with $j_x \in \{0, 1\}$ (the blue dots), all other coefficients are uniquely determined and $\dim(\mathcal{D}) \le (p + 1)^2$.
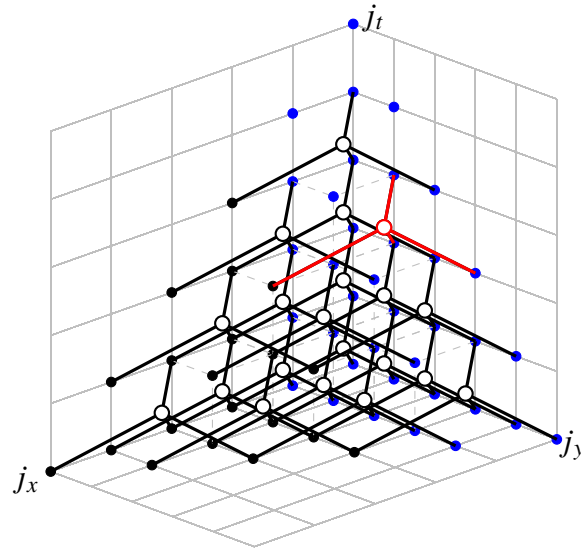
Figure 4.4 A representation of the relations defining the set $\mathcal{D}$ in the (2+1)-dimensional case.

Now, it just remains to prove that $\mathrm{rank}(\mathbf{M}) \ge (p + 1)^2$. We fix an arbitrary ordering $\ell = 1, \dots, (p + 1)^2$ of the "triangular" index set $\{(m, \lambda) : m = 0, \dots, p, \ \lambda = 1, \dots, 2m + 1\}$, so that we can write ϕ_ℓ, k_ℓ and θ_ℓ for the basis functions $\phi_{m,\lambda}$ and the parameters $k_{m,\lambda}$ and $\theta_{m,\lambda}$.

First, from the basis function definition, the matrix $\mathbf{M}$ can be decomposed as $\mathbf{M} = \mathbf{GD}_{\mathbf{z},s}$ for the diagonal matrix $\mathbf{D}_{\mathbf{z},s} = \mathrm{diag}(\phi_1(\mathbf{z}, s), \dots, \phi_{(p+1)^2}(\mathbf{z}, s))$ and a matrix $\mathbf{G} \in \mathbb{C}^{r_{3,p}, n_{3,p}}$ with

entries independent of $(\mathbf{z}, s)$ given by:

$$\mathbf{G}_{\mathbf{j},\ell} = \frac{1}{j_x! j_y! j_t!} (i\epsilon^{-1} k_\ell \cos\theta_\ell)^{j_x} (i\epsilon^{-1} k_\ell \sin\theta_\ell)^{j_y} \left(-i\epsilon^{-1}\left(\frac{k_\ell^2}{2} + V\right)\right)^{j_t}, \quad \mathbf{j} = (j_x, j_y, j_t). \tag{4.14}$$

We define a square matrix $\mathbf{S} \in \mathbb{C}^{n_{3,p} \times n_{3,p}}$ with the block structure $\mathbf{S} = \begin{pmatrix} \mathbf{S}^+ \\ \mathbf{S}^- \end{pmatrix}$, where $\mathbf{S}^+ \in \mathbb{C}^{\frac{(p+1)(p+2)}{2} \times n_{3,p}}$ and $\mathbf{S}^- \in \mathbb{C}^{\frac{p(p+1)}{2} \times n_{3,p}}$ are defined as

$$S^+_{(j_\mathbf{x},j_t),\ell} = \left(\frac{k_\ell^2}{2} + V\right)^{j_t} k_\ell^{j_\mathbf{x}} e^{i\theta_\ell j_\mathbf{x}} \qquad \text{for} \quad \begin{matrix} j_\mathbf{x} = 0, \ldots, p, \\ j_t = 0, \ldots, p - j_\mathbf{x}, \end{matrix} \quad \ell = 1, \ldots, (p+1)^2,$$

$$S^-_{(j_\mathbf{x},j_t),\ell} = \left(\frac{k_\ell^2}{2} + V\right)^{j_t} k_\ell^{j_\mathbf{x}} e^{-i\theta_\ell j_\mathbf{x}} \qquad \text{for} \quad \begin{matrix} j_\mathbf{x} = 1, \ldots, p, \\ j_t = 0, \ldots, p - j_\mathbf{x}, \end{matrix} \quad \ell = 1, \ldots, (p+1)^2.$$

The binomial theorem allows us to relate the blocks $\mathbf{S}^\pm$ to $\mathbf{G}$:

$$S^+_{(j_\mathbf{x},j_t),\ell} = \left(\frac{k_\ell^2}{2} + V\right)^{j_t} (k_\ell \cos\theta_\ell + ik_\ell \sin\theta_\ell)^{j_\mathbf{x}}$$

$$= \sum_{j_x+j_y=j_\mathbf{x}} \frac{j_\mathbf{x}!}{j_x! j_y!} (k_\ell \cos\theta_\ell)^{j_x} (ik_\ell \sin\theta_\ell)^{j_y} \left(\frac{k_\ell^2}{2} + V\right)^{j_t}$$

$$= \sum_{j_x+j_y=j_\mathbf{x}} \frac{j_t! j_\mathbf{x}! \epsilon^{|\mathbf{j}|}}{i^{j_x}(-i)^{j_t}} \mathbf{G}_{\mathbf{j},\ell}, \qquad \begin{matrix} j_\mathbf{x} = 0, \ldots, p, \\ j_t = 0, \ldots, p - j_\mathbf{x}, \\ \ell = 1, \ldots, (p+1)^2, \end{matrix}$$

$$S^-_{(j_\mathbf{x},j_t),\ell} = \left(\frac{k_\ell^2}{2} + V\right)^{j_t} (k_\ell \cos\theta_\ell - ik_\ell \sin\theta_\ell)^{j_\mathbf{x}}$$

$$= \sum_{j_x+j_y=j_\mathbf{x}} \frac{j_\mathbf{x}!}{j_x! j_y!} (k_\ell \cos\theta_\ell)^{j_x} (-ik_\ell \sin\theta_\ell)^{j_y} \left(\frac{k_\ell^2}{2} + V\right)^{j_t}$$

$$= \sum_{j_x+j_y=j_\mathbf{x}} \frac{j_t! j_\mathbf{x}! \epsilon^{|\mathbf{j}|}}{i^{j_x}(-1)^{j_y}(-i)^{j_t}} \mathbf{G}_{\mathbf{j},\ell}, \qquad \begin{matrix} j_\mathbf{x} = 1, \ldots, p, \\ j_t = 0, \ldots, p - j_\mathbf{x}, \\ \ell = 1, \ldots, (p+1)^2. \end{matrix}$$

This means that there exists a matrix $\mathbf{P} \in \mathbb{C}^{n_{3,p} \times r_{3,p}}$ such that $\mathbf{S} = \mathbf{P}\mathbf{G}$.

Next, we prove that $\mathbf{S}$ is not singular. If $\mathbf{S}^\top \mathbf{c} = \mathbf{0}$ for some vector $\mathbf{c} \in \mathbb{C}^{n_{3,p}}$, then for each pair $(k_m, \theta_{m,\lambda})$ in the definition of the basis functions we have

$$0 = \sum_{n=-p}^{p} \left(\sum_{j_t=0}^{p-|n|} c_{n,j_t} \left(\frac{k_m^2}{2} + V\right)^{j_t} k_m^{|n|} \right) e^{in\theta_{m,\lambda}} = \sum_{n=-p}^{p} \zeta_n(k_m) \, e^{in\theta_{m,\lambda}}, \tag{4.15}$$

$$m = 0, \ldots, p, \quad \lambda = 1, \ldots, 2m+1,$$

If we fix $m = p$ in (4.15), the $\zeta_n(k_p)$ are the coefficients of a trigonometric polynomial of degree p with $2p + 1$ different zeros $\theta_{p,\lambda}$, which implies that $\zeta_n(k_p) = 0$, for $n = -p, \ldots, p$. In particular $0 = \zeta_{\pm p}(k_p) = c_{\pm p,0} k_p^p \Rightarrow c_{\pm p,0} = 0$, since $k_p \neq 0$ by hypothesis.

We now proceed by (backward) induction. Assume that for some $\eta \in \{1, \ldots, p\}$ we have:

$$c_{\pm n, j_t} = 0, \quad \text{for } n = p - \eta + 1, \ldots, p \text{ and } j_t = 0, \ldots, p - |n|, \tag{4.16a}$$

$$\zeta_n(k_m) = 0, \quad \text{for } n = -p, \ldots, p \text{ and } m = p - \eta + 1, \ldots, p. \tag{4.16b}$$

Then (4.15) gives

$$\sum_{n=-(p-\eta)}^{p-\eta} \zeta_n(k_m) e^{in\theta_{m,\lambda}} = 0,$$

and for $m = (p - \eta)$, the $\zeta_n(k_{p-\eta})$ are the coefficients of a trigonometric polynomial of degree $p - \eta$ with $2(p - \eta) + 1$ different zeros $\theta_{p-\eta,\lambda}$. Therefore, assumption (4.16b) also holds for $m = p - \eta$. For $n = \pm(p - \eta)$ in (4.16b) we have

$$k_m^{p-\eta} \sum_{j_t=0}^{\eta} c_{\pm(p-\eta), j_t} \left(\frac{k_m^2}{2} + V \right)^{j_t} = 0;$$

therefore, $c_{\pm(p-\eta), j_t}$ are the coefficients of a complex-valued polynomial of degree η with $\eta + 1$ different zeros $\left(k_m^2/2 + V \right)$, $m = p - \eta, \ldots, p$; which implies that assumption (4.16a) holds for $n = p - \eta$. Recursively, this leads to $\mathbf{c} = \mathbf{0}$; therefore the matrix $\mathbf{S}$ is invertible. Since $\mathbf{S} = \mathbf{PG}$ and $\mathbf{M} = \mathbf{GD}_{\mathbf{z},s}$ for a nonsingular matrix $\mathbf{D}_{\mathbf{z},s}$, the matrix $\mathbf{M}$ has rank at least $n_{3,p}$, so it is full rank.

The solution $\mathbf{a}(\mathbf{z}, s)$ of the rectangular linear system $\mathbf{Ma}(\mathbf{z}, s) = \mathbf{b}$ is $\mathbf{a}(\mathbf{z}, s) = \mathbf{D}_{\mathbf{z},s}^{-1} \mathbf{S}^{-1} \mathbf{Pb}$, which satisfies (4.2a). Then, the following bound is obtained

$$|\mathbf{a}(z, s)|_1 \leq |||\mathbf{S}^{-1}|||_1 |||\mathbf{D}_{\mathbf{z},s}^{-1}|||_1 |\mathbf{Pb}|_1 ,$$

where $|||\mathbf{D}_{z,s}^{-1}|||_1 = 1$ for all $(\mathbf{z}, s) \in B$. Writing $b_j = \frac{1}{j!} D^j \psi(\mathbf{z}, s)$ and integrating over B as in the proof of Proposition 4.6, we obtain (4.2b) with $C_\star = \sqrt{r_{3,p}} |||\mathbf{S}^{-1}|||_1 |||\mathbf{P}|||_1$. $\qquad \square$

Analogously to the $(1 + 1)$-dimensional case, the constant $C_\star$ depends on the norm of the inverse of the matrix $\mathbf{S}$, which depends on ϵ, and the choice of both the parameters $\{k_m\}$ and the directions $\{\mathbf{d}_{m,\lambda}\}$. This indicates that finding an appropriate choice of these parameters is crucial in order to get accurate and stable approximations.

Remark 4.8 (Arbitrary space dimension). In order to extend the Trefftz approximation theory to $(d + 1)$-dimensional problems with $d \geq 3$, one has to provide conditions on the parameters defining the local basis functions. Not only one has to determine the minimal number of directions $\{\mathbf{d}_{m,\lambda}\} \subset \mathcal{S}_1^d$ associated to each parameter k_m, but also the mutual relations that the directions have to satisfy (e.g. not too many of them can belong to the same $(d-1)$-dimensional hyperplane). $\qquad \blacksquare$

Remark 4.9. The previous sections show that for $d = 1, 2$ the local space $\mathbb{T}^p(K)$ has dimension $O_{p \to \infty}(p^d)$ and approximates Schrödinger solutions with the same convergence rates as the space $\mathbb{P}^p(K)$ of the degree-p polynomials on K, which has larger dimension $\dim(\mathbb{P}^p(K)) = O_{p \to \infty}(p^{d+1})$. More precisely, $\mathbb{T}^p(K)$ has the dimension of the space of the harmonic polynomials of degree p on K. We expect the same to hold for $d > 2$. $\qquad \blacksquare$

4.5 Numerical experiments

We present some numerical experiments validating the error estimates in the mesh-dependent norm $||| \cdot |||_{\mathrm{DG}}$ derived in Section 4.2. We also numerically assess the $L^2(\Omega)$ error at the final-time, mesh-independent $\|\cdot\|_{L_2(\mathcal{F}_h^T)}$ norm and evaluate the energy dissipation of the proposed method. The κ_2-condition number of the stiffness matrix is studied numerically in Section 5.5.5.

By choosing the same space mesh $\mathcal{T}_{h_{\mathbf{x},n}}^{\mathbf{x}}$ and time step $\tau = t_n - t_{n-1}$ for all $n = 1, \ldots, N$, we can apply a time translation for each time-slab $\Omega \times (t_{n-1}, t_n)$ in the definition of the basis functions (4.1), as $\phi_\ell(\mathbf{x}, t) := \exp\left[i\epsilon^{-1}\left(k_\ell \mathbf{d}_\ell^\top \mathbf{x} - \left(k_\ell^2/2 + V|_K\right)(t - t_{n-1})\right)\right]$. This makes the matrices of the linear systems for all the time-slabs to be the same, which represents a substantial reduction in the computational cost of the method. To solve these systems we perform the LU factorization of such matrix once using the Matlab's function `lu` with scaling and row–column permutations, which produces sparser and stable factorizations; then, we solve for each time slab applying forward and backward substitutions.

The integrals in the assembly of the Galerkin matrix and load vectors are computed with Gauss–Legendre quadratures. On polytopic meshes, thanks to the choice of exponential basis functions (4.1), closed formulas for all the integrals appearing in the matrix assembly could be written, following the ideas in [52, §4.1]: the implementation of these formulas is non-trivial, particularly in higher dimensions, but it could considerably speed up the computations. We recall that the Trefftz-DG formulation (3.2) does not involve $(d + 1)$-dimensional integrals over mesh elements but only on the d-dimensional element faces.

4.5.1 Square-well potential in $(1 + 1)$ dimensions

Let us consider the (1+1)-dimensional Schrödinger equation (1.1) on $Q_T = (-\sqrt{2}, \sqrt{2}) \times (0, 1)$ with $\epsilon = 1$, homogeneous Dirichlet boundary conditions, and the following square-well potential:

$$V(x) = \begin{cases} 0 & x \in (-\sqrt{2}/2, \sqrt{2}/2), \\ V_* & x \in (-\sqrt{2}, \sqrt{2}) \setminus (-\sqrt{2}/2, \sqrt{2}/2), \end{cases} \tag{4.17}$$

for some $V_* > 0$. The initial condition is taken as an eigenfunction (bound state) of $-\frac{1}{2}\partial_x^2 + V$ on $(-\sqrt{2}, \sqrt{2})$:

$$\psi_0(x) = \begin{cases} \cos(k_*\sqrt{2}x), & x \in (-\sqrt{2}, \sqrt{2}), \\ \frac{\cos(k_*)}{\sinh(\sqrt{V_*-k_*^2})} \sinh(\sqrt{V_* - k_*^2}(2 - \sqrt{2}|x|)), & x \in (-\sqrt{2}, \sqrt{2}) \setminus (-\sqrt{2}/2, \sqrt{2}/2), \end{cases}$$

where k_* is a real root of the function $f(k) := \sqrt{V_* - k^2} - k\tan(k)\tanh(\sqrt{V_* - k^2})$. The solution of the corresponding initial boundary value problem (1.1) is $\psi(x, t) = \psi_0(x)\exp(-ik_*^2 t)$. For each V_*, there is a finite number of such values k_*: in the numerical experiments below, we take the largest one, corresponding to faster oscillations in space and time. In Figure 4.5, we present the plot of the function $f(x)$ for $V_* = 20$ and $V_* = 50$ with the values of k_* used in this experiment.

From the definition (4.7) of the Trefftz basis, we observe that if the parameters k_ℓ are chosen very close to one another then the corresponding basis functions approach mutual

(a) $V_* = 20$ **(b)** $V_* = 50$
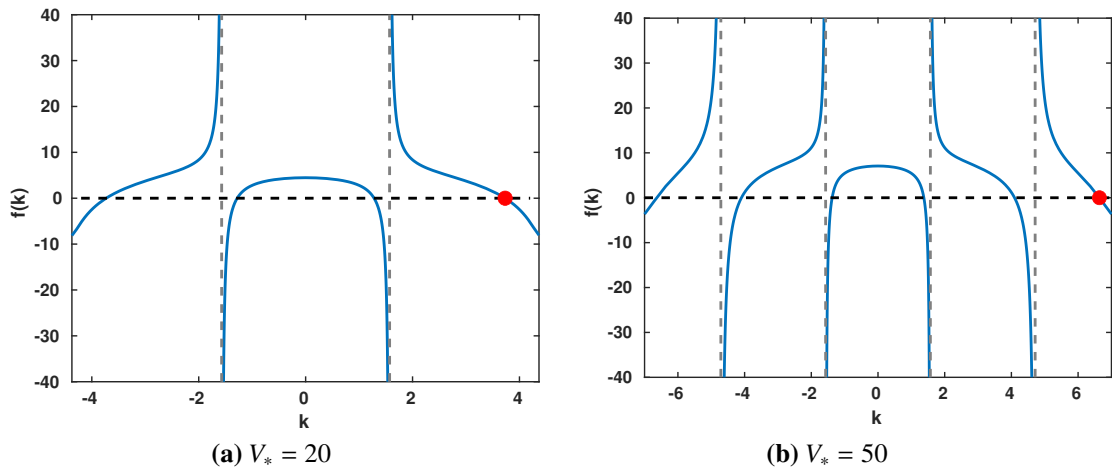
Figure 4.5 Plot of $f(k)$ for different values of V_* in the definition of the square-well potential (4.17). The red dots are the values of k taken in the numerical experiments: $k_* \approx 3.7319$ ($V_* = 20$) and $k_* \approx 6.6394$ ($V_* = 50$) .

linear dependence. The consequence is that the Trefftz-DG method becomes more and more ill-conditioned. This is confirmed by the Vandermonde-like matrix $\mathbf{V}$ in (4.11), which is singular in the limit $|k_{\ell_1} - k_{\ell_2}| \to 0$. In the experiments we take $2p + 1$ equally spaced values $k_\ell/2 \in \{-p, -(p-1), \ldots, 0, \ldots, p-1, p\}$.

In Figure 4.6 we plot the exact solution ψ for $V_* = 20$ and $V_* = 50$. It can be observed that, for increasing V_*, the solution oscillates more with respect to x in $(-\sqrt{2}/2, \sqrt{2}/2)$, while it decays monotonically to 0 in $(-\sqrt{2}, \sqrt{2}) \setminus (-\sqrt{2}/2, \sqrt{2}/2)$.

(a) $\mathfrak{Re}\,(\psi)$ for $V_* = 20$ **(b)** $\mathfrak{Re}\,(\psi)$ for $V_* = 50$
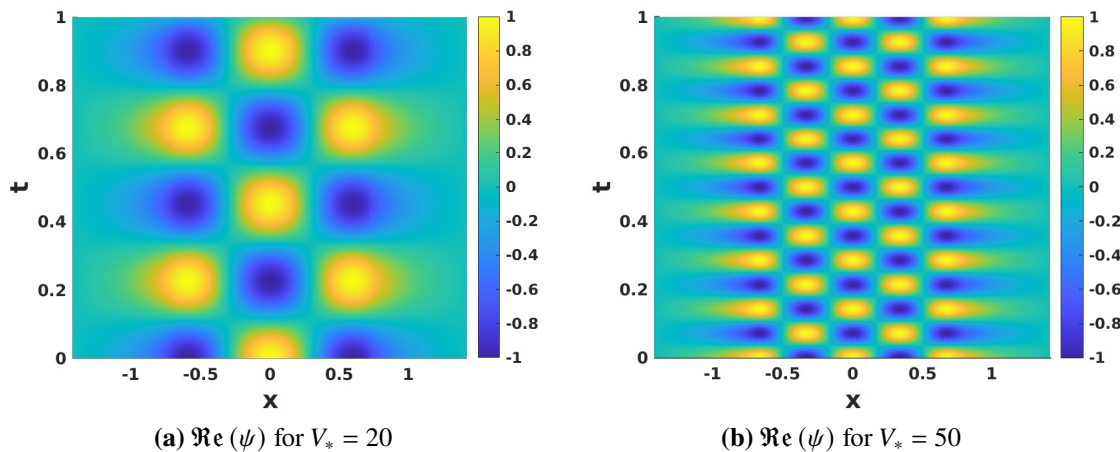
Figure 4.6 Exact solution ψ in the space–time cylinder Q_T for the (1+1)-dimensional square-well potential problem (4.17).

In Figure 4.7a, we plot errors obtained for $V_* = 20$ and a sequence of space–time, uniform, Cartesian meshes with $h_x = 2\sqrt{2}h_t = 0.1 \times 2^{-i}$, $i = 1, \ldots, 4$ and $p = 1, 2, 3$. The expected convergence rates of order $O(h^p)$ are observed for the error in the DG norm, while optimal convergence rates of order $O(h^{p+1})$ are obtained for the $L^2(\Omega)$ error at the final time.

43

(a) Error in DG norm

(b) Error in L_2 norm at $T = 1$
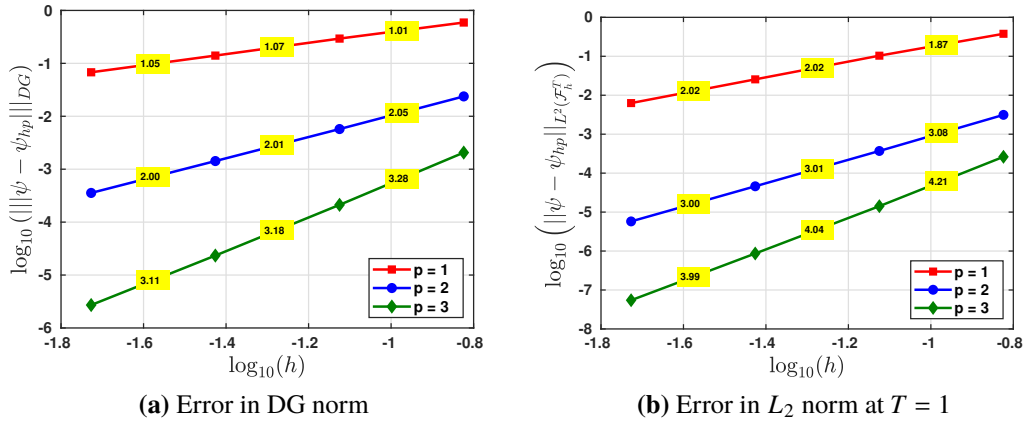
Figure 4.7 h-convergence of the Trefftz-DG method with $\mathbb{T}^p(\mathcal{T}_h)$ defined in (4.1), for the $(1+1)$-dimensional problem with square well potential (4.17) with $V_* = 20$.

Since we have homogeneous Dirichlet boundary conditions, the continuous model preserves the energy functional $\mathcal{E}(t, \psi)$, recall Remark 2.9. In Figure 4.8a, we show the time-evolution of the energy error for the Trefftz-DG approximation for the coarsest mesh, which is smaller for larger p, as expected. Moreover, in Figure 4.8b, we numerically observe that $\mathcal{E}_{loss}$ converges to zero as $O(h^{2p})$, as it can be proved combining Theorem 4.4 and Remark 2.9.

(a) Energy error evolution

(b) Energy loss at $T = 1$
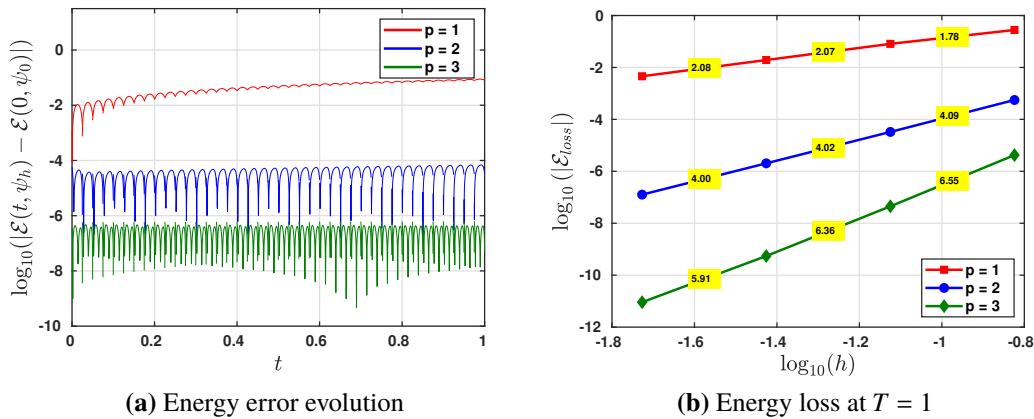
Figure 4.8 Time-evolution of the energy error and dependence on h and p for the problem with square-well potential (4.17) and $V_* = 20$.

In order to see the effect of the choice of the parameters k_ℓ, we first note that, in this experiment, we know the time frequency of the exact solution, which is $\omega = k_*^2$. Therefore it is natural to expect the approximation to be better if our basis functions oscillate at the same time frequency. To numerically illustrate this, in Figures 4.9a and 4.9b, we show that the convergence rates clearly degrade for $p = 1$, $V_* = 50$ and $V_* = 100$, and our previous choice of the parameters k_ℓ, as the time frequencies of the basis functions are too far from those of the exact solution. On the contrary, by taking the parameters k_ℓ as $\{-\sqrt{2}k_*, 0, \sqrt{2}k_*\}$ we recover the expected rates. This clearly suggests that a sensible tuning of the basis function parameters can significantly improve the accuracy of the method.

(a) $V_* = 50$ **(b)** $V_* = 100$
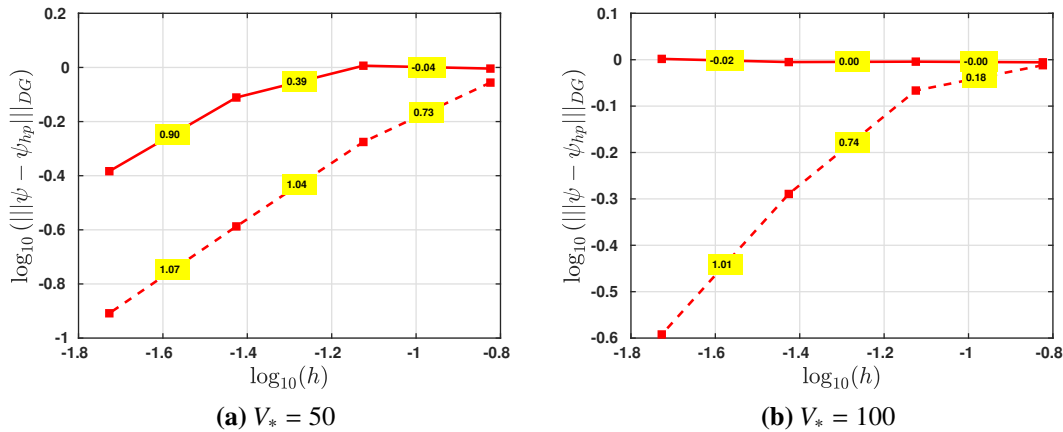
Figure 4.9 h-convergence of the Trefftz-DG method with $\mathbb{T}^p(\mathcal{T}_h)$ defined in (4.1), for the $(1+1)$-dimensional problem with square-well potential (4.17), $p = 1$, $V_* = 50$ ($k_* \approx 6.6394$) and $V_* = 100$ ($k_* \approx 9.6812$). The parameters k_ℓ are chosen as: $k_\ell \in \{-2, 0, 2\}$ (continuous line), which is the same choice of the previous plots, and $k_\ell \in \{0, \pm\sqrt{2}k_*\}$ (dashed line).

4.5.2   Transient Gaussian distribution in $(2 + 1)$-dimensions

We consider the linear Schrödinger equation (1.1) with zero potential $V = 0$ on the space–time cylinder $Q_T = \Omega \times (0, \frac{1}{2})$, with $\Omega = (-1, 1)^2$. Following [2], the initial and boundary conditions are chosen such that the exact solution is

$$\psi(x, y, t) = \frac{i}{i - 2t}\mathrm{e}^{-\frac{i}{i-2t}(x^2+y^2+i(x+\frac{1}{2}t))}. \tag{4.18}$$

The basis function parameters are chosen as $p + 1$ equally spaced space wavenumbers $k_m = 1, 2, \ldots, p + 1$, and the equally spaced angles $\theta_{m,\lambda} = \frac{2\pi(\lambda-1)}{2m+1}$ in $(0, 2\pi]$, $\lambda = 1, \ldots, 2m + 1$, as in Figure 4.3.

In Figure 4.10a, we show the convergence rates of the Trefftz-DG approximation for a set of Cartesian meshes with $h_x = h_y = 4h_t = 0.2 \times 2^{-i}$, $i = 1, \ldots, 4$. The $O(h^p)$ convergence rates obtained are in agreement with Theorem 4.4. Similarly to the $(1 + 1)$-dimensional example in the previous section, in Figure 4.10b, we observe $O(h^{p+1})$ convergence rates in the $L_2(\Omega)$ norm at the final time.

In Figure 4.11, we study the p-convergence of the method: for the coarsest space meshes with $h_x = h_y = 4h_t = 0.1$, and for $p = 1, \ldots, 8$, the DG norm of the error is plotted against the squared root of the total number of degrees of freedom in Q_T. We observe $O(\mathrm{e}^{-b\sqrt{N_{DoFs}}})$ convergence when the "local degree" p is raised. This is in strong contrast with what one might expect from a method based on standard polynomial spaces: in that case only the slower rate $O(\mathrm{e}^{-b\sqrt[3]{N_{DoFs}}})$ can be achieved (see Figures 5.7 and 5.12 in Chapter 5, and recall Remark 4.9). In Figure 4.12, we plot the exact solution ψ in (4.18) at the initial and final times.

(a) Error in DG norm

(b) Error in the L_2 norm at $T = \frac{1}{2}$
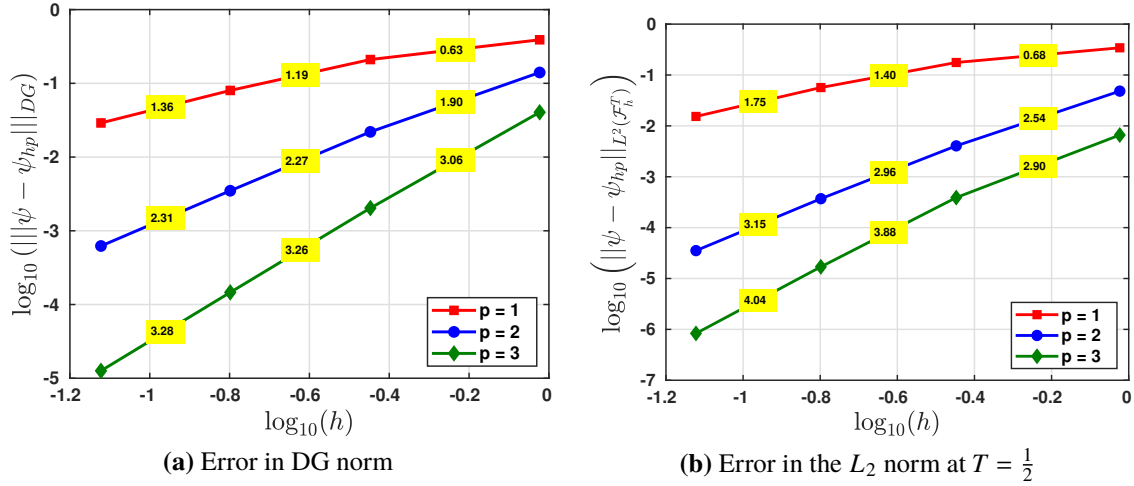
Figure 4.10 h-convergence of the Trefftz-DG method with $\mathbb{T}^p(\mathcal{T}_h)$ defined in (4.1), for the $(2+1)$-dimensional transient Gaussian problem with exact solution ψ in (4.18).

(a) Error in DG norm

(b) Error in the L^2 norm at $T = \frac{1}{2}$
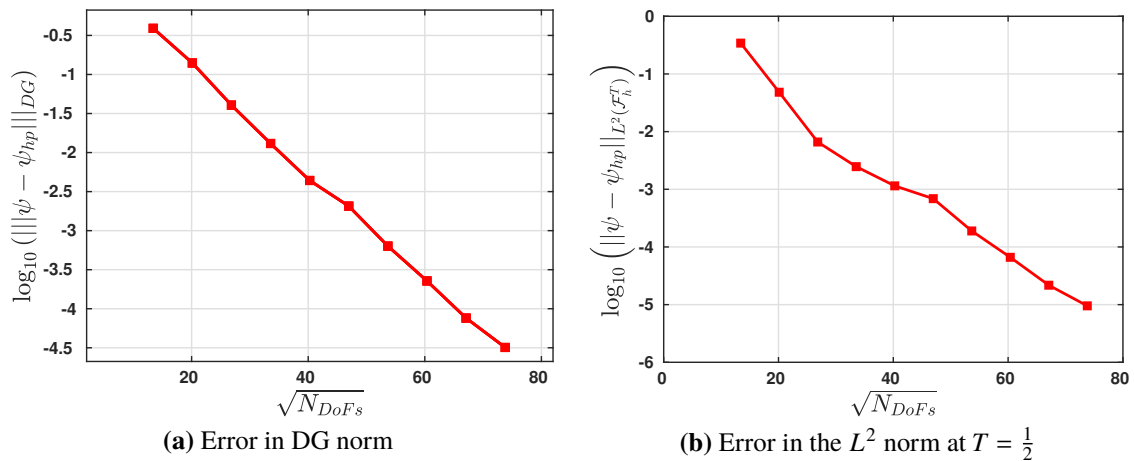
Figure 4.11 p-convergence of the Trefftz-DG error against the squared root of the total number of degrees of freedom for the $(2+1)$-dimensional transient Gaussian problem with exact solution ψ in (4.18) on a Cartesian space–time mesh with $h_x = h_y = 4h_t = 0.1$.

(a) $t = 0$

(b) $t = 0.5$
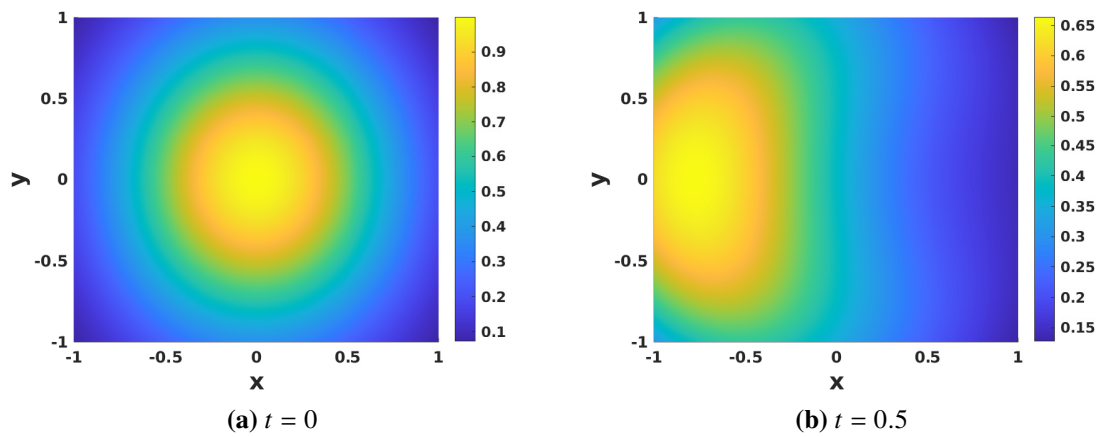
Figure 4.12 Real part of the exact solution ψ in (4.18) of the $(2 + 1)$-dimensional transient Gaussian problem.

Chapter 5

A quasi-Trefftz space for piecewise-smooth potentials

Contents

5.1 Introduction

In this chapter, we consider the time-dependent Schrödinger equation (1.1) with piecewise-smooth potential V. We analyze the convergence of the ultra-weak DG variational formulation (2.3) for two different discrete spaces: *i)* the space of polynomials in $\mathbb{R}^{d+1}$; *ii)* a quasi-Trefftz polynomial subspace with much smaller dimension.

Pure Trefftz methods are essentially limited to problems with piecewise-constant coefficients, as for PDEs with varying coefficients the design of "rich enough" finite-dimensional Trefftz spaces is in general not possible. A way to overcome this limitation is the use of quasi-Trefftz methods, which are based on spaces containing functions that are just approximate local solutions to the PDE. In essence, the earliest quasi-Trefftz spaces are the generalized plane waves used in [54] for the discretization of the Helmholtz equation with smoothly varying coefficients. More recently, a quasi-Trefftz DG method for the acoustic wave equation with piecewise-smooth material parameters was proposed in [55], where some polynomial quasi-Trefftz spaces were introduced. As an alternative idea, the embedded Trefftz DG method proposed in [72] does not require the local basis functions to be known in advance, as they are simply taken as a basis for

the kernel of the local discrete operators in a standard DG formulation. This corresponds to a Galerkin projection of a DG formulation with a predetermined discrete space onto a Trefftz-type subspace. In practice, it requires the computation of singular or eigenvalue decompositions of the local matrices.

Building on [55], for elementwise smooth potentials, we design and analyze a quasi-Trefftz polynomial space with similar approximation properties of full polynomial spaces but with much smaller dimension, thus substantially reducing the total number of degrees of freedom required for a given accuracy.

Structure of the chapter: In Sections 5.2 and 5.3, we prove optimal h-convergence estimates for the method when the test and trial spaces are taken as the space of piecewise polynomials in $\mathbb{R}^{d+1}$ or a novel quasi-Trefftz space, respectively. In Section 5.4, we present some numerical experiments that validate our theoretical results and illustrate the advantages of the proposed method.

5.2 Full polynomial space

In next theorem, we derive *a priori* error estimates for the DG formulation (2.3) with the space of elementwise polynomials

$$\mathbb{V}_{hp}(\mathcal{T}_h) = \prod_{K \in \mathcal{T}_h} \mathbb{P}^p(K). \tag{5.1}$$

Theorem 5.1. *Let $p \in \mathbb{N}$, and fix α, β, δ and μ as in(2.2), and assume that $V \in L^\infty(Q_T)$. Let $\psi \in \mathbf{V}(\mathcal{T}_h) \cap H^{p+1}(\mathcal{T}_h)$ be the exact solution of (1.1) and $\psi_{hp} \in \mathbb{V}_{hp}(\mathcal{T}_h)$ be the solution to the variational formulation (2.3) with $\mathbb{V}_{hp}(\mathcal{T}_h)$ given by (5.1). Then, if $h_{K_\mathbf{x}} \simeq h_{K_t}$ for all $K = K_\mathbf{x} \times K_t \in \mathcal{T}_h$, there exists a positive constant C independent of the element sizes $h_{K_\mathbf{x}}, h_{K_t}$, but depending on the degree p, the parameter ϵ, the $L^\infty(Q_T)$ norm of V, the trace inequality constant C_{tr} in (2.10), the local quasi-uniformity parameter $\mathsf{lqu}(\mathcal{T}_h)$ and the star-shapedness parameter ρ such that*

$$|||\psi - \psi_{hp}|||_{\mathrm{DG}} \le C \sum_{K = K_\mathbf{x} \times K_t \in \mathcal{T}_h} \max\{h_{K_\mathbf{x}}, h_{K_t}\}^p \, |\psi|_{H^{p+1}(K)}.$$

Proof. The proof follows from the choice of the volume penalty function μ and the stabilization functions α, β in (2.2), the quasi-optimality bound (2.8), Proposition 2.8, the inequality $\sqrt{|\mathbf{v}|_1} \le \sum_{i=1}^N \sqrt{|v_i|} \; \forall \mathbf{v} \in \mathbb{R}^N$, the fact that $Q^{p+1}[\psi_{|K}] \in \mathbb{V}_{hp}(K)$ for all elements $K \in \mathcal{T}_h$, and the Bramble-Hilbert lemma 2.11. □

5.3 Quasi-Trefftz spaces

We now introduce a polynomial quasi-Trefftz space. Let $p \in \mathbb{N}$, and assume that $V \in C^{p-2}(K)$. For each $K \in \mathcal{T}_h$, we define the following local polynomial quasi-Trefftz space:

$$\mathbb{QT}^p(K) := \left\{ q_p \in \mathbb{P}^p(K) : D^j \mathcal{S}_\epsilon q_p(\mathbf{x}_K, t_K) = 0, \; |j| \le p - 2 \right\}, \tag{5.2}$$

for some point $(\mathbf{x}_K, t_K)$ in K. We denote the local dimensions $n_{d+1,p} := \dim(\mathbb{QT}^p(K))$ and $r_{d+1,p} := \dim(\mathbb{P}^p(K))$ in dependence of the space dimension d of the problem and the

polynomial degree p, but independent of the element K. We have intentionally used $n_{d+1,p}$ to denote the dimension of the local quasi-Trefftz space, which was also used in Propositions 4.6 and 4.7 to denote the dimension of the local pseudo-plane wave Trefftz spaces $\mathbb{T}^p$ in (4.1). In fact, the dimension of both spaces that guarantees convergence of order $\mathcal{O}(h^p)$ for the error in the $||| \cdot |||_{\mathrm{DG}}$-norm coincides, see Proposition 5.4 below.

We consider the following global discrete space

$$\mathbb{V}_{hp}(\mathcal{T}_h) = \prod_{K \in \mathcal{T}_h} \mathbb{Q}\mathbb{T}^p(K). \tag{5.3}$$

For all $\boldsymbol{j} = (\boldsymbol{j_x}, j_t) \in \mathbb{N}^{d+1}$, if $V \in C^{|\boldsymbol{j}|}(K)$ and $f \in C^{|\boldsymbol{j}|+2}(K)$, then by the multi-index Leibniz product rule for multivariate functions we have

$$\begin{aligned}
D^{\boldsymbol{j}}\mathcal{S}_\epsilon f(\mathbf{x}_K, t_K) = & i\epsilon D^{\boldsymbol{j_x}, j_t+1} f(\mathbf{x}_K, t_K) + \frac{\epsilon^2}{2} \sum_{\ell=1}^d D^{\boldsymbol{j_x}+2\boldsymbol{e}_\ell, j_t} f(\mathbf{x}_K, t_K) \\
& - \sum_{\mathbf{z} \le \boldsymbol{j}} \binom{\boldsymbol{j}}{\mathbf{z}} D^{\boldsymbol{j}-\mathbf{z}} V(\mathbf{x}_K, t_K) D^{\mathbf{z}} f(\mathbf{x}_K, t_K),
\end{aligned} \tag{5.4}$$

where $\{\boldsymbol{e}_\ell\}_{\ell=1}^d \subset \mathbb{R}^d$ is the canonical basis.

The next proposition is the key ingredient to prove optimal convergence rates in Theorem 5.3 for the DG method (2.3) when $\mathbb{V}_{hp}(\mathcal{T}_h)$ is chosen as the quasi-Trefftz polynomial space defined in (5.2).

Proposition 5.2. *Let* $p \in \mathbb{N}$ *and* $K \in \mathcal{T}_h$. *Assume that* $V \in C^{\max\{p-2,0\}}(K)$, *and* $\psi \in C^p(K)$ *satisfies* $\mathcal{S}_\epsilon\psi = 0$ *in* K, *then the Taylor polynomial* $T^{p+1}_{(\mathbf{x}_K, t_K)}[\psi] \in \mathbb{Q}\mathbb{T}^p(K)$.

Proof. By the definition of the Taylor polynomial, $T^{p+1}_{(\mathbf{x}_K, t_K)}[\psi] \in \mathbb{P}^p(K)$. Therefore, it only remains to show that $D^{\boldsymbol{j}}\mathcal{S}_\epsilon T^{p+1}_{(\mathbf{x}_K, t_K)}[\psi](\mathbf{x}_K, t_K) = 0$ for all $|\boldsymbol{j}| \le p-2$. Taking $f = T^{p+1}_{(\mathbf{x}_K, t_K)}[\psi]$ in (5.4), all the derivatives of $T^{p+1}_{(\mathbf{x}_K, t_K)}[\psi]$ at $(\mathbf{x}_K, t_K)$ that appear in (5.4) are at most of total order $|\boldsymbol{j}| + 2 \le p$, so they coincide with the corresponding derivatives of ψ. Furthermore, since $\mathcal{S}_\epsilon\psi = 0$, then

$$D^{\boldsymbol{j}}\mathcal{S}_\epsilon T^{p+1}_{(\mathbf{x}_K, t_K)}[\psi](\mathbf{x}_K, t_K) = D^{\boldsymbol{j}}\mathcal{S}_\epsilon\psi(\mathbf{x}_K, t_K) = 0,$$

which completes the proof. $\qquad\square$

Proposition 5.2 allows for the use of the Taylor error bound (2.15) in the analysis of the quasi-Trefftz DG scheme.

Theorem 5.3. *Let* $p \in \mathbb{N}$, *fix* α, β, δ *and* μ *as in* (2.2), *and assume that* $V \in L^\infty(Q_T) \cap C^{\max\{p-2,0\}}(\mathcal{T}_h)$. *Let* $\psi \in \mathbf{V}(\mathcal{T}_h) \cap C^{p+1}(\mathcal{T}_h)$ *be the exact solution of* (1.1) *and* $\psi_{hp} \in \mathbb{V}_{hp}(\mathcal{T}_h)$ *be the solution to the variational formulation* (2.3) *with* $\mathbb{V}_{hp}(\mathcal{T}_h)$ *given by* (5.3). *Then, if* $h_{K_\mathbf{x}} \simeq h_{K_t}$ *for all* $K = K_\mathbf{x} \times K_t \in \mathcal{T}_h$, *there exists a positive constant* C *independent of the mesh size* h, *but depending on the degree* p, *the parameter* ϵ, *the* $L^\infty(Q_T)$ *norm of* V, *the trace inequality constant* C_{tr} *in* (2.10), *the local quasi-uniformity parameter* $\mathsf{lqu}(\mathcal{T}_h)$ *and the measure of the space–time domain* Q_T *such that*

$$|||\psi - \psi_{hp}|||_{\mathrm{DG}} \le C \sum_{K=K_\mathbf{x}\times K_t \in \mathcal{T}_h} \max\{h_{K_\mathbf{x}}, h_{K_t}\}^p \, |\psi|_{C^{p+1}(K)}.$$

Proof. The proof follows from the choice of the volume penalty function μ and the stabilization functions α, β in (2.2), the quasi-optimality bound (2.8), bound (2.8), the inequality $\sqrt{|\mathbf{v}|_1} \leq \sum_{i=1}^{N} \sqrt{|v_i|} \; \forall \mathbf{v} \in \mathbb{R}^N$, Proposition 5.2, and the estimate (2.15). $\qquad\square$

The *a priori* error estimate in Theorem 5.3 requires stronger regularity assumptions on ψ than Theorem 5.1 (namely $\psi \in C^{p+1}(\mathcal{T}_h)$ instead of $\psi \in H^{p+1}(\mathcal{T}_h)$) due to the fact that $\mathbb{QT}^p(K)$ is tailored to contain the Taylor polynomial $T^{p+1}_{(\mathbf{x}_K, t_K)}[\psi]$, but in general it does not contain the averaged Taylor polynomial $Q^{p+1}[\psi]$.

5.3.1 Basis functions and dimension

So far, we have not specified the dimension and a basis for the space $\mathbb{QT}^p(K)$, which is the aim of this section.

Recalling that $r_{d,p} = \dim\left(\mathbb{P}^p(\mathbb{R}^d)\right) = \binom{p+d}{d}$, let $\{\widehat{m}_\alpha\}_{\alpha=1}^{r_{d,p}}$ and $\{\widetilde{m}_\beta\}_{\beta=1}^{r_{d,p-1}}$ be bases of $\mathbb{P}_p(\mathbb{R}^d)$ and $\mathbb{P}_{p-1}(\mathbb{R}^d)$, respectively. We define

$$n_{d+1,p} := r_{d,p} + r_{d,p-1} = \binom{p+d}{d} + \binom{p+d-1}{d} = \frac{(p+d-1)!(2p+d)}{d!p!},$$

and the following $n_{d+1,p}$ elements of $\mathbb{QT}^p(K)$

$$\left\{ b_J \in \mathbb{QT}^p(K) : \begin{cases} b_J\left(\mathbf{x}_K^{(1)}, \cdot\right) = \widehat{m}_J \text{ and } \partial_{x_1} b_J\left(\mathbf{x}_K^{(1)}, \cdot\right) = 0 & \text{if } J \leq r_{d,p} \\ b_J\left(\mathbf{x}_K^{(1)}, \cdot\right) = 0 \text{ and } \partial_{x_1} b_J\left(\mathbf{x}_K^{(1)}, \cdot\right) = \widetilde{m}_{J-r_{d,p}} & \text{if } r_{d,p} < J \leq n_{d+1,p} \end{cases} \right\},$$
$$\tag{5.5}$$

where $g\left(\mathbf{x}_K^{(1)}, \cdot\right)$ denotes the restriction of $g : K \to \mathbb{C}$ to $x_1 = \mathbf{x}_K^{(1)}$, and $\mathbf{x}_K^{(1)}$ is the first component of $\mathbf{x}_K \in \mathbb{R}^d$.

Any element $q_p \in \mathbb{QT}^p(K)$ can be expressed in the scaled monomial basis as

$$q_p(\mathbf{x}, t) = \sum_{|\boldsymbol{j}| \leq p} C_{\boldsymbol{j}} \left(\frac{\mathbf{x} - \mathbf{x}_K}{h_{K_\mathbf{x}}}\right)^{\boldsymbol{j}_\mathbf{x}} \left(\frac{t - t_K}{h_{K_t}}\right)^{j_t},$$

for some complex coefficients $\{C_{\boldsymbol{j}}\}_{|\boldsymbol{j}| \leq p}$. By the conditions $D^{\boldsymbol{j}} \mathcal{S}_\epsilon q_p(\mathbf{x}_K, t_K) = 0$, for all $|\boldsymbol{j}| \leq p - 2$, in the definition of $\mathbb{QT}^p(K)$, we have the following relations between the coefficients

$$\frac{i\epsilon}{h_{K_t}}(j_t + 1) C_{\boldsymbol{j}_\mathbf{x}, j_t+1} + \frac{\epsilon^2}{2h_{K_\mathbf{x}}^2} \sum_{\ell=1}^{d} (\boldsymbol{j}_{\mathbf{x}_\ell} + 1)(\boldsymbol{j}_{\mathbf{x}_\ell} + 2) C^J_{\boldsymbol{j}_\mathbf{x}+2\boldsymbol{e}_\ell, j_t}$$
$$- \sum_{\mathbf{z} \leq \boldsymbol{j}} \frac{h_{K_\mathbf{x}}^{|\boldsymbol{j}_\mathbf{x}|-|\mathbf{z}_\mathbf{x}|} h_{K_t}^{j_t - z_t}}{(\boldsymbol{j} - \mathbf{z})!} D^{\boldsymbol{j}-\mathbf{z}} V(\mathbf{x}_K, t_K) C^J_{\mathbf{z}} = 0,$$

which can be rewritten as

$$C_{\boldsymbol{j}_\mathbf{x}+2\boldsymbol{e}_1, j_t} = \frac{1}{(\boldsymbol{j}_{\mathbf{x}_1} + 1)(\boldsymbol{j}_{\mathbf{x}_1} + 2)} \left(-\frac{2ih_{K_t}}{\epsilon h_{K_\mathbf{x}}^2}(j_t + 1) C^J_{\boldsymbol{j}_\mathbf{x}, j_t+1} - \sum_{\ell=2}^{d} (\boldsymbol{j}_{\mathbf{x}_\ell} + 1)(\boldsymbol{j}_{\mathbf{x}_\ell} + 2) C^J_{\boldsymbol{j}_\mathbf{x}+2\boldsymbol{e}_\ell, j_t} \right.$$

$$+ \frac{2}{\epsilon^2} \sum_{\mathbf{z} \leq \boldsymbol{j}} \frac{h_{K_{\mathbf{x}}}^{|j_{\mathbf{x}}|-|z_{\mathbf{x}}|+2} h_{K_t}^{j_t+z_t}}{(\boldsymbol{j}-\mathbf{z})!} D^{\boldsymbol{j}-\mathbf{z}} V(\mathbf{x}_K, t_K) C_{\mathbf{z}}^{J} \right) \qquad |\boldsymbol{j}| \leq p - 2, \ \boldsymbol{j} = (\boldsymbol{j}_{\mathbf{x}}, j_t). \quad (5.6)$$

The conditions imposed in (5.5) on the restriction of b_J to $x_1 = \mathbf{x}_K^{(1)}$ fix the coefficients of their expansion for all $\boldsymbol{j}$ with $j_{x_1} \in \{0, 1\}$. The coefficients that are not immediately determined by the conditions in (5.5) (i.e., those for $j_{x_1} \geq 2$) are uniquely defined and can be computed for the $(1 + 1)$- and $(2 + 1)$-dimensional cases using the recurrence relation (5.6) exactly as in Figures 4.2 and 4.4.

Proposition 5.4. *The set of functions* $\{b_J\}_{J=1}^{n_{d+1,p}}$ *defined in* (5.5) *are a basis for the space* $\mathbb{QT}^p(K)$. *Therefore,*

$$\dim\left(\mathbb{QT}^p(K)\right) = n_{d+1,p} = \frac{(p + d - 1)!(2p + d)}{d!p!} = O_{p\to\infty}(p^d)$$

$$\ll \dim(\mathbb{P}^p(K)) = \binom{d + 1 + p}{d + 1} = O_{p\to\infty}(p^{d+1}).$$

Proof. We first observe that the set of polynomials $\{b_J\}_{J=1}^{n_{d+1,p}}$ is linearly independent due to their restrictions to $x_1 = \mathbf{x}_K^{(1)}$. Moreover, the relations (5.6) imply that q_p is uniquely determined by its restriction $q_p(\mathbf{x}_K^{(1)}, \cdot)$, and the restriction of its derivative $\partial_{x_1} q_p(\mathbf{x}_K^{(1)}, \cdot)$. In addition, there exist some complex coefficients $\{\lambda_s\}_{s=1}^{n_{d+1,p}}$ such that

$$q_p\left(\mathbf{x}_K^{(1)}, \cdot\right) = \sum_{s=1}^{r_{d,p}} \lambda_s \widehat{m}_s(\cdot) = \sum_{s=1}^{r_{d,p}} \lambda_s b_s\left(\mathbf{x}_K^{(1)}, \cdot\right),$$

$$\partial_{x_1} q_p\left(\mathbf{x}_K^{(1)}, \cdot\right) = \sum_{s=r_{d,p}+1}^{n_{d+1,p}} \lambda_s \widetilde{m}_{s-r_{d,p}}(\cdot) = \sum_{s=r_{d,p}+1}^{n_{d+1,p}} \lambda_s \partial_{x_1} b_s(\mathbf{x}_K^{(1)}, \cdot),$$

whence $q_p = \sum_{s=1}^{n_{d+1,p}} \lambda_s b_s$, which completes the proof. $\qquad \square$

Remark 5.5 (Quasi-Trefftz basis construction: difference between Schrödinger and wave equations)**.** The definition of the basis functions b_J in (5.5) can be modified by fixing the restriction of b_J and its partial derivative $\partial_{x_\ell} b_J$ to $x_\ell = \mathbf{x}_K^{(\ell)}$ for any $1 \leq \ell \leq d$. However, it is not possible to assign the values for a given time $t = t_K$, as the order of the time derivative appearing in the Schrödinger equation is lower than the order of the space derivatives. How this affects the basis construction is visible from Figure 4.2: the coefficients (colored dots) can be computed sequentially when all the other coefficients of a relation (Y-shaped stencil) are known, so it is possible to reach all dots moving left to right, but not moving bottom to top. Imposing the values at a given time is possible for the wave equation, as it is done in [55, §4.4], precisely because in that case time and space derivatives have the same order. $\qquad \blacksquare$

Remark 5.6 (Constant-potential case)**.** The space $\mathbb{QT}^p(K)$ does not reduce to a Trefftz space for the case of constant potential V. Nonetheless, as discussed in Chapter 3, the pure Trefftz space $\mathbb{P}_{\mathbf{T}}^p(K)$ defined in (3.3) as

$$\mathbb{P}_{\mathbf{T}}^p(K) = \left\{q_p \in \mathbb{P}^p(K) : \mathcal{S}_\epsilon q_p = 0\right\},$$

does not possess strong enough approximation properties to guarantee optimal h-convergence.
$\blacksquare$

Remark 5.7 (Trefftz dimension). As seen in Proposition 5.4, the quasi-Trefftz polynomial space has considerably lower dimension than the full polynomial space of the same degree. This "dimension reduction" is common to all Trefftz and quasi-Trefftz schemes. In particular, the dimension $n_{d+1,p}$ of $\mathbb{QT}^p(K)$ is equal to the dimension of the space of harmonic polynomials of degree $\leq p$ in $\mathbb{R}^{d+1}$, the Trefftz space of complex pseudo-plane wave functions for the Schrödinger equation with piecewise-constant potential in Chapter 4, the Trefftz and quasi-Trefftz polynomial space for the wave equation in [80, Eqns. (42)–(43)] and [55]. ∎

5.4 Numerical aspects

In the next sections we validate the theoretical results regarding the h-convergence of the proposed method, and numerically assess some additional features such as p-convergence and conditioning.

We list some aspects regarding our numerical experiments

- We use Cartesian-product space–time meshes with uniform partitions along each direction, which are a particular case of the situation described in Remark 2.2.

- We choose $(\mathbf{x}_K, t_K)$ in the definition of the quasi-Trefftz space $\mathbb{QT}^p(K)$ in (5.2) as the center of the element K.

- We focus on the dimensionless Schrödinger equation (1.1) with $\epsilon = 1$.

- In all the experiments, we consider Dirichlet boundary conditions.

- The linear systems are solved using Matlab's backslash command.

- The quasi-Trefftz basis functions $\{b_J\}_{J=1}^{n_{d+1,p}}$ are constructed by choosing $\widehat{m}_J$ and $\widetilde{m}_J$ in (5.5) as scaled monomials and by computing the remaining coefficients $C_{\mathbf{j}}$ with the relations (5.6).

- In the h-convergence plots, the numbers in the yellow rectangles are the empirical algebraic convergence rates for the quasi-Trefftz version (continuous lines). The dashed lines correspond to the errors obtained for the full polynomial space.

5.5 Numerical experiments in $(1+1)$ dimensions

We first focus on the $(1+1)$-dimensional case, for which families of explicit solutions are available for some well-known potentials V.

5.5.1 h-convergence

In order to validate the error estimates in Theorems 5.1 and 5.3, we consider a series of problems with different potentials V. No significant difference in terms of accuracy between the quasi-Trefftz and the full polynomial versions of the method with the same polynomial degree p (corresponding to different numbers of DOFs $n_{d+1,p}$ and $r_{d+1,p}$, respectively) is observed in all the experiments.

Harmonic oscillator potential $(V(x) = \frac{\omega^2 x^2}{2})$

For this potential, the Schrödinger equation (1.1) models the situation of a quantum harmonic oscillator for an angular frequency $\omega > 0$. On $Q_T = (-3, 3) \times (0, 1)$, we consider the following well-known family of solutions (see e.g., [47, §2.3])

$$\psi_n(x, t) = \frac{1}{\sqrt{2^n n!}} \left(\frac{\omega}{\pi}\right)^{1/4} \mathcal{H}_n\left(\sqrt{\omega}x\right) \exp\left(-\frac{1}{2}\left(\omega x^2 + (2n+1)i\omega t\right)\right) \qquad n \in \mathbb{N}, \qquad (5.7)$$

where $\mathcal{H}_n(\cdot)$ denotes the n-th physicist's Hermite polynomials as defined in [84, Table 18.3.1, denoted by $H_n(\cdot)$].

In Figure 5.1, we present the errors obtained for $\omega = 10$, $n = 2$ and a sequence of Cartesian meshes with uniform partitions and $h_x = h_t = 0.05 \times 2^{-i}$, $i = 0, \ldots 4$. Rates of convergence of order $O(h^p)$ in the DG norm are observed, as predicted by the error estimate in Theorem 5.3. A convergence of at least order $O(h^{p+1})$ is observed for the $L^2(\Omega)$-error at the final time, which is faster (by a factor h) than the order that can be deduced from the estimates in Theorems 5.1 and 5.3.

(a) DG error

(b) L^2 error at final time
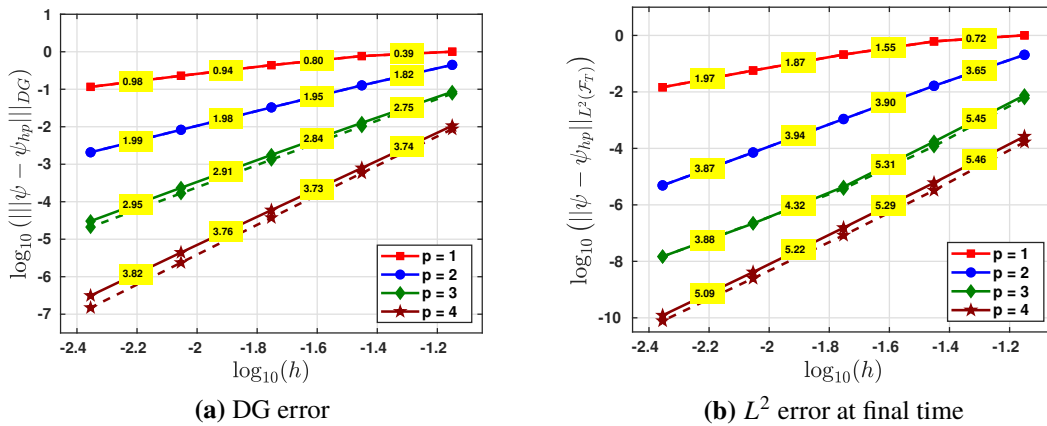
Figure 5.1 h-convergence for the $(1 + 1)$-dimensional quantum harmonic oscillator problem with potential $(V(x) = 50x^2)$ and exact solution ψ_2 in (5.7).

Due to the fast decay of the exact solution close to the boundary (see Figure 5.6a), the energy is expected to be preserved, see Remark 2.9. In Figure 5.2, we show the evolution of the energy error, and the convergence of the energy loss $\mathcal{E}_{loss}$ to zero for the quasi-Trefftz version. In the latter, rates of order $O(h^{2p})$ are observed, which follows from Remark 2.9, and the error estimates in Theorems 5.1 and 5.3.

Reflectionless potential $(V(x) = -a^2 \mathrm{sech}^2(ax))$

This potential was studied in [17] as an example of a reflectionless potential. On the space–time domain $Q_T = (-5, 5) \times (0, 1)$, we consider the Schrödinger equation with exact solution (see [47, Problem 2.48])

$$\psi(x, t) = \left(\frac{\sqrt{2}i - a\tanh(ax)}{\sqrt{2}i + a}\right) \exp\left(i\left(\sqrt{2}x - t\right)\right). \qquad (5.8)$$

(a) Energy error evolution

(b) Energy loss at $T = 1$
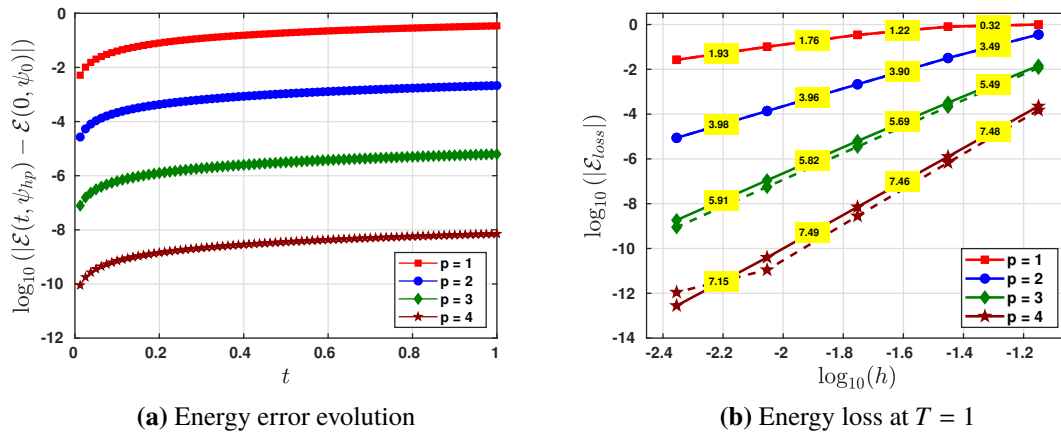
Figure 5.2 Time-evolution of the energy error for the quantum harmonic oscillator problem with potential $\left(V(x) = 50x^2\right)$ and exact solution ψ_2 in (5.7).

In Figure 5.3, we show the errors obtained for a sequence of meshes with $h_x = 2h_t = 0.2 \times 2^{-i}$, $i = 0, \ldots, 4$, and $a = 1$. As in the previous experiment, rates of convergence of order $O\left(h^p\right)$ and $O\left(h^{p+1}\right)$ are observed for the errors in the DG norm and the $L^2(\Omega)$-norm at the final time, respectively. The real part of the exact solution is depicted in Figure 5.6b.

(a) DG error

(b) L^2 error at $T = 1$
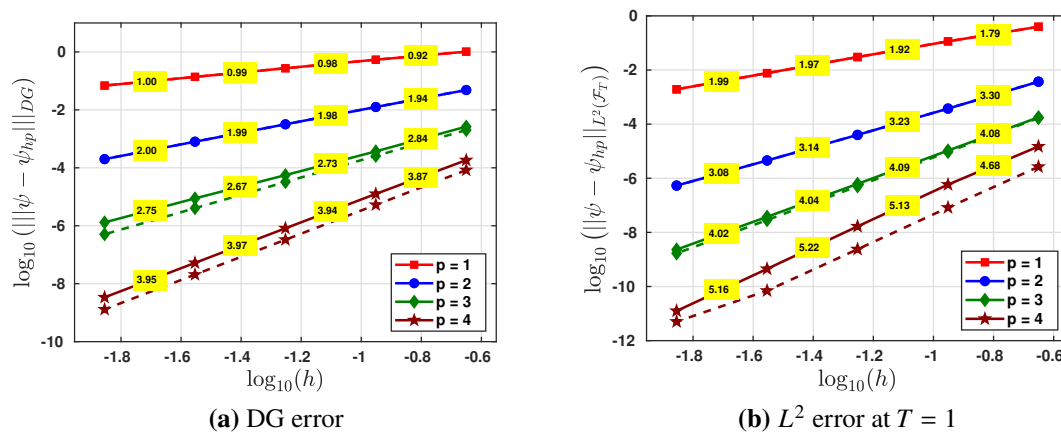
Figure 5.3 h-convergence for the $(1+1)$-dimensional problem with potential $V(x) = -\mathrm{sech}^2(x)$ and exact solution (5.8).

Morse potential $\left(V(x) = D(1 - \mathrm{e}^{-\alpha x})^2\right)$

This potential was introduced by Morse in [82] to obtain a quantum-mechanical energy level spectrum of a vibrating, non-rotating diatomic molecule. There, the following family of solutions was presented (see also [18])

$$
\begin{aligned}
\psi_{\lambda,n}(x,t) = {} & N(\lambda,n)\xi(x)^{\lambda-n-1/2}\mathbb{L}_n^{(2\lambda-2n-1)}(\xi(x)) \\
& \times \exp\left(-\frac{\xi(x)}{2} - it\left[(n+1/2) - \frac{1}{2\lambda}(n+1/2)^2\right]\omega_o\right),
\end{aligned}
\tag{5.9}
$$

where $\lfloor\cdot\rfloor$ is the floor function, $n = 0,\ldots,\lfloor\lambda-1/2\rfloor$, $\mathbb{L}_n^{(\alpha)}$ denote the general associated Laguerre polynomials as defined in [84, Table 18.3.1] and

$$N(\lambda,n) = \left\lfloor \frac{(2\lambda - 2n - 1)\Gamma(n+1)}{\Gamma(2\lambda - n)} \right\rfloor^{1/2}, \quad \lambda = \frac{\sqrt{2D}}{\alpha}, \quad \xi(x) = 2\lambda\exp(-\alpha x), \quad \omega_o = \sqrt{2D}\alpha.$$

In Figure 5.4, we show the errors obtained for the Morse potential problem with $D = 8$, $\alpha = 4$ and exact solution $\psi_{1,1}$ on the space–time domain $Q_T = (-0.5, 1.5)\times(0,1)$ for a sequence of meshes with $h_x = h_t = 0.1 \times 2^{-i}$, $i = 0,\ldots,4$. The observed rates of convergence are in agreement with those obtained in the previous experiments. The real part of the exact solution is depicted in Figure 5.6c.

(a) DG error

(b) L^2 error at $T = 1$
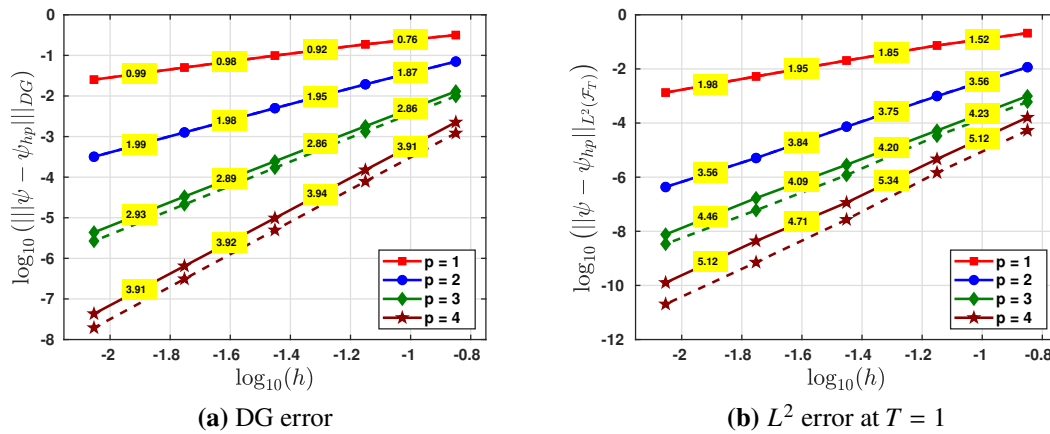
Figure 5.4 h-convergence for the $(1+1)$-dimensional problem with Morse potential $V(x) = D(1 - \exp(-\alpha x))^2$ for $D = 8$ and $\alpha = 4$ with exact solution (5.9).

Square-well potential We now consider the $(1+1)$-dimensional problem from Section 4.5, where the exact solution is not arbitrarily smooth. On the space–time domain $Q_T = (-\sqrt{2}, \sqrt{2})\times(0,1)$, we consider the Schrödinger equation with homogeneous Dirichlet boundary conditions and the following square-well potential

$$V(x) = \begin{cases} 0 & x \in (-\sqrt{2}/2, \sqrt{2}/2), \\ V_* & x \in (-\sqrt{2}, \sqrt{2}) \setminus (-\sqrt{2}/2, \sqrt{2}/2), \end{cases} \tag{5.10}$$

for some fixed $V_* > 0$. The initial condition is taken as an eigenfunction (bound state) of $-\frac{1}{2}\partial_x^2 + V$ on $(-\sqrt{2}, \sqrt{2})$:

$$\psi_0(x) = \begin{cases} \cos\left(k_*\sqrt{2}x\right) & x \in (-\sqrt{2}/2, \sqrt{2}/2), \\ \frac{\cos(k_*)}{\sinh(\sqrt{V_*-k_*^2})}\sinh\left(\sqrt{V_* - k_*^2}(2 - \sqrt{2}|x|)\right) & x \in (-\sqrt{2}, \sqrt{2}) \setminus (-\sqrt{2}/2, \sqrt{2}/2), \end{cases}$$

where k_* is a real root of the function $f(k) := \sqrt{V_* - k^2} - k\tan(k)\tanh(\sqrt{V_* - k^2})$. The solution of the corresponding initial boundary value problem (1.1) is $\psi(x,t) = \psi_0(x)\exp(-ik^2 t)$ and belongs to the space $C^\infty\left(I; C^1(\Omega)\right)\setminus C^\infty\left(I; C^2(\Omega)\right)$. Among the finite set of values k_* for a

given V_*, in this experiment we take the largest one, corresponding to faster oscillations in space and time.

In Figure 5.5, we show the errors obtained for $V_* = 20$ ($k_* \approx 3.73188$) and a sequence of meshes with $h_t = \sqrt{2}h_x = 0.1 \times 2^{-i}$, $i = 0, \dots, 4$. Optimal convergence is observed for the errors of the quasi-Trefftz version of the method in both norms.

(a) DG error

(b) L^2 error at $T = 1$
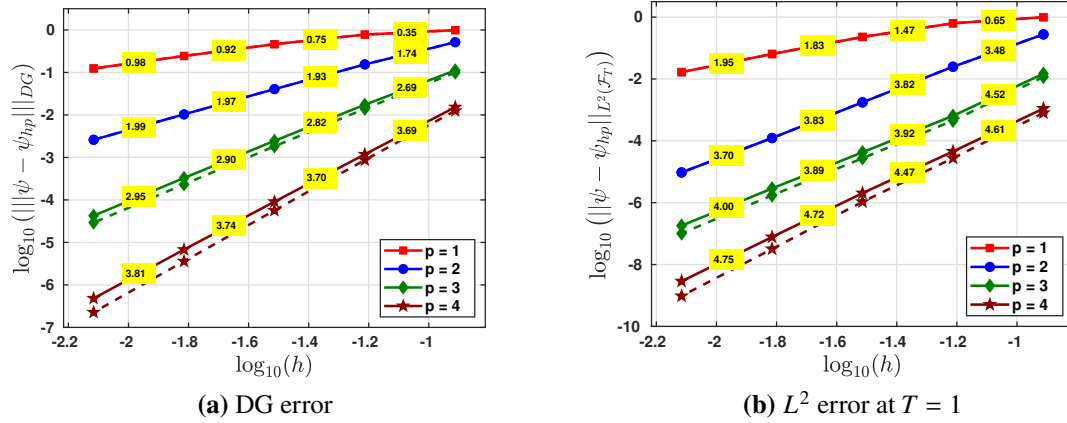
Figure 5.5 h-convergence for the $(1+1)$-dimensional problem with square-well potential $V(x)$ in (5.10).

(a) $V(x) = 50x^2$

(b) $V(x) = -\text{sech}^2(x)$

(c) $V(x) = 8\left(1 - \exp(-4x)\right)^2$

(d) Square-well potential ($V_* = 20$)
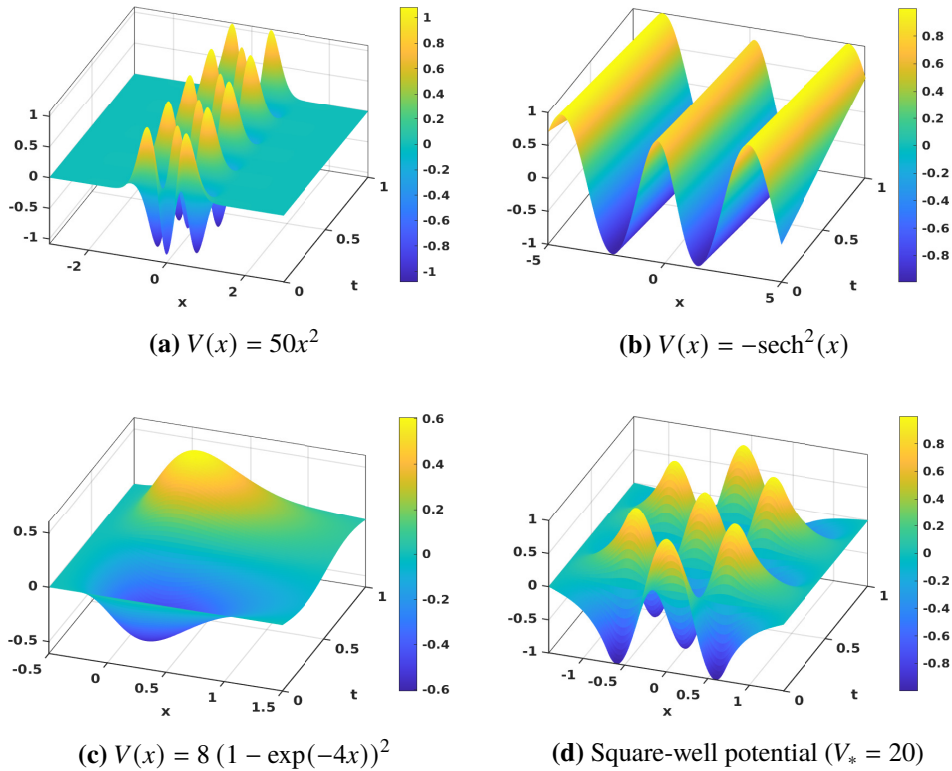
Figure 5.6 Real part of the exact solutions for the $(1+1)$-dimensional problems.

5.5.2 Effect of stabilization and volume penalty terms

In this experiment we are interested in the effect of neglecting some of the terms in the variational formulation (2.3). To do so, we consider the $(1 + 1)$-dimensional quantum harmonic oscillator problem with exact solution (5.7). In Tables 5.1–5.2 (quasi-Trefftz space) and 5.3–5.4 (full polynomial space) we present the errors in the DG-norm obtained for the same sequence of meshes and approximation degrees as in the previous section, for different combinations of the stabilization terms α, β and the volume penalty parameter μ. Although the proof of well-posedness of the method (2.3) relies on the assumption that α, β and μ are strictly positive, in our numerical experiments, the matrices of the arising linear systems are non-singular and optimal convergence rates are observed even when all these parameters are set to zero. Moreover, the errors obtained when $\alpha = 0$ or $\beta = 0$ are smaller as some terms in the definition (2.4a) of $|||\cdot|||_{\mathrm{DG}}$ vanish, while the presence of μ seems to have just a mild effect in the results.

h	$\alpha = \frac{1}{h_{F_\mathbf{x}}}, \beta = h_{F_\mathbf{x}}$		$\mu = \max\{h_{K_t}, h_{K_\mathbf{x}}\}$ $\alpha = 0, \beta = 0$		$\alpha = \frac{1}{h_{F_\mathbf{x}}}, \beta = 0$		$\alpha = 0, \beta = h_{F_\mathbf{x}}$	
	DG error	Rate	DG error	Rate	DG error	Rate	DG error	Rate
				$p = 1$				
7.07e-02	1.00e+00	—	9.81e-01	—	1.01e+00	—	1.00e+00	—
3.54e-02	7.67e-01	**0.39**	4.76e-01	**1.04**	6.72e-01	**0.58**	6.53e-01	**0.62**
1.77e-02	4.40e-01	**0.80**	2.14e-01	**1.15**	3.62e-01	**0.89**	3.40e-01	**0.94**
8.84e-03	2.29e-01	**0.94**	1.01e-01	**1.08**	1.85e-01	**0.97**	1.70e-01	**1.00**
4.42e-03	1.16e-01	**0.98**	4.96e-02	**1.03**	9.31e-02	**0.99**	8.49e-02	**1.00**
				$p = 2$				
7.07e-02	4.47e-01	—	2.59e-01	—	2.99e-01	—	4.37e-01	—
3.54e-02	1.27e-01	**1.82**	6.90e-02	**1.91**	8.24e-02	**1.86**	1.20e-01	**1.87**
1.77e-02	3.28e-02	**1.95**	1.78e-02	**1.96**	2.15e-02	**1.94**	3.05e-02	**1.97**
8.84e-03	8.29e-03	**1.98**	4.50e-03	**1.98**	5.48e-03	**1.97**	7.68e-03	**1.99**
4.42e-03	2.08e-03	**1.99**	1.13e-03	**1.99**	1.38e-03	**1.98**	1.93e-03	**2.00**

Table 5.1 h-convergence for the quasi-Trefftz version applied to the quantum harmonic oscillator problem with potential $V(x) = 50x^2$ and exact solution ψ_2 in (5.7) for different combinations of the stabilization parameters α, β and volume penalty parameter $\mu \neq 0$.

5.5.3 p-Convergence

We now study numerically the p-convergence of the method, i.e., for a fixed space–time mesh $\mathcal{T}_h$, we study the errors when increasing the polynomial degree p. We consider the $(1 + 1)$-dimensional problems above with the same parameters and the coarsest meshes for each case. In Figure 5.7, we compare the errors obtained for the method with the two choices for the discrete space $\mathbb{V}_{hp}(\mathcal{T}_h)$ analyzed in the previous sections: the full polynomial space (5.1) and the quasi-Trefftz polynomial space (5.3). As expected, for the quasi-Trefftz version we observe exponential decay of the error of order $O\left(e^{-bN_{DoFs}}\right)$, where N_{DoFs} denotes the total number of degrees of freedom. As for the full polynomial space, only root-exponential convergence $O\left(e^{-c\sqrt{N_{DoFs}}}\right)$ is expected. The superiority of the quasi-Trefftz version is evident in all cases.

h	$\alpha = \frac{1}{h_{F_\mathbf{x}}}, \beta = h_{F_\mathbf{x}}$		$\alpha = 0, \beta = 0$		$\alpha = \frac{1}{h_{F_\mathbf{x}}}, \beta = 0$		$\alpha = 0, \beta = h_{F_\mathbf{x}}$	
	DG error	Rate	DG error	Rate	DG error	Rate	DG error	Rate
				$p = 1$				
7.07e-02	1.04e+00	—	1.16e+00	—	1.07e+00	—	1.09e+00	—
3.54e-02	7.78e-01	**0.43**	5.02e-01	**1.21**	6.84e-01	**0.64**	6.69e-01	**0.70**
1.77e-02	4.42e-01	**0.81**	2.18e-01	**1.20**	3.64e-01	**0.91**	3.42e-01	**0.97**
8.84e-03	2.29e-01	**0.95**	1.02e-01	**1.09**	1.85e-01	**0.97**	1.71e-01	**1.00**
4.42e-03	1.16e-01	**0.99**	4.98e-02	**1.04**	9.32e-02	**0.99**	8.50e-02	**1.01**
				$p = 2$				
7.07e-02	4.63e-01	—	2.96e-01	—	3.23e-01	—	4.60e-01	—
3.54e-02	1.29e-01	**1.84**	7.38e-02	**2.00**	8.58e-02	**1.91**	1.23e-01	**1.90**
1.77e-02	3.31e-02	**1.97**	1.84e-02	**2.01**	2.19e-02	**1.97**	3.09e-02	**1.99**
8.84e-03	8.33e-03	**1.99**	4.58e-03	**2.00**	5.54e-03	**1.99**	7.73e-03	**2.00**
4.42e-03	2.09e-03	**2.00**	1.14e-03	**2.00**	1.39e-03	**1.99**	1.93e-03	**2.00**

The header row above the data has the centered label $\mu = 0$.

Table 5.2 h-convergence for the quasi-Trefftz version applied to the quantum harmonic oscillator problem with potential $V(x) = 50x^2$ and exact solution ψ_2 in (5.7) for different combinations of the stabilization parameters α, β and volume penalty parameter $\mu = 0$.

h	$\alpha = \frac{1}{h_{F_\mathbf{x}}}, \beta = h_{F_\mathbf{x}}$		$\alpha = 0, \beta = 0$		$\alpha = \frac{1}{h_{F_\mathbf{x}}}, \beta = 0$		$\alpha = 0, \beta = h_{F_\mathbf{x}}$	
	DG error	Rate	DG error	Rate	DG error	Rate	DG error	Rate
				$p = 1$				
7.07e-02	1.00e+00	—	9.81e-01	—	1.01e+00	—		
3.54e-02	7.67e-01	**0.39**	4.76e-01	**1.04**	6.72e-01	**0.58**	1.00e+00	—
1.77e-02	4.40e-01	**0.80**	2.14e-01	**1.15**	3.62e-01	**0.89**	3.40e-01	**0.94**
8.84e-03	2.29e-01	**0.94**	1.01e-01	**1.08**	1.85e-01	**0.97**	1.70e-01	**1.00**
4.42e-03	1.16e-01	**0.98**	4.96e-02	**1.03**	9.31e-02	**0.99**	8.49e-02	**1.00**
				$p = 2$				
7.07e-02	4.46e-01	—	2.55e-01	—	2.96e-01	—	4.34e-01	—
3.54e-02	1.27e-01	**1.81**	6.88e-02	**1.89**	8.22e-02	**1.85**	1.20e-01	**1.86**
1.77e-02	3.28e-02	**1.95**	1.77e-02	**1.95**	2.15e-02	**1.94**	3.05e-02	**1.97**
8.84e-03	8.29e-03	**1.98**	4.50e-03	**1.98**	5.48e-03	**1.97**	7.68e-03	**1.99**
4.42e-03	2.08e-03	**1.99**	1.13e-03	**1.99**	1.38e-03	**1.98**	1.93e-03	**2.00**

The header row above the data has the centered label $\mu = \max\{h_{K_t}, h_{K_\mathbf{x}}\}$.

Table 5.3 h-convergence for the full polynomial version applied to the quantum harmonic oscillator problem with potential $V(x) = 50x^2$ and exact solution ψ_2 in (5.7) for different combinations of the stabilization parameters α, β and volume penalty parameter $\mu \neq 0$.

Exponential convergence of space–time Trefftz and quasi-Trefftz schemes has been observed in several cases [54, 7, 85] (see also Section 4.5) but no proof is available yet (differently from the stationary case, [52, §3]).

h	$\mu = 0$							
	$\alpha = \frac{1}{h_{F_{\mathbf{x}}}}, \beta = h_{F_{\mathbf{x}}}$		$\alpha = 0, \beta = 0$		$\alpha = \frac{1}{h_{F_{\mathbf{x}}}}, \beta = 0$		$\alpha = 0, \beta = h_{F_{\mathbf{x}}}$	
	DG error	Rate	DG error	Rate	DG error	Rate	DG error	Rate
				$p = 1$				
7.07e-02	1.04e+00	—	1.16e+00	—	1.07e+00	—	1.09e+00	—
3.54e-02	7.78e-01	**0.43**	5.02e-01	**1.21**	6.84e-01	**0.64**	6.69e-01	**0.70**
1.77e-02	4.42e-01	**0.81**	2.18e-01	**1.20**	3.64e-01	**0.91**	3.42e-01	**0.97**
8.84e-03	2.29e-01	**0.95**	1.02e-01	**1.09**	1.85e-01	**0.97**	1.71e-01	**1.00**
4.42e-03	1.16e-01	**0.99**	4.98e-02	**1.04**	9.32e-02	**0.99**	8.50e-02	**1.01**
				$p = 2$				
7.07e-02	4.63e-01	—	2.93e-01	—	3.22e-01	—	4.57e-01	—
3.54e-02	1.29e-01	**1.84**	7.36e-02	**1.99**	8.57e-02	**1.91**	1.23e-01	**1.90**
1.77e-02	3.31e-02	**1.97**	1.84e-02	**2.00**	2.19e-02	**1.97**	3.09e-02	**1.99**
8.84e-03	8.33e-03	**1.99**	4.58e-03	**2.00**	5.54e-03	**1.98**	7.72e-03	**2.00**
4.42e-03	2.09e-03	**2.00**	1.14e-03	**2.00**	1.39e-03	**1.99**	1.93e-03	**2.00**

Table 5.4 h-convergence for the full polynomial version applied to the quantum harmonic oscillator problem with potential $V(x) = 50x^2$ and exact solution ψ_2 in (5.7) for different combinations of the stabilization parameters α, β and volume penalty parameter $\mu = 0$.

5.5.4 Singular solution

We consider the $(1 + 1)$-dimensional problem on the space–time domain $Q_T = (0, 1) \times (0, 0.1)$ with homogeneous Dirichlet boundary conditions, and initial condition $\psi_0(x) = Ax(1 - x)$ with normalization constant $A = \sqrt{30}$. The exact solution is given by (see [48, Example 2.2, Ch. 2])

$$\psi(x, t) = \sqrt{30} \left(\frac{2}{\pi}\right)^3 \sum_{n=1}^{\infty} \frac{1}{(2n + 1)^3} \sin\left((2n + 1)\pi x\right) e^{\frac{-i(2n+1)^2 \pi^2 t}{2}}, \qquad (5.11)$$

which belongs to $H^{\frac{5}{4} - \epsilon}(0, T; H_0^1(\Omega))$ for all $\epsilon > 0$; cf. [91, § 7.1]. In Figure 5.8, we show the errors in the DG norm obtained for a sequence of meshes with $h_t = 0.1 h_x = 0.05 \times 2^{-j}$, $j = 0, \ldots, 4$, and different discrete spaces: the polynomial Trefftz space $\mathbb{P}_{\mathbf{T}}^{2p}$ (Fig. 5.8a); the quasi-Trefftz polynomial space (Fig. 5.8b); the full polynomial space $\mathbb{P}^p$ (Fig. 5.8c); the pseudo-plane wave Trefftz space (Fig. 5.8d). To compute the errors, we have truncated the series in (5.11) to $n = 250$. Slightly reduced rates of convergence are observed for the polynomial and the pseudo-plane wave Trefftz spaces, compared to the quasi-Trefftz and full-polynomial spaces; however, for the above parameter choices, the discrete Trefftz spaces give the smallest errors. Such a reduction of the convergence rates is expected for the polynomial Trefftz space, as the error estimate in Theorem 3.6 requires stronger regularity assumptions in space on the exact solution. This motivates to look for sharper error estimates, especially for singular problems.

5.5.5 Conditioning

We now assess the conditioning of the stiffness matrix. In Figure 5.9, we compare the $\kappa_2(\cdot)$-condition number for the stiffness matrix $\mathbf{K}_n$ defined in Remark 2.2, for the free particle prob-
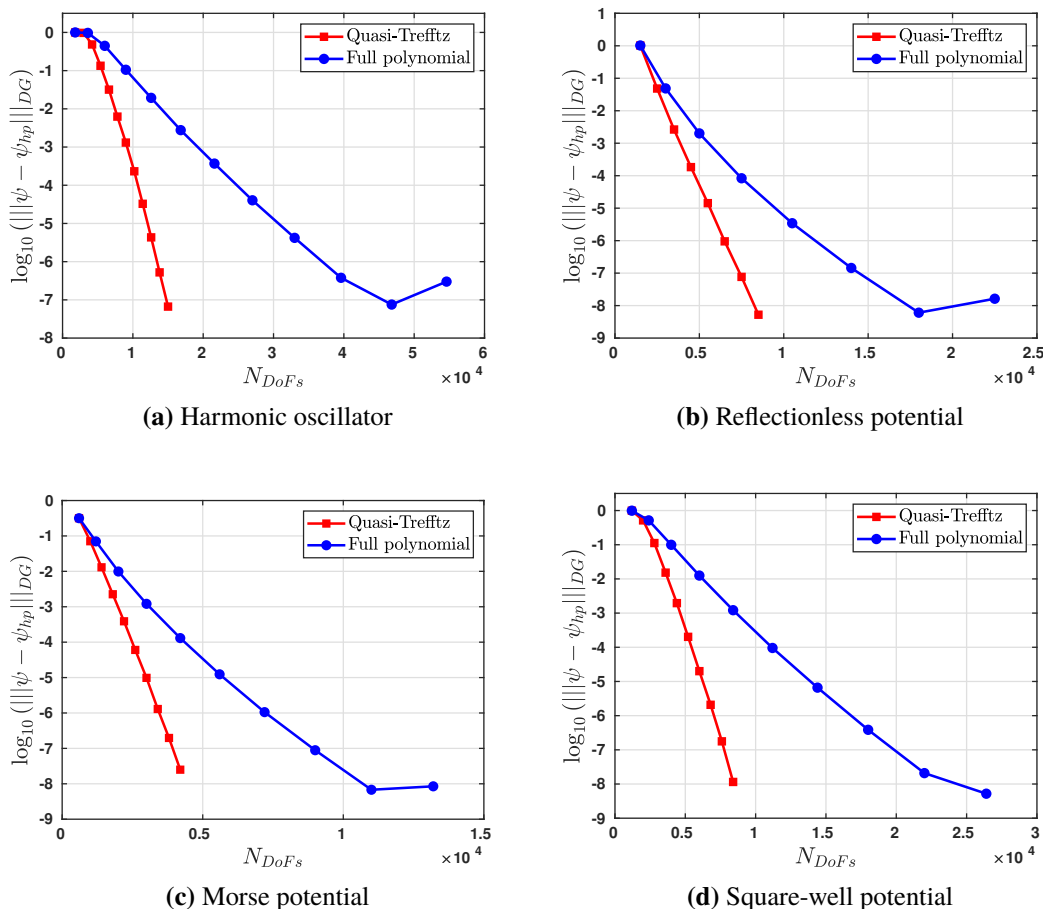
Figure 5.7 *p*-convergence for the coarsest mesh in the $(1+1)$-dimensional problems.

lem ($V = 0$) on the space–time domain $Q_T = (0,1) \times (0,1)$. We consider the polynomial Trefftz space (3.3) of degree $2p$, the pure-Trefftz space of complex pseudo-plane wave functions $\mathbb{T}^p(\mathcal{T}_h)$ in (4.1), the full polynomial space, and the quasi-Trefftz space in (5.2). We present some details for this experiment:

- For the construction of the polynomial Trefftz basis functions $\{b_J^{\mathbf{T}}\}_{J=1}^{r_{1,2p}}$, we consider two choices of $\{m_J\}_{J=1}^{r_{1,2p}}$ in (3.6):

$$m_J(x) := \left(\frac{x - x_K}{h_x}\right)^J \qquad J = 1, \ldots, 2p+1, \tag{5.12a}$$

$$m_J(x) := \frac{(x - x_K)^J}{h_x^{\lfloor J/2 \rfloor}} \qquad J = 1, \ldots, 2p+1, \tag{5.12b}$$

where $\lfloor \cdot \rfloor$ is the floor function. The remaining coefficients C_j are computed with the relations (3.7).

- We consider two choices for the parameters k_ℓ in the definition of the basis functions $\{\phi_\ell\}$ in (4.7): the arbitrary choice used in Section 4.5 $k_\ell = -p, \ldots, p$, and the choice $k_\ell/2 = 2\pi\ell/h_x$ which makes the basis orthogonal in each element.

(a) Polynomial Trefftz space

(b) Quasi-Trefftz space

(c) Full polynomial space

(d) Pseudo-plane wave Trefftz space
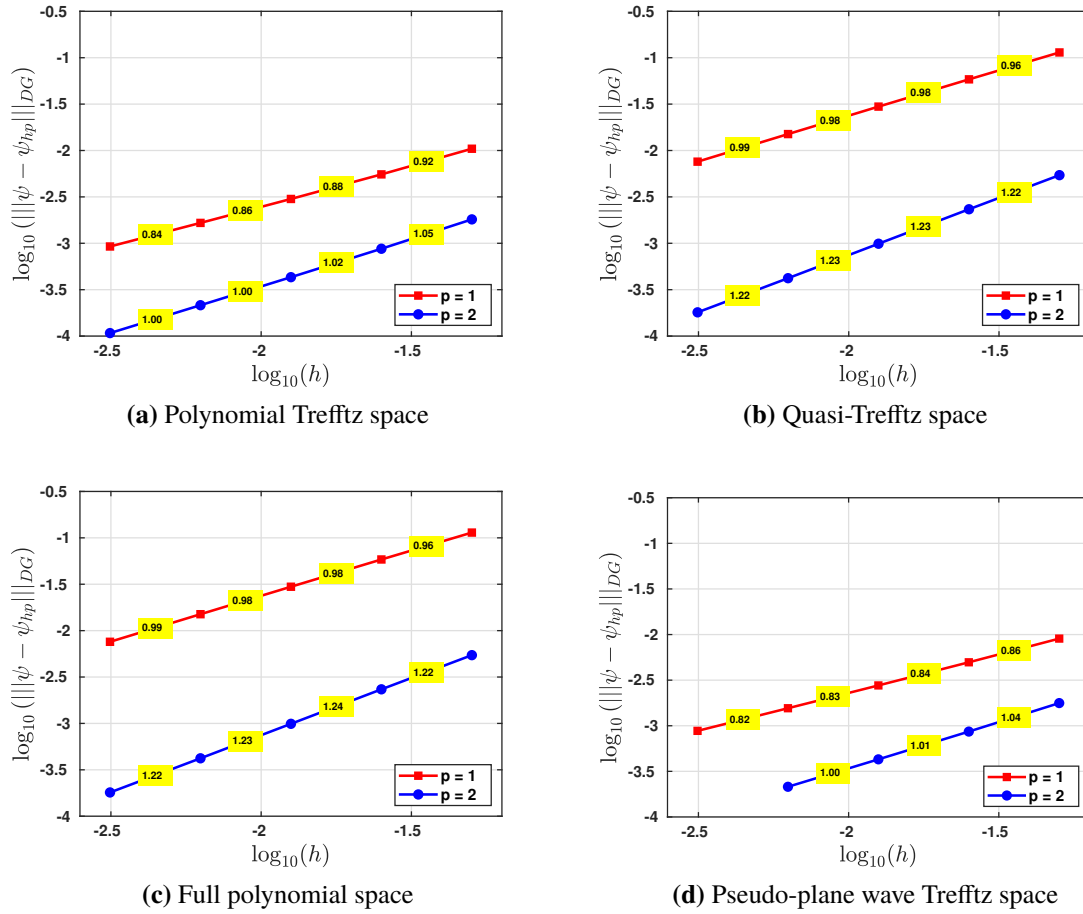
Figure 5.8 h-convergence for the $(1+1)$-dimensional problem with singular solution ψ in (5.11) and different discrete spaces.

The condition number $\kappa_2(\mathbf{K})$ for the polynomial Trefftz space for $\{m_J\}_{J=1}^{r_{1,2p}}$ in (5.12b), the quasi-Trefftz space, the full polynomial space, and the Trefftz space with orthogonal basis asymptotically grows as $O\left(h^{-1}\right)$ for all $p \in \mathbb{N}$, while for the polynomial Trefftz space for $\{m_J\}_{J=1}^{r_{1,2p}}$ in (5.12a) and the Trefftz space with a non-orthogonal basis, asymptotically grows as $O\left(h^{-(2p+1)}\right)$. Unfortunately, the basis functions $\{\phi_\ell\}$ obtained by choosing $k_\ell = 2\pi\ell/h_x$ oscillates faster when h_x decreases, which renders such a choice unsuitable from the approximation point of view.

5.6 Numerical experiments in $(2+1)$ dimensions

We now present some numerical test for space dimension $d = 2$. We recall that we use Cartesian space–time meshes with uniform partitions along each direction.

5.6.1 h-convergence

Singular time-independent potential $(V(x, y) = 1 - 1/x^2 - 1/y^2)$

(a) Polynomial Trefftz space $\mathbb{P}_{\mathbf{T}}^{2p}$ for $\{m_J\}_{J=1}^{r_{1,2p}}$ in (5.12a)

(b) Polynomial Trefftz space $\mathbb{P}_{\mathbf{T}}^{2p}$ for $\{m_J\}_{J=1}^{r_{1,2p}}$ in (5.12b)

(c) Trefftz space $\mathbb{T}^p$ in (4.1) (non-orthogonal basis)

(d) Trefftz space $\mathbb{T}^p$ in (4.1) (orthogonal basis)

(e) Full polynomial space $\mathbb{P}^p$ in (5.1)

(f) Quasi-Trefftz space $\mathbb{QT}^p$ in (5.2)
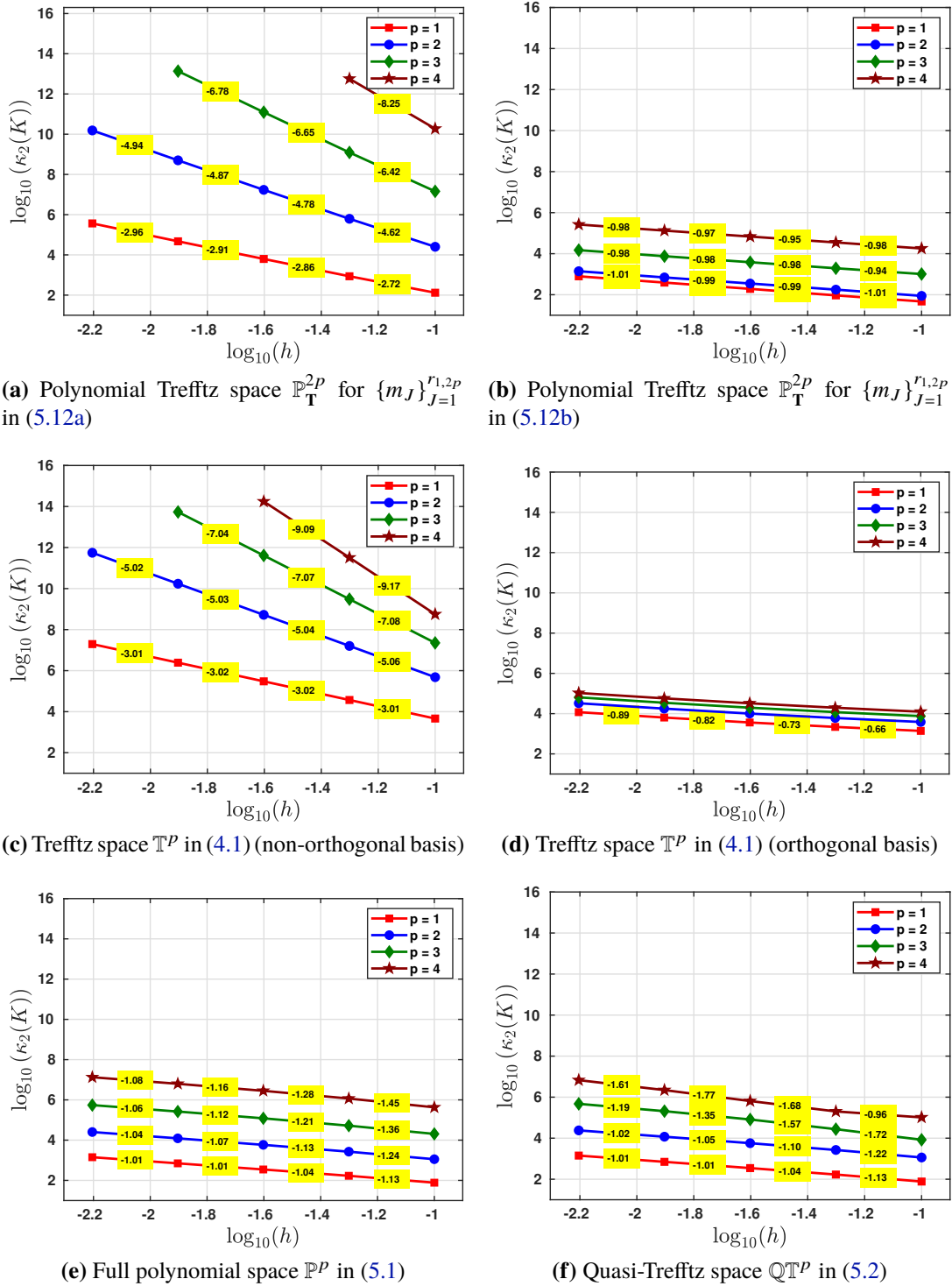
Figure 5.9 Conditioning of the stiffness matrix for the space–time ultra-weak DG method with different discrete spaces.

We consider the $(2 + 1)$-dimensional problem on $Q_T = (0, 1)^2 \times (0, 1)$ with exact solution (see [98])

$$\psi(x, y, t) = x^2 y^2 e^{it}. \tag{5.13}$$

63

In Figure 5.10, we show the errors obtained for a sequence of meshes with $h_x = h_y = h_t = 0.2/(i+1)$, $i = 1, \ldots, 4$, and different degrees of approximation p. As in the numerical results for the $(1+1)$-dimensional problems, we obtain rates of convergence of order $O\left(h^p\right)$ in the DG norm, and $O\left(h^{p+1}\right)$ in the $L^2(\Omega)$-norm at the final time.

(a) DG error (b) L^2 error at $T = 1$
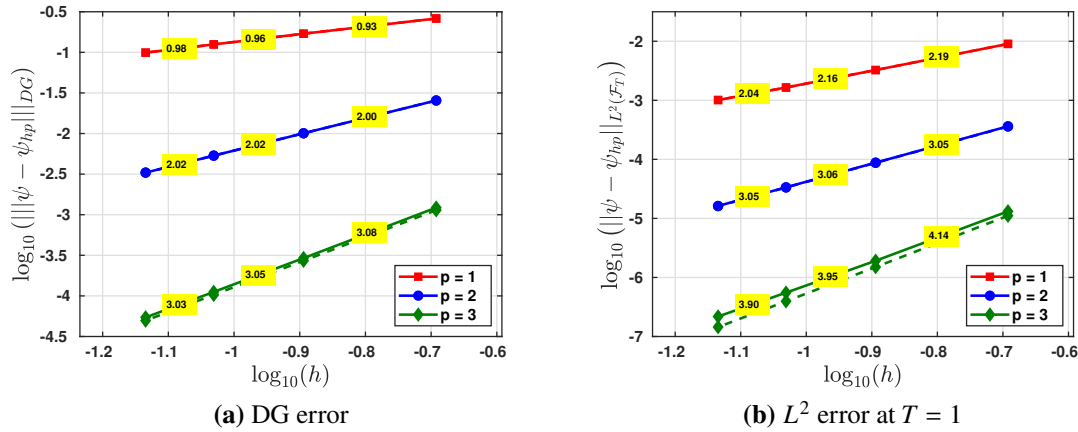
Figure 5.10 h-convergence for the $(2+1)$-dimensional problem with potential $V(x, y) = 1 - 1/x^2 - 1/y^2$ and exact solution (5.13).

Time-dependent potential $(V(x, y, t) = 2\tanh^2(\sqrt{2}x) - 4(t - 1/2)^3 + 2\tanh^2(\sqrt{2}y) - 2)$

We now consider a manufactured problem with a time-dependent potential (see [20]). On the space–time domain $Q_T = (0, 1)^2 \times (0, 1)$ the exact solution is

$$\psi(x, y, t) = ie^{i(t-1/2)^4}\mathrm{sech}(x)\mathrm{sech}(y). \tag{5.14}$$

In Figure 5.11 we show the errors obtained for the sequence of meshes from the previous experiment, and optimal convergence is observed in both norms.

5.6.2 p-convergence

In Figure 5.12 we show the results obtained for the p-version of the method applied to the $(2+1)$-dimensional problems above, on the coarsest mesh. As expected, for the $(2 + 1)$-dimensional case, the error of the quasi-Trefftz version decays root-exponentially as $O\left(e^{-b\sqrt{N_{DoFs}}}\right)$.

(a) DG error

(b) L^2 error at $T = 1$
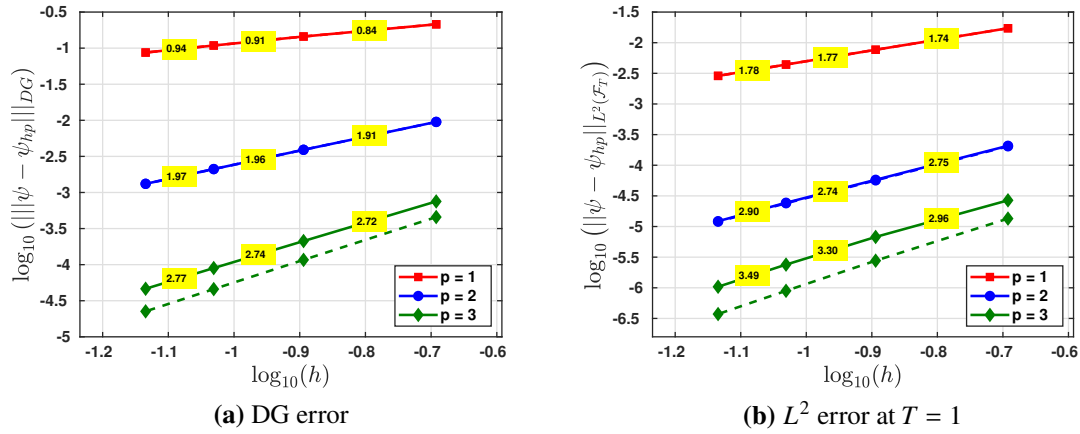
Figure 5.11 h-convergence for the $(2 + 1)$-dimensional problem with time dependent potential $V(x, y, t) = 2 \tanh^2 \left(\sqrt{2}x \right) - 4 \left(t - 1/2 \right)^3 + 2 \tanh^2 \left(\sqrt{2}y \right) - 2$ and exact solution (5.14).

(a) Time independent potential $V(x, y) = 1 - 1/x^2 - 1/y^2$

(b) Time dependent potential $V(x, y, t) = 2 \tanh^2 \left(\sqrt{2}x \right) - 4 \left(t - 1/2 \right)^3 + 2 \tanh^2 \left(\sqrt{2}y \right) - 2$
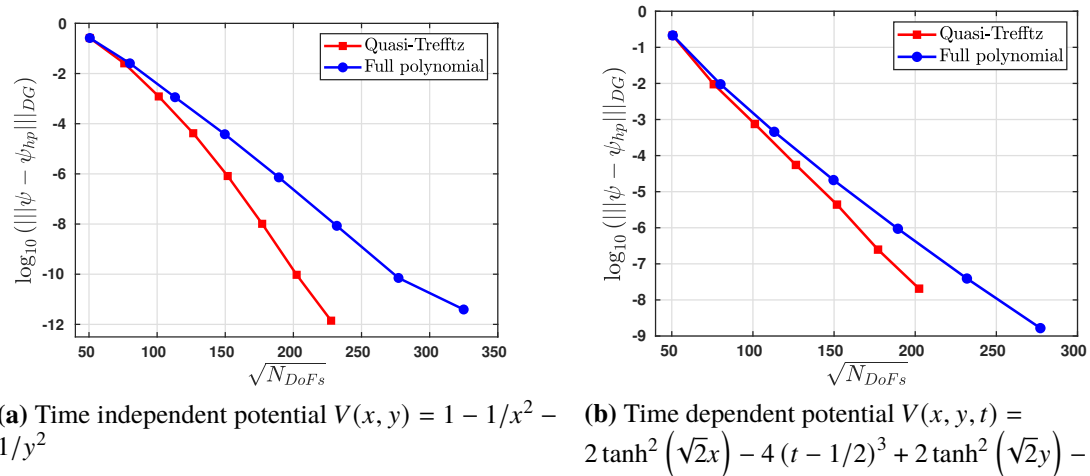
Figure 5.12 p-convergence for the $(2 + 1)$-dimensional problems.

Part II

Heat equation

Chapter 6

Space–time virtual element method for the heat equation

Contents

6.1 Introduction

In this chapter, we design the first space–time virtual element method (VEM) for the solution to a time-dependent PDE, namely, the heat equation; we consider space domains in one, two, and three dimensions. As for the discretization of the Schrödinger equation, we employ prismatic-type elements. This allows us to distinguish two types of mesh facets: *space-like* facets, i.e., facets lying on hyperplanes in space–time that are perpendicular to the time axis; *time-like* facets, i.e., facets whose normals are perpendicular to the time axis. The method we propose is based on the standard space–time variational formulation for the heat equation (1.5) in the space–time cylinder $Q_T = \Omega \times (0, T)$ with trial space $L^2(0, T; H_0^1(\Omega)) \cap H^1(0, T; H^{-1}(\Omega))$ and test space $L^2(0, T; H_0^1(\Omega))$; see [19, Ch. XVIII, §4.1].

Several space–time methods have been designed for the discretization of the heat equation, and they can be classified into two main groups. The first one is based on the discretization of the standard Petrov-Galerkin formulation in [19]; see [4, 93] for continuous finite element methods, [92] for a wavelet method, [99, 14] for discontinuous Galerkin methods, [66] for an isogeometric methods, [96] for a coercive method based on a Hilbert transformation of the test space, and [97] for a mixed finite element method. Residual-type error indicators for the method of [93] were considered in [94, 95]. Possible drawbacks of employing continuous finite elements are that suboptimal convergence rates are obtained for some singular solutions, and incompatible boundary and initial conditions cannot be naturally handled.

The second group is based on first order system least squares discretizations (FOSLS); see [35, 36, 100] for finite element methods and [81] for isogeometric methods. Space–time FOSLS finite elements naturally provide reliable and efficient error indicators; see [35, 90, 36]. However, they require the computation of an additional vector-valued flux variable.

We summarize the main features of the proposed VEM.

- Local virtual element spaces consist of functions that solve a heat equation with polynomial data on each space–time mesh element, which makes the method particularly suitable for further extensions, e.g., to its Trefftz variant.

- Even for prismatic elements with simplicial bases, the proposed local virtual element spaces do not coincide with their standard tensor-product finite element counterparts.

- Global virtual element spaces involve approximating continuity constraints across mesh facets. More precisely, we impose nonconformity conditions on time-like facets analogous to those in [3] for the Poisson problem, and allow for discontinuous functions in time. Across space-like facets, we transmit the information between consecutive time slabs by upwinding, i.e., preserving the causality relation. In the present VEM context, the upwind terms are defined by means of a polynomial projection.

- To keep the presentation and the analysis of the method as simple as possible, the analysis is presented for the particular case of space–time tensor-product meshes. An extension of the method to general prismatic meshes is presented in Chapter 8 below. The method can handle nonmatching time-like or space-like facets, which is greatly advantageous for space–time adaptivity.

We summarize the advantages of the proposed space–time VEM over standard space–time conforming finite element methods.

- The nonconforming VEM setting is of arbitrary order and its design is independent of the spatial dimension.

- Nonmatching space-like and time-like facets, which naturally stem from mesh adaptive procedures, can be handled easily.

- As the discrete test spaces are not continuous in time, we can solve the global problem as a sequence of local problems on time slabs.

Structure of the chapter: In Section 6.2, we introduce some notation and assumptions on the space–time meshes. The new space–time VEM is presented in Section 6.3. First, local VE spaces are introduced in Section 6.3.1 together with their degrees of freedom (DoFs). Based on the choice of such DoFs, in Section 6.3.2, we show that we can compute different polynomial projections of the VE functions. Such polynomial projections are instrumental in the design of the global VE spaces; see Section 6.3.3. Likewise, in Section 6.3.4, we design computable discrete bilinear forms for the VEM introduced in Section 6.3.5. Section 6.4 is devoted to the well-posedness of the method.

6.2 Space–time mesh assumptions

For the sake of presentation, we consider tensor-product-in-time meshes. We postpone to Chapter 8 the extension of the method to general prismatic meshes.

We consider a sequence of polytopic meshes $\{\mathcal{T}_h\}_h$ of Q_T. We require that

(G1) the space domain Ω is split into a mesh $\mathcal{T}_h^{\mathbf{x}}$ of nonoverlapping d-dimensional polytopes with straight facets; the time interval $(0, T)$ is split into N subintervals $I_n := (t_{n-1}, t_n)$ with knots $0 = t_0 < t_1 < \cdots < t_N = T$; each element K in $\mathcal{T}_h$ can be written as $K_{\mathbf{x}} \times I_n$, for some $K_{\mathbf{x}}$ in $\mathcal{T}_h^{\mathbf{x}}$ and $1 \le n \le N$.

Essentially, assumption (G1) states that (*i*) each element is the tensor-product of a d-dimensional polytope with a time interval; (*ii*) each element belongs to a time slab out of the N identified by the partition $\{t_n\}_{n=0}^N$; (*iii*) each time slab is partitioned by the same space mesh; (*iv*) all elements within the same time slab have the same extent in time.

Given an element $K = K_{\mathbf{x}} \times I_n \in \mathcal{T}_h$, we denote its diameter by h_K and the diameter of $K_{\mathbf{x}}$ by $h_{K_{\mathbf{x}}}$, and set $h_{I_n} := t_n - t_{n-1}$. We let $h := \max_{K \in \mathcal{T}_h} h_K$ and $h_{\mathbf{x}} := \max_{K \in \mathcal{T}_h^{\mathbf{x}}} h_{K_{\mathbf{x}}}$. Furthermore, the set of all $(d-1)$-dimensional facets of $K_{\mathbf{x}}$ is denoted by $\mathcal{F}^{K_{\mathbf{x}}}$, and for any $F_{\mathbf{x}} \in \mathcal{F}^{K_{\mathbf{x}}}$ we define

$$h_{F_{\mathbf{x}}} := \begin{cases} \min\{h_{K_{\mathbf{x}}}, h_{\widetilde{K}_{\mathbf{x}}}\} & \text{if } F_{\mathbf{x}} = K_{\mathbf{x}} \cap \widetilde{K}_{\mathbf{x}} \text{ for some } \widetilde{K}_{\mathbf{x}} \in \mathcal{T}_h^{\mathbf{x}}, \\ h_{K_{\mathbf{x}}} & \text{if } F_{\mathbf{x}} \subset \partial\Omega. \end{cases}$$

For $d = 1$, $F_{\mathbf{x}}$ is a point and $\int_{F_{\mathbf{x}}} v(x, t) \mathrm{d}S$ is equal to $v(F_{\mathbf{x}}, t)$. For each facet $F_{\mathbf{x}}$ in $\mathcal{F}^{K_{\mathbf{x}}}$, we introduce the time-like facet $F := F_{\mathbf{x}} \times I_n$; we collect all these time-like facets into the set $\mathcal{F}^K$.

We fix one of the two unit normal d-dimensional vectors associated with $F_{\mathbf{x}}$ and denote it by $\vec{\mathbf{n}}_{F_{\mathbf{x}}}$. For $d \ge 1$, each time-like facet $F = F_{\mathbf{x}} \times I_n$ lies in a d-dimensional hyperplane with unit normal vector $\vec{\mathbf{n}}_F := (\vec{\mathbf{n}}_{F_{\mathbf{x}}}, 0)$.

Next, we require further assumptions on the spatial mesh $\mathcal{T}_h^{\mathbf{x}}$:

(G2) **Uniform star-shapedness:** There exists $0 < \rho \le \frac{1}{2}$ such that, each space element $K_{\mathbf{x}} \in \mathcal{T}_h^{\mathbf{x}}$ is star-shaped with respect to a ball of radius $\rho h_{K_{\mathbf{x}}}$.

(G3) **Bounded number of facets:** The number of $(d-1)$-dimensional facets of $K_{\mathbf{x}}$ is uniformly bounded with respect to the meshsize.

(G4) **Local quasi-uniformity in space:** there exists a number $\mathsf{lqu}(\mathcal{T}_h) > 0$ such that, given two neighbouring elements $K_{\mathbf{x}}$ and $\widetilde{K}_{\mathbf{x}}$ of $\mathcal{T}_h^{\mathbf{x}}$, we have that $\mathsf{lqu}(\mathcal{T}_h^{\mathbf{x}})^{-1} h_{\widetilde{K}_{\mathbf{x}}} \le h_{K_{\mathbf{x}}} \le \mathsf{lqu}(\mathcal{T}_h^{\mathbf{x}}) h_{\widetilde{K}_{\mathbf{x}}}$.

For a given space–time element $K \subset \mathbb{R}^{d+1}$ and any space-like or time-like facet $F \subset \partial K$, we denote the space of polynomials of total degree at most $p \in \mathbb{N}$ on K and F by $\mathbb{P}^p(K)$ and $\mathbb{P}^p(F)$, respectively. For a given time interval I, $\mathbb{P}^p(I)$ denotes the space of polynomials in I of total degree at most p in $\mathbb{N}$.

For a positive natural number k, we define the spaces of broken H^k functions over $\mathcal{T}_h^{\mathbf{x}}$ and $\mathcal{T}_h$, respectively, by

$$H^k(\mathcal{T}_h^{\mathbf{x}}) := \left\{ v \in L^2(Q_T) \mid v_{|K_{\mathbf{x}}} \in H^k(K_{\mathbf{x}}) \ \forall K_{\mathbf{x}} \in \mathcal{T}_h^{\mathbf{x}} \right\},$$
$$H^k(\mathcal{T}_h) := \left\{ v \in L^2(Q_T) \mid v_{|K} \in H^k(K) \ \forall K \in \mathcal{T}_h \right\}.$$

We denote the broken Sobolev k seminorm on $\mathcal{T}_h$ by $|\cdot|_{H^k(\mathcal{T}_h)}$ and the space of piecewise polynomials of degree at most ℓ in $\mathbb{N}$ on $\mathcal{T}_h$ by $\mathcal{S}^\ell(\mathcal{T}_h)$.

6.3 The virtual element method

In this section, we introduce a VEM for the discretization of the heat equation (1.2) with homogeneous initial and Dirichlet boundary conditions. The proposed method is based on the standard Petrov-Galerkin weak formulation (1.5).

6.3.1 Local virtual element spaces

Given an approximation degree $p \in \mathbb{N}$ and an element $K \in \mathcal{T}_h$, we define the following local VE space:

$$V_h(K) := \left\{ v \in L^2(K) \mid \widetilde{c}_H^K \partial_t v_h - \widetilde{\lambda}^K \Delta_{\mathbf{x}} v_h \in \mathbb{P}^{p-1}(K), \quad v_{h|K_{\mathbf{x}} \times \{t_{n-1}\}} \in \mathbb{P}^p(K_{\mathbf{x}}); \right.$$
$$\left. \vec{\mathbf{n}}_{F_{\mathbf{x}}} \cdot \nabla_{\mathbf{x}} v_{h|F} \in \mathbb{P}^p(F) \ \forall F := F_{\mathbf{x}} \times I_n \text{ with } F_{\mathbf{x}} \in \mathcal{F}^{K_{\mathbf{x}}} \right\}, \tag{6.1}$$

where $\widetilde{c}_H^K := h_{I_n}$ and $\widetilde{\lambda}^K := h_{K_{\mathbf{x}}}^2$.

The space $V_h(K)$ contains all polynomials of degree at most p in K, i.e., $\mathbb{P}^p(K) \subset V_h(K)$. The degree p in the Neumann boundary conditions is not necessary for this inclusion to be valid, as $p - 1$ would be sufficient. Nevertheless, the degree p is crucial in the proof of the Poincaré-type inequality in Proposition 6.9 below.

Remark 6.1 (Regularity of local VEM functions)**.** Functions in $V_h(K)$ solve a heat equation problem with polynomial source, initial condition, and Neumann boundary conditions. For this reason, $V_h(K) \subset L^2(I_n; H^1(K_{\mathbf{x}}))$; see [74, Thm. 4.1 and §4.7.2 in Ch. 3] with standard modifications to deal with the nonhomogeneous Neumann data. ∎

Remark 6.2 (Scaling argument)**.** As opposite to the standard VE setting [9], in definition (6.1) we consider solutions to local problems involving some scaling factors ($\widetilde{c}_H^K$ and $\widetilde{\lambda}^K$). The reason for this is that these local problems involve differential operators of different orders. By using a scaling argument and mapping the element K into a "reference" element $\widehat{K} = \widehat{I} \times \widehat{K}_{\widehat{\mathbf{x}}}$, with $|\widehat{I}_n| = \text{diam}(\widehat{K}_{\widehat{\mathbf{x}}}) = 1$, the resulting reference space consists of solutions to a heat equation with both coefficients equal to 1. This allows us to use equivalence of norms results when proving the stability of the scheme. ∎

Let $\{m_\alpha^K\}_{\alpha=1}^{r_{d+1,p-1}}$, $\{m_\beta^F\}_{\beta=1}^{r_{d,p}}$, and $\{m_\gamma^{K_{\mathbf{x}}}\}_{\gamma=1}^{r_{d,p}}$ be any bases of $\mathbb{P}^{p-1}(K)$, $\mathbb{P}^p(F)$, and $\mathbb{P}^p(K_{\mathbf{x}})$, where we recall that $r_{d,p} = \dim(\mathbb{P}^p(\mathbb{R}^d))$. Introduce the following set of linear functionals on $V_h(K)$:

- the *bulk* moments

$$\frac{1}{|K|} \int_{I_n} \int_{K_{\mathbf{x}}} v_h \, m_\alpha^K \mathrm{d}\mathbf{x} \, \mathrm{dt} \qquad \forall \alpha = 1, \ldots, r_{d+1,p-1}; \tag{6.2a}$$

- for all space–time facets $F = F_{\mathbf{x}} \times I_n \in \mathcal{F}^K$, the *time-like* moments

$$\frac{1}{|F|} \int_{I_n} \int_{F_{\mathbf{x}}} v_h \, m_\beta^F \mathrm{d}S \, \mathrm{d}t \qquad \forall \beta = 1, \ldots, r_{d,p}; \tag{6.2b}$$

- the *space-like* moments

$$\frac{1}{|K_{\mathbf{x}}|} \int_{K_{\mathbf{x}}} v_h(\cdot, t_{n-1}) m_\gamma^{K_{\mathbf{x}}} \mathrm{d}\mathbf{x} \qquad \forall \gamma = 1, \ldots, r_{d,p}. \tag{6.2c}$$

Since functions $v_h \in V_h(K)$ are polynomials at time t_{n-1}, then the integrals in (6.2c) are well defined. Moreover, the inclusion $V_h(K) \subset L^2(I_n; H^1(K_{\mathbf{x}}))$, see Remark 6.1, implies that the integrals in (6.2a) and (6.2b) are also well defined.

We introduce the number of functionals in (6.2a)–(6.2c) as

$$\# \mathrm{DoFs} := \dim(\mathbb{P}^{p-1}(K)) + \sum_{F \in \mathcal{F}^K} \dim(\mathbb{P}^{p-1}(F)) + \dim(\mathbb{P}^p(K_{\mathbf{x}})).$$

In the following lemma, we prove that the linear functionals (6.2a)–(6.2c) actually define a set of DoFs for $V_h(K)$.

Lemma 6.3 (Unisolvence of DoFs). *The linear functionals (6.2a)–(6.2c) are a set of unisolvent DoFs for the space $V_h(K)$.*

Proof. Since the right-hand side and the initial and Neumann boundary conditions in (6.1) are independent of each other, the dimension of $V_h(K)$ is equal to the number of the linear functionals (6.2a)–(6.2c). Thus, it suffices to prove that the set of these linear functionals is unisolvent. In other words, we prove that, whenever $v_h \in V_h(K)$ satisfies $\mathrm{DoF}_i(v_h) = 0$ for all $i = 1, \ldots, \#\mathrm{DoFs}$, then $v_h = 0$.

Thanks to the definition of the DoFs (6.2a) and (6.2b), we have

$$0 = \int_{I_n} \int_{K_{\mathbf{x}}} v_h \underbrace{\left(\widetilde{c}_H^K \partial_t v_h - \widetilde{\lambda}^K \Delta_{\mathbf{x}} v_h \right)}_{\in \mathbb{P}^{p-1}(K)} \mathrm{d}\mathbf{x} \, \mathrm{d}t + \widetilde{\lambda}^K \sum_{F_{\mathbf{x}} \in \mathcal{F}^{K_{\mathbf{x}}}} \int_{I_n} \int_{F_{\mathbf{x}}} v_h \underbrace{\vec{\mathbf{n}}_{F_{\mathbf{x}}} \cdot \nabla_{\mathbf{x}} v_h}_{\in \mathbb{P}^p(F)} \mathrm{d}S \, \mathrm{d}t$$

$$= \frac{\widetilde{c}_H^K}{2} \left(\|v_h(\cdot, t_n)\|_{L^2(K_{\mathbf{x}})}^2 - \|v_h(\cdot, t_{n-1})\|_{L^2(K_{\mathbf{x}})}^2 \right) + \widetilde{\lambda}^K \|\nabla_{\mathbf{x}} v_h\|_{L^2(K)^d}^2 .$$

Furthermore, using the definition of the DoFs (6.2c), we have $\|v_h(\cdot, t_{n-1})\|_{L^2(K_{\mathbf{x}})}^2 = 0$ and deduce

$$\widetilde{\lambda}^K \|\nabla_{\mathbf{x}} v_h\|_{L^2(K)^d}^2 = 0 \quad \Rightarrow \quad \nabla_{\mathbf{x}} v_h = 0 \text{ in } K \quad \Rightarrow \quad v_h = v_h(t).$$

From the definition of the space $V_h(K)$, this implies that $\partial_t v_h$ belongs to $\mathbb{P}^{p-1}(I_n)$; equivalently, v_h belongs to $\mathbb{P}^p(I_n)$. On the other hand, we know that the moments (6.2b) are zero, in particular when they are taken with respect to monomials up to degree p in time only. This implies $v_h = 0$. $\qquad \square$

6.3.2 Polynomial projections

Functions in the local VE space $V_h(K)$ are not known in closed form. However, if we have at our disposal the DoFs of a function v_h in $V_h(K)$, then we can compute projections onto polynomial spaces with given maximum degree.

First, for all $K = K_{\mathbf{x}} \times I_n \in \mathcal{T}_h$ and $\varepsilon > 0$, we define the operator $\Pi_p^{\mathcal{N}} : H^{\frac{1}{2}+\varepsilon}(I_n; L^2(K_{\mathbf{x}})) \cap L^2(I_n; H^1(K_{\mathbf{x}})) \to \mathbb{P}^p(K)$ as follows: for any v in $H^{\frac{1}{2}+\varepsilon}(I_n; L^2(K_{\mathbf{x}})) \cap L^2(I_n; H^1(K_{\mathbf{x}}))$,

$$\int_{I_n} \int_{K_{\mathbf{x}}} \nabla_{\mathbf{x}} q_p^K \cdot \nabla_{\mathbf{x}} \left(\Pi_p^{\mathcal{N}} v - v \right) \, \mathrm{d}\mathbf{x} \, \mathrm{d}t = 0 \quad \forall q_p^K \in \mathbb{P}^p(K); \tag{6.3a}$$

$$\int_{I_n} \int_{K_{\mathbf{x}}} q_{p-1}(t) \left(\Pi_p^{\mathcal{N}} v - v \right) \, \mathrm{d}\mathbf{x} \, \mathrm{d}t = 0 \quad \forall q_{p-1} \in \mathbb{P}^{p-1}(I_n); \tag{6.3b}$$

$$\int_{K_{\mathbf{x}}} \left(\Pi_p^{\mathcal{N}} v(\mathbf{x}, t_{n-1}) - v(\mathbf{x}, t_{n-1}) \right) \, \mathrm{d}\mathbf{x} = 0. \tag{6.3c}$$

We have $V_h(K) \subset L^2(I_n; H^1(K_{\mathbf{x}}))$; see Remark 6.1. This and the fact that functions in $V_h(K)$ restricted to the time t_{n-1} are polynomials entail that we can define $\Pi_p^{\mathcal{N}} v$ also for v in $V_h(K)$.

Lemma 6.4. *The operator $\Pi_p^{\mathcal{N}}$ is well defined. Moreover, for any v_h in $V_h(K)$, $\Pi_p^{\mathcal{N}} v_h$ is computable via the* DoFs (6.2a)–(6.2c).

Proof. In order to prove that $\Pi_p^{\mathcal{N}}$ is well defined, we need to show that the number of (linear) conditions in (6.3a)–(6.3c) is equal to $\dim(\mathbb{P}^p(K))$. As (6.3a) is void for all $q_p^K \in \mathbb{P}^p(I_n)$, we have that the number of conditions in (6.3a)–(6.3c) is equal to $\dim(\mathbb{P}^p(K))$. We only need to show that they are linearly independent.

To this aim, assume that $v = 0$. Conditions (6.3a) imply that $\nabla_{\mathbf{x}} \Pi_p^{\mathcal{N}} v = 0$, i.e., $\Pi_p^{\mathcal{N}} v$ belongs to $\mathbb{P}^p(I_n)$. Let $L_p(\cdot)$ be the Legendre polynomial of degree p over $[-1, 1]$. Using conditions (6.3b), we deduce that there exists a constant c such that

$$\Pi_p^{\mathcal{N}} v = c L_p \left(\frac{2t - t_{n-1} - t_n}{t_n - t_{n-1}} \right).$$

Since condition (6.3c) entails $\Pi_p^{\mathcal{N}} v(\cdot, t_{n-1}) = 0$ and $L_p(-1) \neq 0$, we deduce that $c = 0$, whence $\Pi_p^{\mathcal{N}} v = 0$. Therefore, the conditions are linearly independent and so $\Pi_p^{\mathcal{N}}$ is well defined.

As for the computability of $\Pi_p^{\mathcal{N}} v_h$ for v_h in $V_h(K)$, conditions (6.3a) and (6.3b) are available via the bulk moments (6.2a) (up to order $p - 2$) and the time-like moments (6.2b) (up to order $p - 1$); condition (6.3c) is available via the lowest-order space-like moment in (6.2c). $\quad\square$

Next, for all $K \in \mathcal{T}_h$ and $\varepsilon > 0$, we define the operator $\Pi_p^{\star} : C^0(I_n; L^2(K_{\mathbf{x}})) \to \mathbb{P}^p(K)$ as follows: for any v in $H^{\frac{1}{2}+\varepsilon}(I_n; L^2(K_{\mathbf{x}}))$,

$$\int_{I_n} \int_{K_{\mathbf{x}}} q_{p-1}^K \left(\Pi_p^{\star} v - v \right) \, \mathrm{d}\mathbf{x} \, \mathrm{d}t = 0 \quad \forall q_{p-1}^K \in \mathbb{P}^{p-1}(K); \tag{6.4a}$$

$$\int_{K_{\mathbf{x}}} q_p^{K_{\mathbf{x}}} \left(\Pi_p^{\star} v(\mathbf{x}, t_{n-1}) - v(\mathbf{x}, t_{n-1}) \right) \, \mathrm{d}\mathbf{x} = 0 \quad \forall q_p^{K_{\mathbf{x}}} \in \mathbb{P}^p(K_{\mathbf{x}}). \tag{6.4b}$$

Again, we have $V_h(K) \subset L^2(I_n; H^1(K_{\mathbf{x}}))$; see Remark 6.1. This and the fact that functions in $V_h(K)$ restricted to the time t_{n-1} are polynomials entail that we can define $\Pi_p^{\star} v$ also for v in $V_h(K)$.

Lemma 6.5. *The operator $\Pi_p^\star$ is well defined. Moreover, for any v_h in $V_h(K)$, $\Pi_p^\star v_h$ is computable via the* DoFs *(6.2a)–(6.2c).*

Proof. As in the proof of Lemma 6.4, we observe that the number of (linear) conditions in (6.4a)–(6.4b) is equal to $\dim(\mathbb{P}^p(K))$. Thus, it suffices to show that they are linearly independent.

Assume that $v = 0$. Then, taking $q_p^{K_\mathbf{x}} = \Pi_p^\star v(\mathbf{x}, t_{n-1})$ in (6.4b), we get $\Pi_p^\star v(\mathbf{x}, t_{n-1}) = 0$. On the other hand, taking $q_{p-1}^K = \partial_t \Pi_p^\star v$ in (6.4a), we get

$$0 = \frac{1}{2} \left(\left\| \Pi_p^\star v(\cdot, t_n) \right\|_{L^2(K_\mathbf{x})}^2 - \left\| \Pi_p^\star v(\cdot, t_{n-1}) \right\|_{L^2(K_\mathbf{x})}^2 \right) = \frac{1}{2} \left\| \Pi_p^\star v(\cdot, t_n) \right\|_{L^2(K_\mathbf{x})}^2.$$

In addition, we observe that

$$\left\| \partial_t \Pi_p^\star v \right\|_{L^2(K)}^2 = \int_{K_\mathbf{x}} \Pi_p^\star v(\mathbf{x}, t) \, \partial_t \Pi_p^\star v(\mathbf{x}, t) \mathrm{d}\mathbf{x} \, \Big|_{t=t_{n-1}}^{t_n} - \int_{I_n} \int_{K_\mathbf{x}} \Pi_p^\star v \underbrace{\partial_{tt} \Pi_p^\star v}_{\in \mathbb{P}^{p-2}(K)} \, \mathrm{d}\mathbf{x} \, \mathrm{d}t = 0.$$

This implies that $\partial_t \Pi_p^\star v = 0$, which, together with $\Pi_p^\star v(\cdot, t_n) = 0$, gives that $\Pi_p^\star v = 0$. Therefore, the conditions are linearly independent and so $\Pi_p^\star$ is well defined.

As for the computability of $\Pi_p^\star v_h$ for $v_h \in V_h(K)$, conditions (6.4a) are available via the bulk DoFs (6.2a), and conditions (6.4b) are at disposal via the space-like DoFs at time t_{n-1} (6.2c). $\quad\square$

We introduce other polynomial projectors: for all $K \in \mathcal{T}_h$ and v in $L^2(K)$,

$$\Pi_{p-1}^{0,K} : L^2(K) \to \mathbb{P}^{p-1}(K), \qquad (q_{p-1}^K, v - \Pi_{p-1}^{0,K} v)_{L^2(K)} = 0 \quad \forall q_{p-1}^K \in \mathbb{P}^{p-1}(K);$$

for each temporal interval I_n and $v \in L^2(I_n)$,

$$\Pi_{p-1}^{0,I_n} : L^2(I_n) \to \mathbb{P}^{p-1}(I_n), \qquad (q_{p-1}^{I_n}, v - \Pi_{p-1}^{0,I_n} v)_{L^2(I_n)} = 0 \quad \forall q_{p-1}^{I_n} \in \mathbb{P}^{p-1}(I_n);$$

for each element $K_\mathbf{x} \in \mathcal{T}_h^\mathbf{x}$ and $v \in L^2(K_\mathbf{x})$,

$$\Pi_0^{0,K_\mathbf{x}} : L^2(K_\mathbf{x}) \to \mathbb{R}, \qquad (q_0, v - \Pi_0^{0,K_\mathbf{x}} v)_{L^2(K_\mathbf{x})} = 0 \quad \forall q_0 \in \mathbb{R};$$

for all time-like facet F and v in $L^2(F)$,

$$\Pi_p^{0,F} : L^2(F) \to \mathbb{P}^p(F), \qquad (q_p^F, v - \Pi_p^{0,F} v)_{L^2(F)} = 0 \quad \forall q_p^F \in \mathbb{P}^p(F).$$

Given v_h in $V_h(K)$, the computability of the above projectors applied to v_h follows from the definition of the DoFs (6.2a)–(6.2c). The projector $\Pi_{p-1}^{0,K}$ induces the global piecewise L^2 projector Π_{p-1}^{0,Q_T} over $\mathcal{T}_h$.

The following polynomial inverse inequalities are valid.

Lemma 6.6. *For any $p \in \mathbb{N}$, there exist positive constants $c_{\Pi_p^N}$ and $c_{\Pi_p^\star}$ independent of h_{I_n} and $h_{K_\mathbf{x}}$ such that, for all q_p in $\mathbb{P}^p(K)$,*

$$\left\| q_p \right\|_{L^2(K)}^2 + h_{K_\mathbf{x}}^2 \left\| \nabla_\mathbf{x} q_p \right\|_{L^2(K)^d}^2 + h_{I_n}^2 \left\| \partial_t q_p \right\|_{L^2(K)}^2$$

$$\leq c_{\Pi_p^N} \left(h_{K_\mathbf{x}}^2 \left\| \nabla_\mathbf{x} q_p \right\|_{L^2(K)^d}^2 + \left\| \Pi_{p-1}^{0,I_n} q_p \right\|_{L^2(K)}^2 + h_{I_n} \left\| \Pi_0^{0,K_\mathbf{x}} q_p(\cdot, t_{n-1}) \right\|_{L^2(K_\mathbf{x})}^2 \right) \tag{6.5}$$

and

$$\|q_p\|^2_{L^2(K)} + h^2_{K_\mathbf{x}} \|\nabla_\mathbf{x} q_p\|^2_{L^2(K)^d} + h^2_{I_n} \|\partial_t q_p\|^2_{L^2(K)}$$

$$\leq c_{\Pi^\star_p} \left(\left\|\Pi^{0,K}_{p-1} q_p\right\|^2_{L^2(K)} + h_{I_n} \|q_p(\cdot, t_{n-1})\|^2_{L^2(K_\mathbf{x})} \right). \tag{6.6}$$

Proof. The assertion follows from the regularity of the spatial mesh in assumption (G2), the fact that the functionals on the right-hand side of (6.5) and (6.6) are norms for $\mathbb{P}^p(K)$, and the equivalence of norms for spaces of polynomials with fixed maximum degree. □

The presence of the subscripts appearing in the inverse estimate constants $c_{\Pi^\mathcal{N}_p}$ and $c_{\Pi^\star_p}$ is to remind that the norms on the right-hand side of (6.5) and (6.6) are induced by the definition of the operators $\Pi^\mathcal{N}_p$ and $\Pi^\star_p$.

6.3.3 Global virtual element spaces

We construct the global VE spaces in a nonconforming fashion. To this aim, we introduce a jump operator on the time-like facets. Each internal time-like facet F is shared by two elements K_1 and K_2 with outward pointing unit normal vectors $\vec{\mathbf{n}}_{K_1}$ and $\vec{\mathbf{n}}_{K_2}$, whereas each boundary time-like facet belongs to the boundary of a single element K_3 with outward pointing unit normal vector $\vec{\mathbf{n}}_{K_3}$. We denote the d-dimensional vector containing the spatial components of the restriction of $\vec{\mathbf{n}}_{K_j}$ to the time-like facet F by $\vec{\mathbf{n}}^F_{K_j}$. For any piecewise scalar function v, the normal jump on each time-like facet F is defined as

$$[\![v]\!]_F := \begin{cases} v_{|K_1} \vec{\mathbf{n}}^F_{K_1} + v_{|K_2} \vec{\mathbf{n}}^F_{K_2} & \text{if } F \text{ is an internal face;} \\ v_{|K_3} \vec{\mathbf{n}}^F_{K_3} & \text{if } F \text{ is a boundary face.} \end{cases} \tag{6.7}$$

On each time slab I_n, we introduce the nonconforming Sobolev space of order p associated with the mesh $\mathcal{T}^\mathbf{x}_h$:

$$\mathcal{NC}_p(\mathcal{T}^\mathbf{x}_h; I_n) := \left\{ v \in L_2\left(I_n; H^1(\mathcal{T}^\mathbf{x}_h)\right) \,\Big|\, \int_{I_n} \int_{F_\mathbf{x}} q^F_p [\![v]\!]_F \cdot \vec{\mathbf{n}}_{F_\mathbf{x}} \mathrm{d}S\, \mathrm{d}t = 0 \quad \forall q^F_p \in \mathbb{P}^p(F) \right\}. \tag{6.8}$$

This allows us to define the VE discretization Y_h of the space Y in (1.3) as the space of functions that are possibly discontinuous in time across space-like facets and nonconforming in space:

$$Y_h := \left\{ v_h \in L^2(Q_T) \mid v_{h|K} \in V_h(K) \; \forall K \in \mathcal{T}_h, \; v_{h|\mathcal{T}^\mathbf{x}_h \times I_n} \in \mathcal{NC}_p(\mathcal{T}^\mathbf{x}_h; I_n) \; \forall n = 1, \dots, N \right\}.$$

The functions in the space X in (1.3) are continuous in time, namely,

$$X \hookrightarrow C^0([0, T]; L^2(\Omega)); \tag{6.9}$$

see e.g, [101, Thm. 25.5]. Nevertheless, we discretize X with Y_h as well, and impose the time continuity weakly through upwinding. As the functions in the local VE space $V_h(K)$ are not known at the local final time t_n, the upwind fluxes are defined in terms of the traces of their polynomial projections $\Pi^\star_p$; see (6.21) below.

Remark 6.7 (Discretization of the trial space X). Due to the choice of the DoFs, one cannot define a continuous-in-time discretization of X with the local spaces $V_h(K)$. If this were possible, then each VE function on $K = K_{\mathbf{x}} \times I_n$ would be a polynomial of degree p at the local final time t_n. For general choices of the right-hand side, initial condition, and boundary conditions in (6.1), this is not true. ∎

Remark 6.8. A possible VE discretization of the trial function space X could be

$$X_h := \{v_h \in Y_h \mid v_h \text{ satisfies } (6.10)\},$$

where

$$v_h(\cdot, 0) = 0 \text{ in } \Omega; \qquad v_h^{(n)}(\cdot, t_{n-1}) = \Pi_p^\star v_h^{(n-1)}(\cdot, t_{n-1}) \text{ in } K_{\mathbf{x}} \qquad \forall 1 < n < N. \qquad (6.10)$$

In (6.10), $v_h^{(n)}$ and $v_h^{(n-1)}$ denote the restrictions of v_h to the time slabs $\Omega \times I_n$ and $\Omega \times I_{n-1}$. In this way, the continuity in time across space-like facets is approximated directly within the discretization space through the polynomial projection $\Pi_p^\star$ in (6.4). Nevertheless, this leads to an error accumulation in time, which results in a suboptimal convergence. This can be verified numerically and justified theoretically. ∎

6.3.4 Discrete bilinear forms

On each element K, define the local continuous bilinear form in $V_h(K) \times V_h(K)$ and seminorm

$$a^K(u_h, v_h) := \lambda(\nabla_{\mathbf{x}} u_h, \nabla_{\mathbf{x}} v_h)_{L^2(K)}, \qquad\qquad |v_h|^2_{Y(K)} := a^K(v_h, v_h).$$

Proposition 6.9 (Local Poincaré inequality). *If v_h belongs to $V_h(K)$, $K = K_{\mathbf{x}} \times I_n$, then $|v_h|_{Y(K)} = 0$ if and only if $v_h = v_h(t)$ belongs to $\mathbb{P}^p(I_n)$. Moreover, there exists a positive constant C_P^K independent of h_{I_n} and $h_{K_{\mathbf{x}}}$ such that*

$$\inf_{q_p^t \in \mathbb{P}^p(I_n)} \left\| v_h - q_p^t \right\|_{L^2(K)} \leq C_P^K h_{K_{\mathbf{x}}} \|\nabla_{\mathbf{x}} v_h\|_{L^2(K)^d} \qquad \forall v_h \in V_h(K). \qquad (6.11)$$

Proof. If v_h belongs to $V_h(K)$ with $\|\nabla_{\mathbf{x}} v_h\|_{Y(K)} = 0$, then $v_h = v_h(t)$. The definition of $V_h(K)$ in (6.1) implies that $\partial_t v_h$ belongs to $\mathbb{P}^{p-1}(I_n)$ or, equivalently, that v_h belongs to $\mathbb{P}^p(I_n)$. The converse is obviously true.

Inequality (6.11) follows from the equivalence of seminorms with the same kernel on finite dimensional spaces and the scaling argument in Remark 6.2. □

We define $Y(\mathcal{T}_h) := L^2(0, T; H^1(\mathcal{T}_h^{\mathbf{x}}))$ and introduce the global broken seminorms

$$\text{for almost all } t, \quad |v(\cdot, t)|^2_{H^1(\mathcal{T}_h^{\mathbf{x}})} := \sum_{K_{\mathbf{x}} \in \mathcal{T}_h^{\mathbf{x}}} \|\nabla_{\mathbf{x}} v(\cdot, t)\|^2_{L^2(K_{\mathbf{x}})^d};$$

$$|v|^2_{Y(\mathcal{T}_h)} := \int_0^T \lambda |v(\cdot, t)|^2_{H^1(\mathcal{T}_h^{\mathbf{x}})} \mathrm{d}t = \sum_{K \in \mathcal{T}_h} |v|^2_{Y(K)}.$$

Proposition 6.10. *The seminorm $|\cdot|_{Y(\mathcal{T}_h)}$ is a norm in Y_h.* [1]

---

[1]In fact, $|\cdot|_{Y(\mathcal{T}_h)}$ is a norm on $Y + Y_h$. So, for arguments in $Y + Y_h$, we shall denote it by $\|\cdot\|_{Y(\mathcal{T}_h)}$.

Proof. Given v_h in Y_h, we only have to prove that $|v_h|_{Y(\mathcal{T}_h)} = 0$ implies $v_h = 0$. The identity $|v_h|_{Y(\mathcal{T}_h)} = 0$ implies that $|v_h|_{Y(K)} = 0$ for all elements $K = K_\mathbf{x} \times I_n$. Using Proposition 6.9, we deduce that $v_{h|K}$ only depends on time and belongs to $\mathbb{P}^p(I_n)$. The assertion follows using the spatial nonconformity of the space Y_h, see (6.8), which is up to order p. $\square$

On each element K in $\mathcal{T}_h$, $K = K_\mathbf{x} \times I_n$, let

$$S^K : [V_h(K) + L^2(I_n; H^1(K_\mathbf{x})) \cap C^0(I_n; L^2(K_\mathbf{x}))]^2 \to \mathbb{R}$$

be any symmetric positive semidefinite bilinear form that is computable via the DoFs (6.2a)–(6.2c) and satisfies the following properties:

- for any v_h in $V_h(K) \cap \ker(\Pi_p^\mathcal{N})$, we have that

$$S^K(v_h, v_h) = 0 \qquad \Longrightarrow \qquad v_h = 0; \tag{6.12}$$

- the following bound is valid with a positive constant $\widetilde{c}^* > 0$ independent of h_{I_n}, $h_{K_\mathbf{x}}$, and K:

$$S^K(v, v) \le \widetilde{c}^* \left(h_{K_\mathbf{x}}^{-2} \|v\|_{L^2(K)}^2 + \|\nabla_\mathbf{x} v\|_{L^2(K)^d}^2 + h_{K_\mathbf{x}}^{-2} h_{I_n}^2 \|\partial_t v\|_{L^2(K)}^2 \right) \qquad \forall v \in H^1(K). \tag{6.13}$$

Property (6.12) implies that $S^K(\cdot, \cdot)$ induces a norm in $V_h(K) \cap \ker(\Pi_p^\mathcal{N})$. Another consequence of (6.12) and the scaling argument in Remark 6.2, is that there exist two constants $0 < c_* < c^*$ independent of K such that

$$c_* |v_h|_{Y(K)}^2 \le \lambda S^K(v_h, v_h) \le c^* |v_h|_{Y(K)}^2 \qquad \forall v_h \in V_h(K) \cap \ker(\Pi_p^\mathcal{N}). \tag{6.14}$$

In fact, the functional $|\cdot|_{Y(K)}$ is a norm on $V_h(K) \cap \ker(\Pi_p^\mathcal{N})$.

We define the discrete counterpart of the local bilinear forms $a^K(\cdot, \cdot)$:

$$a_h^K(u_h, v_h) := a^K(\Pi_p^\mathcal{N} u_h, \Pi_p^\mathcal{N} v_h) + \lambda S^K((I - \Pi_p^\mathcal{N})u_h, (I - \Pi_p^\mathcal{N})v_h). \tag{6.15}$$

Lemma 6.11. *Property (6.14) implies that there exist two constants $0 < \alpha_* < \alpha^*$ independent of K such that the following local stability bounds are valid:*

$$\alpha_* |v_h|_{Y(K)}^2 \le a_h^K(v_h, v_h) \le \alpha^* |v_h|_{Y(K)}^2 \qquad \forall v_h \in V_h(K). \tag{6.16}$$

Proof. We only show the upper bound as the lower bound follows analogously leading to $\alpha_* := \min(1, c_*)$. We have

$$a_h^K(v_h, v_h) = \lambda \left\| \nabla_\mathbf{x} \Pi_p^\mathcal{N} v_h \right\|_{L^2(K)^d}^2 + \lambda S^K((I - \Pi_p^\mathcal{N})v_h, (I - \Pi_p^\mathcal{N})v_h)$$

$$\le \left| \Pi_p^\mathcal{N} v_h \right|_{Y(K)}^2 + c^* \left| (I - \Pi_p^\mathcal{N})v_h \right|_{Y(K)}^2 \le \max(1, c^*) \left(\left| \Pi_p^\mathcal{N} v_h \right|_{Y(K)}^2 + \left| (I - \Pi_p^\mathcal{N})v_h \right|_{Y(K)}^2 \right).$$

Pythagoras' theorem implies

$$a_h^K(v_h, v_h) \le \max(1, c^*) |v_h|_{Y(K)}^2.$$

This proves the upper bound in (6.16) with $\alpha^* = \max(1, c^*)$. $\square$

The global discrete bilinear form associated with the spatial Laplace operator $-\Delta_{\mathbf{x}}(\cdot)$ reads

$$a_h(u_h, v_h) := \sum_{K \in \mathcal{T}_h} a_h^K(u_h, v_h) \qquad \forall u_h, \ v_h \in Y_h.$$

Taking into account Proposition 6.10, an immediate consequence of (6.16) are the global stability bounds

$$\alpha_* \|v_h\|_{Y(\mathcal{T}_h)}^2 \le a_h(v_h, v_h) \le \alpha^* \|v_h\|_{Y(\mathcal{T}_h)}^2 \qquad \forall v_h \in Y_h. \tag{6.17}$$

For sufficiently smooth functions, we have the following upper bounds.

Proposition 6.12. *For all v in $H^1(\mathcal{T}_h)$, the following local and global bounds are valid: for all K in $\mathcal{T}_h$,*

$$a_h^K(v, v) \le 3\lambda \max(1, \widetilde{c}^*) \left(1 + (1 + c_{tr})c_{\Pi_p^N}\right)$$
$$\times \left(h_{K_{\mathbf{x}}}^{-2} \|v\|_{L^2(K)}^2 + \|\nabla_{\mathbf{x}} v\|_{L^2(K)^d}^2 + h_{K_{\mathbf{x}}}^{-2} h_{I_n}^2 \|\partial_t v\|_{L^2(K)}^2 \right), \tag{6.18}$$

$$a_h(v, v) \le 3\lambda \max(1, \widetilde{c}^*) \left(1 + (1 + c_{tr})c_{\Pi_p^N}\right)$$
$$\times \sum_{K \in \mathcal{T}_h} \left(h_{K_{\mathbf{x}}}^{-2} \|v\|_{L^2(K)}^2 + \|\nabla_{\mathbf{x}} v\|_{L^2(K)^d}^2 + h_{K_{\mathbf{x}}}^{-2} h_{I_n}^2 \|\partial_t v\|_{L^2(K)}^2 \right), \tag{6.19}$$

where $\widetilde{c}^*$ is the stability constant in (6.13), λ is the thermal conductivity, c_{tr} is the constant appearing in the elemental trace (in time) inequality, and $c_{\Pi_p^N}$ is the inverse estimate constant in (6.6).

Proof. The stability of the Π_p^N projector entails

$$a^K(\Pi_p^N v, \Pi_p^N v) = \lambda \left\|\nabla_{\mathbf{x}} \Pi_p^N v\right\|_{L^2(K)^d}^2 \le \lambda \left\|\nabla_{\mathbf{x}} v\right\|_{L^2(K)^d}^2.$$

Using definition (6.15) and bound (6.13), we deduce

$$a_h^K(v, v) = a^K(\Pi_p^N v, \Pi_p^N v) + \lambda S^K((I - \Pi_p^N)v, (I - \Pi_p^N)v)$$
$$\le \max(1, \widetilde{c}^*)\lambda \Big(\|\nabla_{\mathbf{x}} v\|_{L^2(K)^d}^2 + h_{K_{\mathbf{x}}}^{-2} \left\|(I - \Pi_p^N)v\right\|_{L^2(K)}^2$$
$$+ \left\|\nabla_{\mathbf{x}}(I - \Pi_p^N)v\right\|_{L^2(K)^d}^2 + h_{K_{\mathbf{x}}}^{-2} h_{I_n}^2 \left\|\partial_t(I - \Pi_p^N)v\right\|_{L^2(K)}^2 \Big). \tag{6.20}$$

Using the polynomial inverse estimate (6.5) with $q_p = \Pi_p^N v$, we can write

$$h_{K_{\mathbf{x}}}^{-2} \left\|\Pi_p^N v\right\|_{L^2(K)}^2 + \left\|\nabla_{\mathbf{x}} \Pi_p^N v\right\|_{L^2(K)^d}^2 + h_{K_{\mathbf{x}}}^{-2} h_{I_n}^2 \left\|\partial_t \Pi_p^N v\right\|_{L^2(K)}^2$$
$$= h_{K_{\mathbf{x}}}^{-2} \left(\left\|\Pi_p^N v\right\|_{L^2(K)}^2 + h_{K_{\mathbf{x}}}^2 \left\|\nabla_{\mathbf{x}} \Pi_p^N v\right\|_{L^2(K)^d}^2 + h_{I_n}^2 \left\|\partial_t \Pi_p^N v\right\|_{L^2(K)}^2\right)$$
$$\le c_{\Pi_p^N} h_{K_{\mathbf{x}}}^{-2} \left(h_{K_{\mathbf{x}}}^2 \left\|\nabla_{\mathbf{x}} \Pi_p^N v\right\|_{L^2(K)^d}^2 + \left\|\Pi_{p-1}^{0,K} \Pi_p^N v\right\|_{L^2(K)}^2 + h_{I_n} \left\|\Pi_0^{0,K_{\mathbf{x}}} \Pi_p^N v(\cdot, t_{n-1})\right\|_{L^2(K)}^2\right).$$

The definition of Π_p^N, and the stability of the L^2 and Π_p^N projectors entail

$$h_{K_{\mathbf{x}}}^{-2} \left\|\Pi_p^N v\right\|_{L^2(K)}^2 + \left\|\nabla_{\mathbf{x}} \Pi_p^N v\right\|_{L^2(K)^d}^2 + h_{K_{\mathbf{x}}}^{-2} h_{I_n}^2 \left\|\partial_t \Pi_p^N v\right\|_{L^2(K)}^2$$
$$\le c_{\Pi_p^N} h_{K_{\mathbf{x}}}^{-2} \left(h_{K_{\mathbf{x}}}^2 \|\nabla_{\mathbf{x}} v\|_{L^2(K)^d}^2 + \|v\|_{L^2(K)}^2 + h_{I_n} \|v(\cdot, t_{n-1})\|_{L^2(K)}^2\right).$$

Applying a trace inequality along the time variable (with constant c_{tr}) on the last term yields

$$h_{K_\mathbf{x}}^{-2} \left\| \Pi_p^\mathcal{N} v \right\|_{L^2(K)}^2 + \left\| \nabla_\mathbf{x} \Pi_p^\mathcal{N} v \right\|_{L^2(K)^d}^2 + h_{K_\mathbf{x}}^{-2} h_{I_n}^2 \left\| \partial_t \Pi_p^\mathcal{N} v \right\|_{L^2(K)}^2$$

$$\leq (1 + c_{tr}) c_{\Pi_p^\mathcal{N}} \left(h_{K_\mathbf{x}}^{-2} \left\| v \right\|_{L^2(K)}^2 + \left\| \nabla_\mathbf{x} v \right\|_{L^2(K)^d}^2 + h_{K_\mathbf{x}}^{-2} h_{I_n}^2 \left\| \partial_t v \right\|_{L^2(K)}^2 \right).$$

We insert this bound into (6.20) after applying the triangle inequality and obtain (6.18). Adding over all elements gives (6.19). $\qquad \square$

Henceforth, for a given v in $L^2(Q_T)$, we shall write

$$v^{(n)} := v_{|\Omega \times I_n} \qquad \text{for all } n = 1, \ldots, N.$$

For all u_h, v_h in Y_h and K in $\mathcal{T}_h$, $K = K_\mathbf{x} \times I_n$, we set

$$b_h^K(u_h, v_h) := \begin{cases} c_H (\partial_t \Pi_p^\star u_h, v_h)_{L^2(K)} + a_h^K(u_h, v_h) + c_H \left(\Pi_p^\star u_h^{(1)}(\cdot, 0), v_h^{(1)}(\cdot, 0) \right)_{L^2(K_\mathbf{x})} & \text{if } n = 1; \\ c_H (\partial_t \Pi_p^\star u_h, v_h)_{L^2(K)} + a_h^K(u_h, v_h) & \\ \quad + c_H \left(\Pi_p^\star u_h^{(n)}(\cdot, t_{n-1}) - \Pi_p^\star u_h^{(n-1)}(\cdot, t_{n-1}), v_h^{(n)}(\cdot, t_{n-1}) \right)_{L^2(K_\mathbf{x})} & \text{if } 2 \leq n \leq N. \end{cases} \tag{6.21}$$

The bilinear form $b_h^K(\cdot, \cdot)$ is computable through the DoFs. Actually, $\Pi_p^\star u_h^{(n)}(\cdot, t_{n-1}) = u_h^{(n)}(\cdot, t_{n-1})$ for $1 \leq n \leq N$ by the definition of $\Pi_p^\star$ in (6.4) and the definition of the local VE spaces.

We define the discrete counterpart of the global bilinear form $b(\cdot, \cdot)$ in (1.4) as follows:

$$b_h(u_h, v_h) := \sum_{K \in \mathcal{T}_h} b_h^K(u_h, v_h) \qquad \forall u_h, v_h \in Y_h. \tag{6.22}$$

The third terms in the definition of $b_h^K(u_h, v_h)$ in (6.21) stand for upwind fluxes for the weak imposition of the zero initial condition for $n = 1$, or of time continuity for $2 \leq n \leq N$.

An admissible stabilization

Consider the following stabilization, for $K = K_\mathbf{x} \times I_n$:

$$S^K(u_h, v_h) := h_{K_\mathbf{x}}^{-2} (\Pi_{p-1}^{0,K} u_h, \Pi_{p-1}^{0,K} v_h)_{L^2(K)} + h_{K_\mathbf{x}}^{-1} \sum_{F \in \mathcal{F}^K} (\Pi_p^{0,F} u_h, \Pi_p^{0,F} v_h)_{L^2(F)}$$

$$+ h_{K_\mathbf{x}}^{-2} h_{I_n} (u_h(\cdot, t_{n-1}), v_h(\cdot, t_{n-1}))_{L^2(K_\mathbf{x})}. \tag{6.23}$$

This bilinear form is computable via the DoFs.

Proposition 6.13. *The stabilization in* (6.23) *satisfies properties* (6.12) *and* (6.13).

Proof. Property (6.12) follows from the fact that $S^K(v_h, v_h)$ involves the squares of *all* the DoFs. Furthermore, property (6.13) follows from the stability of the L^2 projectors and the trace inequality applied to the time-like and space-like facet terms. $\qquad \square$

As pointed out in Section 6.3.4, property (6.12) and the scaling argument in Remark 6.2 imply that property (6.14) is satisfied as well.

6.3.5 The method

The proposed space–time VEM reads as follows:

$$\text{Find } u_h \in Y_h \text{ such that } b_h(u_h, v_h) = (f, \Pi_{p-1}^{0,Q_T} v_h)_{L^2(Q_T)} \qquad \forall v_h \in Y_h. \tag{6.24}$$

The projector Π_{p-1}^{0,Q_T} makes the right-hand side computable and is L^2 stable, which is used in the proof of the well-posedness of (6.24) in Theorem 6.17 below.

Under assumption (G1), the method can be solved in a time-marching fashion by solving the counterpart of (6.24) restricted to the time-slab I_n, for $n = 1, \dots, N-1$, and then transmitting the information to the subsequent time-slab I_{n+1} through upwinding.

Remark 6.14 (Nonhomogenous data)**.** The case of nonhomogeneous initial or Dirichlet boundary conditions can be dealt with by a standard lifting argument and a suitable modification of the right-hand side of (6.24), see Section 7.5 below. ∎

6.4 Well-posedness of the virtual element method

In this section, we prove well-posedness of the method in (6.24). To this aim, we endow the trial space with a suitable norm, which is defined by means of a VE Newton potential, see Section 6.4.1. In Section 6.4.2, we prove a discrete inf-sup condition. This proof extends that of [93, Thm. 2.1] to our setting, where multiple variational crimes have to be taken into account.

Before that, we prove a global Poincaré-type inequality for functions in the space Y_h.

Proposition 6.15. *Let assumptions* (G1)*–*(G3) *be valid. Then, there exists a positive constant C_P independent of the mesh size h such that*

$$\|v_h\|_{L^2(Q_T)} \le C_P \|v_h\|_{Y(\mathcal{T}_h)} \qquad \forall v_h \in Y_h. \tag{6.25}$$

Proof. It suffices to prove the counterpart of (6.25) over each time slab I_n. We define the scalar jump $[\![v_h]\!]$ as $[\![v_h]\!]_F \cdot \vec{\mathbf{n}}_F$. [2]

We start from the spatial Poincaré inequality in [11, Eqn. (1.3) for $d \ge 2$ and Sect. 8 for $d = 1$] with constant c_{PB} and integrate it in time over the time slab I_n:

$$\|v_h\|_{L^2(\Omega \times I_n)}^2 \le c_{PB} \Big(\sum_{K \in \mathcal{T}_h, K \subset \Omega \times I_n} \|\nabla_{\mathbf{x}} v_h\|_{L^2(K)^d}^2 + \sum_{F_{\mathbf{x}} \in \mathcal{F}_h^{\mathbf{x}}} h_{F_{\mathbf{x}}}^{-1} \int_{I_n} \Big(\int_{F_{\mathbf{x}}} [\![v_h]\!] dS \Big)^2 dt \Big).$$

For $d = 1$, the integral over the point $F_{\mathbf{x}}$ is the evaluation at $F_{\mathbf{x}}$.

To conclude, we have to estimate the second term on the right-hand side. To this aim, we prove estimates on each time-like facet and then collect them together. For simplicity, we further assume that $F = F_{\mathbf{x}} \times I_n$ is an internal (time-like) facet shared by two elements $K_1 = K_{\mathbf{x},1} \times I_n$ and $K_2 = K_{\mathbf{x},2} \times I_n$. The case of a boundary time-like facet can be dealt with similarly. Recall from the nonconformity of the space Y_h in (6.8), that $\Pi_p^{0,F} [\![v_h]\!] = 0$.

---

[2] We have that $[\![\cdot]\!]$ is a scalar function whereas $[\![\cdot]\!]_F$ defined in (6.7) is a vector field.

Using Jensen's inequality, we write

$$h_{F_\mathbf{x}}^{-1} \int_{I_n} \left(\int_{F_\mathbf{x}} [\![v_h]\!] \mathrm{d}S \right)^2 \mathrm{d}t \leq \int_{I_n} \int_{F_\mathbf{x}} [\![v_h]\!]^2 \mathrm{d}S \, \mathrm{d}t = \left\| [\![v_h]\!] - \Pi_p^{0,F} [\![v_h]\!] \right\|_{L^2(F)}^2 .$$

Denote the L^2 projection onto $\mathbb{P}^p(I_n)$ of the restriction of v_h to K_j by $q_p^{t,j}$, $j = 1, 2$. Let q_p^t be defined on $K_1 \cup K_2$ piecewise as $q_{p|K_j}^t = q_p^{t,j}$, $j = 1, 2$. A standard trace inequality in space with constant c_{tr}, and the local Poincaré inequality (6.11) give

$$h_{F_\mathbf{x}}^{-1} \int_{I_n} \left(\int_{F_\mathbf{x}} [\![v_h]\!] \mathrm{d}S \right)^2 \mathrm{d}t \leq c_{tr} \left(h_{K_{1,\mathbf{x}}}^{-1} \left\| v_h - q_p^t \right\|_{L^2(K_1)}^2 + h_{K_{1,\mathbf{x}}} \left\| \nabla_\mathbf{x}(v_h - q_p^t) \right\|_{L^2(K_1)^d}^2 \right.$$

$$\left. + h_{K_{2,\mathbf{x}}}^{-1} \left\| v_h - q_p^t \right\|_{L^2(K_2)}^2 + h_{K_{2,\mathbf{x}}} \left\| \nabla_\mathbf{x}(v_h - q_p^t) \right\|_{L^2(K_2)^d}^2 \right)$$

$$\leq 2 c_{tr} (\max_{K \in \mathcal{T}_h} C_P^K)^2 \left(h_{K_{1,\mathbf{x}}} \left\| \nabla_\mathbf{x}(v_h - q_p^t) \right\|_{L^2(K_1)^d}^2 + h_{K_{2,\mathbf{x}}} \left\| \nabla_\mathbf{x}(v_h - q_p^t) \right\|_{L^2(K_2)^d}^2 \right)$$

$$\leq 2 c_{tr} (\max_{K \in \mathcal{T}_h} C_P^K)^2 \max(h_{K_{1,\mathbf{x}}}, h_{K_{2,\mathbf{x}}}) \sum_{j=1}^2 \left\| \nabla_\mathbf{x} v_h \right\|_{L^2(K_j)^d}^2 .$$

Summing over all the time-like facets of the n-th time slab and recalling that the number of $(d-1)$-dimensional facets of each $K_\mathbf{x}$ is uniformly bounded with respect to the meshsize, see assumption (G3), we get the assertion. $\qquad \square$

6.4.1   A virtual element Newton potential

We define a VE Newton potential $\mathfrak{R}_h : \mathcal{S}^p(\mathcal{T}_h) \to Y_h$ as follows: for any ϕ_h in $\mathcal{S}^p(\mathcal{T}_h)$, $\mathfrak{R}_h \phi_h$ in Y_h solves

$$a_h(\mathfrak{R}_h \phi_h, v_h) = b_h(\phi_h, v_h) - a_h(\phi_h, v_h)$$
$$= c_H \Big[(\partial_t \phi_h, v_h)_{L^2(Q_T)} + \Big(\phi_h^{(1)}(\cdot, 0), v_h^{(1)}(\cdot, 0) \Big)_{L^2(\Omega)}$$
$$+ \sum_{n=2}^N \Big(\phi_h^{(n)}(\cdot, t_{n-1}) - \phi_h^{(n-1)}(\cdot, t_{n-1}), v_h^{(n)}(\cdot, t_{n-1}) \Big)_{L^2(\Omega)} \Big] \qquad \forall v_h \in Y_h. \tag{6.26}$$

Thanks to the stability bounds (6.17), the bilinear form $a_h(\cdot, \cdot)$ is continuous and coercive, and the continuity in the $Y(\mathcal{T}_h)$ norm of the functional on the right-hand side of (6.26) follows from Proposition 6.15. Therefore, the VE Newton potential is well defined.

We introduce the following norm on the sum space $X + Y_h$: for all v in $X + Y_h$,

$$\|v\|_{X(\mathcal{T}_h)}^2 := \|v\|_{Y(\mathcal{T}_h)}^2 + \left\| \mathfrak{R}_h(\Pi_p^\star v) \right\|_{Y(\mathcal{T}_h)}^2 + \frac{c_H}{2} \Big(\left\| \Pi_p^\star v^{(1)}(\cdot, 0) \right\|_{L^2(\Omega)}^2$$
$$+ \sum_{n=2}^N \left\| \Pi_p^\star v^{(n)}(\cdot, t_{n-1}) - \Pi_p^\star v^{(n-1)}(\cdot, t_{n-1}) \right\|_{L^2(\Omega)}^2 + \left\| \Pi_p^\star v^{(N)}(\cdot, T) \right\|_{L^2(\Omega)}^2 \Big). \tag{6.27}$$

Recalling the embedding $X \hookrightarrow C^0(0, T; L^2(\Omega))$ in (6.9), we have that $\Pi_p^\star$ in (6.4) is well-defined for functions in X. In Chapter 7, we shall present the convergence analysis of the method with respect to the $\|\cdot\|_{X(\mathcal{T}_h)}$ norm.

6.4.2 A discrete inf-sup condition and well-posedness of the method

In this section, we prove a discrete inf-sup condition in the spaces $(Y_h, \|\cdot\|_{X(\mathcal{T}_h)})$ for the trial functions and $(Y_h, \|\cdot\|_{Y(\mathcal{T}_h)})$ for the test functions.

Proposition 6.16. *There exists a positive constant γ_I independent of $\mathcal{T}_h$ such that*

$$\sup_{0 \neq v_h \in Y_h} \frac{b_h(u_h, v_h)}{\|v_h\|_{Y(\mathcal{T}_h)}} \geq \gamma_I \|u_h\|_{X(\mathcal{T}_h)} \qquad \forall u_h \in Y_h. \tag{6.28}$$

Proof. For any $u_h \in Y_h$, define $w_h := \mathfrak{R}_h(\Pi_p^\star u_h) \in Y_h$. It suffices to prove that

$$\frac{b_h(u_h, u_h + \delta w_h)}{\|u_h + \delta w_h\|_{Y(\mathcal{T}_h)}} \geq \gamma_I \|u_h\|_{X(\mathcal{T}_h)},$$

for a suitable real parameter $\delta > 0$, which will be fixed below.

The triangle inequality and the definition of the norm $\|\cdot\|_{X(\mathcal{T}_h)}$ in (6.27) imply

$$\|u_h + \delta w_h\|^2_{Y(\mathcal{T}_h)} \leq 2 \left(\|u_h\|^2_{Y(\mathcal{T}_h)} + \delta^2 \|w_h\|^2_{Y(\mathcal{T}_h)} \right) \leq 2\max(1, \delta^2) \|u_h\|^2_{X(\mathcal{T}_h)},$$

whence we deduce

$$\|u_h + \delta w_h\|_{Y(\mathcal{T}_h)} \leq \sqrt{2}\max(1, \delta) \|u_h\|_{X(\mathcal{T}_h)}. \tag{6.29}$$

Next, recalling (6.22) and (6.21), we write

$$b_h(u_h, u_h) = \sum_{K \in \mathcal{T}_h} \left(c_H(\partial_t \Pi_p^\star u_h, u_h)_{L^2(K)} + a_h^K(u_h, u_h) \right) + c_H \left\| \Pi_p^\star u_h^{(1)}(\cdot, 0) \right\|^2_{L^2(\Omega)}$$
$$+ c_H \sum_{n=2}^N \left(\Pi_p^\star u_h^{(n)}(\cdot, t_{n-1}) - \Pi_p^\star u_h^{(n-1)}(\cdot, t_{n-1}), u_h^{(n)}(\cdot, t_{n-1}) \right)_{L^2(\Omega)}. \tag{6.30}$$

For $K = K_{\mathbf{x}} \times I_n$, we have

$$(\partial_t \Pi_p^\star u_h, u_h)_{L^2(K)} \overset{(6.4a)}{=} (\partial_t \Pi_p^\star u_h, \Pi_p^\star u_h)_{L^2(K)}$$
$$= \frac{1}{2} \left(\left\| \Pi_p^\star u_h^{(n)}(\cdot, t_n) \right\|^2_{L^2(K_{\mathbf{x}})} - \left\| \Pi_p^\star u_h^{(n)}(\cdot, t_{n-1}) \right\|^2_{L^2(K_{\mathbf{x}})} \right).$$

By (6.4b), we have $u_h^{(n)}(\cdot, t_{n-1}) = \Pi_p^\star u_h^{(n)}(\cdot, t_{n-1})$. Simple calculations give

$$
\sum_{K \in \mathcal{T}_h} (\partial_t \Pi_p^\star u_h, u_h)_{L^2(K)} + \left\| \Pi_p^\star u_h^{(1)}(\cdot, 0) \right\|_{L^2(\Omega)}^2
$$

$$
+ \sum_{n=2}^{N} \left(\Pi_p^\star u_h^{(n)}(\cdot, t_{n-1}) - \Pi_p^\star u_h^{(n-1)}(\cdot, t_{n-1}), u_h^{(n)}(\cdot, t_{n-1}) \right)_{L^2(\Omega)}
$$

$$
= \sum_{n=1}^{N} \left(\frac{1}{2} \left\| \Pi_p^\star u_h^{(n)}(\cdot, t_n) \right\|_{L^2(\Omega)}^2 - \frac{1}{2} \left\| \Pi_p^\star u_h^{(n)}(\cdot, t_{n-1}) \right\|_{L^2(\Omega)}^2 \right) + \left\| \Pi_p^\star u_h^{(1)}(\cdot, 0) \right\|_{L^2(\Omega)}^2
$$

$$
+ \sum_{n=2}^{N} \left(\left\| \Pi_p^\star u_h^{(n)}(\cdot, t_{n-1}) \right\|_{L^2(\Omega)}^2 - \left(\Pi_p^\star u_h^{(n-1)}(\cdot, t_{n-1}), \Pi_p^\star u_h^{(n)}(\cdot, t_{n-1}) \right)_{L^2(\Omega)} \right)
$$

$$
= \frac{1}{2} \left\| \Pi_p^\star u_h^{(1)}(\cdot, 0) \right\|_{L^2(\Omega)}^2 + \sum_{n=2}^{N} \frac{1}{2} \left\| \Pi_p^\star u_h^{(n)}(\cdot, t_{n-1}) - \Pi_p^\star u_h^{(n-1)}(\cdot, t_{n-1}) \right\|_{L^2(\Omega)}^2
$$

$$
+ \frac{1}{2} \left\| \Pi_p^\star u_h^{(N)}(\cdot, T) \right\|_{L^2(\Omega)}^2.
$$

Therefore, from (6.30) and (6.17), we get

$$
\begin{aligned}
b_h(u_h, u_h) \geq & \alpha_* \left\| u_h \right\|_{Y(\mathcal{T}_h)}^2 + \frac{c_H}{2} \left(\left\| \Pi_p^\star u_h^{(1)}(\cdot, 0) \right\|_{L^2(\Omega)}^2 \right. \\
& \left. + \sum_{n=2}^{N} \left\| \Pi_p^\star u_h^{(n)}(\cdot, t_{n-1}) - \Pi_p^\star u_h^{(n-1)}(\cdot, t_{n-1}) \right\|_{L^2(\Omega)}^2 + \left\| \Pi_p^\star u_h^{(N)}(\cdot, T) \right\|_{L^2(\Omega)}^2 \right).
\end{aligned} \tag{6.31}
$$

Moreover, the definition of $b_h(\cdot, \cdot)$ in (6.21) and (6.22), and the definition of the VE Newton potential in (6.26) imply

$$
b_h(u_h, \delta w_h) = \delta \left(a_h(w_h, w_h) + a_h(u_h, w_h) \right). \tag{6.32}
$$

Since (6.17) gives $a_h(w_h, w_h) \geq \alpha_* \left\| w_h \right\|_{Y(\mathcal{T}_h)}^2$, then Young's inequality entails, for all positive ε,

$$
\begin{aligned}
a_h(u_h, w_h) \geq & - (a_h(u_h, u_h))^{\frac{1}{2}} (a_h(w_h, w_h))^{\frac{1}{2}} \geq -\alpha^* \left\| u_h \right\|_{Y(\mathcal{T}_h)} \left\| w_h \right\|_{Y(\mathcal{T}_h)} \\
\geq & -\frac{\alpha^*}{2\varepsilon} \left\| u_h \right\|_{Y(\mathcal{T}_h)}^2 - \frac{\alpha^* \varepsilon}{2} \left\| w_h \right\|_{Y(\mathcal{T}_h)}^2.
\end{aligned}
$$

Inserting the two above inequalities into (6.32) yields

$$
b_h(u_h, \delta w_h) \geq \delta \left(\alpha_* - \frac{\alpha^* \varepsilon}{2} \right) \left\| w_h \right\|_{Y(\mathcal{T}_h)}^2 - \frac{\alpha^* \delta}{2\varepsilon} \left\| u_h \right\|_{Y(\mathcal{T}_h)}^2. \tag{6.33}
$$

As a final step, we sum (6.31) and (6.33):

$$
\begin{aligned}
b_h(u_h, u_h + \delta w_h) \geq & \left(\alpha_* - \frac{\alpha^* \delta}{2\varepsilon} \right) \left\| u_h \right\|_{Y(\mathcal{T}_h)}^2 + \delta \left(\alpha_* - \frac{\alpha^* \varepsilon}{2} \right) \left\| w_h \right\|_{Y(\mathcal{T}_h)}^2 + \frac{c_H}{2} \left(\left\| \Pi_p^\star u_h^{(1)}(\cdot, 0) \right\|_{L^2(\Omega)}^2 \right. \\
& \left. + \sum_{n=2}^{N} \left\| \Pi_p^\star u_h^{(n)}(\cdot, t_{n-1}) - \Pi_p^\star u_h^{(n-1)}(\cdot, t_{n-1}) \right\|_{L^2(\Omega)}^2 + \left\| \Pi_p^\star u_h^{(N)}(\cdot, T) \right\|_{L^2(\Omega)}^2 \right).
\end{aligned}
$$

Taking $0 < \varepsilon < (2\alpha_*)/\alpha^*$ and $0 < \delta < (2\varepsilon\alpha_*)/\alpha^*$, defining

$$\beta := \min\left(\alpha_* - \frac{\alpha^*\delta}{2\varepsilon}, \delta\left(\alpha_* - \frac{\alpha^*\varepsilon}{2}\right)\right) > 0,$$

and recalling (6.27) and (6.29), we can write

$$b_h(u_h, u_h + \delta w_h) \geq \beta\left(\|u_h\|^2_{Y(\mathcal{T}_h)} + \|w_h\|^2_{Y(\mathcal{T}_h)}\right) + \frac{c_H}{2}\left(\left\|\Pi^\star_p u_h^{(1)}(\cdot, 0)\right\|^2_{L^2(\Omega)}\right.$$

$$+ \sum_{n=2}^{N}\left\|\Pi^\star_p u_h^{(n)}(\cdot, t_{n-1}) - \Pi^\star_p u_h^{(n-1)}(\cdot, t_{n-1})\right\|^2_{L^2(\Omega)} + \left\|\Pi^\star_p u_h^{(N)}(\cdot, T)\right\|^2_{L^2(\Omega)}\right)$$

$$\geq \min(1, \beta)\|u_h\|^2_{X(\mathcal{T}_h)} \geq \frac{\min(1, \beta)}{\sqrt{2}\max(1, \delta)}\|u_h\|_{X(\mathcal{T}_h)}\|u_h + \delta w_h\|_{Y(\mathcal{T}_h)}.$$

The assertion follows with $\gamma_I := \min(1, \beta)/(\sqrt{2}\max(1, \delta))$. $\qquad\square$

We are in a position to prove the well-posedness of the method in (6.24).

Theorem 6.17 (Well-posedness of the space–time VEM)**.** *There exists a unique solution u_h to the method in* (6.24) *with the following continuous dependence on the data:*

$$\|u_h\|_{X(\mathcal{T}_h)} \leq \gamma_I^{-1} C_P \lambda^{-1} \|f\|_{L^2(Q_T)},$$

where γ_I is the discrete inf-sup constant in (6.28)*, C_P is the Poincaré-type inequality constant in* (6.25) *and λ is the thermal conductivity.*

Proof. The discrete inf-sup condition (6.28) implies uniqueness of the solution. The existence follows from the uniqueness, owing to the finite dimensionality of Y_h. As for the stability bound, we apply again the inf-sup condition (6.28) and recall the definition (6.24) of the method:

$$\|u_h\|_{X(\mathcal{T}_h)} \leq \frac{1}{\gamma_I}\sup_{0 \neq v_h \in Y_h}\frac{b_h(u_h, v_h)}{\|v_h\|_{Y(\mathcal{T}_h)}} = \frac{1}{\gamma_I}\sup_{0 \neq v_h \in Y_h}\frac{(f, \Pi^{0,Q_T}_{p-1}v_h)_{L^2(Q_T)}}{\|v_h\|_{Y(\mathcal{T}_h)}}.$$

The Cauchy-Schwarz inequality, the L^2 stability of Π^{0,Q_T}_{p-1}, and the global Poincaré-type inequality (6.25) give the assertion. $\qquad\square$

Chapter 7

Error analysis

Contents

7.1   Introduction

In this chapter, we analyze the convergence of the virtual element method in (6.24). More precisely, we present an *a priori* error estimate and prove quasi-optimal convergence rates in the $X(\mathcal{T}_h)$-norm in (6.27) for the h-version of the method. We first use the discrete inf-sup estimate (6.28) to obtain a Strang-type bound for the error of the method in the $X(\mathcal{T}_h)$-norm, which is the sum of four terms: *i)* a VE interpolation error; *ii)* a term related to the discretization of the right-hand side f; *iii)* a term measuring the spatial nonconformity of the discrete space; *iv)* a term involving polynomial error estimates, which arise from the nonconformity in time and the polynomial inconsistency of the discrete bilinear form. Optimal h-convergence for sufficiently smooth solutions is then proven in Theorem 7.4 by simply estimating each term.

The numerical experiments in Section 7.5 illustrate some additional properties of the proposed method:

- The method converges with optimal rates of order $O(h^{p+1})$ in a discrete $H^1(0, T; H^{-1}(\Omega))$-type norm.

- Although the method is not based on tensor-product polynomial spaces, optimal rates of convergence of order $O(h^{p+1})$ are observed for the error in the $L^2(Q_T)$-norm. When non tensor-product polynomial spaces are used for the interior penalty DG formulation in [14], only suboptimal rates of order $O(h^{p+\frac{1}{2}})$ are obtained in the $L^2(Q_T)$-norm, see [14, Thm. 5.11].

- The method shows the highest convergence rates that can be expected for singular solutions according to their regularity.

- The κ_2-condition number of the stiffness matrix growths as $O(h^{-1})$ for all $p \in \mathbb{N}$.

- Exponential convergence is obtained for the p-version of the method for smooth solutions.

Structure of the chapter: We start by introducing some technical tools in Section 7.2, which are typical of the nonconforming framework. Then, in Sections 7.3 and 7.4, we develop an *a priori* error analysis in two steps: first, we prove a Strang-type bound; next, we derive optimal convergence rates, by using interpolation and polynomial approximation results, assuming sufficient regularity on the solution. We conclude with some numerical experiments in Section 7.5.

7.2 Technical results

Introduce the bilinear form $\mathcal{N}_h : L^2\big(0, T; H^{\frac{3}{2}+\varepsilon}(\Omega)\big) \times Y_h \to \mathbb{R}$ given by

$$\mathcal{N}_h(u, v_h) := \lambda \sum_{n=1}^{N} \int_{I_n} \sum_{F_\mathbf{x} \in \mathcal{F}_h^\mathbf{x}} \int_{F_\mathbf{x}} \nabla_\mathbf{x} u \cdot [\![v_h]\!]_F \, \mathrm{d}S \, \mathrm{dt}. \tag{7.1}$$

This bilinear form encodes information on the nonconformity of the space Y_h across time-like facets.

On $K = I_n \times K_\mathbf{x}$, define the local bilinear form

$$b^K(w, v) := \int_{I_n} \int_{K_\mathbf{x}} (c_H \partial_t w v + \lambda \nabla_\mathbf{x} w \cdot \nabla_\mathbf{x} v) \, \mathrm{d}\mathbf{x} \, \mathrm{dt}.$$

Lemma 7.1 (Nonconformity in space of Y_h)**.** *Assume that the solution u to the continuous problem* (1.5) *belongs to* $L^2(0, T; H^{\frac{3}{2}+\varepsilon}(\Omega))$. *Then, for all v_h in Y_h,*

$$\sum_{K \in \mathcal{T}_h} b^K(u, v_h) = (f, v_h)_{L^2(Q_T)} + \mathcal{N}_h(u, v_h). \tag{7.2}$$

Proof. Integrating by parts in space and recalling the definition of $\mathcal{N}_h$ in (7.1), we can write

$$\sum_{K \in \mathcal{T}_h} b^K(u, v_h) = \sum_{K \in \mathcal{T}_h} \int_{I_n} \left(\int_{K_\mathbf{x}} (c_H \partial_t u - \lambda \Delta_\mathbf{x} u) v_h \mathrm{d}\mathbf{x} + \lambda \sum_{F_\mathbf{x} \in \mathcal{F}^{K_\mathbf{x}}} \int_{F_\mathbf{x}} v_h (\vec{\mathbf{n}}_{F_\mathbf{x}} \cdot \nabla_\mathbf{x} u) \mathrm{d}S \right) \mathrm{dt}$$

$$= (f, v_h)_{L^2(Q_T)} + \mathcal{N}_h(u, v_h). \qquad \square$$

We introduce another preliminary result, which characterizes the polynomial inconsistency of the method in (6.24). To this aim, we define the bilinear form $\mathcal{J}^K : H^1(\mathcal{T}_h) \times Y_h \to \mathbb{R}$ given on each $K = K_\mathbf{x} \times I_n$ in $\mathcal{T}_h$ as follows: for all w in $H^1(\mathcal{T}_h)$, v_h in Y_h,

$$\mathcal{J}^K(w, v_h) := \begin{cases} c_H \left(\Pi_p^\star w^{(1)}(\cdot, 0), v_h^{(1)}(\cdot, 0) \right)_{L^2(K_\mathbf{x})} & \text{if } n = 1; \\ c_H \left(\Pi_p^\star w^{(n)}(\cdot, t_{n-1}) - \Pi_p^\star w^{(n-1)}(\cdot, t_{n-1}), v_h^{(n)}(\cdot, t_{n-1}) \right)_{L^2(K_\mathbf{x})} & \text{if } 2 \leq n \leq N. \end{cases} \tag{7.3}$$

This bilinear form encodes the polynomial inconsistency of the method at space-like facets, as stated in the following lemma.

Lemma 7.2 (Polynomial inconsistency of $b_h(\cdot, \cdot)$)**.** *The local bilinear forms $b_h^K(\cdot, \cdot)$ satisfy*

$$b_h^K(q_p, v_h) = b^K(q_p, v_h) + \mathcal{J}^K(q_p, v_h) \qquad \forall q_p \in \mathcal{S}^p(\mathcal{T}_h), \ \forall v_h \in V_h(K), \ \forall K \in \mathcal{T}_h. \quad (7.4)$$

Proof. Thanks to the definition of the bilinear form $a^K(\cdot, \cdot)$, the orthogonality properties of the projector $\Pi_p^{\mathcal{N}}$, and the fact that the projectors $\Pi_p^{\star}$ and $\Pi_p^{\mathcal{N}}$ preserve polynomials of degree p, we have

$$
\begin{aligned}
b_h^K(q_p, v_h) &= c_H(\partial_t \Pi_p^{\star} q_p, v_h)_{L^2(K)} + a^K(\Pi_p^{\mathcal{N}} q_p, \Pi_p^{\mathcal{N}} v_h) \\
&\quad + \lambda S^K((I - \Pi_p^{\mathcal{N}}) q_p, (I - \Pi_p^{\mathcal{N}}) v_h) + \mathcal{J}^K(\Pi_p^{\star} q_p, v_h) \\
&= c_H(\partial_t \Pi_p^{\star} q_p, v_h)_{L^2(K)} + a^K(q_p, \Pi_p^{\mathcal{N}} v_h) + \mathcal{J}^K(q_p, v_h) \\
&= c_H(\partial_t q_p, v_h)_{L^2(K)} + a^K(q_p, v_h) + \mathcal{J}^K(q_p, v_h) = b^K(q_p, v_h) + \mathcal{J}^K(q_p, v_h). \ \square
\end{aligned}
$$

7.3 A Strang-type result

We prove an *a priori* bound for the error of the method (6.24) in the $X(\mathcal{T}_h)$-norm.

Theorem 7.3. *Let u and u_h be the solutions to (1.5) (with homogeneous initial and boundary conditions) and (6.24), respectively. Assume that u belongs to $L^2(0, T, H^{\frac{3}{2}+\varepsilon}(\Omega))$ for some $\varepsilon > 0$. Denoting by $u_I \in Y_h$ the DoF interpolant of u in Y_h, and γ_I be the discrete inf-sup constant appearing in (6.28). Then, we have*

$$
\begin{aligned}
\|u - u_h\|_{X(\mathcal{T}_h)} &\leq \|u - u_I\|_{Y(\mathcal{T}_h)} + \gamma_I^{-1} \sup_{0 \neq v_h \in Y_h} \left[\frac{|(f - \Pi_{p-1}^{0,Q_T} f, v_h)_{L^2(Q_T)}|}{\|v_h\|_{Y(\mathcal{T}_h)}} + \frac{|\mathcal{N}_h(u, v_h)|}{\|v_h\|_{Y(\mathcal{T}_h)}} \right. \\
&\quad \left. + \inf_{q_p \in \mathcal{S}^p(\mathcal{T}_h)} \frac{\sum_{K \in \mathcal{T}_h} \left(b_h^K(u - q_p, v_h) - b^K(u - q_p, v_h) + \mathcal{J}^K(q_p, v_h) \right)}{\|v_h\|_{Y(\mathcal{T}_h)}} \right].
\end{aligned}
\tag{7.5}
$$

Proof. By the triangle inequality, we have

$$\|u - u_h\|_{X(\mathcal{T}_h)} \leq \|u - u_I\|_{X(\mathcal{T}_h)} + \|u_I - u_h\|_{X(\mathcal{T}_h)} =: T_1 + T_2.$$

Since $\Pi_p^{\star}$ is computable from the DoFs, we have that $\Pi_p^{\star}(u - u_I) = 0$ in each element. Taking into account (6.27), this yields

$$
\begin{aligned}
T_1^2 &= \|u - u_I\|_{Y(\mathcal{T}_h)}^2 + \left\| \mathfrak{R}_h \left(\Pi_p^{\star}(u - u_I) \right) \right\|_{Y(\mathcal{T}_h)}^2 + \frac{c_H}{2} \Big(\left\| \Pi_p^{\star}(u - u_I)(\cdot, 0) \right\|_{L^2(\Omega)}^2 \\
&\quad + \sum_{n=2}^{N} \left\| \Pi_p^{\star}(u - u_I^{(n)})(\cdot, t_{n-1}) - \Pi_p^{\star}(u - u_I^{(n-1)})(\cdot, t_{n-1}) \right\|_{L^2(\Omega)}^2 + \left\| \Pi_p^{\star}(u - u_I)(\cdot, T) \right\|_{L^2(\Omega)}^2 \Big) \\
&= \|u - u_I\|_{Y(\mathcal{T}_h)}^2.
\end{aligned}
$$

The rest of the proof is devoted to estimating the term T_2. The definition of u_I implies

$$b_h(u_I, v_h) = b_h(u, v_h) \qquad \forall v_h \in Y_h.$$

Using this property and the discrete inf-sup condition (6.28), we get

$$\|u_I - u_h\|_{X(\mathcal{T}_h)} \le \gamma_I^{-1} \sup_{0 \ne v_h \in Y_h} \frac{b_h(u_I - u_h, v_h)}{\|v_h\|_{Y(\mathcal{T}_h)}} = \gamma_I^{-1} \sup_{0 \ne v_h \in Y_h} \frac{b_h(u - u_h, v_h)}{\|v_h\|_{Y(\mathcal{T}_h)}}.$$

We recall (6.24), add and subtract any q_p in $\mathcal{S}^p(\mathcal{T}_h)$, use the inconsistency property (7.4), add and subtract u, recall the property of the nonconformity bilinear form $\mathcal{N}_h(\cdot, \cdot)$ in (7.2), and deduce

$$b_h(u - u_h, v_h) = \sum_{K \in \mathcal{T}_h} b_h^K(u, v_h) - (f, \Pi_{p-1}^{0,Q_T} v_h)_{L^2(Q_T)}$$

$$= \sum_{K \in \mathcal{T}_h} \left(b_h^K(u - q_p, v_h) + b_h^K(q_p, v_h) \right) - (f, \Pi_{p-1}^{0,Q_T} v_h)_{L^2(Q_T)}$$

$$= \sum_{K \in \mathcal{T}_h} \left(b_h^K(u - q_p, v_h) + b^K(q_p, v_h) + \mathcal{J}^K(q_p, v_h) \right) - (f, \Pi_{p-1}^{0,Q_T} v_h)_{L^2(Q_T)}$$

$$= \sum_{K \in \mathcal{T}_h} \left(b_h^K(u - q_p, v_h) + b^K(q_p - u, v_h) + b^K(u, v_h) + \mathcal{J}^K(q_p, v_h) \right) - (f, \Pi_{p-1}^{0,Q_T} v_h)_{L^2(Q_T)}$$

$$= \sum_{K \in \mathcal{T}_h} \left(b_h^K(u - q_p, v_h) - b^K(u - q_p, v_h) + \mathcal{J}^K(q_p, v_h) \right) + (f - \Pi_{p-1}^{0,Q_T} f, v_h)_{L^2(Q_T)} + \mathcal{N}_h(u, v_h).$$

The assertion follows from taking the infimum over all q_p in $\mathcal{S}^p(\mathcal{T}_h)$ and then the supremum over all v_h in Y_h. $\qquad\square$

7.4 *A priori* error estimate

The aim of this section is to prove optimal convergence rates for the method in (6.24). So far, we have derived all estimates with explicit constants, so as to track the use of different type of inequalities (Poincaré, trace, inverse estimates, ...). Furthermore, we kept separated the contributions of $h_{K_\mathbf{x}}$ and h_{I_n}. In this section, we shall not keep this level of detail.

We prove error estimates under some regularity assumptions on the exact solution and focus on the case of isotropic space–time meshes, i.e., assume that

$$h_{K_\mathbf{x}} \simeq h_{I_n} \simeq h_K \qquad \forall K = K_\mathbf{x} \times I_n \in \mathcal{T}_h. \tag{7.6}$$

In Theorem 7.3, we proved that the error of the method in (6.24) is bounded by the sum of four terms: *i)* a VE interpolation error; *ii)* a term involving the discretization of the right-hand side f; *iii)* a term measuring the spatial nonconformity of the discrete space Y_h; *iv)* a term involving polynomial error estimates, which arise from the nonconformity in time and the polynomial inconsistency of the discrete bilinear form. Based on that result, we prove the following theorem.

Theorem 7.4. *Let assumptions* (G1)–(G4) *be valid, and* $\mathcal{T}_h$ *be isotropic in the sense of* (7.6). *Assume that* $u \in H^{p+1}(\mathcal{T}_h)$ *and* $f \in H^p(\mathcal{T}_h)$, *where* $p \ge 1$ *denotes the degree of approximation of the method in* (6.24). *Then,*

$$\|u - u_h\|_{X(\mathcal{T}_h)} \lesssim h^p \left(|u|_{H^{p+1}(\mathcal{T}_h)} + |f|_{H^p(\mathcal{T}_h)} \right). \tag{7.7}$$

Proof. We estimate the four terms on the right-hand side of (7.5) separately. The assertion then follows by combining the four bounds we provide below.

Part *i*) VE interpolation error. For any q_p in $\mathcal{S}^p(\mathcal{T}_h)$, the triangle inequality implies

$$\|u - u_I\|_{Y(\mathcal{T}_h)} \leq |u - q_p|_{Y(\mathcal{T}_h)} + |q_p - u_I|_{Y(\mathcal{T}_h)}. \tag{7.8}$$

We focus on the second term on the right-hand side. For any $K \in \mathcal{T}_h$, $(q_p - u_I)_{|K}$ belongs to $V_h(K)$. Therefore, the stability bounds (6.16) entail

$$|q_p - u_I|^2_{Y(K)} \lesssim a_h^K(q_p - u_I, q_p - u_I) \qquad \forall K \in \mathcal{T}_h.$$

Since u_I is the DoFs interpolant of u and $a_h(\cdot, \cdot)$ is computed via the DoFs, the above inequality also implies

$$|q_p - u_I|^2_{Y(\mathcal{T}_h)} \lesssim a_h(q_p - u_I, q_p - u_I) = a_h(q_p - u, q_p - u).$$

Furthermore, using the discrete continuity property (6.19), we arrive at

$$|q_p - u_I|^2_{Y(\mathcal{T}_h)} \lesssim \sum_{K \in \mathcal{T}_h} \left(h_K^{-2} \|u - q_p\|^2_{L^2(K)} + |u - q_p|^2_{H^1(K)} \right).$$

Inserting this into (7.8) and using standard polynomial approximation results yield

$$\|u - u_I\|_{Y(\mathcal{T}_h)} \lesssim h^p |u|_{H^{p+1}(\mathcal{T}_h)}.$$

Part *ii*) Handling the variational crime on the right-hand side f. Using the definition of Π_{p-1}^{0,Q_T}, standard polynomial approximation estimates, and the global discrete Poincaré inequality (6.25) entail

$$(f - \Pi_{p-1}^{0,Q_T} f, v_h)_{L^2(Q_T)} \leq \sum_{K \in \mathcal{T}_h} \left\| f - \Pi_{p-1}^{0,K} f \right\|_{L^2(K)} \|v_h\|_{L^2(K)} \lesssim h^p |f|_{H^p(\mathcal{T}_h)} \|v_h\|_{L^2(Q_T)}$$

$$\lesssim h^p |f|_{H^p(\mathcal{T}_h)} \|v_h\|_{Y(\mathcal{T}_h)}.$$

Part *iii*) Handling the variational crime of the time-like nonconformity. We estimate

$$\sup_{0 \neq v_h \in Y_h} \frac{|\mathcal{N}_h(u, v_h)|}{\|v_h\|_{Y(\mathcal{T}_h)}} = \sup_{0 \neq v_h \in Y_h} \frac{|\lambda \sum_{F \in \mathcal{F}_h} \int_{I_n} \int_{F_\mathbf{x}} \nabla_\mathbf{x} u \cdot [\![v_h]\!]_F \mathrm{d}S \, \mathrm{dt}|}{\|v_h\|_{Y(\mathcal{T}_h)}}.$$

We present estimates on a single facet $F = F_\mathbf{x} \times I_n$. For the sake of simplicity, we assume that F is an internal time-like facet shared by the elements K_1 and K_2. Using the definition of the spatial nonconformity of the space Y_h, see (6.8), and the properties of L^2 projectors, for all $q_p^{t,1}$, $q_p^{t,2}$ in $\mathbb{P}_p(I_n)$, we write [1]

$$\left| \int_{I_n} \int_{F_\mathbf{x}} \nabla_\mathbf{x} u \cdot [\![v_h]\!]_F \mathrm{d}S \, \mathrm{dt} \right| = \left| \int_{I_n} \int_{F_\mathbf{x}} \nabla_\mathbf{x} u \cdot \vec{\mathbf{n}}_{F_\mathbf{x}} [\![v_h]\!] \mathrm{d}S \, \mathrm{dt} \right|$$

$$= \left| \int_{I_n} \int_{F_\mathbf{x}} \left(\nabla_\mathbf{x} u \cdot \vec{\mathbf{n}}_{F_\mathbf{x}} - \Pi_p^{0,F} \left(\nabla_\mathbf{x} u \cdot \vec{\mathbf{n}}_{F_\mathbf{x}} \right) \right) [\![v_h]\!] \mathrm{d}S \, \mathrm{dt} \right|$$

$$= \left| \int_{I_n} \int_{F_\mathbf{x}} \left(\nabla_\mathbf{x} u \cdot \vec{\mathbf{n}}_{F_\mathbf{x}} - \Pi_p^{0,F} \left(\nabla_\mathbf{x} u \cdot \vec{\mathbf{n}}_{F_\mathbf{x}} \right) \right) \left((v_{h|K_2} - q_p^{t,2}) - (v_{h|K_1} - q_p^{t,1}) \right) \mathrm{d}S \, \mathrm{dt} \right|.$$

---

[1] Here we use the scalar normal jump $[\![\cdot]\!]$ defined in the proof of Proposition 6.15.

Next, use the Cauchy-Schwarz inequality, the triangle inequality, standard properties of the L^2 projector, a trace inequality, the local quasi-uniformity of the space–time mesh, and arrive at

$$
\left| \int_{I_n} \int_{F_\mathbf{x}} \nabla_\mathbf{x} u \cdot [\![v_h]\!]_F \mathrm{d}S \, \mathrm{dt} \right|
$$

$$
\leq \left\| \nabla_\mathbf{x} u \cdot \vec{\mathbf{n}}_{F_\mathbf{x}} - \Pi_p^{0,F} \left(\nabla_\mathbf{x} u \cdot \vec{\mathbf{n}}_{F_\mathbf{x}} \right) \right\|_{L^2(F)} \left(\left\| v_{h|K_2} - q_p^{t,2} \right\|_{L^2(F)} + \left\| v_{h|K_1} - q_p^{t,1} \right\|_{L^2(F)} \right)
$$

$$
\leq \left\| \nabla_\mathbf{x} \left(u - \Pi_{p+1}^{0,K_1} u \right) \cdot \vec{\mathbf{n}}_{F_\mathbf{x}} \right\|_{L^2(F)} \left(\left\| v_{h|K_2} - q_p^{t,2} \right\|_{L^2(F)} + \left\| v_{h|K_1} - q_p^{t,1} \right\|_{L^2(F)} \right)
$$

$$
\lesssim \left(h_{K_1}^{-\frac{1}{2}} \left| u - \Pi_{p+1}^{0,K_1} u \right|_{Y(K_1)} + h_{K_1}^{\epsilon} \left| u - \Pi_{p+1}^{0,K_1} u \right|_{H^{\frac{3}{2}+\varepsilon}(K_1)} \right)
$$

$$
\times \left(\sum_{j=1,2} \left(h_{K_j}^{-\frac{1}{2}} \left\| v_h - q_p^{t,j} \right\|_{L^2(K_j)} + h_{K_j}^{\frac{1}{2}} |v_h|_{Y(K_j)} \right) \right).
$$

An application of (6.11) yields

$$
\left| \int_{I_n} \int_{F_\mathbf{x}} \nabla_\mathbf{x} u \cdot [\![v_h]\!]_F \mathrm{d}S \, \mathrm{dt} \right| \lesssim \left(\left| u - \Pi_{p+1}^{0,K_1} u \right|_{Y(K_1)} + h_{K_1}^{\epsilon+\frac{1}{2}} \left| u - \Pi_{p+1}^{0,K_1} u \right|_{H^{\frac{3}{2}+\varepsilon}(K_1)} \right) \sum_{j=1,2} |v_h|_{Y(K_j)}.
$$

Summing up over all the elements and using approximation properties of the L^2 projector, we eventually get

$$
\sup_{0 \neq v_h \in Y_h} \frac{|\mathcal{N}_h(u, v_h)|}{\|v_h\|_{Y(\mathcal{T}_h)}} \lesssim h^{p+1} |u|_{H^{p+1}(\mathcal{T}_h)}.
$$

Part *iv.a*) Polynomial approximation error of $b^K(\cdot, \cdot)$ type. Let q_p be in $\mathcal{S}^p(\mathcal{T}_h)$. Using the Cauchy-Schwarz inequality twice and the definition of the bilinear form $b^K(\cdot, \cdot)$ give

$$
b^K(u - q_p, v_h) = c_H \left(\partial_t (u - q_p), v_h \right)_{L^2(K)} + \lambda \left(\nabla_\mathbf{x}(u - q_p), \nabla_\mathbf{x} v_h \right)_{L^2(K)}
$$

$$
\lesssim |u - q_p|_{H^1(K)} (\|v_h\|_{L^2(K)} + |v_h|_{Y(K)}).
$$

Summing up over all the elements, using an ℓ^2 Cauchy-Schwarz inequality, and recalling the global Poincaré-type inequality (6.25), we can write

$$
\sum_{K \in \mathcal{T}_h} b^K(u - q_p, v_h) \lesssim |u - q_p|_{H^1(\mathcal{T}_h)} \|v_h\|_{Y(\mathcal{T}_h)} . \tag{7.9}
$$

Part *iv.b*) Polynomial approximation error of $b_h^K(\cdot, \cdot) + \mathcal{J}^K(\cdot, \cdot)$ type. Thanks to definitions (6.21) and (7.3) on each element $K = K_\mathbf{x} \times I_n$ in $\mathcal{T}_h$, for all v_h in Y_h, we have

$$
b_h^K(u - q_p, v_h) + \mathcal{J}^K(q_p, v_h) = c_H(\partial_t \Pi_p^\star (u - q_p), v_h)_{L^2(K)} + a_h^K(u - q_p, v_h) + \mathcal{J}^K(u, v_h), \tag{7.10}
$$

where q_p is the same as in Part *iv.a*). We first focus on the second term. Using the stability bound (6.16) and the continuity property (6.18) with $h_{K_\mathbf{x}} \simeq h_{I_n}$, we arrive at

$$
a_h^K(u - q_p, v_h) \leq a_h^K(u - q_p, u - q_p)^{\frac{1}{2}} a_h^K(v_h, v_h)^{\frac{1}{2}} \lesssim \left(h_K^{-1} \|u - q_p\|_{L^2(K)} + |u - q_p|_{H^1(K)} \right) |v_h|_{Y(K)}.
$$

Next, we deal with the first term on the right-hand side of (7.10). The Cauchy-Schwarz inequality yields

$$(\partial_t \Pi_p^\star(u - q_p), v_h)_{L^2(K)} \leq \left\| \partial_t \Pi_p^\star(u - q_p) \right\|_{L^2(K)} \|v_h\|_{L^2(K)}.$$

A polynomial inverse inequality gives

$$\left\| \partial_t \Pi_p^\star(u - q_p) \right\|_{L^2(K)} \lesssim h_K^{-1} \left\| \Pi_p^\star(u - q_p) \right\|_{L^2(K)}. \tag{7.11}$$

By using (6.6), the definition of $\Pi_p^\star$, the stability of the L^2 orthogonal projection, and the trace inequality, we arrive at

$$
\begin{aligned}
\left\| \Pi_p^\star(u - q_p) \right\|_{L^2(K)} &\lesssim \left\| \Pi_{p-1}^{0,K} \Pi_p^\star(u - q_p) \right\|_{L^2(K)} + h_K^{\frac{1}{2}} \left\| \Pi_p^{0,K_\mathbf{x}} \Pi_p^\star(u - q_p)(\cdot, t_{n-1}) \right\|_{L^2(K_\mathbf{x})} \\
&= \left\| \Pi_{p-1}^{0,K}(u - q_p) \right\|_{L^2(K)} + h_K^{\frac{1}{2}} \left\| \Pi_p^{0,K_\mathbf{x}}(u - q_p)(\cdot, t_{n-1}) \right\|_{L^2(K_\mathbf{x})} \\
&\leq \left\| u - q_p \right\|_{L^2(K)} + h_K^{\frac{1}{2}} \left\| (u - q_p)(\cdot, t_{n-1}) \right\|_{L^2(K_\mathbf{x})} \\
&\lesssim \left\| u - q_p \right\|_{L^2(K)} + h_K |u - q_p|_{H^1(K)}.
\end{aligned} \tag{7.12}
$$

Therefore, we obtain

$$(\partial_t \Pi_p^\star(u - q_p), v_h)_{L^2(K)} \lesssim \left(h_K^{-1} \left\| u - q_p \right\|_{L^2(K)} + |u - q_p|_{H^1(K)} \right) \|v_h\|_{L^2(K)}.$$

Finally, we estimate the third term on the right-hand side of (7.10). Since the initial condition $u(\cdot, 0)$ is zero, $\mathcal{J}^K(u, v_h) = 0$ if $n = 1$. So, we consider the case $n \geq 2$:

$$
\begin{aligned}
\mathcal{J}^K(u, v_h) &= c_H(\Pi_p^\star u^{(n)}(\cdot, t_{n-1}) - \Pi_p^\star u^{(n-1)}(\cdot, t_{n-1}), v_h^{(n)}(\cdot, t_{n-1}))_{L^2(K_\mathbf{x})} \\
&\lesssim h_K^{\frac{1}{2}} \Big(\left\| u(\cdot, t_{n-1}) - \Pi_p^\star u^{(n)}(\cdot, t_{n-1}) \right\|_{L^2(K_\mathbf{x})} \\
&\quad + \left\| u(\cdot, t_{n-1}) - \Pi_p^\star u^{(n-1)}(\cdot, t_{n-1}) \right\|_{L^2(K_\mathbf{x})} \Big) h_K^{-\frac{1}{2}} \left\| v_h^{(n)}(\cdot, t_{n-1}) \right\|_{L^2(K_\mathbf{x})}.
\end{aligned}
$$

Proceeding as in Proposition 6.13, it is possible to show that

$$h_K^{-\frac{1}{2}} \left\| v_h^{(n)}(\cdot, t_{n-1}) \right\|_{L^2(K_\mathbf{x})} \lesssim |v_h|_{Y(K)}.$$

Thus, we can focus on the two terms involving u. As for the first one, we use a trace inequality along the time direction, add and subtract the same q_p as above, recall that $\Pi_p^\star$ preserves polynomials of degree at most p, use the triangle inequality, apply the polynomial inverse estimate (7.11), and get

$$
\begin{aligned}
h_K^{\frac{1}{2}} \left\| u(\cdot, t_{n-1}) - \Pi_p^\star u^{(n)}(\cdot, t_{n-1}) \right\|_{L^2(K_\mathbf{x})} &\lesssim \left\| u - \Pi_p^\star u \right\|_{L^2(K)} + h_K \left\| \partial_t(u - \Pi_p^\star u) \right\|_{L^2(K)} \\
&\leq \left\| u - q_p \right\|_{L^2(K)} + h_K \left\| \partial_t(u - q_p) \right\|_{L^2(K)} + \left\| \Pi_p^\star(u - q_p) \right\|_{L^2(K)} + h_K \left\| \partial_t(\Pi_p^\star(u - q_p)) \right\|_{L^2(K)} \\
&\leq \left\| u - q_p \right\|_{L^2(K)} + h_K \left\| \partial_t(u - q_p) \right\|_{L^2(K)} + \left\| \Pi_p^\star(u - q_p) \right\|_{L^2(K)}.
\end{aligned}
$$

Next, we apply estimate (7.12) and get

$$h_K^{\frac{1}{2}} \left\| u(\cdot, t_{n-1}) - \Pi_p^\star u^{(n)}(\cdot, t_{n-1}) \right\|_{L^2(K_\mathbf{x})} \lesssim \left\| u - q_p \right\|_{L^2(K)} + h_K |u - q_p|_{H^1(K)}.$$

For the second term involving u, we proceed analogously. Setting $K' := K_{\mathbf{x}} \times I_{n-1}$ and using the local quasi-uniformity of the space–time mesh, we get

$$h_K^{\frac{1}{2}} \left\| u(\cdot, t_{n-1}) - \Pi_p^\star u^{(n-1)}(\cdot, t_{n-1}) \right\|_{L^2(K_{\mathbf{x}})} \lesssim \left\| u - q_p \right\|_{L^2(K')} + h_{K'} \left| u - q_p \right|_{H^1(K')}.$$

Summing over all the elements, using standard manipulations (including ℓ^2 Cauchy-Schwarz inequalities), and applying the global Poincaré type inequality (6.25) give

$$\sum_{K \in \mathcal{T}_h} \left(b_h^K(u - q_p, v_h) + \mathcal{J}^K(q_p, v_h) \right) \lesssim \sum_{K \in \mathcal{T}_h} \left(h_K^{-2} \left\| u - q_p \right\|_{L^2(K)}^2 + \left| u - q_p \right|_{H^1(K)}^2 \right)^{\frac{1}{2}} \left\| v_h \right\|_{Y(\mathcal{T}_h)}. \tag{7.13}$$

Conclusion of Part *iv*) From (7.9) and (7.13), which are valid for any $q_p \in V_h(K)$, and standard polynomial approximation results, we obtain

$$\sup_{0 \neq v_h \in Y_h} \inf_{q_p \in \mathcal{S}^p(\mathcal{T}_h)} \frac{\sum_{K \in \mathcal{T}_h} \left(b_h^K(u - q_p, v_h) - b^K(u - q_p, v_h) + \mathcal{J}^K(q_p, v_h) \right)}{\left\| v_h \right\|_{Y(\mathcal{T}_h)}} \lesssim h^p |u|_{H^{p+1}(\mathcal{T}_h)}. \quad \square$$

7.5 Numerical results

In this section, we assess the error estimate proven in Theorem 7.4. We briefly mention some relevant computational aspects regarding the numerical results for the $(1 + 1)$-dimensional test cases below.

- In case of nonhomogeneous initial and/or boundary conditions, we set moments at $\Omega \times \{0\}$ and/or at $\partial\Omega \times (0, T)$ accordingly and modify the right-hand side. This corresponds to a standard lifting procedure, where the lifting has all the remaining moments equal to zero. In the presence of incompatible initial and boundary data, no artificial compatibility condition is enforced on the discrete solutions.

- Since the virtual element solution u_h is not known in closed form and the error in the $X(\mathcal{T}_h)$-norm is not computable, we report the following error quantities: given u_h the solution to (6.24),

$$\mathcal{E}^Y := \left\| u - \Pi_p^N u_h \right\|_{Y(\mathcal{T}_h)}, \quad \mathcal{E}^N := \left\| \Pi_p^N (\mathfrak{R}_h \Pi_p^\star(u - u_h)) \right\|_{Y(\mathcal{T}_h)},$$

$$(\mathcal{E}^U)^2 := \frac{c_H}{2} \left(\left\| \Pi_p^\star(u - u_h)(\cdot, 0) \right\|_{L^2(\Omega)}^2 + \left\| \Pi_p^\star(u - u_h)(\cdot, T) \right\|_{L^2(\Omega)}^2 \right.$$
$$\left. + \sum_{n=2}^{N} \left\| \Pi_p^\star(u - u_h)^{(n)}(\cdot, t_{n-1}) - \Pi_p^\star(u - u_h)^{(n-1)}(\cdot, t_{n-1}) \right\|_{L^2(\Omega)}^2 \right). \tag{7.14a}$$

Moreover, we also show the error in the $L^2(Q_T)$-norm

$$\mathcal{E}^L := \left\| u - \Pi_p^\star u_h \right\|_{L^2(Q_T)}, \tag{7.14b}$$

which is not covered by our theory.

- We are also interested in comparing the results with those obtained for the continuous FEM discretization of (1.5) in [93]: find $\widetilde{u}_h \in \widetilde{X}_h$ such that

$$b(\widetilde{u}_h, \widetilde{v}_h) = (f, \widetilde{v}_h)_{L^2(Q_T)} \qquad \forall \widetilde{v}_h \in \widetilde{Y}_h. \tag{7.15}$$

Here, $\widetilde{Y}_h$ is a space of continuous piecewise polynomials over a space–time simplicial tessellation of Q_T, and $\widetilde{X}_h$ is the subspace of $\widetilde{Y}_h$ of functions with zero initial condition.

We consider the following error quantities

$$\widetilde{\mathcal{E}}^Y = \|u - \widetilde{u}_h\|_Y, \qquad \widetilde{\mathcal{E}}^N = \|\mathcal{N}_h(u - \widetilde{u}_h)\|_Y, \tag{7.16}$$

where the discrete Newton potential $\mathcal{N}_h(u - \widetilde{u}_h) \in \widetilde{Y}_h$ is the solution to the following variational problem:

$$a(\mathcal{N}_h(u - \widetilde{u}_h), \widetilde{v}_h) = b(u - \widetilde{u}_h, \widetilde{v}_h) - a(u - \widetilde{u}_h, \widetilde{v}_h) \qquad \forall \widetilde{v}_h \in \widetilde{Y}_h.$$

- In all experiments, we take $c_H = \lambda = 1$ and employ the stabilization in (6.23).

- We use tensor-product meshes and uniform partitions along the space and time directions.

7.5.1 Patch test

The discrete bilinear form $b_h(\cdot, \cdot)$ in (6.22) is polynomial inconsistent; see Lemma 7.2. However, thanks to the error estimate (7.7), the method in (6.24) passes the patch test, i.e., up to round-off errors, polynomial solutions of order p are approximated exactly.

We consider the following family of exact solutions on the space–time domain $Q_T = (0, 1) \times (0, 1)$:

$$u_p(x, t) = \begin{cases} t^{p/2} x^{p/2} & \text{if } p \text{ is even;} \\ t^{(p-1)/2} x^{(p+1)/2} + t^{(p+1)/2} x^{(p-1)/2} & \text{if } p \text{ is odd.} \end{cases} \tag{7.17}$$

For any $p \in \mathbb{N}$, u_p belongs to $\mathbb{P}^p(Q_T)$. In Figure 7.1, for $p = 1, \ldots, 5$, we show the errors in the approximation of u_p obtained using a sequence of meshes with $h_x = h_t = 0.2 \times 2^{-i}$, $i = 1, \ldots, 5$, and approximation degree p. The scale of 10^{-10} in the figures validates the patch test. The growth of the error observed while decreasing the mesh size represents the actual effect of the condition number when solving the linear systems stemming from (6.24).

7.5.2 Smooth solution

On the space–time domain $Q_T = (0, 1) \times (0, 1)$, we consider the nonhomogeneous problem with exact smooth solution

$$u(x, t) = \sin(t) \sin(3\pi x). \tag{7.18}$$

In Figure 7.2, we show the errors in (7.14) obtained using a sequence of meshes with $h_x = h_t = 0.2 \times 2^{-i}$ for $i = 1, \ldots, 5$, and different approximation degrees p. We observe convergence of order $O(h^p)$ for the $\mathcal{E}^Y$ error, of order $O(h^{p+1})$ for the $\mathcal{E}^N$ error, of order $O(h^{p+\frac{1}{2}})$ for the $\mathcal{E}^U$ error, and of order $O(h^{p+1})$ for the $\mathcal{E}^L$ error.

For the conforming finite element method in [93], convergence in the Y-norm (1.3) of both the discrete solution and its Newton potential is of order $O(h^p)$, see Figure 7.3.
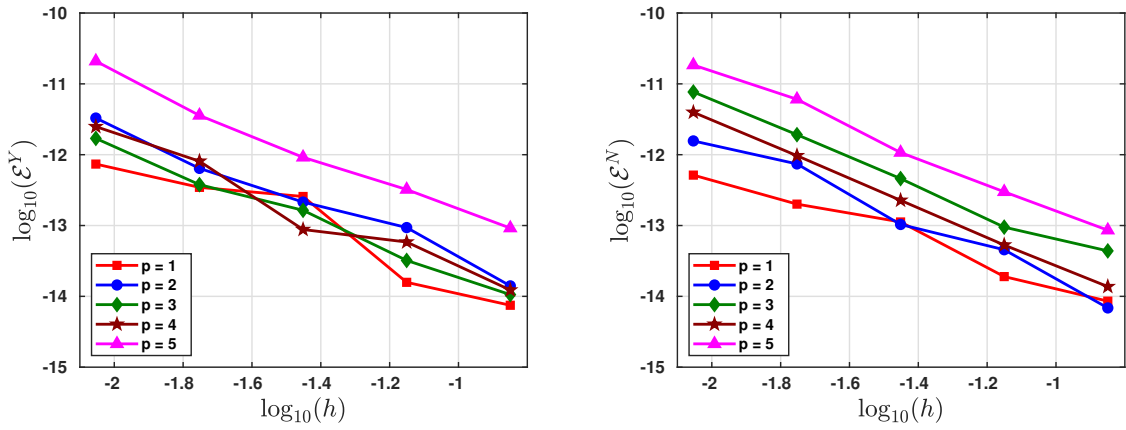
Figure 7.1 *h*-dependence of the VEM errors in (7.14) for the patch tests with solution u_p in (7.17).
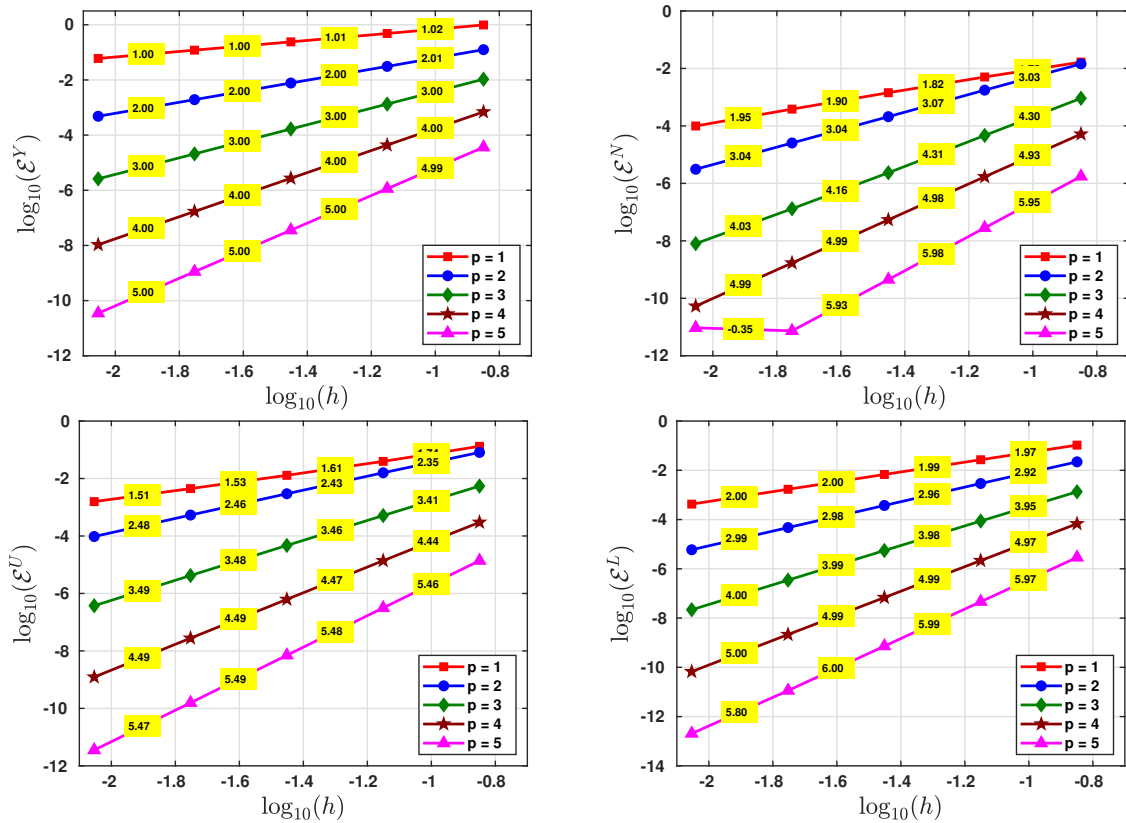

Figure 7.2 *h*-convergence of the VEM errors in (7.14) for the test case with smooth solution (7.18).

7.5.3   Singular solutions

We assess the convergence of the method for test cases with finite Sobolev regularity using the same sequence of meshes as in Section 7.5.1. Given the space–time domain $Q_T = (0,1) \times (0,1)$ and $\alpha > 1/2$, we consider the singular solutions
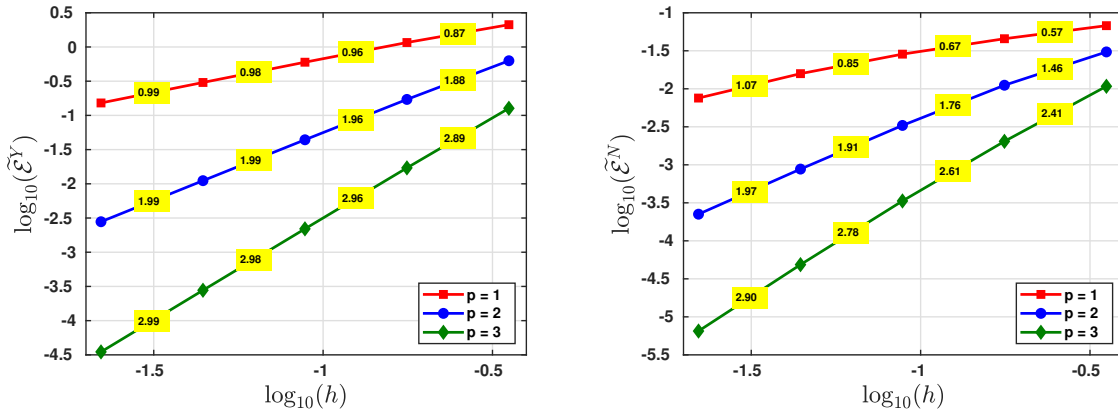
$$u_\alpha(x,t) = t^\alpha \sin(\pi x). \tag{7.19}$$

94

Figure 7.3 h-convergence of the FEM errors $\widetilde{\mathcal{E}}^Y$ and $\widetilde{\mathcal{E}}^N$ in (7.16) for the test case with smooth solution (7.18).

We have that $u_\alpha, \partial_x u_\alpha \in H^{\alpha+1/2-\epsilon}(0,1; C^\infty(0,1))$ and $\partial_t u_\alpha \in H^{\alpha-1/2-\epsilon}$ for any $\epsilon > 0$. The singularity occurs at the initial time. The errors in (7.14) for a sequence of meshes with $h_x = h_t = 0.2 \times 2^{-i}$ for $i = 1, \ldots, 6$, are depicted in Figures 7.4 and 7.5 for $\alpha = 0.55$ and $\alpha = 0.75$, respectively. We observe convergence of order $O(h^{\min\{p,\alpha+1/2\}})$ for the error $\mathcal{E}^Y$, of order $O(h^{\alpha-\frac{1}{2}})$ for the error $\mathcal{E}^N$, of order $O(h^\alpha)$ for the error $\mathcal{E}^U$, and of order $O(h^{\alpha+\frac{1}{2}})$ for the error $\mathcal{E}^L$.

For the conforming FEM in [93], only convergence of order $O(h^\alpha)$ is observed for $\widetilde{\mathcal{E}}^Y$ when $p > 1$.

7.5.4 Incompatible initial and boundary conditions

On the space–time domain $Q_T = (0,1) \times (0,1)$, we consider the heat equation problem (1.2) with zero source term ($f = 0$), homogeneous Dirichlet boundary conditions ($g = 0$), and constant initial condition ($u_0 = 1$). The corresponding exact solution is given by the Fourier series

$$u(x,t) = \sum_{n=0}^{\infty} \frac{4}{(2n+1)\pi} \sin\left((2n+1)\pi x\right) \exp\left(-(2n+1)^2\pi^2 t\right). \tag{7.20}$$

Due to the incompatibility of the initial and boundary conditions, u is discontinuous at $(0,0)$ and $(1,0)$, and does not belong to $H^1(Q_T)$ but belongs to $H^s\left(0,1; H_0^1(0,1)\right)$ for any $s < 1/4$; see [91, §7.1].

In Figure 7.7, we show the errors obtained on a sequence of uniform Cartesian meshes for the proposed VEM and on a sequence of structured triangular meshes for the conforming finite element method in [93].

The conforming finite element method does not converge in the Y-norm, while the $\mathcal{E}^Y$ error of the VEM converges with order $O(h^{1/4})$. For the computation of the error, we truncate the series (7.20) at $n = 250$.
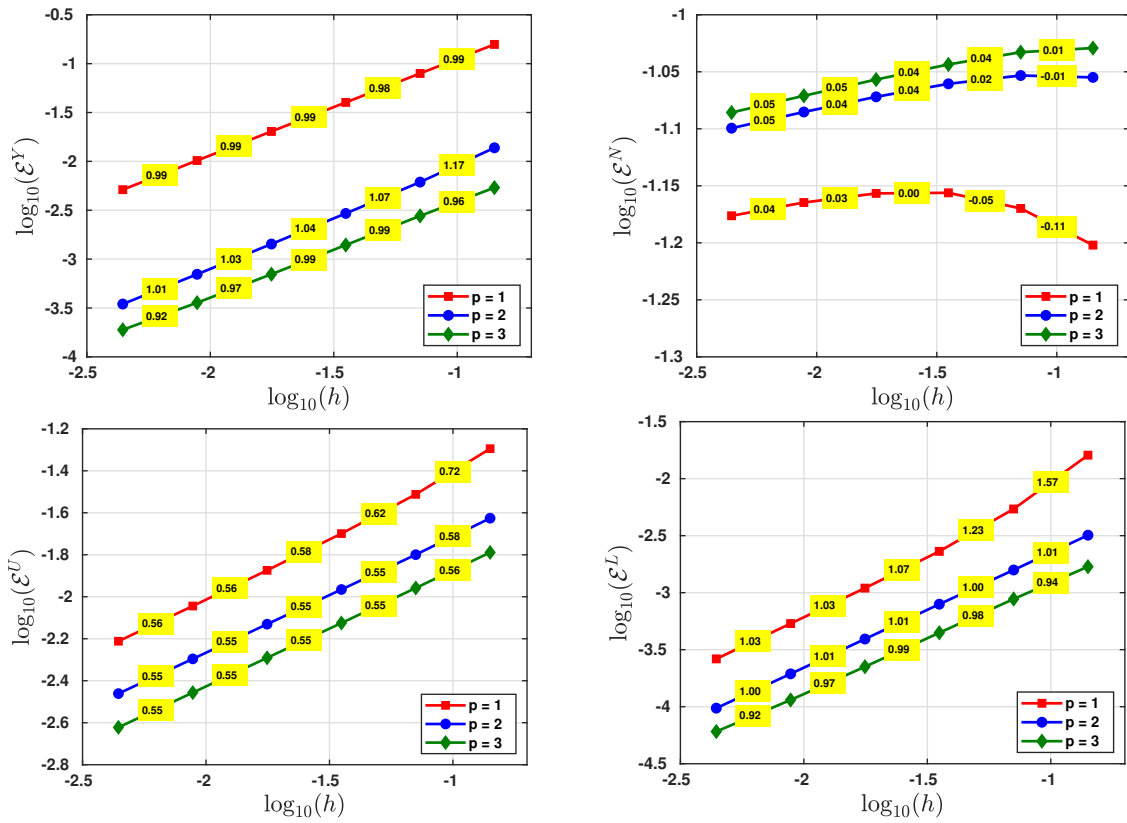
Figure 7.4 h-convergence of the VEM errors in (7.14) for the test case with singular solutions u_α (7.19), $\alpha = 0.55$.

7.5.5 Conditioning

We assess the conditioning of the matrices stemming from the method (6.24). We denote by **K** and **A** the matrices associated with the restrictions of the bilinear forms $b_h(\cdot, \cdot)$ and $a_h(\cdot, \cdot)$ to each time-slab. Observe that **K** is used to compute the solution of the VEM (6.24), while **A** is used only to evaluate the error $\mathcal{E}^N$ in (7.14). In Figure 7.8, we show the condition number $\kappa_2(\cdot)$ of both matrices, where asymptotic growths of order $O(h^{-1})$ and $O(h^{-2})$ are observed for **K** and **A**, respectively.

7.5.6 p-convergence

We are also interested in the performance of the p-version of the method, i.e., we fix a mesh and increase the degree of approximation. This is worth investigating also in view of the design of and hp refinements. We consider the smooth solution test case from Section 7.5.2 with a fixed mesh with $h_x = h_t = 0.1$. The results are shown in Figure 7.9 in *semilogy* scale. We observe the expected exponential convergence of order $O(e^{-b\sqrt{N_{DoFs}}})$.
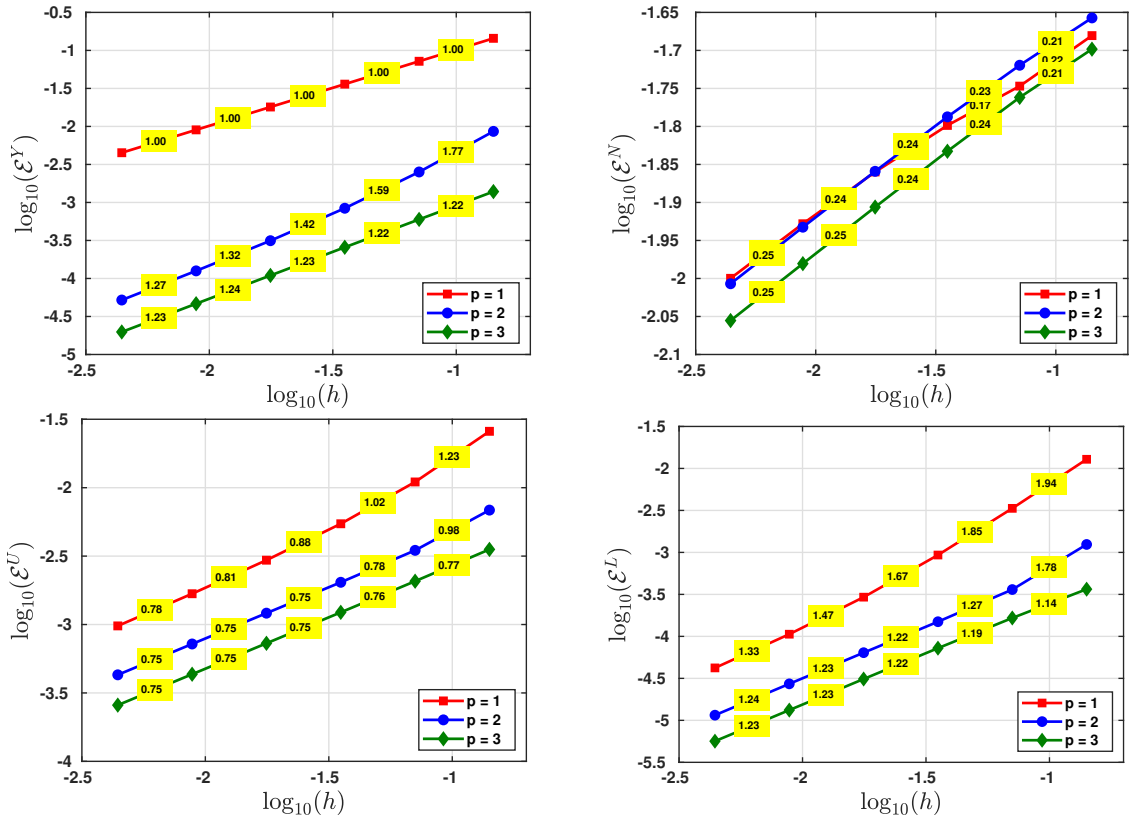
Figure 7.5 *h*-convergence of the VEM errors in (7.14) for the test case with singular solutions u_α (7.19), $\alpha = 0.75$.
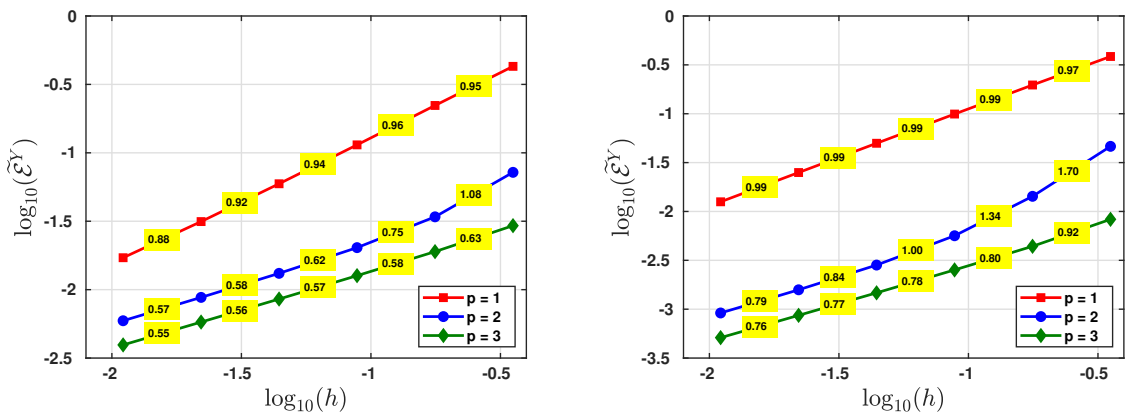

Figure 7.6 *h*-convergence of the FEM error $\widetilde{\mathcal{E}}^Y$ in (7.16) for the test case with singular solutions u_α (7.19), $\alpha = 0.55$ and $\alpha = 0.75$.
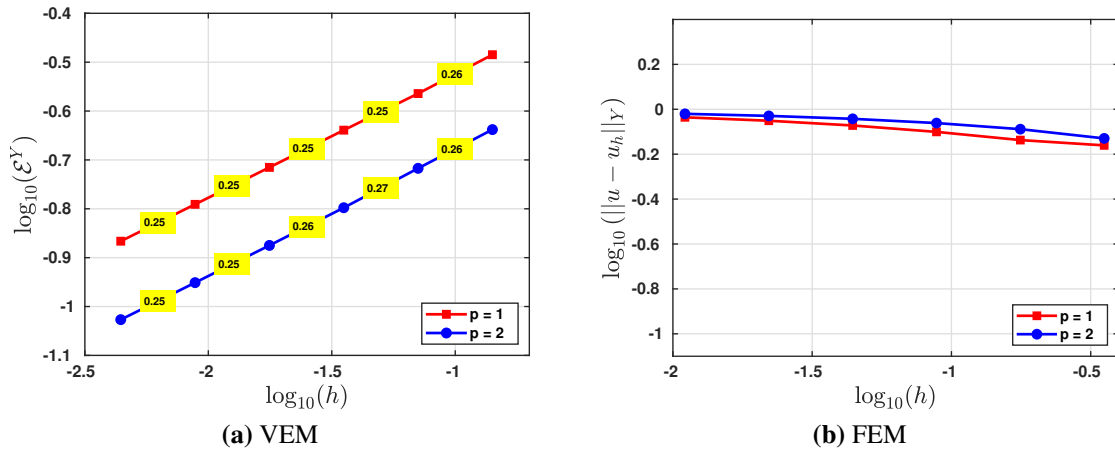
(a) VEM

(b) FEM

Figure 7.7 h-convergence of the VEM error $\mathcal{E}^Y$ in (7.14) and the FEM error $\widetilde{\mathcal{E}}^Y$ in (7.16) for the test case with incompatible initial and boundary conditions.
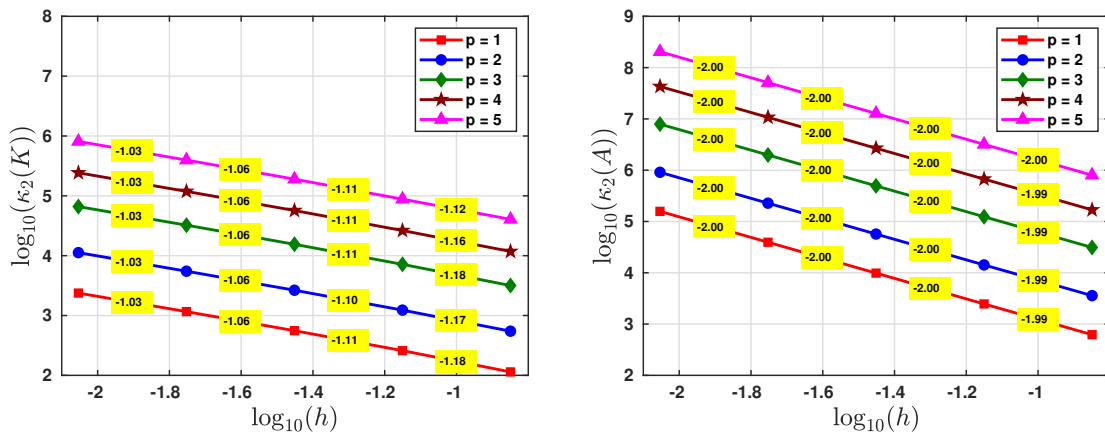

Figure 7.8 κ_2-condition number of the matrices stemming from the VEM in (6.24).
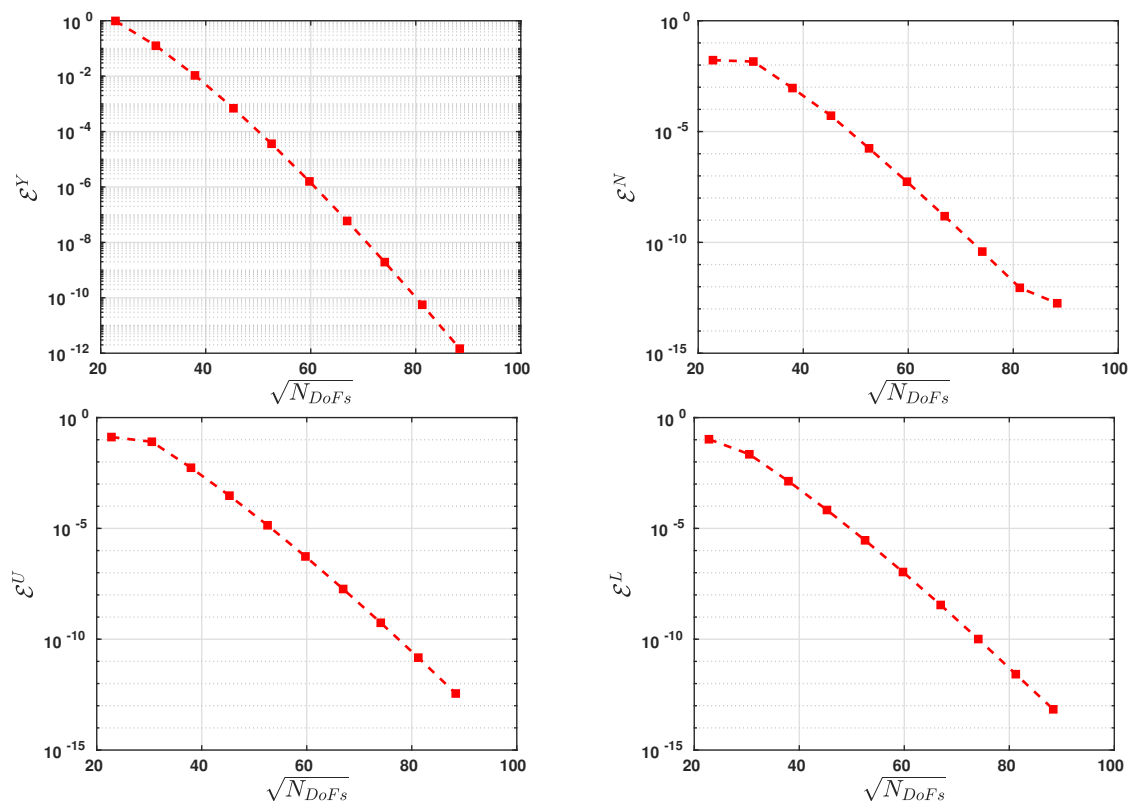
Figure 7.9 *p*-convergence of the VEM errors (7.14) for the smooth solution in (7.18).

Chapter 8

Extension to general prismatic meshes and variable degrees of accuracy

Contents

8.1   Introduction

In this chapter, we extend the nonconforming space–time virtual element method in (6.24) to the case of general prismatic meshes and nonuniform degrees of accuracy. This method allows for the use of space–time meshes with hanging facets (nodes, edges, faces) and is based on possibly discontinuous in time test and trial functions that are solutions to local space–time problems with polynomial data. This approach provides a natural framework for space–time adaptivity without need of re-meshing neighbouring elements as hanging facets are naturally included in the polytopic framework. If the space–time mesh is decomposed into separate time-slabs, it still allows for splitting the global linear system into much smaller systems that can be solved sequentially. Due to the nonconformity across time-like facets, the design and implementation of the method are independent of the space dimension. No artificial compatibility of initial and boundary conditions is enforced on the trial virtual element space, and optimal convergence rates have been achieved also for singular solutions. For tensor-product space–time meshes, the method presented in this chapter reduces to (6.24) in Chapter 6.

Structure of the paper. We introduce general prismatic meshes in Section 8.2; design local virtual element spaces and describe their degrees of freedom (DoFs) in Section 8.3; define global nonconforming space–time virtual element spaces in Section 8.4; detail the discrete bilinear forms in Section 8.5; present the method in Section 8.6. We discuss space–time virtual elements with variable degrees of accuracy in Section 8.7.

8.2   General prismatic meshes

As in Section 2.2, we consider sequences of meshes $\{\mathcal{T}_h\}$ consisting of nonoverlapping prismatic elements covering Q_T in the following sense: each element $\mathcal{K} \in \mathcal{T}_h$ can be written as $\mathcal{K} = \mathcal{K}_\mathbf{x} \times \mathcal{K}_t$, where $\mathcal{K}_\mathbf{x}$ is an d-dimensional open polytope with boundary $\partial \mathcal{K}_\mathbf{x}$ and $\mathcal{K}_t := (a_t, b_t) \subset (0, T)$. We denote the diameter of $\mathcal{K}_\mathbf{x}$ and the length of $\mathcal{K}_t$ by $h_{\mathcal{K}_\mathbf{x}}$ and $h_{\mathcal{K}_t}$, respectively. We call "mesh facet" any intersection $F = \partial \mathcal{K}^1 \cap \partial \mathcal{K}^2$, $F = \partial \mathcal{K}^1 \cap (\partial \Omega \times (0, T))$, $F = \partial \mathcal{K}^1 \cap (\Omega \times \{0\})$ or $F = \partial \mathcal{K}^1 \cap (\Omega \times \{T\})$, for $\mathcal{K}^1$, $\mathcal{K}^2 \in \mathcal{T}_h$, that has positive d-dimensional measure and is contained in a d-dimensional hyperplane. Let $\vec{n}_\mathcal{K}^F = (\vec{n}_{\mathcal{K}_\mathbf{x}}^F, \vec{n}_{\mathcal{K}_t}^F) \in \mathbb{R}^{d+1}$ be the outward-pointing unite normal vector on $\partial \mathcal{K}$ restricted to F. We assume that each mesh facet $F \subset \partial \mathcal{K}$, with $\mathcal{K} = \mathcal{K}_\mathbf{x} \times \mathcal{K}_t \in \mathcal{T}_h$, is either

$$\text{a space-like facet} \quad \text{if } \vec{n}_{\mathcal{K}_\mathbf{x}}^F = 0, \quad \text{or} \quad \text{a time-like facet} \quad \text{if } \vec{n}_{\mathcal{K}_t}^F = 0.$$

Each time-like facet F can be written as $F := F_\mathbf{x} \times F_t$, where $F_\mathbf{x} \subset \partial \mathcal{K}_\mathbf{x}$, and $F_t \subset \mathcal{K}_t$. For each $\mathcal{K} = \mathcal{K}_\mathbf{x} \times \mathcal{K}_t \in \mathcal{T}_h$ with $\mathcal{K}_t = (a_t, b_t)$, we collect the space-like facets at time $t = a_t$ and the time-like facets of $\mathcal{K}$ into the sets $\mathcal{F}_\mathcal{K}^{\text{space}}$ and $\mathcal{F}_\mathcal{K}^{\text{time}}$, respectively.

Associated with each element $\mathcal{K} \in \mathcal{T}_h$, we define K as the $(d+1)$-dimensional closed polytope [1] containing $\mathcal{K}$ and whose boundary is given by the union of the facets in $\mathcal{F}_\mathcal{K}^{\text{space}} \cup \mathcal{F}_\mathcal{K}^{\text{time}}$. For any $F = F_\mathbf{x} \times F_t \in \mathcal{F}_\mathcal{K}^{\text{time}}$, we define

$$h_{F_\mathbf{x}} := \begin{cases} \min\{h_{\mathcal{K}_\mathbf{x}}, h_{\widetilde{\mathcal{K}}_\mathbf{x}}\} & \text{if } F = \mathcal{K} \cap \widetilde{\mathcal{K}} \text{ for some } \widetilde{\mathcal{K}} = \widetilde{\mathcal{K}}_\mathbf{x} \times \widetilde{\mathcal{K}}_t \in \mathcal{T}_h, \\ h_{\mathcal{K}_\mathbf{x}} & \text{if } F_\mathbf{x} \subset \partial \Omega. \end{cases}$$

We denote the sets of *space-like* and *time-like* facets of $\mathcal{T}_h$ as

$$\mathcal{F}_h^{\text{space}} := \bigcup_{\mathcal{K} \in \mathcal{T}_h} \mathcal{F}_\mathcal{K}^{\text{space}}, \qquad \mathcal{F}_h^{\text{time}} := \bigcup_{\mathcal{K} \in \mathcal{T}_h} \mathcal{F}_\mathcal{K}^{\text{time}}.$$

Hanging nodes $(1 + 1$ dimensional case), edges $(2 + 1$ dimensional case), and faces $(3 + 1$ dimensional case) are included within this structure of the mesh. In the absence of hanging facets, it is not necessary to distinguish between $\mathcal{K}$ and K, and in such a case all the definitions in this chapter coincide with those in Chapter 6. For each $F \in \mathcal{F}_h^{\text{time}}$, we fix a normal unit vector $\mathbf{n}_F \in \mathbb{R}^{d+1}$ of the form $(\mathbf{n}_{F_\mathbf{x}}, 0)$, $\mathbf{n}_{F_\mathbf{x}} \in \mathbb{R}^d$. In Figure 8.1, we illustrate the definitions in this section.

8.3   Local space–time virtual element spaces

In this section, we introduce local space–time virtual element spaces, extending the construction in Chapter 6 to the general prismatic meshes described in Section 8.2.

Introduce the scaling factors

$$\widetilde{c}_H^\mathcal{K} := h_{\mathcal{K}_t}, \qquad\qquad \widetilde{\lambda}^\mathcal{K} := h_{\mathcal{K}_\mathbf{x}}^2. \tag{8.1}$$

---

[1] We consider a d-dimensional "polytope" as defined in [16]: a geometrical figure bounded by a finite number of portions of lines (for $d = 2$), planes (for $d = 3$) and hyperplanes (for $d > 3$).
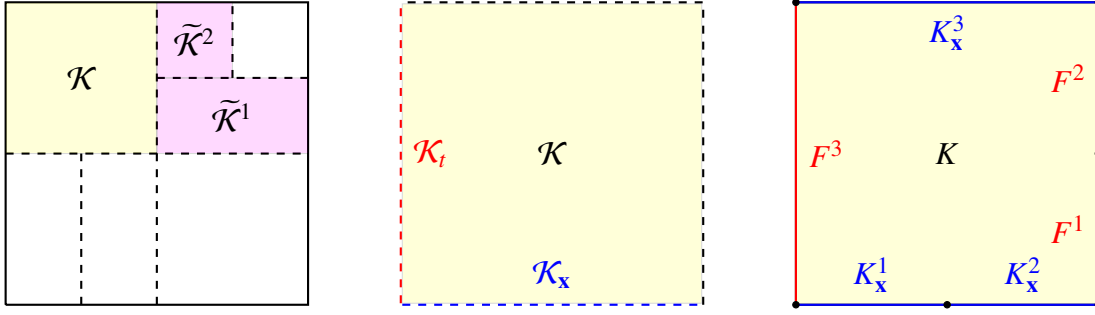
Figure 8.1 Example of a prismatic space–time mesh. **Left panel:** The prismatic partition $\mathcal{T}_h$ of the space–time domain. **Central panel:** Zoom of the element $\mathcal{K} \in \mathcal{T}_h$. **Right panel:** The associated closed hexagon K with three space-like facets $K_{\mathbf{x}}^1$, $K_{\mathbf{x}}^2$ and $K_{\mathbf{x}}^3$ and three time-like facets F^1, F^2 and F^3. By definition, $h_{F_{\mathbf{x}}^1} = \min\{h_{\mathcal{K}_{\mathbf{x}}}, h_{\widetilde{\mathcal{K}}_{\mathbf{x}}^1}\}$, $h_{F_{\mathbf{x}}^2} = \min\{h_{\mathcal{K}_{\mathbf{x}}}, h_{\widetilde{\mathcal{K}}_{\mathbf{x}}^2}\}$ and $h_{F_{\mathbf{x}}^3} = h_{\mathcal{K}_{\mathbf{x}}}$.

Given $p \in \mathbb{N}$, $\mathcal{K} = \mathcal{K}_{\mathbf{x}} \times \mathcal{K}_t \in \mathcal{T}_h$ and its associated K, we define the local space–time virtual element space as

$$V_h(K) := \left\{ v \in L^2(\mathcal{K}) \mid \widetilde{c}_H^{\mathcal{K}} \partial_t v_h - \widetilde{\lambda}^{\mathcal{K}} \Delta_{\mathbf{x}} v_h \in \mathbb{P}^{p-1}(\mathcal{K}) \,;\; v_h|_{\mathcal{K}_{\mathbf{x}}} \in \mathbb{P}^p(\mathcal{K}_{\mathbf{x}}) \,; \right.$$
$$\left. \vec{\mathbf{n}}_{F_{\mathbf{x}}} \cdot \nabla_{\mathbf{x}} v_h|_F \in \mathbb{P}^p(F) \;\forall F \in \mathcal{F}_{\mathcal{K}}^{\text{time}} \right\}. \tag{8.2}$$

Although functions in $V_h(K)$ are not known in closed form, it holds $\mathbb{P}^p(\mathcal{K}) \subset V_h(K)$.

Given an element $\mathcal{K} \in \mathcal{T}_h$ and any of its time-like facets F, let $\{m_\alpha^{\mathcal{K}}\}_{\alpha=1}^{r_{d+1,p-1}}$, $\{m_\beta^F\}_{\beta=1}^{r_{d,p}}$, and $\{m_\gamma^{\mathcal{K}_{\mathbf{x}}}\}_{\gamma=1}^{r_{d,p}}$ be given bases of $\mathbb{P}^{p-1}(\mathcal{K})$, $\mathbb{P}^p(F)$, and $\mathbb{P}^p(\mathcal{K}_{\mathbf{x}})$, respectively.

Introduce the following set of linear functionals:

- the *bulk* moments

$$\frac{1}{|\mathcal{K}|} \int_{\mathcal{K}} v_h \, m_\alpha^{\mathcal{K}} \mathrm{d}\mathbf{x} \; \mathrm{d}t, \qquad \alpha = 1, \ldots, r_{d+1,p-1}; \tag{8.3}$$

- for all *time-like* facets $F \in \mathcal{F}_{\mathcal{K}}^{\text{time}}$, the *time-like* moments

$$\frac{1}{|F|} \int_F v_h \, m_\beta^F \mathrm{d}S \, \mathrm{d}t, \qquad \beta = 1, \ldots, r_{d,p}; \tag{8.4}$$

- the *space-like* moments

$$\frac{1}{|\mathcal{K}_{\mathbf{x}}|} \int_{\mathcal{K}_{\mathbf{x}}} v_h(\cdot, t_{n-1}) m_\gamma^{\mathcal{K}_{\mathbf{x}}} \mathrm{d}\mathbf{x}, \qquad \gamma = 1, \ldots, r_{d,p}. \tag{8.5}$$

These linear functionals constitute a set of unisolvent degrees of freedom. To see this, it suffices to apply Lemma 6.3 to the case of the general prismatic-type elements introduced in Section 8.2. In Figure 8.2, we illustrate the degrees of freedom for the element $\mathcal{K} = \mathcal{K}_{\mathbf{x}} \times \mathcal{K}_t$ and its associated closed polytope K from Figure 8.1. Observe that, if $\mathcal{K}_t = (a_t, b_t)$, we define space-like moments on $\mathcal{K}_{\mathbf{x}} \times \{a_t\}$ and not on the space-like facets $K_{\mathbf{x}} \in \mathcal{F}_{\mathcal{K}}^{\text{space}}$. The latter will be used in the definition of the upwind terms.
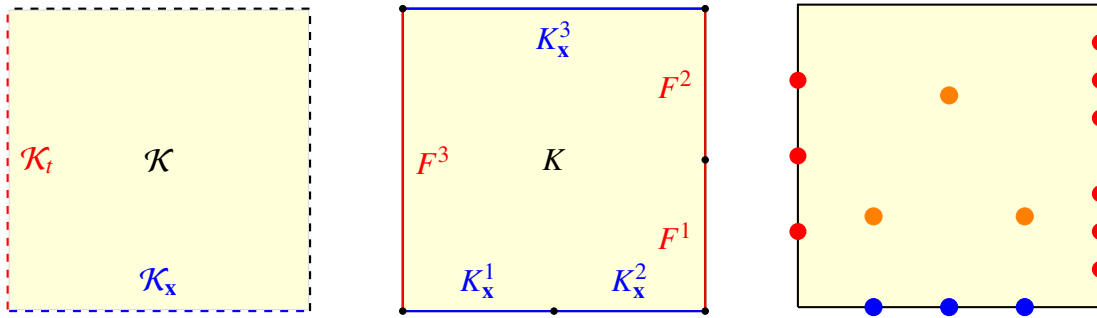
Figure 8.2 Example of degrees of freedom for $V_h(K)$ with $p = 2$ ($\mathcal{K}$ and K are those from Figure 8.1). **Left panel:** the element $\mathcal{K} \in \mathcal{T}_h$. **Central panel:** the associated closed hexagon K with three space-like facets $K_{\mathbf{x}}^1, K_{\mathbf{x}}^2$ and $K_{\mathbf{x}}^3$ and three time-like facets F^1, F^2 and F^3. **Right panel:** corresponding degrees of freedom; the **orange** dots denote the bulk moments; the **blue** dots denote the *space-like* moments; the **red** dots denote the *time-like* moments.

8.4 Global nonconforming space–time virtual element spaces

In this section, we design a global virtual element space Y_h, consisting of functions that are possibly discontinuous-in-time and nonconforming in space.

We define the Sobolev nonconforming space of order p associated with the mesh $\mathcal{T}_h$ as

$$\mathcal{N}\mathcal{C}_p(\mathcal{T}_h) := \left\{ v \in L^2(Q_T) \ \middle| \ v_{|\mathcal{K}} \in L^2(\mathcal{K}_t, H^1(\mathcal{K}_{\mathbf{x}})) \ \forall \mathcal{K} = \mathcal{K}_{\mathbf{x}} \times \mathcal{K}_t \in \mathcal{T}_h; \right.$$
$$\left. \int_F q_p^F [\![v]\!]_F \cdot \vec{\mathbf{n}}_F \, \mathrm{d}S = 0 \quad \forall q_p^F \in \mathbb{P}^p(F), \ \forall F \in \mathcal{F}_h^{\text{time}} \right\}. \tag{8.6}$$

We define the global virtual element space as follows:

$$Y_h := \left\{ v_h \in \mathcal{N}\mathcal{C}_p(\mathcal{T}_h) \ \middle| \ v_{h|\mathcal{K}} \in V_h(K) \ \forall \mathcal{K} \in \mathcal{T}_h \right\}.$$

8.5 The discrete bilinear forms

The definition of the discrete bilinear forms in Section 6.3.4 extends naturally to the general prismatic meshes in Section 8.2 with the same polynomial projections in Section 6.3.2, by suitably modifying the upwind-type terms. We recall that such terms are used to weakly enforce continuity in time on the discrete trial space. We define them as: for all space-like facets $K_{\mathbf{x}} \subset \Omega \times \{t^*\}$, $t^* \in [0, T]$,

$$\mathcal{U}^{K_{\mathbf{x}}}(u_h) := \begin{cases} c_H \Pi_p^\star u_{h|\mathcal{K}}(\cdot, t^*) & \text{if } t^* = 0, \\ c_H \left(\Pi_p^\star u_{h|\mathcal{K}^+}(\cdot, t^*) - \Pi_p^\star u_{h|\mathcal{K}^-}(\cdot, t^*) \right) & \text{if } t^* > 0, \end{cases} \tag{8.7}$$

where $K_{\mathbf{x}} \in \mathcal{F}_{\mathcal{K}}^{\text{space}}$ if $t^* = 0$, and $K_{\mathbf{x}} \in F_{\mathcal{K}^+}^{\text{space}}$ is a subset of $\partial \mathcal{K}^+ \cap \partial \mathcal{K}^-$ if $t^* > 0$.

Remark 8.1 (Inhomogeneous initial conditions). The case of nonzero initial conditions can be dealt with by modifying the definition of the upwind functional at time $t = 0$ as follows: for all space-like facets $K_{\mathbf{x}} \subset \Omega \times \{0\}$,

$$\mathcal{U}^{K_{\mathbf{x}}}(u_h) := c_H \left(\Pi_p^\star u_{h|\mathcal{K}}(\cdot, 0) - u_0 \right).$$

The computability of the upwind terms follows from the definition of the space-like moment degrees of freedom (8.5) and the computability of the projector $\Pi_p^\star$.

We extend the definition of the bilinear form $b_h(\cdot, \cdot)$ in (6.21)-(6.22), which is a discrete counterpart of the continuous bilinear form $b(\cdot, \cdot)$ in (1.4), to general prismatic meshes, as follows:

$$b_h(u_h, v_h) := \sum_{\mathcal{K} \in \mathcal{T}_h} \left[c_H (\partial_t \Pi_p^\star u_h, v_h)_{L^2(\mathcal{K})} + a_h^{\mathcal{K}}(u_h, v_h) \right] + \sum_{K_{\mathbf{x}} \in \mathcal{F}_h^{\text{space}}} \left(\mathcal{U}^{K_{\mathbf{x}}}(u_h), v_{h|\mathcal{K}^+} \right)_{L^2(K_{\mathbf{x}})}, \tag{8.8}$$

where $a_h^{\mathcal{K}}(\cdot, \cdot)$ is the bilinear form defined in (6.15), with the following computable stabilization satisfying (6.14)

$$S^{\mathcal{K}}(u_h, v_h) := p^2 h_{K_{\mathbf{x}}}^{-2} (\Pi_{p-1}^{0,\mathcal{K}} u_h, \Pi_{p-1}^{0,\mathcal{K}} v_h)_{L^2(\mathcal{K})} + \sum_{F \in \mathcal{F}_{\mathcal{K}}^{\text{time}}} p h_{K_{\mathbf{x}}}^{-1} (\Pi_p^{0,F} u_h, \Pi_p^{0,F} v_h)_{L^2(F)}$$
$$+ p h_{K_t} h_{K_{\mathbf{x}}}^{-2} (u_h, v_h)_{L^2(K_{\mathbf{x}})}. \tag{8.9}$$

8.6 The virtual element method

Assuming that f belongs to $L^2(Q_T)$, the space-time virtual element method for the approximation of solutions to (1.5) reads: find $u_h \in Y_h$ such that

$$b_h(u_h, v_h) = (f, \Pi_{p-1}^{0,Q_T} v_h)_{L^2(\Omega)} \qquad \forall v_h \in Y_h. \tag{8.10}$$

The discrete inf-sup argument in Section 6.4 that is used to prove the well-posedness and *a priori* error estimates can be extended to the case of general prismatic meshes introduced in Section 8.2, using the norms

$$\|v\|_{Y(\mathcal{T}_h)}^2 := \sum_{K \in \mathcal{T}_h} \|v\|_{Y(\mathcal{K})}^2, \qquad \|v\|_{X(\mathcal{T}_h)}^2 := \|v\|_{Y(\mathcal{T}_h)}^2 + \left\|\mathfrak{R}_h \Pi_p^\star v\right\|_{Y(\mathcal{T}_h)}^2$$

$$+ \frac{c_H}{2} \left(\sum_{K_{\mathbf{x}} \in \mathcal{F}_h^{\text{space}}} \left\|\mathcal{U}^{K_{\mathbf{x}}}(v)\right\|_{L^2(K_{\mathbf{x}})}^2 + \left\|\Pi_p^\star v(\cdot, T)\right\|_{L^2(\Omega)}^2 \right),$$

where the discrete Newton potential $\mathfrak{R}_h : \mathcal{S}^p(\mathcal{T}_h) \to Y_h$ is defined as follows: for all $v_h \in Y_h$,

$$a_h(\mathfrak{R}_h \phi_h, v_h) = c_H \left((\partial_t \phi_h, v_h)_{L^2(Q_T)} + \sum_{K_{\mathbf{x}} \in \mathcal{F}_h^{\text{space}}} \left(\mathcal{U}^{K_{\mathbf{x}}}(\phi_h), v_{h|\mathcal{K}^+} \right)_{L^2(K_{\mathbf{x}})} \right).$$

The well-posedness of the method follows as in Theorem 6.17 with these new definitions.

8.7 Space–time virtual element spaces with variable degrees of accuracy

Let $\mathcal{T}_h$ be a given space–time polytopic mesh consisting of $N_{\mathcal{T}_h}$ elements, and let $\mathbf{p} \in \mathbb{N}^{N_{\mathcal{T}_h}}$ be a given distribution of degrees of accuracy. More precisely, we sort the elements of $\mathcal{T}_h$ as $\{\mathcal{K}_j\}_{j=1}^{N_{\mathcal{T}_h}}$, and denote the degree of accuracy in each element $\mathcal{K}_j$ by p_j, $j = 1, \ldots, N_{\mathcal{T}_h}$.

Given the vector **p**, we fix the degrees of freedom associated with each element according to the following "*maximum strategy*":

- in each element $\mathcal{K}_j$, $j = 1, \ldots, N_{\mathcal{T}_h}$, we take bulk moments (8.3) up to degree $p_j - 1$;

- on each internal *time-like* facet F shared by two different elements $\mathcal{K}_j$ and $\mathcal{K}_\ell$ for given $j, \ell = 1, \ldots, N_{\mathcal{T}_h}$, we take *time-like* moments (8.4) up to degree $\max(p_j, p_\ell)$;

- on each *time-like* facet $F \subset \partial\Omega \times (0, T)$ on the boundary of the element $\mathcal{K}_j$ for a given $j = 1, \ldots, N_{\mathcal{T}_h}$, we take *time-like* moments (8.4) up to degree p_j;

- if $\mathcal{K}_j = \mathcal{K}_{\mathbf{x},j} \times \mathcal{K}_{t,j}$ for a given $j = 1, \ldots, N_{\mathcal{T}_h}$, we take the *space-like* moments (8.5) on $\mathcal{K}_{\mathbf{x},j}$ up to degree p_j.

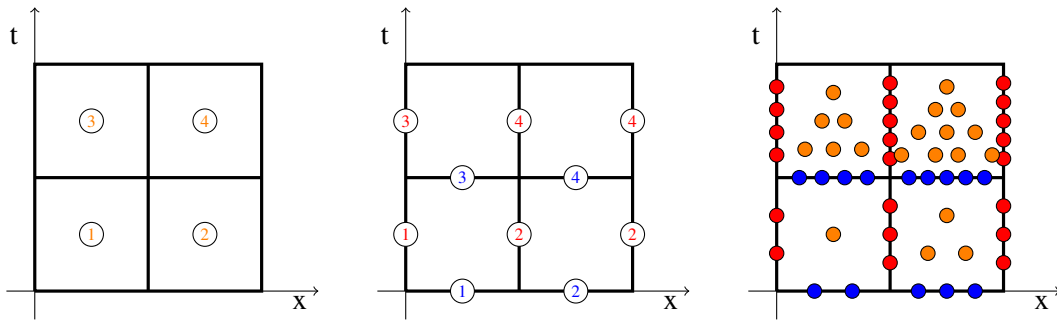To illustrate the "maximum strategy", we provide an example in Figure 8.3.

Figure 8.3 Example of the maximum strategy used in the definition of the local VEM space (8.11). **Left panel:** initial distribution of degrees of accuracy over the elements. **Central panel:** "polynomial degrees" on *space-like* (**blue**) and *time-like* (**red**) facets induced by the degrees of accuracy of the elements in the left panel. **Right panel:** corresponding degrees of freedom; the **orange** dots denote the bulk moments; the **blue** dots denote the *space-like* moments; the **red** dots denote the *time-like* moments.

We collect the *time-like* polynomial degrees in the vector $\mathbf{p}^{\text{time}} \in \mathbb{N}^{\text{card}(\mathcal{F}_h^{\text{time}})}$ and the *space-like* polynomial degrees in the vector $\mathbf{p}^{\text{space}} \in \mathbb{N}^{N_{\mathcal{T}_h}}$. We order the *time-like* facets in $\mathcal{F}_h$ as $\{F_j\}_{j=1}^{\text{card}(\mathcal{F}_h^{\text{time}})}$.

The corresponding local space on K_j reads

$$V_h(K_j) := \left\{ v \in L^2(\mathcal{K}_j) \mid \widetilde{c}_H^{\mathcal{K}} \partial_t v_h - \widetilde{\lambda}^{\mathcal{K}} \Delta_{\mathbf{x}} v_h \in \mathbb{P}^{p_j - 1}(\mathcal{K}_j); \; v_{h|\mathcal{K}_{\mathbf{x},j}} \in \mathbb{P}^{p_j^{\text{space}}}(\mathcal{K}_{\mathbf{x},j}); \right.$$
$$\left. \vec{\mathbf{n}}_{F_k,\mathbf{x}} \cdot \nabla_{\mathbf{x}} v_{h|F_k} \in \mathbb{P}^{p_k^{\text{time}}}(F_k) \; \forall F_k \in \mathcal{F}_{K_j}^{\text{time}} \right\}, \tag{8.11}$$

where we recall that $\widetilde{c}_H^{\mathcal{K}}$ and $\widetilde{\lambda}^{\mathcal{K}}$ are given in (8.1).

The global space Y_h is constructed by the nonconforming coupling of the *time-like* degrees of freedom (8.4). An immediate consequence of this "*maximum strategy*" is that $\mathbb{P}^{p_j}(\mathcal{K}_j)$ is contained in $V_h(K_j)$.

Chapter 9

Numerical aspects

Contents

9.1   Introduction

In this chapter, we discuss flagging strategies to handle the space–time mesh data structure, which improve the performance of the virtual element method (8.10). More precisely, we present *i)* a time-slab flagging that allows us to identify a time-slab decomposition of the space–time mesh, and is used to decompose the global linear system as a sequence of smaller problems associated with each time-slab; *ii)* an element-topology flagging, which allows for a reduction in the computational cost of assembling the linear system, by identifying elements with the same topology (i.e., equivalent up to scalings in **x** or t and translations, and with the same distribution of space-like and time-like facets) and using a set of reference elements to compute the projections and evaluate the discrete operators. We investigate numerically the hp-version of the method and demonstrate the expected exponential convergence in terms of suitable roots of the number of degrees of freedom for singular solutions.

A residual-type error indicator is introduced, and an h-adaptive refinement procedure based on it is tested. This error indicator appears to be reliable and efficient for the error $\mathcal{E}^Y$ in (9.1). However, the main obstacle to the proof of reliability and efficiency of residual error estimators for space–time methods based on the Petrov-Galerkin formulation (1.5) is that, starting from an inf-sup stability bound as that in Proposition 6.16, the use of standard Clément interpolants leads to bounds involving $H^1(Q_T)$-type norms, which are stronger than the Y-type norm in the denominator of the right-hand side of (6.28). Therefore, we only investigate numerically the overall virtual element adaptive procedure. The results obtained for certain singular solutions show that it outperforms the corresponding one in [94, 95] for a continuous finite element method.

To the best of our knowledge, just a few works on hp space–time methods for parabolic problems are available in the literature. In [14], Cangiani *et al.*, proposed an hp space–time interior-penalty discontinuous Galerkin method for prismatic meshes. An hp finite element version of the space–time isogeometrical analysis method in [66] was presented by Devaud and Schwab in [23]. In [22], Devaud proposed an hp Petrov-Galerkin formulation involving fractional Riemann-Liouville derivatives in time of order 1/2. For tensor-product meshes, some hp methods can be interpreted either as time-stepping or as space–time methods, see e.g., [6, 91, 86]. An hp-version of the conforming finite element discretization in [4, 93] of the variational formulation (1.5) has been used in [62] to model laser powder bed fusion.

The most representative results on residual error estimators for space–time methods are those for FOSLS finite element methods [35, 37], which naturally provide reliable and efficient error estimators at the expense of including an additional vector-valued flux variable. A residual-type error indicator for the conforming finite element method in [93] was introduced in [94, 95], but no proof of efficiency and reliability was provided. Some recent efforts have been made to get *a posteriori* error estimates for DG time discretizations of parabolic problems, see [38, 39]; however, the extension of such results to space–time methods on general prismatic meshes is a nontrivial open issue. Alternative space–time adaptive schemes for parabolic problems include parabolic duality methods [29, 30], functional-type estimators [88, 64, 65, 67, 68], and flux reconstruction methods [33, 32].

Structure of the chapter: We describe refinements procedures in Section 9.2; a time-slab flagging strategy to split method (8.10) into smaller linear systems in Section 9.3; an element-topology flagging strategy in Section 9.4. In Section 9.5, we assess the convergence of the h- and hp-versions of the method on some singular solutions. In Section 9.6, a computable residual-type error indicator is defined and used to lead adaptive mesh refinements.

9.2 Mesh refinements

We describe a procedure to refine general prismatic space–time elements. For the sake of presentation, we consider the $(1 + 1)$ dimensional case only; the extension to to any spatial dimensions follows with a minor effort.

As discussed in Section 8.2, each mesh consists of rectangular elements $\mathcal{K}$, with boundary given by the union of four (two space-like, two time-like) straight segments. Each straight segment may be the union of aligned edges of the element, which constitute the boundary of an associated closed polytope K. Regardless of the number of existing hanging nodes from previous refinements, a given K to be refined is split into four siblings by connecting the centroid of K with the midpoints of each straight segment of the boundary.

In Figure 9.1, we show an example of an element refinement. By this procedure, at most new five nodes are generated, fewer in presence of previously generated hanging nodes.

9.3 Time-slab flagging strategy

We present a flagging strategy that allows for the decomposition of the linear system stemming from (8.10) into smaller linear systems. To this aim, we assume that the first mesh of a sequence $\{\mathcal{T}_h\}$ is a "tensor-product-in-time" mesh, which can be arranged into time-slabs. On

Figure 9.1 Element refinement strategy. We connect the centroid of the element with the midpoints of each straight segment of the boundary, regardless of the presence of previously generated hanging nodes. The **red** dots denote the nodes of K; the **blue** dots denote the newly created nodes.

this mesh, method (8.10) can be assembled and solved sequentially with respect to the time-slabs.

In Algorithm 1, we explain how to identify a time-slab partition of a given space–time mesh.

Algorithm 1: Time-slab flagging

1 **Input** a prismatic space–time mesh $\mathcal{T}_h$.
2 **Get** a partition $0 =: t_0 < \cdots < t_\ell := T$ of the time interval $[0, T]$, so that there exists at least one element $\mathcal{K} = \mathcal{K}_{\mathbf{x}} \times \mathcal{K}_t \in \mathcal{T}_h$ with $\mathcal{K}_{\mathbf{x}} \subset \Omega \times \{t_j\}$ for each $j = 0, \ldots, \ell - 1$, and $\mathcal{K}_t = (t_i, t_j)$ for given $i, j = 0, \ldots, \ell, i < j$.
3 **for** $j = 1$ **to** $\ell - 1$ **do**
4   **if** *all the elements $\mathcal{K} \in \mathcal{T}_h$ are either below or above t_j* **then**
5   **set** a new time-slab flag for all the elements that lie below t_j but have not been previously flagged.

Based on the new flagging, at each refinement step, we can assemble and solve method (8.10) sequentially on the newly created time-slabs. In Figure 9.2, we illustrate this flagging procedure with an example.
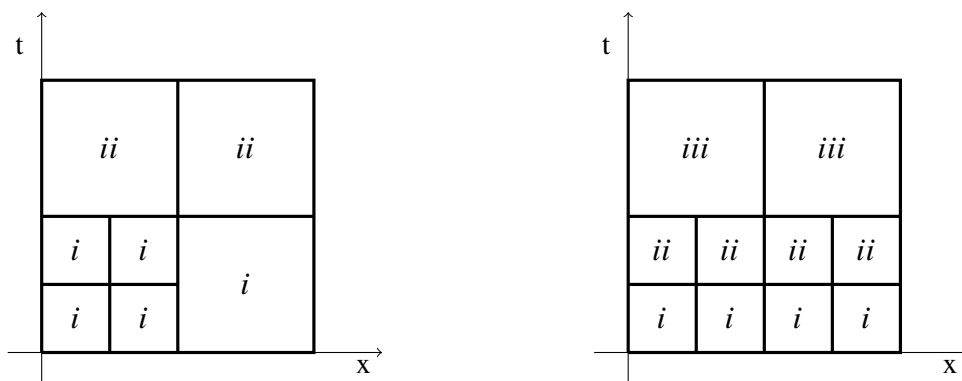

Figure 9.2 Left panel: we start with a given space–time mesh with prescribed time-slab structure and flagging. **Right panel:** time-slab flagging based on the proposed strategy after one refinement step.

9.4 Element-topology flagging strategy

A common issue in the implementation of polytopic methods is the lack of a reference element; this makes the computation of the local matrices more expensive than for methods based on simplicial or Cartesian meshes.

Assume that the elements of a given mesh $\mathcal{T}_h$ can be grouped into a uniformly finite number of classes of equivalence (up to scalings in $\mathbf{x}$ or t and translations), i.e., there exists a set $C = \{\hat{K}\}$ of "reference elements" with card(C) bounded uniformly for all meshes such that the associated closed polytope K of each element $\mathcal{K} \in \mathcal{T}_h$ can be obtained through an affine map $\mathbb{T}_K : \hat{K} \to K$, for some $\hat{K} \in C$, with $\mathbb{T}_K$ involving only scalings in $\mathbf{x}$ or t and translations (but not rotations), and preserving the distribution of space-like and time-like facets. In this case, local matrices need to be computed only for the "reference elements", thus speeding up the assembling of the final system. In adaptive mesh refinements, new "reference elements" may be required. For this reason, we introduce a flag associated with the element topology that identifies the corresponding "reference element".

For instance, in Figure 9.3 (left panel), we only have one "reference element" (a space–time square); in Figure 9.3 (central panel), we have two reference elements (a space–time square and a space–time pentagon with two time-like facets on the left); in Figure 9.3 (right panel), we have two reference elements (a space–time square and a space–time pentagon with two space-like facets at the bottom). The scaling factors $\widetilde{c}_H^{\mathcal{K}}$ and $\widetilde{\lambda}^{\mathcal{K}}$ in (8.1), which are used in the definition of the local VEM space $V_h(K)$ in (8.2), are instrumental to preserve the discrete space under dilations in $\mathbf{x}$ or t, see Remark 6.2. Inside each element, we denote the associated "topology-flag" with a natural number.
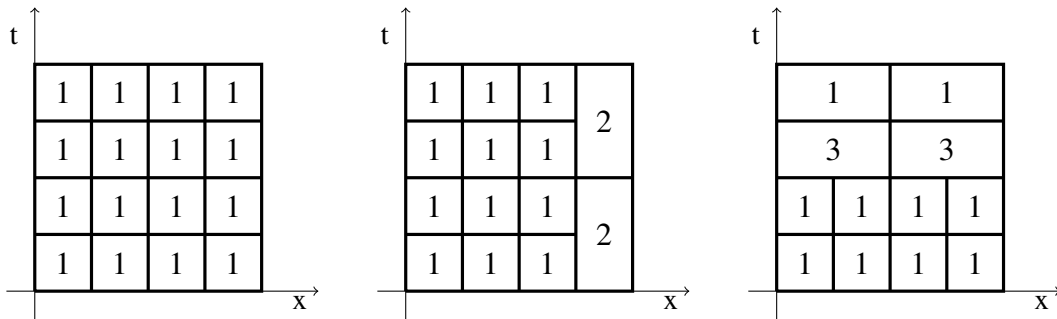

Figure 9.3 Left panel: a mesh consisting of equivalent elements. **Central panel:** a mesh with nonmatching time-like facets. **Right panel:** a mesh with nonmatching space-like facets.

9.5 Numerical investigation: convergence tests

We assess the convergence of the h- and hp-versions of the virtual element method in (8.10). Since the virtual element solution u_h is not known in closed form and the error in the $X(\mathcal{T}_h)$ norm is not computable, we report the following error quantities, which reduce to those defined in (7.14) for tensor product space–time meshes: given u_h the solution to (8.10),

$$\mathcal{E}^Y := \left\| u - \Pi_p^N u_h \right\|_{Y(\mathcal{T}_h)}, \qquad \mathcal{E}^N := \left\| \Pi_p^N (\mathfrak{R}_h \Pi_p^\star (u - u_h)) \right\|_{Y(\mathcal{T}_h)},$$

$$(\mathcal{E}^U)^2 := \frac{c_H}{2} \left(\left\| \Pi_p^\star (u - u_h)(\cdot, T) \right\|_{L^2(\Omega)}^2 + \sum_{K_{\mathbf{x}} \in \mathcal{F}_h^{\text{space}}} \left\| \mathcal{U}^{K_{\mathbf{x}}} (\Pi_p^\star (u - u_h)) \right\|_{L^2(K_{\mathbf{x}})}^2 \right), \qquad (9.1)$$

$$(\mathcal{E}^X)^2 := (\mathcal{E}^Y)^2 + (\mathcal{E}^N)^2 + (\mathcal{E}^U)^2.$$

9.5.1   Test cases

We recall the test cases in Section 7.5. The right-hand side f, and the boundary and initial conditions are computed accordingly to the exact solutions below.

Test case 1.   On the space–time domain $Q_T = (0, 1) \times (0, 1)$, we define the function

$$u_1(x, t) := \sin(t) \sin(\pi x). \qquad (9.2)$$

Test case 2.   On the space–time domain $Q_T = (0, 1) \times (0, 0.1)$, for $\alpha > \frac{1}{2}$, we define the function

$$u_2(x, t) := t^\alpha \sin(\pi x), \qquad (9.3)$$

which belongs to $H^{\alpha+1/2-\varepsilon}(0, 1; C^\infty(\Omega))$, $\varepsilon > 0$.

Test case 3.   On the space–time domain $Q_T = (0, 1) \times (0, 1)$, we define the function

$$u_3(x, t) := \sum_{n=0}^{\infty} \frac{4}{(2n + 1)\pi} \sin((2n + 1)\pi\, x) \exp(-(2n + 1)^2 \pi^2 t), \qquad (9.4)$$

which is the Fourier series of the solution to (1.2) with zero source term f, initial condition $u_0 = 1$, and homogeneous Dirichlet boundary conditions g. In the numerical experiments, the series in (9.3) is truncated at $n = 250$. The function u_3 belongs to $H^s(0, 1; H_0^1(0, 1))$ for any $s < 1/4$ and to $H^1(0, 1; H^{-1}(\Omega)) \cap L^2(0, 1; H_0^1(\Omega)) \backslash H^1(Q_T)$, see [91]. In particular, it is singular at the interface between the (incompatible) initial and boundary conditions, and has a fast decay behaviour close to $t = 0$.

9.5.2 h- and hp-versions for singular solutions

The performance of the h-version of the method on smooth solutions is already investigated in Section 7.5. Test case 1 is used in Section 9.6 below.

We focus on the convergence of the h- and hp-versions of the method for singular solutions. To that aim, we consider the test cases 2 and 3. Notably, we want to assess exponential convergence in terms of the cubic root of the number of degrees of freedom for the hp-version on certain geometrically refined space–time meshes.

First, we consider the test case 2. For the h-version of the method, we consider uniform degree of accuracy $p = 1$ and a sequence of uniform Cartesian space–time meshes with $h_t = 2h_x = 0.2 \times 2^{-i}$, $i = 1, \ldots, 6$. For the hp-version of the method, we proceed similarly as in [14, Example 2]: we fix a partition of the spatial domain with $h_x = 0.05$ and consider a sequence of temporal meshes geometrically graded towards $t = 0$ with grading factor $\sigma_t = 0.1$. In addition, the degree of accuracy p is increased by 1 from one time slab to the next one. In Figure 9.4, we depict the first three meshes with varying degrees of accuracy.
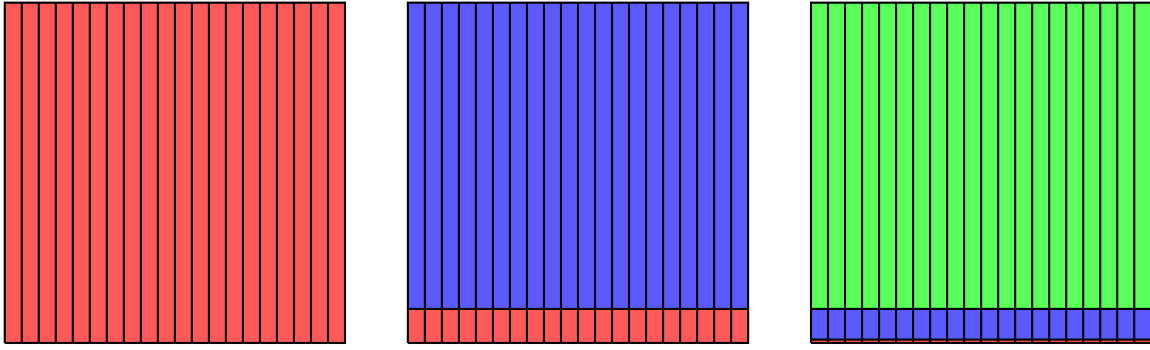
Figure 9.4 First three meshes employed in the hp refinements for test case 2 with exact solution u_2 in (9.3). The space–time domain is $Q_T = (0, 1) \times (0, 0.1)$. For a better understanding of the figure, we scale the t-coordinates by 10. In colours, we represent the local degrees of accuracy: **red:** $p = 1$; **blue:** $p = 2$; **green:** $p = 3$.

In Figures 9.5 and 9.6, we show the errors in (9.1) for $\alpha = 0.55$ and $\alpha = 0.75$ in *semilogy* scale, respectively. Exponential convergence in terms of the cubic root of the number of degrees of freedom is observed for the hp-version of the method and both values of α. In all cases, the h-version is outperformed and displays only the expected algebraic decay of the error.
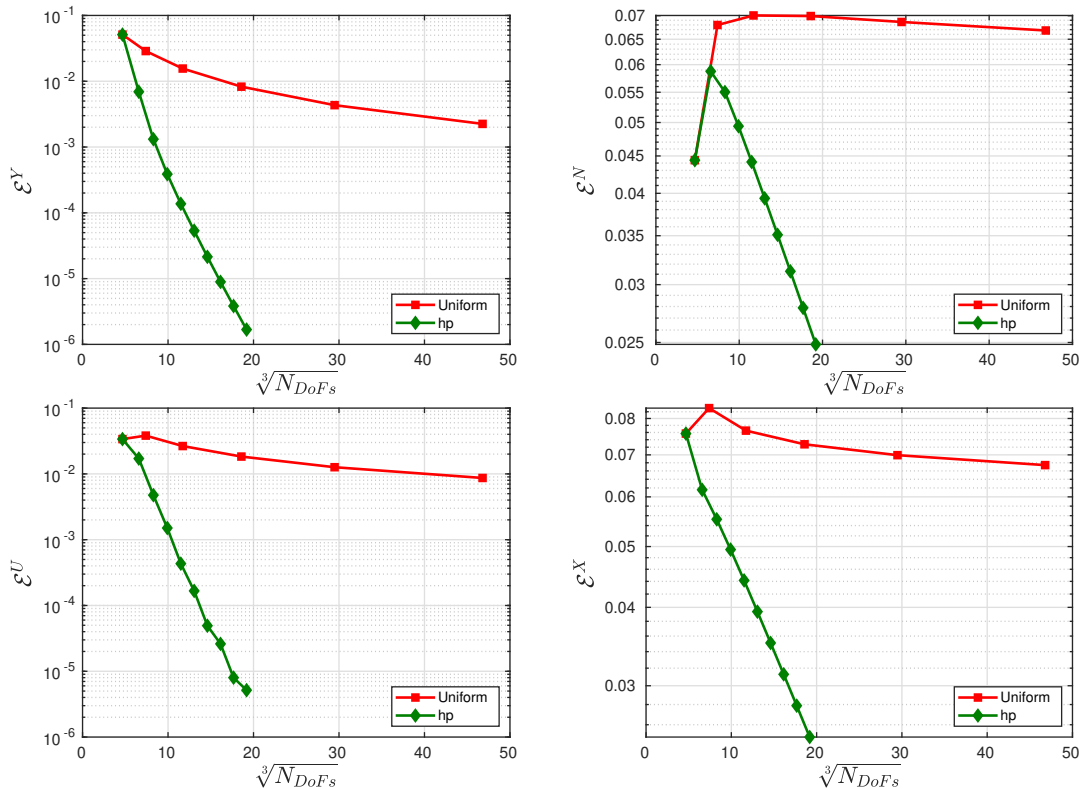

Figure 9.5 Convergence of the errors in (9.1) for the h- and hp-versions of the method. We consider test case 2 with exact solution u_2 in (9.3), $\alpha = 0.55$.

Next, we focus on test case 3. For the h-version of the method, we consider uniform degree of accuracy $p = 1$, and a sequence of uniform Cartesian space–time meshes with $h_x = 0.5 h_t =$
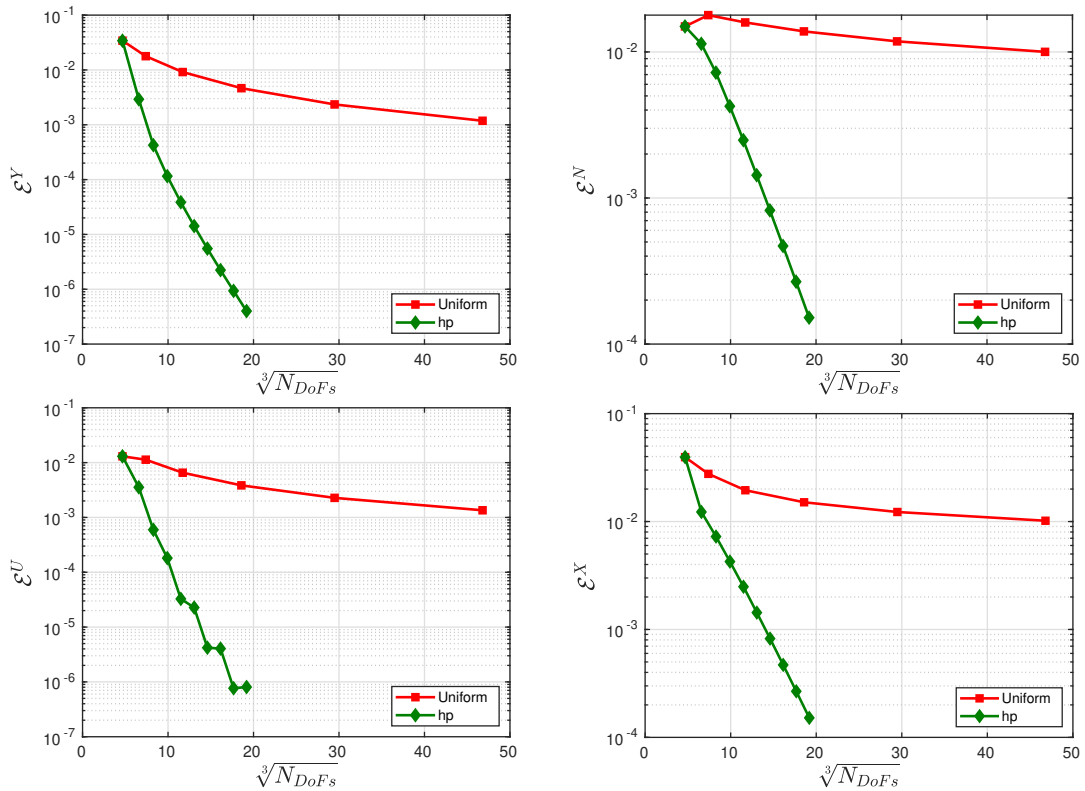
Figure 9.6 Convergence of the errors in (9.1) for the h- and hp-versions of the method. We consider test case 2 with exact solution in (9.3), $\alpha = 0.75$.

2^{-i}, $i = 1, \ldots, 8$. For the hp-version of the method, we proceed similarly as in [91, §7.4]: we consider a sequence of space–time meshes geometrically graded towards $x = 0$, $x = 1$ and $t = 0$ with grading factors $\sigma_x = \sigma_t = 0.25$. In Figure 9.7, we depict the first three meshes with varying degrees of accuracy. In Figure 9.8, we show the errors in (9.1) in *semilogy* scale.
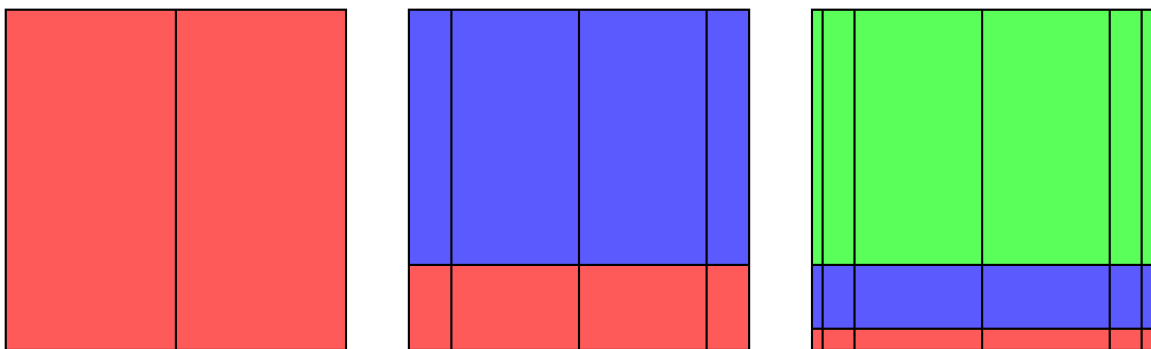

Figure 9.7 First three meshes employed in the hp refinements for test case 2 with exact solution u_2 in (9.3). The space–time domain is $Q_T = (0, 1) \times (0, 1)$. In colours, we represent the local degrees of accuracy: **red:** $p = 1$; **blue:** $p = 2$; **green:** $p = 3$.

Exponential convergence in terms of the cubic root of the number of degrees of freedom is observed for the hp-version of the method; as before, only algebraic convergence is observed for the h-version.
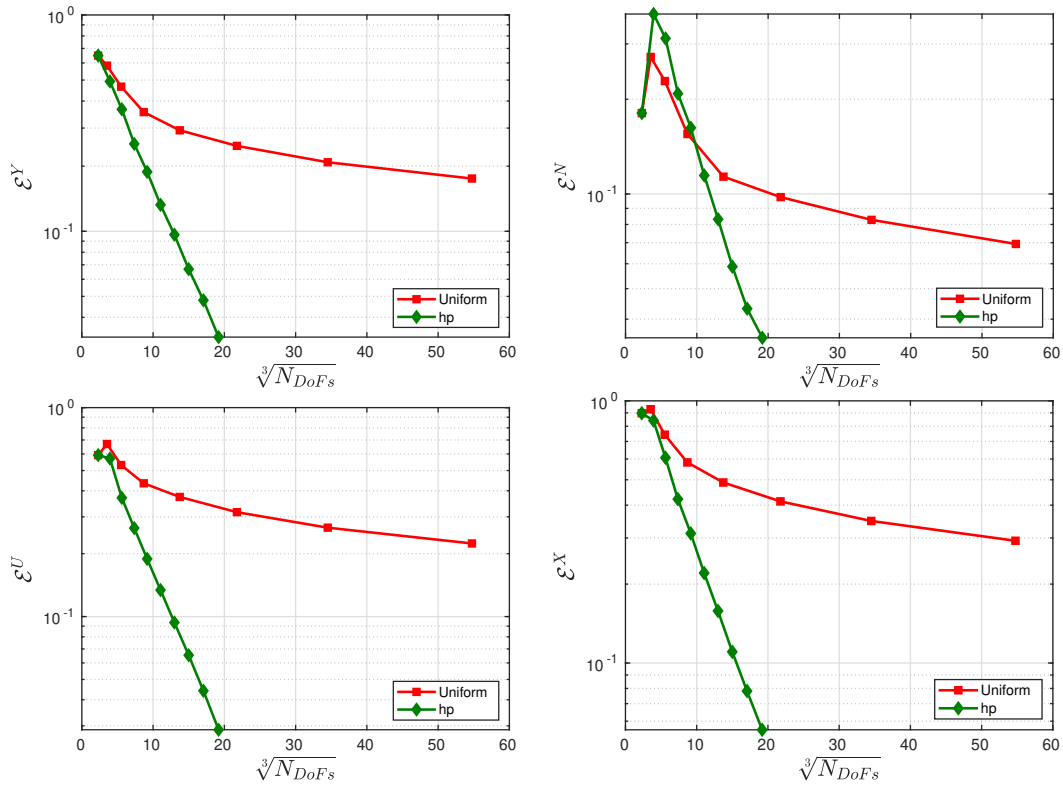
Figure 9.8 Convergence of the errors in (9.1) for the h- and hp-versions of the method. We consider test case 3 with exact solution in (9.4).

9.6 Numerical investigation: an adaptive procedure

We consider a standard adaptive algorithm of the form

$$\boxed{\textbf{SOLVE}} \implies \boxed{\textbf{ESTIMATE}} \implies \boxed{\textbf{MARK}} \implies \boxed{\textbf{REFINE}} \tag{9.5}$$

We base the **MARK** step on a Dörfler marking strategy [25] with parameter θ that we shall specify at each occurrence. The **REFINE** step is based on the h-refinement procedure described in Section 9.2, and local p-refinements are not considered. For the **ESTIMATE** step, we use the following local, computable residual-type error indicator: given an element $\mathcal{K} \in \mathcal{T}_h$,

$$\eta_\mathcal{K}^2 := \sum_{i=1}^{5} \eta_{\mathcal{K},i}^2,$$

where

$$\eta_{\mathcal{K},1}^2 := \frac{h_{\mathcal{K}_\mathbf{x}}^2}{\lambda p^2} \left\| f + \lambda \Delta_\mathbf{x} \Pi_p^\mathcal{N} u_h - c_H \partial_t \Pi_p^\star u_h \right\|_{L^2(\mathcal{K})}^2, \quad \eta_{\mathcal{K},2}^2 := \frac{1}{2\lambda} \sum_{\substack{F \in \mathcal{F}_\mathcal{K}^{\text{time}}, \\ F \not\subset \partial\Omega \times (0,T)}} \frac{h_{F_\mathbf{x}}}{p} \left\| \lambda [\![\nabla_\mathbf{x} \Pi_p^\mathcal{N} u_h]\!] \right\|_{L^2(F)}^2,$$

$$\eta_{\mathcal{K},3}^2 := \lambda \sum_{\substack{F \in \mathcal{F}_\mathcal{K}^{\text{time}}, \\ F \subset \partial\Omega \times (0,T)}} \frac{p}{h_{F_\mathbf{x}}} \left\| \Pi_p^\mathcal{N} u_h \right\|_{L^2(F)}^2 + \frac{\lambda}{2} \sum_{\substack{F \in \mathcal{F}_\mathcal{K}^{\text{time}}, \\ F \not\subset \partial\Omega \times (0,T)}} \frac{p}{h_{F_\mathbf{x}}} \left\| [\![\Pi_p^\mathcal{N} u_h]\!] \right\|_{L^2(F)}^2, \tag{9.6}$$

$$\eta_{\mathcal{K},4}^2 := \sum_{K_{\mathbf{x}} \in \mathcal{F}_{\mathcal{K}}^{\text{space}}} c_H^{-1} \left\| \mathcal{U}^{K_{\mathbf{x}}}(u_h) \right\|_{L^2(K_{\mathbf{x}})}^2, \qquad \eta_{\mathcal{K},5}^2 := \lambda S^{\mathcal{K}}((I - \Pi_p^{\mathcal{N}})u_h, (I - \Pi_p^{\mathcal{N}})u_h).$$

The above local residual-type error indicator consists of five terms: $\eta_{\mathcal{K},1}$ is the internal residual of the projected discrete solution; $\eta_{\mathcal{K},2}$ is the boundary residual involving the normal trace of the gradient of the projected discrete solution; $\eta_{\mathcal{K},3}$ is a term due to the nonconformity in space involving the jump of traces on time-like facets; $\eta_{\mathcal{K},4}$ is related to the upwind terms in the scheme; $\eta_{\mathcal{K},5}$ is a correction term due to the virtual element stabilization of the method.

The global error indicator and its parts read

$$\eta^2 := \sum_{i=1}^5 \eta_i^2, \qquad \eta_i^2 := \sum_{\mathcal{K} \in \mathcal{T}_h} \eta_{\mathcal{K},i}^2. \tag{9.7}$$

9.6.1 Assessment of the reliability and efficiency of the error indicator

We test the performance of the proposed error indicator η in (9.7). To this aim, we introduce the effectivity index related to the computable error $\mathcal{E}^Y$:

$$\text{effectivity index} := \frac{\eta}{\mathcal{E}^Y}. \tag{9.8}$$

In Figure 9.9 (left panel), we assess numerically the reliability and efficiency of η for the test case 1 with exact solution u_1 in (9.2) under uniform mesh refinements, starting from a uniform space–time Cartesian mesh with $h_x = h_t = 0.1$. The effectivity indices for $p = 1, 2, 3$, tend to constant values, whence the error indicator η appears to be efficient and reliable with respect to the error $\mathcal{E}^Y$.

In Figure 9.9 (right panel), we show all the errors in (9.1) and the five terms appearing in the error indicator (9.7) for degree of accuracy $p = 1$. All the contributions η_i for $i \in \{1, 2, 3, 5\}$ decay with the same order $O(N_{DoFs}^{-\frac{1}{2}})$, as the error $\mathcal{E}^Y$. The term η_4 decays faster with order $O(N_{DoFs}^{-\frac{3}{4}})$, but this does not affect the decay of the total error indicator η. The error indicator η decays with order $O(N_{DoFs}^{-\frac{1}{2}})$, i.e., slower than that for the error $\mathcal{E}^N$ in (9.1), which decays with order $O(N_{DoFs}^{-1})$. This suggests that the error indicator η is not efficient with respect to the error $\mathcal{E}^N$.

Next, we assess the efficiency and effectivity of the error indicator η also for the test cases 2 and 3 with singular solutions u_2 ($\alpha = 0.55$ and 0.75) in (9.3) and u_3 in (9.4) on the space–time domains $Q_T = (0, 1) \times (0, 1)$ and $Q_T = (0, 1) \times (0, 0.1)$, respectively. In Figures 9.10–9.12, we show the decay of the errors in (9.1) and of the five terms appearing in the error indicator under uniform mesh refinements, starting with Cartesian meshes with $h_x = 10h_t = 0.1$ (for the test case 2) and $h_x = h_t = 0.1$ (for the test case 3). In all cases, the effectivity indices shown on the left panels tend to constant values, which suggests that the error indicator (9.7) is efficient and reliable with respect to the error $\mathcal{E}^Y$ also for singular solutions.

For the test case 2, the errors $\mathcal{E}^Y$ and $\mathcal{E}^N$ decay with orders $O(N_{DoFs}^{-\frac{1}{2}(\alpha+\frac{1}{2})})$ and $O(N_{DoFs}^{-\frac{1}{2}(\alpha-\frac{1}{2})})$, respectively; see Figures 9.10 and 9.11. The error indicator η decays with the same order as that of the error $\mathcal{E}^Y$, while it it not reliable with respect to the error $\mathcal{E}^N$.
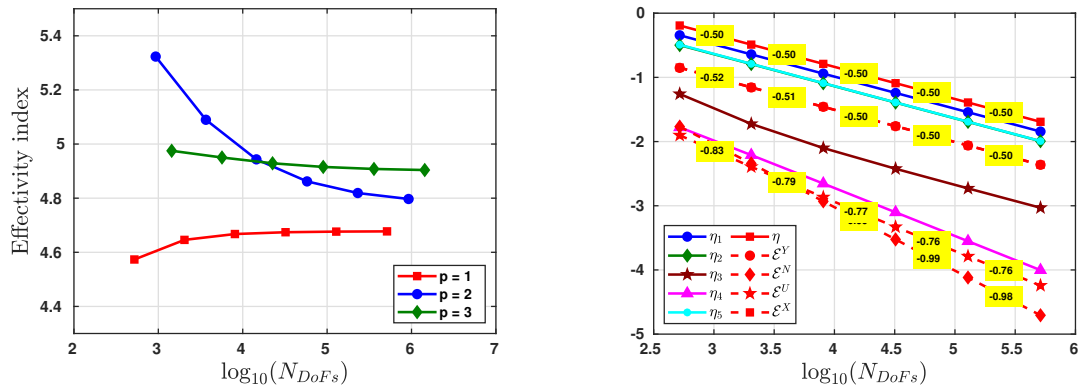
Figure 9.9 The test case 1 with exact solution u_1 in (9.2). **Left panel:** Effectivity index. **Right panel:** Comparison of the errors in (9.1) with the terms appearing in the error indicator (9.7) for $p = 1$.

For the test case 3, the error indicator η and all the errors in (9.1) decay with the same order, namely $O(N_{DoFs}^{-\frac{s}{2}})$ with $s = \frac{1}{4}$, see Figure 9.12. This is in agreement with the regularity of the exact solution u_3 which belongs to $H^s(0, 1; H_0^1(0, 1))$.

In summary, the above experiments seem to indicate that the error indicator η is reliable and efficient for the error $\mathcal{E}^Y$ but not for the error $\mathcal{E}^N$.
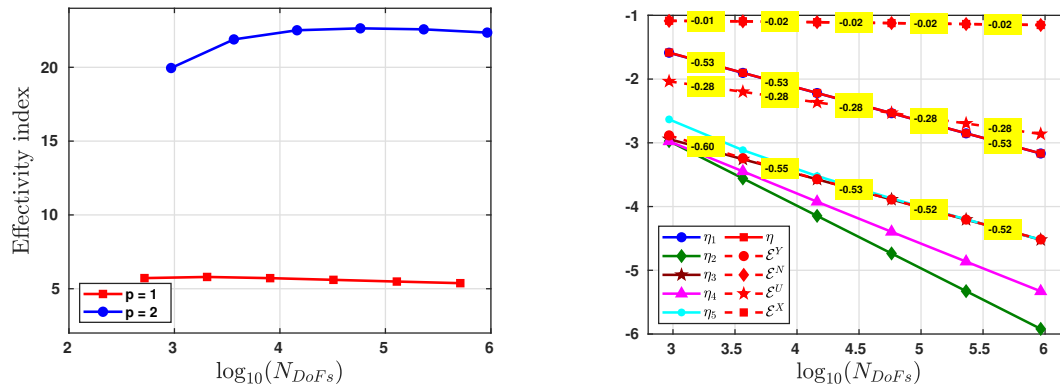

Figure 9.10 The test case 2 with exact solution u_2 in (9.3) and $\alpha = 0.55$. **Left panel:** Effectivity index. **Right panel:** Comparison of the errors in (9.1) with the terms appearing in the error indicator (9.7) for $p = 2$

.

9.6.2 Adaptive mesh refinements

We test the performance of method (8.10) under adaptive mesh refinements as described in (9.5). We consider the test cases 2 and 3 with exact solutions u_2 ($\alpha = 0.55$) in (9.3) and u_3 in (9.4), respectively. The marking step is dictated by the error indicator in (9.7).

In order to compare our results with those obtained with an adaptive procedure for the continuous FEM (7.15), we recall the residual-type error indicator introduced in [94, 95]:
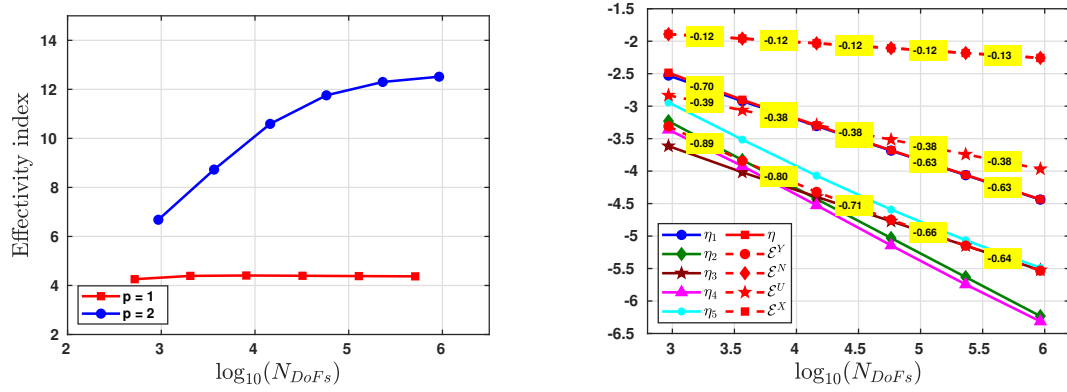
Figure 9.11 The test case 2 with exact solution u_2 in (9.3) and $\alpha = 0.75$. **Left panel:** Effectivity index. **Right panel:** Comparison of the errors in (9.1) with the terms appearing in the error indicator (9.7) for $p = 2$

.

Figure 9.12 The test case 3 with exact solution u_3 in (9.4). **Left panel:** Effectivity index. **Right panel:** Comparison of the errors in (9.1) with the terms appearing in the error indicator (9.7) for $p = 2$.
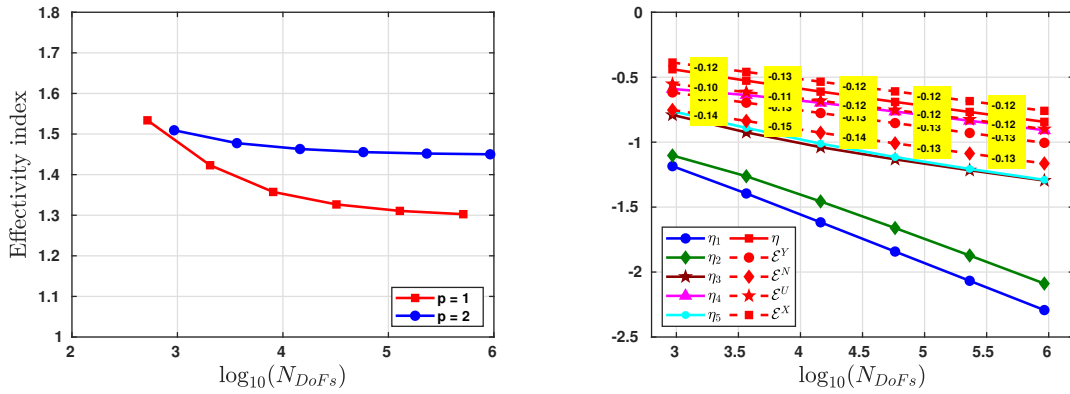
$$\widetilde{\eta}^2 := \sum_{i=1}^{2} \widetilde{\eta}_i^2, \qquad \widetilde{\eta}_i^2 = \sum_{\mathcal{K} \in \mathcal{T}_h} \widetilde{\eta}_{\mathcal{K},i}^2,$$

$$\widetilde{\eta}_{\mathcal{K},1}^2 := \frac{h_{\mathcal{K}}^2}{p^2} \|f + \lambda \Delta_{\mathbf{x}} \widetilde{u}_h - c_H \partial_t \widetilde{u}_h\|_{L^2(\mathcal{K})}^2, \qquad \widetilde{\eta}_{\mathcal{K},2}^2 := \frac{1}{2} \sum_{F \in \mathcal{F}_{\mathcal{K}}^{\text{time}}} \frac{h_F}{p} \|\lambda [\![\nabla_{\mathbf{x}} \widetilde{u}_h]\!]\|_{L^2(F)}^2 . \tag{9.9}$$

For the VEM, we start with a mesh with 1 element; for the continuous FEM, we start with a structured simplicial mesh with 2 elements.

In Figure 9.13, we show the errors of both methods under uniform and adaptive mesh refinements for the test case 2 with $\alpha = 0.55$. Uniform and adaptive mesh refinements for the VEM in (8.10) lead to higher convergence rates that those for the continuous FEM in (7.15). For the Dörfler marking strategy, we set $\theta = 0.99$ for the VEM and $\theta = 0.9$ for the FEM.
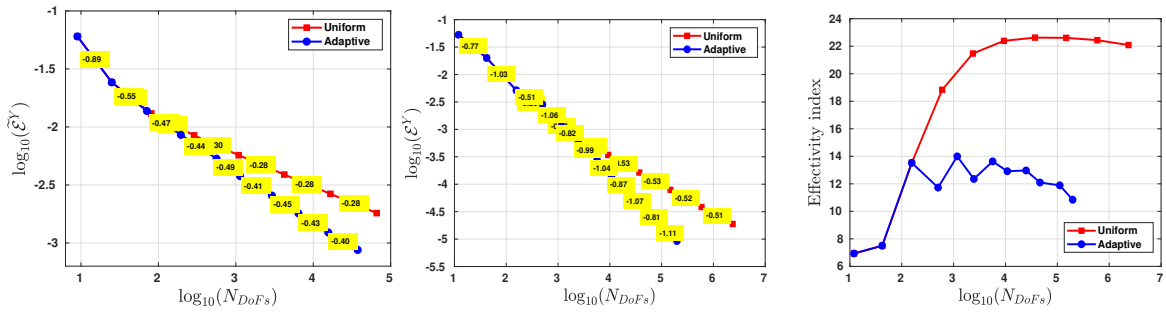
Figure 9.13 The test case 2 with exact solution u_2 in (9.3) and $\alpha = 0.55$. **Left panel:** Error $\widetilde{\mathcal{E}}^Y$ for the continuous FEM with $p = 2$ under uniform and adaptive mesh refinements, with convergence rates of approximately $O(N_{DoFs}^{-0.28})$ and $O(N_{DoFs}^{-0.40})$, respectively. **Central panel:** Error $\mathcal{E}^Y$ for the VEM with $p = 2$ under uniform and adaptive mesh refinements, with convergence rates of approximately $O(N_{DoFs}^{-0.52})$ and $O(N_{DoFs}^{-1})$, respectively. **Right panel:** The effectivity index for the adaptive VEM.

In Figure 9.14, we show the Y-type errors for both methods under uniform and adaptive mesh refinements with Dörfler marking parameter $\theta = 0.9$ for the test case 3 and $p = 1$. The adaptive procedure for the VEM (8.10) leads to higher convergence rates compared to those obtained for uniform refinements. For the continuous FEM (7.15), although the adaptive procedure produces meshes that are refined towards the lower corners as in Figure 9.16 (right panels), the error $\widetilde{\mathcal{E}}^Y$ does not converge to zero.
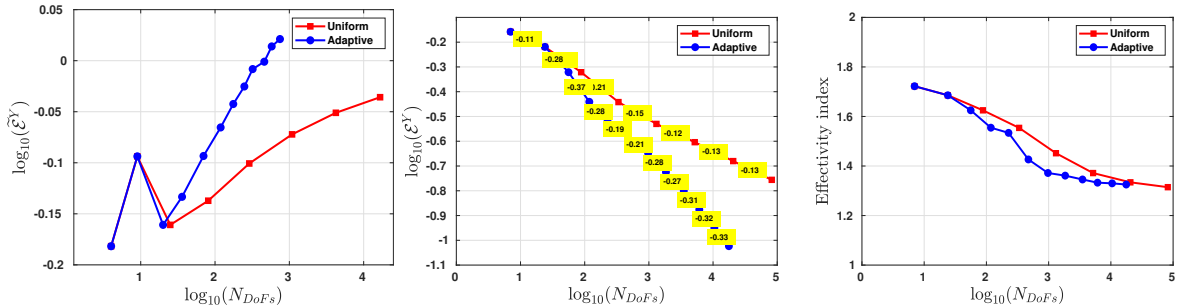

Figure 9.14 The test case 3 with exact solution u_3 in (9.4). **Left panel:** Error $\widetilde{\mathcal{E}}^Y$ for the continuous finite element with $p = 1$ under uniform and adaptive refinements. **Central panel:** Error $\mathcal{E}^Y$ for the VEM with $p = 1$ under uniform and adaptive mesh refinements, with convergence rates of approximately $O(N_{DoFs}^{-0.13})$ and $O(N_{DoFs}^{-0.33})$, respectively. **Right panel:** The effectivity index for the adaptive VEM.

In Figures 9.15 and 9.16, we plot some meshes for the test cases with exact solutions u_2 ($\alpha = 0.55$) and u_3, respectively, produced by the adaptive procedure driven by the VEM error indicator in (9.7) and the continuous FEM error indicator in (9.9).

Next, in Tables 9.1 and 9.2, we focus on the adaptive mesh refinements driven by the VEM error indicator in (9.7) and report the number of time-slabs and "reference elements" as in Sections 9.3 and 9.4 for some adaptively generated meshes. The time-slab and the element-topology strategies allow us to speed up considerably the assembly and solve time of method (8.10).
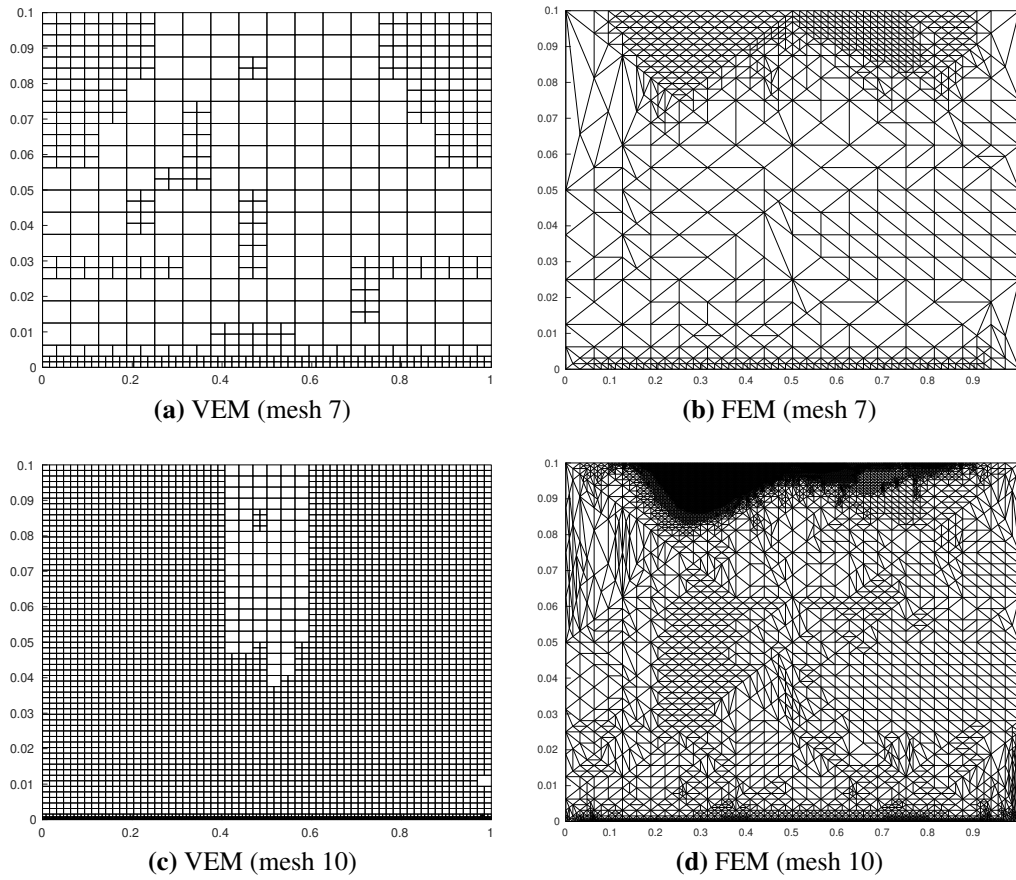
(a) VEM (mesh 7)

(b) FEM (mesh 7)

(c) VEM (mesh 10)

(d) FEM (mesh 10)

Figure 9.15 Meshes generated by the adaptive schemes driven by the VEM error indicator η in (9.7) (left panel) and the continuous FEM error indicator $\widetilde{\eta}$ (right panel) for the test case with exact solutions u_2 ($\alpha = 0.55$) in (9.3).

Mesh	Number of reference elements	Total number of elements	Number of time-slabs
m_2	1	4	2
m_4	1	52	5
m_6	16	265	10
m_8	19	1189	21
m_{10}	24	5110	45
m_{12}	24	21883	103

Table 9.1 Performance of flagging strategies for the test case 2 with $\alpha = 0.55$ and $p = 2$. The corresponding errors $\mathcal{E}^Y$ are shown in Figure 9.13.
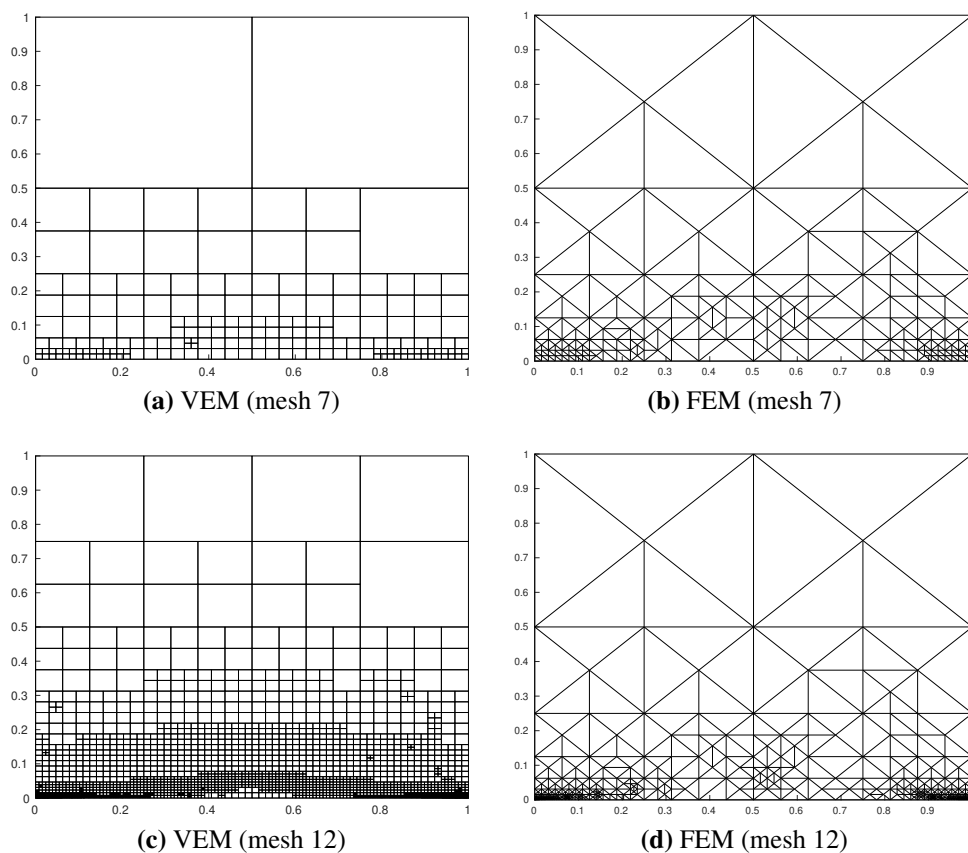
Figure 9.16 Meshes generated by the adaptive schemes driven by the VEM error indicator η in (9.7) (left panel) and the continuous FEM error indicator $\widetilde{\eta}$ in (9.9) (right panel) for the test case with exact solutions u_3 in (9.4).

Mesh	Number of reference elements	Total number of elements	Number of time-slabs
m_2	1	4	2
m_4	3	22	4
m_6	9	91	5
m_8	21	361	9
m_{10}	33	1204	13
m_{12}	35	3493	21

Table 9.2 Performance of flagging strategies for the test case 3 with $p = 1$. The corresponding errors $\mathcal{E}^Y$ are shown in Figure 9.14.

119

References

[1] D. R. Adams and L. I. Hedberg. *Function spaces and potential theory*, volume 314 of *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, Berlin, 1996.

[2] X. Antoine, C. Besse, and V. Mouysset. Numerical schemes for the simulation of the two-dimensional Schrödinger equation using non-reflecting boundary conditions. *Math. Comp.*, 73(248):1779–1799, 2004.

[3] B. Ayuso de Dios, K. Lipnikov, and G. Manzini. The nonconforming virtual element method. *ESAIM Math. Model. Numer. Anal.*, 50(3):879–904, 2016.

[4] A. K. Aziz and P. Monk. Continuous finite elements in space and time for the heat equation. *Math. Comp.*, 52(186):255–274, 1989.

[5] I. Babuška and T. Janik. The *h-p* version of the finite element method for parabolic equations. I. The *p*-version in time. *Numer. Methods Partial Differential Equations*, 5(4):363–399, 1989.

[6] I. Babuška and T. Janik. The *h-p* version of the finite element method for parabolic equations. II. The *h-p* version in time. *Numer. Methods Partial Differential Equations*, 6(4):343–369, 1990.

[7] L. Banjai, E. H. Georgoulis, and O. Lijoka. A Trefftz polynomial space-time discontinuous Galerkin method for the second order wave equation. *SIAM J. Numer. Anal.*, 55(1):63–86, 2017.

[8] P. Bansal, A. Moiola, I. Perugia, and C. Schwab. Space-time discontinuous Galerkin approximation of acoustic waves with point singularities. *IMA J. Numer. Anal.*, 41(3):2056–2109, 2021.

[9] L. Beirão da Veiga, F. Brezzi, A. Cangiani, G. Manzini, L. D. Marini, and A. Russo. Basic principles of virtual element methods. *Math. Models Methods Appl. Sci.*, 23(1):199–214, 2013.

[10] M. Born and R. Oppenheimer. On the quantum theory of molecules. In *Quantum Chemistry: Classic Scientific Papers*, pages 1–24. World Scientific, 2000.

[11] S. C. Brenner. Poincaré-Friedrichs inequalities for piecewise H^1 functions. *SIAM J. Numer. Anal.*, 41(1):306–324, 2003.

[12] S. C. Brenner and L. R. Scott. *The mathematical theory of finite element methods*, volume 15 of *Texts in Applied Mathematics*. Springer, New York, third edition, 2008.

[13] J. Callahan. *Advanced calculus: a geometric view*. Springer Science & Business Media, 2010.

[14] A. Cangiani, Z. Dong, and E. H. Georgoulis. *hp*-version space-time discontinuous Galerkin methods for parabolic problems on prismatic meshes. *SIAM J. Sci. Comput.*, 39(4):A1251–A1279, 2017.

[15] O. Cessenat and B. Després. Application of an ultra weak variational formulation of elliptic PDEs to the two-dimensional Helmholtz problem. *SIAM J. Numer. Anal.*, 35(1):255–299, 1998.

[16] H. S. M. Coxeter. *Regular polytopes.* Dover Publications, third edition, 1973.

[17] R. E. Crandall and B. R. Litt. Reassembly and time advance in reflectionless scattering. *Ann. Physics*, 146(2):458–469, 1983.

[18] J. P. Dahl and M. Springborg. The Morse oscillator in position space, momentum space, and phase space. *J. Chem. Phys.*, 88(7):4535–4547, 1988.

[19] R. Dautray and J.-L. Lions. *Mathematical analysis and numerical methods for science and technology. Vol. 5.* Springer-Verlag, Berlin, 1992.

[20] M. Dehghan and A. Shokri. A numerical method for two-dimensional Schrödinger equation using collocation and radial basis functions. *Comput. Math. Appl.*, 54(1):136–146, 2007.

[21] L. Demkowicz, J. Gopalakrishnan, S. Nagaraj, and P. Sepúlveda. A spacetime DPG method for the Schrödinger equation. *SIAM J. Numer. Anal.*, 55(4):1740–1759, 2017.

[22] D. Devaud. Petrov-Galerkin space-time *hp*-approximation of parabolic equations in $H^{1/2}$. *IMA J. Numer. Anal.*, 40(4):2717–2745, 2020.

[23] D. Devaud and C. Schwab. Space-time *hp*-approximation of parabolic equations. *Calcolo*, 55(3):35, 2018.

[24] D. A. Di Pietro and A. Ern. *Mathematical aspects of discontinuous Galerkin methods*, volume 69 of *Mathématiques & Applications (Berlin).* Springer, Heidelberg, 2012.

[25] W. Dörfler. A convergent adaptive algorithm for Poisson's equation. *SIAM J. Numer. Anal.*, 33(3):1106–1124, 1996.

[26] R. G. Durán. On polynomial approximation in Sobolev spaces. *SIAM J. Numer. Anal.*, 20(5):985–988, 1983.

[27] H. Egger, F. Kretzschmar, S. M. Schnepp, I. Tsukerman, and T. Weiland. Transparent boundary conditions for a discontinuous Galerkin Trefftz method. *Appl. Math. Comput.*, 267:42–55, 2015.

[28] H. Egger, F. Kretzschmar, S. M. Schnepp, and T. Weiland. A space-time discontinuous Galerkin Trefftz method for time dependent Maxwell's equations. *SIAM J. Sci. Comput.*, 37(5):B689–B711, 2015.

[29] K. Eriksson and C. Johnson. Adaptive finite element methods for parabolic problems. I. A linear model problem. *SIAM J. Numer. Anal.*, 28(1):43–77, 1991.

References

[30] K. Eriksson and C. Johnson. Adaptive finite element methods for parabolic problems. II. Optimal error estimates in $L_\infty L_2$ and $L_\infty L_\infty$. *SIAM J. Numer. Anal.*, 32(3):706–740, 1995.

[31] K. Eriksson, C. Johnson, and V. Thomée. Time discretization of parabolic problems by the discontinuous Galerkin method. *RAIRO Modél. Math. Anal. Numér.*, 19(4):611–643, 1985.

[32] A. Ern, I. Smears, and M. Vohralík. Guaranteed, locally space-time efficient, and polynomial-degree robust a posteriori error estimates for high-order discretizations of parabolic problems. *SIAM J. Numer. Anal.*, 55(6):2811–2834, 2017.

[33] A. Ern and M. Vohralík. A posteriori error estimation based on potential and flux reconstruction for the heat equation. *SIAM J. Numer. Anal.*, 48(1):198–223, 2010.

[34] D. A. French and T. E. Peterson. A continuous space-time finite element method for the wave equation. *Math. Comp.*, 65(214):491–506, 1996.

[35] T. Führer and M. Karkulik. Space-time least-squares finite elements for parabolic equations. *Comput. Math. Appl.*, 92:27–36, 2021.

[36] G. Gantner and R. Stevenson. Further results on a space-time FOSLS formulation of parabolic PDEs. *ESAIM Math. Model. Numer. Anal.*, 55(1):283–299, 2021.

[37] G. Gantner and R. Stevenson. Improved rates for a space-time FOSLS of parabolic PDEs. *arXiv:2208.10824*, 2022.

[38] E. H. Georgoulis, O. Lakkis, and T. P. Wihler. A posteriori error bounds for fully-discrete hp-discontinuous Galerkin timestepping methods for parabolic problems. *Numer. Math.*, 148(2):363–386, 2021.

[39] E. H. Georgoulis and C. G. Makridakis. Lower bounds, elliptic reconstruction and a posteriori error control of parabolic problems. *IMA J. Num. Anal.*, page drac080, 2023.

[40] C. J. Gittelson, R. Hiptmair, and I. Perugia. Plane wave discontinuous Galerkin methods: analysis of the h-version. *M2AN Math. Model. Numer. Anal.*, 43(2):297–331, 2009.

[41] S. Gómez, L. Mascotto, A. Moiola, and I. Perugia. Space-time virtual elements for the heat equation. *To appear in SIAM J. Numer. Anal.*, 2023.

[42] S. Gómez, L. Mascotto, and I. Perugia. Design and performance of a space–time virtual element method for the heat equation on prismatic meshes. *Comput. Methods Appl. Mech. Eng.*, 418(A):116491, 2024.

[43] S. Gómez and A. Moiola. A space-time Trefftz discontinuous Galerkin method for the linear Schrödinger equation. *SIAM J. Numer. Anal.*, 60(2):688–714, 2022.

[44] S. Gómez and A. Moiola. Space–time DG method for the Schrödinger equation with variable potential. *arXiv:2306.05780*, 2023.

[45] S. Gómez, A. Moiola, I. Perugia, and P. Stocker. On polynomial Trefftz spaces for the linear time-dependent Schrödinger equation. *Appl. Math. Lett.*, 146(C):108824, 2023.

[46] R. Grella. Fresnel propagation and diffraction and paraxial wave equation. *J. of Optics*, 13(6):367, 1982.

[47] D. J. Griffiths. *Introduction to Quantum Mechanics*. Prentice-Hall, New York, 1995.

[48] D. J. Griffiths and D. F. Schroeter. *Introduction to quantum mechanics*. Cambridge university press, 2018.

[49] S. Hain and K. Urban. An ultra-weak space-time variational formulation for the Schrödinger equation. 2022. arXiv:2212.14398.

[50] N. J. Higham. *Accuracy and stability of numerical algorithms*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, second edition, 2002.

[51] R. Hiptmair, A. Moiola, and I. Perugia. Error analysis of Trefftz-discontinuous Galerkin methods for the time-harmonic Maxwell equations. *Math. Comp.*, 82(281):247–268, 2013.

[52] R. Hiptmair, A. Moiola, and I. Perugia. A survey of Trefftz methods for the Helmholtz equation. In *Building bridges: connections and challenges in modern approaches to numerical partial differential equations*, volume 114 of *Lect. Notes Comput. Sci. Eng.*, pages 237–278. Springer, Cham, 2016.

[53] T. J. R. Hughes and J. R. Stewart. A space-time formulation for multiscale phenomena. *J. Comput. Appl. Math.*, 74(1-2):217–229, 1996.

[54] L.-M. Imbert-Gérard and B. Després. A generalized plane-wave numerical method for smooth nonconstant coefficients. *IMA J. Numer. Anal.*, 34(3):1072–1103, 2014.

[55] L.-M. Imbert-Gérard, A. Moiola, and P. Stocker. A space-time quasi-Trefftz DG method for the wave equation with piecewise-smooth coefficients. *Math. Comp.*, 92(341):1211–1249, 2023.

[56] L.-M. Imbert-Gérard and P. Monk. Numerical simulation of wave propagation in inhomogeneous media using generalized plane waves. *ESAIM Math. Model. Numer. Anal.*, 51(4):1387–1406, 2017.

[57] P. Jamet. Galerkin-type approximations which are discontinuous in time for parabolic equations in a variable domain. *SIAM J. Numer. Anal.*, 15(5):912–928, 1978.

[58] S. Jin, P. Markowich, and C. Sparber. Mathematical and computational methods for semiclassical Schrödinger equations. *Acta Numer.*, 20:121–209, 2011.

[59] O. Karakashian and C. Makridakis. A space-time finite element method for the nonlinear Schrödinger equation: the discontinuous Galerkin method. *Math. Comp.*, 67(222):479–499, 1998.

References

[60] O. Karakashian and C. Makridakis. A space-time finite element method for the non-linear Schrödinger equation: the continuous Galerkin method. *SIAM J. Numer. Anal.*, 36(6):1779–1807, 1999.

[61] J. Keller and J. Papadakis. *Wave propagation and underwater acoustics*. Springer, 1977.

[62] P. Kopp, V. Calo, E. Rank, and S. Kollmannsberger. Space-time hp-finite elements for heat evolution in laser powder bed fusion additive manufacturing. *Eng. Comput.*, 38(6):4879–4893, 2022.

[63] L. D. Landau and E. M. Lifshitz. *Quantum Mechanics; Non-relativistic Theory*. Pergamon Press, 1965.

[64] U. Langer, S. Matculevich, and S. Repin. A posteriori error estimates for space-time IgA approximations to parabolic initial boundary value problems. *arXiv:1612.08998*, 2016.

[65] U. Langer, S. Matculevich, and S. Repin. Functional type error control for stabilised space-time IgA approximations to parabolic problems. In *Large-scale scientific computing*, volume 10665 of *Lecture Notes in Comput. Sci.*, pages 55–65. Springer, Cham, 2018.

[66] U. Langer, S. E. Moore, and M. Neumüller. Space-time isogeometric analysis of parabolic evolution problems. *Comput. Methods Appl. Mech. Engrg.*, 306:342–363, 2016.

[67] U. Langer and A. Schafelner. Adaptive space-time finite element methods for non-autonomous parabolic problems with distributional sources. *Comput. Methods Appl. Math.*, 20(4):677–693, 2020.

[68] U. Langer and A. Schafelner. Space-time hexahedral finite element methods for parabolic evolution problems. In *Domain Decomposition Methods in Science and Engineering XXVI*, pages 515–522. Springer International Publishing, 2023.

[69] U. Langer and O. Steinbach. *Space-Time Methods: Applications to Partial Differential Equations*, volume 25. Walter de Gruyter GmbH & Co KG, 2019.

[70] S. Larsson, R. H. Nochetto, S. A Sauter, and C. Wieners. Space-time methods for time-dependent partial differential equations. *Oberwolfach reports*, 19(1):303–381, 2023.

[71] C. Lasser and C. Lubich. Computing quantum dynamics in the semiclassical regime. *Acta Numer.*, 29:229–401, 2020.

[72] C. Lehrenfeld and P. Stocker. Embedded Trefftz discontinuous Galerkin methods. *Int. J. Numer. Methods Eng.*, 124(17):3637–3661, 2023.

[73] M. Levy. *Parabolic equation methods for electromagnetic wave propagation*, volume 45 of *IEE Electromagnetic Waves Series*. Institution of Electrical Engineers (IEE), London, 2000.

[74] J.-L. Lions and E. Magenes. *Non-homogeneous boundary value problems and applications. Vol. I*. Die Grundlehren der mathematischen Wissenschaften, Band 181. Springer-Verlag, New York-Heidelberg, 1972.

[75] L. Mascotto. The role of stabilization in the virtual element method: a survey. *Comput. Math. with Appl.*, 151:244–251, 2023.

[76] L. Mascotto, I. Perugia, and A. Pichler. Non-conforming harmonic virtual element method: *h*- and *p*-versions. *J. Sci. Comput.*, 77(3):1874–1908, 2018.

[77] L. Mascotto, I. Perugia, and A. Pichler. A nonconforming Trefftz virtual element method for the Helmholtz problem. *Math. Models Methods Appl. Sci.*, 29(9):1619–1656, 2019.

[78] W. McLean. *Strongly elliptic systems and boundary integral equations*. Cambridge University Press, Cambridge, 2000.

[79] A. Moiola, R. Hiptmair, and I. Perugia. Plane wave approximation of homogeneous Helmholtz solutions. *Z. Angew. Math. Phys.*, 62(5):809–837, 2011.

[80] A. Moiola and I. Perugia. A space-time Trefftz discontinuous Galerkin method for the acoustic wave equation in first-order formulation. *Numer. Math.*, 138(2):389–435, 2018.

[81] M. Montardini, M. Negri, G. Sangalli, and M. Tani. Space-time least-squares isogeometric method and efficient solver for parabolic problems. *Math. Comp.*, 89(323):1193–1227, 2020.

[82] P. M. Morse. Diatomic molecules according to the wave mechanics. II. Vibrational levels. *Phys. Rev.*, 34:57–64, 1929.

[83] R. H. Nochetto, S. A Sauter, and C. Wieners. Space-time methods for time-dependent partial differential equations. *Oberwolfach Reports*, 14(1):863–947, 2018.

[84] F. Olver, Daniel W. Lozier, R. F Boisvert, and C. Clark. *NIST handbook of mathematical functions*. Cambridge university press, 2010.

[85] I. Perugia, J. Schöberl, P. Stocker, and C. Wintersteiger. Tent pitching and Trefftz-DG method for the acoustic wave equation. *Comput. Math. Appl.*, 79(10):2987–3000, 2020.

[86] I. Perugia, C. Schwab, and M. Zank. Exponential convergence of *hp*-time-stepping in space-time discretizations of parabolic PDEs. *ESAIM Math. Model. Numer. Anal.*, 57(1):29–67, 2023.

[87] Q.-H. Qin. Trefftz Finite Element Method and Its Applications. *Appl. Mech. Rev.*, 58(5):316–337, 2005.

[88] S. Repin. Estimates of deviations from exact solutions of initial-boundary value problem for the heat equation. *Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. Rend. Lincei (9) Mat. Appl.*, 13(2):121–133, 2002.

[89] P. C. Rosenbloom and D. V. Widder. Expansions in terms of heat polynomials and associated functions. *Trans. Amer. Math. Soc.*, 92:220–266, 1959.

[90] A. Schafelner and P. S. Vassilevski. Numerical results for adaptive (negative norm) constrained first order system least squares formulations. *Comput. Math. Appl.*, 95:256–270, 2021.

References

[91] D. Schötzau and C. Schwab. Time discretization of parabolic problems by the hp-version of the discontinuous Galerkin finite element method. *SIAM J. Numer. Anal.*, 38(3):837–875, 2000.

[92] C. Schwab and R. Stevenson. Space-time adaptive wavelet methods for parabolic evolution problems. *Math. Comp.*, 78(267):1293–1318, 2009.

[93] O. Steinbach. Space-time finite element methods for parabolic problems. *Comput. Methods Appl. Math.*, 15(4):551–566, 2015.

[94] O. Steinbach and H. Yang. Comparison of algebraic multigrid methods for an adaptive space-time finite-element discretization of the heat equation in 3D and 4D. *Numer. Linear Algebra Appl.*, 25(3):e2143, 17, 2018.

[95] O. Steinbach and H. Yang. Space-time finite element methods for parabolic evolution equations: discretization, a posteriori error estimation, adaptivity and solution. In *Space-Time Methods: Application to Partial Differential Equations*, volume 25 of *Radon Ser. Comput. Appl. Math.*, pages 207–248. De Gruyter, Berlin, 2019.

[96] O. Steinbach and M. Zank. Coercive space-time finite element methods for initial boundary value problems. *Electron. Trans. Numer. Anal.*, 52:154–194, 2020.

[97] R. Stevenson and J. Westerdiep. Stability of Galerkin discretizations of a mixed space-time variational formulation of parabolic evolution equations. *IMA J. Numer. Anal.*, 41(1):28–47, 2021.

[98] M. Subaşi. On the finite-differences schemes for the numerical solution of two dimensional Schrödinger equation. *Numer. Methods Partial Differential Equations*, 18(6):752–758, 2002.

[99] J. J. Sudirham, J. J. W. van der Vegt, and R. M. J. van Damme. Space-time discontinuous Galerkin method for advection-diffusion problems on time-dependent domains. *Appl. Numer. Math.*, 56(12):1491–1518, 2006.

[100] K. Voronin, C. S. Lee, M. Neumüller, P. Sepulveda, and P. S. Vassilevski. Space-time discretizations using constrained first-order system least squares (CFOSLS). *J. Comput. Phys.*, 373:863–876, 2018.

[101] J. Wloka. *Partial differential equations*. Cambridge University Press, Cambridge, 1987.