# Parabolic PDEs in Space-Time Formulations: Stability for Petrov-Galerkin Discretizations with B-Splines and Existence of Moments for Problems with Random Coefficients

I n a u g u r a l - D i s s e r t a t i o n

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

**Christian Mollet**

aus Paderborn

Köln, 2016

**Berichterstatter:** Prof. Dr. Ulrich Trottenberg

Prof. Dr. Guido Sweers

Prof. Dr. Olaf Steinbach (TU Graz)

**Tag der mündlichen Prüfung:** 01. 07. 2016

# Preface

Mein größter Dank gilt Prof. Dr. Angela Kunoth für die hervorragende Betreuung schon zu meiner Paderborner Studentenzeit. Sowohl fachlich als auch persönlich stand sie mir immer mit Rat und Tat zur Seite. Ohne sie wäre die Arbeit in dieser Form sicher gar nicht erst zustande gekommen. Darüber hinaus möchte ich mich auch für die finanzielle Unterstützung durch die Anstellung als wissenschaftlicher Mitarbeiter bedanken.

Ein weiterer besonderer Dank gebührt den Gutachtern meiner Dissertation Prof. Dr. Ulrich Trottenberg, Prof. Dr. Guido Sweers und Prof. Dr. Olaf Steinbach: Für ihren nicht selbstverständlichen und spontanen Einsatz und ihre unvoreingenommene Bereitschaft, mich auf der Zielgraden der Promotion zu begleiten.

Auch die weiteren Mitgliedern der Prüfungskommission, Prof. Dr. George Marinescu und Dr. Stephanie Friedhoff, sowie das gesamte Team des mathematischen Instituts, möchte ich an dieser Stelle dankend erwähnen, insbesondere Herrn apl. Prof. Dr. Dirk Horstmann für seinen unermüdlichen Einsatz in meiner Sache.

I want to thank Prof. Dr. Stig Larsson for all his mathematical ideas and discussions during my stay at Chalmers. He has strongly influenced important parts of my thesis. The same holds true for Matteo Molteni. I had a really great time in Göteborg. Tak!

Bedanken möchte ich mich auch bei sämtlichen Kollegen aus Paderborner und Kölner Zeiten, die mich auf dem Weg zur Promotion begleitet haben.

Zuallererst möchte ich meinem langjährigen Bürokollegen Dr. Roland Pabel aus der Zeit an der Uni Paderborn danken: Er war eine große Bereicherung für den Arbeitsalltag, sowohl fachlich aber auch persönlich.

Danken möchte ich auch der aktuellen Besetzung der AG an der Universität zu Köln: Meinem Bürokollegen Stephan Gerster gilt dabei ein besonderer Dank für seine Ausdauer und Beharrlichkeit beim Korrekturlesen und was nicht vergessen werden darf, für unsere unterhaltsamen mathematischen, sowie vor allem unmathematischen Gespräche. Besonders für das Korrekturlesen gilt mein Dank auch Sam Leweke und Sandra Boschert. Nicht zu vergessen sind natürlich die restlichen Mitglieder der AG: Boqian Huang, Enis Sen, José Licón und Anke Müller (sortiert im Uhrzeigersinn nach Arbeitsplätzen), die mir ebenfalls immer mit Rat und Tat zur Seite standen.

Nicht vergessen darf man auch den kollegialen Beistand meiner „Flur“-Kollegen der AG von Prof. Dr. Gregor Gassner. Besonders für die tiefgründigen Gespräche fernab der Mathematik beim Mittagessen.

Abseits von der fachlichen Seite gilt mein größter Dank Andrea Eckhoff-Rosenbaum. Sie musste sicher am meisten unter meinen wirren mathematischen Erzählungen leiden. Aber während der gesamten Zeit meiner Promotion stand sie mir immer zur Seite und war insbesondere in den letzten Wochen vor der Abgabe die größte Stütze.

Ein sehr großes Dankeschön gilt hier auch meinen Eltern, die mich bei meinem akademischen Lebensweg immer uneingeschränkt unterstützt haben und mir besonders in schwierigen Zeit immer zu Seite standen und mich aufgemuntert haben. An dieser Stelle dürfen auch meine Großeltern nicht vergessen werden: Besonders für meinen Opa Hans hört mit der Abgabe dieser Arbeit das Warten bald auf.

Last but not least möchte ich allen meinen Freunden danken. Danke, dass es euch gibt und für all die schönen Zeiten.

# Contents

# Zusammenfassung

Der Inhalt dieser Arbeit enthält Einflüsse der numerischen Mathematik, der Funktionalanalysis und der Approximationstheorie, sowie der Stochastik beziehungsweise der Uncertainty Quantification. Im Fokus stehen neben den voll schwachen Raum-Zeit-Formulierungen linearer parabolischer partieller Differentialgleichungen und deren Eigenschaften vor allem die Stabilität ihrer jeweiligen Petrov-Galerkin-Approximationen. Darüber hinaus wird das Konzept der schwachen Raum-Zeit-Formulierungen auf parabolische Differentialgleichungen mit stochastischen Koeffizienten erweitert, wobei ein weiterer zentraler Aspekt dieser Arbeit die Existenz von Momenten der Lösung und die Optimalität der entsprechenden Petrov-Galerkin-Lösung sein wird. Zur numerischen Umsetzung der Petrov-Galerkin-Approximationen mit B-Splines beliebiger Ordnung wurde im Rahmen dieser Arbeit ein Programmcode in Matlab entwickelt.

Durch die Formulierung in voll schwacher Form bezüglich der Zeit und dem Ort erhält man eine einzige Operatorgleichung. Diese Formulierung hat den Vorteil, dass sich die Existenz und Eindeutigkeit einer Lösung durch die Isomorphie des Raum-Zeit-Operators folgern lässt und darüber hinaus auch explizite Schranken des Operators gegeben sind. Um die Stabilität einer Petrov-Galerkin-Approximation gewährleisten zu können, muss sichergestellt werden, dass eine diskrete inf-sup Bedingung unabhängig von der Feinheit der Diskretisierung erfüllt ist. Die Validität der diskreten inf-sup Bedingung wird, anders als bei elliptischen Operatoren, nicht von ihrem kontinuierlichen Pendant geerbt. Die Stabilitätsanalyse wird zunächst für allgemeine Operatorgleichungen durchgeführt und anschließend explizit auf die Raum-Zeit-Formulierungen angewandt. Ziel ist es eine möglichst allgemeingültige Konstruktionsvorschrift für stabile Diskretisierungen zu erarbeiten.

Ferner werden schwache Raum-Zeit-Formulierungen auf parabolische Gleichungen mit stochastischen Koeffizienten erweitert. Hierbei wird das Hauptaugenmerk auf die Existenz von Momenten der Lösung, sowie der quasi Optimalität und Stabilität der Petrov-Galerkin-Approximation gelegt. Die Leitidee wird es sein, die Schranken des räumlichen Differenzialoperators als Zufallsvariablen anstatt als feste Konstanten zu betrachten.

Im Zusammenhang mit dieser Arbeit wurde zusätzlich ein Programmpaket in Matlab implementiert. Die Umsetzung basiert auf Diskretisierungen mittels B-Splines und ist auf eine breite Anwendung ausgelegt. Im Vordergrund stehen Petrov-Galerkin-Diskretisierungen von parabolischen Gleichungen mit stochastischen Parametern in schwacher Raum-Zeit-Formulierung. Das Programmpaket lässt sich in beliebigen Raumdimensionen, Ordnungen und Diskretisierungsleveln der B-Splines, sowie für uniforme und nicht uniforme Gitter, unabhängig für Test- sowie Lösungsraum anwenden. Zur benutzerfreundlichen Anwendung wurde zusätzlich eine grafische Oberfläche zur Petrov-Galerkin-Approximation von parabolischen Gleichungen in den vorgestellten schwachen Raum-Zeit-Formulierungen bereitgestellt.

# Abstract

The topic of this thesis contains influences of numerical mathematics, functional analysis and approximation theory, as well as stochastic respectively uncertainty quantification. The focus is beside the full space-time weak formulation of linear parabolic partial differential equations and their properties, especially the stability of their Petrov-Galerkin approximations. Moreover, the concept of space-time weak formulations is extended to parabolic differential equations with random coefficients, where an additional central aspect of this thesis is the existence of moments of the solution and the optimality of its Petrov-Galerkin solution. In the scope of this work a Matlab code for the numerical realization of the Petrov-Galerkin approximation with B-splines of arbitrary order was developed.

Due to the formulation in full weak form with respect to space and time one obtains a single operator equation. This formulation has the advantage that the existence and uniqueness of a solution follows from the isomorphism of the space-time operator and furthermore yields explicit bounds of the operator. In order to guarantee stability of a Petrov-Galerkin approximation, one needs to ensure that a discrete inf-sup condition is fulfilled independent of the refinement of the discretization. The validity of the discrete inf-sup is, contrary to elliptic operators, not inherited from its continuous pendant. The stability analysis is performed for generic operator equations first and is applied afterwards explicitly to space-time formulations. The goal is to develop a preferably universal construction rule for stable discretizations.

Furthermore, the space-time formulations are extended to parabolic problems with random coefficients. The main focus lies on the existence of moments of the solution as well as the quasi-optimality and stability of the Petrov-Galerkin approximation. A central idea will be to consider the bounds of the spatial differential operator as random variables rather than fixed constants.

In connection with this thesis, a program package in Matlab was implemented. The realization bases on discretizations with B-splines and is developed for a broad functionality. It focuses on Petrov-Galerkin discretizations of parabolic partial differential equations with random coefficients in space-time weak formulation. The code can be applied for different space dimensions, orders and level of discretizations of B-splines, as well as for uniform and non-uniform grids, independent for test and solution space. In addition, for a user-friendly handling it was implemented a graphical user interface for Petrov-Galerkin approximations of parabolic equations in the presented space-time weak formulations.

# 1 Introduction

## 1.1 Motivation and Overview

Consider the following well-known heat equation

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}u(t,x) - a\Delta_x u(t,x) &= g(t,x), \quad \text{for } (t,x) \in (0,T] \times D, \\ u(0,x) &= u_0(x), \quad \text{for } x \in D, \\ u(t,x) &= 0, \quad \text{for } (t,x) \in (0,T] \times \partial D, \end{aligned} \tag{1.1.1}$$

on a bounded spatial domain $D \subset \mathbb{R}^n$ and finite time interval $(0,T]$, $0 < T < \infty$, with Laplacian $\Delta_x$ acting on the $x$ variable and sufficiently smooth given right hand side $g$, initial value (function) $u_0$ and homogeneous boundary conditions as well as constant thermal diffusivity or diffusion coefficient $a \in \mathbb{R}^+$. The heat equation is of parabolic type and describes the distribution of heat in $D$. It is stated in *strong* form. A classical way to approximately solve such kind of equations is via finite differences for instance. In finite difference schemes, one defines a set of finite grid points and approximates the derivatives with a difference quotient. A drawback is that the approximation is only given at the finite grid points and also requires rather strong regularity assumptions to yield accurate approximations. A different approach is to reformulate equation (1.1.1) in a variational form. To this end, one multiplies (1.1.1) with a spatial test function $\phi \in \mathcal{C}_0^\infty(D)$ and integrate over $D$. In this way one arrives at the standard weak formulation

$$\int_D (\dot{u}(t,x)\phi(x) + a\nabla_x u(t,x) \cdot \nabla_x \phi(x))\,\mathrm{d}x = \int_D g(t,x)\phi(x)\,\mathrm{d}x, \quad \text{for all } \phi \in H_0^1(D) \tag{1.1.2}$$

with $u(0,x) = u_0(x)$ for $x \in D$, since $\mathcal{C}_0^\infty(D) \subset H_0^1(D)$ is a dense subset and by using integration by parts. We have used the shorthand notations

$$\dot{u}(t,x) := \frac{\mathrm{d}}{\mathrm{d}t}u(t,x), \quad \nabla_x u(t,x) \cdot \nabla_x \phi(x) := \sum_{i=1}^n \left(\frac{\partial}{\partial x_i}u(t,x)\right)\left(\frac{\partial}{\partial x_i}\psi(x)\right).$$

Typical approaches to solve parabolic problems in variational formulation (1.1.2) are the method of lines or Rothe's method. The idea of the method of lines for instance is to consider the equation on a finite spatial subspace, e.g., a finite element space, which yields a system of ordinary differential equations (ODEs) in time. This system of ODEs can be solved, e.g., with Runge-Kutta methods. Although we have formulated the heat equation in a variational sense, the time variable $t$ is still involved explicitly leading, in particular, to time stepping methods.

I will pick up a different approach in this thesis. Instead of working with a classical weak formulation (1.1.2) in space only, a full weak formulation in space *and* time will

be considered. The main idea is to multiply the PDE (1.1.1) with space-time test functions and to integrate over both space *and* time. In this way one arrives at the full *space-time weak problem*:

$$\text{Find } u \in L_2(I;V) \cap H^1(I;V') : b(u,v) = \mathcal{F}(v) \qquad \forall\, v \in L_2(I;V) \times H, \tag{1.1.3}$$

with $H := L_2(D)$, $V := H_0^1(D)$, $I := (0,T)$ and

$$\begin{aligned} b(u,(v_1,v_2)) &:= \int_I \int_D (\dot{u}(t,x)v_1(t,x) + a\nabla_x u(t,x) \cdot \nabla_x v_1(t,x)) \,\mathrm{d}x\mathrm{d}t \\ &\quad + \int_D u(0,x)v_2(x) \,\mathrm{d}x, \\ \mathcal{F}(v) &:= \int_I \int_D g(t,x)v_1(t,x) \,\mathrm{d}x\mathrm{d}t + \int_D u_0(x)v_2(x) \,\mathrm{d}x. \end{aligned}$$

The explicit time dependency is eliminated in this full space-time weak formulation (1.1.3). Similar to the method of lines mentioned above, one can consider the equation on finite subspaces, but now with respect to the time domain *and* the spatial domain. This approach leads to *Petrov-Galerkin discretizations*, which will be discussed detailed in this thesis. In a Petrov-Galerkin approach, linear systems of equations can be solved without any time stepping. Therefore, depending on the properties of the system matrix, one can make use of well-known solvers from linear algebra for instance. But one of the main motivations to consider parabolic evolution problems in a full weak formulation was the fact that theoretical tools are available to derive explicit bounds for both, the solution in energy norm as well as the error of a Petrov-Galerkin approximation. These trackable constants will turn out the be essential for the development of the results in this thesis, not only concerning stability of Petrov-Galerkin discretizations, but also in the context of the extension to random PDEs.

In a more general setting with linear spatial differential operators $A(t)\colon V \to V'$ replacing $-a\Delta_x$, one obtains

$$b(u,(v_1,v_2)) := \int_I \big({}_{V'}\langle \dot{u}(t), v_1(t)\rangle_V + {}_{V'}\langle A(t)u(t), v_1(t)\rangle_V\big) \,\mathrm{d}t + (u(0),v_2)_H, \tag{1.1.4}$$

with Hilbert spaces $V \hookrightarrow H \cong H' \hookrightarrow V'$ in a Gelfand triple and duality pairing ${}_{V'}\langle\cdot,\cdot\rangle_V$ on $V' \times V$. This kind of approach was already mentioned in, e.g., [DL92] in order to prove existence and uniqueness of solutions of parabolic problems in weak formulation, but not primary considered as a stand-alone full weak formulation and without explicit bounds. A full weak space-time formulation was revisited in [SS09] to treat it numerically with adaptive wavelet methods. I will solely consider linear equations in this thesis, such that $b(\cdot,\cdot)$ is a bilinear form. In addition, a homogenization as well as another slightly different full weak formulation by using integration by parts is discussed in this thesis, cf. [Sta11] and [CS11].

Although the parabolic problem is condensed to one linear equation, the well-known Lax-Milgram Theorem cannot be applied since it is inevitable to consider different solution and test spaces. But there is a generalization applying also to non-coercive bilinear forms, where the coercivity is replaced by an *inf-sup condition*. Due to the close connection to the coercivity, indeed coercivity implies the inf-sup condition, it is also called weak coercivity. The theorem is called *Banach-Nečas-Babuška Theorem*, cf. [Bab71], and will be one of the main tools in this thesis. In addition to the well-posedness of the full space-time formulation, one also obtains explicit lower and upper bounds for the bilinear form, which will turn out to be essential for the existence of moments of the solution of parabolic PDEs with *random parameters*. Since the bounds on the solution in energy norm enter the analysis of random PDEs directly, it is important to have preferably sharp estimates. Therefore, a precise analysis of all formulations introduced before is given for different types of involved spatial differential operators $A(t)$. In this way, one aim is to obtain improved bounds, in particular with respect to more restricted spatial differential operators, like, e.g., the important class of self-adjoint and time independent spatial differential operators. These explicitly given bounds were one of the main motivations for considering a full weak formulation in space and time. Additionally, the regularity with explicit bounds for time-independent and self-adjoint operators are analyzed. Here the main interest is how the regularity of the spatial differential operator transfers to the parabolic space-time operator.

An important issue for the numerics is the stability of discretizations, ensuring that the approximations stay bounded for finer discretizations. The stability of Petrov-Galerkin discretizations is characterized by a *discrete inf-sup condition*, which is the discrete counterpart of the inf-sup condition required in the Banach-Nečas-Babuška Theorem. In the inf-sup regime, there also holds an optimality statement concerning the approximation error, namely an analog to Céa's lemma, where the discrete inf-sup constant plays the role of the coercivity constant. But contrary to the coercivity, does the inf-sup condition *not* pass to its discrete counterpart in general. That means that one has to ensure the validity of the inf-sup condition on the discrete subspaces even if the non-discrete problem is well-posed. Therefore, one has to make sure that there exists a strict lower bound $\beta > 0$ for the discrete inf-sup condition which does *not* depend on the dimension of the discrete subspaces. Otherwise a stability problem would appear, for instance, when the discrete inf-sup constant decreases with the refinement level. Such a stability problem is not a pure theoretical one, but can be observed practically even for discretizations with piecewise polynomials.

In this work, the idea is to stabilize the discretization by allowing the test space to have a *higher* dimension. This in turn yields an overdetermined system of equations. The idea to increase the degrees of freedom in the test space was inspired by [DK01] and developed for the present context of generic operator equations with focus on weak space-time formulations in my former work [Mol13b]. In order to still guarantee solvability, one aims at minimizing the residual $\frac{|b(\cdot,q_\ell)-\mathcal{F}(q_\ell)|}{\|q_\ell\|_Y}$ for all discrete test function $q_\ell$ and given test space $Y$, cf. [And13]. If a Petrov-Galerkin solution exists, then it

obviously also minimizes this residual, but not necessarily vice versa.

The stability is analyzed for a very general setting first and then specialized for space-time formulation as one model example. The stability Theorem 4.2.10, which I have worked out in [Mol13b], yields an explicit enrichment of the test space to stabilize the discretization. It will turn out that one needs to assume a slightly shifted regularity of the operator and standard Bernstein and Jackson inequalities on the discrete subspaces, which are known to hold for a broad class of functions. After having such a universal statement at hand, it will be specified how it applies to the particular situation of space-time formulations considered in this thesis. The discretization under consideration will be of tensor product type. The idea is to assume Bernstein and Jackson inequalities only with respect to the temporal and spatial component separately and work out what exactly needs to be prescribed in order to ensure the validity of the assumptions of the general theorem. In this way, one arrives at a recipe how to construct stable tensor product subspaces, cf. Table 4.3.12, 4.3.29. The construction rules apply to rather wide families of bases, in particular to B-splines on hierarchical grids as used in my implementation.

In the next part of the work, the deterministic problems considered so far are extended to parabolic PDEs with random coefficients. In order to derive more realistic models, it is often useful to introduce an additional *stochastic influence*, leading to random PDEs. Both, random PDEs as well as stochastic PDEs (SDEs), became more and more popular in the recent years. Due to the massively increasing computing power and development of more sophisticated algorithms, one can attack complicated problems with more data. Many models are meant to be simplifications of reality, describing the behavior for particular idealized situations. There are often additional influences which cannot be classified uniquely, as in quantum physics, or depend on too many events and parameters to derive a perfectly exact description. At this stage, one can use random parameters to handle the uncertainty. Stochastic influences appear directly from the theory, as in quantum physics, or can be used to model uncertainty effects due to lack of information. There are many further reasons, why uncertainties and inaccuracies can appear, as, e.g., due to measurement errors, human failures, etc. In this way, one can also enhance existing idealized models. Considering a random function as a right hand side or initial value of a PDE, allows to model certain amount of uncertainty in the given data. The uncertainty is then controlled via an additional parameter $\omega$ living in some suitable sample space. This means that the PDE is assumed to be perturbed by a random influence. The thermal diffusivity $a$ of a heat equation (1.1.1), respectively (1.1.3) in full weak formulation, can also be perturbed. This perturbation implies that the diffusion depends on the uncertainty parameter, i.e., $a = a(\omega)$. In this way, the distribution of heat in a material can be modeled, where the thermal conductivity of the material is not know exactly but only estimated, e.g., by measurements. Another example is given in [MKM13, Mol11, Mol13a], where the Schrödinger equation is perturbed by a disorder potential representing uncertainty. In this particular example, the disorder comes into play, since the physical domain of a

quantum wire is very small (nano-meter scale). Such quantum wires are impossible to produce perfectly in practice and always lead to imperfections on the surface, which are not negligible.

To be more precise, in a parabolic random PDE the right hand side, initial condition and even the involved spatial differential operator itself is assumed to depend on an additional *random parameter* introducing uncertainty. Consequently, the solution of a parabolic random PDE depends on the random parameter, such that the previous deterministic solution becomes a random field on a probability space $(\Omega, \Sigma, \mathbb{P})$. To this end, the full space-time problem (1.1.3) or (1.1.4) extends to the parameter dependent problem:

$$\text{Find } U_\omega := U(\omega) \in L_2(I;V) \cap H^1(I;V') : b_\omega(U_\omega, v) = \mathcal{F}_\omega(v) \quad \forall v \in L_2(I;V) \times H,$$

for almost every $\omega \in \Omega$ with

$$\begin{aligned} b_\omega(U_\omega, (v_1, v_2)) := & \int_I \Big( \Big\langle \frac{\mathrm{d}U_\omega(t)}{\mathrm{d}t}, v_1(t) \Big\rangle + \langle A(t,\omega)U_\omega(t), v_1(t)\rangle \Big) \, \mathrm{d}t \\ & + (U_\omega(0), v_2)_H, \end{aligned}$$

$$\mathcal{F}_\omega(v) := \int_I \langle g(t,\omega), v_1(t)\rangle \, \mathrm{d}t + (U_0(\omega), v_2)_H.$$

Parabolic random PDEs in full space-time weak formulations can be treated in the same way as the deterministic pendant (1.1.4). In particular, one needs to assume uniform bounds for the spatial differential operator with respect to every realization given by $\omega$ in order to ensure existence and uniqueness according to the Banach-Nečas-Babuška Theorem. But in this thesis, instead of assuming uniform boundedness of $A(t,\omega)$ in $\Omega$ from below or above, the bounds on $A(t,\omega)$ are assumed to be *random variables* instead. That is, the novel approach is that the lower bound is not considered to be constant, but a random variable $A_{\min} = A_{\min}(\omega)$ and the same for the upper bound $A_{\max} = A_{\max}(\omega)$. The main idea is to show existence of $p$-moments of the solution depending on the moments of $A_{\min}$ and $A_{\max}$ as well as the moments of the initial condition and right hand side. This is indeed a substantial extension since, in particular, the spatial differential operator is allowed to tend to infinity or zero, respectively, provided that a certain number of moments exist. The fundamental results concerning the existence of moments of the solution of parabolic random PDEs were worked out in my very recent work [LMM16] with Stig Larsson and Matteo Molteni. The existence of moments of the solution relies on the estimates of the norm of the solution for each realization, which is worked out in detail for the deterministic counterpart. Since the estimates are sharper the more we restrict the spatial differential operator, one gets also sharper results concerning the existence of moments. It applies to several situations not covered so far, for instance to *log-normal* coefficients $a \sim \mathcal{LN}(\mu, \sigma^2)$, which are obviously unbounded and take values arbitrary close to zero, with $A(t,\omega) := a(\omega)A(t)$. But also coefficients like $a(X) := |X|^{-\alpha}$ with uniformly distributed random variables $X$

are covered by the theory for certain values of $\alpha \in \mathbb{R}$. In addition to (almost sure) coercive spatial differential operators, I will also present a similar result for non-coercive spatial differential operators which only satisfy a Gårding inequality almost surely.

Finally, the intention is to combine the stability properties derived for the deterministic PDE with the ideas used to prove existence of moments of the solution of random PDEs. In this way, one can try to extend the *$\omega$-wise quasi-optimality* to $L_p(\Omega;\cdot)$. Ideally, stable discretizations can be constructed which do *not* depend on the particular realization, i.e., on $\omega \in \Omega$. The stability of a discretization depend on the PDE, such that it generally will depend on $\omega \in \Omega$. Unfortunately, the stability result derived before might theoretically not be optimal for the pathwise treatment. Such a behavior is not surprising since the result is consciously kept general and dependent on the operator, although the numerical results suggest that already a fixed number of extra layers would be sufficient. To this end, another stability approach from [And13] is revisited, which was designed specifically for a particular space-time weak formulation. Even if this approach is rather restrictive compared to the universal stability result in this thesis, it will turn out that a provable construction of stable discretizations independent of $\omega$ can be derived. The idea is to modify the approaches from [And13] to prove stability for a whole family of parameter dependent PDEs. It should be highlighted again that ideally the same discretization can be used for *all* random parameters (almost surely), such that there is no need to adapt it to each simulation in order to gain stability. Moreover, it allows to derive quasi-optimality of Petrov-Galerkin solutions of parabolic random PDEs with respect to $L_p(\Omega;\cdot)$ on a fixed, a-priori given, subspace. It is worth mentioning again that the spatial differential operator is not necessarily bounded uniformly neither from above nor from below.

The theoretical considerations will be substantiated by several numerical examples. In this regard, I have developed a new code in Matlab aiming at solving prototypes of Petrov-Galerkin approaches with B-splines. Splines are piecewise polynomials and B-splines form compactly supported bases of these spline spaces. Splines are known to have good approximation properties and its B-spline bases are well suited for Petrov-Galerkin approximations, not only due to the compact support. Piecewise polynomials are also the foundation of common finite elements and wavelets. The code is designed for a very broad functionality. It is deliberately kept general in order to cover many situations, in particular, the first *and* second formulation of parabolic PDEs in space-time weak form with B-spline ansatz functions of arbitrary order in each coordinate direction. Due to the stabilization presented in this thesis, one needs to enrich the test space such that the code also should be able to calculate with different levels in the test and solution space. The program is based on *B-spline tensor products* of *arbitrary order* and *arbitrary dimensions*. Furthermore, one can choose different refinement levels in each coordinate direction and independent for the solution and test space, leading to overlappings of very different supports and resulting in rectangular system matrices. Even the underlying discrete grid, respectively the knots defining the basis sets, are basically allowed to be *non-uniform* for most routines. The point evaluations of splines

are based on a *multidimensional de Boor scheme*. Due to the tensor product structure, all matrices break down to sums of Kronecker products of matrices with respect to one coordinate direction. Since the derivatives of splines can be expressed as a sum of B-splines of lower order, the stiffness matrices can be set up efficiently by using the de Boor scheme, combined with an appropriate quadrature rule depending on the order of B-spline. Since splines are by definition piecewise polynomial, many system matrices can be set up *exactly*. In this way, one can assemble matrices of very general structure with arbitrary derivatives and with different bases in solution and test space. In addition to these main routines, there are some help functions like quadrature rules or a plot routine for splines. Moreover, I have included a *graphical user interface* (GUI) for full space-time parabolic problems of arbitrary dimension, order, right hand side etc. Due to the generality of the code, its use is not restricted to (Petrov-)Galerkin approaches, but can also be used efficiently for collocation methods with point evaluations relying on the de Boor scheme. Furthermore, it is also not limited to parabolic problems, but is applicable to elliptic, Schrödinger and hyperbolic equations as well, although the theoretical considerations go beyond the scope of this thesis.

## 1.2 Outline

The thesis is structured in the following chapters.

**Chapter 2** First I introduce some fundamentals which are used throughout the thesis. I briefly give some useful properties of operators and of intersections of Hilbert spaces, which appear in parabolic problems. Beside intersection spaces I will also introduce interpolation spaces of Sobolev spaces. That means, in this regard, Sobolev spaces of fractional order, which will be used later to prove existence and uniqueness of solutions with shifted spatial regularity. The important Banach-Nečas-Babuška Theorem is also stated in this chapter. Finally, fundamental properties on B-splines are presented, which will be the building block for the Matlab-code.

**Chapter 3** This chapter is devoted to full weak formulations in space and time. I will present basically three different types of formulations, namely the first formulation, its homogenization and the second formulation. I also distinguish between coercive spatial differential operators and spatial differential operators fulfilling a Gårding-inequality. Since the explicit dependency of the spatial differential operator on the continuity constant as well as on the coercivity/Gårding constant is needed later, explicit and preferably sharp bounds are worked out. I will also present a regularity result showing how the spatial regularity inherits to the space-time weak formulation and how it influences the space-time continuity and inf-sup bounds.

**Chapter 4** After having presented space-time weak formulations of parabolic PDEs so far, I will deal in this chapter with its Petrov-Galerkin discretization, respectively the stability of Petrov-Galerkin discretizations in general. The idea is to enriching the test space in order to obtain a uniformly stable discretization. How to archive such a stable subspace will be the main work of this chapter. The rest of the chapter shows how the general strategy applies to space-time weak formulations, in particular in the second form.

**Chapter 5** In this chapter, deterministic parabolic PDEs are extended to parabolic PDEs with random coefficients. After applying the results of chapter 3 pathwise, existence of moments of the solution for possibly unbounded spatial differential operators $A(t, \omega)$ in parabolic problems is proven. Moreover, quasi-optimality of Petrov-Galerkin solutions in certain Lebesgue-Bochner spaces $L_p(\Omega; \mathcal{X})$ with respect to the stochastic parameter $\omega \in \Omega$ and deterministic solution space $\mathcal{X}$ can be derived.

**Chapter 6** Finally, this chapter gives some numerical results concerning the theoretical statements of the previous chapter. First I will consider the stability issue of Petrov-Galerkin discretizations for deterministic parabolic PDEs in space-time weak formulation. After establishing stable discretizations for deterministic problems, I focus on its stochastic counterparts from chapter 5. The calculations illustrate the main results concerning the existence of moments of the solutions and also the quasi-optimality of its Petrov-Galerkin solutions.

**Chapter 7** I conclude the body of the thesis with a brief summary of the presented results and give an outlook on future work.

**Chapter A** A documentation of the code, implemented to produce most of the numerical results, is given in the appendix. I will present the structure of the code, describe its main routines and explain the general technical implementation ideas.

# 2 Fundamentals and Preliminaries

At the very beginning we introduce the mathematical background which is indispensable for the theoretical treatments in this thesis. To this end, we need to give a few definitions and recall some known, mostly functional analytical, results. We start with general definitions, then we consider Sobolev spaces, Bochner spaces and dual spaces in more detail and give useful properties of linear operator equations. Moreover, intersection spaces and the important Banach-Nečas-Babuška Theorem as a generalization of the well known Lax-Milgram theorem are introduced. In view of arbitrary regularity results later, we also need interpolation spaces. Finally, some spline and B-spline fundamentals are presented, which are required for the implementation of the code as well as for the analysis of the numerical results.

## 2.1 General Definitions

In this section we want to recall some general vocabulary and notation, which will be used throughout this work. The content of this section can be found for instance in standard textbooks on functional analysis as, e.g., [Aub00, Rud91, DL88, DL92, Zei95a, Zei95b] and also in textbooks on partial differential equations as, e.g., [Ste10, Eva10, RR04, Wlo82] just to name a few.

**Definition 2.1.1.** *We call two norms $\|\cdot\|$ and $|||\cdot|||$ on a vector space $X$* equivalent*, if there exist constants $m, M \in \mathbb{R}$ with $0 < m \leq M < \infty$ such that*

$$m\|x\| \leq |||x||| \leq M\|x\| \quad \text{for all } x \in X.$$

*We will use the following shorthand notations:*

- $|||x||| \lesssim \|x\|$ *if there exists a constant $M > 0$ such that $|||x||| \leq M\|x\|$,*
- $|||x||| \gtrsim \|x\|$ *if there exists a constant $m > 0$ such that $|||x||| \geq m\|x\|$,*
- $|||x||| \sim \|x\|$ *if there exists constants $m, M > 0$ such that $m\|x\| \leq |||x||| \leq M\|x\|$.*

**Definition 2.1.2.** *Let $X$ and $Y$ be two normed vector spaces with norm $\|\cdot\|_X$ and $\|\cdot\|_Y$, respectively.*

*The set of all linear operators from $X$ to $Y$ with finite operator norm is denoted by $\mathcal{L}(X, Y)$. The* operator norm *is defined by*

$$\|A\|_{X\to Y} := \sup_{x\in X\setminus\{0\}} \frac{\|A(x)\|_Y}{\|x\|_X} = \sup_{\substack{x\in X,\\ \|x\|_X=1}} \|A(x)\|_Y,$$

*where one often omits the subscript if it is clear that the operator acts from $X$ to $Y$ or uses the notation $\|A\|_{\mathcal{L}(X,Y)}$.*

One generally simply writes $Ax$ instead of $A(x)$ for linear operators.

In the remainder of this section, we will consider certain function spaces in more detail. We start with the definition of dual spaces.

**Definition 2.1.3.** *Let $(X, \|\cdot\|_X)$ be a Banach space. The set of all bounded and linear mappings from $X$ into $\mathbb{R}$ is called the* dual space *of $X$ and will be denoted by*

$$X' := \mathcal{L}(X, \mathbb{R}).$$

*A norm of the dual space $X'$ is given by the operator norm*

$$\|v\|_{X'} := \sup_{x \in X \setminus \{0\}} \frac{|v(x)|}{\|x\|_X},$$

*which is called* dual norm. *Moreover, we introduce the* duality pairing *on $X' \times X$ by*

$$_{X'}\langle \cdot, \cdot \rangle_X \colon X' \times X \to \mathbb{R}, \qquad {}_{X'}\langle v, x \rangle_X := v(x),$$

*where we often omit the subscript and write $\langle \cdot, \cdot \rangle$, if the space is clear from the context.*

The notation differs slightly in the literature. The dual space is sometimes denoted by $X^*$ and the duality pairing by $\langle \cdot, \cdot \rangle_{X' \times X}$. The following well-known Riesz representation Theorem gives an important connection between Hilbert spaces and its dual space.

**Theorem 2.1.4. Riesz representation theorem**
*Let $X$ be a Hilbert space with inner product $(\cdot, \cdot)_X$. Furthermore, assume that we have a bounded linear functional $z \in X'$, then there exists a uniquely determined element $v_z \in X$ such that*

$$z(w) = (v_z, w)_X \quad \text{for all } w \in X.$$

*Moreover, each element $v \in X$ induces a linear functional $z_v \in X'$ with $\|z_v\|_{X'} = \|v\|_X$ by*

$$z_v(w) := (v, w)_X \quad \text{for all } w \in X.$$

*The mapping*

$$R_X \colon X \to X', \quad v \mapsto z_v$$

*is an* isometric isomorphism *so that one can identify each Hilbert space $X$ with its dual $X'$ and one writes $X \cong X'$. The operator $R_X$ is called* Riesz isomorphism *or Riesz mapping.*

With the aid of the Riesz representation Theorem 2.1.4, we can formulate the definition of a *Gelfand triple*, which is important for weak formulations of PDEs. Before we can define such Gelfand triples, we need the following definition.

**Definition 2.1.5.** *Let $X$ be a Banach space with norm $\|\cdot\|_X$ and $Y \subset X$ a closed subspace with norm $\|\cdot\|_Y$. Then $Y$ is* continuously embedded *in $X$ if*

$$\|v\|_X \lesssim \|v\|_Y \qquad \text{for all } v \in Y,$$

*and one writes $Y \hookrightarrow X$. Furthermore, one says that $Y$ is* densely embedded *in $X$ if in addition $Y$ is dense in $X$.*

**Definition 2.1.6.** *Let $X$ be a Hilbert space and $Y \hookrightarrow X$ a reflexive Banach space which is densely and continuously embedded in $X$. Identifying the* pivot space *$X$ with its dual $X'$ by the Riesz representation Theorem 2.1.4 yields the* Gelfand triple

$$Y \hookrightarrow X \cong X' \hookrightarrow Y'.$$

Note that if the embedding $Y \hookrightarrow X$ is dense and continuous, the embedding $X' \hookrightarrow Y'$ is dense and continuous as well. Moreover, notice that in a Gelfand triple the space $Y$ is *not* identified with its dual $Y'$, but the inner product on $X$ is used to define the duality pairing ${}_{Y'}\langle\cdot,\cdot\rangle_Y$, that is,

$${}_{Y'}\langle z, y\rangle_Y = {}_{X'}\langle z, y\rangle_X := (z, y)_X, \qquad \text{if } \ z \in X \cong X' \subset Y', \ y \in Y$$

where $(\cdot,\cdot)_X$ denotes the inner product on $X$. Typical choices are $Y = H^m(D)$ and $X = L_2(D)$ with respect to an open domain $D \subset \mathbb{R}^n$ with possibly additional boundary conditions. We conclude with an extension of the well-known Hölder inequality, which is used frequently in chapter 5 to prove existence of moments of solutions of random PDEs.

**Theorem 2.1.7.** *Assume that $f_1, \dots, f_m$ are functions on domain $D \subset \mathbb{R}^n$ such that*

$$f_i \in L_{p_i}(D), \quad 1 \le i \le m, \quad \text{with} \quad \frac{1}{p} = \frac{1}{p_1} + \frac{1}{p_2} + \cdots \frac{1}{p_m} \le 1.$$

*Then the product $f = f_1 f_2 \cdots f_m$ belongs to $L_p(D)$ with*

$$\|f\|_{L_p(D)} \le \|f_1\|_{L_{p_1}(D)} \|f_2\|_{L_{p_2}(D)} \cdots \|f_m\|_{L_{p_m}(D)}.$$

*Proof.* A prove of this general form can be found in [Bre11, Th. 4.6/Remark 2]. □

## 2.2 Sobolev and Bochner Spaces

This section introduces Sobolev spaces and their dual spaces as well as Bochner spaces. The solution and test spaces for partial differential equations in variational formulation are typically Sobolev spaces. Usually, operators implied by the variational formulation are arranged to map from one Sobolev space into a dual Sobolev space. Suitable solution and test spaces of space-time weak formulations of parabolic evolution problems

are Bochner spaces or intersections of them. Bochner spaces can be interpreted as generalized Sobolev spaces in some sense or as tensor product Sobolev spaces as we will illustrate in this section. Since Sobolev spaces are very standard in the present framework of variational problems, we will only give their definition here. A very substantial treatment of Sobolev spaces is given in [AF03]. For more details and further properties of Bochner spaces and vector valued Sobolev spaces we refer to [Aub00, DL92, ABHN01, Wlo82].

**Definition 2.2.1.** *Let $f \in L_2(D)$ with $D \subset \mathbb{R}^n$ open and $\alpha \in \mathbb{N}_0^n$. Suppose that there is an element $g \in L_2(D)$ such that*

$$\int_D g(\mathbf{x})\phi(\mathbf{x})\ \mathrm{d}\mathbf{x} = (-1)^{|\alpha|} \int_D f(\mathbf{x}) \frac{\partial^{|\alpha|}}{\partial \mathbf{x}^\alpha}\phi(\mathbf{x})\ \mathrm{d}\mathbf{x} \quad \text{for all } \phi \in \mathcal{C}_0^\infty(D),$$

*where the integrals are considered as Lebesgue integrals, then $g$ is the $\alpha$-th* weak derivative *of $f$ in the $L_2(D)$ sense and one defines $\frac{\partial^{|\alpha|}}{\partial \mathbf{x}^\alpha} f := g \in L_2(D)$. We denote by $\mathcal{C}_0^\infty(D)$ the space of infinitely often differentiable functions with compact support in $D$. Here we have used the multiindex notation*

$$\frac{\partial^{|\alpha|}}{\partial \mathbf{x}^\alpha} f(\mathbf{x}) := \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}} f(\mathbf{x}),$$

*with $\mathbf{x} := (x_1, \ldots, x_n)$, $\alpha := (a_1, \ldots, a_n)$ and $|\alpha| := \sum_{i=1}^n a_i$.*

**Definition 2.2.2.** *The Hilbert space of elements $f \in L_2(D)$ for which the weak derivatives $D^\alpha f := \frac{\partial^{|\alpha|}}{\partial \mathbf{x}^\alpha} f \in L_2(D)$ according to Definition 2.2.1 exist for all $|\alpha| \leq m \in \mathbb{N}$ is called* Sobolev space *of order $m$ and will be denoted by $H^m(D)$. The Sobolev space $H^m(D)$ is equipped with the inner product*

$$\begin{aligned}(f,g)_{H^m(D)} &:= \sum_{|\alpha|\leq m} \Big(\frac{\partial^{|\alpha|}}{\partial \mathbf{x}^\alpha} f, \frac{\partial^{|\alpha|}}{\partial \mathbf{x}^\alpha} g\Big)_{L_2(D)} \\ &= \sum_{|\alpha|\leq m} \int_D \left(\frac{\partial^{|\alpha|}}{\partial \mathbf{x}^\alpha} f(\mathbf{x})\right)\left(\frac{\partial^{|\alpha|}}{\partial \mathbf{x}^\alpha} g(\mathbf{x})\right)\ \mathrm{d}\mathbf{x},\end{aligned}$$

*and the induced norm*

$$\|f\|_{H^m(D)} := \sqrt{(f,f)_{H^m(D)}} = \sqrt{\sum_{|\alpha|\leq m} \left\| \frac{\partial^{|\alpha|}}{\partial \mathbf{x}^\alpha} f \right\|_{L_2(D)}^2}.$$

*A seminorm is given by*

$$|f|_{H^m(D)} := \sqrt{\sum_{|\alpha|= m} \left\| \frac{\partial^{|\alpha|}}{\partial \mathbf{x}^\alpha} f \right\|_{L_2(D)}^2}.$$

*The dual space $(H^m(D))'$ of a Sobolev space $H^m(D)$ is usually denoted by $\dot{H}^{-m}(D)$.*

The Sobolev space $H_0^m(D)$ with zero boundary conditions is defined as the closure of $\mathcal{C}_0^\infty(D)$ with respect to the norm $\|\cdot\|_{H^m(D)}$ on $H^m(D)$, i.e.,

$$H_0^m(D) = \overline{\mathcal{C}_0^\infty(D)}^{\|\cdot\|_{H^m(D)}}. \tag{2.2.3}$$

The dual space $(H_0^m(D))'$ is denoted by $H^{-m}(D)$. The definition of Sobolev spaces according to Definition 2.2.2 is only meaningful for integer exponents $m \in \mathbb{N}$, respectively $m \in \mathbb{Z}$ due to the notation of its dual spaces. But Sobolev spaces can also be extended to a scale of spaces with real valued smoothness indices, which will be important later. We will state the definition along the lines of [AF03, Ch. VII] and [Pab15, sec. 1] at this point and explain the role of fractional Sobolev spaces later, when we deal with interpolation spaces.

**Definition 2.2.4.** *For $f \in L_2(\mathbb{R}^n)$ the* Fourier transform $\mathcal{F}(f)$ *can be defined with a density argument as*

$$\mathcal{F}(f)(\mathbf{x}) := \lim_{R\to\infty} \int_{\|\xi\|_{\ell_2(\mathbb{R}^n)}\leq R} f(\xi)\exp(-2\pi \mathrm{i}\mathbf{x}\cdot\xi)\ \mathrm{d}\xi,$$

*in the $L_2$-sense, with imaginary unit* i *and inner product $\mathbf{x}\cdot\xi$. For every $s \in \mathbb{R}$ we define the* Sobolev space of order $s$ *as*

$$H^s(\mathbb{R}^n) := \{v : (1+|\cdot|^2)^{s/2}\mathcal{F}(v) \in L_2(\mathbb{R}^n)\},$$

*with inner product*

$$(u,v)_{H^s(\mathbb{R}^n)} := ((1+|\cdot|^2)^{s/2}\mathcal{F}(u), (1+|\cdot|^2)^{s/2}\mathcal{F}(v))_{L_2(\mathbb{R}^n)}.$$

These spaces are sometimes called *Bessel potential function spaces*, cf. [Ste10, sec. 2.4]. Due to the Plancherel theorem, the Fourier transform $\mathcal{F}\colon L_2(\mathbb{R}^n) \to L_2(\mathbb{R}^n)$ is a unitary operator, i.e.,

$$\|\mathcal{F}(\cdot)\|_{L_2(\mathbb{R}^n)} = \|\cdot\|_{L_2(\mathbb{R}^n)}.$$

Moreover, it can be shown that

$$\mathcal{F}\left(D^\alpha f\right)(\mathbf{x}) = (2\pi\mathrm{i})^{|\alpha|}\mathbf{x}^\alpha \mathcal{F}(f), \qquad \alpha \in \mathbb{R}^n,\ \mathbf{x}\in\mathbb{R}^n$$

and also that the two characterizations $(\cdot)^\alpha\mathcal{F}(f) \in L_2(\mathbb{R}^n)$ and $(1+|\cdot|^2)^{s/2}\mathcal{F}(f) \in L_2(\mathbb{R}^n)$ are equivalent for $|\alpha| \leq s$ with integer $s \in \mathbb{N}$, so that Definition 2.2.4 is consistent with Definition 2.2.2 for integer orders $s = m \in \mathbb{N}$. There are several different ways to define fractional order Sobolev spaces which are not necessarily equivalent on bounded domains, as ,e.g., Sobolev-Slobodeckij spaces, see [Ste10, sec. 2.3].

Sobolev spaces form the basis for Bochner spaces which come into play when dealing, e.g., with evolution problems due to the additional time variable.

**Definition 2.2.5.** *For a given Banach space $X$ with norm $\|\cdot\|_X$, the* Bochner space *$L_2(I;X)$ on a domain $I \subset \mathbb{R}^n$ is the space of all strongly measurable functions $f\colon I \to X$ such that the norm*

$$\|f\|^2_{L_2(I;X)} := \int_I \|f(t)\|^2_X \, \mathrm{d}t \tag{2.2.6}$$

*is finite. Analogously, we define the* vector valued Sobolev space *$H^m(I;X)$ with $m \geq 0$ as the space of all functions $f \in L_2(I;X)$ with weak derivatives $D^\alpha f \in L_2(I;X)$ for all $|\alpha| \leq m$. A norm on $H^m(I;X)$ is given by*

$$\|f\|_{H^m(I;X)} := \left( \sum_{|\alpha| \leq m} \|D^\alpha f\|^2_{L_2(I;X)} \right)^{\frac{1}{2}}.$$

The weak derivatives $D^\alpha f$ are defined similar as in the Sobolev case from Definition 2.2.1, as the element $g \in L_2(I;X)$ such that

$$\int_I g(t)\phi(t) \, \mathrm{d}t = (-1)^{|\alpha|} \int_I f(t) \frac{\partial^{|\alpha|}}{\partial t^\alpha} \phi(t) \, \mathrm{d}t \in X \quad \text{for all } \phi \in \mathcal{C}_0^\infty(I).$$

Typically, the Banach space $X$ in Definition 2.2.5 is a Sobolev space $X = H^m(D)$, with $m \in \mathbb{R}$ and domain $D \subset \mathbb{R}^n$. Due to the definition of Bochner spaces, the elements $f \in L_2(I; H^m(D))$ are functions defined on the domain $I$ with values in $H^m(D)$, what means that $f(t)\colon D \to \mathbb{R}$ are functions itself for all $t \in I$, now defined on the domain $D$. Considering a function $g\colon I \times D \to \mathbb{R}$ which is $m$-times weak differentiable with respect to the coordinates on $D$, then $g(t,\cdot) \in H^m(D)$ for $t \in I$. Due to this connection we define $g(t,\mathbf{x}) := g(t)(\mathbf{x})$, for $t \in I$ and $\mathbf{x} \in D$. The denotation Bochner space is often used synonymously also for vector valued Sobolev spaces. For Hilbert spaces $X$ with inner product $(\cdot,\cdot)_X$ an inner product inducing the norm (2.2.6) on the Bochner space is given by

$$(f,g)_{L_2(I;X)} := \int_I (f(t), g(t))_X \, \mathrm{d}t, \qquad f, g \in L_2(I;X) \tag{2.2.7}$$

and similar for vector valued Sobolev spaces. Vector valued Sobolev spaces as well as Bochner spaces are Hilbert spaces with respect to this inner product (2.2.7), see [LM72, Rem. 1.5] or [Wlo82, Prop. 24.5]. As already mentioned before, vector valued Sobolev spaces and Bochner spaces can be identified via Hilbert tensor products.

**Proposition 2.2.8.** *For a separable Hilbert space $X$, we can identify $H^m(I;X) \cong H^m(I) \otimes X$ isometrically.*

*Proof.* See [Aub00, Th. 12.7.1]. □

For $f_1 \in X$ and $f_2 \in Y$, we define $f := f_1 \otimes f_2 \in X \otimes Y$, with $f(t)(x) := f(t,x) := f_1(t)f_2(x)$, and equip the tensor product space $X \otimes Y$ with the inner product

$$(f_1 \otimes f_2, g_1 \otimes g_2)_{X\otimes Y} := (f_1, g_1)_X (f_2, g_2)_Y. \tag{2.2.9}$$

It is not hard to see that the inner product defined in (2.2.7) coincides with the inner product defined in (2.2.9) for $f_1 \otimes f_2,\ g_1 \otimes g_2 \in L_2(I) \otimes X \cong L_2(I;X)$:

$$\begin{aligned}(f_1 \otimes f_2, g_1 \otimes g_2)_{L_2(I;X)} &= \int_I (f_1(t)f_2(\cdot), g_1(t)g_2(\cdot))_X \ \mathrm{d}t \\ &= \left(\int_I f_1(t)g_1(t)\ \mathrm{d}t\right)(f_2, g_2)_X \\ &= (f_1, g_1)_{L_2(I)}(f_2, g_2)_X \\ &= (f_1 \otimes f_2, g_1 \otimes g_2)_{L_2(I)\otimes X}.\end{aligned}$$

The mathematically precise treatment of Hilbert tensor products via Hilbert-Schmidt operators is rather technical. We refer to [Aub00, Pis03] for a detailed description.

In the context of, e.g., parabolic problems, one additionally considers intersections of Hilbert spaces.

**Definition 2.2.10.** *Let $X$ and $Y$ be two Hilbert spaces with norms $\|\cdot\|_X$ and $\|\cdot\|_Y$. Then, the* intersection space *is defined as*

$$X \cap Y := \{f \in X, Y : \|f\|^2_{X\cap Y} := \|f\|^2_X + \|f\|^2_Y < \infty\}. \tag{2.2.11}$$

The denotation *intersection* is reasonable, since the definition of the norm (2.2.11) implies that elements from the intersection space $X \cap Y$ are contained in both component spaces, $X$ and $Y$, and vice versa. Moreover, the intersection space is continuously embedded in both component spaces and also defines a Hilbert space with inner product

$$(f, g)_{X\cap Y} := (f, g)_X + (f, g)_Y, \qquad f, g \in X \cap Y,$$

where $(\cdot,\cdot)_X$ and $(\cdot,\cdot)_Y$ are inner products on $X$ and $Y$, respectively, cf. [AF03, sec. 7.7] for an equivalent norm. A very prominent example of intersection spaces is of the form $L_2(I;V) \cap H^1(I;\tilde{V})$ on an interval $I$ with two separable Hilbert spaces $V$ and $\tilde{V}$, where one often considers $\tilde{V} := V'$. These type of function spaces are used in chapter 3 in order to state well-posed parabolic PDEs in an appropriate way.

Before we can give an important embedding theorem, which is crucial for well-posedness of parabolic PDEs in space-time weak formulations, we briefly introduce *interpolation spaces* or intermediate spaces. As the name already suggests, interpolation spaces describe spaces which lie "in between" two spaces in a certain manner, which has to be defined properly. Since the mathematically precise definition of these spaces is rather technical, we first want to give an important example of interpolation spaces of Sobolev

spaces. For two Sobolev spaces $H^\alpha(\mathbb{R}^n) \subset H^\beta(\mathbb{R}^n)$, the interpolation space of order $\frac{1}{2}$ coincides with

$$H^{\frac{\alpha+\beta}{2}}(\mathbb{R}^n) = [H^\alpha(\mathbb{R}^n), H^\beta(\mathbb{R}^n)]_{1/2}. \tag{2.2.12}$$

The spaces (2.2.12) are well defined for $\frac{\alpha+\beta}{2} \in \mathbb{Z}$ according to Definition 2.2.2, but also for fractional orders via Definition 2.2.4. We see that, for instance, the interpolation of order $\frac{1}{2}$ between a Sobolev space and its dual is the pivot space $L_2(\mathbb{R}^n)$, which agrees with the suggestion that pivot spaces lie "exactly between" pairs of primal and dual Sobolev spaces. We orientate ourselves at [DL90, Ch. VIII, §3, Def. 8] and give a precise but not too technical description.

**Definition 2.2.13.** *Let $V \hookrightarrow \tilde{V}$ be separable Hilbert spaces with dense and continuous embedding equipped with inner products $(\cdot,\cdot)_V$ and $(\cdot,\cdot)_{\tilde{V}}$, respectively. Then the inner product $(\cdot,\cdot)_V$ defines a unique unbounded operator $A$ in $\tilde{V}$ with domain*

$$D(A) = \{u \in V : v \mapsto (u,v)_V \text{ is continuous on } V \text{ for the topology on } \tilde{V}\}, \tag{2.2.14}$$

*satisfying*

$$(Au,u)_{\tilde{V}} := (u,u)_V, \qquad \text{for all } u \in D(A),$$

*cf. also [DL88, Ch. VI/VII]. Using the spectral decomposition of $A$ in $\tilde{V}$, one can define roots of $A$ and we set $\Lambda := A^{1/2}$. For $\theta \in [0,1]$ we can now define the* interpolation of order $\theta$ between $V$ and $\tilde{V}$ *as*

$$[V,\tilde{V}]_\theta := D(\Lambda^{1-\theta}),$$

*with inner product*

$$(u,v)_\theta := (u,v)_{\tilde{V}} + (\Lambda^{1-\theta}u, \Lambda^{1-\theta}v)_{\tilde{V}}.$$

It can be seen that we obtain in particular

$$[V,\tilde{V}]_0 = V \qquad \text{and} \qquad [V,\tilde{V}]_1 = \tilde{V}.$$

Note that (2.2.14) comes into play in order to ensure that the composition of Riesz mappings

$$A := R_{\tilde{V}}^{-1} R_V \colon D(A) \to \tilde{V},$$

is well-defined, where $R_V \colon V \to V'$ and $R_{\tilde{V}} \colon \tilde{V} \to \tilde{V}'$ are the Riesz mappings with respect to $V$ and $\tilde{V}$, respectively, according to the Riesz representation Theorem 2.1.4. This is the case for the set $\{u \in V : R_V u \in \tilde{V}' \subset V'\} = R_V^{-1}(\tilde{V}')$, which is equivalent to the set (2.2.14). One can relate the interpolation spaces between two Sobolev spaces with other (fractional order) Sobolev spaces as in example (2.2.12).

**Theorem 2.2.15.** *Let $\theta \in [0,1]$, $\alpha, \beta \in \mathbb{R}^+$, $\alpha \neq \beta$. Then the following identity holds*

$$[H^\alpha(\mathbb{R}^n), H^\beta(\mathbb{R}^n)]_\theta = H^{(1-\theta)\alpha+\theta\beta}(\mathbb{R}^n). \tag{2.2.16}$$

*Proof.* See [BS88, Ch. 5, Theorem 4.17] and [BL76, 6.4.5. Theorem]. □

One can also use (2.2.16) directly to define Sobolev spaces of fractional order instead of using Definition 2.2.4. In order to cover also Sobolev spaces on arbitrary domains, we define

$$H^s(D) := [H^m(D), L_2(D)]_\theta, \qquad s = (1-\theta)m,\ m \in \mathbb{N},\ 0 \le \theta \le 1, \tag{2.2.17}$$

according to [DL90, Example 4]. Due to Theorem 2.2.15 we deduce that definition (2.2.17) is consistent with Definition 2.2.4 on $\mathbb{R}^n$. Moreover, it is not hard to show that (2.2.17) yields the same Sobolev spaces as Definition 2.2.2 when $s \in \mathbb{N}^+$. It is worth mentioning that Theorem 2.2.15 stays also true for sufficiently regular domains $D \subset \mathbb{R}^n$, when we define the Sobolev spaces of fractional order according to Definition 2.2.4 restricted to $D$, cf. [AF03, Ch VII, 7.66 ]. With the previous Definition 2.2.13 we can formulate the important embedding theorem from [DL92, Ch. XVIII, §1, Th. 1 and Sec. 3].

**Theorem 2.2.18.** *For every interval $I \subset \mathbb{R}$ and separable Hilbert spaces $V \hookrightarrow \tilde{V}$ with dense and continuous embedding, every $v \in L_2(I;V) \cap H^1(I;\tilde{V})$ is almost everywhere equal to a continuous function of $\overline{I}$ in $H := [V,\tilde{V}]_{1/2}$. Further, we have*

$$L_2(I;V) \cap H^1(I;\tilde{V}) \hookrightarrow \mathcal{C}^0(\overline{I};H),$$

*with continuous embedding.*

In particular, Theorem 2.2.18 implies that point evaluations on the interval $\overline{I}$ exist for functions from the intersection space $L_2(I;V) \cap H^1(I;\tilde{V})$. This fact will turn out to allow point evaluations at given initial times in the context of parabolic PDEs, which is crucial for the well-posedness of its full space-time weak formulation.

## 2.3 Operator Equations and Properties

In this section we briefly introduce some operator theoretical aspects, which will be needed in the present work.

First, we will introduce the concept of dual operators.

**Definition 2.3.1.** *Let $X, Y$ be two Banach spaces and $A \in \mathcal{L}(X,Y)$. Then the dual operator $A'\colon Y' \to X'$ is defined via*

$$_{X'}\langle A'\tilde{v}, w\rangle_X := {}_{Y'}\langle \tilde{v}, Aw\rangle_Y \quad \text{for all } w \in X, \tilde{v} \in Y'.$$

It is well known that dual operators fulfill

$$A' \in \mathcal{L}(Y',X') \quad \text{with } \|A'\| = \|A\|,$$

see, e.g., [Wlo82, sec. 12.1]. Assuming $X$ and $Y$ to be reflexive Banach spaces, the dual operator of $A \in \mathcal{L}(X, Y')$ is consequently given by

$$A' \in \mathcal{L}(Y, X'), \qquad {}_{X'}\langle A'v, w\rangle_X := {}_Y\langle v, Aw\rangle_{Y'} \quad \text{for all } w \in X, v \in Y,$$

where $Y$ and its bidual $Y''$ are isometrically isomorph and we identify $Y'' \cong Y$. Bounded bilinear forms $a(\cdot,\cdot)$ on Hilbert spaces $X$ and $Y$ uniquely define operators $A \in \mathcal{L}(X, Y')$ by

$${}_{Y'}\langle Av, w\rangle_Y := a(v, w), \qquad v \in X,\ w \in Y, \tag{2.3.2}$$

cf. [Wlo82, Prop. 17.8].

The following Theorem gives a general criterion for the existence and uniqueness of solutions of linear operator equations. It is well-known (see, e.g., [Bab71, Th 2.1], [NSV09, Th. 2.1]) that an operator $A \in \mathcal{L}(X, Y')$ is boundedly invertible if and only if the operator is bounded, surjective and if an inf-sup condition is fulfilled. This is formulated in the following Theorem 2.3.3

**Theorem 2.3.3.** *Let $X$ and $Y$ be two Hilbert spaces with norm $\|\cdot\|_X$ and $\|\cdot\|_Y$, respectively. A linear operator $A \in \mathcal{L}(X, Y')$ is boundedly invertible if and only if*

$$\sup_{v\in X\setminus\{0\}} \sup_{w\in Y\setminus\{0\}} \frac{|\langle Av, w\rangle|}{\|v\|_X\|w\|_Y} < \infty \qquad \textit{(boundedness)} \tag{2.3.4}$$

$$\inf_{v\in X\setminus\{0\}} \sup_{w\in Y\setminus\{0\}} \frac{|\langle Av, w\rangle|}{\|v\|_X\|w\|_Y} > 0 \qquad \textit{(inf-sup condition)} \tag{2.3.5}$$

$$\sup_{v\in X\setminus\{0\}} |\langle Av, w\rangle| > 0 \quad \forall w \in Y \setminus \{0\} \qquad \textit{(surjectivity).} \tag{2.3.6}$$

Obviously, the boundedness (2.3.4) is already incorporated in the Definition 2.1.2 of $\mathcal{L}(X, Y')$. Notice that linear operators on Hilbert spaces are bounded if and only if they are continuous. To this end, the constant (2.3.4) is usually referred to as continuity constant. The formulation in [Bab71, Th. 2.1] differs slightly from our formulation, but by taking trivial reformulations of the assumptions, linear forms and dual spaces, one can straightforwardly derive Theorem 2.3.3 from [Bab71, Th. 2.1]. Theorem 2.3.3 is often called Babuška Theorem, Babuška-Lax-Milgram Theorem, *Banach-Nečas-Babuška (BNB) Theorem* or generalized Lax-Milgram Theorem. This BNB Theorem is the building block for our analysis of well-posedness of parabolic PDEs in space-time weak formulations and is therefore also important for the results on the existence of moments of solutions of random PDEs in chapter 5. It can be shown that the following equivalence holds.

**Proposition 2.3.7.** *Let $A \in \mathcal{L}(X, Y')$. Then the following conditions are equivalent:*

*(i)*

$$\inf_{v\in X\setminus\{0\}} \sup_{w\in Y\setminus\{0\}} \frac{|\langle Av, w\rangle|}{\|v\|_X\|w\|_Y} > 0,$$
$$\sup_{v\in X\setminus\{0\}} |\langle Av, w\rangle| > 0 \quad \forall w \in Y \setminus \{0\}.$$

*(ii)*
$$\inf_{w\in Y\setminus\{0\}} \sup_{v\in X\setminus\{0\}} \frac{|\langle Av, w\rangle|}{\|v\|_X\|w\|_Y} > 0,$$
$$\sup_{w\in Y\setminus\{0\}} |\langle Av, w\rangle| > 0 \quad \forall v \in X \setminus \{0\}.$$

*(iii)*
$$\inf_{v\in X\setminus\{0\}} \sup_{w\in Y\setminus\{0\}} \frac{|\langle Av, w\rangle|}{\|v\|_X\|w\|_Y} = \inf_{w\in Y\setminus\{0\}} \sup_{v\in X\setminus\{0\}} \frac{|\langle Av, w\rangle|}{\|v\|_X\|w\|_Y}.$$

*Proof.* The equivalence between $(i)$ and $(iii)$ is proven in [NSV09, Theorem 2.2] and is a consequence of the fact that $\|(A')^{-1}\| = \|A^{-1}\|$, with $Y'' \cong Y$ since Hilbert spaces are reflexive. The equivalence between $(ii)$ and $(iii)$ follows by considering the dual $A' \in \mathcal{L}(Y, X')$. $(iii)$ is obviously equivalent to

$$\inf_{w\in Y\setminus\{0\}} \sup_{v\in X\setminus\{0\}} \frac{|\langle v, A'w\rangle|}{\|v\|_X\|w\|_Y} = \inf_{v\in X\setminus\{0\}} \sup_{w\in Y\setminus\{0\}} \frac{|\langle v, A'w\rangle|}{\|v\|_X\|w\|_Y},$$

which in turn is equivalent to $(i)$ with respect to the dual operator $A'$, i.e.,

$$\inf_{w\in Y\setminus\{0\}} \sup_{v\in X\setminus\{0\}} \frac{|\langle v, A'w\rangle|}{\|v\|_X\|w\|_Y} > 0,$$
$$\sup_{w\in Y\setminus\{0\}} |\langle v, A'w\rangle| > 0 \quad \forall v \in X \setminus \{0\}.$$

□

This property is very useful, since it allows to swap the arguments of the infimum and supremum. The next corollary associates the operator norms of $A$ and $A^{-1}$ with the continuity and inf-sup constants form Theorem 2.3.3.

**Corollary 2.3.8.** *Let $X$ and $Y$ be two Hilbert-spaces with norm $\|\cdot\|_X$ and $\|\cdot\|_Y$, respectively, and assume that $A \in \mathcal{L}(X, Y')$ fulfills the three conditions* (2.3.4)*,* (2.3.5) *and* (2.3.6) *of the Babuška Theorem 2.3.3. Then, besides the bounded invertibility, the operator norms are given by*

$$\begin{aligned} \|A\|_{X\to Y'} &= \sup_{v\in X\setminus\{0\}} \sup_{w\in Y\setminus\{0\}} \frac{|\langle Av, w\rangle|}{\|v\|_X\|w\|_Y} =: A_{\max} \\ \|A^{-1}\|_{Y'\to X} &= \left( \inf_{v\in X\setminus\{0\}} \sup_{w\in Y\setminus\{0\}} \frac{|\langle Av, w\rangle|}{\|v\|_X\|w\|_Y} \right)^{-1} =: A_{\min}^{-1}. \end{aligned} \tag{2.3.9}$$

*Proof.* These identities follow directly from the definition of the operator norm

$$\|A\|_{X\to Y'} := \sup_{v\in X\setminus\{0\}} \frac{\|Av\|_{Y'}}{\|v\|_X} = \sup_{v\in X\setminus\{0\}} \sup_{w\in Y\setminus\{0\}} \frac{|\langle Av, w\rangle|}{\|v\|_X\|w\|_Y},$$

and

$$\begin{aligned}\|A^{-1}\|_{Y'\to X} &:= \sup_{\tilde{w}\in Y'\setminus\{0\}} \frac{\|A^{-1}\tilde{w}\|_X}{\|\tilde{w}\|_{Y'}} = \left(\inf_{\tilde{w}\in Y'\setminus\{0\}} \frac{\|\tilde{w}\|_{Y'}}{\|A^{-1}\tilde{w}\|_X}\right)^{-1} \\ &= \left(\inf_{Av\in Y'\setminus\{0\}} \frac{\|Av\|_{Y'}}{\|v\|_X}\right)^{-1} = \left(\inf_{v\in X\setminus\{0\}} \sup_{w\in Y\setminus\{0\}} \frac{|\langle Av, w\rangle|}{\|v\|_X\|w\|_Y}\right)^{-1},\end{aligned}$$

where we have exploited the surjectivity. □

Another important consequence is a stability result for the solution of the corresponding operator equation.

**Corollary 2.3.10.** *Let $X$ and $Y$ be two Hilbert-spaces with norm $\|\cdot\|_X$ and $\|\cdot\|_Y$, respectively, and assume that $A \in \mathcal{L}(X, Y')$ fulfills the three conditions* (2.3.4)*,* (2.3.5) *and* (2.3.6) *of the Babuška Theorem 2.3.3. Considering the operator equation*

$$Au = f, \qquad u \in X,\ f \in Y',$$

*the solution $u$ is bounded by*

$$\|u\|_X \le A_{\min}^{-1}\, \|f\|_{Y'},$$

*with inf-sup constant*

$$A_{\min} := \inf_{v\in X\setminus\{0\}} \sup_{w\in Y\setminus\{0\}} \frac{|\langle Av, w\rangle|}{\|v\|_X\|w\|_Y}.$$

*Proof.* The proof follows directly by equality (2.3.9) and BNB Theorem 2.3.3. □

## 2.4 B-Splines

We conclude this chapter with some essential properties on splines and B-splines. We refer to [dB01, Sch07, HÖ3, Sch15] for a detailed treatment of splines. In connection with [Sch15] there is also a program package called `SplinePak` for spline interpolation available.

The code implemented for this thesis relies heavily on B-splines and is used for most of the numerical results in chapter 6. It is designed for Petrov-Galerkin approximations of PDEs, respectively minimal residual Petrov-Galerkin approximations, and does not focus on interpolations of given functions itself. First of all, we need to specify how we define *splines*.

**Definition 2.4.1.** *Let $\Delta := \{x_i\}_{i=0,\dots,M+1}$ be a (not necessarily uniform) grid on the interval $[a,b]$ with pairwise disjoint knots such that*

$$a =: x_0 < x_1 < \ldots < x_{M+1} := b.$$

*Then we define the* space of splines of order $k \in \mathbb{N}$ *with respect to $\Delta$ as*

$$SP_{\Delta,k} := \{S \in \mathcal{C}^{k-2}([a,b]) : S|_{[x_i,x_{i+1})} \in \Pi_k,\ i = 0,\dots,M\},$$

*where $\Pi_k$ denotes the space of polynomials of order $k \in \mathbb{N}$, i.e., of degree at most $k-1$.*

That means that we define the spline space as the space of piecewise polynomials of order $k$ with global smoothness $k-2$. It is the smoothest space with piecewise polynomials of order $k$ without degenerating to a single polynomial. One could also require another global smoothness, even vary the smoothness at each knot, but we restrict ourselves here to this particular situation. The splines defined in the way of Definition 2.4.1 are sometimes called the splines of order $k$ with *simple knots*, cf. [Sch07, Example 4.3]. We can see by counting the degrees of freedom that $\dim SP_{\Delta,k} = k+M$.

There are different bases for the space $SP_{\Delta,k}$, but since we are using them as ansatz functions in a Petrov-Galerkin discretization, we would like to have bases functions with *local support*. To this end, we consider B-splines.

**Proposition 2.4.2.** *Let $T := \{\theta_i\}_{i=1,\dots,N+k}$ be the extended sequence of knots with fixed $k \in \mathbb{N}$ such that*

$$\theta_1 = \ldots = \theta_k = a < \theta_{k+1} < \ldots < \theta_N < b = \theta_{N+1} = \ldots = \theta_{N+k}.$$

*Then we define the B-splines $N_{i,k}(x)$ of order $k$ with respect to $\theta_i,\dots,\theta_{i+k}$ for $i = 1,\dots,N$ recursively as*

$$\begin{aligned} N_{i,1}(x) &:= \begin{cases} 1 & \text{for } x \in [\theta_i, \theta_{i+1}) \\ 0 & \text{else} \end{cases} \\ N_{i,k}(x) &:= \tfrac{x-\theta_i}{\theta_{i+k-1}-\theta_i} N_{i,k-1}(x) + \tfrac{\theta_{i+k}-x}{\theta_{i+k}-\theta_{i+1}} N_{i+1,k-1}(x), \end{aligned} \tag{2.4.3}$$

*for $\theta_{i+k-1} \neq \theta_i$ and $\theta_{i+k} \neq \theta_{i+1}$. Since $N_{i,1} \equiv 0$ if $\theta_i = \theta_{i+1}$, we set the quotients in (2.4.3) to zero when the knots in the denominator coincide. The set of B-splines is a basis for*

$$SP_{\Delta_T,k} = \operatorname{span}\{N_{i,k} : i = 1,\dots,N\},$$

*with respect to the pairwise disjoint* inner knots *of $T$, i.e., $\Delta_T := \{\theta_i\}_{i=k,\dots,N+1}$.*

These bases obviously have local support, are non-negative and are piecewise polynomial of order $k$. It is worth mentioning that one can control the global smoothness, respectively the smoothness at each knot, by inserting additional coinciding inner knots.

In this way, one can also construct bases with local support for more general spaces of splines as pointed out above. These B-spline bases serve as the centerpiece of the program code developed for this thesis. In the following, we will state some very useful basic properties which are also inevitable for an efficient implementation. Since we deal with piecewise polynomials which are given recursively, one can express the derivative also recursively and with respect to piecewise polynomials of less degree.

**Corollary 2.4.4.** *Let $N_{i,k}$ be a B-spline with $i \in \{1, \dots, N\}$, $k \in \mathbb{N}$ with respect to the extended sequence of knots as in Proposition 2.4.2. Then the derivatives are given recursively by*

$$N'_{i,k}(x) = (k-1)\left[\frac{N_{i,k-1}(x)}{\theta_{i+k-1} - \theta_i} - \frac{N_{i+1,k-1}(x)}{\theta_{i+k} - \theta_{i+1}}\right],$$

*where we again set the quotients to zero when there are two coinciding knots in the denominator.*

Due to the recursive definition based on the characteristic function $N_{i,1}$ and the local support, point evaluations of a given spline expressed with respect to B-splines can be computed very efficiently with the following scheme, cf. [Sch07, Th. 5.7].

**Theorem 2.4.5.** *Let*

$$S(x) = \sum_{i=1}^{N} c_i N_{i,k}(x), \qquad x \in [a, b),$$

*with respect to an extended sequence of knots as in Proposition 2.4.2. Then the spline can be expressed equivalently as*

$$S(x) = \sum_{i=r+1}^{N} c_i^{[r]}(x) N_{i,k-r}(x),$$

*for $0 \le r \le k-1$, where*

$$c_i^{[r]}(x) := \begin{cases} c_i & \text{if } r = 0, \\ \frac{x-\theta_i}{\theta_{i+k-r}-\theta_i} c_i^{[r-1]}(x) + \frac{\theta_{i+k-r}-x}{\theta_{i+k-r}-\theta_i} c_{i-1}^{[r-1]}(x) & \text{if } r > 0, \\ 0 & \text{if } \theta_{i+k-r} = \theta_i \end{cases}.$$

*In particular, if $\theta_i \le \bar{x} < \theta_{i+1}$, then*

$$S(\bar{x}) = c_i^{[k-1]}(\bar{x}).$$

*The calculation can be performed with a triangle, Neville-like, scheme*

$$
\begin{array}{ccccccccc}
c_{j-k+1} & & & & & & & & \\
& \searrow & & & & & & & \\
c_{j-k+2} & \to & c^{[1]}_{j-k+2}(x) & & & & & & \\
& \searrow & & \searrow & & & & & \\
c_{j-k+3} & \to & c^{[1]}_{j-k+3}(x) & \to & c^{[2]}_{j-k+3}(x) & & & & \\
\vdots & & & & & & & & \\
& \searrow & & \searrow & & \cdots & \searrow & & \\
c_j & \to & c^{[1]}_j(x) & \to & c^{[2]}_j(x) & \cdots & \to & c^{[k-1]}_j(x). &
\end{array}
$$

The scheme described in the previous Theorem 2.4.5 is also called *de Boor algorithm* and is implemented in the routine `Nev`, see appendix A. Due to Corollary 2.4.4, one can also derive a similar scheme for efficient point evaluations of the derivatives of a B-spline, cf. [DR08, (9.20)]. The scheme holds also true for coinciding inner knots.

Let us consider a hierarchy of spline spaces on uniform grids now. First of all, we define the space of splines on $[0,1]$ of order $k \in \mathbb{N}$ on a dyadic uniform mesh of grid size $2^{-j}$, $j \in \mathbb{N}$, as $SP_{j,k}$. That means

$$SP_{j,k} := \operatorname{span}\{N_{i,k} : i = 1, \ldots, N\}, \tag{2.4.6}$$

with respect to an extended sequence of knots $\{\theta_i\}_{i=1,\ldots,N+k}$ such that

$$\theta_1 = \ldots = \theta_k = 0 < \theta_{k+1} < \ldots < \theta_N < 1 = \theta_{N+1} = \ldots = \theta_{N+k},$$

where $N := 2^j + k - 1$ with $\theta_{i+1} - \theta_i = 2^{-j}$ for inner knots with $i = k, \ldots, N$. We call the index $j$ the *level of resolution*, refinement level or simply level. It is not hard to see that the spaces are nested $SP_{j,k} \subset SP_{j+1,k} \subset SP_{j+2,k} \ldots$ since each set of nodes is contained in the sets of nodes on the higher levels.Moreover, the B-splines are refinable.

**Proposition 2.4.7.** *The uniform B-splines $N_{i,k} \in SP_{j,k}$, $j \in \mathbb{N}$, $k \in \mathbb{N}$, $i = 1, \ldots, N$ are* refinable. *Moreover, for the (inner) B-splines the refinement relation*

$$N_{i,k}(x) = \sum_{m=0}^{k} 2^{1-k} \binom{k}{m} N^+_{2i-k+m,k}(x), \quad i = k, \ldots, N-k+1$$

*holds, where $N^+_{i,k} \in SP_{j+1,k}$ denote the B-splines with respect to the next finer grid.*

*Proof.* See [H03, 3.9] adapted to our notation. □

One can use Proposition 2.4.7 to express a spline on $SP_{j,k}$, represented with respect to a B-spline basis, also in terms of B-splines on the refined space $SP_{j+1,k}$. That is, there exists a matrix $M \in \mathbb{R}^{N \times (2N+1-k)}$ such that

$$\mathbf{N}_k = M \mathbf{N}^+_k,$$

with $\mathbf{N}_k := (N_{1,k}, \ldots, N_{N,k})^T$ and $\mathbf{N}_k^+ := (N_{1,k}^+, \ldots, N_{2N+1-k,k}^+)^T$. By Proposition 2.4.7 $M$ takes the form

$$M = \left(\begin{array}{c|cccccc} & 0 & & & & \\ & \vdots & & & & \\ M_{\mathrm{l}}^* & 0 & & & & \\ & a_0 & & & & \\ & a_1 & & & & \\ \hline & a_2 & a_0 & & & \\ & \vdots & \vdots & & & \\ & a_k & \vdots & & a_0 & \\ & & a_{k-1} & \ddots & a_1 & \\ & & a_k & & \vdots & \\ & & & & a_k & \\ & & & & 0 & M_{\mathrm{r}}^* \\ & & & & \vdots & \\ & & & & 0 & \end{array}\right),$$

where $a_m := 2^{1-k}\binom{k}{m}$, $m = 0, \ldots, k$ and $M_{\mathrm{l}}^*$, $M_{\mathrm{r}}^*$ are boundary adaptations. This transformation is done in the implementation with the routine `RefineMat`, cf. appendix A. The refinement coefficients for the boundary functions are calculated via interpolation. In addition to this practically useful transformation, one can also prove Bernstein inequalities for classes of refinable functions.

Since the code implemented in connection with this thesis and most of the stability analyses rely on multivariate tensor product functions, we introduce tensor product spaces before we state some smoothness and approximation results. Let $\Delta_{T_j}$ be sequences of knots as in Proposition 2.4.2 for $j = 1, \ldots, n$ with dimension $n \in \mathbb{N}$. Define the *space of tensor-product spline* as

$$SP^n_{\Delta_T, \mathbf{k}} := \bigotimes_{j=1}^{n} SP_{\Delta_{T_j}, k_j}. \tag{2.4.8}$$

In analogy to the shorthand notation (2.4.6) on uniform grids, we define the multidimensional pendant as $SP_{\mathbf{j},\mathbf{k}}$ in an obvious way. One can extend the considerations above straightforwardly to tensor products in most cases by introducing a Kronecker product. The scheme from Theorem 2.4.5 can also be extended. To this end, we consider a tensor product spline

$$S(x_1, \ldots, x_n) = \sum_{i_1=1}^{N_1} \cdots \sum_{i_n=1}^{N_n} c_{i_1,\ldots,i_n}\ N_{i_1,k_1}(x_1) \cdots N_{i_n,k_n}(x_n) \in SP^n_{\Delta_T, \mathbf{k}},$$

with given expansion coefficients $c_{i_1,\dots,i_n}$. This sum is reducible to the one-dimensional case by the following reordering

$$S(\mathbf{x}) = \sum_{i_1=1}^{N_1} N_{i_1}(x_1)\Big(\cdots\Big(\sum_{i_{n-1}=1}^{N_{n-1}} N_{i_{n-1}}(x_{n-1})\Big(\sum_{i_n=1}^{N_n} c_{\mathbf{i}}\ N_{i_n}(x_n)\Big)\cdots\Big), \tag{2.4.9}$$

where $\mathbf{x} := (x_1,\dots,x_n)$ and $\mathbf{i} := (i_1,\dots,i_n)$. The evaluation at a given point in $\mathbb{R}^n$ can now be calculated by recursively applying the scheme from Theorem 2.4.5, starting with the inner term in (2.4.9) $\sum_{i_n=1}^{N_n} c_{\mathbf{i}}\ N_{i_n}(x_n)$ for fixed remaining indices $i_1,\dots,i_{n-1}$. This yields the new expansion coefficients for the remaining coordinate directions and so on. To illustrate the rather technical description in arbitrary dimensions, we consider a two-dimensional example now, i.e.,

$$S(x,y) = \sum_{i=1}^{N_1}\sum_{j=1}^{N_2} c_{i,j} N_{i,k_1}(x)\ N_{j,k_2}(y) = \sum_{i=1}^{N_1} N_{i,k_1}(x)\Big(\sum_{j=1}^{N_2} c_{i,j}\ N_{j,k_2}(y)\Big).$$

Let us define the inner part as

$$d_i(y) := \sum_{j=1}^{N_2} c_{i,j}\ N_{j,k_2}(y), \quad \text{such that } S(x,y) = \sum_{i=1}^{N_1} d_i(y)\ N_{i,k_1}(x).$$

So one can apply the scheme from Theorem 2.4.5 to $S(x,y)$ with expansion coefficients $d_i(y)$ if $d_i(y)$ is known for all $i = 1,\dots,N_1$ at a given evaluation point. This in turn can be done by applying the scheme from Theorem 2.4.5 to $d_i(y) := \sum_{j=1}^{N_2} c_{i,j}\ N_{j,k_2}(y)$ for each index $i = 1,\dots,N_1$. In this way one obtains successively the value of the two-dimensional spline at a given evaluation point.

We conclude this chapter with approximation and smoothness properties of B-splines, namely so called *direct and inverse estimates*, also known as Bernstein and Jackson estimates.

**Theorem 2.4.10.** *The splines $S \in SP_{\mathbf{j},\mathbf{k}}$, $\mathbf{j},\mathbf{k} \in \mathbb{N}^n$, satisfy the Bernstein inequality*

$$\|S\|_{H^s((0,1)^n)} \lesssim 2^{sj}\|S\|_{L_2((0,1)^n)}, \tag{2.4.11}$$

*for all $0 \le s \le \gamma$ with $\gamma := \sup\{s \in \mathbb{R} : SP_{\mathbf{j},\mathbf{k}} \subset H^s((0,1)^n)\}$.*

*Proof.* The proof follows with [Urb09, Lemma 5.11] and Proposition 2.4.7. □

The spline spaces $SP_{\mathbf{j},\mathbf{k}}$ with $k_{\min} := \min\{k_1,\dots,k_n\}$ are by construction $(k_{\min}-2)$-times differentiable. Therefore, since they are piecewise polynomial, it holds $SP_{\mathbf{j},\mathbf{k}} \subset H^{k_{\min}-1}((0,1)^n)$, such that the Bernstein estimate (2.4.11) is satisfied at least with $\gamma \ge k_{\min} - 1$. It can be shown that one generally obtains an even higher Sobolev smoothness of certain fractional order, cf. [DKU99, sec. 3.4].

**Theorem 2.4.12.** *The approximation error with respect to tensor-product splines of an arbitrary function $f \in H^{\mathbf{r}}(D)$ with $r_j \leq k_j$ on uniform grids with grid sizes $h_j$ for $j = 1, \ldots, n$ is given by*

$$\inf_{S \in SP^n_{\Delta_T, \mathbf{k}}} \|f - S\|_{L_2(D)} \leq C \sum_{j=1}^{n} h_j^{r_j} \left\| \frac{\partial^{r_j}}{\partial x_j^{r_j}} f \right\|_{L_2(D)},$$

*on a rectangular domain $D := \otimes_{j=1}^{n} (a_j, b_j)$ given by the boundary points of $T_j$ with a constant $C$ only depending on $\mathbf{r}, \mathbf{k}, n$. We denoted the* tensor Sobolev space *by*

$$H^{\mathbf{r}}(D) := \left\{ f : \frac{\partial^{|\alpha|}}{\partial \mathbf{x}^{\alpha}} f \in L_2(D), 0 \leq \alpha \leq \mathbf{r}, j = 1, \ldots, n \right\},$$

*with multiindex notation as in Definition 2.2.1 and $0 \leq \alpha \leq \mathbf{r} :\Leftrightarrow 0 \leq \alpha_j \leq r_j$ for all $j = 1, \ldots, n$.*

*Proof.* The theorem is a direct consequence of [Sch07, Th.12.7] applied to the particular setting. □

The previous theorem not only gives a Jackson inequality, which needs to be assumed later in our stability analysis in section 4.2, but also states (optimal) approximation rates with tensor product B-splines. One can also extend the measure on the left hand side to mixed Sobolev norms with decreased powers in the right hand side correspondingly, i.e.,

$$\inf_{S \in SP^n_{\Delta_T, \mathbf{k}}} \|f - S\|_{H^{\mathbf{q}}(D)} \lesssim \sum_{j=1}^{n} h_j^{r_j - q_j} \left\| \frac{\partial^{r_j}}{\partial x_j^{r_j}} f \right\|_{L_2(D)}, \tag{2.4.13}$$

with $q_j \leq r_j \leq k_j$ for $j = 1, \ldots, n$, cf. [Sch07, Th. 13.20].

# 3 Full Space-Time Weak Formulation for Parabolic Problems

In this chapter, we introduce full weak formulations of PDEs in space *and* time. The idea of formulating a PDE weakly in space and time is not new, see, e.g., [DL92], but was revisited by Schwab and Stevenson in [SS09] for numerical purposes in the context of adaptive wavelet methods. A full space-time weak formulation was essentially also used before in [BJ89, BJ90] for *h-p* finite element methods. The basic idea is to multiply a parabolic PDE with a space-time test function and to integrate over space and time. In this way one obtains a single operator equation in both variables. From a numerical point of view, one can eliminate the explicit time dependency, as it appears in classical semi-discretizations, and discretize in space and time simultaneously when considering a full space and time weak formulation. That means, one does not need any time stepping, but can calculate, for instance, the *Petrov-Galerkin solution* directly in space and time. We refer to chapter 4 for a detailed description of Petrov-Galerkin approaches and in particular their stability. Classical approaches to solve parabolic evolution problems are the *method of lines* or *Rothe's method*. Both methods are applied to operator equations of the form

$$\frac{\mathrm{d}u(t)}{\mathrm{d}t} + A(t)u(t) = g(t), \qquad u(0) = u_0,$$

with a spatial differential operator $A(t)$. The idea of the method of lines is to solve a coupled system of ordinary differential equations (ODEs) obtained by a spatial semidiscretization, see [Sch91, Tho06], and Rothe's method is based on a time semidiscretization, see [Lan01, Kac86]. That is, these two methods as well as, e.g., the *discontinuous Galerkin method* (see, e.g., [EG04]), which has become more and more popular in the last years, are based on semidiscretizations and are time marching methods. The stability of explicit methods with time marching generally requires a CFL condition, that is, a restriction on the time step size. It can be shown that the classical time-stepping methods as, for example, Euler method or Crank-Nicolson method, but also the discontinuous Galerkin method, can be interpreted as a Petrov-Galerkin approach of a space-time weak formulation with respect to certain ansatz functions. That means, Petrov-Galerkin discretizations of full space-time formulations cover suitable time stepping methods in some sense, but are much more flexible in the choice of ansatz functions. To this end, theoretically, one can derive numerical methods of arbitrary order in space and time. We focus on the stability of these discretizations in section 4.2.

In view of adaptive solution methods, a space-time weak formulation has the advantage that one obtains adaptivity directly in space and time simultaneously, and does not need an adaptive time stepping combined with adaptive spatial discretizations. Nevertheless, for smooth solutions, a uniform grid provides an approximation with optimal error rates. We will see that one obtains a well-posed equation on Hilbert

spaces defined on the whole space-time cylinder under rather mild conditions. Moreover, the corresponding operator can be shown to be boundedly invertible with, in many cases, *explicitly predictable bounds*. These bounds play an important role for the (quasi-optimal) approximation error and the well-posedness of the problem, cf. Theorem 4.1.8. Having explicit bounds is also mandatory for deriving the results of chapter 5 for random PDEs. There are basically two reasonable, slightly different formulations, which are referred to as *first* and *second* formulation following the denotation in [LMM16, LM16b].

For parabolic PDEs, an approach using space-time sparse grid multilevel methods was introduced in [GO07]. The authors interpreted space-time sparse grid discretization schemes as the sparse counterparts of the Crank-Nicolson scheme and the discontinuous Galerkin scheme with piecewise constant or linear functions in time. It was shown that the additional temporal dimension in the complexity estimates can be avoided with these constructions. The interpretation of essentially time-stepping methods via a space-time formulation with suitable ansatz functions was also exploited for reduced basis approximations [UP12, UP14] as well as for (quantized) tensor train (QTT) low rank tensor approximations [DKO12], just to mention a few. In that regard, we would like to mention [UP12, UP14] in the context of the Crank-Nicolson method and [MB97] in the context of the discontinuous Galerkin method for the heat equation. A space-time finite element approach in full weak formulation was considered in [Ste15]. The author also provides stability and a priori error estimates. A space-time weak formulation for stochastic PDEs, namely the stochastic heat equation, was considered in [LM16b]. Moreover, in [GK11, KM15] the full space-time weak formulation was considered for the constraints of control problems. It is very well suited for these problems, since they lead to systems of PDEs which are coupled in space and time, so that one has to store the information for all time steps anyway. Another different approach using half derivatives was introduced in [LW13, LS15, KM15]. Using half derivatives yields a "symmetricification" of the space-time operator in a certain sense. In this respect, we also want to mention [SS16], where fractional derivatives are used for space-time formulations of Navier-Stokes equations.

A rather detailed analysis of the first formulation, its homogenization and the second formulation are presented. Moreover, two different types of coercive spatial differential operators are discussed, first the most general case, then a restriction to self-adjoint and time-independent operators. The considerations are not restricted to coercive spatial differential operators, but also include spatial differential operators which only satisfy a *Gårding inequality*. The general coercive case can basically be found in [Tan13, SS09] for the first form and in [Tan13, CS11, LM16a] for the second form. We worked out improved estimates for coercive, self-adjoint, as well as time-independent spatial differential operators in second form in our very recent work [LMM16]. A homogenization was introduced in [Sta11], for which the explicit bounds are derived in this thesis. A Gårding inequality was also part of [SS09]. The remaining cases with respect to the homogenization mentioned above, with a regularity result on more general spaces, are

worked out for this thesis to have a completed overview.

## 3.1 First Formulation

We start with the *first* space-time weak formulation. This formulation was already considered, e.g., in [SS09, Ste15, Tan13, And13, And12, UP12, UP14]. Let $(V, (\cdot,\cdot)_V)$ and $(H, (\cdot,\cdot)_H)$ be two separable Hilbert spaces with scalar product $(\cdot,\cdot)_V$ and $(\cdot,\cdot)_H$, respectively, with $V \hookrightarrow H$ assumed to be densly embedded in $H$. Taking $H$ as the pivot space by identifying $H$ with its dual $H'$, we obtain the Gelfand triple $V \hookrightarrow H \cong H' \hookrightarrow V'$ with corresponding duality pairing on $V' \times V$ denoted by ${}_{V'}\langle\cdot,\cdot\rangle_V$. We often drop the indices and simply write $\langle\cdot,\cdot\rangle$ if the spaces are clear from the context.

Consider the time interval $I := (0,T)$, $0 < T < \infty$ and, for $t \in I$ almost everywhere (a.e.), a bilinear form $a(t;\cdot,\cdot)\colon V \times V \to \mathbb{R}$, where the mapping $t \mapsto a(t;u,v)$ is measurable on $I$ for any $u, v \in V$. A.e. is a shorthand notation for "almost everywhere" and means that the set of elements for which a property does not hold is a set of measure zero, where we consider the Lebesgue measure here. Moreover, we assume boundedness and coercivity for $t \in I$ a.e., that is, there exists constants $0 < A_{\min} \le A_{\max} < \infty$ such that for $t \in I$ a.e.,

$$\begin{aligned} |a(t;v,w)| &\le A_{\max}\|v\|_V\|w\|_V && \text{for all } v,w \in V \text{ (boundedness)},\\ a(t;v,v) &\ge A_{\min}\|v\|_V^2 && \text{for all } v \in V \text{ (coercivity)}. \end{aligned} \tag{3.1.1}$$

It is worth mentioning that one can prove well-posedness when we only require a Gårding inequality for $t \in I$ a.e.

$$a(t;v,v) + \lambda\|v\|_H^2 \ge A_{\min}\|v\|_V^2 \qquad \text{for all } v \in V, \tag{3.1.2}$$

with a $\lambda \ge 0$, instead of the coercivity, see [SS09]. Notice that the denotation is not always consistent in the literature. The coercivity (3.1.1) is sometimes referred to as ellipticity, whereas the Gårding inequality (3.1.2) is also referred to as coercivity. We will consistently use the denotation introduced above in this thesis. Unless specified differently, we restrict ourselves to the truely coercive case (3.1.1) first and refer to Corollary 3.1.16, 3.2.16, 3.3.16 for the inf-sup and continuity constants of non-coercive bilinear forms. Moreover, we are able to show improved constants in the coercive case.

This bilinear form $a(t;\cdot,\cdot)$ defines, for $t \in I$ a.e., a unique operator $A(t) \in \mathcal{L}(V,V')$, cf. (2.3.2), via

$$\langle A(t)u, v\rangle := a(t;u,v), \qquad u, v \in V. \tag{3.1.3}$$

Now the corresponding evolution problem is formulated as

$$\begin{aligned} \frac{\mathrm{d}u}{\mathrm{d}t}(t) + A(t)u(t) &= g(t) \quad \text{in } V',\ t \in (0,T] \text{ a.e.}\\ u(0) &= u_0 \quad \text{in } H, \end{aligned} \tag{3.1.4}$$

for a given right hand side $g \in L_2(I;V')$ and initial value (function) $u_0 \in H$. This is the standard weak formulation with respect to space, formulated as a generic evolutionary equation with bounded and linear *spatial differential* operator $A(t) \in \mathcal{L}(V,V')$. A classical approach would be to test equation (3.1.4) with spatial test functions and to discretize the corresponding problem:

$$\begin{aligned} {}_{V'}\left\langle \frac{\mathrm{d}u}{\mathrm{d}t}(t) + A(t)u(t), v_1 \right\rangle_V &= {}_{V'}\langle g(t), v_1\rangle_V \quad \text{for all } v_1 \in V,\ t \in (0,T] \text{ a.e.} \\ (u(0), v_2)_H &= (u_0, v_2)_H \qquad\qquad \text{for all } v_2 \in H. \end{aligned}$$

Instead of considering a spatial semidiscretization of (3.1.4) with, e.g., finite elements or wavelets, and to solve the resulting system of ordinary differential equations (ODEs) by time-stepping methods, we want to find a *full space-time weak formulation*.

First, we test the operator equation (3.1.4) with *space-time* test functions $v_1 \in L_2(I;V)$ and integrate over $t \in I$. In this way we arrive at the variational formulation

$$\int_I ({}_{V'}\langle \dot{u}(t), v_1(t)\rangle_V + {}_{V'}\langle A(t)u(t), v_1(t)\rangle_V)\ \mathrm{d}t = \int_I {}_{V'}\langle g(t), v_1(t)\rangle_V\ \mathrm{d}t, \tag{3.1.5}$$

for all $v_1 \in L_2(I;V)$, where we have used the usual shorthand notation $\dot{u}(t) := \frac{\mathrm{d}}{\mathrm{d}t}u(t)$ for time derivatives. In order to ensure the initial condition, we additionally need to add the condition tested by another test function $v_2 \in H$ which yields the *first* full space-time variational formulation:

Find a solution $u \in \mathcal{X}$ such that

$$b(u,v) = \mathcal{F}(v) \qquad \text{for all } v := (v_1, v_2) \in \mathcal{Y}, \tag{3.1.6}$$

where the test space is given by

$$\mathcal{Y} := L_2(I;V) \times H, \tag{3.1.7}$$

and the solution space by

$$\mathcal{X} := L_2(I;V) \cap H^1(I;V'), \tag{3.1.8}$$

with bilinear form $b(\cdot,\cdot)\colon \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ defined by

$$b(u,(v_1,v_2)) := \int_I \big({}_{V'}\langle \dot{u}(t), v_1(t)\rangle_V + {}_{V'}\langle A(t)u(t), v_1(t)\rangle_V\big)\ \mathrm{d}t + (u(0), v_2)_H \tag{3.1.9}$$

and right hand side $\mathcal{F}(\cdot)\colon \mathcal{Y} \to \mathbb{R}$ given by

$$\mathcal{F}(v) := \int_I {}_{V'}\langle g(t), v_1(t)\rangle_V\ \mathrm{d}t + (u_0, v_2)_H. \tag{3.1.10}$$

We equip the spaces $\mathcal{X}$ and $\mathcal{Y}$ with norms

$$\|v\|_{\mathcal{X}}^2 := \|v\|_{L_2(I;V)}^2 + \|\dot{v}\|_{L_2(I;V')}^2, \qquad \|v\|_{\mathcal{Y}}^2 := \|v_1\|_{L_2(I;V)}^2 + \|v_2\|_H^2.$$

Recall that point evaluations of $u \in \mathcal{X}$ are well-defined with $u(0) \in H$ by the embedding Theorem 2.2.18. We define an operator $B \in \mathcal{L}(\mathcal{X}, \mathcal{Y}')$ by

$$_{\mathcal{Y}'}\langle Bu, w\rangle_{\mathcal{Y}} := b(u, w) \tag{3.1.11}$$

in analogy to (3.1.3). The following theorem from [SS09] ensures existence and uniqueness of a solution of the full weak formulation (3.1.6).

**Theorem 3.1.12.** *The operator $B \in \mathcal{L}(\mathcal{X}, \mathcal{Y}')$ from* (3.1.11) *with $\mathcal{X}$ and $\mathcal{Y}$ from* (3.1.8) *and* (3.1.7) *is boundedly invertible.*

A proof was essentially already given in [DL92, Ch. 13, §3] and [Wlo82, Ch. IV, §26], but not in this particular form and without error bounds for operator $B$. A closer inspection of the alternative proof in [SS09, Appendix A] yields the error bounds

$$\sup_{v\in\mathcal{X}\setminus\{0\}} \sup_{w\in\mathcal{Y}\setminus\{0\}} \frac{|\langle Bv, w\rangle|}{\|v\|_{\mathcal{X}}\|w\|_{\mathcal{Y}}} \leq \sqrt{2\max\{1, A_{\max}^2\} + \rho^2} =: C_B \tag{3.1.13}$$

as well as

$$\inf_{v\in\mathcal{X}\setminus\{0\}} \sup_{w\in\mathcal{Y}\setminus\{0\}} \frac{|\langle Bv, w\rangle|}{\|v\|_{\mathcal{X}}\|w\|_{\mathcal{Y}}} \geq \frac{\min\{A_{\min}A_{\max}^{-2}, A_{\min}\}}{\sqrt{2\max\{A_{\min}^{-2}, 1\}+\rho^2}} =: c_B,$$

with constant $\rho$ defined as

$$\rho := \sup_{0\neq w\in\mathcal{Y}} \frac{\|w(0,\cdot)\|_H}{\|w\|_{\mathcal{Y}}}, \tag{3.1.14}$$

which is finite due to Theorem 2.2.18. That is, operator $B$ and its inverse $B^{-1}$ are bounded by

$$\|B\|_{\mathcal{L}(\mathcal{X},\mathcal{Y}')} \leq C_B, \qquad \|B^{-1}\|_{\mathcal{L}(\mathcal{Y}',\mathcal{X})} \leq c_B^{-1},$$

see Corollary 2.3.8, so that we have a well-conditioned problem with (spectral) condition

$$\kappa(B) \leq \frac{C_B}{c_B}.$$

Problem 3.1.6 is well-defined for $g \in L_2(I; V')$ and $u_0 \in H$, since it implies $\mathcal{F} \in \mathcal{Y}'$. Improved error bounds were proven in [Tan13, UP12, UP14]. In [Tan13, Prop. 2.2] a slightly modified norm was used, but it is not hard to see that one also obtains

$$\inf_{v\in\mathcal{X}\setminus\{0\}} \sup_{w\in\mathcal{Y}\setminus\{0\}} \frac{|\langle Bv, w\rangle|}{\|v\|_{\mathcal{X}}\|w\|_{\mathcal{Y}}} \geq \frac{\min\{A_{\min}, A_{\max}^{-1}, A_{\min}A_{\max}^{-1}\}}{2}. \tag{3.1.15}$$

Notice, that the previous continuity and inf-sup constants are derived with respect to the standard Sobolev norms, and, different from, e.g., [And13, UP12, UP14], do *not*

use norms induced by the spatial differential operator. That is, there is no "hidden" operator dependence involved. This will be essential when dealing with (stochastic) parameter dependent operators later in chapter 5.

As already mentioned before, one can relax the condition to have a coercive spatial differential operator and only assume a Gårding inequality (3.1.2).

**Corollary 3.1.16.** *The operator $B \in \mathcal{L}(\mathcal{X}, \mathcal{Y}')$ from (3.1.11) is still boundedly invertible, when we assume a Gårding inequality* (3.1.2) *instead of coercivity* (3.1.1) *with respect to $A(t)$ for $t \in I$ a.e. In that case, the operator can be bounded as*

$$\begin{aligned}&\sup_{v\in\mathcal{X}\setminus\{0\}}\sup_{w\in\mathcal{Y}\setminus\{0\}}\frac{|\langle Bv,w\rangle|}{\|v\|_{\mathcal{X}}\|w\|_{\mathcal{Y}}}\\ \leq& e^{\lambda T}\max\{\sqrt{1+2\lambda^2\varrho^4},\sqrt{2}\}\sqrt{2\max\{1,(A_{\max}+\lambda)^2\}+\rho^2}\end{aligned}$$

*as well as*

$$\begin{aligned}&\inf_{v\in\mathcal{X}\setminus\{0\}}\sup_{w\in\mathcal{Y}\setminus\{0\}}\frac{|\langle Bv,w\rangle|}{\|v\|_{\mathcal{X}}\|w\|_{\mathcal{Y}}}\\ \geq\ & e^{-\lambda T}\frac{\min\{A_{\min},(A_{\max}+\lambda)^{-1},A_{\min}(A_{\max}+\lambda)^{-1}\}}{2\max\{\sqrt{1+2\lambda^2\varrho^4},\sqrt{2}\}},\end{aligned}$$

*where $\varrho := \sup_{0\neq v\in V}\frac{\|v\|_H}{\|v\|_V}$ denotes the embedding constant of $V \hookrightarrow H$.*

*Proof.* The proof is basically already given without emphasizing the bounds explicitly in [SS09, Appendix A] and with bounds, but for slightly different norms and spaces, in [UP14, Cor. 2.7]. But, an examination of the proof in [SS09, Appendix A] reveled that there might be a problem with two relevant estimates, which would consequently leads to incorrect bounds in [UP14, Cor. 2.7]. To this end, we go similar lines as in [SS09, Appendix A] and derive the final estimates carefully in detail, just to be on the safe side and to obtain explicit bounds.

Defining $\hat{u}(t) := e^{-\lambda t}u(t)$, $\hat{v}_1(t) := e^{\lambda t}v_1(t)$ and $\hat{g}(t) := e^{-\lambda t}g(t)$, one can immediately observe that the problem

$$\begin{aligned}\langle\hat{B}\hat{u},\hat{v}\rangle\ :=\ &\int_I \left({}_{V'}\langle\tfrac{\mathrm{d}}{\mathrm{d}t}\hat{u}(t),\hat{v}_1(t)\rangle_V+{}_{V'}\langle A(t)\hat{u}(t),\hat{v}_1(t)\rangle_V\right)\mathrm{d}t\\ &+\lambda\int_I(\hat{u}(t),\hat{v}_1(t))_H\,\mathrm{d}t+(\hat{u}(0),v_2)_H\\ =\ &\int_I{}_{V'}\langle\hat{g}(t),\hat{v}_1(t)\rangle_V\,\mathrm{d}t+(u_0,v_2)_H,\end{aligned}$$

for all $\hat{v} := (\hat{v}_1, v_2) \in \mathcal{Y}$, is equivalent to (3.1.6).

Next, we want to estimate the ratios $\|w\|_{\mathcal{X}}/\|\hat{w}\|_{\mathcal{X}}$ and $\|v\|_{\mathcal{Y}}/\|\hat{v}\|_{\mathcal{Y}}$ for all $w \in \mathcal{X}$ and $v \in \mathcal{Y}$, with $\hat{w} := e^{-\lambda t}w$ and $\hat{v} := (e^{\lambda t}v_1, v_2)$.

One can estimate for $v \in \mathcal{Y} \setminus \{0\}$

$$\|v\|_{\mathcal{Y}}^2 = \|e^{-\lambda t}\hat{v}_1\|_{L_2(I;V)}^2 + \|v_2\|_H^2 \leq \|\hat{v}_1\|_{L_2(I;V)}^2 + \|v_2\|_H^2$$

and

$$\|v\|_{\mathcal{Y}}^2 = \|e^{-\lambda t}\hat{v}_1\|_{L_2(I;V)}^2 + \|v_2\|_H^2 \geq e^{-2\lambda T}(\|\hat{v}_1\|_{L_2(I;V)}^2 + \|v_2\|_H^2),$$

since $\lambda \geq 0$ by assumption (3.1.2), so that

$$e^{-\lambda T} \leq \|v\|_{\mathcal{Y}}/\|\hat{v}\|_{\mathcal{Y}} \leq 1,$$

by the definition of $\hat{v}$. The other ratio is more technical but also straightforward. For $w \in \mathcal{X} \setminus \{0\}$, using Cauchy-Schwarz, we can estimate

$$\begin{aligned}
\|w\|_{\mathcal{X}}^2 &= \|e^{\lambda t}\hat{w}\|_{L_2(I;V)}^2 + \|e^{\lambda t}\lambda\hat{w} + e^{\lambda t}\tfrac{\mathrm{d}}{\mathrm{d}t}\hat{w}\|_{L_2(I;V')}^2 \\
&\leq e^{2\lambda T}\left(\|\hat{w}\|_{L_2(I;V)}^2 + 2(\lambda^2\|\hat{w}\|_{L_2(I;V')}^2 + \|\tfrac{\mathrm{d}}{\mathrm{d}t}\hat{w}\|_{L_2(I;V')}^2)\right) \\
&\leq e^{2\lambda T}\left((1+2\varrho^4\lambda^2)\|\hat{w}\|_{L_2(I;V)}^2 + 2\|\tfrac{\mathrm{d}}{\mathrm{d}t}\hat{w}\|_{L_2(I;V')}^2\right) \\
&\leq e^{2\lambda T}\max\{1+2\varrho^4\lambda^2, 2\}\|\hat{w}\|_{\mathcal{X}}^2,
\end{aligned}$$

since

$$\|z\|_{V'} = \sup_{0\neq x\in V}\frac{(z,x)_H}{\|x\|_V} \leq \varrho \sup_{0\neq x\in V}\frac{(z,x)_H}{\|x\|_H} \leq \varrho \sup_{0\neq x\in H}\frac{(z,x)_H}{\|x\|_H} = \varrho\|z\|_{H'} \leq \varrho^2\|z\|_V,$$

for any $z \in V \subset H \cong H' \subset V'$. For the lower bound we similar obtain

$$\begin{aligned}
\|\hat{w}\|_{\mathcal{X}}^2 &= \|e^{-\lambda t}w\|_{L_2(I;V)}^2 + \|-e^{-\lambda t}\lambda w + e^{-\lambda t}\dot{w}\|_{L_2(I;V')}^2 \\
&\leq \|w\|_{L_2(I;V)}^2 + 2(\lambda^2\|w\|_{L_2(I;V')}^2 + \|\dot{w}\|_{L_2(I;V')}^2) \\
&\leq (1+2\varrho^4\lambda^2)\|w\|_{L_2(I;V)}^2 + 2\|\dot{w}\|_{L_2(I;V')}^2 \\
&\leq \max\{1+2\varrho^4\lambda^2, 2\}\|w\|_{\mathcal{X}}^2.
\end{aligned}$$

Using these estimates one can easily derive

$$\begin{aligned}
&\sup_{w\in\mathcal{X}\setminus\{0\}}\sup_{v\in\mathcal{Y}\setminus\{0\}}\frac{|\langle Bw, v\rangle|}{\|w\|_{\mathcal{X}}\|v\|_{\mathcal{Y}}} = \sup_{w\in\mathcal{X}\setminus\{0\}}\sup_{v\in\mathcal{Y}\setminus\{0\}}\frac{|\langle \hat{B}\hat{w}, \hat{v}\rangle|}{\|w\|_{\mathcal{X}}\|v\|_{\mathcal{Y}}} \\
\leq\ & e^{\lambda T}\max\{\sqrt{1+2\lambda^2\varrho^4}, \sqrt{2}\}\sup_{w\in\mathcal{X}\setminus\{0\}}\sup_{v\in\mathcal{Y}\setminus\{0\}}\frac{|\langle \hat{B}\hat{w}, \hat{v}\rangle|}{\|\hat{w}\|_{\mathcal{X}}\|\hat{v}\|_{\mathcal{Y}}},
\end{aligned}$$

and

$$\begin{aligned}
&\inf_{w\in\mathcal{X}\setminus\{0\}}\sup_{v\in\mathcal{Y}\setminus\{0\}}\frac{|\langle Bw, v\rangle|}{\|w\|_{\mathcal{X}}\|v\|_{\mathcal{Y}}} = \inf_{w\in\mathcal{X}\setminus\{0\}}\sup_{v\in\mathcal{Y}\setminus\{0\}}\frac{|\langle \hat{B}\hat{w}, \hat{v}\rangle|}{\|w\|_{\mathcal{X}}\|v\|_{\mathcal{Y}}} \\
\geq\ & e^{-\lambda T}\max\{\sqrt{1+2\lambda^2\varrho^4}, \sqrt{2}\}^{-1}\inf_{w\in\mathcal{X}\setminus\{0\}}\sup_{v\in\mathcal{Y}\setminus\{0\}}\frac{|\langle \hat{B}\hat{w}, \hat{v}\rangle|}{\|\hat{w}\|_{\mathcal{X}}\|\hat{v}\|_{\mathcal{Y}}}.
\end{aligned}$$

Now the proof follows with (3.1.13) and (3.1.15), since the spatial part of $\hat{B}$, i.e., ${}_{V'}\langle A(t)\cdot,\cdot\rangle_V + \lambda(\cdot,\cdot)_H$, is obviously coercive on $V$ with

$$|{}_{V'}\langle A(t)x, y\rangle_V + \lambda(x,y)_H| \leq (A_{\max} + \lambda)\|x\|_V\|y\|_V, \quad \text{for all } x, y \in V.$$

□

Notice that the lower and upper bounds decrease or increase, respectively, *exponentially* in the final time $T$. That is, the estimates become very bad and unsuitable for long-term integration. Nevertheless, it proves existence and uniqueness of a solution in the first place, but also yields worst case scenarios. The bounds are also suitable to derive existence of moments of solutions of spatial differential random operators, as we will see later in section 5.

Especially from a computational point of view, it is quite uneconomical to have the additional Cartesian product included in the test space $\mathcal{Y} = L_2(I;V) \times H$ stemming from the initial condition. We will present two possibilities to get rid of this product. One idea is to go one step back and to reformulate the equation (3.1.5) in a slighty different way, such that the initial condition is incorporated naturally by using integration by parts, cf. [CS11, Tan13, LMM16]. Another idea is to consider a *homogenized* problem with zero initial conditions as done in the next section.

## 3.2 Homogenization

A homogenization of the first formulation (3.1.6) was, to my knowledge, first described in [Sta11]. The idea is to eliminate the initial condition in the operator from (3.1.6). This in turn would also lead to a slightly different test space $\mathcal{Y}$ without Cartesian product.

To this end, assume that we have available a function $u_0^* \in \mathcal{X}$ such that

$$u_0^*(0) = u_0.$$

We redefine the solution and test space as

$$\mathcal{X}_0 := L_2(I;V) \cap H^1_{0,\{0\}}(I;V') := \{u \in L_2(I;V) \cap H^1(I;V') : u(0) \equiv 0\}, \tag{3.2.1}$$

and

$$\mathcal{Y}_0 := L_2(I;V), \tag{3.2.2}$$

since we want to incorporate zero initial conditions. Recall that (3.2.1) is well defined by the embedding Theorem 2.2.18.

Now we can formulate the *homogenized problem* as follows:
Find a solution $u^* \in \mathcal{X}_0$ such that

$$b_0(u^*, v) = \mathcal{F}_0(v), \quad \text{for all } v \in \mathcal{Y}_0, \tag{3.2.3}$$

where $b_0(\cdot,\cdot)\colon \mathcal{X}_0 \times \mathcal{Y}_0 \to \mathbb{R}$ is now defined by

$$b_0(u,v) := \int_I ({}_{V'}\langle \dot{u}(t), v(t)\rangle_V + {}_{V'}\langle A(t)u(t), v(t)\rangle_V)\,\mathrm{d}t \tag{3.2.4}$$

and the right hand side $\mathcal{F}_0\colon \mathcal{Y}_0 \to \mathbb{R}$ by

$$\mathcal{F}_0(v) := \int_I {}_{V'}\langle g^*(t), v(t)\rangle_V\,\mathrm{d}t, \tag{3.2.5}$$

with

$$g^*(t) := g(t) - (\dot{u}_0^*(t) + A(t)u_0^*(t)) \in L_2(I,V'), \quad t \in I \text{ a.e.}$$

The bilinear form $b_0(\cdot,\cdot)$ defines an operator $B_0\colon \mathcal{X}_0 \to \mathcal{Y}_0'$ in the same way as in (3.1.11)

$${}_{\mathcal{Y}_0'}\langle B_0 u, w\rangle_{\mathcal{Y}_0} := b_0(u,w). \tag{3.2.6}$$

The next proposition, essentially from [Sta11, Prop 3.15], yields the equivalence of this homogenized problem (3.2.3) and the original problem (3.1.6).

**Proposition 3.2.7.** *$u^* \in \mathcal{X}_0$ solves the homogenized problem (3.2.3) if and only if $u := u^* + u_0^* \in \mathcal{X}$ solves problem (3.1.6).*

*Proof.* It is assumed in [Sta11, Prop. 3.15] that $H = L_2(D)$ on a bounded domain $D \subset \mathbb{R}^n$, but it is not hard to show that it stays true for more general $H$. Choosing test functions $(0, v_2) \in L_2(I;V) \times H$ in (3.1.6) with arbitrary $v_2 \in H$, one obtains $(u(0) - u_0, v_2)_H = 0$ for all $v_2 \in H$. In particular, with $v_2 := u(0) - u_0$ we have $\|u(0) - u_0\|_H^2 = 0$ and therefore $u(0) \equiv u_0$ in $H$. Now the proof follows along the same lines as in [Sta11, Prop 3.15]. □

Let us focus now on the assumption of having a function $u_0^* \in \mathcal{X}$ with $u_0^*(0) = u_0$. Since obviously $H^1(I;V) \subset L_2(I;V) \cap H^1(I;V')$, we can easily specify a suitable choice of $u_0^*$ when we assume $u_0 \in V \subset H$. With this little drawback, we can immediately see that $u_0^* := 1 \otimes u_0$ meets the assumptions $u_0^*(0) = u_0$ and $u_0^* \in \mathcal{X}$ for any given initial value $u_0 \in V$.

Straightforward calculations along the lines of [Tan13, Prop. 2.2 and Prop. 2.3] show that the inf-sup condition as well as boundedness hold with the following constants

$$\sup_{v\in\mathcal{X}_0\setminus\{0\}} \sup_{w\in\mathcal{Y}_0\setminus\{0\}} \frac{|\langle B_0 v, w\rangle|}{\|v\|_{\mathcal{X}_0}\|w\|_{\mathcal{Y}_0}} \le \sqrt{2}\max\{1, A_{\max}\} \qquad \text{(boundedness)}$$

as well as

$$\inf_{v\in\mathcal{X}_0\setminus\{0\}} \sup_{w\in\mathcal{Y}_0\setminus\{0\}} \frac{|\langle B_0 v, w\rangle|}{\|v\|_{\mathcal{X}_0}\|w\|_{\mathcal{Y}_0}} \ge A_{\min}\frac{\min\{1, A_{\max}^{-1}\}}{\sqrt{2}} \qquad \text{(inf-sup condition).}$$

We can eliminate the factor $A_{\max}^{-1}$ in the inf-sup condition on more general spaces for time independent spatial differential operators $A$. To this end, we generalize the previous setting to shifted spaces

$$\begin{aligned} \bar{\mathcal{X}}_0 &:= L_2(I;W_+) \cap H^1_{0,\{0\}}(I;W_-), \\ \bar{\mathcal{Y}}_0 &:= L^2(I;W_-') \end{aligned} \tag{3.2.8}$$

with continuous and densely embedded separable Hilbert spaces $W_+ \hookrightarrow W_0 \hookrightarrow W_-$ and interpolation space $W_0 := [W_-, W_+]_{1/2}$. See (2.2.17) for the definition of interpolation spaces. Considering the operator $B_0 \in \mathcal{L}(\bar{\mathcal{X}}_0, \bar{\mathcal{Y}}_0{}')$ on these redefined spaces (3.2.8), we can prove the inf-sup and continuity condition for time-independent spatial differential operators $A \in \mathcal{L}(W_+, W_-)$ even with improved estimates.

**Theorem 3.2.9.** *Let $W_+ \hookrightarrow W_-$ be separable Hilbert spaces with scalar products $(\cdot,\cdot)_{W_+}$ and $(\cdot,\cdot)_{W_-}$, respectively, and $W_0 := [W_-, W_+]_{1/2}$ the interpolation space endowed with scalar product $(\cdot,\cdot)_{W_0}$. Assume that the operator $A$ is time-independent and self-adjoint with*

$$A_{\min} \le \|A\|_{W_+ \to W_-} \le A_{\max}, \tag{3.2.10}$$

*with constants $0 < A_{\min} \le A_{\max} < \infty$. Then $B_0 \colon \bar{\mathcal{X}}_0 \to \bar{\mathcal{Y}}_0{}'$ is boundedly invertible with*

$$\sup_{v \in \bar{\mathcal{X}}_0 \setminus \{0\}} \sup_{w \in \bar{\mathcal{Y}}_0 \setminus \{0\}} \frac{|\langle B_0 v, w\rangle|}{\|v\|_{\bar{\mathcal{X}}_0} \|w\|_{\bar{\mathcal{Y}}_0}} \le \sqrt{2} \max\{1, A_{\max}\}, \tag{3.2.11}$$

*as well as*

$$\inf_{v \in \bar{\mathcal{X}}_0 \setminus \{0\}} \sup_{w \in \bar{\mathcal{Y}}_0 \setminus \{0\}} \frac{|\langle B_0 v, w\rangle|}{\|v\|_{\bar{\mathcal{X}}_0} \|w\|_{\bar{\mathcal{Y}}_0}} \ge \frac{\min\{1, A_{\min}\}}{\sqrt{2}}. \tag{3.2.12}$$

*Proof.* The proof is splitted mainly into three parts. First, we prove the explicit bounds (3.2.11) and (3.2.12). Then, we briefly recall modified parts of the proof [Tan13, Prop. 2.2] to show surjectivity (2.3.6) of the operator $B_0$.

The proof of the continuity estimate follows with Cauchy-Schwarz inequality, boundedness of the spatial differential operator and Hölder inequality.

$$\begin{aligned} |\langle B_0 x, y\rangle| &\le \int_0^T (\|\dot{x}\|_{W_-} \|y\|_{W_-'} + \|Ax\|_{W_-} \|y\|_{W_-'}) \,\mathrm{d}t \\ &\le \int_0^T (\|\dot{x}\|_{W_-} \|y\|_{W_-'} + A_{\max} \|x\|_{W_+} \|y\|_{W_-'}) \,\mathrm{d}t \\ &= \int_0^T \|y\|_{W_-'} (\|\dot{x}\|_{W_-} + A_{\max} \|x\|_{W_+}) \,\mathrm{d}t \\ &\le \left( \int_0^T \|y\|^2_{W_-'} \,\mathrm{d}t \right)^{1/2} \left( \int_0^T (\|\dot{x}\|_{W_-} + A_{\max} \|x\|_{W_+})^2 \,\mathrm{d}t \right)^{1/2} \\ &\le \left( \int_0^T \|y\|^2_{W_-'} \,\mathrm{d}t \right)^{1/2} \left( \int_0^T 2 \, (\|\dot{x}\|^2_{W_-} + A^2_{\max} \|x\|^2_{W_+}) \,\mathrm{d}t \right)^{1/2} \\ &\le \sqrt{2} \max\{1, A_{\max}\} \|y\|_{\bar{\mathcal{Y}}_0} \|x\|_{\bar{\mathcal{X}}_0}. \end{aligned}$$

Next we prove the inf-sup condition. For arbitrary $x \in \bar{\mathcal{X}}_0$, choose $y_x := R_{W_-}(Ax + \dot{x}) \in \bar{\mathcal{Y}}_0$, with Riesz isomorphism $R_{W_-} : W_- \to W_-'$, according to Riesz representation Theorem 2.1.4. With this choice we obtain

$$\begin{aligned}
{}_{\bar{\mathcal{Y}}_0'}\langle B_0 x, y_x\rangle_{\bar{\mathcal{Y}}_0} &= \int_0^T \left( {}_{W_-}\langle \dot{x}, R_{W_-} Ax\rangle_{W_-'} + {}_{W_-}\langle \dot{x}, R_{W_-}\dot{x}\rangle_{W_-'} \right) \mathrm{d}t \\
&\quad + \int_0^T \left( {}_{W_-}\langle Ax, R_{W_-} Ax\rangle_{W_-'} + ({}_{W_-}\langle Ax, R_{W_-}\dot{x}\rangle_{W_-'} \right) \mathrm{d}t \\
&= \int_0^T \left( (\dot{x}, Ax)_{W_-} + \|\dot{x}\|_{W_-}^2 + \|Ax\|_{W_-}^2 + (Ax, \dot{x})_{W_-} \right) \mathrm{d}t \\
&= \int_0^T (\|\dot{x}\|_{W_-}^2 + \|Ax\|_{W_-}^2)\mathrm{d}t + 2\int_0^T (Ax, \dot{x})_{W_-}\mathrm{d}t.
\end{aligned}$$

Therefore, by the embedding Theorem 2.2.18 $\bar{\mathcal{X}}_0 \hookrightarrow \mathcal{C}^0([0,T]; W_0)$ and the self-adjointness, we can estimate

$$\begin{aligned}
{}_{\bar{\mathcal{Y}}_0'}\langle B_0 x, y_x\rangle_{\bar{\mathcal{Y}}_0} &= \int_0^T (\|\dot{x}\|_{W_-}^2 + \|Ax\|_{W_-}^2)\mathrm{d}t + \int_0^T \frac{\mathrm{d}}{\mathrm{d}t}(Ax(t), x(t))_{W_-}\mathrm{d}t \\
&= \int_0^T (\|\dot{x}\|_{W_-}^2 + \|Ax\|_{W_-}^2)\mathrm{d}t + (Ax(T), x(T))_{W_-} \\
&= \int_0^T (\|\dot{x}\|_{W_-}^2 + \|Ax\|_{W_-}^2)\mathrm{d}t + \|A^{1/2}x(T)\|_{W_-}^2 \\
&\geq \int_0^T (\|\dot{x}\|_{W_-}^2 + \|Ax\|_{W_-}^2)\mathrm{d}t, \qquad (3.2.13)
\end{aligned}$$

since $A$ is assumed to be independent of time, where we considered $(\cdot,\cdot)_{W_-}$ as its unique extension by continuity. This can further be estimated as

$$\int_0^T (\|\dot{x}\|_{W_-}^2 + \|Ax\|_{W_-}^2)\mathrm{d}t \geq \int_0^T (\|\dot{x}\|_{W_-}^2 + A_{\min}^2\|x\|_{W_+}^2)\mathrm{d}t \geq \min\{1, A_{\min}^2\}\|x\|_{\bar{\mathcal{X}}_0}^2. \qquad (3.2.14)$$

Combining (3.2.13) and (3.2.14) with

$$\|y_x\|_{\bar{\mathcal{Y}}_0}^2 = \int_0^T \|R_{W_-}(Ax + \dot{x})\|_{W_-'}^2 \,\mathrm{d}t \leq 2\left(\int_0^T (\|Ax\|_{W_-}^2 + \|\dot{x}\|_{W_-}^2)\mathrm{d}t\right)$$

yields

$${}_{\bar{\mathcal{Y}}_0'}\langle Bx, y_x\rangle_{\bar{\mathcal{Y}}_0} \geq \frac{\min\{1, A_{\min}\}}{\sqrt{2}}\|y_x\|_{\bar{\mathcal{Y}}_0}\|x\|_{\bar{\mathcal{X}}_0}.$$

Finally, we mimic the proof of the non-degeneracy from [Tan13, Prop. 2.2] and adept it to our generalized situation by carefully tracking the appearing spaces. We also refer to [LM16b, Ch. 5] and section 5.1 for a slightly different approach including fractional

powers of $A$ and to [CS11, Th. 2.4] for a semigroup approach. To this end, we assume that there is a $y^* \in \bar{\mathcal{Y}}_0 \setminus \{0\}$ satisfying

$$ {}_{\bar{\mathcal{Y}}_0'}\langle B_0 x, y^* \rangle_{\bar{\mathcal{Y}}_0} = 0 \qquad \text{for all } x \in \bar{\mathcal{X}}_0. $$

We follow that

$$ \int_I {}_{W_-}\langle \dot{x}, y^* \rangle_{W_-'} \, \mathrm{d}t = - \int_I {}_{W_-}\langle Ax, y^* \rangle_{W_-'} \, \mathrm{d}t \lesssim \int_I \|x\|_{W_+} \|y^*\|_{W_-'} \, \mathrm{d}t < \infty $$

for all $x \in \bar{\mathcal{X}}_0$, where we drop the constants since they are not important at this stage. Therefore, we have that, additionally, $y^* \in L_2(I; W_-') \cap H_1(I; W_+')$ by the definition of weak derivatives and since

$$ \int_I {}_{W_-}\langle Ax, y^* \rangle_{W_-'} \, \mathrm{d}t = \int_I {}_{W_+}\langle x, A' y^* \rangle_{W_+'} \, \mathrm{d}t. $$

We keep the notation of the dual $A'$, though $A$ is assumed to be self-adjoint, since this part of the proof also holds for spatial differential operators which are not necessarily self-adjoint. Due to this regularity, we can integrate by parts and obtain

$$ \int_I \left( {}_{W_+}\langle x, -\dot{y}^* \rangle_{W_+'} + \int_I {}_{W_+}\langle x, A' y^* \rangle_{W_+'} \right) \, \mathrm{d}t + {}_{W_0}\langle x(T), y^*(T) \rangle_{W_0'} = 0 $$

for all $x \in \bar{\mathcal{X}}_0$. We arrive at the problem $-\dot{y}^*(t) + A' y^*(t) = 0$ in $W_+'$ for $t \in I$ a.e. with $y^*(T) = 0$ in $W_0'$ by testing with appropriate test functions. The affine transformation $\bar{y}^*(\cdot) := y(T - \cdot)$ yields, for $t \in I$ a.e.,

$$ \frac{\mathrm{d}}{\mathrm{d}t} \bar{y}^*(t) + A' \bar{y}^*(t) = 0, \qquad \bar{y}^*(0) = 0. \tag{3.2.15} $$

Interchanging to role $W_+ \leftrightarrow W_-'$ and $W_- \leftrightarrow W_+'$ in the proof of the inf-sup condition before, we arrive at

$$ 0 = {}_{L_2(I;W_+')}\langle \bar{B}_0 \bar{y}^*, z \rangle_{L_2(I;W_+)} \gtrsim \|\bar{y}^*\|^2_{L_2(I;W_-') \cap H_1(I;W_+')}, $$

for appropriately chosen $z \in L_2(I; W_+)$ and $\bar{B}_0$ defined as space-time weak operator according to (3.2.15). So we can follow that $\bar{y}^* \equiv 0$, what is a contradiction and therefore finishes the proof by the BNB-Theorem 2.3.3. □

It is not hard to see that the lower bound in (3.2.10) is indeed given by the inf-sup condition on $W_+$ and $W_-'$. By Corollary 2.3.8, one can express the inf-sup constant with respect to the norm of the inverse $A^{-1}$. Therefore, we can conclude

$$ 1 = \|A^{-1} A\|_{W_+ \to W_+} \leq \|A^{-1}\|_{W_- \to W_+} \|A\|_{W_+ \to W_-}, $$

such that

$$\|A\|_{W_+\to W_-} \geq \frac{1}{\|A^{-1}\|_{W_-\to W_+}} = \inf_{v\in W_+\setminus\{0\}} \sup_{w\in W'_-\setminus\{0\}} \frac{|\langle Av, w\rangle|}{\|v\|_{W_+}\|w\|_{W'_-}}.$$

The considered spatial spaces in the previous Theorem 3.2.9 are rather abstract. Reasonable choices for spatial differential operators $A$ of order $2m$ on a bounded domain $D \subset \mathbb{R}^n$ are Sobolev spaces of the form

$$W_+ := H^{m+\alpha}(D) \hookrightarrow W_0 = H^\alpha(D) \hookrightarrow W_- := H^{-m+\alpha}(D),$$

with suitable regularity shift $\alpha \in \mathbb{R}$ and possibly some boundary conditions. Without regularity shift $\alpha := 0$, we arrive at the initial problem with $A\colon H^m(D) \to H^{-m}(D)$. We also arrive at the case considered in [CS11] when choosing $W_+ := W$, $W_- := H$ according to the notation from [CS11], but additionally with explicitly determined bounds in Theorem 3.2.9.

We can also prove existence and uniqueness of a solution, when only a Gårding-inequality (3.1.2) is assumed together with error bounds in the very same way as in the previous chapter 3.1. Indeed, the proof of Corollary 3.1.16 applies also for the homogenization and yields explicit constants together with the boundedness and inf-sup estimates in this chapter. For the sake of completeness, we collect the results in the following corollary.

**Corollary 3.2.16.** *The operator $B_0 \in \mathcal{L}(\mathcal{X}_0, \mathcal{Y}_0')$ from (3.2.6) is still boundedly invertible, when we assume a Gårding inequality* (3.1.2) *instead of a coercivity* (3.1.1)*. The operator can be bounded as*

$$\sup_{v\in\mathcal{X}_0\setminus\{0\}} \sup_{w\in\mathcal{Y}_0\setminus\{0\}} \frac{|\langle B_0 v, w\rangle|}{\|v\|_{\mathcal{X}_0}\|w\|_{\mathcal{Y}_0}} \leq e^{\lambda T} \max\{\sqrt{1+2\lambda^2\varrho^4}, \sqrt{2}\}\sqrt{2}\max\{1, A_{\max}+\lambda\},$$

*as well as*

$$\inf_{v\in\mathcal{X}_0\setminus\{0\}} \sup_{w\in\mathcal{Y}_0\setminus\{0\}} \frac{|\langle B_0 v, w\rangle|}{\|v\|_{\mathcal{X}_0}\|w\|_{\mathcal{Y}_0}} \geq e^{-\lambda T} \frac{A_{\min}\min\{1, (A_{\max}+\lambda)^{-1}\}}{\sqrt{2}\max\{\sqrt{1+2\lambda^2\varrho^4}, \sqrt{2}\}},$$

*where $\varrho := \sup\limits_{0\neq v\in V} \frac{\|v\|_H}{\|v\|_V}$ the embedding constant of $V \hookrightarrow H$.*

*Proof.* The proof follows in the same way as the proof of Corollary 3.1.16 together with the relevant continuity and inf-sup estimates from this chapter. □

By combining Corollary 3.2.16 with Theorem 3.2.9, one can derive improved estimates in the setting of Corollary 3.2.16 for self-adjoint and time-independent spatial differential operators. Solving the homogenization instead of the standard first formulation, has the advantage, that there is no need to deal with the Cartesian product in the solution space any more. This in turn also simplifies the implementation. Another advantage is that we are able to improve our error estimates, since we do not have to consider the additional term in the norm of the solution space.

## 3.3 Second Formulation

The idea of the following *second formulation* aims at incorporating the initial condition in a natural way by using integration by parts. In this way, one also needs to require less smoothness of the solution. This is, in particular, essential to derive a well posed space-time weak formulation for the stochastic heat equation, see [LM16b], since the Wiener process is not weakly differentiable. This reformulation was basically also used in [CS11, LMN15, Tan13, LM16a, BJ89, BJ90]. Consider again the variational formulation (3.1.5). Instead of adding the initial condition as in section 3.1, one uses integration by parts and arrives at

$$\begin{aligned}\int_I {}_{V'}\langle \dot{u}(t), v(t)\rangle_V \,\mathrm{d}t &= -\int_I {}_V\langle u(t), \dot{v}(t)\rangle_{V'} \,\mathrm{d}t + (u(T), v(T))_H - (u(0), v(0))_H. \\ &= -\int_I {}_V\langle u(t), \dot{v}(t)\rangle_{V'} \,\mathrm{d}t - (u(0), v(0))_H, \end{aligned} \tag{3.3.1}$$

for test functions $v \in L_2(I;V) \cap H^1_{0,\{T\}}(I;V')$, where $H^1_{0,\{T\}}(I)$ denotes the set of functions in $H^1(I)$ which vanish at final time $T$, i.e.,

$$L_2(I;V) \cap H^1_{0,\{T\}}(I;V') := \{u \in L_2(I;V) \cap H^1(I;V') : u(T) \equiv 0\} \tag{3.3.2}$$

similar to (3.2.1). Therefore, we define our solution and test spaces as

$$\tilde{\mathcal{X}} := L_2(I;V), \qquad \tilde{\mathcal{Y}} := L_2(I;V) \cap H^1_{0,\{T\}}(I;V') \tag{3.3.3}$$

and arrive at the second full space-time weak formulation:
Find a solution $u \in \tilde{\mathcal{X}}$ such that

$$\tilde{b}(u,v) = \tilde{\mathcal{F}}(v) \qquad \text{for all } v \in \tilde{\mathcal{Y}}, \tag{3.3.4}$$

where the bilinear form $\tilde{b}(\cdot,\cdot)\colon \tilde{\mathcal{X}} \times \tilde{\mathcal{Y}} \to \mathbb{R}$ is defined as

$$\tilde{b}(u,v) := \int_I \left(-{}_V\langle u(t), \dot{v}(t)\rangle_{V'} + {}_{V'}\langle A(t)u(t), v(t)\rangle_V\right) \,\mathrm{d}t \tag{3.3.5}$$

and the right hand side $\tilde{\mathcal{F}}(\cdot)\colon \tilde{\mathcal{Y}} \to \mathbb{R}$ is given by

$$\tilde{\mathcal{F}}(v) := \int_I {}_{V'}\langle g(t), v(t)\rangle_V \,\mathrm{d}t + (u_0, v(0))_H. \tag{3.3.6}$$

Indeed, with (3.3.1) we immediately obtain

$$\begin{aligned}\tilde{b}(u,v) &= \int_I (\langle \dot{u}(t), v(t)\rangle + \langle A(t)u(t), v(t)\rangle) \,\mathrm{d}t + (u(0), v(0))_H \\ &= \int_I \langle g(t), v(t)\rangle \,\mathrm{d}t + (u_0, v(0))_H,\end{aligned}$$

for sufficiently smooth functions, so that the initial condition is naturally incorporated in the variational formulation (3.3.4) and only the test space is restricted. This second formulation is therefore also called *natural formulation*, e.g., in [Tan13]. The advantage is that the additional Cartesian product in the test space is eliminated. On the other hand, we need to restrict our test space to functions which vanish at the final time $T$, in order to obtain this formulation. If we think of a Petrov-Galerkin discretization, the ansatz functions of the test space need to satisfy other boundary conditions than the ansatz functions of the solution space in this formulation. This circumstance could lead to an under-determined system of equations due to the restriction of the test space. In this case, one would need to introduce additional ansatz functions for the test space, which generally destroy the structure of the system matrix. For instance, if we think of a hierarchical basis, one would need to add additional linear independent functions, e.g., with respect to finer levels. Nevertheless, it is not really a drawback, since one needs to choose a finer discretization for the test space anyway in order to stabilize our discrete problem according to section 4.2.

The existence and uniqueness of a solution is directly implied by the first formulation and the explicit bounds can be calculated along similar lines as the proof in [SS09, Appendix A], see also [CS11, Th. 2.2]. Again, there are improved bounds in [Tan13, Prop. 2.3], namely

$$\sup_{v\in\tilde{\mathcal{X}}\setminus\{0\}}\sup_{w\in\tilde{\mathcal{Y}}\setminus\{0\}}\frac{|\langle\tilde{B}v,w\rangle|}{\|v\|_{\tilde{\mathcal{X}}}\|w\|_{\tilde{\mathcal{Y}}}}\leq\sqrt{2}\max\{1,A_{\max}\} \tag{3.3.7}$$

as well as

$$\inf_{v\in\tilde{\mathcal{X}}\setminus\{0\}}\sup_{w\in\tilde{\mathcal{Y}}\setminus\{0\}}\frac{|\langle\tilde{B}v,w\rangle|}{\|v\|_{\tilde{\mathcal{X}}}\|w\|_{\tilde{\mathcal{Y}}}}\geq A_{\min}\frac{\min\{1,A_{\max}^{-1}\}}{\sqrt{2}}, \tag{3.3.8}$$

where $\tilde{B}\in\mathcal{L}(\tilde{\mathcal{X}},\tilde{\mathcal{Y}})$ is the unique operator defined by $\tilde{b}(\cdot,\cdot)\colon\tilde{\mathcal{X}}\times\tilde{\mathcal{Y}}\to\mathbb{R}$ analog to (3.1.11). So we obtain the same bounds as for the homogenized first version from section 3.2.

Moreover, we can prove a similar regularity result to Theorem 3.2.9 with improved estimates for time-independent spatial differential operators. To this end, we consider the spaces

$$\begin{aligned}\bar{\tilde{\mathcal{X}}} &:= L_2(I;W_-'),\\ \bar{\tilde{\mathcal{Y}}} &:= L_2(I;W_+)\cap H^1_{0,\{T\}}(I;W_-)\end{aligned} \tag{3.3.9}$$

with continuous and dense embedding $W_+\hookrightarrow W_0\hookrightarrow W_-$ of separable Hilbert spaces and interpolation space $W_0:=[W_-,W_+]_{1/2}$, recalling (2.2.17) for the definition of interpolation spaces. The following corollary is the counterpart of Theorem 3.2.9 for the second formulation.

**Corollary 3.3.10.** *Let $W_+\hookrightarrow W_-$ be separable Hilbert spaces with scalar products $(\cdot,\cdot)_{W_+}$ and $(\cdot,\cdot)_{W_-}$, respectively, and $W_0:=[W_-,W_+]_{1/2}$ the interpolation space endowed with scalar product $(\cdot,\cdot)_{W_0}$. Assume that the operator $A$ is time-independent and*

*self-adjoint with*

$$A_{\min} \leq \|A\|_{W_+ \to W_-} \leq A_{\max},$$

*with constants* $0 < A_{\min} \leq A_{\max} < \infty$. *Then* $\tilde{B}\colon \bar{\tilde{\mathcal{X}}} \to \bar{\tilde{\mathcal{Y}}}'$ *is boundedly invertible with*

$$\sup_{v \in \bar{\tilde{\mathcal{X}}} \setminus \{0\}} \sup_{w \in \bar{\tilde{\mathcal{Y}}} \setminus \{0\}} \frac{|\langle \tilde{B}v, w \rangle|}{\|v\|_{\bar{\tilde{\mathcal{X}}}} \|w\|_{\bar{\tilde{\mathcal{Y}}}}} \leq \sqrt{2} \max\{1, A_{\max}\}, \tag{3.3.11}$$

*as well as*

$$\inf_{v \in \bar{\tilde{\mathcal{X}}} \setminus \{0\}} \sup_{w \in \bar{\tilde{\mathcal{Y}}} \setminus \{0\}} \frac{|\langle \tilde{B}v, w \rangle|}{\|v\|_{\bar{\tilde{\mathcal{X}}}} \|w\|_{\bar{\tilde{\mathcal{Y}}}}} \geq \frac{\min\{1, A_{\min}\}}{\sqrt{2}}. \tag{3.3.12}$$

*Proof.* The proof of the continuity estimate follows in the same way as in the proof of Proposition 3.2.9. Next we prove the inf-sup condition, but with swapped solution and test spaces first. Although $A$ is assumed to be self-adjoint, it is convenient to keep the notation $A'$ to make clear that $A$ naturally maps from $W_-'$ to $W_+'$ in this setting, such that consequently $A'$ maps from $W_+$ to $W_-$. Different from the proof of Theorem 3.2.9, we choose $x_y := R_{W_-}(A'y - \dot{y}) \in \bar{\tilde{\mathcal{X}}}$ (solution space), with Riesz isomorphism $R_{W_-}\colon W_- \to W_-'$, according to Riesz representation Theorem 2.1.4 for arbitrary $y \in \bar{\tilde{\mathcal{Y}}}$ (test space). With this choice we obtain

$$\begin{aligned}
{}_{\bar{\tilde{\mathcal{Y}}}^*}\langle \tilde{B}x_y, y \rangle_{\bar{\tilde{\mathcal{Y}}}} &= \int_0^T \Big( -{}_{W_-'}\langle R_{W_-}A'y, \dot{y} \rangle_{W_-} + {}_{W_-'}\langle R_{W_-}\dot{y}, \dot{y} \rangle_{W_-} \Big) \,\mathrm{d}t \\
&\quad + \int_0^T \Big( {}_{W_+'}\langle AR_{W_-}A'y, y \rangle_{W_+} - {}_{W_+'}\langle AR_{W_-}\dot{y}, y \rangle_{W_+} \Big) \,\mathrm{d}t \\
&= \int_0^T \big( -(A'y, \dot{y})_{W_-} + \|\dot{y}\|_{W_-}^2 + \|A'y\|_{W_-}^2 - (\dot{y}, A'y)_{W_-} \big) \,\mathrm{d}t \\
&= \int_0^T (\|\dot{y}\|_{W_-}^2 + \|A'y\|_{W_-}^2) \mathrm{d}t - 2 \int_0^T (A'y, \dot{y})_{W_-} \mathrm{d}t.
\end{aligned}$$

Therefore, by the embedding Theorem 2.2.18 $\bar{\tilde{\mathcal{Y}}} \hookrightarrow \mathcal{C}^0([0,T]; W_0)$ and the self-adjointness, we can estimate

$$\begin{aligned}
{}_{\bar{\tilde{\mathcal{Y}}}^*}\langle Bx_y, y \rangle_{\bar{\tilde{\mathcal{Y}}}} &= \int_0^T (\|\dot{y}\|_{W_-}^2 + \|A'y\|_{W_-}^2) \mathrm{d}t - \int_0^T \frac{\mathrm{d}}{\mathrm{d}t} (A'y(t), y(t))_{W_-} \mathrm{d}t \\
&= \int_0^T (\|\dot{y}\|_{W_-}^2 + \|A'y\|_{W_-}^2) \mathrm{d}t + \|A^{1/2}y(0)\|_{W_-}^2 \\
&\geq \int_0^T (\|\dot{y}\|_{W_-}^2 + \|A'y\|_{W_-}^2) \mathrm{d}t.
\end{aligned} \tag{3.3.13}$$

This can further be estimated as

$$\begin{aligned}
\int_0^T (\|\dot{y}\|_{W_-}^2 + \|A'y\|_{W_-}^2) \mathrm{d}t &\geq \int_0^T (\|\dot{y}\|_{W_-}^2 + A_{\min}^2 \|y\|_{W_+}^2) \mathrm{d}t \\
&\geq \min\{1, A_{\min}^2\} \|y\|_{\bar{\tilde{\mathcal{Y}}}}^2.
\end{aligned} \tag{3.3.14}$$

Combining (3.3.13) and (3.3.14) with

$$\|x_y\|_{\bar{\tilde{\mathcal{X}}}}^2 = \int_0^T \|R_{W_-}(A'y - \dot{y})\|_{W_-'}^2 \,\mathrm{d}t \leq 2\left(\int_0^T (\|A'y\|_{W_-}^2 + \|\dot{y}\|_{W_-}^2)\mathrm{d}t\right)$$

yields

$${}_{\bar{\tilde{\mathcal{Y}}}'}\langle Bx_y, y\rangle_{\bar{\tilde{\mathcal{Y}}}} \geq \frac{\min\{1, A_{\min}\}}{\sqrt{2}} \|x_y\|_{\bar{\tilde{\mathcal{X}}}'} \|y\|_{\bar{\tilde{\mathcal{Y}}}}.$$

Now we have proven that

$$\inf_{w \in \bar{\tilde{\mathcal{Y}}}\setminus\{0\}} \sup_{v \in \bar{\tilde{\mathcal{X}}}\setminus\{0\}} \frac{|\langle \tilde{B}v, w\rangle|}{\|v\|_{\bar{\tilde{\mathcal{X}}}} \|w\|_{\bar{\tilde{\mathcal{Y}}}}} \geq \frac{\min\{1, A_{\min}\}}{\sqrt{2}}.$$

What remains to show is the surjectivity from Theorem 2.3.3 with swapped solution and test space, see Proposition 2.3.7. Again we prove it along the lines of the proof of Proposition 3.2.9 relying itself on [Tan13, Prop. 2.2 and Prop. 2.3].

To this end, we assume that there is a $x^* \in \bar{\tilde{\mathcal{X}}} \setminus \{0\}$ satisfying

$${}_{\bar{\tilde{\mathcal{Y}}}'}\langle \tilde{B}x^*, y\rangle_{\bar{\tilde{\mathcal{Y}}}} = 0 \qquad \text{for all } y \in \bar{\tilde{\mathcal{Y}}}.$$

We follow that

$$\int_I {}_{W_-'}\langle x^*, \dot{y}\rangle_{W_-} \,\mathrm{d}t = \int_I {}_{W_-'}\langle x^*, A'y\rangle_{W_-} \,\mathrm{d}t \lesssim \int_I \|x^*\|_{W_-'} \|y\|_{W_+} \,\mathrm{d}t < \infty$$

for all $y \in \bar{\tilde{\mathcal{Y}}}$, where we drop the constants since they are not important at this stage. Therefore, we have that, additionally, $x^* \in L_2(I; W_-') \cap H_1(I; W_+')$ by the definition of weak derivatives and since

$$\int_I {}_{W_-'}\langle x^*, A'y\rangle_{W_-} \,\mathrm{d}t = \int_I {}_{W_+'}\langle Ax^*, y\rangle_{W_+} \,\mathrm{d}t.$$

Due to this regularity, we can integrate by parts and obtain

$$\int_I \left( {}_{W_+'}\langle \dot{x}^*, y\rangle_{W_+} + \int_I {}_{W_+'}\langle Ax^*, y\rangle_{W_+} \right) \mathrm{d}t + {}_{W_0'}\langle x^*(0), y(0)\rangle_{W_0} = 0$$

for all $y \in \bar{\tilde{\mathcal{Y}}}$. So we see that $\dot{x}^*(t) + Ax^*(t) = 0$ in $W_+'$ for $t \in I$ a.e. with $x^*(0) = 0$ in $W_0'$. This is exactly the homogenization from subsection 3.2, where the invertibility was already proven. To this end, we can conclude immediately that $x^* \equiv 0$, what is a contradiction and therefore yields surjectivity.

Putting everything together, we have proven the second condition from Proposition 2.3.7, so that we are allowed to swap the spaces in the inf-sup conditions. Therefore, we obtain

$$\inf_{v \in \bar{\tilde{\mathcal{X}}}\setminus\{0\}} \sup_{w \in \bar{\tilde{\mathcal{Y}}}\setminus\{0\}} \frac{|\langle \tilde{B}v, w\rangle|}{\|v\|_{\bar{\tilde{\mathcal{X}}}} \|w\|_{\bar{\tilde{\mathcal{Y}}}}} = \inf_{w \in \bar{\tilde{\mathcal{Y}}}\setminus\{0\}} \sup_{v \in \bar{\tilde{\mathcal{X}}}\setminus\{0\}} \frac{|\langle \tilde{B}v, w\rangle|}{\|v\|_{\bar{\tilde{\mathcal{X}}}} \|w\|_{\bar{\tilde{\mathcal{Y}}}}} \geq \frac{\min\{1, A_{\min}\}}{\sqrt{2}},$$

what finishes the proof. □

Reasonable choices for $W_-$ and $W_+$ are already mentioned at the end of subsection 3.2. In [LM16a] the second formulation was also given in a slightly different form

$$\begin{aligned}&\int_I (-\langle u_1(t), \dot{v}(t)\rangle + \langle A(t)u_1(t), v(t)\rangle)\ \mathrm{d}t + \langle u_2, v(T)\rangle \\ &= \int_I \langle g(t), v(t)\rangle\ \mathrm{d}t + (u_0, v(0))_H,\end{aligned} \tag{3.3.15}$$

with $(u_1, u_2) \in L_2(I;V) \times H$ and $v \in L_2(I;V) \cap H^1(I;V')$. That means, the authors, different from our formulation, do not assume zero final time conditions in the test space. This formulation is very similar to the first formulation since it contains a Cartesien product and treat the initial condition essentially. It can be shown, that $u_1(T) = u_2$ for sufficiently smooth right hand side. Otherwise, one can interpret $u_2$ as a continuous version of $u_1$ evaluated at final time $t = T$. It can be convenient to keep the additional element $v_2$ in order to prove different error estimates in $L_\infty(I;H)$ also containing point wise defined norms as done in [LM16a].

Again, we also want to state the counterpart of Corollary 3.1.16 and 3.2.16 for non-coercive spatial differential operators.

**Corollary 3.3.16.** *The operator $\tilde{B} \in \mathcal{L}(\tilde{\mathcal{X}}, \tilde{\mathcal{Y}}')$ defined via (3.3.5) is still boundedly invertible, when we assume a Gårding inequality* (3.1.2) *instead of coercivity* (3.1.1). *The operator can be bounded as*

$$\sup_{v\in\tilde{\mathcal{X}}\setminus\{0\}} \sup_{w\in\tilde{\mathcal{Y}}\setminus\{0\}} \frac{|\langle Bv, w\rangle|}{\|v\|_{\tilde{\mathcal{X}}}\|w\|_{\tilde{\mathcal{Y}}}} \leq e^{\lambda T} \max\{\sqrt{1+2\lambda^2\varrho^4}, \sqrt{2}\}\sqrt{2}\max\{1, A_{\max}+\lambda\},$$

*as well as*

$$\inf_{v\in\tilde{\mathcal{X}}\setminus\{0\}} \sup_{w\in\tilde{\mathcal{Y}}\setminus\{0\}} \frac{|\langle Bv, w\rangle|}{\|v\|_{\tilde{\mathcal{X}}}\|w\|_{\tilde{\mathcal{Y}}}} \geq e^{-\lambda T} \frac{A_{\min}\min\{1, (A_{\max}+\lambda)^{-1}\}}{\sqrt{2}\max\{\sqrt{1+2\lambda^2\varrho^4}, \sqrt{2}\}},$$

*where $\varrho := \sup_{0\neq v\in V} \frac{\|v\|_H}{\|v\|_V}$ denotes the embedding constant of $V \hookrightarrow H$.*

*Proof.* The proof follows in a similar way as the proof of Corollary 3.1.16 together with the relevant continuity and inf-sup estimates from this chapter. □

We conclude this chapter with some remarks about the connection between the different formulations introduced in the present chapter. Considering the second formulation, we have that for $h \in \tilde{\mathcal{X}}'$ and $u \in \tilde{\mathcal{Y}}$

$${}_{\tilde{\mathcal{X}}'}\langle \tilde{B}'u, v\rangle_{\tilde{\mathcal{X}}} = {}_{\tilde{\mathcal{Y}}}\langle u, \tilde{B}v\rangle_{\tilde{\mathcal{Y}}'} = \int_I {}_{V'}\langle -\dot{u}(t) + A'(t)u(t), v(t)\rangle_V\ \mathrm{d}t = h(v)$$

for all $v \in \tilde{\mathcal{X}}$, is the full weak formulation of the dual problem of finding, for $t \in I$ a.e., a solution $u(t) \in V'$ such that

$$-\dot{u}(t) + A'(t) = h \quad \text{in } V', \qquad u(T) = 0.$$

With the affine transformations $\bar{u} := u(T - \cdot)$, $\bar{A} := A'(T - \cdot)$ and $\bar{h} := h(T - \cdot)$ it is equivalent to

$$\frac{\mathrm{d}}{\mathrm{d}t}\bar{u}(t) + \bar{A}(t)\bar{u}(t) = \bar{h} \quad \text{in } V', \qquad \bar{u}(0) = 0,$$

which is the homogenized problem of section 3.2 with spatial differential operator $\bar{A}$ and right hand side $\bar{h}$. This means that the second formulation (3.3.4) is connected with the homogenization of the first formulation (3.2.3) via its dual problem. The same holds for the first formulation (3.1.6) and the modified second formulation (3.3.15). Nevertheless, note that neither (3.2.3) coincides with the dual problem of (3.3.4) nor is (3.3.15) the dual problem of (3.1.6), but closely related.

We have summarized the particular estimates of the continuity and inf-sup constants of the different formulations above in Table 3.3.17.

Table 3.3.17: Continuity and inf-sup estimates for different formulations.

| Formulation | Continuity | Inf-sup |
|---|---|---|
| First form, coercive | $\sqrt{2}\max\{1, A_{\max}^2\} + \rho^2$ | $\frac{\min\{A_{\min}, A_{\max}^{-1}, A_{\min}A_{\max}^{-1}\}}{2}$ |
| First form, Gårding-inequality | $e^{\lambda T}\max\{\sqrt{1+2\lambda^2\varrho^4}, \sqrt{2}\}$ $\times \sqrt{2}\max\{1, (A_{\max}+\lambda)^2\} + \rho^2$ | $e^{-\lambda T} \times$ $\frac{\min\{A_{\min}, (A_{\max}+\lambda)^{-1}, A_{\min}(A_{\max}+\lambda)^{-1}\}}{2\max\{\sqrt{1+2\lambda^2\varrho^4}, \sqrt{2}\}}$ |
| Homogenization, coercive | $\sqrt{2}\max\{1, A_{\max}\}$ | $A_{\min}\frac{\min\{1, A_{\max}^{-1}\}}{\sqrt{2}}$ |
| Homogenization, coercive, time-independent, self-adjoint | $\sqrt{2}\max\{1, A_{\max}\}$ | $\frac{\min\{1, A_{\min}\}}{\sqrt{2}}$ |
| Homogenization, Gårding-inequality | $e^{\lambda T}\max\{\sqrt{1+2\lambda^2\varrho^4}, \sqrt{2}\}$ $\times \sqrt{2}\max\{1, A_{\max}+\lambda\}$ | $e^{-\lambda T}\frac{A_{\min}\min\{1, (A_{\max}+\lambda)^{-1}\}}{\sqrt{2}\max\{\sqrt{1+2\lambda^2\varrho^4}, \sqrt{2}\}}$ |
| Second form, coercive | $\sqrt{2}\max\{1, A_{\max}\}$ | $A_{\min}\frac{\min\{1, A_{\max}^{-1}\}}{\sqrt{2}}$ |
| Second form, coercive, time-independent, self-adjoint | $\sqrt{2}\max\{1, A_{\max}\}$ | $\frac{\min\{1, A_{\min}\}}{\sqrt{2}}$ |
| Second form, Gårding-inequality | $e^{\lambda T}\max\{\sqrt{1+2\lambda^2\varrho^4}, \sqrt{2}\}$ $\times \sqrt{2}\max\{1, A_{\max}+\lambda\}$ | $e^{-\lambda T}\frac{A_{\min}\min\{1, (A_{\max}+\lambda)^{-1}\}}{\sqrt{2}\max\{\sqrt{1+2\lambda^2\varrho^4}, \sqrt{2}\}}$ |

# 4 Petrov-Galerkin Approach

This chapter deals with the discretization of generic PDEs as given in the previous section for instance. The descriptions are kept rather general first, but are also applied to the results concerning parabolic problems in space-time formulation introduced in chapter 3. A Petrov-Galerkin approach is known to be *quasi-optimal* provided that certain stability assumptions are fulfilled, as we will see in this chapter. After introducing Petrov-Galerkin approaches and, more precisely, *minimal residual Petrov-Galerkin approaches*, we focus on the stability of such a discretization. This is by no means trivial when dealing with different solution and test spaces as they appear in parabolic problems and it needs to be elaborated very precisely. This difficulty appears since the operator induced by the PDE only satisfies an inf-sup condition (2.3.5), which is not inherited to its Petrov-Galerkin discretization in general. The stability results stem from [Mol13b], revised and described in more detail for this thesis.

## 4.1 General Setting

**Definition 4.1.1.** *Considering a generic operator equation*

$$B(u) = f, \quad u \in X, \quad f \in Y', \tag{4.1.2}$$

*with unknown solution $u \in X$, given right hand side $f \in Y'$ and boundedly invertible operator $B\colon X \to Y'$ mapping from a Hilbert space $X$ into a dual Hilbert space $Y'$. The* Petrov-Galerkin solution $u_j \in S_j$ *is given as the solution of the variational problem*

$$ {}_{Y'}\langle B(u_j), q_\ell\rangle_Y = {}_{Y'}\langle f, q_\ell\rangle_Y \qquad \text{for all } q_\ell \in Q_\ell, \tag{4.1.3}$$

*with respect to discrete subspaces $S_j \subset X$ and $Q_\ell \subset Y$.*

For $X \neq Y$ the test functions $q_\ell \in Q_\ell$ are generally different from the ansatz functions in $S_j$. If the solution and test spaces $X = Y$ are equal, one usually chooses $Q_\ell = S_j$ and calls (4.1.3) *Galerkin discretization* and *Ritz-Galerkin discretization* if additionally $\langle B(\cdot), \cdot\rangle$ is symmetric and coercive.

In order to arrive at our stability results in section 4.2, we allow discrete test spaces of higher dimension than that of the solution space. That means, that (4.1.3) is generally not solvable exactly, so we need to introduce a least-square like approach, cf. [And13].

**Definition 4.1.4.** *The* minimal residual Petrov-Galerkin solution $u_j$ *of an operator equation* (4.1.2) *is given as the minimizer of the functional residual*

$$u_j := \operatorname*{arg\,min}_{v_j \in S_j} \sup_{q_\ell \in Q_\ell \setminus \{0\}} \frac{|\langle B(v_j) - f, q_\ell\rangle|}{\|q_\ell\|_Y}. \tag{4.1.5}$$

It can be easily seen that the minimizer (4.1.5) is equivalent to

$$u_j = \arg\min_{v_j \in S_j} \|f - B(v_j)\|_{Q'_\ell},$$

where

$$\|\tilde{q}_\ell\|_{Q'_\ell} := \sup_{q_\ell \in Q_\ell \setminus \{0\}} \frac{|\langle \tilde{q}_\ell, q_\ell \rangle|}{\|q_\ell\|_Y}, \qquad \tilde{q}_\ell \in Q'_\ell$$

defines a dual norm on $Q'_\ell \supset Y'$. Although the minimizer (4.1.5) also depends on the level $\ell$ of the test space, we only index it by the level $j$ of the solution space in order to point out that $u_j \in S_j$. In case of linear operators, the computation typically leads to solving an overdetermined linear system of equation, so that the numerical solution can be obtained by solving the associated least squares problem respectively the modified Gauss normal equation, see (4.1.14). A Petrov-Galerkin solution of (4.1.3) obviously also solves the minimal residual approach (4.1.5) but not necessarily vice versa. Nevertheless, we usually also call the minimal residual Petrov-Galerkin solution simply Petrov-Galerkin solution for convenience.

For *linear* operators, the stability of a (minimal residual) Petrov-Galerkin approach is characterized by the *discrete inf-sup condition*

$$\inf_{v_j \in S_j \setminus \{0\}} \sup_{q_\ell \in Q_\ell \setminus \{0\}} \frac{|\langle B v_j, q_\ell \rangle|}{\|v_j\|_X \|q_\ell\|_Y} =: \beta_{j,\ell} > 0. \tag{4.1.6}$$

Even if a linear operator $B$ from (4.1.2) satisfies the inf-sup condition (2.3.5) according to the BNB Theorem 2.3.3, it generally does *not* imply its discrete counterpart (4.1.6). The continuity (2.3.4), on the other hand, is inherited to the discrete subspaces

$$\sup_{v_j \in S_j \setminus \{0\}} \sup_{q_\ell \in Q_\ell \setminus \{0\}} \frac{|\langle B v_j, q_\ell \rangle|}{\|v_j\|_X \|q_\ell\|_Y} \leq \sup_{v \in X \setminus \{0\}} \sup_{q \in Y \setminus \{0\}} \frac{|\langle B v, q \rangle|}{\|v\|_X \|q\|_Y} \leq B_{\max} < \infty. \tag{4.1.7}$$

It is of particular importance for the uniform stability, that the constants $\beta_{j,\ell}$ can be bounded uniformly from below by a constant $\beta > 0$ which is *independent* of the discretization represented by $j$ and $\ell$. Otherwise, a stability problem appears for decreasing $\beta_{j,\ell}$, when $j, \ell \to \infty$. So the discrete inf-sup condition (4.1.6) plays an important role for quasi-optimality of Petrov-Galerkin solutions respectively minimal residual Petrov-Galerkin solutions, since the quasi-optimality constant depends reciprocally on the discrete inf-sup constant as shown below. This follows from the following theorem, see [And13, Th. 3.1].

**Theorem 4.1.8.** *Let $B \in \mathcal{L}(X, Y')$ and assume that the discrete inf-sup condition (4.1.6) is satisfied with respect to $S_j \subset X$ and $Q_\ell \subset Y$. Then for any $u \in X$ there exists a unique $u_j \in S_j$ which satisfies*

$$\mathcal{R}_\ell(u_j) = \inf_{v_j \in S_j} \mathcal{R}_\ell(v_j), \qquad \mathcal{R}_\ell(v_j) := \sup_{q_\ell \in Q_\ell \setminus \{0\}} \frac{|\langle B v_j - B u, q_\ell \rangle|}{\|q_\ell\|_Y}.$$

*Moreover, there holds the* quasi-optimality *estimate*

$$\|u - u_j\|_X \leq \frac{B_{\max}}{\beta_{j,\ell}} \inf_{v_j \in S_j} \|u - v_j\|_X, \tag{4.1.9}$$

*with discrete inf-sup constant* $\beta_{j,\ell}$ *given by* (4.1.6) *and continuity constant* $B_{\max}$ *given by* (4.1.7).

Estimate (4.1.9) in Theorem 4.1.8 is strongly connected to the Céa-Lemma, which yields essentially the same quasi-optimality result for bounded and *coercive* bilinear forms, cf. [Bra07, Lem. 4.2], and also [Bab71, Th. 2.2] or [XZ03, Th. 2]. Thus, if the discrete inf-sup condition (4.1.6) is satisfied uniformly, i.e., $\inf_{j,\ell} \beta_{j,\ell} \geq \beta > 0$ for some $\beta$ independent of $j$ and $\ell$, then $u_j$ is the quasi-optimal approximation of $u$ with

$$\|u - u_j\|_X \leq \frac{B_{\max}}{\beta} \inf_{v_j \in S_j} \|u - v_j\|_X.$$

Otherwise, if $\beta_{j,\ell}$ tends to zero for $j, \ell \to \infty$, the quasi-optimality constants $\frac{B_{\max}}{\beta_{j,\ell}}$ in (4.1.9) blows up.

Finally, we will present how a (minimal residual) Petrov-Galerkin solution can be computed. Let $\Phi := \{\phi_1, \dots, \phi_{N_j}\}$ be a basis of $S_j$ with dimension $\dim S_j = N_j$ and $\Theta := \{\theta_1, \dots, \theta_{N_\ell}\}$ a basis of $Q_\ell$ with dimension $\dim Q_\ell = N_\ell$. In order to highlight the difference between a standard Petrov-Galerkin approach (4.1.3) and the minimal residual Petrov-Galerkin approach (4.1.5), we first want to consider the case (4.1.3) with $N_j = N_\ell$.

We can write the solution $u_j \in S_j$ as a linear combination

$$u_j = \sum_{i=1}^{N_j} c_i \, \phi_i, \qquad c_1, \dots, c_{N_j} \in \mathbb{R}.$$

Since equation (4.1.3) holds for all basis functions we obtain

$$\begin{aligned} \langle B(u_j), \theta_1 \rangle &= \langle f, \theta_1 \rangle \\ &\vdots \\ \langle B(u_j), \theta_{N_\ell} \rangle &= \langle f, \theta_{N_\ell} \rangle, \end{aligned}$$

with Petrov-Galerkin solution $u_j \in S_j$. So we have to solve the equation

$$\mathbf{B}_{j,\ell}(\mathbf{u}_j) := (\langle B(u_j), \theta_i \rangle)_{i=1}^{N_\ell} = (\langle B(\mathbf{u}_j^T \Phi), \theta_i \rangle)_{i=1}^{N_\ell} = (\langle f, \theta_i \rangle)_{i=1}^{N_\ell} =: \mathbf{f}_\ell \in \mathbb{R}^{N_\ell}, \tag{4.1.10}$$

with $u_j = \mathbf{u}_j^T \Phi := \sum_{i=1}^{N_j} c_i \, \phi_i$ and $\mathbf{u}_j := (c_1, \dots, c_{N_j})^T$, where we used the shorthand notation of $\Phi$ synonymously also for the vector of basis functions $(\phi_1, \dots, \phi_{N_j})^T$. For linear operators $B \in \mathcal{L}(X, Y')$, we can simplify

$$\langle B u_j, q_\ell \rangle = \langle B(\sum_{i=1}^{N_j} c_i \, \phi_i), q_\ell \rangle = \sum_{i=1}^{N_j} c_i \, \langle B\phi_i, q_\ell \rangle, \quad q_\ell \in Q_\ell,$$

so that (4.1.10) can be written as a matrix-vector equation

$$\mathbf{B}_{j,\ell}\mathbf{u}_j = \mathbf{f}_\ell, \qquad \text{with } \mathbf{B}_{j,\ell} \in \mathbb{R}^{N_\ell \times N_j}, \tag{4.1.11}$$

where $(\mathbf{B}_{j,\ell})_{i,k} := \langle B\phi_k, \theta_i\rangle$ for $k = 1, \dots, N_j$ and $i = 1, \dots, N_\ell$. The solution of (4.1.10) or (4.1.11) can be computed by various methods from numerical linear algebra as, e.g., Newton's method or conjugate gradient method, depending on the properties of the operator. As already mentioned before, we restrict ourselves to linear problems. To this end, the subsequent considerations are for linear equations, i.e., (4.1.11). Nevertheless, the following description can be extended to non-linear operators straightforwardly.

Next we want to derive an algebraic residual minimization problem connected with (4.1.5). In this regard, notice first that we want to minimize the residual with respect to the dual norm $\|\cdot\|_{Q'_\ell}$. Therefore, we introduce the Riesz-map $R_Y \colon Y \to Y'$ defined via the Riesz representation theorem by

$${}_{Y'}\langle R_Y v, w\rangle_Y = (v, w)_Y, \qquad v, w \in Y. \tag{4.1.12}$$

Using this Riesz-map, we can express the norm of an element $v \in Y$ by $\|v\|_Y^2 = {}_{Y'}\langle\, R_Y v, v\rangle_Y$, so that we obtain

$$u_j = \operatorname*{arg\,min}_{v_j \in S_j} \sup_{q_\ell \in Q_\ell \setminus \{0\}} \frac{|\langle Bv_j - f, q_\ell\rangle|}{\sqrt{\langle R_Y q_\ell, q_\ell\rangle}}.$$

An analog discretization as in the standard Petrov-Galerkin case yields

$$\mathbf{q}_\ell^T(\mathbf{B}_{j,\ell}\mathbf{v}_j - \mathbf{f}_\ell) = \langle B(v_j) - f, q_\ell\rangle, \quad \mathbf{q}_\ell^T\mathbf{R}_Y\mathbf{q}_\ell = \langle R_y q_\ell, q_\ell\rangle, \tag{4.1.13}$$

with $\mathbf{B}_{j,\ell}$, $\mathbf{v}_j$, $\mathbf{q}_\ell$, $\mathbf{f}_\ell$ and $\mathbf{R}_Y$ defined analogously as in (4.1.10). The equality follows indeed directly by inserting the definition of the coefficients. The discretized Riesz operator $\mathbf{R}_Y$ is the Gram matrix with respect to the bilinear form $(\cdot,\cdot)_Y$ and (4.1.12). Now it can be deduced that

$$\mathbf{u}_j = \operatorname*{arg\,min}_{\mathbf{v}_j \in \mathbb{R}^{N_j}} \|\mathbf{B}_{j,\ell}\mathbf{v}_j - \mathbf{f}_\ell\|_{\mathbf{R}_Y^{-1}},$$

with $\|\mathbf{w}_\ell\|_{\mathbf{R}_Y^{-1}}^2 := \mathbf{w}_\ell^T\mathbf{R}_Y^{-1}\mathbf{w}_\ell = \|\mathbf{R}_Y^{-\frac{1}{2}}\mathbf{w}_\ell\|_{\ell_2(\mathbb{R}^{N_\ell})}^2$, for $\mathbf{w}_\ell \in \mathbb{R}^{N_\ell}$, so that the minimizer $\mathbf{u}_j \in \mathbb{R}^{N_j}$ is given as the unique solution of the (modified) Gauss normal equation

$$\mathbf{B}_{j,\ell}^T\mathbf{R}_Y^{-1}\mathbf{B}_{j,\ell}\mathbf{u}_j = \mathbf{B}_{j,\ell}^T\mathbf{R}_Y^{-1}\mathbf{f}_\ell. \tag{4.1.14}$$

This can be seen by noting that

$$\begin{aligned}
\|\mathbf{B}_{j,\ell}\mathbf{v}_j - \mathbf{f}_\ell\|_{\mathbf{R}_Y^{-1}} &= \|\mathbf{R}_Y^{-1/2}(\mathbf{B}_{j,\ell}\mathbf{v}_j - \mathbf{f}_\ell)\|_{\ell_2(\mathbb{R}^{N_\ell})} = \sup_{0\neq\tilde{\mathbf{q}}_\ell\in\mathbb{R}^{N_\ell}} \frac{|\tilde{\mathbf{q}}_\ell^T\mathbf{R}_Y^{-1/2}(\mathbf{B}_{j,\ell}\mathbf{v}_j - \mathbf{f}_\ell)|}{\|\tilde{\mathbf{q}}_\ell\|_{\ell_2(\mathbb{R}^{N_\ell})}} \\
&= \sup_{0\neq\tilde{\mathbf{q}}_\ell\in\mathbb{R}^{N_\ell}} \frac{|(\mathbf{R}_Y^{-1/2}\tilde{\mathbf{q}}_\ell)^T(\mathbf{B}_{j,\ell}\mathbf{v}_j - \mathbf{f}_\ell)|}{\|\mathbf{R}_Y^{-1/2}\tilde{\mathbf{q}}_\ell\|_{\mathbf{R}_Y}} = \sup_{0\neq\mathbf{q}_\ell\in\mathbb{R}^{N_\ell}} \frac{|\mathbf{q}_\ell^T(\mathbf{B}_{j,\ell}(\mathbf{v}_j) - \mathbf{f}_\ell)|}{\|\mathbf{q}_\ell\|_{\mathbf{R}_Y}} \\
&= \sup_{0\neq\mathbf{q}_\ell\in\mathbb{R}^{N_\ell}} \frac{|\mathbf{q}_\ell^T(\mathbf{B}_{j,\ell}\mathbf{v}_j - \mathbf{f}_\ell)|}{\sqrt{\mathbf{q}_\ell^T\mathbf{R}_Y\mathbf{q}_\ell}} = \sup_{q_\ell\in Q_\ell\setminus\{0\}} \frac{|\langle Bv_j - f, q_\ell\rangle|}{\sqrt{\langle R_Y q_\ell, q_\ell\rangle}},
\end{aligned}$$

where we have used that $\mathbf{R}_Y$ is symmetric positive definite, induced by a Riesz-map. Comparing (4.1.11) with (4.1.14), we observe that we have to deal additionally with the discretized Riesz-map $\mathbf{R}_Y$, when solving the minimal residual Petrov-Galerkin approach with $N_\ell > N_j$. It is sometimes helpful to replace the norm $\|\cdot\|_Y^2 := \langle R_Y \cdot, \cdot \rangle$ in the denominator of (4.1.5) by a spectrally equivalent norm, say $\|\cdot\|_{\mathcal{N}}^2 := \langle \mathcal{N} \cdot, \cdot \rangle \sim \|\cdot\|_Y^2$. The minimum obviously stays the same due the equivalence, but the discretization of the Riesz-map can be replaced by any other spectrally equivalent operator. Using, for instance, a wavelet discretization, one can replace $\mathbf{R}_Y$ by a diagonal scaling.

## 4.2 Stability

Remark that the best approximation $u_j^* \in S_j$ of the solution $u := B^{-1} f$ is given by

$$u_j^* = \arg\min_{v_j \in S_j} \|u - v_j\|_X = \arg\min_{v_j \in S_j} \|f - B v_j\|_{Y'},$$

since $B$ is boundedly invertible and linear. To this end, the best approximation is given by the normal equation

$$\left\langle R_Y^{-1}(f - B u_j^*), B v_j \right\rangle = 0, \quad \text{for all } v_j \in S_j,$$

with Riesz-map $R_Y$ defined via (4.1.12). Moreover, it is not hard to see that this is equivalent to the Petrov-Galerkin approach

$$\langle B u_j^*, w \rangle = \langle f, w \rangle, \quad \text{for all } w \in R_Y^{-1} B(S_j),$$

where $R_Y^{-1} B(S_j)$ is the *optimal test space* associated with $S_j$. The general idea now is to replace the optimal test space by a numerically feasible sufficiently large test space $Q_\ell$. This motivates the approach to enrich the test space $Q_\ell$ in order to stabilize the discrete problem. Different stabilization techniques with a similar starting point are considered in [DHSW12].

Instead of recursively stabilizing the bases, one constructs fixed standard bases such as spline functions or wavelets, where the regularity is also controlled by the number of extra layers in the test space. The test spaces constructed in this way yield *a-priori fixed* stable bases.

The aim of the construction in [Mol13b] is to construct suitable families of subspaces $\{S_j\}_{j \geq j_0}$ and $\{Q_\ell\}_{\ell \geq \ell_0}$ which uniformly satisfy the discrete inf-sup condition (4.1.6) a-priori with strict lower bound $0 < \beta \leq \beta_{j,\ell}$ independent of the discretization parameters $j$ and $\ell$. More precisely, we will give a general criterion for choosing suitable subspaces a-priori under classical Jackson and Bernstein conditions. This general result will be applied later in section 4.3 to the second space-time weak formulation of parabolic evolution problems from section 3.3 as an important model example. To this end, we

follow the lines of [Mol13b]. A brief overview can also be found in [Mol13c]. The basic idea how to enrich the test spaces was inspired by [DK01] dealing with the Ladyzenkaja-Bǎbuska-Brezzi condition (LBB-condition) of saddle point problems. The LBB-condition, as it appears in saddle point problems, is closely related to the inf-sup condition from the BNB-Theorem 2.3.3. As in [DK01] essential boundary conditions can be enforced with the aid of Lagrange multipliers leading to saddle point problems. For instance in [JK16] the authors presented a different approach to treat essential boundary conditions with Lagrange multipliers, but avoiding a saddle point formulation and therefore the LBB-condition.

We start with some assumptions on the operator $B$ from (4.1.2). Besides the boundedness

$$\|Bv\|_{Y'} \leq B_{\max}\|v\|_X, \quad v \in X \quad \text{and} \quad \|B^{-1}\tilde{q}\|_X \leq B_{\min}^{-1}\|\tilde{q}\|_{Y'}, \quad \tilde{q} \in Y', \tag{4.2.1}$$

with constants $0 < B_{\min} \leq B_{\max} < \infty$, we require a slightly shifted regularity

$$(B')^{-1} \in \mathcal{L}(X'_-, Y_+), \quad \text{with} \quad \|(B')^{-1}\|_{\mathcal{L}(X'_-,Y_+)} \leq C_+, \tag{4.2.2}$$

with a constant $0 < C_+ < \infty$ and continuously and densely embedded separable Hilbert subspaces $X'_- \hookrightarrow X'$ and $Y_+ \hookrightarrow Y$, where $B'$ denotes the dual operator of $B$, see Definition 2.3.1. Note that the notation is meaningful since $X \subset X_-$ implies $X'_- \subset X'$ and vice versa. It is different from the notation in [Mol13b] but in accordance with the one in, e.g., Corollary 3.3.10. It is well known, that $B \in \mathcal{L}(X, Y')$ implies $B' \in \mathcal{L}(Y, X')$ with

$$\|B'\|_{Y\to X'} = \|B\|_{X\to Y'} \quad \text{and} \quad \|(B')^{-1}\|_{X'\to Y} = \|B^{-1}\|_{Y'\to X},$$

see, e.g., [Aub00, Prop. 3.3.1]. That means, the regularity assumption (4.2.2) could equivalently be stated for the primal operator $B$ instead of $B'$, but since we need to consider the dual operator in order to prove Lemma 4.2.7 and Theorem 4.2.10, we directly formulated (4.2.2) in this form. So the two bounds $B_{\min}^{-1}$ and $C_+$ are related. This assumption is very similar to the shift theorem in [Ver95, (A1)]. The next assumptions are Jackson and Bernstein estimates as well as an often called reverse Cauchy-Schwarz inequality with respect to sequences of subspaces of $X$ and $Y$ representing the discrete solution and test spaces. Let $\{S_j\}_{j=j_0}^{\infty} \subset X$, $\{\tilde{S}_j\}_{j=j_0}^{\infty} \subset X'_-$ and $\{Q_\ell\}_{\ell=\ell_0}^{\infty} \subset Y_+$ such that for some $\nu > 1$ the Bernstein estimate

$$\|\tilde{v}_j\|_{X'_-} \leq C_{B,X'}\nu^j\|\tilde{v}_j\|_{X'}, \qquad \tilde{v}_j \in \tilde{S}_j \tag{4.2.3}$$

as well as the Jackson estimate

$$\inf_{q_\ell \in Q_\ell} \|q - q_\ell\|_Y \leq C_{J,Y}\nu^{-\ell}\|q\|_{Y_+}, \qquad q \in Y_+ \tag{4.2.4}$$

are satisfied with constants $C_{B,X'}, C_{J,Y} > 0$. Moreover, we assume that the *reverse Cauchy-Schwarz inequality* holds on $X$:
For every $v_j \in S_j$ there exists an element $\tilde{v}_j^* \in \tilde{S}_j$, depending on $v_j$, such that

$$\|v_j\|_X\|\tilde{v}_j^*\|_{X'} \leq C_{CS}\ {}_X\langle v_j, \tilde{v}_j^*\rangle_{X'}, \tag{4.2.5}$$

with a constant $C_{CS} > 0$. The reverse Cauchy-Schwarz inequality (4.2.5) can be formulated equivalently as

$$\inf_{v_j \in S_j \setminus \{0\}} \sup_{\tilde{v}_j \in \tilde{S}_j \setminus \{0\}} \frac{{}_X\langle v_j, \tilde{v}_j \rangle_{X'}}{\|v_j\|_X \|\tilde{v}_j\|_{X'}} \geq (C_{CS})^{-1} > 0. \tag{4.2.6}$$

That is, the reverse Cauchy-Schwarz inequality can be seen as a stability property of $S_j$ and $\tilde{S}_j$ with respect to the duality pairing ${}_X\langle \cdot, \cdot \rangle_{X'}$.

Of course, the previous assumptions are somewhat abstract and seem to be quite restrictive. But as we will see later, even for intersections of tensor product Hilbert spaces, as they appear in space-time weak formulations from chapter 3, we can easily state families of spaces which satisfy the assumptions. Moreover, assumption (4.2.2) on the operator is a rather standard regularity result for many operators, cf. also Corollary 3.3.10. Before we formulate the abstract main result, we need the following lemma.

**Lemma 4.2.7.** *Assume that* (4.2.2) *as well as* (4.2.3), (4.2.4) *and* (4.2.6) *are fulfilled. Then, for arbitrary $\tilde{v}_j \in \tilde{S}_j$ there exists an element $q_\ell^* \in Q_\ell$, depending on $\tilde{v}_j$, such that*

$$\|\tilde{v}_j - B' q_\ell^*\|_{X'} \leq C_{J,X'} \nu^{-(\ell - j)} \|\tilde{v}_j\|_{X'}, \tag{4.2.8}$$

*with a constant $C_{J,X'} := B_{\max}\, C_+\, C_{J,Y}\, C_{B,X'}$ and*

$$\|q_\ell^*\|_Y \leq B_{\min}^{-1} (C_{J,X'} \nu^{-(\ell-j)} + 1)\, \|\tilde{v}_j\|_{X'}. \tag{4.2.9}$$

*Proof.* Given $\tilde{v}_j \in \tilde{S}_j$ choose $q_\ell^* \in Q_\ell$ such that

$$\|(B')^{-1}\tilde{v}_j - q_\ell^*\|_Y = \min_{q_\ell \in Q_\ell \setminus \{0\}} \|(B')^{-1}\tilde{v}_j - q_\ell\|_Y.$$

With this choice it follows

$$\|(B')^{-1}\tilde{v}_j - q_\ell^*\|_Y \geq B_{\max}^{-1} \|\tilde{v}_j - B' q_\ell^*\|_{X'},$$

by using the boundedness (4.2.1). By the Jackson estimate (4.2.4), the regularity (4.2.2), and the Bernstein estimate (4.2.3) it follows

$$\begin{aligned} \|\tilde{v}_j - B' q_\ell^*\|_{X'} &\leq B_{\max} C_{J,Y} \nu^{-\ell} \|(B')^{-1}\tilde{v}_j\|_{Y_+} \leq B_{\max} C_+ C_{J,Y} \nu^{-\ell} \|\tilde{v}_j\|_{X'_-} \\ &\leq B_{\max} C_+ C_{J,Y} C_{B,X'} \nu^{-(\ell-j)} \|\tilde{v}_j\|_{X'}, \end{aligned}$$

proving (4.2.8). Similar, we can prove (4.2.9) by

$$\begin{aligned} \|q_\ell^*\|_Y &\leq \|q_\ell^* - (B')^{-1}\tilde{v}_j\|_Y + \|(B')^{-1}\tilde{v}_j\|_Y \leq B_{\min}^{-1} \|B' q_\ell^* - \tilde{v}_j\|_{X'} + B_{\min}^{-1} \|\tilde{v}_j\|_{X'} \\ &\leq B_{\min}^{-1} (C_{J,X'} \nu^{-(\ell-j)} + 1)\, \|\tilde{v}_j\|_{X'}. \end{aligned}$$

□

Finally, we can state our stability result.

**Theorem 4.2.10.** *Assume that* (4.2.2) *as well as* (4.2.3), (4.2.4) *and* (4.2.6) *are fulfilled. Choose* $L \in \mathbb{N}$ *such that*

$$C_{CS} C_{J,X'} \nu^{-L} < 1, \tag{4.2.11}$$

*with constants* $C_{J,X'}$ *and* $C_{CS}$ *as in Lemma 4.2.7 and* (4.2.5) *and set*

$$\ell \geq j + L, \tag{4.2.12}$$

*for any refinement level* $j$*. Then the discrete inf-sup condition*

$$\inf_{v_j \in S_j \setminus \{0\}} \sup_{q_\ell \in Q_\ell \setminus \{0\}} \frac{|\langle B v_j, q_\ell \rangle|}{\|v_j\|_X \|q_\ell\|_Y} \geq \beta > 0$$

*is satisfied with a constant* $\beta$ *uniformly bounded away from* $0$ *as* $j \to \infty$*. In particular,* $\beta$ *is given by*

$$\beta := \frac{C_{CS}^{-1} - C_{J,X'} \nu^{-L}}{B_{\min}^{-1}(C_{J,X'} \nu^{-L} + 1)}. \tag{4.2.13}$$

*Proof.* Let $v_j \in S_j$ be arbitrary, then by (4.2.5) there exists an element $\tilde{v}_j^* \in \tilde{S}_j$ such that

$$\begin{aligned} \|v_j\|_X \|\tilde{v}_j^*\|_{X'} &\leq C_{CSX} \langle v_j, \tilde{v}_j^* \rangle_{X'} \\ &= C_{CS} \left( {}_X\langle v_j, (\tilde{v}_j^* - B' q_\ell) \rangle_{X'} + {}_X\langle v_j, B' q_\ell \rangle_{X'} \right) \\ &\leq C_{CS} \left( \|v_j\|_X \|\tilde{v}_j^* - B' q_\ell\|_{X'} + {}_{Y'}\langle B v_j, q_\ell \rangle_Y \right), \end{aligned}$$

for arbitrary $q_\ell \in Q_\ell$. Next, we choose $q_\ell := q_\ell^* \in Q_\ell$ according to Lemma 4.2.7 such that

$$\|\tilde{v}_j^* - B' q_\ell^*\|_{X'} \leq C_{J,X'} \nu^{-(\ell - j)} \|\tilde{v}_j^*\|_{X'}.$$

Then it follows

$$\left( C_{CS}^{-1} - C_{J,X'} \nu^{-(\ell-j)} \right) \|\tilde{v}_j^*\|_{X'} \|v_j\|_X \leq {}_{Y'}\langle B v_j, q_\ell^* \rangle_Y.$$

Using (4.2.9) directly yields

$$\frac{C_{CS}^{-1} - C_{J,X'} \nu^{-(\ell-j)}}{B_{\min}^{-1}(C_{J,X'} \nu^{-(\ell-j)} + 1)} \|q_\ell^*\|_Y \|v_j\|_X \leq {}_{Y'}\langle B v_j, q_\ell^* \rangle_Y,$$

where

$$\frac{C_{CS}^{-1} - C_{J,X'} \nu^{-(\ell-j)}}{B_{\min}^{-1}(C_{J,X'} \nu^{-(\ell-j)} + 1)} \geq \frac{C_{CS}^{-1} - C_{J,X'} \nu^{-L}}{B_{\min}^{-1}(C_{J,X'} \nu^{-L} + 1)} =: \beta,$$

with the choice (4.2.12) of level $\ell$. Finally with the definition (4.2.11) of $L$ we obtain $\beta > 0$ what finishes the proof. □

This theorem shows how to choose level $\ell$ on the test space depending on the choice of level $j$ on the solution space in order to obtain a uniform discrete inf-sup condition. Obviously, (4.2.11) is satisfied when choosing $L \in \mathbb{N}$ with

$$L > \lceil \log_\nu(C_{CS} C_{J,X'}) \rceil \,. \tag{4.2.14}$$

Note that it is crucial to have explicit values of $L$ and $\beta$ given by (4.2.14) and (4.2.13) in order to ensure existence of moments of solutions of random PDEs in chapter 5.

## 4.3 Stability of Parabolic PDEs in Space-Time Weak Formulation

The next step is to apply the result of Theorem 4.2.10 to the space-time weak formulations introduced in chapter 3. To be more precise, we will verify the regularity assumption (4.2.2) and specify suitable families of discrete subspaces $\{S_j\}_{j=j_0}^\infty \subset X$, $\{\tilde{S}_j\}_{j=j_0}^\infty \subset X'_-$ and $\{Q_\ell\}_{\ell=\ell_0}^\infty \subset Y_+$, which satisfy (4.2.3), (4.2.4) and (4.2.6). Following the lines of [Mol13b], we will consider the second space-time weak form from section 3.3 with a spatial differential operator of order $2m$. That is, we consider

$$\begin{aligned} X &:= \tilde{\mathcal{X}} = L_2(I; H^m(D)), \\ Y &:= \tilde{\mathcal{Y}} = L_2(I; H^m(D)) \cap H^1_{0,\{T\}}(I; H^{-m}(D)), \end{aligned}$$

where $I := (0,T)$ with finite $0 < T < \infty$ as in chapter 3. The regularity (4.2.2) can be ensured for time-independent operators by Corollary 3.3.10, where the inf-sup condition (3.3.12) gives the required constant $C_+$. Recalling the considerations at the end of section 3.2, a reasonable choice would be

$$\begin{aligned} X'_- &:= \bar{\tilde{\mathcal{X}}}' = L_2(I; H^{-m+\alpha}(D)), \\ Y_+ &:= \bar{\tilde{\mathcal{Y}}} = L_2(I; H^{m+\alpha}(D)) \cap H^1_{0,\{T\}}(I; H^{-m+\alpha}(D)), \end{aligned}$$

with regularity shift $\alpha$ determined by the spatial regularity. In the particular case of shifting the spatial regularity in the test spaces by $\alpha := m$, i.e.,

$$\begin{aligned} X'_- &:= L_2(I; L_2(D)), \\ Y_+ &:= L_2(I; H^{2m}(D)) \cap H^1_{0,\{T\}}(I; L_2(D)), \end{aligned} \tag{4.3.1}$$

we have the regularity result from [CS11, Th. 2.4] which is known to hold also for time-dependent operators but without explicit bounds.

In view of (4.2.2) recall that $B^{-1} \in \mathcal{L}(Y'_+, X_-)$ is equivalent to $(B')^{-1} \in \mathcal{L}(X'_-, Y_+)$. Due to the definition of $X_-$, we have to construct dual bases with higher regularity since $X'_- \subset X'$. We will restrict ourselves first to spaces defined in (4.3.1), where the

regularity can also be proven for time-dependent spatial differential operators, but will give summarized results for more general spaces in Corollary 4.3.28.

In order to verify the assumptions (4.2.3), (4.2.4) and (4.2.6) on the discrete subspaces, we need to specify their choices. We choose $S_j^x \subset H^m(D)$, $S_j^t \subset L_2(I)$, $Q_\ell^x \subset H^{2m}(D)$ and $Q_\ell^t \subset H^1_{0,\{T\}}(I)$ and define

$$\begin{aligned} S_j &:= S_j^t \otimes S_j^x \subset L_2(I; H^m(D)) = X, \\ Q_\ell &:= Q_\ell^t \otimes Q_\ell^x \subset L_2(I; H^{2m}(D)) \cap H^1_{0,\{T\}}(I; L_2(D)) = Y_+. \end{aligned} \tag{4.3.2}$$

We arrange the spaces in such a way that they form sequences of nested subspaces

$$S_{j_0}^x \subset S_{j_0+1}^x \subset \cdots \subset S_j^x \subset S_{j+1}^x \subset \cdots \subset H^m(D), \tag{4.3.3}$$

with $\overline{\bigcup_{j=j_0}^\infty S_j^x}^{\|\cdot\|_{H^m(D)}} = H^m(D)$ and similar for $S_j^t, Q_\ell^x$ and $Q_\ell^t$.

Next we need to make some general assumptions on the discrete subspaces. We consider a possibly different sequence $\{\tilde{S}_j^x\}_{j=j_0}^\infty$ of nested and closed subspaces of $L_2(D)$ with $\dim \tilde{S}_j^x = \dim S_j^x$ such that there exists a constant $c_S^{(x)} > 0$ so that

$$\inf_{v \in S_j^x \setminus \{0\}} \sup_{\tilde{v} \in \tilde{S}_j^x \setminus \{0\}} \frac{|(v, \tilde{v})_{L_2(D)}|}{\|v\|_{L_2(D)} \|\tilde{v}\|_{L_2(D)}} \geq c_S^{(x)}, \qquad \text{for all } j \geq j_0. \tag{4.3.4}$$

This condition is often referred to as $L_2(D)$-*stability relation*. This assumption is obviously fulfilled with $c_S^{(x)} = 1$ when choosing $\tilde{S}_j^x = S_j^x$, but using (4.3.4) allows to use biorthogonal bases, for instance, cf. [DKU99]. That is, we have more flexibility in the choice of discretization spaces. We consider sequences $\{\tilde{S}_j^t\}_{j=j_0}^\infty$, $\{\tilde{Q}_\ell^t\}_{\ell=\ell_0}^\infty$ and $\{\tilde{Q}_\ell^x\}_{\ell=\ell_0}^\infty$ in a similar way. It is important to note that the dual subspaces $\tilde{S}_j^t$, $\tilde{S}_j^x$, $\tilde{Q}_\ell^t$ and $\tilde{Q}_\ell^x$ are only needed for the analysis of the discrete inf-sup condition, but do *not* enter the implementation concerning the discrete operator discretization. The next lemma shows that the $L_2(D)$-stability is inherited to the tensor product space in space and time.

**Lemma 4.3.5.** *The sequence* $\{\tilde{S}_j\}_{j=j_0}^\infty := \{\tilde{S}_j^t \otimes \tilde{S}_j^x\}_{j=j_0}^\infty$ *satisfies*

$$\inf_{v \in S_j \setminus \{0\}} \sup_{\tilde{v} \in \tilde{S}_j \setminus \{0\}} \frac{|(v, \tilde{v})_{L_2(I;L_2(D))}|}{\|v\|_{L_2(I;L_2(D))} \|\tilde{v}\|_{L_2(I;L_2(D))}} \geq c_S, \qquad \textit{for all } j \geq j_0, \tag{4.3.6}$$

*where* $c_S := c_S^{(t)} c_S^{(x)}$. *The same holds for the sequence* $\{\tilde{Q}_\ell\}_{\ell=\ell_0}^\infty := \{\tilde{Q}_\ell^t \otimes \tilde{Q}_\ell^x\}_{\ell=\ell_0}^\infty$ *with* $c_Q := c_Q^{(t)} c_Q^{(x)}$.

*Proof.* See, for example, [And13, Cor. 5.3]. □

An important consequence is the existence of uniformly bounded biorthogonal projectors according to the following proposition [DS99, Th. 2.1].

**Proposition 4.3.7.** *The stability* (4.3.4) *implies the existence of sequences* $\{P_{S_j^x}\}_{j=j_0}^\infty$ *of biorthogonal projectors* $P_{S_j^x}\colon L_2(D) \to S_j^x$ *such that for* $j \geq j_0$, $\operatorname{range}(I - P_{S_j^x}) = (\tilde{S}_j^x)^{\perp_{L_2(D)}}$, *while the adjoints* $P'_{S_j^x}\colon L_2(D) \to \tilde{S}_j^x$ *satisfy* $\operatorname{range}(I - P'_{S_j^x}) = (S_j^x)^{\perp_{L_2(D)}}$. *The projectors are uniformly bounded with respect to* $L_2(D)$ *with*

$$\|P_{S_j^x} v\|_{L_2(D)} \leq (c_S^{(x)})^{-1} \|v\|_{L_2(D)}, \qquad \text{for } j \geq j_0,\ v \in L_2(D). \tag{4.3.8}$$

*Furthermore, one has*

$$P_{S_j^x} P_{S_i^x} = P_{S_i^x}, \quad P'_{S_j^x} P'_{S_i^x} = P'_{S_i^x}, \quad i \leq j. \tag{4.3.9}$$

*The same holds for projectors* $P_{S_j^t}, P_{Q_\ell^t}, P_{Q_\ell^x}, P'_{S_j^t} P'_{Q_\ell^t}$ *and* $P'_{Q_\ell^x}$ *defined in a similar fashion and also for the tensor products due to Lemma 4.3.5.*

In particular in view of the Bernstein and Jackson estimates on the subspaces $\tilde{S}_j \subset X'_-$ and $Q_\ell \subset Y_+$, respectively, we have to arrange the spatial and temporal spaces $\{S_j^t\}_{j=j_0}^\infty, \{S_j^x\}_{j=j_0}^\infty, \{Q_\ell^t\}_{\ell=\ell_0}^\infty$ and $\{Q_\ell^x\}_{\ell=\ell_0}^\infty$ as well as its dual spaces such that they satisfy suitable approximation and smoothness properties. The aim is to give standard approximation and smoothness assumptions on the temporal and spatial subspaces separately, so that the tensor products satisfy the desired properties (4.2.3) and (4.2.4) with respect to the Hilbert spaces on the whole space-time cylinder. To this end, for fixed $\mu > 1$ consider the generic Jackson inequality with respect to a sequence of spaces $\{F_k\}_{k=k_0}^\infty$ on a domain $D' \subset \mathbb{R}^d$:

$$\inf_{f_k \in F_k} \|f - f_k\|_{L_2(D')} \lesssim \mu^{-sk} \|f\|_{H^s(D')}, \quad f \in H^s(D'), \quad 0 \leq s \leq d_F \tag{4.3.10}$$

and the generic Bernstein estimate with respect to a sequence of spaces $\{F_k\}_{k=k_0}^\infty$ on a domain $D' \subset \mathbb{R}^d$:

$$\|f_k\|_{H^s(D')} \lesssim \mu^{sk} \|f_k\|_{L_2(D')}, \quad f_k \in F_k, \quad 0 \leq s < \gamma_F. \tag{4.3.11}$$

The parameter $d_F$ characterizes the approximation order and $\gamma_F$ the smoothness of the space $F_k$. In view of the required Bernstein and Jackson estimate (4.2.3) and (4.2.4), we arrange the spaces $\{S_j^t\}_{j=j_0}^\infty, \{S_j^x\}_{j=j_0}^\infty, \{Q_\ell^t\}_{\ell=\ell_0}^\infty$ and $\{Q_\ell^x\}_{\ell=\ell_0}^\infty$ as well as its dual versions, such that they fulfill the Bernstein and Jackson estimates with parameters listed in Table 4.3.12. Although it is not trivial to prove that the choices according to Table 4.3.12 yield the desired properties (4.2.3) and (4.2.4) for the tensor product spaces, one can, heuristically, intuitively explain the required parameters. Namely, since the solution space $X := L_2(I; H^m(D))$ can be identified isometrically by the tensor product space $L_2(I) \otimes H^m(D)$, cf. Proposition 2.2.8, we require that $\gamma_{S^t}, d_{S^t} > 0$

Table 4.3.12: Approximation and smoothness parameters according to (4.3.10) and (4.3.11).

| space | primal | dual |
|---|---|---|
| $S_j^t$ | $\gamma_{S^t}, d_{S^t} > 0$ | $\gamma_{\tilde{S}^t}, d_{\tilde{S}^t} > 0$ |
| $S_j^x$ | $\gamma_{S^x}, d_{S^x} > m$ | $\gamma_{\tilde{S}^x}, d_{\tilde{S}^x} > 0$ |
| $Q_\ell^t$ | $\gamma_{Q^t}, d_{Q^t} > 1$ | $\gamma_{\tilde{Q}^t}, d_{\tilde{Q}^t} > 0$ |
| $Q_\ell^x$ | $\gamma_{Q^x}, d_{Q^x} > 2m$ | $\gamma_{\tilde{Q}^x}, d_{\tilde{Q}^x} > m$ |

due to $L_2(I)$ and that $\gamma_{S^x}, d_{S^x} > m$ due to $H^m(D)$. Similar, according to the test space $Y \cong (L_2(I) \otimes H^m(D)) \cap (H^1_{\{T\}}(I) \otimes H^{-m}(D))$, we choose $\gamma_{Q^t}, d_{Q^t} > 1$ and $\gamma_{Q^x}, d_{Q^x} > m$ as well as $\gamma_{\tilde{Q}^x}, d_{\tilde{Q}^x} > m$. Moreover, due to the choice of $Y_+ \cong (L_2(I) \otimes H^{2m}(D)) \cap (H^1_{\{T\}}(I) \otimes L_2(D))$, we have to enlarge $\gamma_{Q^x}, d_{Q^x} > 2m$.

**Remark 4.3.13.** *The approximation and smoothness properties given in Table 4.3.12 are known to hold for hierarchical spline spaces on uniform grids, see Theorem 2.4.10 and 2.4.12. In particular, it also holds for finite element as well as for spline-wavelet spaces, where we also have in mind space-time* sparse grid *spaces [GO07] instead of uniform full grid spaces. Using spline discretizations, the smoothness is determined by the global smoothness of the elements and the approximation property by the piecewise polynomial degree of the elements, see section 2.4. In a dyadic partitioning of the domain, the parameter generally can be chosen as $\mu = 2$. For more details on splines, finite elements and wavelets we refer to [dB01], [Bra07] and [Dah97], respectively.*

In order to give a mathematically rigorous proof of the three assumptions (4.2.3), (4.2.4) and (4.2.6), we will make extensive use of the following Theorem from [DS99, Th. 2.1] respectively [Dah96, Th. 3.2].

**Theorem 4.3.14.** *Assume stability property* (4.3.4) *for all involved spaces as well as the Jackson and Bernstein inequalities* (4.3.10) *and* (4.3.11) *with associated approximation and smoothness parameters from Table 4.3.12. Then, with $P_{S^x_{j_0-1}} := 0$, the following norm equivalence holds:*

$$\left( \sum_{j=j_0}^{\infty} \mu^{2sj} \| (P_{S_j^x} - P_{S_{j-1}^x}) v \|^2_{L_2(D)} \right)^{1/2} \sim \|v\|_{H^s(D)}, \qquad v \in H^s(D),$$

*for all $s \in (-\min\{\gamma_{\tilde{S}^x}, d_{\tilde{S}^x}\}, \min\{\gamma_{S^x}, d_{S^x}\})$. Analog equivalences hold for the sequence of spaces $\{S_j^t\}_{j=j_0}^\infty$, $\{Q_\ell^t\}_{\ell=\ell_0}^\infty$ and $\{Q_\ell^x\}_{\ell=\ell_0}^\infty$, as well as for $\{\tilde{S}_j^t\}_{j=j_0}^\infty$, $\{\tilde{S}_j^x\}_{j=j_0}^\infty$, $\{\tilde{Q}_\ell^t\}_{\ell=\ell_0}^\infty$ and $\{\tilde{Q}_\ell^x\}_{\ell=\ell_0}^\infty$ with interchanged roles of $(\gamma_{\tilde{S}^x}, d_{\tilde{S}^x})$ and $(\gamma_{S^x}, d_{S^x})$.*

Similar equivalences also hold for tensor products of temporal and spatial projectors as they will be used, e.g., to prove Proposition 4.3.17. First, we will verify the Bernstein estimate (4.2.3).

**Proposition 4.3.15.** *The Bernstein estimate*

$$\|\tilde{v}_j\|_{X'_-} \leq C_{B,X'}\mu^{jm}\|\tilde{v}_j\|_{X'}, \qquad \tilde{v}_j \in \tilde{S}_j \tag{4.3.16}$$

*holds for spaces $\tilde{S}_j$ constructed according to Table 4.3.12 and $X'_-$ defined in* (4.3.1).

*Proof.* For every $\tilde{v}_j \in \tilde{S}_j$ there holds

$$\begin{aligned}
\|\tilde{v}_j\|^2_{L_2(I;L_2(D))} &= \int_0^T \|\tilde{v}_j(t)\|^2_{L_2(D)}\,\mathrm{d}t = \int_0^T \sup_{\|u\|_{L_2(D)}=1} |\langle \tilde{v}_j(t), P_{S_j^x}u\rangle|^2\,\mathrm{d}t \\
&\leq \int_0^T \sup_{\|u\|_{L_2(D)}=1} \|\tilde{v}_j(t)\|^2_{H^{-m}(D)}\|P_{S_j^x}u\|^2_{H^m(D)}\,\mathrm{d}t \\
&\lesssim \int_0^T \sup_{\|u\|_{L_2(D)}=1} \|\tilde{v}_j(t)\|^2_{H^{-m}(D)}\mu^{2jm}\|P_{S_j^x}u\|^2_{L_2(D)}\,\mathrm{d}t \\
&\lesssim \mu^{2jm}\int_0^T \|\tilde{v}_j(t)\|_{H^{-m}(D)}\,\mathrm{d}t = \mu^{2jm}\|\tilde{v}_j\|^2_{L_2(I;H^{-m}(D))},
\end{aligned}$$

since $\gamma_{S^x} > m$ by the choice of parameters in Table 4.3.12, where we have used $L_2(D)$-stability (4.3.8). □

A more rigorous tracking of the involved constants would yield $C_{B,X'} \leq C^{(x)}_{B,S}\,(c^{(x)}_S)^{-1}$, where $c^{(x)}_S$ denotes the $L_2$-stability constant from (4.3.4) and $C^{(x)}_{B,S}$ the constant in the Bernstein estimate (4.3.11) for $S^x_j$ as well as $\nu = \mu^m$.

The next proposition verifies the Jackson estimate (4.2.4).

**Proposition 4.3.17.** *The Jackson estimate*

$$\inf_{q_\ell \in Q_\ell} \|q - q_\ell\|_Y \leq C_{J,Y}\mu^{-\ell m}\|q\|_{Y_+}, \qquad q \in Y_+ \tag{4.3.18}$$

*holds for spaces $Q_\ell$ constructed according to Table 4.3.12 and $Y_+$ defined in* (4.3.1).

*Proof.* Due to Theorem 4.3.14 with Bernstein and Jackson parameters from Table 4.3.12 and by using [GO95, Prop. 1 and Prop. 2] the splittings

$$\begin{aligned}
&\{Y;(\cdot,\cdot)_Y\} \\
=\ &\sum_{k=\ell_0}^{\infty}\sum_{i=\ell_0}^{\infty}\{\mathrm{range}(D_{Q,k,i});(\mu^{2mi}+\mu^{2k-2im})(\cdot,\cdot)_{L_2(I)}\otimes(\cdot,\cdot)_{L_2(D)}\},
\end{aligned}$$

and

$$\begin{aligned}&\{Y_+;(\cdot,\cdot)_{Y_+}\}\\ =\ &\sum_{k=\ell_0}^{\infty}\sum_{i=\ell_0}^{\infty}\{\mathrm{range}(D_{Q,k,i});(\mu^{2(2m)i}+\mu^{2k})(\cdot,\cdot)_{L_2(I)}\otimes(\cdot,\cdot)_{L_2(D)}\},\end{aligned}$$

are stable. Here we have used the abbreviation $D_{Q,k,i}:=(P_{Q_k^t}-P_{Q_{k-1}^t})\otimes(P_{Q_i^x}-P_{Q_{i-1}^x})$, with projectors defined in Proposition 4.3.7 and $P_{Q_{\ell_0-1}^t}=0$ as well as $P_{Q_{\ell_0-1}^x}=0$. That means, we have the norm equivalences

$$\left(\sum_{k=\ell_0}^{\infty}\sum_{i=\ell_0}^{\infty}(\mu^{2mi}+\mu^{2k-2im})\|D_{Q,k,i}q\|^2_{L_2(I)\otimes L_2(D)}\right)^{1/2}\sim\|q\|_Y,\qquad q\in Y,\tag{4.3.19}$$

as well as

$$\left(\sum_{k=\ell_0}^{\infty}\sum_{i=\ell_0}^{\infty}(\mu^{2(2m)i}+\mu^{2k})\|D_{Q,k,i}q\|^2_{L_2(I)\otimes L_2(D)}\right)^{1/2}\sim\|q\|_{Y_+},\qquad q\in Y_+.\tag{4.3.20}$$

For more details about tensor product subspace splittings we refer to [GO95]. Using these equivalences, we can conclude

$$\begin{aligned}\inf_{q_\ell\in Q_\ell}\|q-q_\ell\|_Y^2\ &\le\ \|q-\sum_{k=\ell_0}^{\ell}\sum_{i=\ell_0}^{\ell}D_{Q,k,i}q\|_Y^2=\|\sum_{k=\ell+1}^{\infty}\sum_{i=\ell+1}^{\infty}D_{Q,k,i}q\|_Y^2\\ &\lesssim\ \sum_{k=\ell+1}^{\infty}\sum_{i=\ell+1}^{\infty}(\mu^{2im}+\mu^{2k-2im})\|D_{Q,k,i}q\|^2_{L_2(I)\otimes L_2(D)}\\ &=\ \sum_{k=\ell+1}^{\infty}\sum_{i=\ell+1}^{\infty}\big[(\mu^{2(2m)i}+\mu^{2k})^{-1}(\mu^{2im}+\mu^{2k-2im})\\ &\qquad\times(\mu^{2(2m)i}+\mu^{2k})\|D_{Q,k,i}q\|^2_{L_2(I)\otimes L_2(D)}\big]\\ &\le\ \max_{i>\ell}\{\mu^{-2im}\}\sum_{k=\ell+1}^{\infty}\sum_{i=\ell+1}^{\infty}\big[(\mu^{2(2m)i}+\mu^{2k})\\ &\qquad\times\|D_{Q,k,i}q\|^2_{L_2(I)\otimes L_2(D)}\big]\\ &\lesssim\ \mu^{-2\ell m}\|q\|^2_{Y_+},\end{aligned}$$

for $q\in Y_+$. □

Again, we can explicitly bound the constant $C_{J,Y}$ by tracking the involved constants. One obtains $C_{J,Y}\le(c_{Y_+})^{-1}C_Y$, where $C_Y$ denotes the upper bound in (4.3.19) and $c_{Y_+}$ the lower bound in (4.3.20) as well as $\nu=\mu^m$.

Finally, we verify the reverse Cauchy-Schwarz inequality (4.2.5).

**Proposition 4.3.21.** *Assume that the spaces $S_j$ and $\tilde{S}_j$ are constructed according to Table 4.3.12. Then for every $v_j \in S_j$ there exists an element $\tilde{v}_j^* \in \tilde{S}_j$, depending on $v_j$, such that*

$$\|v_j\|_X \|\tilde{v}_j^*\|_{X'} \leq C_{CS}(v_j, \tilde{v}_j^*)_{L_2(I;H)}, \tag{4.3.22}$$

*with a constant $0 < C_{CS} < \infty$.*

*Proof.* Let $k \leq i \leq j$ and $v \in S_j$, then $D_{S,k,i}v \in S_i$, by the nestedness (4.3.3) with $D_{S,k,i} := (P_{S_k^t} - P_{S_{k-1}^t}) \otimes (P_{S_i^x} - P_{S_{i-1}^x})$, with $P_{S_{j_0-1}^t} = 0$ as well as $P_{S_{j_0-1}^x} = 0$, and due to the stability (4.3.6) there exists an element $\tilde{v}_{k,i} \in \tilde{S}_i \setminus \{0\}$ such that

$$(D_{S,k,i}v, \tilde{v}_{k,i})_{L_2(I)\otimes L_2(D)} \geq c_S \, \|D_{S,k,i}v\|_{L_2(I)\otimes L_2(D)} \|\tilde{v}_{k,i}\|_{L_2(I)\otimes L_2(D)}.$$

Defining $\tilde{v}_{k,i}^* := \mu^{2mi} \frac{\|D_{S,k,i}v\|_{L_2(I)\otimes L_2(D)}}{\|\tilde{v}_{k,i}\|_{L_2(I)\otimes L_2(D)}} \, \tilde{v}_{k,i}$ yields

$$(D_{S,k,i}v, \tilde{v}_{k,i}^*)_{L_2(I)\otimes L_2(D)} \geq c_S \, \mu^{2mi} \|D_{S,k,i}v\|^2_{L_2(I)\otimes L_2(D)}.$$

Setting $\tilde{v}^* := \sum_{k=j_0}^{j} \sum_{i=j_0}^{j} D'_{S,k,i} \tilde{v}_{k,i}^* \in \tilde{S}_j$ yields

$$\begin{aligned} (v, \tilde{v}^*)_{L_2(I)\otimes L_2(D)} &= \sum_{k=j_0}^{j} \sum_{i=j_0}^{j} \left(D_{S,k,i}v, \tilde{v}_{k,i}^*\right)_{L_2(I)\otimes L_2(D)} \\ &\geq c_S \, \sum_{k=j_0}^{j} \sum_{i=j_0}^{j} \mu^{2mi} \|D_{S,k,i}v\|^2_{L_2(I)\otimes L_2(D)}. \end{aligned} \tag{4.3.23}$$

Due to the construction of $S_j$ and $\tilde{S}_j$, we have the norm equivalence

$$\begin{aligned} &c_X \left( \sum_{k=j_0}^{\infty} \sum_{i=j_0}^{\infty} \mu^{2mi} \|D_{S,k,i}v_j\|^2_{L_2(I)\otimes L_2(D)} \right)^{1/2} \\ \leq &\|v_j\|_X \leq C_X \left( \sum_{k=j_0}^{\infty} \sum_{i=j_0}^{\infty} \mu^{2mi} \|D_{S,k,i}v_j\|^2_{L_2(I)\otimes L_2(D)} \right)^{1/2}, \end{aligned} \tag{4.3.24}$$

for any $v_j \in S_j$ and constants $C_X \geq c_X > 0$. Furthermore for the dual norm we obtain

$$\|v_j\|_{X'} \leq c_X^{-1} \left( \sum_{k=j_0}^{\infty} \sum_{i=j_0}^{\infty} \mu^{-2mi} \|D'_{S,k,i}v_j\|^2_{L_2(I)\otimes L_2(D)} \right)^{1/2}. \tag{4.3.25}$$

Using norm equivalence (4.3.24) as well as (4.3.25), we obtain

$$\|v\|_X \|\tilde{v}^*\|_{X'} \leq 4c_S^{-1}\kappa \sum_{k=j_0}^{j}\sum_{i=j_0}^{j} \mu^{2mi}\|D_{S,k,i}v\|^2_{L_2(I)\otimes L_2(D)}, \tag{4.3.26}$$

with $\kappa := \frac{C_x}{c_X}$, where we have used that

$$\begin{aligned} & \|D'_{S,k,i}\tilde{v}^*\|_{L_2(I)\otimes L_2(D)} \\ = \;& \mu^{2mi}\frac{\|D_{S,k,i}v\|_{L_2(I)\otimes L_2(D)}}{\|\tilde{v}_{k,i}\|_{L_2(I)\otimes L_2(D)}}\|D'_{S,k,i}\tilde{v}_{k,i}\|_{L_2(I)\otimes L_2(D)} \\ \leq \;& \mu^{2mi}\|D_{S,k,i}v\|_{L_2(I)\otimes L_2(D)}\; 4c_S^{-1}, \end{aligned}$$

since $D'_{S,k,i}$ is stable according to (4.3.8) with

$$\begin{aligned} & \|(P'_{S_k^t} - P'_{S_{k-1}^t}) \otimes (P'_{S_i^x} - P'_{S_{i-1}^x})\| \\ \leq \;& \left(\|P'_{S_k^t}\| + \|P'_{S_{k-1}^t}\|\right)\left(\|P'_{S_i^x}\| + \|P'_{S_{i-1}^x}\|\right) \\ \leq \;& 4c_S^{-1}. \end{aligned}$$

Combining (4.3.23) and (4.3.26) finishes the proof. □

That means, the reverse Cauchy-Schwarz inequality (4.2.6) holds with $C_{CS} \leq 4c_S^{-2}c_X^{-1}C_X$, with $L_2$-stability constant $c_S = c_S^{(t)}c_S^{(x)}$ from (4.3.6), and the lower and upper bound $c_X$ and $C_X$, respectively, in (4.3.24).

Finally, the previous Propositions 4.3.15, 4.3.17 and 4.3.21 together with the regularity result of Corollary 3.3.10, respectively [CS11, Theorem 2.4] without explicit bounds, but also for time dependent spatial differential operators, prove that the assumptions (4.2.2) as well as (4.2.3), (4.2.4) and (4.2.6) are satisfied for an operator $B$ defined via (3.3.5) and for the families of spaces introduced here. That means that Theorem 4.2.10 can be applied to this situation. So the discrete inf-sup condition holds with a constant which does not dependent on the discretization when choosing the levels $j$ and $\ell$ according to Theorem 4.2.10. An overview of all relevant constants is given in Table 4.3.27.

To some extent, the following corollary generalizes our previous results. The choices of parameters we made (Table 4.3.12) as well as the corresponding results are only restricted due to the choices of subspaces $X'_+ \subset X'$ and $Y_+ \subset Y$ according to (4.3.1). As already mentioned before, one can derive similar results in an analog fashion for more general spaces by Corollary 3.3.10.

**Corollary 4.3.28.** *Assuming generally that the regularity* (4.2.2) *holds for subspaces*

$$\begin{aligned} X'_- &:= H^{d_t}(I; H^{d_x-m}(D)), \\ Y_+ &:= H^{d_t}(I; H^{d_x+m}(D)) \cap H^{d_t+1}_{0,\{T\}}(I; H^{d_x-m}(D)), \end{aligned}$$

Table 4.3.27: Relevant constants

| constant | value | description and reference |
|---|---|---|
| $L$ | $\left\lceil \frac{\log_\mu(C_{CS}C_{J,X'})}{m} \right\rceil$ | number of extra layers, (4.2.14) |
| $\beta$ | $\frac{C_{CS}^{-1} - C_{J,X'}\mu^{-Lm}}{B_{\min}^{-1}(C_{J,X'}\mu^{-Lm}+1)}$ | stability constant, (4.2.13) |
| $C_{J,X'}$ | $B_{\max}C_+C_{J,Y}C_{B,X'}$ | auxiliary constant from Lemma 4.2.7, (4.2.8) |
| $C_{CS}$ | $4c_s^{-2}c_X^{-1}C_X$ | constant in reverse Cauchy-Schwarz inequality, (4.3.22) |
| $C_{B,X'}$ | $C_{B,S}^{(x)}(c_S^{(x)})^{-1}$ | constant in Bernstein estimate, (4.3.16) |
| $C_{J,Y}$ | $(c_{Y_+})^{-1}C_Y$ | constant in Jackson estimate, (4.3.18) |
| $B_{\min}$ | see (3.3.8) | inf-sup constant, (4.2.1) |
| $B_{\max}$ | cf. (3.3.7) | continuity constant, (4.2.1) |
| $C_+$ | cf. (3.3.12) | constant for shifted regularity, (4.2.2) |
| $c_s$ | $c_S^{(t)}c_S^{(x)}$ | $L_2$-stability constant, (4.3.6) |
| $c_X$, $C_X$ | - | lower and upper bound in norm equivalence, (4.3.24) |
| $C_{B,S}^{(x)}$ | - | constant in Bernstein estimate for $S_j^x$, (4.3.11) |
| $c_S^{(t)}$, $c_S^{(x)}$ | - | $L_2$-stability constant, (4.3.4) |
| $c_{Y_+}$ | - | lower bound in norm equivalence, (4.3.20) |
| $C_Y$ | - | upper bound in norm equivalence, (4.3.19) |

*with $d_t \geq 0$, $d_x > 0$, then Theorem 4.2.10 holds with $L \in \mathbb{N}$ such that*

$$L > \frac{\log_\mu(C_{CS}C_{J,X'})}{d_x + d_t},$$

*when choosing the parameters according to Table 4.3.29.*

The spaces in Corollary 3.3.10 would imply $d_t := 0$ and $d_x := \alpha$.

We conclude this chapter with a brief outlook. The recipe for obtaining stable subspaces for the first space-time formulation may be developed along the same lines. Furthermore, the spaces $S_j := S_j^t \otimes S_j^x$ and $Q_\ell := Q_\ell^t \otimes Q_\ell^x$, respectively, are constructed such that the spatial and temporal resolutions are equal. Obviously, one could also choose

Table 4.3.29: Approximation and smoothness parameters (general setting).

| space | primal | dual |
|---|---|---|
| $S_j^t$ | $\gamma_{S^t}, d_{S^t} > 0$ | $\gamma_{\tilde{S}^t}, d_{\tilde{S}^t} > d_t$ |
| $S_j^x$ | $\gamma_{S^x}, d_{S^x} > m,$ | $\gamma_{\tilde{S}^x}, d_{\tilde{S}^x} > \max\{0, d_x - m\}$ |
| $Q_\ell^t$ | $\gamma_{Q^t}, d_{Q^t} > d_t + 1$ | $\gamma_{\tilde{Q}^t}, d_{\tilde{Q}^t} > 0$ |
| $Q_\ell^x$ | $\gamma_{Q^x}, d_{Q^x} > d_x + m$ | $\gamma_{\tilde{Q}^x}, d_{\tilde{Q}^x} > m$ |

different resolutions and replace $S_j$ by $S_{(j_1,j_2)} := S_{j_1}^t \otimes S_{j_2}^x$ and $Q_\ell$ by $Q_{(\ell_1,\ell_2)} := Q_{\ell_1}^t \otimes Q_{\ell_2}^x$, respectively. It is most likely that considering such more general spaces would yield a sharper estimate (4.2.11) for the number of extra layers $L$ and, therefore, would improve the result given by Theorem 4.2.10 for such particular situations. Moreover, one obtains a sharper estimate for $L$ the larger we can choose parameter $\nu$ in (4.2.11). Assuming higher shift of regularity of the operator $B$ allows larger values for $\nu$ if we choose discrete subspaces with corresponding stronger approximation and smoothness properties.

Related results concerning the stability of Petrov-Galerkin discretizations for parabolic evolution equations can be found in [And13, And12]. Moreover, in this regard, we would also like to mention [GO07, MB97, UP12, UP14, DKO12] exploiting the fact that time-stepping methods such as, for example, Crank-Nicolson and discontinuous Galerkin methods, can be interpreted as space-time discretizations. We refer to [Mol13b] for a careful comparison of the quoted works. Stability results for the first formulation described in section 3.1 are given in [And13]. In particular, it was proven that only one extra layer in the temporal test space suffices to ensure stability for piecewise linear and continuous splines, see [And13, Prop. 6.1 and 6.3].

We apply the previous stability results pathwisely to parabolic problems with random coefficients in chapter 5, in particular section 5.3. It will turn out that the number of extra layers $L$ will depend on the stochastic parameter. This is not very surprising, since our previous results are very general and depend highly on the operator under consideration. To this end, a more convenient choice is to restrict ourselves to the first formulation and modify the results from Andreev [And13] to the particular situation with random coefficients. Moreover, we can combine ideas on subspace dependent norms from [UP14] and [And12] to obtain satisfactory stability results also for parabolic PDEs with random coefficients. A detailed description will be given in section 5.3.

# 5 Random PDEs in Space-Time Weak Formulation

Chapter 3 was devoted to deterministic generic parabolic PDEs of the form

$$\frac{\mathrm{d}u(t)}{\mathrm{d}t} + A(t)u(t) = g(t), \qquad u(0) = u_0, \qquad t \in I \text{ a.e.},$$

with (possibly time dependent) spatial differential operators $A(t)$ and time interval $I := (0, T)$. These problems are extended in the way that the right hand side, the initial condition, and the spatial differential operator are allowed to depend on an additional *random parameter* $\omega$. That is, a generic parabolic evolution problem with random coefficients

$$\frac{\mathrm{d}U(t,\omega)}{\mathrm{d}t} + A(t,\omega)U(t,\omega) = g(t,\omega), \qquad U(0,\omega) = U_0(\omega), \qquad t \in I \text{ a.e.}, \quad \mathbb{P}\text{-a.s.}$$

will be considered. The notation $\mathbb{P}$-a.s. is an abbreviation of almost surely with respect to the measure $\mathbb{P}$, where "almost surely" means "almost everywhere" and is used when dealing with probability measures. That is, an event occurs $\mathbb{P}$-almost surely, if it happens with probability one with respect to the probability measure $\mathbb{P}$. Therefore, $\mathbb{P}$-a.s. is the same as $\omega \in \Omega$ a.s. (respectively a.e.), where we use both denotations in the following. The equation is defined on a *probability space* to treat the random parameter properly. The solution of stochastic partial differential equations are commonly denoted by capital letters, and we do the same here for the random PDEs in order to distinguish between deterministic and random PDEs. The uncertainty enters the PDE in form of the random parameter $\omega$. The initial value $U_0$ as well as the right hand side $g$ are not deterministic functions, but *random fields* or random functions. The solution $U$ is also a random field. The $L_p$-regularity of the solution with respect to the random parameter $\omega$ classifies the existence of *moments* and vice versa. A random PDE can be seen as a parameter dependent PDE, where the parameter itself is an element of a sample space $\Omega$. The present random PDE differs from a "true" stochastic partial differential equation (SDE) by the way how the stochastic influence enters the equation. An SDE is of the form

$$\begin{aligned} \mathrm{d}U(t,\omega) + A(t,\omega)U(t,\omega)\mathrm{d}t + C(t,\omega)\mathrm{d}W(t,\omega) &= g(t,\omega)\mathrm{d}t, \\ U(0,\omega) &= U_0(\omega), \end{aligned}$$

with Wiener process or Brownian motion $W(t,\omega)$ as an additional (additive) noise term. A space-time weak formulation of stochastic PDEs was introduced in [LM16b], where it was shown that the stochastic counterpart of the $\omega$-wise second space-time weak formulation from section 3.3 yields existence and uniqueness of a solution.

The precise stochastic details would go much beyond the scope of this thesis. So we focus on the required assumptions only and try to keep the probability theory on a moderate level as far as possible. Since we are dealing exclusively with random PDEs

and not with SDEs, no Itô calculus is required and the parameter $\omega$ could also be interpreted as an infinitely dimensional deterministic parameter via a Karhunen-Loève expansion. Moreover, since we are interested in the number of existing moments of the solution, it is basically sufficient to work with Lebesgue spaces $L_p(\Omega)$ on a sample space $\Omega$ and refrain from very deep pure stochastic treatments. We refer to, e.g., [Bau95] for a rigorous stochastic introduction and to [AT07, GS91] for details in particular on random PDEs, random fields, Karhunen-Loève expansion, etc.

A random PDE is, due to the lack of noise, a simplification of a stochastic PDE, so that existence and uniqueness is obtained in the same way. Nevertheless, in order to prove existence and uniqueness of a solution of a SDE, one needs to assume that the spatial differential operator $A(t,\omega)$ is bounded from above and below *uniformly* in $\omega$. In our approach, we allow the bounds of $A(t,\omega)$ to be random variables itself $A_{\min} = A_{\min}(\omega)$ and $A_{\max} = A_{\max}(\omega)$, where $A_{\max}$ and $A_{\min}$ denote an upper and lower bound for $A$, respectively, see chapter 3. This is, of course, an essential improvement, since it will be sufficient that the bounds, respectively its reciprocal, are only in $L_p(\Omega)$ for certain $p \in \mathbb{R}^+$ with $p \geq 1$ and herewith allow $A(t,\omega)$ to tend to infinity and zero. This means, in particular, one may allow $A_{\max}, \frac{1}{A_{\min}} \notin L_\infty(\Omega)$. Therefore, one can treat relevant cases not covered so far by preceding papers on the same topic, cf. [GAS14] for instance. A similar idea for elliptic random PDEs was already considered in [Tec13, Cha12].

The basic ideas in this chapter are based on our novel work [LMM16]. Section 5.1 deals with the existence of moments of the solution of the continuous, non-discrete problem in space-time formulation from chapter 3. The focus is placed on the second formulation with the sharpest estimates, but also results concerning the other formulations are summarized in Table 5.1.22 and 5.1.23. Section 5.1 proceeds along the lines of [LMM16, ch. 3], where we detail the explanations and additionally give precise results for different kinds of coercive spatial differential operators as well as operators which satisfy a Gårding-inequality only. After having considered the existence of moments for the continuous problems, we are facing the semidiscrete and fully-discrete case. First, we present the results from [LMM16, sec 4.1] for the semi-discrete case in section 5.2. Then we focus on the quasi-optimality in $L_p(\Omega;\mathcal{X})$ in section 5.3. We go a different way for the fully discrete case, as done in [LMM16]. It was already observed in chapter 4 that the deterministic Petrov-Galerkin approach is unstable in general, but can be stabilized by enriching the dimension of the test space and considering a minimal residual Petrov-Galerkin approach instead. This applies to each trajectory of the random counterpart almost surely. In contrary to [LMM16], we exploit this behavior in this thesis and do not need to assume a CFL-condition.

## 5.1 Existence of $p$-Moments

We recall the formulation of parabolic PDEs from chapter 3, but equip it with an additional stochastic parameter $\omega$. Given two separable Hilbert spaces $V \hookrightarrow H$ with continuous and dense embedding, arranged in a Gelfand triple $V \subset H \cong H' \subset V'$, we consider a linear parabolic random PDE

$$\begin{aligned} \frac{\mathrm{d}U(t,\omega)}{\mathrm{d}t} + A(t,\omega)U(t,\omega) &= g(t,\omega) \qquad t \in I \text{ a.e., } \mathbb{P}\text{-a.s.} \\ U(0,\omega) &= U_0(\omega), \qquad\qquad \mathbb{P}\text{-a.s.,} \end{aligned} \tag{5.1.1}$$

specified as follows. Let $I := (0,T) \subset \mathbb{R}$ be a finite interval with $0 < T < \infty$ and $(\Omega, \Sigma, \mathbb{P})$ a complete probability space with normal filtration $\Sigma = (\Sigma_t)_{t\in I}$. We consider a progressively measurable random function with Bochner integrable trajectories $g(\cdot,\omega) \in L_2(I;V')$ almost surely and a progressively measurable initial data $U_0(\omega) \in H$ almost surely. Moreover, we assume that $A(t,\omega)\colon I \times \Omega \to \mathcal{L}(V,V')$ is progressively measurable, coercive and bounded uniformly in $I$, but *not* necessarily in $\Omega$. That means

$$\begin{aligned} |\langle A(t,\omega)u, v\rangle| &\le A_{\min}(\omega)\|u\|_V\|v\|_V \qquad \text{for all } u,v \in V,\ t \in I \text{ a.e. (boundedness)}, \\ \langle A(t,\omega)u, u\rangle &\ge A_{\max}(\omega)\|u\|_V^2 \qquad\quad \text{for all } u \in V,\ t \in I \text{ a.e. (coercivity)}, \end{aligned}$$

for $\omega \in \Omega$ a.s., such that the lower and upper bounds $A_{\min}$ and $A_{\max}$ are $\mathbb{P}$-a.s. positive and $\mathbb{P}$-a.s. finite *random variables*, respectively. These are exactly the same assumptions as in section 3, except that we have introduced an additional random parameter $\omega \in \Omega$ and allow our bounds of $A$ to be random variables too. Random fields can be seen as function valued random variables, in this way generalizing the concept of stochastic processes. More precisely, $U_0$ is a $H$-valued random variable and $U$ and $g$ are $\mathcal{X}$ and $\mathcal{Y}'$-valued random variables, respectively. A random function $F$ is a collection of random variables $F(x)$ with values $x$ from the function domain. Stochastic processes for instances are random functions in time. In this context, $U_0$ is a random function on the spatial variable $x \in D$, where $U$ and $g$ are random functions on $(t,x) \in I \times D$, with $D$ denoting the spatial domain. That means, we can see $U_0$, $g$ and $U$ as random fields and also as random functions. One should keep that in mind, since there are both concepts used in the literature. As in chapter 3, we assume that $A(t)$ is bounded and coercive for the whole chapter, but also treat the non-coercive case explicitly at the end.

Considering the random PDE (5.1.1) $\omega$-wise, we arrive at the second weak formulation

$$\tilde{b}_\omega(U(\cdot,\omega), v) = \tilde{\mathcal{F}}_\omega(v) \qquad \text{for all } v \in \tilde{\mathcal{Y}},\ \mathbb{P}\text{-a.s.} \tag{5.1.2}$$

where the (parameter dependent) bilinear form $\tilde{b}_\omega(\cdot,\cdot)\colon \tilde{\mathcal{X}} \times \tilde{\mathcal{Y}} \to \mathbb{R}$ is defined as

$$\tilde{b}_\omega(u,v) := \int_I \left(-{}_V\langle u(t), \dot{v}(t)\rangle_{V'} + {}_{V'}\langle A(t,\omega)u(t), v(t)\rangle_V\right) \mathrm{d}t, \quad \mathbb{P}\text{-a.s.} \tag{5.1.3}$$

and the (parameter dependent) right hand side $\tilde{\mathcal{F}}_\omega(\cdot)\colon \tilde{\mathcal{Y}} \to \mathbb{R}$ is given by

$$\tilde{\mathcal{F}}_\omega(v) := \int_I \langle g(t,\omega), v(t)\rangle \,\mathrm{d}t + (U_0(\omega), v(0))_H, \quad \mathbb{P}\text{-a.s.}, \tag{5.1.4}$$

in the same way as done in section 3.3. The solution and test spaces for fixed trajectories are given as

$$\tilde{\mathcal{X}} := L_2(I;V), \qquad \tilde{\mathcal{Y}} := L_2(I;V) \cap H^1_{0,\{T\}}(I;V').$$

We conclude from section 3.3, that the operator $\tilde{B}_\omega\colon \tilde{\mathcal{X}} \to \tilde{\mathcal{Y}}'$ defined by

$${}_{\tilde{\mathcal{Y}}'}\langle \tilde{B}_\omega u, v\rangle_{\tilde{\mathcal{Y}}} := \tilde{b}_\omega(u,v) \qquad u \in \tilde{\mathcal{X}},\ v \in \tilde{\mathcal{Y}}$$

is boundedly invertible $\mathbb{P}$-a.s. since $A_{\min}$, $A_{\max}$ are $\mathbb{P}$-a.s. positive and finite. We can deduce weak space-time formulations in the first form, as well as the homogenization of the first form, in the same way. We restrict ourselves here illustratively to the second formulation, since we can derive the sharpest results and have naturally incorporated initial conditions. The coercivity constant $A_{\min}$ and the continuity constant $A_{\max}$ are very important in the parameter dependent case. Indeed, since we do not assume uniform boundedness in $\Omega$, but only finiteness almost surely and consider the bounds as random variables, the existence of $p$-moments for the solution depends on the existence of possibly higher moments for $A_{\min}$ and $A_{\max}$, aside from similar requests on $U_0$ and $g$. Therefore, having the sharpest possible bounds is crucial for the existence of higher moments of the solution. To this end, we recall here the sharpest bounds derived in section 3.3. In the general setting, we have proven for $\omega \in \Omega$ a.s.

$$\sup_{v\in\tilde{\mathcal{X}}\setminus\{0\}} \sup_{w\in\tilde{\mathcal{Y}}\setminus\{0\}} \frac{|\langle \tilde{B}_\omega v, w\rangle|}{\|v\|_{\tilde{\mathcal{X}}}\|w\|_{\tilde{\mathcal{Y}}}} \le \sqrt{2}\max\{1, A_{\max}(\omega)\} \tag{5.1.5}$$

as well as

$$\inf_{v\in\tilde{\mathcal{X}}\setminus\{0\}} \sup_{w\in\tilde{\mathcal{Y}}\setminus\{0\}} \frac{\langle \tilde{B}_\omega v, w\rangle}{\|v\|_{\tilde{\mathcal{X}}}\|w\|_{\tilde{\mathcal{Y}}}} \ge \frac{\min\{A_{\min}(\omega), A_{\min}(\omega)A^{-1}_{\max}(\omega)\}}{\sqrt{2}}. \tag{5.1.6}$$

In Proposition 3.3.10 an improved inf-sup estimate was proven for time-independent spatial differential operators, such that

$$\inf_{v\in\tilde{\mathcal{X}}\setminus\{0\}} \sup_{w\in\tilde{\mathcal{Y}}\setminus\{0\}} \frac{\langle \tilde{B}_\omega v, w\rangle}{\|v\|_{\tilde{\mathcal{X}}}\|w\|_{\tilde{\mathcal{Y}}}} \ge \frac{\min\{1, A_{\min}(\omega)\}}{\sqrt{2}}, \quad \text{for } \omega \in \Omega \text{ a.s.}$$

without the factor $A^{-1}_{\max}(\omega)$, where we do not consider the generalized test and solution spaces here for simplicity. With $g(\cdot,\omega) \in L_2(I;V')$ and $U_0(\omega) \in H$ $\mathbb{P}$-a.s., we obtain

$\tilde{\mathcal{F}}_\omega \in \tilde{\mathcal{Y}}'$ $\mathbb{P}$-a.s., since

$$\begin{aligned}
|\tilde{\mathcal{F}}_\omega(v)| &= |\int_I \langle g(t,\omega), v(t)\rangle \, \mathrm{d}t + (U_0(\omega), v(0))_H| \\
&\leq |\int_I \langle g(t,\omega), v(t)\rangle \, \mathrm{d}t| + \|U_0(\omega)\|_H \|v(0)\|_H \\
&\leq \|g(\cdot,\omega)\|_{L_2(I;V')} \|v\|_{L_2(I;V)} + \|U_0(\omega)\|_H \|v\|_{\tilde{\mathcal{Y}}} \\
&\leq \left(\|g(\cdot,\omega)\|_{\tilde{\mathcal{X}}'} + \|U_0(\omega)\|_H\right) \|v\|_{\tilde{\mathcal{Y}}}, \qquad (5.1.7)
\end{aligned}$$

since $\tilde{\mathcal{Y}} \hookrightarrow \mathcal{C}^0(\bar{I}; H)$ and therefore $\|\tilde{\mathcal{F}}_\omega\|_{\tilde{\mathcal{Y}}'} \leq \|g(\cdot,\omega)\|_{\tilde{\mathcal{X}}'} + \|U_0(\omega)\|_H$ $\mathbb{P}$-a.s. We can easily conclude

$$\begin{aligned}
\|U(\cdot,\omega)\|_{\tilde{\mathcal{X}}} &\leq \frac{\sqrt{2}}{\min\{A_{\min}(\omega), A_{\min}(\omega)A_{\max}^{-1}(\omega)\}} \left(\|g(\cdot,\omega)\|_{\tilde{\mathcal{X}}'} + \|U_0(\omega))\|_H\right) \\
&= \sqrt{2}\max\{A_{\min}^{-1}(\omega), A_{\max}(\omega)A_{\min}^{-1}(\omega)\}\left(\|g(\cdot,\omega)\|_{\tilde{\mathcal{X}}'} + \|U_0(\omega))\|_H\right), \qquad (5.1.8)
\end{aligned}$$

$\mathbb{P}$-a.s., for the solution $U$ of the random PDE (5.1.1) in second space-time weak formulation. Assuming uniform bounds $A_{\min}$ and $A_{\max}$ as well as $g \in L_\infty(\Omega; \tilde{\mathcal{Y}}')$ and $U_0 \in L_\infty(\Omega; H)$, the right hand side of (5.1.8) would be independent of $\omega \in \Omega$, or at least there would exist an upper bound independent of $\omega \in \Omega$, so that nothing is left to show for the existence of moments of the solution. To this end, we are mainly interested in the "non-trivial" case of non-uniform bounds. Notice that we arrange the functions, or random field to be precise, such that, exemplarily, $g(\omega) \in \tilde{\mathcal{Y}}'$ for $\omega \in \Omega$ a.s., whereas $g(\omega)(t) := g(t,\omega)$ for $\omega \in \Omega$ a.s. and $t \in I$ a.e., cf. discussion after Proposition (2.2.8).

Provided the almost sure existence of a solution to (5.1.2), we want to give some sufficient conditions on the existence of moments of the data and of the two random variables $A_{\max}$ and $A_{\min}$, bounding the operator $A$, such that $p$-moments of the solutions exist, for some $p \in [1, \infty]$.

**Theorem 5.1.9.** *Assume that there exist parameters $\alpha, \beta, \gamma \in [1, \infty]$ with*

$$\alpha\beta\gamma \geq \alpha\beta + \alpha\gamma + \beta\gamma. \qquad (5.1.10)$$

*Let the data $g$ and $U_0$, and the random variables $A_{\min}$ and $A_{\max}$ are such that:*

*(i)* $\tilde{\mathcal{F}}_\omega$ *belongs to* $L_\alpha(\Omega; \tilde{\mathcal{Y}}')$, *that is,* $g \in L_\alpha(\Omega; \tilde{\mathcal{X}}')$ *and* $U_0 \in L_\alpha(\Omega; H)$,

*(ii)* $A_{\max} \in L_\beta(\Omega)$,

*(iii)* $\frac{1}{A_{\min}} \in L_\gamma(\Omega)$.

*Then the solution $U$ to problem (5.1.2) belongs to $L_p(\Omega;\tilde{\mathcal{X}})$, for $p := \frac{\alpha\beta\gamma}{\alpha\beta+\alpha\gamma+\beta\gamma}$ or consequently its limit value if any $\alpha,\beta,\gamma$ equal $\infty$. Moreover, we can estimate*

$$\|U\|_{L_p(\Omega;\tilde{\mathcal{X}})} \leq \sqrt{2}\left\|\frac{1}{A_{\min}}\right\|_{L_\gamma(\Omega)}(\|A_{\max}\|_{L_\beta(\Omega)}+1)\|\tilde{\mathcal{F}}_\omega\|_{L_\alpha(\Omega;\tilde{\mathcal{Y}}')},$$

*with norms measured in its respective function spaces (i),(ii),(iii).*

*Proof.* We notice preliminary that

$$\frac{1}{p} = \frac{\alpha\beta+\alpha\gamma+\beta\gamma}{\alpha\beta\gamma} = \frac{1}{\alpha}+\frac{1}{\beta}+\frac{1}{\gamma}$$

by the definition of $p$ and that the consistency condition (5.1.10) implies $p \geq 1$. Therefore, this choice allows to apply a generalization of Hölder's inequality 2.1.7 and one can directly estimate

$$\begin{aligned}
\|U\|_{L_p(\Omega;\tilde{\mathcal{X}})} &\leq \sqrt{2}\left\|\frac{1}{A_{\min}}\right\|_{L_\gamma(\Omega)}\|\max\{1,A_{\max}\}\|_{L_\beta(\Omega)}\|\tilde{\mathcal{F}}_\omega\|_{L_\alpha(\Omega;\tilde{\mathcal{Y}}')}\\
&\leq \sqrt{2}\left\|\frac{1}{A_{\min}}\right\|_{L_\gamma(\Omega)}\|1+|A_{\max}|\|_{L_\beta(\Omega)}\|\tilde{\mathcal{F}}_\omega\|_{L_\alpha(\Omega;\tilde{\mathcal{Y}}')}\\
&\leq \sqrt{2}\left\|\frac{1}{A_{\min}}\right\|_{L_\gamma(\Omega)}(1+\|A_{\max}\|_{L_\beta(\Omega)})\|\tilde{\mathcal{F}}_\omega\|_{L_\alpha(\Omega;\tilde{\mathcal{Y}}')},
\end{aligned}$$

by the deterministic estimate (5.1.8), which proves the claim. □

The previous Theorem 5.1.9 guarantees existence of $p$-moments of the solution $U$ of the random PDE (5.1.2) depending on the existence of moments for $A_{\min}^{-1}$, $A_{\max}$, $g$ and $U_0$, whose regularity or number of moments is classified by the parameters $\alpha$, $\beta$ and $\gamma$. Indeed, under the consistency condition (5.1.10) on the parameters, it was proven that the solution has at least $p := \frac{\alpha\beta\gamma}{\beta\gamma+\alpha\beta+\alpha\gamma} \geq 1$ moments. Notice, that one has the embedding $L_q(\Omega) \subset L_{q'}(\Omega)$ for $1 \leq q' \leq q \leq \infty$, since the probability measure is per definition finite. This justifies to speak of the existence of $p$ moments instead of the $p$-th moment only.

Due to the embedding $L_q(\Omega) \subset L_{q'}(\Omega)$ for $1 \leq q' \leq q \leq \infty$, one could also replace $p$ by any $p' \leq p$ with $p' \in [1,\infty]$ and each $\alpha,\beta,\gamma$ by any $\alpha',\beta',\gamma'$ with $\alpha' \geq \alpha$, $\beta' \geq \beta$ and $\gamma' \geq \gamma$ with $\alpha',\beta',\gamma' \in [1,\infty]$.

In order to get a clearer insight, we give a simple example. Having $A_{\min}^{-1}$, $A_{\max} \in L_3(\Omega)$ and $\tilde{\mathcal{F}}_\omega \in L_3(\Omega;\tilde{\mathcal{Y}}')$, i.e., if the third moments exist, then the consistency condition (5.1.10) is satisfied and we obtain $U \in L_1(\Omega;\tilde{\mathcal{X}})$. That means, that $U$ has finite expectation $\mathbb{E}(\|U\|_{\tilde{\mathcal{X}}}) := \int_\Omega \|U\|_{\tilde{\mathcal{X}}}\,\mathrm{d}\mathbb{P} < \infty$ for these choices.

Next, we want to discuss some limit cases, meaning uniform boundedness. In the "trivial" case of only uniformly bounded mappings $\alpha,\beta,\gamma = \infty$, we immediately obtain the

existence of arbitrary moments of $U$, as already discussed earlier. Since Theorem 5.1.9 also holds true for arbitrary permutations of $(\alpha, \beta, \gamma)$, we only have to consider two further limit cases in order to cover all possible ones. To this end, assume without loss of generality first that $\beta = \gamma = \infty$. Then we can immediately conclude that $A_{\max}, A_{\min}^{-1} \in L_{\infty(\Omega)}$ yields

$$\|U\|_{L_p(\Omega;\tilde{\mathcal{X}})} \leq \sqrt{2}\, \|\max\{A_{\min}^{-1}(\omega), A_{\max}(\omega)A_{\min}^{-1}(\omega)\}\|_{L_\infty(\Omega)} \|\tilde{f}_\omega\|_{L_p(\Omega;\tilde{\mathcal{Y}}')}$$

or the limit case of Theorem 5.1.9

$$\|U\|_{L_p(\Omega;\tilde{\mathcal{X}})} \leq \sqrt{2}\, \|\frac{1}{A_{\min}}\|_{L_\infty(\Omega)} (\|A_{\max}\|_{L_\infty(\Omega)} + 1) \|\tilde{f}_\omega\|_{L_p(\Omega;\tilde{\mathcal{Y}}')}.$$

Therefore, the finiteness of the $p$-moments of the solution $U$ coincides with the one of the $p$-moments of $\tilde{\mathcal{F}}_\omega$, that is, of the initial data $U_0$ and of the right hand side $g$. Finally, for uniformly bounded initial condition $U_0$ and right hand side $g$ with respect to $\Omega$, that is, $\alpha = \infty$, we can conclude

$$\|U\|_{L_p(\Omega;\tilde{\mathcal{X}})} \leq \sqrt{2}\, \left\| \frac{1}{A_{\min}} \max\{1, A_{\max}\} \right\|_{L_p(\Omega)} \|\tilde{f}_\omega\|_{L_\infty(\Omega;\tilde{\mathcal{Y}}')}.$$

So the existence of $p$-moments of the solution $U$ coincides with the one of the $p$-moments of the quotient $\frac{\max\{1, A_{\max}\}}{A_{\min}}$. By using Hölder's inequality, we arrive at $\frac{1}{p} = \frac{1}{\beta} + \frac{1}{\gamma}$, such that $U$ has $p = \frac{\beta\gamma}{\beta+\gamma}$ moments. This is consistent with Theorem 5.1.9, since

$$\lim_{\alpha\to\infty} \frac{\alpha\beta\gamma}{\alpha\beta + \alpha\gamma + \beta\gamma} = \frac{\beta\gamma}{\beta+\gamma}, \quad \lim_{\beta,\gamma\to\infty} \frac{\alpha\beta\gamma}{\alpha\beta + \alpha\gamma + \beta\gamma} = \alpha,$$

$$\lim_{\alpha,\beta,\gamma\to\infty} \frac{\alpha\beta\gamma}{\alpha\beta + \alpha\gamma + \beta\gamma} = \infty.$$

An overview of all combinations are summarized in Table 5.1.11.

Table 5.1.11: Parameters according to Theorem 5.1.9 without permutations.

| $\alpha$ | $\beta$ | $\gamma$ | $p$ |
|---|---|---|---|
| finite | finite | finite | $\frac{\alpha\beta\gamma}{\alpha\beta+\alpha\gamma+\beta\gamma}$ |
| $\infty$ | finite | finite | $\frac{\beta\gamma}{\beta+\gamma}$ |
| finite | $\infty$ | $\infty$ | $\alpha$ |
| $\infty$ | $\infty$ | $\infty$ | $\infty$ |

As in section 3.3, we can improve the result of Theorem 5.1.9 for time-independent spatial differential operators. To see this, we recall the inf-sup estimates (3.3.12)

$$\inf_{v\in\tilde{\mathcal{X}}\setminus\{0\}}\sup_{w\in\tilde{\mathcal{Y}}\setminus\{0\}}\frac{|\langle\tilde{B}_\omega v,w\rangle|}{\|v\|_{\tilde{\mathcal{X}}}\|w\|_{\tilde{\mathcal{Y}}}}\geq\frac{\min\{1,A_{\min}(\omega)\}}{\sqrt{2}},\quad\omega\in\Omega\text{ a.s.},$$

from Proposition 3.3.10 adapted to the context of random PDEs. We formulate the result in the following Corollary.

**Corollary 5.1.12.** *Let $\bar{\tilde{\mathcal{X}}}$, $\bar{\tilde{\mathcal{Y}}}$ as well as $W_-$, $W_0$, $W_+$ be as in Proposition 3.3.10. Assume that the operator $A$ is time-independent and self-adjoint with*

$$A_{\min}(\omega)\leq\|A'\|_{W_+\to W_-}\leq A_{\max}(\omega),\quad\omega\in\Omega\ a.s.,$$

*with* random variables $A_{\min}^{-1}\in L_\gamma(\Omega)$ *and* $A_{\max}<\infty$ $\mathbb{P}$*-a.s.. Moreover, let* $\tilde{\mathcal{F}}_\omega\in L_\alpha(\Omega;\bar{\tilde{\mathcal{Y}}}')$ *with parameters* $\alpha,\gamma\in[1,\infty]$ *arranged such that*

$$\alpha\gamma\geq\alpha+\gamma,$$

*then the solution $U$ to problem* (5.1.2) *belongs to* $L_p(\Omega;\bar{\tilde{\mathcal{X}}})$ *for* $p:=\frac{\alpha\gamma}{\alpha+\gamma}$ *and there holds*

$$\|U\|_{L_p(\Omega;\bar{\tilde{\mathcal{X}}})}\leq\sqrt{2}\,\left(1+\left\|\frac{1}{A_{\min}}\right\|_{L_\gamma(\Omega)}\right)\,\|\tilde{\mathcal{F}}_\omega\|_{L_\alpha(\Omega;\bar{\tilde{\mathcal{Y}}}')}.$$

*Proof.* The proof follows by Theorem 5.1.9 combined with Proposition 3.3.10. Proposition 3.3.10 yields

$$\inf_{v\in\tilde{\mathcal{X}}\setminus\{0\}}\sup_{w\in\tilde{\mathcal{Y}}\setminus\{0\}}\frac{|\langle\tilde{B}_\omega v,w\rangle|}{\|v\|_{\tilde{\mathcal{X}}}\|w\|_{\tilde{\mathcal{Y}}}}\geq\frac{\min\{1,A_{\min}(\omega)\}}{\sqrt{2}},\quad\omega\in\Omega\text{ a.s.},$$

such that

$$\|U\|_{L_p(\Omega;\bar{\tilde{\mathcal{X}}})}\leq\sqrt{2}\,\left\|\max\{1,\frac{1}{A_{\min}}\}\tilde{\mathcal{F}}_\omega\right\|_{L_p(\Omega;\bar{\tilde{\mathcal{Y}}}')}.$$

Using Hölder's inequality 2.1.7 with $\frac{1}{p}=\frac{1}{\alpha}+\frac{1}{\gamma}$ proves the claim in the same way as in the proof of Theorem 5.1.9. □

It is important to note that, other than in Theorem 5.1.9, the existence of moments for solution $U$ in Corollary 5.1.12 does *not* depend on $A_{\max}$. This means that the spatial differential operator may tend to infinity arbitrary fast, as long as only $A_{\max}<\infty$ almost surely. We are able to prove such results independent of $A_{\max}$ also for time-dependent spatial differential operators.

To this end, we first introduce *$\omega$-dependent norms* on $\tilde{\mathcal{X}}$ and $\tilde{\mathcal{Y}}$, namely

$$\begin{aligned}\|v\|_{\tilde{\mathcal{X}}_\omega}^2 &:= \int_0^T \|A^{\frac{1}{2}}(t,\omega)v(t)\|_H^2 \,\mathrm{d}t, \quad \mathbb{P}\text{-a.s.},\\ \|w\|_{\tilde{\mathcal{Y}}_\omega} &:= \|w(0)\|_H^2 + \int_0^T \left(\|A^{\frac{1}{2}}(t,\omega)w(t)\|_H^2 + \|A^{-\frac{1}{2}}(t,\omega)\dot{w}(t)\|_H^2\right)\mathrm{d}t, \quad \mathbb{P}\text{-a.s.},\end{aligned} \tag{5.1.13}$$

which are well-defined under the assumption that $A(t,\omega)$ is self-adjoint, cf. also Definition 2.2.13. These are indeed equivalent norms on $\tilde{\mathcal{X}}$ and $\tilde{\mathcal{Y}}$, respectively, with equivalence constants which will be derived later. In this way we move the $\omega$-dependence into the norms. Now we can derive optimal bounds with respect to these norms and examine the $\omega$-dependence afterwards. The following lemma will be useful to arrive at these optimal bounds.

**Lemma 5.1.14.** *The norm $|\cdot|_{\tilde{\mathcal{Y}}_\omega}$, defined by*

$$|w|_{\tilde{\mathcal{Y}}_\omega}^2 := \int_0^T \left\|A^{\frac{1}{2}}(t,\omega)w(t) - A^{-\frac{1}{2}}(t,\omega)\dot{w}(t)\right\|_H^2 \mathrm{d}t, \quad \mathbb{P}\text{-}a.s.$$

*is equal to the norm $\|\cdot\|_{\tilde{\mathcal{Y}}_\omega}$ defined in* (5.1.13).

*Proof.* A straightforward calculation shows

$$\begin{aligned}|w|_{\tilde{\mathcal{Y}}_\omega}^2 &= \int_0^T \left\|A^{\frac{1}{2}}(t,\omega)w(t) - A^{-\frac{1}{2}}(t,\omega)\dot{w}(t)\right\|_H^2 \mathrm{d}t\\ &= \int_0^T \left(\|A^{\frac{1}{2}}(t,\omega)w(t)\|_H^2 + \|A^{-\frac{1}{2}}(t,\omega)\dot{w}(t)\|_H^2 - 2\,{}_V\langle w(t),\dot{w}(t)\rangle_{V'}\right)\mathrm{d}t\\ &= \|w(0)\|_H^2 + \int_0^T \left(\|A^{\frac{1}{2}}(t,\omega)w(t)\|_H^2 + \|A^{-\frac{1}{2}}(t,\omega)\dot{w}(t)\|_H^2\right)\mathrm{d}t\\ &= \|w\|_{\tilde{\mathcal{Y}}_\omega}^2, \quad \mathbb{P}\text{-a.s.},\end{aligned}$$

since

$${}_{V'}\langle\dot{w}(t),w(t)\rangle_V + {}_V\langle w(t),\dot{w}(t)\rangle_{V'} = \frac{\mathrm{d}}{\mathrm{d}t}(w(t),w(t))_H,$$

cf. [SS09, Appx. A]. □

Using this lemma, we can prove the following theorem.

**Theorem 5.1.15.** *The bilinear form $\tilde{b}_\omega(\cdot,\cdot)\colon \tilde{\mathcal{X}}\times\tilde{\mathcal{Y}}\to\mathbb{R}$ defined in* (5.1.3) *satisfies*

$$\begin{aligned}\sup_{v\in\tilde{\mathcal{X}}\setminus\{0\}}\sup_{w\in\tilde{\mathcal{Y}}\setminus\{0\}}\frac{|\tilde{b}_\omega(v,w)|}{\|v\|_{\tilde{\mathcal{X}}_\omega}\|w\|_{\tilde{\mathcal{Y}}_\omega}} &= 1, && \mathbb{P}\text{-}a.s.,\\ \inf_{v\in\tilde{\mathcal{X}}\setminus\{0\}}\sup_{w\in\tilde{\mathcal{Y}}\setminus\{0\}}\frac{\tilde{b}_\omega(v,w)}{\|v\|_{\tilde{\mathcal{X}}_\omega}\|w\|_{\tilde{\mathcal{Y}}_\omega}} &= 1, && \mathbb{P}\text{-}a.s.,\end{aligned}$$

*with norms defined in* (5.1.13).

*Proof.* The proof is based on the same idea as used in [UP14]. We first notice that the following inequality holds $\mathbb{P}$-a.s.:

$$\begin{aligned}
|\tilde{b}_\omega(v,w)| &\leq \int_0^T |{}_V\langle v(t), -\dot{w}(t) + A(t,\omega)w(t)\rangle_{V'}| \mathrm{d}t \\
&= \int_0^T \left| \left( A^{\frac{1}{2}}(t,\omega)v(t), -A^{-\frac{1}{2}}(t,\omega)\dot{w}(t) + A^{\frac{1}{2}}(t,\omega)w(t) \right)_H \right| \mathrm{d}t \\
&\leq \int_0^T \left( \|A^{\frac{1}{2}}(t,\omega)v(t)\|_H \ \ \| - A^{-\frac{1}{2}}(t,\omega)\dot{w}(t) + A^{\frac{1}{2}}(t,\omega)w(t)\|_H \right) \mathrm{d}t \\
&\leq \|v\|_{\tilde{\mathcal{X}}_\omega} |w|_{\tilde{\mathcal{Y}}_\omega} = \|v\|_{\tilde{\mathcal{X}}_\omega} \|w\|_{\tilde{\mathcal{Y}}_\omega},
\end{aligned}$$

by Lemma 5.1.14 and Cauchy-Schwarz/Hölder inequality in the last inequality. In order to prove the inf-sup condition, we swap the test and solution spaces according to Proposition 2.3.7. We choose

$$v_w(t) := w(t) - A^{-1}(t,\omega)\dot{w}(t)$$

for arbitrary $w \in \tilde{\mathcal{Y}}$ and obtain $\mathbb{P}$-a.s.

$$\begin{aligned}
\tilde{b}_\omega(v_w, w) &= \int_0^T {}_V\langle w(t) - A^{-1}(t,\omega)\dot{w}(t), -\dot{w}(t) + A(t,\omega)w(t)\rangle_{V'} \ \mathrm{d}t \\
&= \int_0^T \left\| A^{\frac{1}{2}}(t,\omega)w(t) - A^{-\frac{1}{2}}(t,\omega)\dot{w}(t) \right\|_H^2 \ \mathrm{d}t \\
&= |w|_{\tilde{\mathcal{Y}}_\omega}^2.
\end{aligned}$$

By the choice of $v_w$ there holds $\mathbb{P}$-a.s.

$$\|v_w\|_{\tilde{\mathcal{X}}_\omega}^2 = \|A^{\frac{1}{2}}(\cdot,\omega)(w - A^{-1}(\cdot,\omega)\dot{w})\|_{L_2(I;H)}^2 = |w|_{\tilde{\mathcal{Y}}_\omega}^2 = \|w\|_{\tilde{\mathcal{Y}}_\omega}^2,$$

such that

$$\tilde{b}_\omega(v_w, w) = \|v_w\|_{\tilde{\mathcal{X}}_\omega} \|w\|_{\tilde{\mathcal{Y}}_\omega}, \quad \mathbb{P}\text{-a.s.}$$

Having a lower inf-sup bound of one as well as an upper sup-sup bound of one implies equality. □

The upper and lower bound in Theorem 5.1.15 are *optimal* since they are obviously the best possible bounds and in particular independent of $\omega$. Having these optimal inf-sup bound, we also obtain an optimal estimate of the solution in the $\omega$-dependent norms, namely

$$\|U\|_{\tilde{\mathcal{X}}_\omega} \leq \|\tilde{\mathcal{F}}_\omega\|_{\tilde{\mathcal{Y}}_\omega'}, \quad \mathbb{P}\text{-a.s.},$$

where $\|\cdot\|_{\tilde{\mathcal{Y}}_\omega'}$ denotes the dual norm with respect to the modified primal norm $\|\cdot\|_{\tilde{\mathcal{Y}}_\omega}$. Now we can estimate both sides directly with respect to the standard norms by "pulling

out" the $\omega$-dependence again. The right hand side can be calculated similar to (5.1.7) as

$$\begin{aligned}
|\tilde{\mathcal{F}}_\omega(w)| &= |\int_I {}_{V'}\langle g(t,\omega), w(t)\rangle_V \, \mathrm{d}t + (U_0(\omega), w(0))_H| \\
&= |\int_I (A^{-\frac{1}{2}}(t,\omega)g(t,\omega), A^{\frac{1}{2}}(t,\omega)w(t))_H \, \mathrm{d}t + (U_0(\omega), w(0))_H| \\
&\leq \|A^{-\frac{1}{2}}(\cdot,\omega)g(\cdot,\omega)\|_{L_2(I;H)}\|A^{\frac{1}{2}}(\cdot,\omega)w\|_{L_2(I;H)} + \|U_0(\omega)\|_H\|w(0)\|_H \\
&\leq \left(\|A^{-\frac{1}{2}}(\cdot,\omega)g(\cdot,\omega)\|_{L_2(I;H)} + \|U_0(\omega)\|_H\right)\|w\|_{\tilde{\mathcal{Y}}_\omega} \\
&\leq \Big(\frac{1}{\sqrt{A_{\min}}}\|g(\cdot,\omega)\|_{\tilde{\mathcal{X}}'} + \|U_0(\omega)\|_H\Big)\|w\|_{\tilde{\mathcal{Y}}_\omega}, \quad \mathbb{P}\text{-a.s.},
\end{aligned}$$

for arbitrary $w \in \tilde{\mathcal{Y}}$ and the left hand side as

$$\|U\|^2_{\tilde{\mathcal{X}}_\omega} = \int_0^T {}_{V'}\langle A(t,\omega)U(t), U(t)\rangle_V \mathrm{d}t \geq A_{\min}\|U\|^2_{\tilde{\mathcal{X}}}, \quad \mathbb{P}\text{-a.s.},$$

where we have used that

$$\begin{aligned}
\|A^{-\frac{1}{2}}(\cdot,\omega)g(\cdot,\omega)\|^2_{L_2(I;H)} &= \int_0^T {}_V\langle A(t,\omega)^{-1}g(t,\omega), g(t,\omega)\rangle_{V'}\mathrm{d}t \\
&\leq \frac{1}{A_{\min}}\|g(\cdot,\omega)\|^2_{\tilde{\mathcal{X}}'}, \quad \mathbb{P}\text{-a.s.}
\end{aligned}$$

Putting these estimates together, we end up with

$$\|U\|_{\tilde{\mathcal{X}}} \leq \frac{1}{A_{\min}}\|g(\cdot,\omega)\|_{\tilde{\mathcal{X}}'} + \frac{1}{\sqrt{A_{\min}}}\|U_0(\omega)\|_H, \quad \mathbb{P}\text{-a.s.} \tag{5.1.16}$$

Again we can see that the inequality does not depend on $A_{\max}$. Moreover, we have slightly *improved estimates* and, in particular, no maximum involved but different factors in front of both terms $g$ and $U_0$. Notice that the same approach would not be directly applicable to the first formulation, since we exploit that we only have to estimate the norm on $\tilde{\mathcal{X}}$ and not on an intersection space. The norm on the intersection space is directly calculated for our particular right hand side $\tilde{\mathcal{F}}_\omega$ here. We can also give a pendent of Theorem 5.1.9 for this improved result for self-adjoint spatial differential operators.

**Corollary 5.1.17.** *Assume that $A(t)$ is self-adjoint for $t \in I$ a.e. and that there exist parameters $\alpha, \beta, \gamma \in [1, \infty]$ with*

$$\alpha\gamma \geq \alpha + \gamma, \quad 2\beta\gamma \geq \beta + 2\gamma. \tag{5.1.18}$$

*Let the data $g$ and $U_0$, and the random variable $A_{\min}$ are such that:*

*(i)* $g \in L_\alpha(\Omega;\tilde{\mathcal{X}}')$,

*(ii)* $U_0 \in L_\beta(\Omega;H)$,

*(iii)* $\frac{1}{A_{\min}} \in L_\gamma(\Omega)$.

*Then the solution $U$ to problem (5.1.2) belongs to $L_p(\Omega)$, for $p := \min\{\frac{\alpha\gamma}{\alpha+\gamma}, \frac{2\beta\gamma}{\beta+2\gamma}\}$ or consequently its limit value if any $\alpha, \beta, \gamma$ equal $\infty$. Moreover, we can estimate*

$$\|U\|_{L_p(\Omega;\tilde{\mathcal{X}})} \leq \left\|\frac{1}{A_{\min}}\right\|_{L_\gamma(\Omega)} \|g\|_{L_\alpha(\Omega;\tilde{\mathcal{X}}')} + \left\|\frac{1}{A_{\min}}\right\|^{\frac{1}{2}}_{L_\gamma(\Omega)} \|U_0\|_{L_\beta(\Omega;H)},$$

*with norms measured in its respective function spaces (i),(ii),(iii).*

*Proof.* With (5.1.16) and Minkowski's inequality we can estimate

$$\|U\|_{L_p(\Omega;\tilde{\mathcal{X}})} \leq \left\|\frac{1}{A_{\min}} g\right\|_{L_p(\Omega;\tilde{\mathcal{X}}')} + \left\|\frac{1}{\sqrt{A_{\min}}} U_0\right\|_{L_p(\Omega;H)}.$$

For the first term we have by Hölder's inequality:

$$\left\|\frac{1}{A_{\min}} g\right\|_{L_p(\Omega;\tilde{\mathcal{X}}')} \leq \left\|\frac{1}{A_{\min}}\right\|_{L_\gamma(\Omega)} \|g\|_{L_\alpha(\Omega;\tilde{\mathcal{X}}')},$$

due to the consistency condition (5.1.18) implying $\frac{1}{p} \geq \frac{1}{\alpha} + \frac{1}{\gamma}$. The second term can be estimated analogously as

$$\left\|\frac{1}{\sqrt{A_{\min}}} U_0\right\|_{L_p(\Omega;H)} = \left\|\frac{1}{\sqrt{A_{\min}}}\right\|_{L_{2\gamma}(\Omega)} \|U_0\|_{L_\beta(\Omega;H)},$$

again due to the consistency condition (5.1.18) implying $\frac{1}{p} \geq \frac{1}{\beta} + \frac{1}{2\gamma}$. By

$$\left\|\frac{1}{\sqrt{A_{\min}}}\right\|_{L_{2\gamma}(\Omega)} \leq \left\|\frac{1}{A_{\min}}\right\|^{\frac{1}{2}}_{L_\gamma(\Omega)},$$

we finally obtain

$$\|U\|_{L_p(\Omega;\tilde{\mathcal{X}})} \leq \left\|\frac{1}{A_{\min}}\right\|_{L_\gamma(\Omega)} \|g\|_{L_\alpha(\Omega;\tilde{\mathcal{X}}')} + \left\|\frac{1}{A_{\min}}\right\|^{\frac{1}{2}}_{L_\gamma(\Omega)} \|U_0\|_{L_\beta(\Omega;H)}.$$

□

It is worth stressing that the assumptions for the existence of moments in Theorem 5.1.9, Corollary 5.1.12 and Corollary 5.1.17 are quite weak and general, but sufficient and not a necessary condition. If we have more knowledge how $\omega$ enters explicitly, we can often expect better estimates.

**Example 5.1.19.** *Let $V := H_0^1(D)$, $H := L_2(D)$, $D := (0,1)$, $\eta_0 \neq \eta_1 \in (0,1)$, $\alpha \in (0,1)$ and $\mathbb{P}$ the Lebesgue measure, $\Omega := [0,1]$ and Borel $\sigma$-algebra $\Sigma$. Define the spatial differential operator as*

$$A(t,\omega) := -|\omega - \eta_0|^\alpha \Delta_x,$$

*with spatial Laplacian $\Delta_x$ and the right hand side as*

$$g(t,x,\omega) := \Big(\frac{1}{|\omega - \eta_1|}\Big)^\alpha \bar{g}(t,x),$$

*for some $\bar{g} \in L_2(I; H^{-1}(D))$ and zero initial condition $U_0 \equiv 0$. We have that $g \in L_p(\Omega; L_2(I; H^{-1}(D))$ and $\frac{1}{A_{\min}} \in L_p(\Omega)$ for any $p < \frac{1}{\alpha}$. For each trajectory, that is, for fixed $\omega \in \Omega$, one can show that*

$$\|U\|_{L_2(I;H_0^1(D))} \lesssim \Big(\frac{1}{|\omega - \eta_0|}\Big)^\alpha \Big(\frac{1}{|\omega - \eta_1|}\Big)^\alpha \|\bar{g}\|_{L_2(I;H^{-1}(D))}.$$

*Since $\eta_0 \neq \eta_1$, there are two singularities at* different *points, such that $U \in L_p(\Omega; L_2(I; H_0^1(D)))$ for any $p < \frac{1}{\alpha}$ as well. With Corollary 5.1.17 we could only conclude $U \in L_q(\Omega; L_2(I; H_0^1(D)))$ for $q < \frac{1}{2\alpha}$.*

Finally, we consider also non-coercive spatial differential operators, which only satisfy a Gårding inequality, cf. (3.1.2). Recalling Corollary 3.3.16, we can go similar lines as before to derive existence of moments for the solution. Assuming a fixed parameter $\lambda$ independent of $\omega$ would result in a very similar behavior as elaborated in Theorem 5.1.9, but with worse (true) constants. However, since we assume the spatial differential operator $A(t,\omega)$ to be a random operator bounded by random variables, it is more meaningful to also consider $\lambda = \lambda(\omega)$ as a random variable. Again we are mainly interested in the case of not uniformly bounded $\lambda \notin L_\infty(\Omega)$. Combining the ideas of this chapter, in particular of Theorem 5.1.9, with Corollary 3.3.16, one obtains the following corollary in a straightforward manner.

**Corollary 5.1.20.** *Assume that $A$ satisfies a Gårding-inequality* (3.1.2) *$\mathbb{P}$-a.s. and is not necessarily coercive. Let the data $g$ and $U_0$, and the random variables $A_{\min}$ and $A_{\max}$ as well as $\lambda$ are such that:*

*(i)* $\tilde{\mathcal{F}}_\omega$ *belongs to* $L_{\alpha_1}(\Omega; \tilde{\mathcal{Y}}')$, *that is,* $g \in L_{\alpha_1}(\Omega; \tilde{\mathcal{Y}}')$ *and* $U_0 \in L_{\alpha_1}(\Omega; H)$,

*(ii)* $A_{\max} \in L_{\alpha_2}(\Omega)$,

*(iii)* $\frac{1}{A_{\min}} \in L_{\alpha_3}(\Omega)$,

*(iv)* $e^{\lambda T} \in L_{\alpha_4}(\Omega)$,

*(v)* $\lambda \in L_{\alpha_5}(\Omega) \cap L_{\alpha_2}(\Omega)$,

*where* $\alpha_i \in [1,\infty]$ *for all* $i = 1,\dots,5$*, such that*

$$\prod_{j=1}^{5} \alpha_j \geq \sum_{j=1}^{5} \prod_{\substack{i=1 \\ i\neq j}}^{5} \alpha_i.$$

*Then the solution* $U$ *to problem* (5.1.2) *belongs to* $L_p(\Omega)$*, for*

$$p := \frac{\prod_{j=1}^{5} \alpha_j}{\sum_{j=1}^{5} \prod_{\substack{i=1 \\ i\neq j}}^{5} \alpha_i}.$$

*Moreover, we can estimate*

$$\begin{aligned}\|U\|_{L_p(\Omega;\tilde{\mathcal{X}})} \leq & \sqrt{2}\, \|e^{\lambda T}\|_{L_{\alpha_4}(\Omega)} (\sqrt{2}\varrho^2 \|\lambda\|_{L_{\alpha_5}(\Omega)} + \sqrt{2} + 1) \\ & \times (\|A_{\max}\|_{L_{\alpha_2}(\Omega)} + \|\lambda\|_{L_{\alpha_2}(\Omega)} + 1) \|A_{\min}^{-1}\|_{L_{\alpha_3}(\Omega)} \|\tilde{\mathcal{F}}_\omega\|_{L_{\alpha_1}(\Omega;\tilde{\mathcal{Y}}')},\end{aligned}$$

*with norms measured in its respective function spaces (i) – (v).*

*Proof.* The proof follows in a similar way as the proof of Theorem 5.1.9. We first obtain

$$\|U\|_{\tilde{\mathcal{X}}} \leq \sqrt{2} e^{\lambda} \max\{\sqrt{1+2\lambda^2\varrho^4}, \sqrt{2}\} \frac{1}{A_{\min}} \max\{1, A_{\max} + \lambda\} \|\tilde{\mathcal{F}}_\omega\|_{\tilde{\mathcal{Y}}'}, \quad \mathbb{P}\text{-a.s.},$$

according to Corollary 3.3.16. Applying a generalized Hölder-inequality 2.1.7 with

$$\frac{1}{p} = \frac{\sum_{j=1}^{5} \prod_{\substack{i=1 \\ i\neq j}}^{5} \alpha_i}{\prod_{j=1}^{5} \alpha_j} = \sum_{j=1}^{5} \frac{1}{\alpha_i}$$

yields

$$\begin{aligned}\|U\|_{L_p(\Omega;\tilde{\mathcal{X}})} \leq & \sqrt{2} \|e^{\lambda}\|_{L_{\alpha_4}(\Omega)} \| \max\{\sqrt{1+2\lambda^2\varrho^4}, \sqrt{2}\}\|_{L_{\alpha_5}(\Omega)} \Big\| \frac{1}{A_{\min}} \Big\|_{L_{\alpha_3}(\Omega)} \\ & \times \| \max\{1, A_{\max} + \lambda\}\|_{L_{\alpha_2}(\Omega)} \|\tilde{\mathcal{F}}_\omega\|_{L_{\alpha_1}(\Omega;\tilde{\mathcal{Y}}')}.\end{aligned}$$

Estimating the maximums similar as done in the proof of Theorem 5.1.9 together with Minkowski's inequality and

$$\| \max\{\sqrt{1+2\lambda^2\varrho^4}, \sqrt{2}\}\|_{L_{\alpha_5}(\Omega)} \leq \| \max\{1 + \sqrt{2}\lambda\varrho^2, \sqrt{2}\}\|_{L_{\alpha_5}(\Omega)}$$

proves the claim. □

We need to ensure that $e^{\lambda(\omega)T}$ has at least finite expectation, that is, $e^{\lambda(\omega)T} \in L_1(\Omega)$. Such kind of random variables are closely connected with *moment-generating functions*. In probability theory, the moment-generating function of a random variable $\lambda$ is given

as $M_\lambda(T) := \mathbb{E}\big[e^{T\lambda}\big]$ for $T \in \mathbb{R}$. Even though moment-generating functions are used to determine moments of random variables $\lambda$, one can make use of the well known results and knowledge of them here. One very prominent example, for instance, is the *log-norm distribution* $e^{\bar{\lambda}}$, with normally distributed $\bar{\lambda} \sim \mathcal{N}(\mu, \sigma^2)$. It is well known that log-normal random variables have arbitrary many moments with $\mathbb{E}\big[e^{p\bar{\lambda}}\big] = \exp(p\mu + \frac{p^2\sigma^2}{2})$, so they are suitable for our present representation. It is worth mentioning that, for example, for normal distributed $\bar{\lambda}$, meaning log-normal $e^{T\bar{\lambda}}$, no strict uniform upper bound exists and takes values arbitrary close to zero. This is consistent with the observation that all moments of $e^{T\bar{\lambda}}$ exist, but $\|e^{T\bar{\lambda}}\|_{L_p(\Omega)} = \mathbb{E}\big[e^{pT\bar{\lambda}}\big]^{1/p} = \exp(T\mu + \frac{pT^2\sigma^2}{2})$ tends to infinity for higher order moments $p \to \infty$, that is, $e^{T\bar{\lambda}} \notin L_\infty(\Omega)$. In the "trivial" case of $\lambda \in L_\infty(\Omega)$ one directly obtains $e^{T\lambda} \in L_p(\Omega)$ for arbitrary $p \geq 1$, since a probability measure is finite and also $T < \infty$. We refer to chapter 6 for further examples.

We could proceed in a similar manner to prove also sufficient conditions for existence of moments of the solution for the first formulation and its homogenization from section 3.1 and 3.2. But since each of the other formulations under consideration in section 3, and its continuity and inf-sup constants, fit perfectly into the context of Theorem 5.1.9, or Corollary 5.1.12, respectively, we omit an ongoing detailed description, but collect the results in Table 5.1.22 and 5.1.23. A treatment like in Corollary 5.1.17 cannot be adapted straightforwardly to the first formulation or its homogenization as already mentioned before.

A summary of the results above for different operators $A(t)$ ("arbitrary", self-adjoint or time-independent) is given in Table 5.1.21. Moreover, the results extended to the other formulations introduced in chapter 3 are illustrated in Table 5.1.22 and 5.1.23.

## 5.2 Quasi Optimality of Spatial Semidiscretization

Next, we want to make a first attempt to analyze existence and uniqueness of (semi-) discrete solutions and also their quasi-optimality measured in a suitable $L_p(\Omega;\cdot)$-norm. To this end, we introduce a proper spatial (finite) subspace $S \subset V$, where its dual $S'$ can be equipped with two different norms

$$\|s\|_{V'} := \sup_{\substack{v \in V \\ \|v\|_V = 1}} {}_{V'}\langle s, v\rangle_V \quad \text{and} \quad \|s\|_{S'} := \sup_{\substack{v \in S \\ \|v\|_V = 1}} {}_{S'}\langle s, v\rangle_S.$$

The norm $\|\cdot\|_{S'}$ is obviously weaker then $\|\cdot\|_{V'}$. The reversed estimate $\|\cdot\|_{V'} \lesssim \|\cdot\|_{S'}$ is classified by a *subspace dependent* constant

$$c_S := \sup_{0 \neq s \in V'} \frac{\|s\|_{V'}}{\|s\|_{S'}} \tag{5.2.1}$$

Table 5.1.21: Existence of moments for different $A(t)$ in second form, cf. section 3.3.

| Property of $A(t)$ | Input | # finite moments $p$ |
|---|---|---|
| bounded, coercive | $\tilde{\mathcal{F}}_\omega \in L_\alpha(\Omega;\tilde{\mathcal{Y}}')$,<br>$A_{\max} \in L_\beta(\Omega)$,<br>$A_{\min}^{-1} \in L_\gamma(\Omega)$ | $\frac{\alpha\beta\gamma}{\alpha\beta+\alpha\gamma+\beta\gamma}$ |
| bounded, coercive,<br>self-adjoint | $g \in L_\alpha(\Omega;\tilde{\mathcal{X}}')$,<br>$U_0 \in L_\beta(\Omega;H)$,<br>$A_{\min}^{-1} \in L_\gamma(\Omega)$ | $\min\{\frac{\alpha\gamma}{\alpha+\gamma}, \frac{2\beta\gamma}{\beta+2\gamma}\}$ |
| bounded, coercive,<br>self-adjoint,<br>time-independent | $\tilde{\mathcal{F}}_\omega \in L_\alpha(\Omega;\tilde{\bar{\mathcal{Y}}}')$,<br>$A_{\min}^{-1} \in L_\gamma(\Omega)$ | $\frac{\alpha\gamma}{\alpha+\gamma}$ |
| bounded,<br>Gårding-inequality | $\tilde{\mathcal{F}}_\omega \in L_{\alpha_1}(\Omega;\tilde{\mathcal{Y}}')$,<br>$A_{\max} \in L_{\alpha_2}(\Omega)$,<br>$A_{\min}^{-1} \in L_{\alpha_3}(\Omega)$,<br>$e^{\lambda T} \in L_{\alpha_4}(\Omega)$,<br>$\lambda \in L_{\alpha_5}(\Omega) \cap L_{\alpha_2}(\Omega)$ | $\frac{\prod_{j=1}^5 \alpha_j}{\sum_{j=1}^5 \prod_{\substack{i=1\\ i\neq j}}^5 \alpha_i}$ |

Table 5.1.22: Existence of moments for different $A(t)$ in first form, cf. section 3.1.

| Property of $A(t)$ | Input | # finite moments $p$ |
|---|---|---|
| bounded, coercive | $\mathcal{F}_\omega \in L_\alpha(\Omega;\mathcal{Y}')$,<br>$A_{\max} \in L_\beta(\Omega)$,<br>$A_{\min}^{-1} \in L_\gamma(\Omega)$ | $\frac{\alpha\beta\gamma}{\alpha\beta+\alpha\gamma+\beta\gamma}$ |
| bounded,<br>Gårding-inequality | $\mathcal{F}_\omega \in L_{\alpha_1}(\Omega;\mathcal{Y}')$,<br>$A_{\max} \in L_{\alpha_2}(\Omega)$,<br>$A_{\min}^{-1} \in L_{\alpha_3}(\Omega)$,<br>$e^{\lambda T} \in L_{\alpha_4}(\Omega)$,<br>$\lambda \in L_{\alpha_5}(\Omega) \cap L_{\alpha_2}(\Omega)$ | $\frac{\prod_{j=1}^5 \alpha_j}{\sum_{j=1}^5 \prod_{\substack{i=1\\ i\neq j}}^5 \alpha_i}$ |

and will play an important role in the upcoming considerations. By [Tan13, Prop. 3.2] and [XZ03], we have the identity

$$c_S = \|P_S\|_{\mathcal{L}(V',V')} = \|I - P_S\|_{\mathcal{L}(V',V')} = \|I - P_S\|_{\mathcal{L}(V,V)} = \|P_S\|_{\mathcal{L}(V,V)},$$

Table 5.1.23: Existence of moments for different $A(t)$ in homogenized form, cf. section 3.2.

| Property of $A(t)$ | Input | # finite moments $p$ |
|---|---|---|
| bounded, coercive | $\mathcal{F}_\omega \in L_\alpha(\Omega;\mathcal{Y}')$, $A_{\max} \in L_\beta(\Omega)$, $A_{\min}^{-1} \in L_\gamma(\Omega)$ | $\frac{\alpha\beta\gamma}{\alpha\beta+\alpha\gamma+\beta\gamma}$ |
| bounded, coercive, self-adjoint, time-independent | $\mathcal{F}_\omega \in L_\alpha(\Omega;\bar{\mathcal{Y}}')$, $A_{\min}^{-1} \in L_\gamma(\Omega)$ | $\frac{\alpha\gamma}{\alpha+\gamma}$ |
| bounded, Gårding-inequality | $\mathcal{F}_\omega \in L_{\alpha_1}(\Omega;\mathcal{Y}')$, $A_{\max} \in L_{\alpha_2}(\Omega)$, $A_{\min}^{-1} \in L_{\alpha_3}(\Omega)$, $e^{\lambda T} \in L_{\alpha_4}(\Omega)$, $\lambda \in L_{\alpha_5}(\Omega) \cap L_{\alpha_2}(\Omega)$ | $\frac{\prod_{j=1}^5 \alpha_j}{\sum_{j=1}^5 \prod_{\substack{i=1 \\ i\neq j}}^5 \alpha_i}$ |

where $P_S$ denotes the extension to $V'$ of the $H$-orthogonal projection onto $S$. The $H$-orthogonal projector $P_S\colon H \to S$ is defined as

$$P_S v \in S \quad \text{with } (P_s v, w)_H = (v, w)_H \qquad \text{for all } v \in H, w \in S, \tag{5.2.2}$$

which can be extended to $V'$, since $S \subset V$. Now we consider problem (5.1.2) with respect to the semidiscrete spaces

$$\tilde{\mathcal{X}}_S := L_2(I;S) \qquad \text{and} \qquad \tilde{\mathcal{Y}}_S := L_2(I;S) \cap H^1_{0,\{T\}}(I;S'), \tag{5.2.3}$$

with $\|v\|^2_{\tilde{\mathcal{Y}}_S} := \int_0^T (\|v(t)\|_V^2 + \|\dot{v}(t)\|_{S'}^2)\,\mathrm{d}t$ equipped with the weaker norm $\|\cdot\|_{S'}$ on $S'$. Moreover, we replace the spatial differential operator $A$ by its discrete counterpart $A_S$ defined by

$${}_{S'}\langle A_S(\omega,t)v, w\rangle_S := {}_{V'}\langle A(\omega,t)v, w\rangle_V \quad \text{for } v, w \in S, \quad \mathbb{P}\text{-a.s.}$$

Collecting the definitions above, we can state our semidiscrete problem.
Find a solution $u_S \in \tilde{\mathcal{X}}_S$ such that

$$\tilde{b}_{S,\omega}(U_S, v) = \bar{\mathcal{F}}_\omega(v) \qquad \text{for all } v \in \tilde{\mathcal{Y}}_S, \quad \mathbb{P}\text{-a.s.}, \tag{5.2.4}$$

where $\tilde{b}_{S,\omega}(\cdot,\cdot)$ is defined as (5.1.3) with $V$ replaced by $S$, $V'$ replaced by $S'$ and $A$ replaced by $A_S$ and $\bar{\mathcal{F}}_\omega$ defined by (5.1.4) with the approximation $U_{0,S} := P_S U_0 \in S$ of the initial value $U_0$. A proof of existence and uniqueness of the deterministic pendant similar to section 3.3 was also given for this semidiscrete problem (5.2.4) in [Tan13, Prop. 3.8].

**Lemma 5.2.5.** *The operator $\tilde{B}_{S,\omega} \in \mathcal{L}(\tilde{\mathcal{X}}_S, \tilde{\mathcal{Y}}_S')$ defined via the bilinear form in* (5.2.4) *as $\langle \tilde{B}_{S,\omega} v, w\rangle := \tilde{b}_{S,\omega}(v,w)$ for $v \in \tilde{\mathcal{X}}_S$ and $w \in \tilde{\mathcal{Y}}_S$ is boundedly invertible $\mathbb{P}$-a.s. on the semidiscrete subspaces $\tilde{\mathcal{X}}_S$ and $\tilde{\mathcal{Y}}_S$ with*

$$\sup_{v\in\tilde{\mathcal{X}}_S\setminus\{0\}} \sup_{w\in\tilde{\mathcal{Y}}_S\setminus\{0\}} \frac{\langle \tilde{B}_{S,\omega} v, w\rangle}{\|v\|_{\tilde{\mathcal{X}}}\|w\|_{\tilde{\mathcal{Y}}_S}} \leq \sqrt{2}\max\{1, A_{\max}\}, \quad \mathbb{P}\text{-a.s.},$$

*and*

$$\inf_{v\in\tilde{\mathcal{X}}_S\setminus\{0\}} \sup_{w\in\tilde{\mathcal{Y}}_S\setminus\{0\}} \frac{\langle \tilde{B}_{S,\omega} v, w\rangle}{\|v\|_{\tilde{\mathcal{X}}}\|w\|_{\tilde{\mathcal{Y}}_S}} \geq A_{\min}\frac{\min\{1, A_{\max}^{-1}\}}{\sqrt{2}}, \quad \mathbb{P}\text{-a.s.},$$

*where $A_{\min}$ and $A_{\max}$ are the random variables bounding the spatial differential operator $A$.*

*Proof.* Although the proof is already given pathwise in [Tan13, Prop 3.8], we want to point out some facts for a better understanding and therefore sketch the proof. First notice, that even though $\tilde{\mathcal{X}}_S$ and $\tilde{\mathcal{Y}}_S$ are subspaces of $\tilde{\mathcal{X}}$ and $\tilde{\mathcal{Y}}$, respectively, the continuity estimate does not follow immediately, since $\tilde{\mathcal{Y}}_S$ is endowed with a weaker norm. But by replacing $\|\cdot\|_{V'}$ consequently by $\|\cdot\|_{S'}$ and since the spatial part of the solution space is also restricted to the subspace $S \subset V$ one obtains

$$\begin{aligned}
|\tilde{b}_{S,\omega}(v,w)| &\leq \int_I (\|v\|_S\|\dot{w}\|_{S'} + \|A_S(t)v\|_{V'}\|w\|_V)\,\mathrm{d}t \\
&\leq \int_I \|v\|_V(\|\dot{w}\|_{S'} + A_{\max}\|w\|_V)\,\mathrm{d}t \\
&\leq \left(\int_I \|v\|_V^2\,\mathrm{d}t\right)^{1/2}\left(\int_I 2(\|\dot{w}\|_{S'}^2 + A_{\max}^2\|w\|_V^2)\,\mathrm{d}t\right)^{1/2}\,\mathrm{d}t \\
&\leq \sqrt{2}\max\{1, A_{\max}\}\|v\|_{\tilde{\mathcal{X}}}\|w\|_{\tilde{\mathcal{Y}}_S}, \quad \mathbb{P}\text{-a.s.},
\end{aligned}$$

for $v \in \tilde{\mathcal{X}}_S$ and $w \in \tilde{\mathcal{Y}}_S$ with $\|\cdot\|_S := \|\cdot\|_V$. That is, considering the weaker norm $\|\cdot\|_{S'}$ is compensated by having test functions with spatial part in $S \subset V$.

Since

$$A_{\max}^{-1}\|\tilde{w}\|_{S'} \leq \|A_s(t)^{-1}\tilde{w}\|_V \leq A_{\min}^{-1}\|\tilde{w}\|_{S'}$$

and

$$\langle \tilde{w}, A_s(t)^{-1}\tilde{w}\rangle \geq A_{\min}\|A_s(t)^{-1}\tilde{w}\|_V^2,$$

$\mathbb{P}$-a.s., one obtains the desired estimate by choosing

$$v_w := w - A_s(t)^{-1}\dot{w},$$

for arbitrary $w \in \tilde{\mathcal{Y}}_S$ and interchanged arguments according to Proposition 2.3.7.

The proof of the non-degeneracy, i.e., the surjectivity with swapped spaces, follows along the same lines as in the non-discrete situation. □

The bounds in the previous lemma are exactly the same as in the non-discrete formulation from section 3.3. To this end, we can derive the existence of moments of the *semidiscrete solution* exactly as for the continuous counterpart Theorem 5.1.9.

The next step is to show *quasi-optimality* measured with respect to a suitable probability space depending on the regularity of the input data. It was shown in [XZ03, Th. 2] that one can also prove a Céa-like result for non-coercive bilinear forms, provided that the BNB-conditions from Theorem 2.3.3 are fulfilled with respect to the discrete subspaces. Unfortunately, we cannot apply this result directly, since we deal with different norms $\|\cdot\|_{S'} \neq \|\cdot\|_{V'}$ for the discrete and non-discrete space. At this point, the constant $c_S$ defined above (5.2.1), which connects both norms, comes into play. It was shown in [LMM16, Th. 7] and (in the deterministic setting) in [Tan13, Th. 3.9], based on [Hac81, Th. 3.4], that the following quasi-optimality result holds.

**Lemma 5.2.6.** *The semidiscrete Galerkin-solution $U_S$ of* (5.2.4) *is quasi-optimal with*

$$\|U - U_S\|_{\tilde{\mathcal{X}}} \leq c_S\Big(1 + \frac{A_{\max}}{A_{\min}}\Big) \inf_{v \in \tilde{\mathcal{X}}_S} \|U - v\|_{\tilde{\mathcal{X}}}, \qquad \mathbb{P}\text{-}a.s.,$$

*where $U$ denotes the solution of problem* (5.1.2).

It is worth mentioning that since $c_S = \|P_S\|_{\mathcal{L}(V,V)}$ with $H$-orthogonal projection $P_S$, one needs to ensure $V$-stability of the projection. Moreover, it can even be shown, that the stability of $P_S$ is also necessary for the quasi-optimality of the semidiscrete Galerkin-solution, see [Tan13, Th. 3.10]. It is important to mention that the orthogonal projection and therefore the constant $c_S$ does obviously *not* depend on the stochastic parameter $\omega$.

The result of the previous Lemma 5.2.6 allows a similar analysis for the $L_p(\Omega;\cdot)$-boundedness as the one made in section 5.1. Indeed, instead of considering the quasi-optimality almost surely, one can consider its restriction to some space $L_p(\Omega;\cdot)$ of suitable order $p \in [1,\infty]$, which has to be elaborated in the following. First, one obviously still obtains

$$\|U - U_S\|_{\tilde{\mathcal{X}}} \leq c_S(1 + \frac{A_{\max}}{A_{\min}}) \inf_{v \in L_p(\Omega;\tilde{\mathcal{X}}_S)} \|U - v\|_{\tilde{\mathcal{X}}}, \quad \mathbb{P}\text{-a.s.}, \tag{5.2.7}$$

but we can also classify a suitable measure for the norms on the left hand side, which is done in the following theorem.

**Theorem 5.2.8.** *Assume there exist parameters $\alpha, \beta, \gamma \in [1,\infty]$ such that*

$$\alpha\beta\gamma \geq 2\alpha(\beta + \gamma) + \beta\gamma. \tag{5.2.9}$$

*Let the data $g$ and $u_0$, and the random variables $A_{\max}$ and $A_{\min}$ are such that:*

*(i) $g$ belongs to $L_\alpha(\Omega;\tilde{\mathcal{Y}}')$ and $U_0$ to $L_\alpha(\Omega;H)$*

*(ii)* $A_{\max} \in L_\beta(\Omega)$,

*(iii)* $\frac{1}{A_{\min}} \in L_\gamma(\Omega)$.

*Then the error between the semidiscrete Galerkin-solution $U_S$ of* (5.2.4) *and the solution $U$ of problem* (5.1.2) *can be measured in $L_{\bar{p}}(\Omega;\tilde{\mathcal{X}})$, i.e.,*

$$\|U - U_S\|_{\tilde{\mathcal{X}}} \in L_{\bar{p}}(\Omega), \qquad \text{with } \bar{p} := \frac{\alpha\beta\gamma}{\beta\gamma + 2\alpha(\beta+\gamma)}.$$

*Moreover, the approximation is quasi-optimal with*

$$\|U - U_S\|_{L_{\bar{p}}(\Omega;\tilde{\mathcal{X}})} \leq 2c_S \Big( \Big\| \frac{1}{A_{\min}} \Big\|_{L_\gamma(\Omega)} \|A_{\max}\|_{L_\beta(\Omega)} \Big) \inf_{v \in L_p(\Omega;\tilde{\mathcal{X}}_S)} \|U - v\|_{L_p(\Omega;\tilde{\mathcal{X}})},$$

*with $p := \frac{\alpha\beta\gamma}{\beta\gamma+\alpha(\beta+\gamma)}$ as in Theorem 5.1.9.*

*Proof.* Notice first that condition (5.2.9) implies in particular that the solution $U$ to the Problem 5.1.2 as well as its semidiscrete solution $U_s$ belongs to $L_p(\Omega;\tilde{\mathcal{X}})$ by Theorem 5.1.9 and Lemma 5.2.5, since

$$\alpha\beta\gamma \geq 2\alpha(\beta+\gamma) + \beta\gamma > \alpha\beta + \alpha\gamma + \beta\gamma.$$

That is, we can measure the solution $U$ in $L_p(\Omega;\tilde{\mathcal{X}})$. The remaining of the proof follows in a similar way as the proof of Theorem 5.1.9 by interchange the right hand side $\|\tilde{\mathcal{F}}_\omega\|_{\tilde{\mathcal{Y}}'}$ with $\left(\inf_{v\in L_p(\Omega;\tilde{\mathcal{X}}_S)} \|U - v\|_{\tilde{\mathcal{X}}}\right) \in L_p(\Omega)$. In order to do so, we first conclude from the choice of parameters $\beta, \gamma$ and $\bar{p}$ that

$$\frac{1}{\beta} + \frac{1}{\gamma} + \frac{1}{p} = \frac{1}{\alpha} + \frac{2}{\beta} + \frac{2}{\gamma} = \frac{1}{\bar{p}},$$

so that we can apply a generalization of Hölder's inequality Theorem 2.1.7. By Lemma 5.2.6 or better (5.2.7), we can estimate

$$\begin{aligned}
\|U - u_S\|_{L_{\bar{p}}(\Omega;\tilde{\mathcal{X}})} &\leq c_S \mathbb{E}\left[ \left(1 + \frac{A_{\max}}{A_{\min}}\right)^{\bar{p}} \inf_{v\in L_p(\Omega;\tilde{\mathcal{X}}_S)} \|u - v\|_{\tilde{\mathcal{X}}}^{\bar{p}} \right]^{1/\bar{p}} \\
&\leq 2c_S \mathbb{E}\left[ \left(\frac{A_{\max}}{A_{\min}}\right)^{\bar{p}} \inf_{v\in L_p(\Omega;\tilde{\mathcal{X}}_S)} \|u - v\|_{\tilde{\mathcal{X}}}^{\bar{p}} \right]^{1/\bar{p}},
\end{aligned}$$

since $\frac{A_{\max}}{A_{\min}} \geq 1$. Applying Hölder's inequality yields

$$\|U - U_S\|_{L_{\bar{p}}(\Omega;\tilde{\mathcal{X}})} \leq 2c_S \Big( \|\frac{1}{A_{\min}}\|_{L_\gamma(\Omega)} \|A_{\max}\|_{L_\beta(\Omega)} \Big) \inf_{v \in L_p(\Omega;\tilde{\mathcal{X}}_S)} \|U - v\|_{L_p(\Omega;\tilde{\mathcal{X}})},$$

what finishes the proof. □

We notice that in case of $A_{\max}$ and $\frac{1}{A_{\min}}$ uniformly bounded in $\Omega$, we can easily achieve that $p = \bar{p} = \alpha$. In the general case, we have instead that $p > \bar{p}$, and in particular the following representation might help to understand how the two parameters are connected to each other. The discrepancy between $p$ and $\bar{p}$ can be measured by the quotient

$$\theta := \frac{p}{\bar{p}} = \frac{1 + 2\frac{\alpha}{r}}{1 + \frac{\alpha}{r}},$$

where $r := \frac{\beta\gamma}{\beta+\gamma}$. It is thus clear that the quasi-optimality result is not given in the standard sense, that is, the norm on the right-hand side and on the left-hand side do not match. In order to have a quasi-optimality result that involves a $\bar{p}$-norm on $\Omega$, we need to require the existence of further moments for the solution, up to a certain $p = \theta\bar{p}$. For example, if we assume that $\alpha = \beta = \gamma$, the consistency condition (5.2.9) requires that $\alpha \geq 5$, and we obtain that

$$U \in L_{\frac{\alpha}{3}}(\Omega; \tilde{\mathcal{X}}) \Rightarrow U \in L_{\frac{\alpha}{5}}(\Omega; \tilde{\mathcal{X}}), \quad \theta = \frac{5}{3}$$

$$\begin{aligned}\|U - U_S\|_{L_{\frac{\alpha}{5}}(\Omega;\tilde{\mathcal{X}})} \leq 2c_S \Big( \|\frac{1}{A_{\min}}\|_{L_\alpha(\Omega)} \|A_{\max}\|_{L_\alpha(\Omega)} \Big) \\ \times \inf_{v \in L_{\frac{\alpha}{3}}(\Omega;\tilde{\mathcal{X}}_S)} \|U - v\|_{L_{\frac{\alpha}{3}}(\Omega;\tilde{\mathcal{X}})},\end{aligned}$$

which means that, in order to have a least-square estimate, we need $\alpha = 10$, thus obtaining

$$U \in L_{\frac{10}{3}}(\Omega; \tilde{\mathcal{X}}) \Rightarrow U \in L_2(\Omega; \tilde{\mathcal{X}}), \quad \theta = \frac{5}{3}$$

$$\begin{aligned}\|U - U_S\|_{L_2(\Omega;\tilde{\mathcal{X}})} \leq 2c_S \Big( \|\frac{1}{A_{\min}}\|_{L_{10}(\Omega)} \|A_{\max}\|_{L_{10}(\Omega)} \Big) \\ \times \inf_{v \in L_{\frac{10}{3}}(\Omega;\tilde{\mathcal{X}}_S)} \|U - v\|_{L_{\frac{10}{3}}(\Omega;\tilde{\mathcal{X}})}.\end{aligned}$$

It is worth mentioning that the error rate in (5.2.7) does only depend on the *ratio* $\frac{A_{\max}}{A_{\min}}$. This ratio is uniformly bounded in many applications, that is, $\frac{A_{\max}}{A_{\min}} \in L_\infty(\Omega)$. This gives the following important remark.

**Remark 5.2.10.** *Let the quotient* $\frac{A_{\max}}{A_{\min}} \leq C$ *be bounded* uniformly *with a constant* $C < \infty$ *independent of* $\omega$*. Then the error is controlled by*

$$\|U - U_S\|_{L_p(\Omega;\tilde{\mathcal{X}})} \leq 2c_S C \inf_{v \in L_p(\Omega;\tilde{\mathcal{X}}_S)} \|U - v\|_{L_p(\Omega;\tilde{\mathcal{X}})},$$

*provided that* $U \in L_p(\Omega; \tilde{\mathcal{X}})$ *and* $U_S \in L_p(\Omega; \tilde{\mathcal{X}}_S)$*.*

The ratio is uniformly bounded for instance if we consider random coefficients having $A(t,\omega) := a(\omega)\bar{A}(t)$ with deterministic spatial differential operator $\bar{A}(t)$ and random variable $a(\omega)$, meaning that the stochastic parameter decouples with the deterministic part. Such kind of coefficients were basically considered for instance in the disorder potential from [MKM13].

## 5.3 Quasi Optimality of Petrov-Galerkin Approach

In this section, we want to discuss general Petrov-Galerkin approaches for random PDEs and in particular their stability and approximation quality. We proceed similar to the previous sections and reuse the known deterministic behavior $\omega$-wise.

A Petrov-Galerkin discretization was introduced in chapter 4 for deterministic problems. We restate this approach briefly here already in the framework of full space-time formulated parabolic problems with random coefficients. A Petrov-Galerkin approach of the random PDE (5.1.2) in the second formulation is to find a solution $u_j \in S_j$ of the *fully discrete* equation

$$\tilde{b}_\omega(U_j, q_\ell) = \langle \tilde{\mathcal{F}}_\omega, q_\ell \rangle \qquad \text{for all } q_\ell \in Q_\ell, \quad \mathbb{P}\text{-a.s.},$$

with respect to discrete subspaces $S_j \subset \tilde{\mathcal{X}}$ and $Q_\ell \subset \tilde{\mathcal{Y}}$ as well as $\tilde{b}_\omega$ and $\tilde{\mathcal{F}}_\omega$ defined in (5.1.3) and (5.1.4). For stability reasons, recall the observations from chapter 3. In particular, the discretization can be stabilized by enriching the test space and no CFL-condition is required as assumed in [LMM16]. We are interested in the minimal residual Petrov-Galerkin solution given by

$$U_j := \operatorname*{arg\,min}_{v_j \in S_j} \sup_{q_\ell \in Q_\ell \setminus \{0\}} \frac{|\langle \tilde{B}_\omega v_j - \tilde{\mathcal{F}}_\omega, q_\ell \rangle|}{\|q_\ell\|_{\tilde{\mathcal{Y}}}} \quad \mathbb{P}\text{-a.s.} \tag{5.3.1}$$

Although the minimal residual Petrov-Galerkin approach can be seen as a generalization of a Petrov-Galerkin approach, they are closely connected and we often denote both simply by Petrov-Galerkin approach. Recall from Theorem 4.1.8 that one can prove *quasi-optimality* for the minimal residual Petrov-Galerkin solution

$$\|U - U_j\|_{\tilde{\mathcal{X}}} \leq \frac{\tilde{B}_{\max}}{\beta_{j,\ell}} \inf_{v_j \in S_j} \|U - v_j\|_{\tilde{\mathcal{X}}}, \quad \mathbb{P}\text{-a.s.},$$

with continuity constant

$$\tilde{B}_{\max}(\omega) := \sup_{v \in \tilde{\mathcal{X}} \setminus \{0\}} \sup_{w \in \tilde{\mathcal{Y}} \setminus \{0\}} \frac{|\tilde{b}_\omega(v,w)|}{\|v\|_{\tilde{\mathcal{X}}} \|w\|_{\tilde{\mathcal{Y}}}}, \quad \omega \in \Omega \text{ a.s.},$$

and *discrete* inf-sup constant

$$\beta_{j,\ell}(\omega) := \inf_{v_j \in S_j \setminus \{0\}} \sup_{q_\ell \in Q_\ell \setminus \{0\}} \frac{|\tilde{b}_\omega(v_j, q_\ell)|}{\|v_j\|_{\tilde{\mathcal{X}}} \|q_\ell\|_{\tilde{\mathcal{Y}}}}, \quad \omega \in \Omega \text{ a.s..}$$

We can see that the quasi-optimality depends on the discrete inf-sup constant, respectively random variable, which does *not* coincide with its continuous counterpart. To this end, if we want to follow the lines of the previous sections, we have to specify the random variable $\beta_{j,\ell}(\omega)$ and in particular have to track the dependency on the spatial differential operator $A(t,\omega)$ carefully. We have already shown that $B_{\max} \leq \sqrt{2}\max\{1, A_{\max}\}$ $\mathbb{P}$-a.s., cf. (3.3.7).

In section 4.2 we have applied a general stability result to the second space-time weak formulation of parabolic PDEs, so that we can derive a result similar to Theorem 5.2.8 for fully discrete Petrov-Galerkin solutions. Theorem 4.2.10 yields explicit constants, which turn into random variables. An overview of all appearing constants is given by Table 4.3.27. It is important to mention that the number of extra layers $L$ needs to satisfy estimate (4.2.11) in order to guarantee uniform stability. Different from the subspace dependent constant $c_S$ in (5.2.1), $L$ does not only depend on the subspaces, but also on properties of the operator, and therefore on the random parameter $\omega$. Nevertheless, the ansatz functions itself can stay $\omega$-independent but only its level of refinement depend on the random parameter. In analogy to Corollary 4.3.28 we can derive the following.

**Corollary 5.3.2.** *Let the assumptions from Corollary 4.3.28 hold with the same bounds* (5.1.5) *and* (5.1.6) *also for the shifted problem and define* $L \in \mathbb{N}$ *with*

$$L(\omega) > \frac{C + \log_\nu(\frac{\max\{A_{\max}^2(\omega),1\}}{A_{\min}(\omega)})}{d_x + d_t}, \quad \omega \in \Omega \text{ a.e.},$$

*where* $C := \log_\nu(2C_{CS}C_{J,Y}C_{B,X'})$ *is a constant independent of* $\omega$ *with notations from Corollary 4.3.28, respectively Table 4.3.27. Then the discrete inf-sup condition*

$$\inf_{v_j \in S_j \setminus \{0\}} \sup_{q_\ell \in Q_\ell \setminus \{0\}} \frac{|\langle \tilde{B}_\omega v_j, q_\ell \rangle|}{\|v_j\|_X \|q_\ell\|_Y} \geq \beta(\omega) > 0, \qquad \mathbb{P}\text{-a.s.}$$

*is satisfied with a random variable* $\beta(\omega)$ *independent of* $j$ *and* $\ell$, $\omega \in \Omega$ *a.s.*

*Proof.* We have to specify the constants given in Corollary 4.3.28 and Table 4.3.27 in view of the dependency on $\omega$. According to Table 4.3.27, we obtain that $C_{CS}$ is independent of the particular equation and only depends on the choice of subspaces. Moreover, we see that $C_{J,X'} = \tilde{B}_{\max}(\omega)\tilde{B}_{\min}^{-1}(\omega)C_{J,Y}C_{B,X'}$ by assumption on the bounds of $\tilde{B}$, with (non-discrete) inf-sup constant

$$\tilde{B}_{\min} := \inf_{v \in \tilde{\mathcal{X}} \setminus \{0\}} \sup_{w \in \tilde{\mathcal{Y}} \setminus \{0\}} \frac{|\tilde{b}_\omega(v,w)|}{\|v\|_{\tilde{\mathcal{X}}} \|w\|_{\tilde{\mathcal{Y}}}}, \quad \mathbb{P}\text{-a.s.},$$

where the constants $C_{J,Y}$ and $C_{B,X'}$ stem from Jackson and Bernstein estimates, respectively, and are thus independent of $\omega \in \Omega$. The choice of $L(\omega)$ now follows from

(5.1.5) and (5.1.6), since

$$\log_\nu(2C_{CS}C_{J,Y}C_{B,X'}\tilde{B}_{\max}(\omega)\tilde{B}^{-1}_{\min}(\omega)) \leq D + \log_\nu\Big(\frac{\max\{A_{\max}(\omega),1\}^2}{A_{\min}(\omega)}\Big)$$

for $\omega \in \Omega$ a.s., where we assumed the same bounds also for the shifted problem by assumption, meaning that $C_+ = \tilde{B}^{-1}_{\min}$. The proof of the discrete inf-sup condition now follows directly from Corollary 4.3.28. □

Note that the assumption to have the same bounds (5.1.5) and (5.1.6) also for the shifted spaces from Corollary 4.3.28 is rather natural. Indeed, it even holds with improved bounds for time-independent spatial differential operators $A(\omega)$ according to Corollary 3.3.10. One could, of course, also specify the discrete inf-sup bound $\beta(\omega)$ according to Table 4.3.27, but, although it would be straightforward, we decided to keep it on this technical level since we are not using this approach in the following.

The fact that the number of extra layers depends on the explicit realization, that is, on $\omega$, is clearly a drawback, but this is the price one has to pay for the generality of the stability result of Theorem 4.2.10. To this end, one could try to construct $\omega$-independent stable subspaces which are more tailored to the particular situation.

To our knowledge there is no construction of unconditionally stable subspaces for this particular second space-time weak formulation. But a very detailed treatment of the construction of stable subspaces for parabolic problems in the *first* space-time weak formulation for *self-adjoint* spatial differential operators is given in [And13] and [And12]. It will turn out that, indeed, one can construct stable subspaces along these lines with some modifications and a careful tracking of the hidden constants.

Therefore, we consider the minimal residual Petrov-Galerkin solution

$$U_j := \operatorname*{arg\,min}_{v_j \in S_j} \sup_{q_\ell \in Q_\ell \setminus \{0\}} \frac{|\langle B_\omega v_j - \mathcal{F}_\omega, q_\ell\rangle|}{\|q_\ell\|_{\mathcal{Y}}}, \quad \mathbb{P}\text{-a.s.}, \tag{5.3.3}$$

with respect to the first formulation (3.1.6) and pathwise operator (3.1.11). That means, we consider in the following the bilinear form $b(\cdot,\cdot)$, the right hand side $\mathcal{F}$ and the spaces $\mathcal{X}$, $\mathcal{Y}$ from section 3.1 equipped with additional random parameter $\omega$ and $S_j \subset \mathcal{X}$ and $Q_\ell \subset \mathcal{Y}$. To be more precise, we consider the problem

$$U(\cdot,\omega) \in \mathcal{X}: \quad b_\omega(U(\cdot,\omega), v) = \mathcal{F}_\omega(v) \qquad \text{for all } v \in \mathcal{Y},\ \omega \in \Omega \text{ a.s.}, \tag{5.3.4}$$

where the (parameter dependent) bilinear form $b_\omega(\cdot,\cdot)\colon \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ is defined as

$$b_\omega(u,v) := \int_I ({}_{V'}\langle \dot{u}(t), v(t)\rangle_V + {}_{V'}\langle A(t,\omega)u(t), v(t)\rangle_V)\ \mathrm{d}t + (u(0), v_2)_H, \quad \mathbb{P}\text{-a.s.},$$

and the (parameter dependent) right hand side $\mathcal{F}_\omega(\cdot)\colon \mathcal{Y} \to \mathbb{R}$ is given by

$$\mathcal{F}_\omega(v) := \int_I {}_{V'}\langle g(t,\omega), v(t)\rangle_V\ \mathrm{d}t + (U_0(\omega), v_2)_H, \quad \mathbb{P}\text{-a.s.},$$

where $v = (v_1, v_2)$, in the same way as done in section 3.3. In the subsequent, we will make use of the ideas in [And13] to construct stable bases for the first space-time weak formulation *independent* of $\omega \in \Omega$. The parameter independent construction is inspired by [And13, Th. 4.1], modified and adapted to our present situation with explicitly tracked parameter dependency and improved bounds with respect to the natural norms on $\mathcal{X}$ and $\mathcal{Y}$.

**Theorem 5.3.5.** *Let $A(t)$ be self adjoint for $t \in I$ a.e. Then for any pair of closed subspaces $S_j \subset \mathcal{X}$ and $Q_\ell := Q_\ell^1 \times Q_\ell^2 \subset \mathcal{Y} := \mathcal{Y}_1 \times \mathcal{Y}_2 := L_2(I;V) \times H$, such that*

$$S_j \times \{v_j(0) : v_j \in S_j\} \subset Q_\ell, \tag{5.3.6}$$

$$\inf_{\dot{v}_j \in \partial_t S_j \setminus \{0\}} \sup_{q_\ell^1 \in Q_\ell^1 \setminus \{0\}} \frac{{}_{\mathcal{Y}_1'}\langle \dot{v}_j, q_\ell^1 \rangle_{\mathcal{Y}_1}}{\|\dot{v}_j\|_{\mathcal{Y}_1'} \|q_\ell^1\|_{\mathcal{Y}_1}} \geq \kappa > 0 \text{ for some } \kappa > 0, \tag{5.3.7}$$

*the discrete inf-sup constant is bounded by*

$$\inf_{v_j \in S_j \setminus \{0\}} \sup_{q_\ell \in Q_\ell \setminus \{0\}} \frac{{}_{\mathcal{Y}'}\langle B_\omega v_j, q_\ell \rangle_{\mathcal{Y}}}{\|v_j\|_{\mathcal{X}} \|q_\ell\|_{\mathcal{Y}}} \geq \min \left\{ A_{\min}^{\frac{1}{2}} \kappa, A_{\max}^{-\frac{1}{2}} \kappa, \left(\frac{A_{\min}}{A_{\max}}\right)^{\frac{1}{2}} \kappa, A_{\min}^{\frac{1}{2}}, A_{\min} \right\},$$

$\mathbb{P}$*-a.s. We have defined*

$$\partial_t S_j := \{\partial_t v_j : v_j \in S_j\} \subset L_2(I;V') \tag{5.3.8}$$

*as the space of time derivatives of $S_j$.*

*Proof.* We consider $\omega \in \Omega$ to be fixed. Since $A(t,\omega)$ is self-adjoint, bounded and coercive for $t \in I$ a.e., the bilinear mapping

$$(w, \tilde{w})_A := \int_0^T {}_{V'}\langle A(t,\omega) w_1(t), \tilde{w}_1(t) \rangle_V \, \mathrm{d}t + (w_2, \tilde{w}_2)_H,$$

for $w := (w_1, w_2),\ \tilde{w} := (\tilde{w}_1, \tilde{w}_2) \in \mathcal{Y}$, defines a scalar product on $\mathcal{Y}$ and also

$$(w_1, \tilde{w}_1)_{A,1} := \int_0^T {}_{V'}\langle A(t,\omega) w_1(t), \tilde{y}_1(t) \rangle_V \, \mathrm{d}t, \quad \text{for } w_1, \tilde{w}_1 \in \mathcal{Y}_1$$

is a scalar product on $\mathcal{Y}_1$. Notice that these norm are very similar to the $\omega$-dependent norm introduced in section 5.1 for the second formulation. Define

$$\tilde{\kappa}_A := \inf_{\dot{v}_j \in \partial_t S_j \setminus \{0\}} \sup_{q_\ell \in Q_\ell \setminus \{0\}} \frac{{}_{\mathcal{Y}_1'}\langle \dot{v}_j, q_\ell^1 \rangle_{\mathcal{Y}_1}}{\|\dot{v}_j\|_{\mathcal{Y}_1'} \|q_\ell\|_A}, \tag{5.3.9}$$

for $q_\ell := (q_\ell^1, q_\ell^2) \in Q_\ell$. Additionally, we introduce the embedding $I \colon \mathcal{X} \to \mathcal{Y}$ as $v \mapsto Iv := (v, v(0))$ as well as the unique operator $S_\omega \in \mathcal{L}(\mathcal{X}, Q_\ell)$ defined by

$$(S_\omega v, q_\ell)_A := {}_{\mathcal{Y}'}\langle B_\omega v, q_\ell \rangle_{\mathcal{Y}} \qquad \text{for all } (v, q_\ell) \in \mathcal{X} \times Q_\ell.$$

Notice that although the operator has its range in $\mathcal{Y}$, it also depends on its codomain $Q_\ell$. The existence and uniqueness of such an operator $S$ is ensured by the Riesz representation Theorem 2.1.4 on the Hilbert space $Q_\ell$ endowed with the modified scalar product $(\cdot,\cdot)_A$. Indeed, having $B_\omega v \in \mathcal{Y}' \subset Q_\ell'$ for $v \in \mathcal{X}$, there is a unique isometry $S_\omega := R_A^{-1} B_\omega \colon \mathcal{X} \to Q_\ell$ with Riesz mapping $R_A \colon Q_\ell \to Q_\ell'$ with respect to the modified norm $(\cdot,\cdot)_A$.

Let $v_j \in S_j \setminus \{0\}$ be arbitrary. By assumption we have $Iv_j \subset Q_\ell$, so that

$$\|S_\omega v_j - Iv_j\|_A = \sup_{q_\ell \in Q_\ell \setminus \{0\}} \frac{(S_\omega v_j - Iv_j, q_\ell)_A}{\|q_\ell\|_A},$$

by using Riesz representation Theorem on $Q_\ell$ again. Since

$$\begin{aligned}
(S_\omega v_j - Iv_j, q_\ell)_A &= {}_{\mathcal{Y}'}\langle B_\omega v_j, q_\ell\rangle_{\mathcal{Y}} - (Iv_j, q_\ell)_A \\
&= \int_0^T \left( {}_{V'}\langle \dot v_j(t), q_\ell^1(t)\rangle_V + {}_{V'}\langle A(t,\omega) v_j(t), q_\ell^1(t)\rangle_V \right) \mathrm{d}t \\
&\quad + (v_j(0), q_\ell^2)_H - \int_0^T {}_{V'}\langle A(t,\omega) v_j(t), q_\ell^1(t)\rangle_V \,\mathrm{d}t \\
&\quad - (v_j(0), q_\ell^2)_H \\
&= {}_{\mathcal{Y}_1'}\langle \dot v_j, q_\ell^1\rangle_{\mathcal{Y}_1},
\end{aligned}$$

we obtain

$$\|Sv_j - Iv_j\|_A = \sup_{q_\ell \in Q_\ell \setminus \{0\}} \frac{{}_{\mathcal{Y}_1'}\langle \dot v_j, q_\ell^1\rangle_{\mathcal{Y}_1}}{\|q_\ell\|_A} \geq \tilde\kappa_A \|\dot v_j\|_{\mathcal{Y}_1'},$$

by definition (5.3.9). Moreover, we have

$$\begin{aligned}
\|S_\omega v_j\|_A^2 &= (S_\omega v_j - Iv_j, S_\omega v_j - Iv_j)_A + (S_\omega v_j - Iv_j, Iv_j)_A + (Iv_j, S_\omega v_j)_A \\
&= \|S_\omega v_j - Iv_j\|_A^2 + 2(S_\omega v_j, Iv_j)_A - \|Iv_j\|_A^2 \\
&\geq \|S_\omega v_j - Iv_j\|_A^2 + \|v_j\|_{A,1}^2,
\end{aligned}$$

by using that the operator $A(t)$ is assumed to be self-adjoint and

$$\begin{aligned}
2(S_\omega v_j, Iv_j)_A &= 2\ {}_{\mathcal{Y}'}\langle B_\omega v_j, Iv_j\rangle_{\mathcal{Y}} \\
&= 2\int_0^T ({}_{V'}\langle \dot v_j(t), v_j(t)\rangle_V + {}_{V'}\langle A(t,\omega) v_j(t), v_j(t)\rangle_V) \,\mathrm{d}t \\
&\quad + 2\|v_j(0)\|_H^2 \\
&= \|v_j\|_{A,1}^2 + \|Iv_j\|_A^2 + \|v_j(T)\|_H^2.
\end{aligned}$$

In the last equality, we exploited that $2\ {}_{V'}\langle \dot v_j(t), v_j(t)\rangle_V = \frac{\mathrm{d}}{\mathrm{d}t}\|v_j(t)\|_H^2$. Combining the equations and estimates above yields

$$\begin{aligned}
\|S_\omega v_j\|_A^2 &\geq \tilde\kappa_A^2 \|\dot v_j\|_{\mathcal{Y}_1'}^2 + \|v_j\|_{A,1}^2 \\
&\geq \min\{\tilde\kappa_A^2, A_{\min}(\omega)\} \|v_j\|_{\mathcal{X}}^2,
\end{aligned}$$

since

$$\|v_j\|_{A,1}^2 \geq A_{\min}(\omega)\|v_j\|_{\mathcal{Y}_1}^2,$$

due to the coercivity of $A(t,\omega)$. Similar we can derive

$$\begin{aligned}\sup_{q_\ell \in Q_\ell \setminus \{0\}} \frac{{}_{\mathcal{Y}'}\langle B_\omega v_j, q_\ell\rangle_{\mathcal{Y}}}{\|q_\ell\|_{\mathcal{Y}}} &\geq \min\{1, \sqrt{A_{\min}(\omega)}\} \sup_{q_\ell \in Q_\ell \setminus \{0\}} \frac{{}_{\mathcal{Y}'}\langle B_\omega v_j, q_\ell\rangle_{\mathcal{Y}}}{\|q_\ell\|_A} \\ &= \min\{1, \sqrt{A_{\min}(\omega)}\}\|S_\omega v_j\|_A\end{aligned}$$

and, by assumption

$$\begin{aligned}\tilde{\kappa}_A &\geq \min\{1, A_{\max}^{-\frac{1}{2}}(\omega)\} \inf_{\dot{v}_j \in \partial_t S_j \setminus \{0\}} \sup_{q_\ell \in Q_\ell \setminus \{0\}} \frac{{}_{\mathcal{Y}_1'}\langle \dot{v}_j, q_\ell^1\rangle_{\mathcal{Y}_1}}{\|\dot{v}_j\|_{\mathcal{Y}_1'}\|q_\ell\|_{\mathcal{Y}}} \\ &= \min\{1, A_{\max}^{-\frac{1}{2}}(\omega)\} \inf_{\dot{v}_j \in \partial_t S_j \setminus \{0\}} \sup_{q_\ell^1 \in Q_\ell^1 \setminus \{0\}} \frac{{}_{\mathcal{Y}_1'}\langle \dot{v}_j, q_\ell^1\rangle_{\mathcal{Y}_1}}{\|\dot{v}_j\|_{\mathcal{Y}_1'}\|q_\ell^1\|_{\mathcal{Y}_1}} \\ &\geq \min\{1, A_{\max}^{-\frac{1}{2}}(\omega)\}\ \kappa.\end{aligned}$$

Combing the previous estimates and sorting and simplifying yields

$$\sup_{q_\ell \in Q_\ell \setminus \{0\}} \frac{{}_{\mathcal{Y}'}\langle B_\omega v_j, q_\ell\rangle_{\mathcal{Y}}}{\|q_\ell\|_{\mathcal{Y}}} \geq \min\left\{A_{\min}^{\frac{1}{2}}\kappa, A_{\max}^{-\frac{1}{2}}\kappa, \left(\frac{A_{\min}}{A_{\max}}\right)^{\frac{1}{2}}\kappa, A_{\min}^{\frac{1}{2}}, A_{\min}\right\}\|v_j\|_{\mathcal{X}}.$$

This finishes the proof since $v_j \in S_j \setminus \{0\}$ was chosen arbitrary. □

It is important to note that the two assumption (5.3.6) and (5.3.7) are *independent* of the equation under consideration, that is, independent of $B_\omega$ or $A(t,\omega)$ and therefore in particular independent of $\omega \in \Omega$, but only depend on the particular discretization.

In order to obtain a uniformly bounded discrete inf-sup constant, the discrete spaces $S_j$ and $Q_\ell$ need to be constructed in a way that the two conditions (5.3.6) and (5.3.7) are fulfilled with a constant $\kappa$ which does not depend on the mesh parameter $j$ and $\ell$. A possible guideline for constructing suitable families of spaces is given by the following proposition from [And13, Prop. 4.2].

**Proposition 5.3.10.** *Let $S_m^t \subset H^1(I)$ and $Q_m^t \subset L_2(I)$, $m \in \mathbb{N}_0$, be sequences of closed, nested subspaces $S_m^t \subset S_{m+1}^t$ and $Q_m^t \subset Q_{m+1}^t$ such that*

$$\tau := \inf_{m \in \mathbb{N}_0} \inf_{\dot{v}_m \in \partial_t S_m^t \setminus \{0\}} \sup_{q_m \in Q_m^t \setminus \{0\}} \frac{(\dot{v}_m, q_m)_{L_2(I)}}{\|\dot{v}_m\|_{L_2(I)}\|q_m\|_{L_2(I)}} > 0. \tag{5.3.11}$$

*Further, let $S_i^x \subset V'$ and $Q_i^x \subset V$, $i \in \mathbb{N}_0$, be sequences of closed, nested subspaces $S_i^x \subset S_{i+1}^x$ and $Q_i^x \subset Q_{i+1}^x$ such that*

$$\eta := \inf_{i \in \mathbb{N}_0} \inf_{v_i \in S_i^x \setminus \{0\}} \sup_{q_i \in Q_i^x \setminus \{0\}} \frac{{}_{V'}\langle v_i, q_i\rangle_V}{\|v_i\|_{V'}\|q_i\|_V} > 0. \tag{5.3.12}$$

*Let $L \in \mathbb{N}_0$ be fixed, and define*

$$S_L := \sum_{0\le m+i\le L} S_m^t \otimes S_i^x \quad \text{and} \quad Q_L := \sum_{0\le m+i\le L} Q_m^t \otimes Q_i^x \times Q_i^x.$$

*Then*

$$\inf_{\dot v_L \in \partial_t S_L \setminus \{0\}} \sup_{q_L^1 \in Q_L^1 \setminus \{0\}} \frac{{\mathcal{Y}_1'}\langle \dot v_L, q_L^1\rangle_{\mathcal{Y}_1}}{\|\dot v_L\|_{\mathcal{Y}_1'} \|q_L^1\|_{\mathcal{Y}_1}} \ge \tau\eta > 0,$$

*with the notation from Theorem 5.3.5. Moreover, if $S_m^t \subset Q_m^t$ and $S_i^x \subset Q_i^x$ then $S_L \times \{v_L(0) : x_L \in \mathcal{X}_L\} \subset Q_L$.*

As already mentioned above, we see that the particular construction of subspaces can be done independently of $\omega \in \Omega$. The previous proposition yields a guideline how to construct suitable subspaces, namely by constructing them in such a way the assumptions are met. The easiest choice for temporal subspaces is obviously $Q_m^t := S_m^t + \partial_t S_m^t$. But a more important example is given by the following construction based on spline spaces, see [And13, Prop. 6.1 and 6.3].

As temporal solution and test space choose the space of continuous and piecewise linear functions (splines) on $2^m$, respectively $2^{m+1}$, uniform subintervals of $I$, i.e.,

$$\begin{aligned} S_m^t &:= \{v \in \mathcal{C}^0(I) : v|_{[(2^{-k}T)i,(2^{-k}T)(i+1))} \in \Pi_2 \quad \text{for all } i = 0, \dots, 2^k - 1\}, \\ Q_m^t &:= S_{m+1}^t, \end{aligned}$$

where $\Pi_j$ denotes the space of polynomials of order $j \in \mathbb{N}$. That is, we set

$$S_m^t := SP_{m,2}, \quad Q_m^t := SP_{m+1,2},$$

according to notation (2.4.6). Notice that we choose a finer level for the discrete test space as also predicted in the general setting in section 4, so that one has to solve a least square problem. It was shown in [And13, Prop. 6.1] that this choice of temporal subspaces satisfies assumption (5.3.11). It is worth mentioning that the proof does not seem to be adaptable to arbitrary spline spaces in a straightforward fashion. There are also other examples as for instance polynomials (not splines), trigonometric polynomials and exponentials given in [And12, Example 5.2.16 and Sec. 7.3].

Concerning the spatial discretization, we choose $U_\ell := V_\ell$. This implies that (5.3.12) is met when the $H$-orthogonal projection onto $V_\ell$ is $V$-stable, cf. also [And13, Lemma 6.2]. These stability can be concluded by Bernstein and Jackson estimates and is known for many finite element, wavelet or general spline spaces, cf. Theorem 2.4.10 and 2.4.12. Similarly, we can argue for biorthogonal spaces and a corresponding biorthogonal projection, see also Proposition 4.3.7 and 4.3.21. Indeed, assumption (5.3.12) coincides with the reverse Cauchy-Schwarz inequality (4.2.6), so we refer to chapter 4 for details.

By probably reindexing the subspaces, Proposition 5.3.10 also holds true, e.g., for full tensor product spaces and for sparse-grid tensor product spaces.

Now we are in the position to formulate the main result concerning quasi-optimality along the lines of Theorem 5.2.8. To this end, recall again from Theorem 4.1.8 that we have quasi-optimality with respect to the deterministic spaces $S_j$ and $\mathcal{X}$. The previous Theorem 5.3.5 now implies a uniform quasi-optimality, independent of the mesh size indexed by $j$ and $\ell$. Combining all ingredients, we can formulate the following theorem.

**Theorem 5.3.13.** *Assume that the subspaces $S_j \subset \mathcal{X}$ and $Q_\ell \subset \mathcal{Y}$ are constructed such that the assumptions* (5.3.6) *and* (5.3.7) *are satisfied. Moreover, let $\alpha, \gamma \in [1, \infty]$ and $\beta \in \left[\frac{3}{2}, \infty\right]$ be parameters such that*

$$2\alpha\beta\gamma \geq 4\alpha\beta + 5\alpha\gamma + 2\beta\gamma \tag{5.3.14}$$

*and let the data $f$ and $U_0$, and the random variables $A_{\max}$ and $A_{\min}$ are such that:*

*(i)* $\mathcal{F}_\omega$ *belongs to* $L_\alpha(\Omega; \mathcal{Y}')$ *and* $U_0$ *to* $L_\alpha(\Omega; H)$

*(ii)* $A_{\max} \in L_\beta(\Omega)$,

*(iii)* $\frac{1}{A_{\min}} \in L_\gamma(\Omega)$.

*Then the error between the Petrov-Galerkin solution $U_j$* (5.3.3) *and the solution $U$ of problem* (5.3.4) *can be measured in $L_{\bar{p}}(\Omega; \mathcal{X})$ and is quasi optimal with*

$$\|U - U_j\|_{L_{\bar{p}}(\Omega;\mathcal{X})} \lesssim \inf_{v \in L_p(\Omega; S_j)} \|U - v\|_{L_p(\Omega;\mathcal{X})},$$

*with $\bar{p} = \frac{2\alpha\beta\gamma}{4\alpha\beta+5\alpha\gamma+2\beta\gamma}$ and $p := \frac{\alpha\beta\gamma}{\alpha\beta+\alpha\gamma+\beta\gamma}$ as in Table 5.1.22, and a constant that does not depend on the particular discretization.*

*Proof.* Notice first that condition (5.3.14) implies in particular that the solution $U$ to the Problem 5.1.1 in second space-time weak form belongs to $L_p(\Omega; \mathcal{X})$ according to Table 5.1.22, since

$$\alpha\beta\gamma \geq 2\alpha\beta + \tfrac{5}{2}\alpha\gamma + \beta\gamma > \alpha\beta + \alpha\gamma + \beta\gamma.$$

and thus $U \in L_{\bar{p}}(\Omega; \mathcal{X})$ as well. That is, we can measure the solution $U$ in $L_p(\Omega; \mathcal{X})$. Moreover, $A_{\max} \in L_\beta(\Omega)$ implies $A_{\max}^{\frac{3}{2}} \in L_{\beta'}(\Omega)$, with $\beta' := \frac{2}{3}\beta \geq 1$. We conclude from the choice of parameters that

$$\frac{1}{\beta'} + \frac{1}{\gamma} + \frac{1}{p} = \frac{1}{\alpha} + \frac{5}{2\beta} + \frac{2}{\gamma} = \frac{1}{\bar{p}},$$

so that we can apply a generalization of Hölder's inequality 2.1.7. By Theorem 5.3.5 and (3.1.13) combined with (4.1.9), we can estimate

$$\begin{aligned}\|U-U_j\|_{L_{\bar{p}}(\Omega;\mathcal{X})} &\leq \Bigg\| \Bigg( \max\{A_{\min}^{-\frac{1}{2}}\kappa, A_{\max}^{\frac{1}{2}}\kappa, \Big(\frac{A_{\max}}{A_{\min}}\Big)^{\frac{1}{2}}\kappa, A_{\min}^{-\frac{1}{2}}, A_{\min}^{-1}\} \\ &\qquad \times \sqrt{(2\max\{A_{\max}^2,1\}+\rho^2)}\Bigg) \inf_{v\in L_p(\Omega;S_j)} \|u-v\|_{\mathcal{X}} \Bigg\|_{L_{\bar{p}}(\Omega;\mathcal{X})} \\ &\leq \Bigg\| \Bigg( A_{\min}^{-\frac{1}{2}}\kappa + A_{\max}^{\frac{1}{2}}\kappa + \Big(\frac{A_{\max}}{A_{\min}}\Big)^{\frac{1}{2}}\kappa + A_{\min}^{-\frac{1}{2}} + A_{\min}^{-1}\Bigg) \\ &\qquad \times \Big(\sqrt{2}A_{\max} + (\sqrt{2}+\rho)\Big) \inf_{v\in L_p(\Omega;S_j)} \|U-v\|_{\mathcal{X}} \Bigg\|_{L_{\bar{p}}(\Omega;\mathcal{X})},\end{aligned}$$

since all appearing random variables and constants are positive almost surely.

Multiplying out each term, one gets a sum where the largest appearing exponent with respect to $A_{\max}$ is $\frac{3}{2}$ and with respect to $\frac{1}{A_{\min}}$ is 1. Applying a generalization of Hölder's inequality 2.1.7 similar to the proof of Theorem 5.2.8 as well as Minkowsi's inequality proves the claim since $\big\|A_{\max}^{\frac{3}{2}}\big\|_{L_{\beta'}(\Omega)} = \|A_{\max}\|_{L_\beta(\Omega)}^{\frac{3}{2}}$, with the definition of $\beta'$ above, and since the Lebesgue spaces are nested $L_q(\Omega)\subset L_{q'}(\Omega)$ for $1\leq q'\leq q\leq\infty$. In the same way, we can see that in particular $\|U_j\|_{L_{\bar{p}}(\Omega;\mathcal{X})}$ is finite, i.e., $U_j\in L_{\bar{p}}(\Omega;\mathcal{X})$ due to Theorem 5.3.5 and Corollary 2.3.10. □

We could also calculate an explicit bound for $\|U-U_j\|_{L_{\bar{p}}(\Omega;\mathcal{X})}$, but the explicit value would be rather long. It would follow directly by writing out each term in the proof explicitly.

We conclude the section with a pendent of Theorem 5.3.5 and Theorem 5.3.13 for a homogenized version according to section 3.2. One can basically treat the homogenization as a particular case of the first formulation. But one does not need to deal with the Cartesian product in the test spaces, so it does not apply directly to the homogenization. We define $b_{0,\omega}(\cdot,\cdot)$ for $\omega\in\Omega$ a.s. as the $\omega$ dependent pendant of (3.2.4) in the course of (5.1.3). Therefore, we give the following two corollaries 5.3.15 and 5.3.16 following the lines of Theorem 5.3.5 and Theorem 5.3.13.

**Corollary 5.3.15.** *Let $A(t)$ be self adjoint for $t\in I$ a.e. Then for any pair of closed subspaces $S_j\subset\mathcal{X}_0$ and $Q_\ell\subset\mathcal{Y}_0$, such that*

$$S_j\subset Q_\ell,$$

$$\inf_{\dot{v}_j\in\partial_t S_j\setminus\{0\}}\sup_{q_\ell\in Q_\ell\setminus\{0\}}\frac{\mathcal{Y}_0'\langle\dot{v}_j,q_\ell\rangle_{\mathcal{Y}_0}}{\|\dot{v}_j\|_{\mathcal{Y}_0'}\|q_\ell\|_{\mathcal{Y}_0}}\geq\kappa>0 \text{ for some } \kappa>0,$$

*the discrete inf-sup constant is bounded by*

$$\inf_{v_j\in S_j\setminus\{0\}}\sup_{q_\ell\in Q_\ell\setminus\{0\}}\frac{b_{0,\omega}(v_j,q_\ell)}{\|v_j\|_{\mathcal{X}_0}\|q_\ell\|_{\mathcal{Y}_0}}\geq\min\left\{\Big(\frac{A_{\min}(\omega)}{A_{\max}(\omega)}\Big)^{\frac{1}{2}}\kappa, A_{\min}(\omega)\right\},\quad \mathbb{P}\text{-a.s.}.$$

*We have defined*

$$\partial_t S_j := \{\partial_t v_j : v_j \in S_j\} \subset L_2(I; V')$$

*as the space of time derivatives of $S_j$ and $\mathcal{X}_0$ and $\mathcal{Y}_0$ are defined as in section 3.2.*

*Proof.* The claim follows along the lines of the proof of Theorem 5.3.5. The inner product $(\cdot,\cdot)_{A,1}$ and $(\cdot,\cdot)_A$ coincide, $Q_\ell^1$ becomes $Q_\ell$ = and $I$ is exchanged simply by the identity id. With this resetting and exploiting zero initial condition, the proof follows as in Theorem 5.3.5, where

$$\sup_{q_\ell \in Q_\ell \setminus \{0\}} \frac{{}_{\mathcal{Y}'}\langle B_\omega v_j, q_\ell\rangle_{\mathcal{Y}}}{\|q_\ell\|_{\mathcal{Y}}} \geq \sqrt{A_{\min}(\omega)}\|S_\omega v_j\|_A, \quad \mathbb{P}\text{-a.s.},$$

as well as

$$\tilde{\kappa}_A \geq A_{\max}^{-\frac{1}{2}}(\omega)\ \kappa, \quad \mathbb{P}\text{-a.s.},$$

with denotations from the proof of Theorem 5.3.5, since $(\cdot,\cdot)_A$ equals $(\cdot,\cdot)_{A,1}$. □

One can see that considering a homogenization simplifies the result, but improves the estimate only slightly. Proposition 5.3.10 applies to the homogenization in an obvious fashion as well. In the same regard, we can prove quasi-optimality in certain $L_p(\Omega;\mathcal{X}_0)$ spaces with exactly the same requirements. We state the result in the following corollary for sake of completeness.

**Corollary 5.3.16.** *Assume that the subspaces $S_j \subset \mathcal{X}_0$ and $Q_\ell \subset \mathcal{Y}_0$ are constructed according to the assumptions of Corollary 5.3.15. Moreover, let $\alpha, \gamma \in [1,\infty]$ and $\beta \in \left[\frac{3}{2}, \infty\right]$ be parameters such that*

$$2\alpha\beta\gamma \geq 4\alpha\beta + 5\alpha\gamma + 2\beta\gamma$$

*and let the data $f$ and $U_0$, and the random variables $A_{\max}$ and $A_{\min}$ are such that:*

*(i) $\mathcal{F}_\omega$ belongs to $L_\alpha(\Omega;\mathcal{Y}_0')$ and $U_0$ to $L_\alpha(\Omega;H)$*

*(ii) $A_{\max} \in L_\beta(\Omega)$,*

*(iii) $\frac{1}{A_{\min}} \in L_\gamma(\Omega)$.*

*Then the error between the solution $U$ of the homogenization of problem (5.3.4) and its Petrov-Galerkin solution $U_j$ can be measured in $L_{\bar{p}}(\Omega;\mathcal{X}_0)$ and is quasi optimal with*

$$\|U - U_j\|_{L_{\bar{p}}(\Omega;\mathcal{X}_0)} \lesssim \inf_{v \in L_p(\Omega;S_j)} \|U - v\|_{L_p(\Omega;\mathcal{X}_0)},$$

*with $\bar{p} = \frac{2\alpha\beta\gamma}{4\alpha\beta+5\alpha\gamma+2\beta\gamma}$ and $p := \frac{\alpha\beta\gamma}{\alpha\beta+\alpha\gamma+\beta\gamma}$ as in Table 5.1.23, and a constant that does not depend on the particular discretization.*

*Proof.* The proof follows as the proof of its pendant Theorem 5.3.13, since the largest appearing exponents for $\frac{1}{A_{\min}}$ and $A_{\max}$ in the inf-sup estimate of Theorem 5.3.5 are the same as in Corollary 5.3.15 as well as the largest exponents for $A_{\max}$ in the continuity constant for the first formulationd and its homogenization according to Table 3.3.17. □

# 6 Numerical Results and Examples

After presenting the theoretical results in the previous chapters, the aim of this chapter is to illustrate them by some numerical results. In order to do so, we focus first on the stability of deterministic parabolic problems presented in chapter 4, respectively section 4.3 to be precise. Afterwards, we discuss the numerical behavior of solutions of parabolic random PDEs and their Petrov-Galerkin approximation with regard to chapter 5. Finally, further results are shown which are beyond the scope of the theoretical part, but demonstrate the functionality of the developed Matlab package for B-spline Petrov-Galerkin methods.

## 6.1 Stability Examples

We start with considering numerical examples concerning our general stability results of Petrov-Galerkin approaches discussed in chapter 4. The aim is to illustrate the theoretically proven stability of Petrov-Galerkin discretizations of space-time weak formulated parabolic PDEs from section 4.3, which essentially means to check for a uniformly bounded discrete inf-sup constant. We will analyze, in particular, how many extra layers $L$ for the test spaces are practically necessary in order to obtain stability, since the theoretical value is hard to estimate, see Table 4.3.27. It was shown in [And13] that already one extra layer in time is sufficient for the first formulation (cf. section 3.1), when choosing continuous and piecewise linear polynomials for time discretization. Although we considered the second formulation in section 4.3 and do not restrict ourselves to the ansatz functions in [And13], we would expect to observe a similar behavior for spline ansatz functions. It is worth mentioning again that our stability analysis is designed for very general operator equations and worked out for the second formulation of full space-time weak formulations of parabolic PDEs as a model example, which is of particular interest in this thesis. We will use the notation from chapter 4 for consistency. The examples in this section with *wavelet discretization* are taken from my former work [Mol13b] and the other examples using *B-spline discretization* are calculated with the Matlab program implemented in connection with this thesis.

**Wavelet discretization**
Before we give first numerical results, we need to consider some algebraic properties in order to work out how to calculate the discrete inf-sup constant (4.1.6) and the (discrete) continuity constant (4.1.7) in a wavelet setting. For details on wavelets for PDEs we refer to [Urb09] and the references therein.

We choose a hierachical Riesz basis $\Psi^{\tilde{\mathcal{X}}} := \{\psi_\lambda^{\tilde{\mathcal{X}}} : \lambda \in \nabla_{\tilde{\mathcal{X}}}\}$ of $\tilde{\mathcal{X}}$ with infinite index set $\nabla_{\tilde{\mathcal{X}}} := \nabla_0^{\tilde{\mathcal{X}}} \cup \nabla_1^{\tilde{\mathcal{X}}} \cup \ldots$, where we define $\nabla_{(j)}^{\tilde{\mathcal{X}}} := \nabla_0^{\tilde{\mathcal{X}}} \cup \cdots \cup \nabla_j^{\tilde{\mathcal{X}}}$. That means

that each element in $\tilde{\mathcal{X}}$ has a unique expansion in terms of $\Psi^{\tilde{\mathcal{X}}}$ and that there exist Riesz-constants $0 < r_{\tilde{\mathcal{X}}} \leq R_{\tilde{\mathcal{X}}} < \infty$ such that

$$r_{\tilde{\mathcal{X}}} \|\mathbf{v}\|_{\ell_2(\nabla_{\tilde{\mathcal{X}}})} \leq \|v\|_{\tilde{\mathcal{X}}} \leq R_{\tilde{\mathcal{X}}} \|\mathbf{v}\|_{\ell_2(\nabla_{\tilde{\mathcal{X}}})}, \tag{6.1.1}$$

for each function $v = \sum_{\lambda \in \nabla_{\tilde{\mathcal{X}}}} v_\lambda \psi_\lambda^{\tilde{\mathcal{X}}}$ and $\mathbf{v} := \{v_\lambda\}_{\lambda \in \nabla_{\tilde{\mathcal{X}}}}$. Analogously, we choose a Riesz basis $\Psi^{\tilde{\mathcal{Y}}} := \{\psi_\lambda^{\tilde{\mathcal{Y}}} : \lambda \in \nabla_{\tilde{\mathcal{Y}}}\}$ for $\tilde{\mathcal{Y}}$ with Riesz-constants $0 < r_{\tilde{\mathcal{Y}}} \leq R_{\tilde{\mathcal{Y}}} < \infty$. Moreover, let the discrete spaces $S_j := \operatorname{span} \Psi_j^{\tilde{\mathcal{X}}}$ with $\Psi_j^{\tilde{\mathcal{X}}} := \{\psi_\lambda^{\tilde{\mathcal{X}}} : \lambda \in \nabla_{(j)}^{\tilde{\mathcal{X}}}\} \subset \Psi^{\tilde{\mathcal{X}}}$. Analogously, we choose $Q_\ell := \operatorname{span} \Psi_\ell^{\tilde{\mathcal{Y}}}$ with $\Psi_\ell^{\tilde{\mathcal{Y}}} := \{\psi_\lambda^{\tilde{\mathcal{Y}}} : \lambda \in \nabla_{(\ell)}^{\tilde{\mathcal{Y}}}\} \subset \Psi^{\tilde{\mathcal{Y}}}$ for the discrete test space. It is well known that the Riesz basis property (6.1.1) implies the existence of another $L_2$-stable biorthogonal Riesz basis according to (4.3.4) respectively (4.3.6), see [Dah94, §3].

Using these bases, we end up with the system matrix

$$\mathbf{B}_{j,\ell} := \{\langle B\psi_\mu^{\tilde{\mathcal{X}}}, \psi_\lambda^{\tilde{\mathcal{Y}}} \rangle\}_{\lambda \in \nabla_{(\ell)}^{\tilde{\mathcal{Y}}}, \mu \in \nabla_{(j)}^{\tilde{\mathcal{X}}}}. \tag{6.1.2}$$

Due to the Riesz stability (6.1.1) and the definition of the system matrix (6.1.2), we obtain the equivalence

$$\min_{\mathbf{v}_j \in \ell_2(\nabla_{(j)}^{\tilde{\mathcal{X}}}) \setminus \{0\}} \max_{\mathbf{q}_\ell \in \ell_2(\nabla_{(\ell)}^{\tilde{\mathcal{Y}}}) \setminus \{0\}} \frac{\mathbf{q}_\ell^\top \mathbf{B}_{j,\ell} \mathbf{v}_j}{\|\mathbf{q}_\ell\|_{\ell_2(\nabla_{(\ell)}^{\tilde{\mathcal{Y}}})} \|\mathbf{v}_j\|_{\ell_2(\nabla_{(j)}^{\tilde{\mathcal{X}}})}} \sim \inf_{v_j \in S_j \setminus \{0\}} \sup_{q_\ell \in Q_\ell \setminus \{0\}} \frac{\langle Bv_j, q_\ell \rangle}{\|q_\ell\|_{\tilde{\mathcal{Y}}} \|v_j\|_{\tilde{\mathcal{X}}}} \tag{6.1.3}$$

of the discrete inf-sup condition (4.1.6) with constants that do *not* dependent on the levels $j$ and $\ell$. That is, using Riesz bases, we can analyze the qualitative behavior of the discrete inf-sup condition (4.1.6) via the norm equivalence (6.1.3). Moreover, the equivalent representation has a simple algebraic interpretation. We rewrite the left hand side of (6.1.3) as

$$\begin{aligned} &\min_{\mathbf{v}_j \in \ell_2(\nabla_{(j)}^{\tilde{\mathcal{X}}}) \setminus \{0\}} \max_{\mathbf{q}_\ell \in \ell_2(\nabla_{(\ell)}^{\tilde{\mathcal{Y}}}) \setminus \{0\}} \frac{\mathbf{q}_\ell^\top \mathbf{B}_{j,\ell} \mathbf{v}_j}{\|\mathbf{q}_\ell\|_{\ell_2(\nabla_{(\ell)}^{\tilde{\mathcal{Y}}})} \|\mathbf{v}_j\|_{\ell_2(\nabla_{(j)}^{\tilde{\mathcal{X}}})}} \\ = &\min_{\mathbf{v}_j \in \ell_2(\nabla_j^{\tilde{\mathcal{X}}}) \setminus \{0\}} \frac{\|\mathbf{B}_{j,\ell} \mathbf{v}_j\|_{\ell_2(\nabla_{(\ell)}^{\tilde{\mathcal{Y}}})}}{\|\mathbf{v}_j\|_{\ell_2(\nabla_{(j)}^{\tilde{\mathcal{X}}})}} = \lambda_{\min}(\mathbf{B}_{j,\ell}^\top \mathbf{B}_{j,\ell})^{\frac{1}{2}} =: \sigma_{\min}(\mathbf{B}_{j,\ell}), \end{aligned} \tag{6.1.4}$$

where $\lambda_{\min}(\mathbf{B}_{j,\ell}^\top \mathbf{B}_{j,\ell})$ denotes the smallest eigenvalue of $\mathbf{B}_{j,\ell}^\top \mathbf{B}_{j,\ell}$, that is, $\sigma_{\min}(\mathbf{B}_{j,\ell})$ is the smallest singular value of $\mathbf{B}_{j,\ell}$.

For the subsequent examples we use B-spline wavelets of primal order $d = 2$ and dual order $\tilde{d} = 4$ (2-4 DKU wavelets) in space for $S_j^x$ as well as for $Q_\ell^x$ and B-spline wavelets with $d = 2$ and $\tilde{d} = 2$ (2-2 DKU wavelets) in time for $S_j^t$ as well as for $Q_\ell^t$. These are both piecewise linear and globally continuous (primal) functions, see [DKU99] for details.

The tensor space-time bases are thus given as $S_j := S_j^t \otimes S_j^x$ and $Q_\ell := Q_\ell^t \otimes Q_\ell^x$. The dimension of the spaces $S_j^t,\ S_j^x,\ Q_\ell^t,\ Q_\ell^x$ spanned by these wavelet bases is proportional to $2^j$ respectively $2^\ell$ and are known to meet the Jackson- and Bernstein estimates according to Table 4.3.12 with $m = 1$ for $\mu = 2$ except the condition for $\gamma_{Q^x}, d_{Q^x}$, see [DKU99, esp. Cor. 3.6 and Prop 3.7]. Nevertheless, one can show that $\gamma_{Q^x}, d_{Q^x} \geq \frac{3}{2}$ (see [DKU99, sec. 3.4]), which is truly larger than $m = 1$, such that the statement of Theorem 4.2.10 stays true by Corollary 4.3.28 with $d_x := \frac{1}{2}$. To this end, according to Theorem 4.2.10, we have uniformly bounded discrete inf-sup constants when choosing levels $\ell \geq j + L$ as in (4.2.12) with a constant $L$ defined in (4.2.11), provided that the operator is sufficiently regular according to (4.2.2). The regularity can be deduced from Corollary 3.3.10. For a detailed description of biorthogonal spline-wavelets we refer again to [DKU99]. After a suitable renormalization, it can be shown that the tensor product of these bases are Riesz bases for $\tilde{\mathcal{X}}$ and $\tilde{\mathcal{Y}}$, see [GO95, Prop.1 and Prop. 2] and [SS09, Ch. 6]. This in turn means that the smallest singular value of the stiffness matrix $\mathbf{B}_{j,\ell}$ is uniformly bounded from below for $\ell \geq j + L$ due to the equality (6.1.4) and the equivalence (6.1.3). Moreover, using these wavelet bases yields an optimally preconditioned sequence of matrices $\{\mathbf{B}_{j,\ell}^\top \mathbf{B}_{j,\ell}\}_{\ell \geq j+L}$, since the largest singular values are uniformly bounded as well. To this regard, recall that the continuity of an operator is inherited by the subspaces, so that the discrete operator is still bounded independently of the discretization, cf. (4.1.7).

The computations were performed using the adaptive wavelet C++ package written by Roland Pabel ([Pab15]) and Matlab (R2012a). The assembling of system matrices was implemented with the aid of the adaptive wavelet C++ package and imported to Matlab, where the eigenvalues were computed by the standard Matlab routine `eigs`. For the computations, a computer with four Intel(R) Xeon(TM) CPU 2.00GHz processors and 16GB Ram on a 64-Bit Linux system was employed.

First, we want to consider the easiest case

$$\begin{aligned} \frac{\mathrm{d}u(t)}{\mathrm{d}t} &= g(t), \qquad t \in (0,1] \\ u(0) &= 0, \end{aligned}$$

with given right hand side $g$. That means, we consider the following second variational problem.

**Example 6.1.5.** *Find a solution $u \in \tilde{\mathcal{X}}$ such that*

$$\langle \tilde{B}u, w\rangle = \tilde{\mathcal{F}}(w) \qquad \forall w \in \tilde{\mathcal{Y}}, \tag{6.1.6}$$

*with*

$$\langle \tilde{B}u, w\rangle := -\int_0^1 u(t)\dot{w}(t)\,\mathrm{d}t, \quad \tilde{\mathcal{F}}(w) := \int_0^1 g(t)w(t)\,\mathrm{d}t$$

*and the spaces $\tilde{\mathcal{X}} := L_2(0,1)$, $\tilde{\mathcal{Y}} := H^1_{0,\{1\}}(0,1)$ and given right hand side $g \in H^{-1}_{0,\{1\}}(0,1) := (H^1_{0,\{1\}}(0,1))'$.*

Due to (6.1.3) and (6.1.4) we have calculated the smallest singular values of $\mathbf{B}_{j,\ell}$ for different levels $j$ and $\ell$ as an indicator for the discrete inf-sup constants $\beta_{j,\ell}$. Since we are interested in the discrete inf-sup and continuity constant of the operator, the particular value of the right hand side does not need to be specified. Notice that, although the test spaces are restricted to zero final-time condition, the system of equation is not underdetermined for $j = \ell$, since we can choose ansatz functions with zero initial conditions for the solution space.

The results for Example 6.1.5 are presented in Figure 6.1.7, where we have plotted the slope of the smallest and largest singular values of $\mathbf{B}_{j,\ell}$ for a fixed level $\ell = 12$ and for the same levels $\ell = j$.

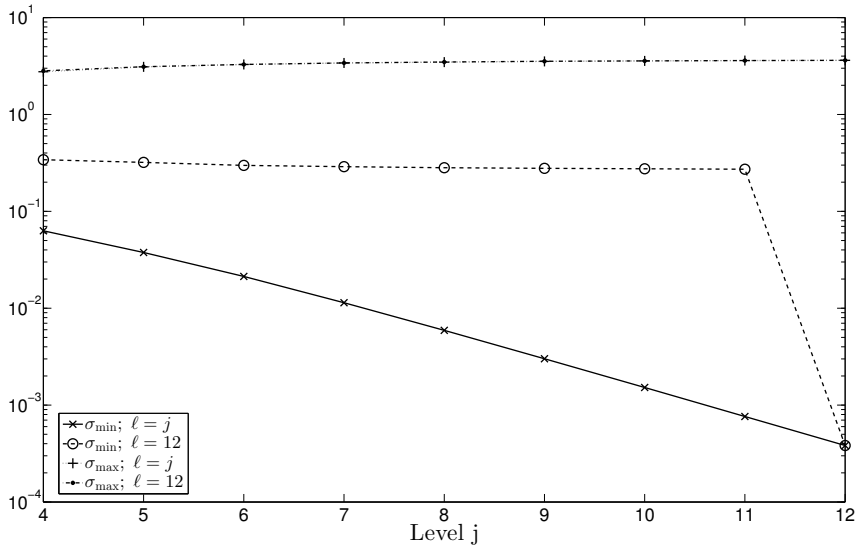



Figure 6.1.7: Plot of the smallest singular values of $\mathbf{B}_{j,\ell}$ according to Example 6.1.5 and of the largest singular values for same levels $j = \ell$ and for fixed level $\ell = 12$.

It shows that $\sigma_{\min}(\mathbf{B}_{j,j})$ and, therefore, the discrete inf-sup constants $\beta_{j,j}$ decrease with increasing level $j$. Moreover, we can observe, that the values of $\sigma_{\min}(\mathbf{B}_{j,\ell})$ and therefore the discrete inf-sup constants stay approximately constant, i.e., $\sigma_{\min}(\mathbf{B}_{j,\ell}) \sim 1$, if $\ell > j$ and is much smaller if $\ell = j$. This behavior confirms that a uniformly bounded sequence of discrete inf-sup constants can be guaranteed if we choose a higher resolution for the test space, and that otherwise, indeed, stability problems occur. Such a behavior was suggested by Theorem 4.2.10. Recall that the stability result of Theorem 4.2.10 says that the sequence of discrete inf-sup constants is uniformly bounded away from zero as long as $\ell \geq j + L$ with $L$ defined in (4.2.11). Moreover, we see that the largest singular values $\sigma_{\max}(\mathbf{B}_{j,\ell})$ are asymptotically constant, so that the sequence of system matrices $\{\mathbf{B}_{j,\ell}^{\top}\mathbf{B}_{j,\ell}\}_{\ell \geq j+L}$ is optimally preconditioned. Both values of $\sigma_{\max}(\mathbf{B}_{j,\ell})$ for $\ell = j$ and fixed $\ell = 12$ are almost equal, such that they are hard to distinguish in

the figure. Concerning the results in [And13], although given for the first formulation, it is not too surprising that already $L = 1$ yields a very satisfactory stable behavior of $\beta_{j,\ell}$ bounded away from zero. This observation is also underlined by Figure 6.1.8. One cannot transfer the stability results from [And13] straightforwardly to the second formulation since one has to enrich the test space having derivatives involved in the second form, whereas the derivatives in the first form appear in the solution space. That means, nevertheless, it was not clear that already one extra layer in time seems to be sufficient also in the second form. In Figure 6.1.8 we have plotted the slope of

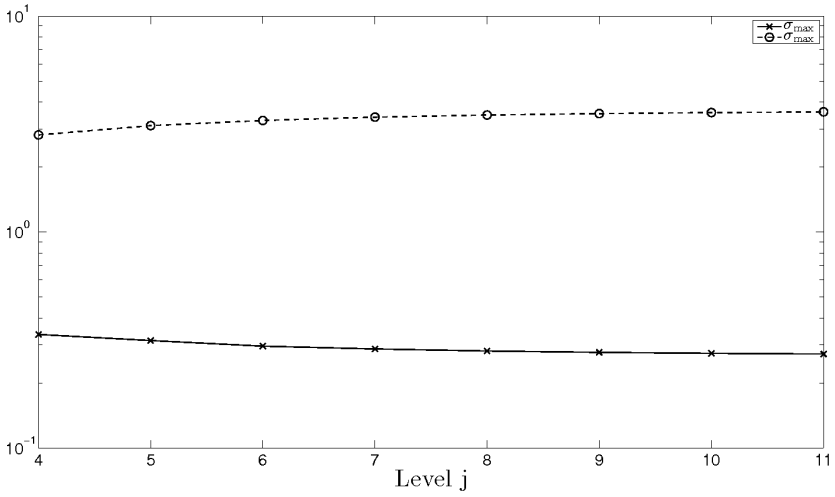



Figure 6.1.8: Plot of the smallest singular values of $\mathbf{B}_{j,\ell}$ as an indicator for the slope of the discrete inf-sup constants $\beta_{j,\ell}$ and of the largest singular values for $\ell = j + 1$ w.r.t to (6.1.6).

the smallest and largest singular values for a fixed number of extra layers $L = 1$, i.e., for increasing $j$ and $\ell$ with $\ell = j + 1$. One can observe that both, the smallest and largest singular values, stay asymptotically constant for increasing levels. For the sake of completeness, we also provide the error of the minimal Petrov-Galerkin solution in the natural norm $\|\cdot\|_{\tilde{\mathcal{X}}} = \|\cdot\|_{L_2(0,1)}$ on the solution space, see Figure 6.1.9. To this end, we choose a right hand side $f(t) := t^2$, with exact solution $u(t) = \frac{1}{3}t^3$.

We compute the approximate solution $u_j$ respectively its expansion coefficients $\mathbf{u}_j$ by using the `mldivide` operation of Matlab. Due to the Riesz basis property, the $\tilde{\mathcal{X}}$-norm of the error can be estimated by the $\ell_2$-norm of its expansion coefficients. As reference solution we took the best approximation of the exact solution on level $j = 12$. By Theorem 2.4.12 the best approximation rate for the used piecewise linear spline wavelets measured in $\|\cdot\|_{\tilde{\mathcal{X}}} = \|\cdot\|_{L_2(0,1)}$ is of order two. As expected due to quasi optimality (4.1.9), the optimal rate of convergence is attained.

We have deliberately chosen an ODE example, since one already observes stability problems even in this comparatively simple example. In principle, all theoretical predictions were already confirmed by the ODE example. Nevertheless, in view of "true"

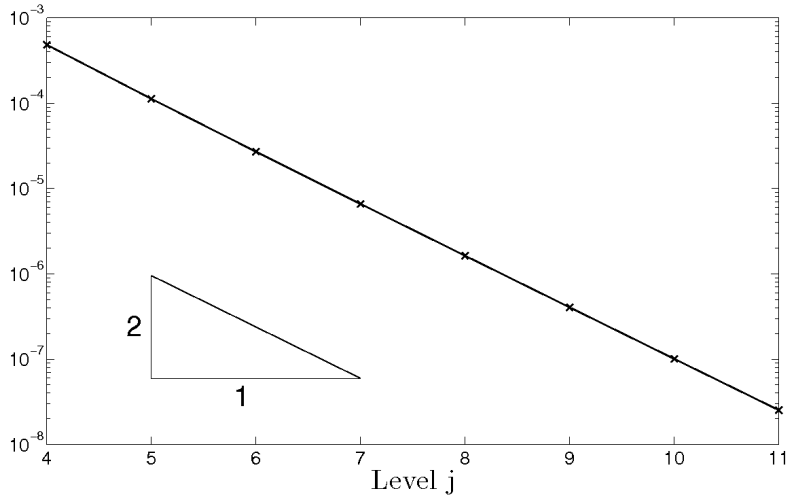



Figure 6.1.9: Estimated errors of the minimal residual Petrov-Galerkin solutions $u_j$ w.r.t. the norm $\|\mathbf{u}_j - \mathbf{u}_{\mathrm{ref}}\|_{\ell_2(\nabla_{\tilde{\mathcal{X}}})} \sim \|u_j - u_{\mathrm{ref}}\|_{\tilde{\mathcal{X}}}$ for (6.1.6) with $f(t) := t^2$ and $\ell = j + 1$.

space-time weak formulations with tensor products and intersections of them, we would like to underline our numerical results also with an example of a PDE, where truly multidimensional spaces are involved.

As a model example, we consider the parabolic evolution problem in strong form

$$\begin{aligned} \frac{\partial u(t,x)}{\partial t} - \frac{\partial^2 u(t,x)}{\partial x^2} + u(t,x) &= g(t,x), \qquad t \in (0,1],\ x \in (0,1) \\ \left.\frac{\partial u(t,x)}{\partial x}\right|_{x=0,1} &= 0, \qquad t \in [0,1] \\ u(0,x) &= 0, \qquad x \in (0,1), \end{aligned}$$

with given right hand side $g$. The second space-time weak form reads as follows.

**Example 6.1.10.** *Find a solution $u \in \tilde{\mathcal{X}}$, such that*

$$\langle \tilde{B}u, w\rangle = \tilde{\mathcal{F}}(w) \qquad \forall w \in \tilde{\mathcal{Y}}, \tag{6.1.11}$$

*with*

$$\langle \tilde{B}u, w\rangle := \int_0^1 \int_0^1 \left(-u\frac{\partial w}{\partial t} + \frac{\partial u}{\partial x}\frac{\partial w}{\partial x} + uw\right) \mathrm{d}x\ \mathrm{d}t,$$

$$\tilde{\mathcal{F}}(w) := \int_0^1 \int_0^1 g(t,x)w(t,x)\ \mathrm{d}x\ \mathrm{d}t,$$

*and the spaces $\tilde{\mathcal{X}} \cong L_2(0,1) \otimes H^1(0,1)$; $\tilde{\mathcal{Y}} \cong (L_2(0,T) \otimes H^1(0,1)) \cap (H^1_{0,\{1\}}(0,1) \otimes \dot{H}^{-1}(0,1))$. We have dropped the dependency on $t$ and $x$ of $u$ and $w$ in the definition of $\tilde{B}$ for a better readability.*

We are only interested in the discrete inf-sup constants of the corresponding stiffness matrices, so again we do not need to specify the right hand side $\tilde{\mathcal{F}} \in \tilde{\mathcal{Y}}'$. It can be easily seen that the assumptions of Corollary 3.3.10 are fulfilled with $A' = A$ time-independent and spaces $W_+ = H^2(\Omega) \hookrightarrow W_0 = H^1(\Omega) \hookrightarrow W_- = L_2(\Omega)$. That is, the regularity property (4.2.2) is satisfied.

Again, we use the smallest singular values of $\mathbf{B}_{j,\ell}$ as an indicator for the discrete inf-sup constants. The results are illustrated in Figure 6.1.12.

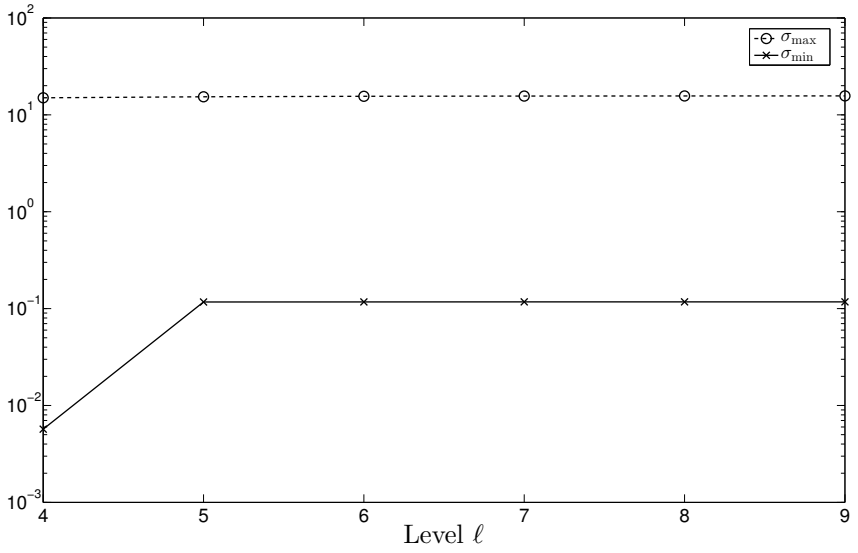



Figure 6.1.12: Plot of the smallest singular values of $\mathbf{B}_{j,\ell}$ and of the largest singular values for fixed level $j = 4$ w.r.t to (6.1.11).

Different from Figure 6.1.7 we fixed the minimum level $j = 4$ and successively increase the level $\ell$ of the trial space in Figure 6.1.12 to stabilize the pair of subspaces. For this PDE Example 6.1.10, we see qualitatively the same behavior as in the ODE Example 6.1.5. We can observe that for a level difference of one or more levels, the smallest singular values stay almost equal. This confirms also in this case that a level difference of one already seems to be sufficient for $\beta_{j,\ell}$ being uniformly bounded away from zero. For the same level $j = \ell$ we can clearly see, that the value is much smaller than for different levels. As in Figure 6.1.7 the largest singular values are asymptotically constant.

**B-Splines of higher order**

We have considered globally continuous, piecewise linear functions for discretization so far. Now we will analyze the stability, in particular in view of the number of required extra layers $L$, for discretizations with smoother ansatz functions. To this end, we

use B-splines of higher order, cf. section 2.4. Since B-spline bases do not provide Riesz-stability (6.1.1) for a scale of Sobolev spaces, we cannot compute the inf-sup constants as done before in (6.1.3). Nevertheless, we can derive a similar expression with an approach as in section 4.1 by introducing Riesz isomorphisms respectively Gram matrices. Let $S_j^t := S_{j,k}^t := SP_{j,k} \subset L_2(0,1)$ be the set spanned by B-splines of order $k \in \mathbb{N}$ on a uniform grid of size $2^{-j}$, $j \in \mathbb{N}$, according to Proposition 2.4.2 and (2.4.6). That means

$$S_{j,k}^t = \operatorname{span}\{N_{i,k} : i = 1, \dots, N\}, \tag{6.1.13}$$

with respect to an extended sequence of knots $\{\theta_i\}_{i=1,\dots,N+k}$ such that

$$\theta_1 = \dots = \theta_k = 0 < \theta_{k+1} < \dots < \theta_N < 1 = \theta_{N+1} = \dots = \theta_{N+k},$$

where $\theta_{i+1} - \theta_i = 2^{-j}$ for inner knots $i = k, \dots, N$. We arrange the spaces $S_{j,k}^x$, $Q_{\ell,k}^t$ and $Q_{\ell,k}^x$ in a similar way, where the order and spacing of the B-splines may vary in each space. Finally, our discrete subspaces of $\tilde{\mathcal{X}}$ and $\tilde{\mathcal{Y}}$ are tensor product spaces of these spaces on one-dimensional domains, i.e.,

$$S_{\mathbf{j},\mathbf{k}} := S_{j_1,k_1}^t \otimes S_{j_2,k_2}^x \subset \tilde{\mathcal{X}}, \quad Q_{\boldsymbol{\ell},\bar{\mathbf{k}}} = Q_{\ell_1,\bar{k}_1}^t \otimes Q_{\ell_2,\bar{k}_2}^x \subset \tilde{\mathcal{Y}}, \tag{6.1.14}$$

with sufficiently high order $k_1, k_2, \bar{k}_1, \bar{k}_2 \in \mathbb{N}$, level $j_1, j_2, \ell_1, \ell_2 \in \mathbb{N}$ and possibly boundary adaptations. For spatial domains of dimension larger than one, the spaces $S_{j_2,k_2}^x$ and $Q_{\ell_2,\bar{k}_2}^x$ are constructed also as tensor products, where $j_2, \ell_2, k_2, \bar{k}_2$ become vectors accordingly. An adaptation to zero boundary condition can be done easily be omitting the very first and/or very last B-spline from the set of B-splines spanning the spline space. That means,

$$\begin{aligned} (S_{j,k}^t)_{0,\{0\}} &:= \operatorname{span}\{N_{i,k},\ i = 2, \dots, N\}, \\ (S_{j,k}^t)_{0,\{1\}} &:= \operatorname{span}\{N_{i,k},\ i = 1, \dots, N-1\}, \\ (S_{j,k}^t)_0 &:= \operatorname{span}\{N_{i,k},\ i = 2, \dots, N-1\}, \end{aligned} \tag{6.1.15}$$

exemplarily for the temporal discretization, depending on the boundary conditions. Notice that the stability analysis of chapter 4 is restricted to have the same level for the temporal and spatial part. This would mean $\mathbf{j} = (j, j, \dots, j)$ and $\boldsymbol{\ell} = (\ell, \ell, \dots, \ell)$ and we denote those spaces simply as $S_{j,\mathbf{k}}$ and $Q_{\ell,\bar{\mathbf{k}}}$, respectively. With these bases we can arrange system matrices in a similar way as in (6.1.2). Although the notation might look overloaded, we expect it to be more convenient. Nevertheless, we usually omit the particular indication of the boundary adaptation in the discrete spaces from (6.1.15), when it is clear from the context which boundary conditions are required. Moreover, let $R_{\tilde{\mathcal{X}}}$ and $R_{\tilde{\mathcal{Y}}}$ be the Riesz mappings on $\tilde{\mathcal{X}}$ and $\tilde{\mathcal{Y}}$, respectively. Its discretizations are Gram matrices $\mathbf{R}_j^{\tilde{\mathcal{X}}}$ and $\mathbf{R}_\ell^{\tilde{\mathcal{Y}}}$, where we refer to section 4.1 again. Having these discretized Riesz mappings at hand, we can see that

$$\frac{\mathbf{q}_\ell^T \mathbf{B}_{j,\ell} \mathbf{v}_j}{\|\mathbf{v}_j\|_{\mathbf{R}_j^{\tilde{\mathcal{X}}}} \|\mathbf{q}_\ell\|_{\mathbf{R}_\ell^{\tilde{\mathcal{Y}}}}} = \frac{\langle \tilde{B} v_j, q_\ell \rangle}{\|v_j\|_{\tilde{\mathcal{X}}} \|q_\ell\|_{\tilde{\mathcal{Y}}}},$$

for $v_j := \sum_{\phi_i \in S_{j,\mathbf{k}}} v_i \phi_i$ and $q_\ell := \sum_{\bar{\phi}_i \in Q_{j,\bar{\mathbf{k}}}} q_i \bar{\phi}_i$, with $\|\mathbf{v}_j\|^2_{\mathbf{R}_j^{\tilde{\mathcal{X}}}} := \mathbf{v}_j^T \mathbf{R}_j^{\tilde{\mathcal{X}}} \mathbf{v}_j$. Since there holds

$$\frac{\mathbf{q}_\ell^T \mathbf{B}_{j,\ell} \mathbf{v}_j}{\|\mathbf{v}_j\|_{\mathbf{R}_j^{\tilde{\mathcal{X}}}} \|\mathbf{q}_\ell\|_{\mathbf{R}_\ell^{\tilde{\mathcal{Y}}}}} = \frac{((\mathbf{R}_\ell^{\tilde{\mathcal{Y}}})^{\frac{1}{2}} \mathbf{q}_\ell)^T ((\mathbf{R}_\ell^{\tilde{\mathcal{Y}}})^{-\frac{1}{2}} \mathbf{B}_{j,\ell} (\mathbf{R}_j^{\tilde{\mathcal{X}}})^{-\frac{1}{2}}) ((\mathbf{R}_j^{\tilde{\mathcal{X}}})^{\frac{1}{2}} \mathbf{v}_j)}{\|(\mathbf{R}_j^{\tilde{\mathcal{X}}})^{\frac{1}{2}} \mathbf{v}_j\|_{\ell_2} \|(\mathbf{R}_\ell^{\tilde{\mathcal{Y}}})^{\frac{1}{2}} \mathbf{q}_\ell\|_{\ell_2}},$$

we can conclude

$$\min_{\tilde{\mathbf{v}}_j \in \ell_2(\mathbb{R}^{N_j}) \setminus \{0\}} \max_{\tilde{\mathbf{q}}_\ell \in \ell_2(\mathbb{R}^{N_j}) \setminus \{0\}} \frac{\tilde{\mathbf{q}}_\ell^T (\mathbf{R}_\ell^{\tilde{\mathcal{Y}}})^{-\frac{1}{2}} \mathbf{B}_{j,\ell} (\mathbf{R}_j^{\tilde{\mathcal{X}}})^{-\frac{1}{2}} \tilde{\mathbf{v}}_j}{\|\tilde{\mathbf{v}}_j\|_{\ell_2} \|\tilde{\mathbf{q}}_\ell\|_{\ell_2}} = \inf_{v_j \in S_{j,\mathbf{k}} \setminus \{0\}} \sup_{q_\ell \in Q_{\ell,\bar{\mathbf{k}}} \setminus \{0\}} \frac{\langle B v_j, w_\ell \rangle}{\|q_\ell\|_{\tilde{\mathcal{Y}}} \|v_j\|_{\tilde{\mathcal{X}}}}.$$

Therefore, the discrete inf-sup constant is given exactly by the smallest singular value of

$$(\mathbf{R}_\ell^{\tilde{\mathcal{Y}}})^{-\frac{1}{2}} \mathbf{B}_{j,\ell} (\mathbf{R}_j^{\tilde{\mathcal{X}}})^{-\frac{1}{2}}, \tag{6.1.16}$$

similar to (6.1.4), where we have to compute square roots of inverse matrices. From a practical point of view, we see that a basis of wavelet type has a big advantage since one does not need to compute these inversions and square roots, because of the norm equivalence on Sobolev spaces (6.1.1). Our interest here is mainly of theoretical nature, that is, analyzing the stability. For a practically more efficient and preconditioned approach with piecewise linear functions for the first formulation we refer to [And14, And16]. Using tensor B-splines of high order is of crucial importance in *isogeometric analysis*. It connects computer aided design and finite element methods. The area of research is comparably new and was mainly introduced in [BCH05, BCH09] and the reader is referred to it for details. There is also a recent work [LMN15] on isogeometric analysis for parabolic problems in space-time variational form. For a finite element approach without tensor product structure we refer to [Ste15].

Recall that $\tilde{\mathcal{Y}} = L_2(I;V) \cap H^1_{0,\{T\}}(I;V')$ with given interval $I := (0,T)$. In order to assemble the corresponding Gram matrix induced by the Riesz isomorphism with respect to $\tilde{\mathcal{Y}}$, one needs, in particular, to determine the discrete Riesz mapping $\mathbf{R}_\ell^{V'}$ on the dual space. It is worth mentioning that $\mathbf{R}_\ell^{V'} \neq (\mathbf{R}_\ell^V)^{-1}$ as one might expect. But since $H$ is the pivot space, one can deduce that

$$\mathbf{R}_\ell^{V'} = \mathbf{R}_\ell^H (\mathbf{R}_\ell^V)^{-1} \mathbf{R}_\ell^H, \tag{6.1.17}$$

due to the fact that $\mathbf{R}_\ell^H$ is the exact canonical embedding according to $v \mapsto R_H v = (v, \cdot)_H$ of a function $v \in Q^x_{\ell_2,\bar{k}_2} \subset H$. Indeed,

$$(v,w)_{V'} = {}_V\langle R_V^{-1} R_H v, R_H w \rangle_{V'} = (R_H R_V^{-1} R_H v, w)_H$$

for all $v, w \in Q^x_{\ell_2,\bar{k}_2}$.

Another possibility to circumvent this problem would be to shift the regularity of the spaces and consider the problem with respect to $\tilde{\mathcal{Y}}_+ := L_2(I;W) \cap H^1(I;H)$ and $\tilde{\mathcal{X}}_- := L_2(I;H)$ with $W \subset V$ such that $[W,H]_{\frac{1}{2}} = V$, provided that the problem is

sufficiently smooth. These are also conform spaces according to Theorem 3.3.10 and the consideration from section 4.3 applies also for these slightly shifted spaces in a straightforward way. Since the prototype spaces (6.1.14) are tensor product spline spaces, the conditions from Table 4.3.12 respectively Table 4.3.29 are satisfied with Theorem 2.4.10 and 2.4.12, where one can simply set the dual spaces as the primal spaces, see remark after (4.3.4).

All the following examples are calculated with the Matlab code for Petrov-Galerkin discretizations with B-splines implemented in the course of this thesis. We used the Matlab routine `svd` to calculate a singular value decomposition and, therefore, the singular values. The program was running on different computers and with Matlab versions Matlab 8.5 (R2015a), Matlab 8.6 (R2015b), and Matlab 9.0 (R2016a). Since the explicit evaluation times are not of interest in any of the subsequent results, we will refrain from their specification for each calculation. A detailed overview of all involved routines is given in appendix A and in the comments in the code itself.

We consider the following heat equation in second space-time weak form and its Petrov-Galerkin solution.

**Example 6.1.18.** *Let $I := (0,1)$ and $D := (0,1)$. Find $u \in \tilde{\mathcal{X}} = L_2(I; H_0^1(D))$ such that*

$$\tilde{b}(u,w) = \tilde{\mathcal{F}}(w) \qquad \text{for all } w \in \tilde{\mathcal{Y}} = L_2(I; H_0^1(D)) \cap H^1_{0,\{1\}}(I; H^{-1}(D)),$$

*with*

$$\begin{aligned}\tilde{b}(v,w) &:= \int_I \left(-\langle v(t), \dot{w}(t)\rangle + \langle \nabla v(t), \nabla w(t)\rangle\right) \,\mathrm{d}t \\ \tilde{\mathcal{F}}_\omega(w) &:= \int_I \langle g(t), w(t)\rangle \,\mathrm{d}t,\end{aligned}$$

*where again $g \in \tilde{\mathcal{Y}}'$ does not need to be specified here.*

Now we want to analyze the behavior of the discrete inf-sup constant in the way explained in (6.1.16) by calculating the smallest singular value of

$$(\mathbf{R}_\ell^{\tilde{\mathcal{Y}}})^{-\frac{1}{2}} \mathbf{B}_{j,\ell} (\mathbf{R}_j^{\tilde{\mathcal{X}}})^{-\frac{1}{2}}$$

for different B-spline bases.

Recalling the discrete subspaces from (6.1.14), we start with a B-spline basis of fixed level and order in both, solution and test space and in each coordinate direction. That means

$$S_{j,\mathbf{k}} = S^t_{j,k} \otimes S^x_{j,k}, \quad Q_{j,\mathbf{k}} = Q^t_{j,k} \otimes Q^x_{j,k},$$

for different order of splines $k = 3, \dots, 7$. Then we increase the resolution in the test space by one extra level, i.e.,

$$S_{j,\mathbf{k}} = S^t_{j,k} \otimes S^x_{j,k}, \quad Q_{j+1,\mathbf{k}} = Q^t_{j+1,k} \otimes Q^x_{j+1,k}. \tag{6.1.19}$$

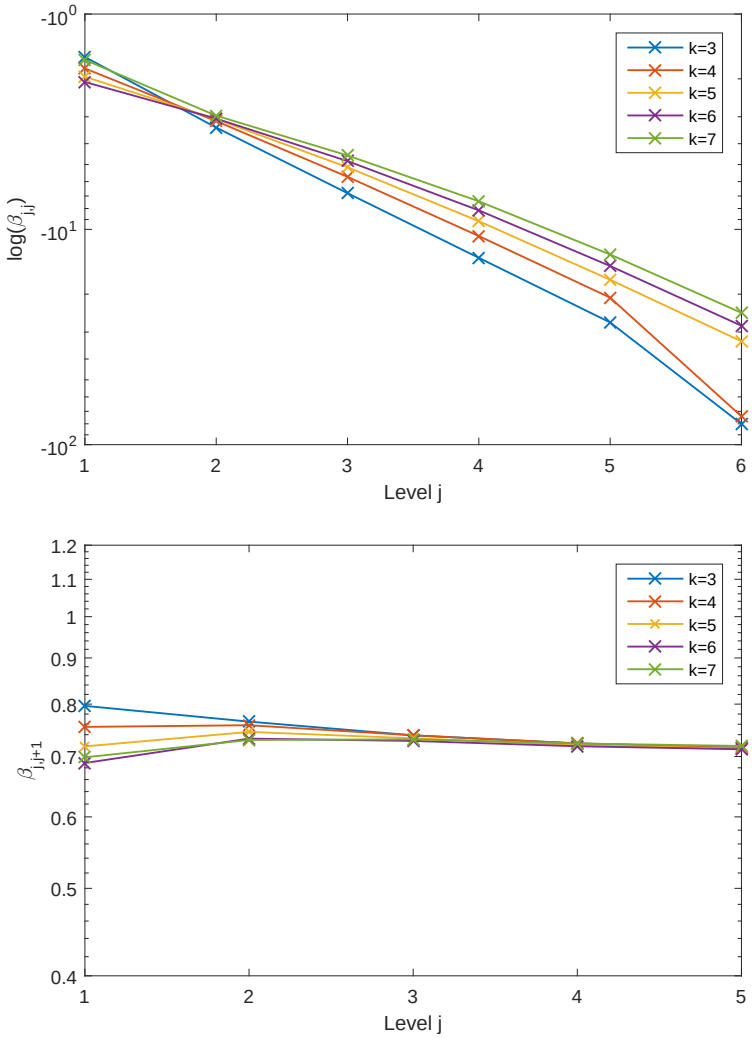



Figure 6.1.20: Discrete inf-sup constants for $S_{j,\mathbf{k}} = S^t_{j,k} \otimes S^x_{j,k}$ and $Q_{j,\mathbf{k}} = Q^t_{j,k} \otimes Q^x_{j,k}$ (top), $S_{j,\mathbf{k}} = S^t_{j,k} \otimes S^x_{j,k}$ and $Q_{j+1,\mathbf{k}} = Q^t_{j+1,k} \otimes Q^x_{j+1,k}$ (bottom), according to example (6.1.18).

The corresponding slope of the inf-sup constants are presented in Figure 6.1.20.

One can observe that the Petrov-Galerkin discretization is extremely unstable for a usual discretization with equal level of refinement in solution and test space. The inf-sup constant even seems to decrease double exponentially. But as we have already observed in the piecewise linear example above, the discretization is stabilized, when we enrich the test space with one extra layer. Therefore, one can observe qualitatively the same behavior also for splines of higher order.

Next we want to see how the inf-sup constant behaves when we increase the order of B-splines in the test space, but keep the order in the solution space as low as possible.

Therefore, we consider first

$$S_{j,\mathbf{k}} = S^t_{j,1} \otimes S^x_{j,2}, \quad Q_{j,\mathbf{k}} = Q^t_{j,k} \otimes Q^x_{j,k}$$

and second

$$S_{j,\mathbf{k}} = S^t_{j,1} \otimes S^x_{j,2}, \quad Q_{j+1,\mathbf{k}} = Q^t_{j+1,k} \otimes Q^x_{j+1,k}$$

with increasing order $k = 3, \ldots, 7$. The results are illustrated in Figure 6.1.21.

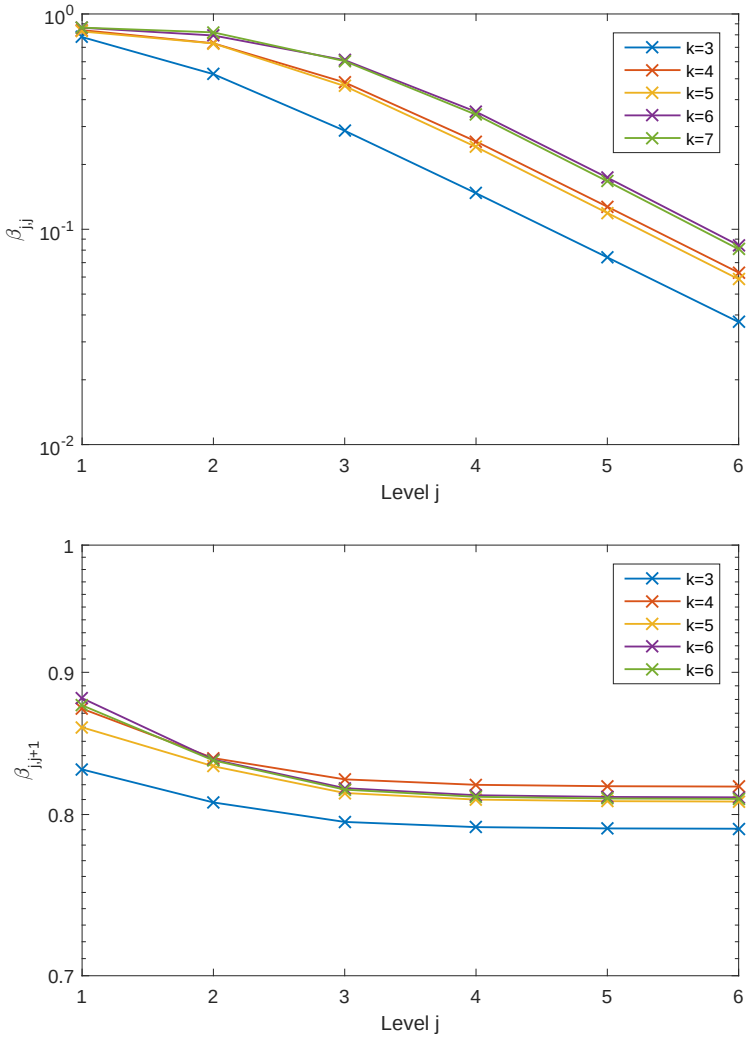



Figure 6.1.21: Discrete inf-sup constants for $S_{j,\mathbf{k}} = S^t_{j,1} \otimes S^x_{j,2}$ and $Q_{j,\mathbf{k}} = Q^t_{j,k} \otimes Q^x_{j,k}$ (top), $S_{j,\mathbf{k}} = S^t_{j,1} \otimes S^x_{j,2}$ and $Q_{j+1,\mathbf{k}} = Q^t_{j+1,k} \otimes Q^x_{j+1,k}$ (bottom), according to example (6.1.18).

We can see that, indeed, increasing the *order* of splines in the test space improves the stability quantitatively. Nevertheless, the inf-sup constant still seems to decrease

exponentially, but not double exponentially as before. For the sake of completeness, we have also plotted the slope with one extra layer in the test space. One can see that this stabilizes the discretization also for this case.

Finally, we would like to investigate how sensible the stability is under perturbations of the diffusion operator. To this end, we modify Example 6.1.18 in the following way.

**Example 6.1.22.** *Let $I := (0,1)$ and $D := (0,1)$. Find $u \in \tilde{\mathcal{X}} = L_2(I; H_0^1(D))$ such that*

$$\tilde{b}(u,w) = \tilde{\mathcal{F}}(w) \qquad \text{for all } w \in \tilde{\mathcal{Y}} = L_2(I; H_0^1(D)) \cap H^1_{0,\{1\}}(I; H^{-1}(D)),$$

*with*

$$\begin{aligned}\tilde{b}(v,w) &:= \int_I \left(-\langle v(t), \dot{w}(t)\rangle + \varepsilon \langle \nabla v(t), \nabla w(t)\rangle\right) \,\mathrm{d}t \\ \tilde{\mathcal{F}}(w) &:= \int_I \langle g(t), w(t)\rangle \,\mathrm{d}t,\end{aligned}$$

*for any $g \in \tilde{\mathcal{Y}}'$ with $\varepsilon \in \mathbb{R}^+$.*

This is also a first attempt into the direction of random PDEs as discussed in the next section. We scale the diffusion by successively decreasing $\varepsilon$, which in turn scales the coercivity constant $A_{\min}$ of the spatial differential operator. In this way, the continuous (non-discrete) problem gets more and more ill-conditioned, but still well-conditioned for each fixed value of $\varepsilon > 0$. Figure 6.1.23 shows the inf-sup constants according to Example 6.1.22 with respect to the subspaces (6.1.19) and fixed order $k = 3$.

One can observe that the discrete inf-sup constant gets worse when $\varepsilon$ gets smaller, but stays asymptotically bounded. The fact that the inf-sup constants get worse is simply due to the fact that the underlying non-discrete problem becomes more and more ill-conditioned. That means that one extra layer in the test space seems to suffice in order to obtain a stable discretization even for ill-conditioned problems.

## 6.2 Existence of Moments for Parabolic Random PDEs

This section is devoted to illustrating the results for parabolic *random* PDEs from chapter 5. We will demonstrate the existence of moments of the solution depending on the moments of the lower and upper bounds $A_{\min}$ and $A_{\max}$ of the spatial differential operator as well as of the right hand side $\tilde{\mathcal{F}}_\omega$ and initial value $U_0$. The following numerical examples are intended to underline mainly Theorem 5.1.9 respectively Corollary 5.1.17 or rather Corollary 5.1.12 and Corollary 5.1.20 numerically. We refer to Table 5.1.21 for an overview.

First, we will present some numerical examples to illustrate the existence of moments. In order to do so, we calculate reference solutions *for each sample point* given by

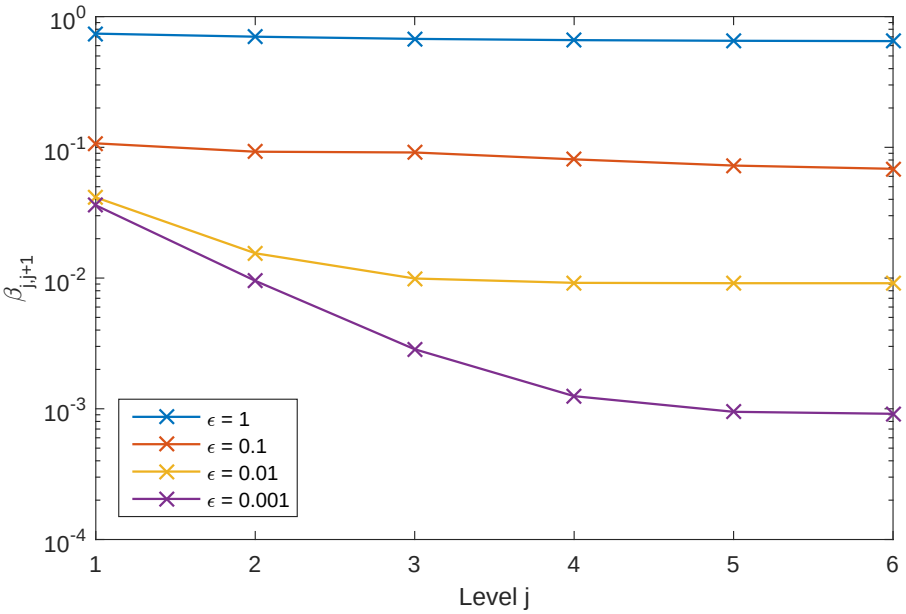



Figure 6.1.23: Discrete inf-sup constants for $S_{j,\mathbf{k}} = S^t_{j,3} \otimes S^x_{j,3}$ and $Q_{j+1,\mathbf{k}} = Q^t_{j+1,3} \otimes Q^x_{j+1,3}$ (bottom), according to example (6.1.22).

the random parameter $\omega$, which is very expensive computationally. We consider the following problem:

**Example 6.2.1.** *Let $I := (0,1)$ and $D := (0,1)^2$. Find $U_\omega := U(\cdot,\omega) \in \tilde{\mathcal{X}} = L_2(I; H^1_0(D))$ $\mathbb{P}$-a.s. such that for $\omega \in \Omega$ a.s.*

$$\tilde{b}_\omega(U_\omega, w) = \tilde{\mathcal{F}}_\omega(w) \qquad \text{for all } w \in \tilde{\mathcal{Y}} = L_2(I; H^1_0(D)) \cap H^1_{0,\{1\}}(I; H^{-1}(D)),$$

*with*

$$\tilde{b}_\omega(v,w) := \int_I \left(-\langle v(t), \dot{w}(t)\rangle + a(\omega)\langle \nabla v(t), \nabla w(t)\rangle\right) \,\mathrm{d}t$$
$$\tilde{\mathcal{F}}_\omega(w) := c(\omega) \int_I \langle g(t), w(t)\rangle \,\mathrm{d}t,$$

*for $g(t,x,y) = \sin(\pi t)\sin(\pi x)\sin(\pi y)$ and random variables $a(\omega)$ and $c(\omega)$.*

The previous Example 6.2.1 is the second space-time weak formulation of the parabolic random heat equation with random diffusion and right hand side. In the following, we will analyze how the moments of the solutions of Example 6.2.1 depend on the parameters $a(\omega)$ and $c(\omega)$. These results were calculated by Matteo Molteni[1] and are taken from [LMM16]. For the discretization it was used globally continuous B-splines of order

[1]Chalmers University of Technology, Gothenburg, Sweden

two in time for the solution space, that is, $S^t_{j,2}$, and globally discontinuous B-splines of order one in time with the same refinement, that is, $Q^t_{j,1}$. For the spatial discretization piecewise linear finite elements on a triangulation were used. The calculations are performed in python with the software package `FEniCS`. Notice that solving an approximation with this discretization is equivalent to a modified Crank-Nicolson scheme and in this way a time stepping was applied. These discretizations from [LMM16] differ slightly from the ones introduced here in the context of the (deterministic) stability of Petrov-Galerkin approaches of the second formulation of chapter 4. Nevertheless, the stability of the particular choice here can be guaranteed since a CFL condition is met as proven in [LMM16]. Beside the number of existing moments of the solution of the continuous problems (3.3.4), it is also ensured that its approximation has the same number of finite moments, see again [LMM16, Th. 12]. Therefore, the discretization is well suited for our purposes.

We want to analyze Example 6.2.1 for different values of $a(\omega)$ and $b(\omega)$, such that we control the number of moments $p_1$, $p_2$ and $p_3$ of $A^{-1}_{\min}$, $A_{\max}$ and $\tilde{\mathcal{F}}$, that means, $A^{-1}_{\min} \in L_{p_1}(\Omega)$, $A_{\max} \in L_{p_2}(\Omega)$ and $\tilde{\mathcal{F}} \in L_{p_3}(\Omega; \tilde{\mathcal{Y}}')$. To this end, we consider

$$a := 1 + \frac{1}{X^2}, \qquad c := 1 + X^3, \tag{6.2.2}$$

$$a := |X|^{0.99}, \qquad c := 1 + X^3, \tag{6.2.3}$$

$$a := |X|^{0.99}, \qquad c := \frac{1}{\sqrt{|X|}}, \tag{6.2.4}$$

$$a := |X|^{0.99}, \qquad c := \frac{1}{\sqrt{|X - 0.4|}}, \tag{6.2.5}$$

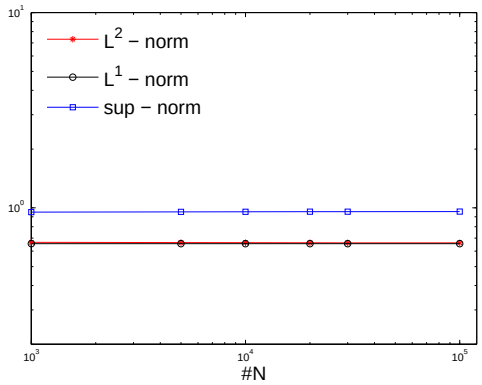
with uniformly distributed random variable $X := X(\omega) \sim \mathcal{U}(-0.5, 0.5)$ for all examples. Having uniformly distributed $X$, one can identify $\mathbb{P}$ with the Lebesgue measure. Moreover, one can specify the probability space as $\Omega := [-0.5, 0.5]$, its $\sigma$-Algebra $\Sigma$ as the Borel $\sigma$-Algebra on $\Omega$ and Lebesgue measure $\mathbb{P}$ and set $X \colon [-0.5, 0.5] \to \mathbb{R}$ as the identity $X(\omega) := \omega$. In this way we can formally treat $\omega$ as a deterministic parameter. The moments, respectively the $L_p(\Omega; \tilde{\mathcal{X}})$-norm of a function $v \in L_p(\Omega; \tilde{\mathcal{X}})$ with $v \colon I \times D \times \Omega \to \mathbb{R}$ is given as

$$\|v\|^p_{L_p(\Omega;\tilde{\mathcal{X}})} = \int_{-0.5}^{0.5} \|v(\cdot, \cdot, z)\|^p_{\tilde{\mathcal{X}}} \, \mathrm{d}z.$$

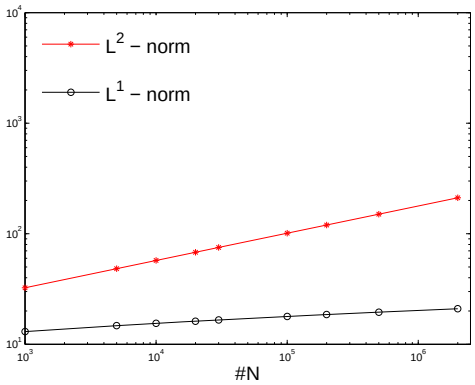
For the computation of the $L_p(\Omega; \cdot)$-norm we can, thus, use standard quadrature rules and do not need to use Monte-Carlo methods which have a rather slow convergence. A trapezoidal rule was used for the quadrature. The results are presented in Figure 6.2.6.

We have chosen the random variables such that

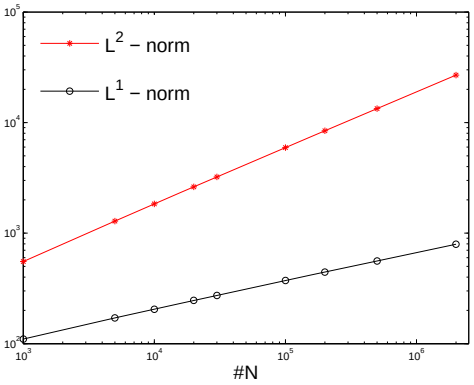
- (6.2.2): $A^{-1}_{\min} \in L_\infty(\Omega)$, $A_{\max} \notin L_1(\Omega)$ and $\tilde{\mathcal{F}}_\omega \in L_\infty(\Omega; \tilde{\mathcal{X}})$,



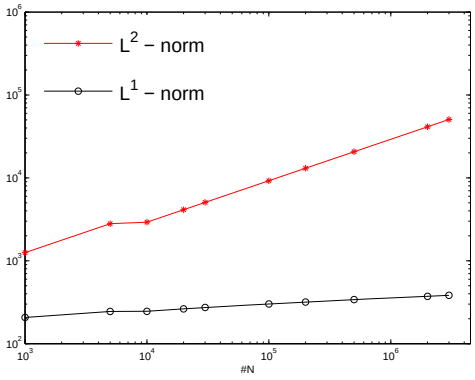

(a) Choice (6.2.2)




(b) Choice (6.2.3)




(c) Choice (6.2.4)




(d) Choice (6.2.5)

Figure 6.2.6: $L_p(\Omega;\tilde{\mathcal{X}})$-norms of the numerical solution of Example 6.2.1 with varying number of quadrature point $N$

- (6.2.3): $A_{\min}^{-1} \in L_1(\Omega) \setminus L_2(\Omega)$, $A_{\max} \in L_\infty(\Omega)$ and $\tilde{\mathcal{F}}_\omega \in L_\infty(\Omega;\tilde{\mathcal{X}})$,
- (6.2.4): $A_{\min}^{-1} \in L_1(\Omega) \setminus L_2(\Omega)$, $A_{\max} \in L_\infty(\Omega)$ and $\tilde{\mathcal{F}}_\omega \in L_1(\Omega;\tilde{\mathcal{X}}) \setminus L_2(\Omega;\tilde{\mathcal{X}})$,
- (6.2.5): $A_{\min}^{-1} \in L_1(\Omega) \setminus L_2(\Omega)$, $A_{\max} \in L_\infty(\Omega)$ and $\tilde{\mathcal{F}}_\omega \in L_1(\Omega;\tilde{\mathcal{X}}) \setminus L_2(\Omega;\tilde{\mathcal{X}})$.

As expected from Corollary 5.1.17, for instance, we see in Figure 6.2.6a that the solution has arbitrary many moments if $A_{\min}$ is uniformly bounded away from zero and if the right hand side is uniformly bounded from above, although the spatial differential operator itself is unbounded with respect to $\omega$. The upper bound $A_{\max}$ does not even has a finite expectation. The second example (6.2.3) has a uniformly bounded upper bound for the spatial differential operator and right hand side, but only the expectation of $A_{\min}^{-1}$ is finite whereas the second moment is infinite. By taking the limit in Theorem 5.1.9, we would expect the solution to have one finite moment. This is confirmed by Figure 6.2.6b. One can even observe that the estimate is sharp since the second

moment does not exist as the norm tends to infinity when increasing the number of quadrature points $N$. The slope of the $L_1$-norm is not perfect in this example because $\mathbb{E}\left[\frac{1}{|X|^{0.99}}\right] < \infty$ but $\mathbb{E}\left[\frac{1}{|X|}\right] = \infty$, where $\mathbb{E}[Y] := \int_\Omega Y(\omega)\mathrm{d}\mathbb{P}(\omega)$ denotes the expectation value of a random variable $Y \in L_1(\Omega)$. Moreover, all examples (6.2.3) – (6.2.5) have a singularity in zero, such that the convergence/divergence speed is expected to be very slow. In the next example (6.2.4) we additionally introduce an unbounded right hand side with respect to $\omega$. Therefore, we would expect that the expectation of the solution does not exist any more. Indeed, we can see in Figure 6.2.6c that both, the $L_1$-norm and the $L_2$-norm tend to infinity, when we increase the number of quadrature points. This is in accordance with the fact that the consistency condition is not satisfied for Theorem 5.1.9, Corollary 5.1.12 or 5.1.17. For the last example (6.2.5), we have shifted the singularity in the right hand side. In (6.2.4) the singularity in $A_{\min}^{-1}$ and $\tilde{\mathcal{F}}_\omega$ appeared at the same point $X(\omega) = 0$. Figure 6.2.6d shows that the expectation is finite but the second moment tends to infinity. This is neither covered by Theorem 5.1.9 nor by Corollary 5.1.12 or 5.1.17, but was analyzed explicitly in Example 5.1.19 before. As in Figure 6.2.6b, the slope of the $L_1$-norm is not perfect, but in comparison with Figure 6.2.6c, one can see the difference when the singularities do not coincide.

We conclude the analysis of the existence of moments in this section with an example without uniformly distributed random parameters. Furthermore, we want to illustrate the validity of Corollary 5.1.20 with non-coercive spatial differential operators. Therefore, we consider the following problem.

**Example 6.2.7.** *Let* $I := (0,1)$ *and* $D := (0,1)$. *Find* $U_\omega := U(\cdot,\omega) \in \tilde{\mathcal{X}} = L_2(I; H_0^1(D))$ $\mathbb{P}$*-a.s. such that for* $\omega \in \Omega$ *a.s.*

$$\tilde{b}_\omega(U_\omega, w) = \tilde{\mathcal{F}}_\omega(w) \qquad \text{for all } w \in \tilde{\mathcal{Y}} = L_2(I; H_0^1(D)) \cap H^1_{0,\{1\}}(I; H^{-1}(D)),$$

*with*

$$\begin{aligned}\tilde{b}_\omega(v,w) &:= \int_I -\left(\langle v(t), \dot{w}(t)\rangle + \langle \nabla v(t), \nabla w(t)\rangle + a(\omega)(v(t), w(t))_H\right) \mathrm{d}t \\ \tilde{\mathcal{F}}_\omega(w) &:= \int_I \langle g(t), w(t)\rangle \,\mathrm{d}t,\end{aligned}$$

*with* $g \equiv 1$ *and* normally distributed *random variable* $a \sim \mathcal{N}(0,16)$.

It is easy to see that the random variable $a(\omega)$ now plays the role of the Gårding shift random variable $\lambda(\omega)$. A possible choice was already discussed after Corollary 5.1.20. In compliance with the discussion, we choose $a \sim \mathcal{N}(0,16)$ normally distributed. Recall that a normally distributed random variable may take values on the whole real line $\mathbb{R}$, so that $a$ can neither be bounded uniformly from below nor from above. But as mentioned after Corollary 5.1.20, it is feasible since the probability density function

decays exponentially fast, meaning that $\exp(a)$ has finite moments. We set the variance of the distribution to 16 in order to broaden the probability distribution function and to have a stronger deviation. Contrary to the uniformly distributed parameters above, we cannot express the random variable one to one as a deterministic parameter with Lebesgue measure. Nevertheless, since we have a probability density function

$$f_{\mathcal{N}(0,\sigma^2)}(x) := \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}, \quad x \in \mathbb{R} \tag{6.2.8}$$

for the normal distribution, one can arrange a probability space as $\Omega := \mathbb{R}$, its Borel $\sigma$-Algebra $\Sigma$ and weighted Lebesgue measure according to $\mathrm{d}\mathbb{P} := f_{\mathcal{N}(0,\sigma^2)}(\lambda)\mathrm{d}\lambda$ with standard Lebesgue measure $\lambda$. The $L_p(\Omega;\tilde{\mathcal{X}})$-norm of a function $v \in L_p(\Omega;\tilde{\mathcal{X}})$ with $v\colon I \times D \times \Omega \to \mathbb{R}$ is given as

$$\|v\|^p_{L_p(\Omega;\tilde{\mathcal{X}})} = \int_{\mathbb{R}} \|v(\cdot,\cdot,z)\|^p_{\tilde{\mathcal{X}}} \; f_{\mathcal{N}(0,\sigma^2)}(z)\mathrm{d}z.$$

With a slight renormalization one could actually use Gauss-Hermite quadrature for the calculation of the $L_p(\Omega;\cdot)$-norm, but since we do not know anything about the smoothness of the solution $U_\omega$ on $a$, we found it more convenient to use a Monte-Carlo simulation. Contrary to a Gauss quadrature with Legendre points, one cannot simply split the interval into small subintervals and apply the Gauss-Hermite quadrature piecewise. The accuracy can only be improved by increasing the order of the quadrature. The numerical calculations are done with the B-spline Petrov-Galerkin developed for this work. For the (deterministic) discretization we have chosen

$$S_{5,2} = S^t_{5,2} \otimes S^x_{5,2}, \quad Q_{6,2} = Q^t_{6,2} \otimes Q^x_{6,2},$$

according to the notation (6.1.14). In order to calculate the energy norm $\|\cdot\|_{\tilde{\mathcal{X}}}$ of the approximate solution, we made use of the observation (4.1.13) together with (4.1.12). The corresponding Gram matrices are assembled by the function `RieszGen` or `RieszMat` of the provided code and the system matrices by `StiffGen` or `StiffMat`. The numerical results are plotted in Figure 6.2.9.

It shows the approximation of the $L_1(\Omega;\tilde{\mathcal{X}})$ and $L_2(\Omega;\tilde{\mathcal{X}})$ norm, but also a higher order $L_{50}(\Omega;\tilde{\mathcal{X}})$-norm. As expected due to Corollary 5.1.20 and the discussions afterwards, we see that the norms stay asymptotically constant, meaning that all the norms are finite and corresponding moments exist.

## 6.3 Quasi-Optimality of Petrov-Galerkin Discretizations

Next, we will take a closer look at the Petrov-Galerkin solution with respect to the random parameter. We have already seen the stability for deterministic PDEs in space-time form in practice in the previous section. Here, we focus on the quasi-optimality

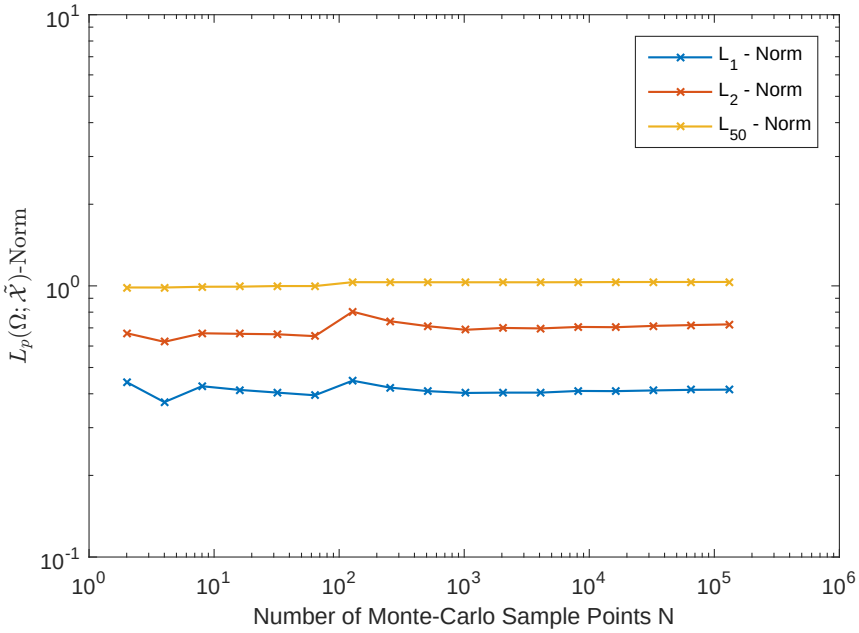



Figure 6.2.9: Numerical results of Example 6.2.7

with respect to the stochastic parameter. We will investigate how the Petrov-Galerkin solution converges not only pathwise for each sample, but in a $L_p(\Omega;\cdot)$-sense. Since we have already observed results with increasing norms, that is, solutions which are not in $L_p(\Omega;\cdot)$ for certain $p$, we restrict ourselves to examples with finite norm. Talking about error convergence would be meaningless in these cases anyway. Recall our prototype Examples 6.2.1. Although it was shown numerically in section 6.1 that we have stability also for the second space-time formulation with only one extra layer, we will consider the first (homogenized) space-time formulation in the following, because it was shown in this form theoretically in Theorem 5.3.13, respectively Corollary 5.3.16. But we want to underline that everything works for the second formulation too. Therefore, we consider the prototype Example 6.2.1 in the first (homogenized) formulation.

**Example 6.3.1.** *Let* $I := (0,1)$ *and* $D := (0,1)$. *Find* $U_\omega := U(\cdot,\omega) \in \mathcal{X}_0 = L_2(I;H_0^1(D)) \cap H^1_{0,\{0\}}(I;H^{-1}(D))$ $\mathbb{P}$*-a.s. such that for* $\omega \in \Omega$ *a.s.*

$$b_{0,\omega}(U_\omega, w) = \mathcal{F}_{0,\omega}(w) \qquad \text{for all } w \in \mathcal{Y}_0 = L_2(I;H_0^1(D)),$$

*with*

$$b_{0,\omega}(v,w) := \int_I \left(\langle \dot{v}(t), w(t)\rangle + a(\omega)\langle \nabla v(t), \nabla w(t)\rangle\right)\,\mathrm{d}t$$
$$\mathcal{F}_{0,\omega}(w) := c(\omega)\int_I \langle g(t), w(t)\rangle\,\mathrm{d}t,$$

*for $g \equiv 1$ and random variables $a(\omega)$ and $c(\omega)$.*

Since we only deal with zero initial conditions for simplicity here, we may consider the canonical homogenization of the first formulation without additional Cartesian product. We fix the discrete subspaces as

$$S_{j,\mathbf{k}} = S^t_{j,2} \otimes S^x_{j,2}, \quad Q_{j+1,\mathbf{k}} = Q^t_{j+1,2} \otimes Q^x_{j+1,2}$$

for all problems concerning quasi-optimality in $L_p(\Omega;\mathcal{X}_0)$-spaces, where we have used the notation from the previous section 6.1 again. As a reference solution we take the solution on level $J = 7$. For the calculation we used the B-spline Petrov-Galerkin code. The error in the energy norm $\|\cdot\|_{\mathcal{X}_0}$ is calculated with the aid of Gram matrices as discrete Riesz mappings, see (4.1.12) and (4.1.13). Nevertheless, we cannot apply them directly since the approximation and the reference solution are defined with respect to different refinement levels. But since B-splines are refinable according to Proposition 2.4.7, the splines on the coarser grid can be transfered successively to a representation on the finest grid of the reference solution. This is done with the routine `RefineMat` component-by-component. Having a representation on the same grid, we can simply calculate the difference and apply the Gram matrices to obtain the value of the energy norm.

In our first examples we choose the random variables in Example 6.3.1 as

$$a := |X|^{\frac{1}{7}}, \qquad c := |X|^{-\frac{1}{10}}, \tag{6.3.2}$$

$$a := |X|^{\frac{1}{4}}, \qquad c := |X|^{-\frac{1}{7}}, \tag{6.3.3}$$

$$a := |X|^{0.99}, \qquad c := \frac{1}{\sqrt{|X-0.4|}}, \tag{6.3.4}$$

with uniformly distributed random variable $X := X(\omega) \sim \mathcal{U}(-0.5, 0.5)$ for all these examples. Therefore, we can compute the $\|\cdot\|_{L_p(\Omega;\cdot)}$-norm with suitable quadrature rules rather than with Monte-Carlo approximations. We have used a Gauss-Legendre quadrature of order 4 on 64 subintervals in our simulations. Note that this means that one has to calculate a reference solution, as well as all other approximations on minor levels, for these different realizations given by the quadrature points. Although we do not know how exactly the Petrov-Galerkin solution depends on the stochastic parameter, the subsequent results suggest that the quadrature is accurate enough. With these choices there holds

- (6.3.2): $A^{-1}_{\min} \in L_6(\Omega) \setminus L_7(\Omega)$, $A_{\max} \in L_\infty(\Omega)$ and $\mathcal{F}^\omega_0 \in L_9(\Omega;\mathcal{X}_0) \setminus L_{10}(\Omega;\mathcal{X}_0)$,
- (6.3.3): $A^{-1}_{\min} \in L_3(\Omega) \setminus L_4(\Omega)$, $A_{\max} \in L_\infty(\Omega)$ and $\mathcal{F}^\omega_0 \in L_6(\Omega;\mathcal{X}_0) \setminus L_7(\Omega;\mathcal{X}_0)$,
- (6.3.4): $A^{-1}_{\min} \in L_1(\Omega) \setminus L_2(\Omega)$, $A_{\max} \in L_\infty(\Omega)$ and $\mathcal{F}^\omega_0 \in L_1(\Omega;\mathcal{X}_0) \setminus L_2(\Omega;\mathcal{X}_0)$.

In addition to the examples above, we will also consider an example with non-uniformly distributed coefficients. To this end, we set $a \sim \mathcal{LN}(0,1)$ log-normally distributed, that is, a random variable with probability density function

$$f_{\mathcal{LN}(0,1)}(x) := \begin{cases} \frac{1}{\sqrt{2\pi}x} \exp\left(-\frac{\ln(x)^2}{2}\right) & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases}. \tag{6.3.5}$$

and $c \equiv 1$. This is equivalent to $a = \exp(Y)$ with normally distributed $Y \sim \mathcal{N}(0,1)$. It was already mentioned in chapter 5 that a log-normal random variable $a \sim \mathcal{LN}(0,1)$ has arbitrary many moments with $\mathbb{E}\left[a^s\right] = \exp\left(\frac{s^2}{2}\right)$, $s \in \mathbb{Z}$. This even holds true for negative values $s < 0$, so it also applies to the moments of $\frac{1}{a}$. But, obviously, no strict uniform upper bound and no strict lower bound away from zero exist for the values of $a(\omega)$. That means $a, \frac{1}{a} \notin L_\infty(\Omega)$, but $a, \frac{1}{a} \in L_p(\Omega)$ for arbitrary fixed $p \in \mathbb{R}^+$. The probability $\mathrm{d}\mathbb{P}$ can again not be identified directly with the Lebesgue measure as before, so we cannot use the Gauss-Legendre quadrature in the way as done above. But since we can express $a$ with a normal distribution, we have for any random variable $v(a) \in L_1(\Omega)$

$$\mathbb{E}[v(a)] = \int_\Omega v(a(\omega))\ \mathrm{d}\mathbb{P}(\omega) = \int_{-\infty}^{\infty} v(z) f_{\mathcal{LN}(0,1)}(z)\ \mathrm{d}z = \int_{-\infty}^{\infty} v(e^z) f_{\mathcal{N}(0,1)}(z)\ \mathrm{d}z,$$

with probability density function of the normal distribution

$$f_{\mathcal{N}(0,1)}(z) := \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \quad z \in \mathbb{R}.$$

By a suitable substitution we have

$$\mathbb{E}[v(a)] = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} v\left(e^{\sqrt{2}x}\right)\ e^{-x^2} \mathrm{d}x,$$

which can be efficiently approximated with a Gauss-Hermite quadrature. We have used a Gauss-Hermite quadrature of order 15 in our results. We have plotted the approximation of the $L_1(\Omega; \mathcal{X}_0)$-error of the examples above in Figure 6.3.6.

We can see that the error decreases with optimal order in all cases, cf. Theorem 2.4.12 and (2.4.13). This was expected from Corollary 5.3.16 for (6.3.2), (6.3.3) as well as for the log-normal case, since $\lim_{\beta\to\infty} \frac{2\alpha\beta\gamma}{4\alpha\beta+5\alpha\gamma+2\beta\gamma} = \frac{\alpha\gamma}{2\alpha+\gamma} = \bar{p}$ in Corollary 5.3.16. But, although the assumptions of Theorem 5.3.13, respectively Corollary 5.3.16, are not fulfilled for (6.3.4), we observe an optimal approximation rate as well. Note that we have already seen in Figure 6.2.6d that the $L_1(\Omega; \tilde{\mathcal{X}})$-norm of the solution (in second form) itself exists. But this does not immediately imply that also the Petrov-Galerkin solution is quasi-optimal with respect to the same norm. We cannot strictly prove why we have optimality here, but it is very likely that it is due to the fact that although $A_{\min}$ is not uniformly bounded, the ratio $\frac{A_{\max}}{A_{\min}}$ is uniformly bounded with this choice. A first approach in this way was done in [LMM16], see also Remark 5.2.10.

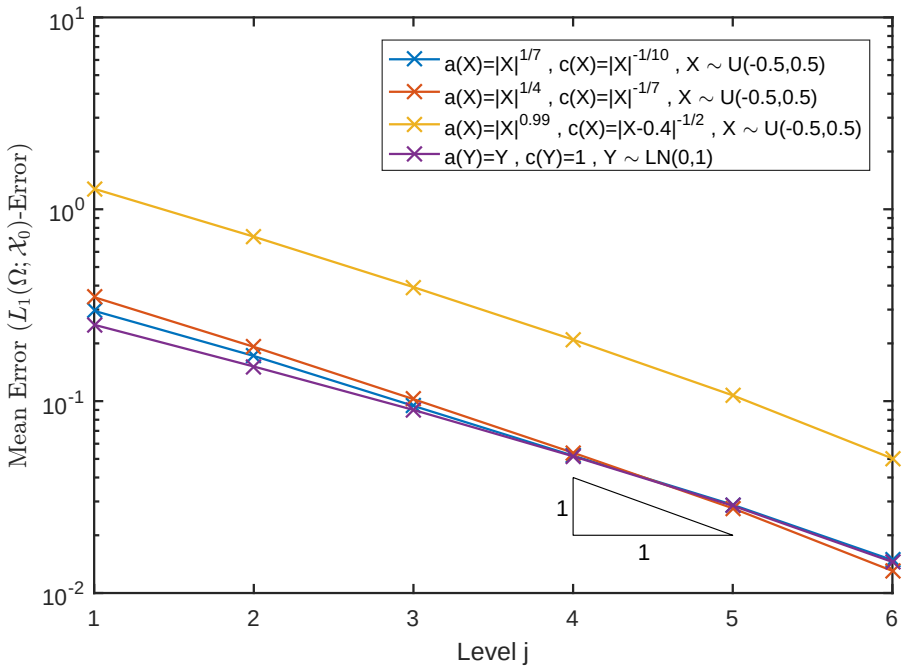



Figure 6.3.6: $L_1(\Omega;\mathcal{X}_0)$-error according to Example 6.3.1.

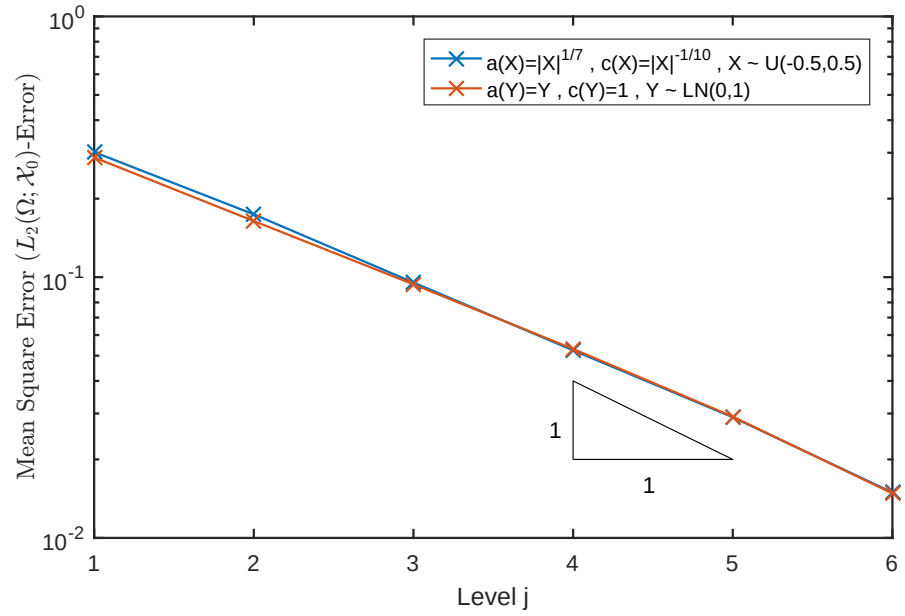



Figure 6.3.7: $L_2(\Omega;\mathcal{X}_0)$-error according to Example 6.3.1.

Additionally, we have calculated the error for the choices (6.3.2) and the log-normally distributed parameter in the $L_2(\Omega;\mathcal{X})$-norm, illustrated in Figure 6.3.7. As in the

Figure 6.3.6 we observe an optimal decay of the error. This is in compliance with Corollary 5.3.16.

## 6.4 Further Examples

We conclude this chapter with some numerical results which further illustrate the functionality of the B-spline code implemented for this work. First, we want to illustrate the well-posedness of the homogenization of a random PDE in first formulation introduced in section 3.2 and also show in this way that the code can treat non-zero initial conditions. To this end, consider the following problem in strong form.

Find a solution $u\colon [0,1]\times[0,1]\to\mathbb{R}$ such that

$$\begin{aligned} \tfrac{\partial}{\partial t}u(t,x) - \tfrac{\partial^2}{\partial x^2}u(t,x) &= x(x-1) - 2(1+t) && \text{for } (t,x)\in(0,1]\times(0,1),\\ u(0,x) &= x(x-1) && \text{for } x\in(0,1)\\ u(t,0) = u(t,1) &= 0 && \text{for } t\in(0,T]. \end{aligned}$$

The PDE is posed such that the exact solution is polynomial and explicitly given as

$$u(t,x) = (1+t)(x^2-x), \quad \text{for } (t,x)\in[0,1]\times[0,1].$$

Since the initial value is smooth enough, we can set $u_0^* := 1\otimes u_0$ in Proposition 3.2.7 and obtain the following homogenized first formulation.

**Example 6.4.1.** *Let $I := (0,1)$ and $D := (0,1)$. Find $u\in\mathcal{X}_0 = L_2(I;H_0^1(D))\cap H^1_{0,\{0\}}(I;H^{-1}(D))$ such that*

$$b_0(u,w) = \mathcal{F}_0(w) \qquad \textit{for all } w\in\mathcal{Y}_0 = L_2(I;H_0^1(D)),$$

*with*

$$\begin{aligned} b_0(v,w) &:= \int_I \langle\dot v(t), w(t)\rangle + \langle\nabla v(t), \nabla w(t)\rangle \,\mathrm{d}t\\ \tilde{\mathcal{F}}_0(w) &:= \int_I\int_D (x(x-1) - 2(1+t) + 2)w(t,x)\,\mathrm{d}x\,\mathrm{d}t. \end{aligned}$$

Apart from this homogenized form, we also consider the example in the second formulation

**Example 6.4.2.** *Let $I := (0,1)$ and $D := (0,1)$. Find $u\in\tilde{\mathcal{X}} = L_2(I;H_0^1(D))$ such that*

$$\tilde b(u,w) = \tilde{\mathcal{F}}(w) \qquad \textit{for all } w\in\tilde{\mathcal{Y}} = L_2(I;H_0^1(D))\cap H^1_{0,\{1\}}(I;H^{-1}(D)),$$

*with*

$$\begin{aligned}\tilde{b}(v,w) &:= \int_I -\langle v(t), \dot{w}(t)\rangle + \langle \nabla v(t), \nabla w(t)\rangle \, \mathrm{d}t \\ \tilde{\mathcal{F}}(w) &:= \int_I \int_D (x(x-1) - 2(1+t))w(t,x) \, \mathrm{d}x \, \mathrm{d}t, \\ &\quad + \int_D x(x-1)w(0,x) \, \mathrm{d}x.\end{aligned}$$

For the discretization we use B-splines of order two in time and space for the solution and test space in both cases, Example 6.4.1 and 6.4.2. We increase the level in the test space by one for stability reasons, i.e.,

$$S_{j,\mathbf{k}} = S^t_{j,2} \otimes S^x_{j,2}, \quad Q_{j+1,\mathbf{k}} = Q^t_{j+1,2} \otimes Q^x_{j+1,2}.$$

The results are shown in Figure 6.4.3.

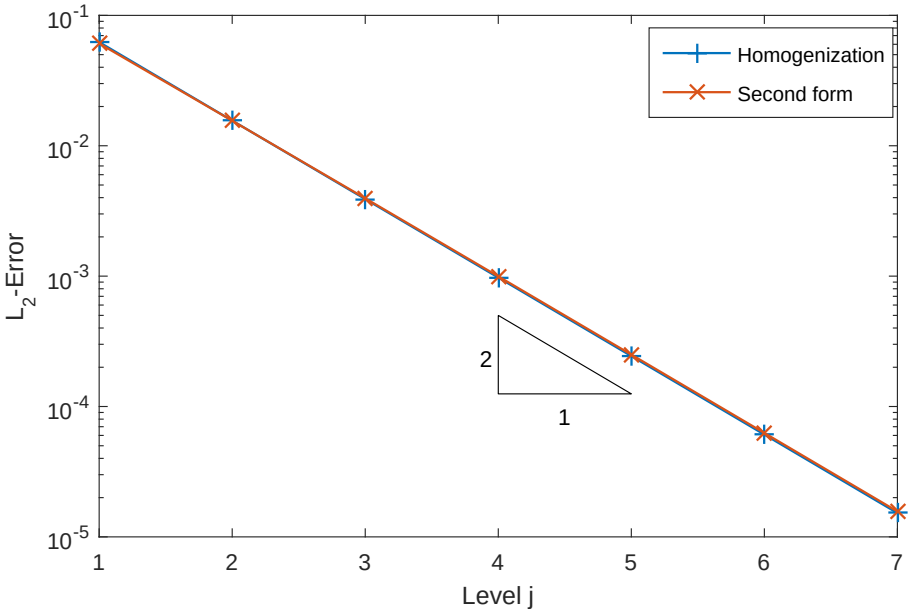



Figure 6.4.3: $L_2(I \times D)$-error for Example 6.4.1 and 6.4.2

First of all, one directly observes that the error decreases in both cases. Moreover, it can be seen that the rate of convergence is optimal according to Theorem 2.4.12 and that the particular error is also quantitatively almost the same.

Next, we also want to consider a three dimensional problem, that is, one dimensional in time and two dimensional in space, and for different order of splines. To this end, we consider the following problem in homogenized first formulation.

**Example 6.4.4.** *Let* $I := (0,1)$ *and* $D := (0,1)^2$. *Find* $u \in \mathcal{X}_0 = L_2(I; H_0^1(D)) \cap H^1_{0,\{0\}}(I; H^{-1}(D))$ *such that*

$$b_0(u,w) = \mathcal{F}_0(w) \qquad \text{for all } w \in \mathcal{Y}_0 = L_2(I; H_0^1(D)),$$

*with*

$$b_0(v,w) := \int_I \langle \dot{v}(t), w(t)\rangle + \langle \nabla v(t), \nabla w(t)\rangle \,\mathrm{d}t$$
$$\mathcal{F}_0(w) := \int_I \int_0^1 \int_0^1 (8\pi^2 t + 1)\sin(2\pi x)\sin(2\pi y)\; w(t,x,y)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}t.$$

Problem 6.4.4 is constructed such that the exact solution is given as

$$u(t,x,y) = t\sin(2\pi x)\sin(2\pi y).$$

We discretized the PDE with tensor product B-splines in space and time, where we fixed the minimum required order in the test space and successively increase the order in the solution space, that means

$$S_{j,k} = S^t_{j,k} \otimes S^x_{j,k} \otimes S^y_{j,k}, \quad Q_{j+1,\mathbf{k}} = Q^t_{j+1,1} \otimes Q^x_{j+1,2} \otimes Q^y_{j+1,2}.$$

This choice also demonstrates the universality of the implemented B-spline code, since we not only arranged the spaces in three dimensions, but also with different orders and levels of splines within the test as well as solution spaces themselves. It is worth mentioning that changing the order and levels of each part of the tensor products, can be managed simply with parameters in the code and is, therefore, very user friendly, see appendix A. The results are illustrated in Figure 6.4.5.

We can see that also in the three dimensional case, the error decreases as expected. Moreover, we can see that the approximation rate improves when increasing the order $k$, with fixed order in the test space. We can indeed observe that the rate is optimal according to Theorem 2.4.12.

We conclude this chapter with two examples which go beyond the scope of this thesis. The implementation is rather general such that it can also deal with different types of PDEs. We choose exemplary a Schrödinger type equation. This type of equation needs to handle complex values as well. Besides, the order of differentiation is the same as in the parabolic case and the way to derive a full space and time weak formulation seems to be straightforward. Although we are not able to prove existence and uniqueness in a full space-time weak formulation explicitly, the corresponding operator should be boundedly invertible at least when the existence and uniqueness of a strong solution is known. To this end, we consider a Schrödinger equation derived from a well-known physical problem, namely a free particle in an infinite potential well, where the solution can be calculated exactly. The position wave function of a particle in the ground state is given as the solution of the following Schrödinger equation.

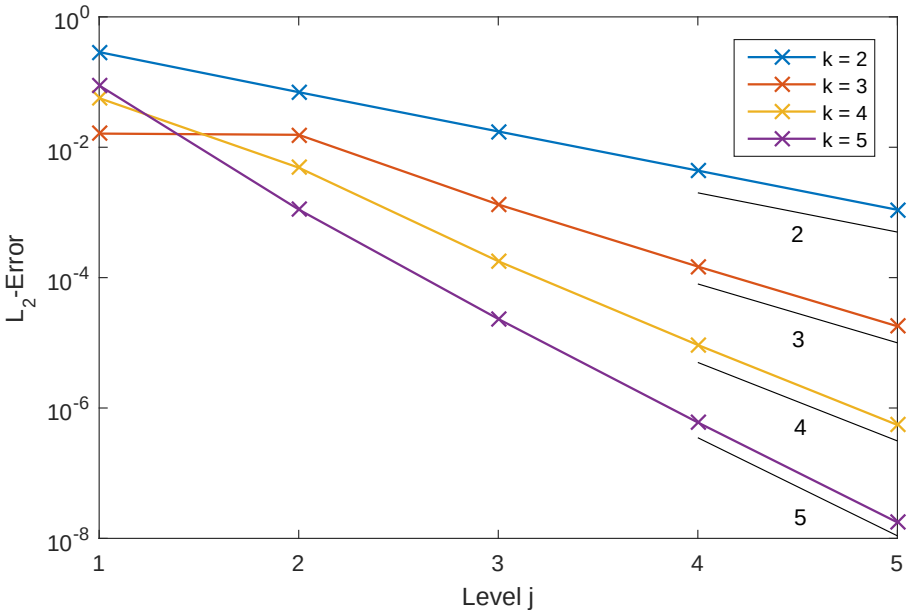



Figure 6.4.5: $L_2(I \times D)$-error for Example 6.4.4.

Let $I := (0, \infty)$ and $\tilde{D} := \mathbb{R}$. Find a solution $\psi$ such that

$$\begin{aligned} \mathrm{i}\hbar \frac{\partial}{\partial t}\psi(t,x) &= -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2}\psi(t,x) + V(x)\psi(t,x), \quad t \in I,\ x \in \tilde{D} \\ \psi(0,x) &= \sqrt{2}\sin(\pi x), \quad x \in \tilde{D}, \end{aligned}$$

where $\hbar$ is the reduced Planck constant, $m$ the mass of the particle and i the imaginary unit. The potential $V$ is defined as the infinite potential well

$$V(x) := \begin{cases} 0, & \text{for } 0 < x < 1, \\ \infty, & \text{otherwise} \end{cases}.$$

We restrict ourselves to electrons and restate the equation in *atomic units*, that is, $\hbar := 1$ and $m_e := 1$, as it is usual in quantum physics. Then we obtain the equivalent problem of finding $\psi$ in atomic units such that

$$\begin{aligned} \mathrm{i}\frac{\partial}{\partial t}\psi(t,x) &= -\frac{1}{2}\frac{\partial^2}{\partial x^2}\psi(t,x), \quad t \in I,\ x \in D := (0,1) \\ \psi(0,x) &= \sqrt{2}\sin(\pi x), \quad x \in D \\ \psi(t,0) = \psi(t,1) &= 0, \quad t \in I. \end{aligned}$$

The exact solution of the previous problem is given as

$$\psi(t,x) := \begin{cases} \sqrt{2}\sin(\pi x)\exp(-\frac{\pi^2\hbar}{2m}\mathrm{i}t), & \text{for } t \geq 0,\ 0 \leq x \leq 1, \\ 0, & \text{otherwise} \end{cases}$$

in SI units and

$$\psi(t,x) := \begin{cases} \sqrt{2}\sin(\pi x)\exp(-\frac{\pi^2}{2}\mathrm{i}t), & \text{for } t \geq 0,\ 0 \leq x \leq 1, \\ 0, & \text{otherwise} \end{cases}$$

in atomic units, respectively. We would like to note that the wave function of a free particle in an infinite potential well is usually stated without an initial condition, but is uniquely determined by the state of the particle and a normalization, since $|\psi|^2$ plays the role of a position density to locate the particle. With our (normalized) initial wave function $\sqrt{2}\sin(\pi x)$, we capture the ground state. If we restrict to a finite time interval $I := (0,1)$, for instance, we can formulate the problem in a (homogenized) first space-time weak formulation in a straightforward manner.

**Example 6.4.6.** *Let $I := (0,1)$ and $D := (0,1)$. Find $u \in \mathcal{X}_0 = L_2(I; H_0^1(D)) \cap H^1_{0,\{0\}}(I; H^{-1}(D))$ such that*

$$b_0(u,w) = \mathcal{F}_0(w) \qquad \textit{for all } w \in \mathcal{Y}_0 = L_2(I; H_0^1(D)),$$

*with*

$$b_0(v,w) := \int_I \mathrm{i}\langle \dot{v}(t), w(t)\rangle - \frac{1}{2}\langle \nabla v(t), \nabla w(t)\rangle \,\mathrm{d}t$$
$$\mathcal{F}_0(w) := \frac{1}{\sqrt{2}}\pi^2 \int_I \int_D \sin(\pi x)\cdot w(t,x)\,\mathrm{d}x.$$

We have chosen the same solution and test spaces as in the parabolic case, although we might generally need to assume stronger smoothness assumptions to obtain existence and uniqueness. As done several times before, we discretize Problem 6.4.6 with B-splines of fixed order two both in space and time, and for the solution as well as test space. Moreover, we increase the resolution in the test space by one extra level exactly as discussed for the parabolic setting. The $L_2(I \times D)$-error is plotted in Figure 6.4.7.

One can observe that, indeed, the optimal rate of convergence is achieved. Although we are not able to prove it, we expected such a behavior due to the similarity to the parabolic framework. Nevertheless, one should have in mind that the behavior of the solution is different from the parabolic nature, since a wave function $\psi$ behaves like both, a particle *and* a wave and is of course also complex valued. This is due to the wave-particle dualism in quantum physics. For basics on quantum mechanics we refer to standard books on theoretical physics such as [Nol04, Nol12].

The second and last example is a collocation method to numerically solve a PDE. It should rather serve as an illustration of the broad functionality of the implementation, than as a theoretical treatment. Collocation methods with splines of higher order have

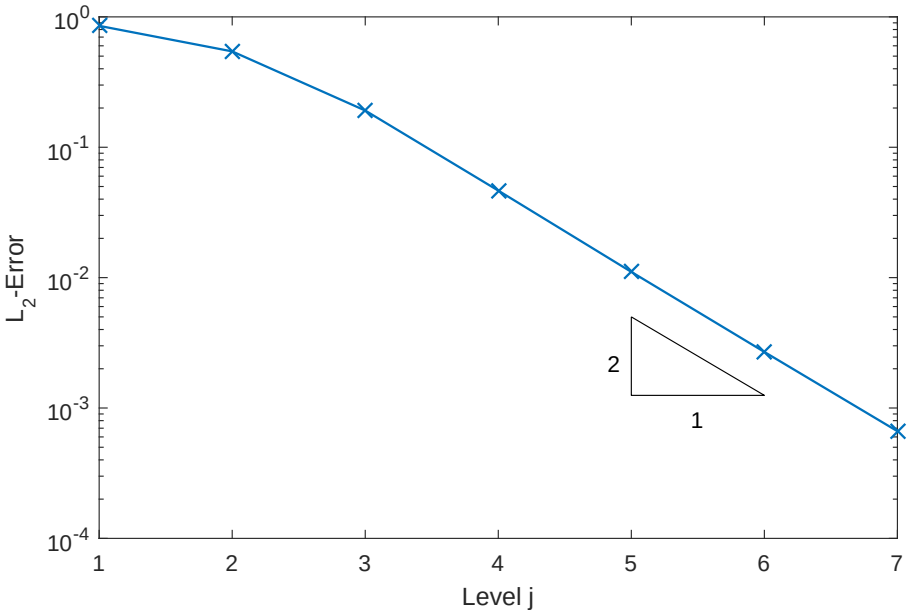



Figure 6.4.7: $L_2(I \times D)$-error for Example 6.4.6.

become more and more relevant in the last years in isogeometric analysis. For (Petrov-Galerkin) isogeometric analysis of parabolic problems in space-time weak formulation, we would like to highlight [LMN15] again.

As a model problem, we consider an example similar to Example 6.4.4. In strong form, for given time interval $I := (0, 1]$ and spatial domain $D := (0, 1)$, we want to find a solution $u \colon I \times \bar{D} \to \mathbb{R}$ such that

$$\begin{aligned} \frac{\partial}{\partial t} u(t,x) - \frac{\partial^2}{\partial x^2} u(t,x) &= (1 + 4\pi^2 t)\sin(2\pi x) =: g(t,x), && (t,x) \in I \times D \\ u(0,x) &= 0, && x \in D \\ u(t,0) = u(t,1) &= 0, && t \in I. \end{aligned}$$

The exact solution is given explicitly as

$$u(t,x) = t \sin(2\pi x). \tag{6.4.8}$$

Our collocation approach now reads as:

**Example 6.4.9.** *Let $I := (0, 1]$ and $D := (0, 1)$. Find $U_{j,k} \in (S^t_{j,k})_{0,\{0\}} \otimes (S^x_{j,k})_0$ such that*

$$\mathcal{L} U_{j,k}(t_{i_1}, x_{i_2}) = g(t_{i_1}, x_{i_2}), \quad i_s = 1, \dots, M_{i_s},\ s = 1, 2,$$

*for given collocation points $(t_{i_1}, x_{i_2}) \in I \times D$ and $M_{i_1},\ M_{i_2} \in \mathbb{N}$, with differential operator $\mathcal{L} := \frac{\partial}{\partial t} - \frac{\partial^2}{\partial x^2}$ and right hand side $g(t,x) := (1 + 4\pi^2)\sin(2\pi x)$.*

As collocation points we set

$$t_i = \frac{\theta_{i+1} + \cdots + \theta_{i+k-1}}{k-1}, \quad i = 2, \ldots, N,$$
$$x_i = \frac{\theta_{i+1} + \cdots + \theta_{i+k-1}}{k-1}, \quad i = 2, \ldots, N-1,$$

where we have chosen, only for simplicity, $\dim S^t_{j,k} = \dim S^x_{j,k} = N \in \mathbb{N}$. These collocation points are chosen in each coordinate direction as Greville abscissae from [EHR+13, eq. (25)]. The solution of Example 6.4.9 is calculated numerically by solving the corresponding matrix vector equation of Kronecker products of matrices consisting of point evaluation of derivatives of B-splines. This is done efficiently with the de Boor scheme described in Theorem 2.4.5 implemented in `Nev`. For our examples we have chosen different orders of splines $k = 3, \ldots, 7$ and calculated the $L_2(I \times D)$-error with respect to the known exact solution (6.4.8). The error is illustrated in Figure 6.4.10.

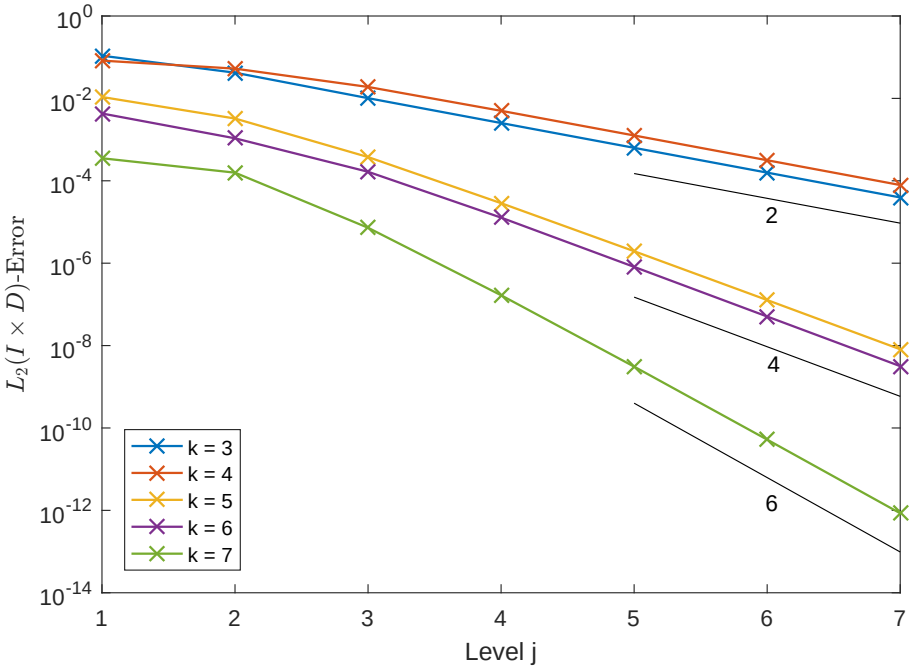



Figure 6.4.10: $L_2(I \times D)$-error for Example 6.4.9.

One can observe that the implemented code also works well for collocation methods. Approximations with multidimensional collocation methods is still an open area of research and go much beyond the scope of this thesis. Therefore, we will only give a brief description of what can be seen in the plot. Other than in the Petrov-Galerkin cases before, only even approximation orders appear. Precisely, splines of order 3 and 4 approximate with an order of 2, splines of order 5 and 6 with an order of 4 and splines

of order 7 with an order of 6. So one approximation order seems to be skipped when increasing the order of the splines. Moreover, one also does not reach the optimal order of Theorem 2.4.12. Anyway, from a computational point of view, one should have in mind that assembling a collocation matrix is much faster than assembling a Petrov-Galerkin stiffness matrix consisting of inner products.

# 7 Conclusion and Outlook

## 7.1 Conclusion

I have presented in this thesis a novel stability analysis of (minimal residual) Petrov-Galerkin approaches for space-time formulations of parabolic PDEs as well as an extension to random PDEs with possibly non-coercive spatial differential operators. Starting from a generic parabolic equation, two different space-time weak formulations were considered, as well as a homogenization. The basic idea was to treat the time variable in the same way as the spatial variables and, consequently, choose space-time test functions in order to derive a full weak formulation. With this approach, a single operator equation was derived. With the Banach-Nečas-Babuška Theorem as the main tool, existence and uniqueness was ensured with explicit estimates of the space-time operator and its inverse. These estimates have turned out to be crucial for the treatment of random PDEs. A summary of all lower and upper estimates of the space-time operator for the different types of formulations and for different types of spatial differential operators was presented in Table 3.3.17. It was also shown in Theorem 3.2.9 and Corollary 3.3.10 that the spatial regularity is transfered to full space-time weak parabolic problems, at least for time-independent spatial differential operators.

After having ensured bounded invertibility of the space-time operator coming from full weak formulations of parabolic PDEs, the next step was to consider their numerical approximation. The main issue was to construct stable discretizations. To find such a stable discretization is by no means a trivial task, since the inf-sup condition, which is required in the Banach-Nečas-Babuška Theorem, is not inherited by discrete subspaces. This means, one has to prove the validity of a *discrete* inf-sup condition, even if the inf-sup condition is known to hold for the *non-discrete* problem. A way to get unconditionally stable discretizations considered in this thesis, was to allow the test space to have a higher resolution. In order to obtain stable spaces for preferably wide ranges of functions, I kept the prerequisites on a rather moderate level. The main stability result was given in Theorem 4.1.8, showing how to construct stable discretizations depending mainly on approximation and regularity properties of the subspaces as well as the regularity of the operator. An abstract but explicitly given number of additional degrees of freedom in the test space was derived in order to ensure stability. This very general statement was applied to the second space-time formulation of parabolic PDEs. A construction rule for stable subspaces for such space-time formulations of parabolic PDEs was derived. Indeed, a recipe only relying on standard Bernstein and Jackson estimates on Sobolev spaces with respect to the temporal and spatial component was given, allowing for rather broad discretizations.

The next step was to extend the parabolic problems considered so far to parabolic problems with random coefficients. A similar space-time weak formulation was given, which immediately led to almost sure existence and uniqueness of a solution by applying

the deterministic results pathwise. The key idea in the context of the existence of moments of solutions was to treat the lower and upper bound of the spatial differential operator as random variables and not as strict uniform constants. By reusing the estimates of the space-time operator worked out in the deterministic case, a regularity result with respect to the random parameter was deduced. Indeed, it turned out that the solution has a certain number of finite moments, even if the spatial differential operator is not uniformly bounded from above or below. The number of existing moments of the solution depending on the number of finite moments of the lower and upper bounds of the spatial differential operator as well as the initial condition and right hand side was derived for different types of spatial differential operators, summarized in Table 5.1.21, 5.1.22, and 5.1.23.

In a similar fashion, quasi-optimality of Petrov-Galerkin approaches were shown for a construction of subspaces independent of the random parameter itself. In Theorem 5.3.13 a possible error measure ensuring quasi-optimality with respect to both, the deterministic variables as well as the random parameter, was given. Again, spatial differential operators which are not necessarily uniformly bounded can be treated too.

All the theoretical results from the stability of Petrov-Galerkin discretizations of deterministic parabolic PDEs, over the existence of moments of the solution of their random counterparts, to the quasi-optimality of their Petrov-Galerkin approximation were also illustrated numerically. The calculations were performed with a Matlab code implemented for this thesis, designed for Petrov-Galerkin approaches with B-splines. I have sketched the broad functionality of the implementation by two examples beyond the scope of parabolic Petrov-Galerkin discretizations.

## 7.2 Outlook

Concerning the stability itself, one could try to construct more classes of suitable subspaces in the spirit of [And13] treating also the second formulation. An adaptation to the second form cannot be done straightforwardly, since, in contrast to the first form, one has to enrich the test space $\tilde{\mathcal{Y}}$, containing temporal derivatives. One can, however, exploit the idea of subspace dependent norms from [And13, UP12] to work out stable subspaces for the second formulation. In this way, one can prove stability with respect to the second formulation when a CFL condition is fulfilled, see [LMM16].

Motivated by the physical model of excitons in semiconductor quantum wires described in [MKM13] and detailed in [Mol11], one could formulate Schrödinger-type equations in a full weak space-time formulation. This can be done heuristically, but I am not aware of works proving bounded invertibility of the resulting space-time operator. Nevertheless, I would expect a similar behavior at least under slightly changed regularity assumptions. First simple numerical examples underline these expectations as shown in Figure 6.4.7.

As the numerical results and Example 5.1.19, for instance, already suggest, the conditions concerning the existence of moments are not sharp, but give general worst-case conditions on $A_{\min}$ and $A_{\max}$ as well as the input data. This is, of course, not surprising, but one could try to focus on operators $A(t,\omega)$, where $A_{\min}$ and $A_{\max}$ behave in a way that $\frac{A_{\max}}{A_{\min}} \leq C < \infty$ with a constant $C$ which does *not* depend on $\omega$. That means, one could try to analyze the particular behavior of spatial differential operators which are not necessarily uniformly bounded from above and below, but where the ratio $\frac{A_{\max}}{A_{\min}}$ is uniformly bounded. There are first results in this direction which suggest that if $p$ moments exist for the solution and its approximation, that is, $U,\ U_j \in L_p(\Omega;\tilde{\mathcal{X}})$, then we also obtain quasi-optimality in the same metric without any further requirements. I refer to [LMM16] and Remark 5.2.10 for details.

The thesis is rather theoretically motivated and not focused on developing novel numerical solution methods. Although I have implemented an efficient and very general solution routine with B-splines, there are, of course, problem adapted numerical methods which are much more efficient for particular settings. A sparse-grid implementation of parabolic problems in first form was described, for instance, in [And14]. Since the code deals with hierarchical B-splines, one could make use of it and implement a *multigrid* solver for our particular problems, cf. [TOS01] for details on multigrid methods. I have already implemented a very first attempt into this direction. Moreover, the program could be expanded to spline *wavelet methods* as well. I refer to [DKU99] for the construction of spline wavelets on the interval. The stability result from chapter 4.2 was worked out such that it can be applied to wavelet type bases. This was indeed one of the main motivations to distinguish between primal and dual bases. Another advantage would be that the inverse Gram matrix in (4.1.14) can be replaced by a simple diagonal scaling due to norm equivalences of wavelets. Moreover, beside, constructing different bases like wavelet type bases, one could also extend the utilities of the existing implementation. For instance the user interface could be generalized to treat other kinds of PDEs and, beside user specified right hand sides, also non-zero initial conditions as well as parameter functions in the (spatial differential) operator itself.

I also have *adaptive* wavelet methods in mind. Since most adaptive wavelet methods are constructed as perturbed methods applied to an infinite equivalent problem, one usually has a build-in stability inherited from the infinite dimensional spaces. Such adaptive methods were considered, for example, by Cohen, Dahmen and DeVore in [CDD01, CDD02, CDD03a, CDD03b] and also in [SS09] for linear problems in the context of full space-time weak formulations. Having spline wavelets and also nonlinear problems in mind, one has the efficient inexact operator application from [MP13] at hand. Although the theoretical background was given in [SS09], to my knowledge, there is no rigorous implementation available, at least for nonlinear problems. There is a novel adaptive wavelet code written in C++ available, which can deal with nonlinear problems in arbitrary dimensions using spline wavelets. The implementation is based on the PhD thesis [Pab15] of Roland Pabel.

The methods mentioned before are solely deterministic methods which apply pathwise to random PDEs. But there are more sophisticated methods to treat the parameter dependence such as *reduced basis methods*, instead of only solving the equations pathwise in combination with a Monte-Carlo or quadrature method, for instance. Parametric PDEs would basically also fit into the context of operator equations with random coefficients, since the latter can be expressed as a deterministic parameter dependent equation via an (infinite) Karhunen-Loève expansion. An overview of approximation methods for parameter dependent PDEs is given in [CD15]. I also refer to [UP14] for reduced basis methods for parabolic problems in (first) space-time weak formulation. In [DPW14] a stabilization of the subspaces was introduced, by connecting the pathwise stability with the parameter dependent stability to some extend. One central point in [DPW14] is to relate a Petrov-Galerkin scheme to an *equivalent saddle-point problem*. This equivalent saddle-point problem gives rise to so called $\delta$-proximal (finite) test spaces ensuring discrete inf-sup stability. Similar to our general stability result in Theorem 4.2.10 the approximation is stabilized by enlarging the test space, but the spaces are not a-priori fixed and are nonstandard spaces depending directly on the operator. Nevertheless, one could try to develop a mutual stability analysis with respect to both, the sample space and the deterministic space, instead of a pathwise deterministic treatment.

The analysis of the influence of random parameters would also be interesting for other types of problems like control problems. The well posedness of linear-quadratic control problems with parabolic PDE constraints formulated in a full space-time weak form was already proven, e.g., in [GK11, KM15, KS13, KS15]. By introducing the Karush-Kuhn-Tucker system, one can consider linear-quadratic control problems as *one* operator equation. Therefore, one could try to work with such a formulation of control problems with random parameters in a similar way as done in this thesis to prove existence of moments of the solution of control problems.

# A Code Documentation

This chapter provides a brief overview of the structure of the Matlab-package, which was implemented in connection with this thesis. We will present the main functionality of the code and its structure. For details on the implementation of each routine, we refer to the code itself, which is accompanied and includes detailed comments. For the mathematical basics of the used properties, definitions, and schemes, we refer to section 2.4. The implementation predominantly relies on the recursive definition of B-splines, its definition of derivatives from Corollary 2.4.4, the point evaluation de Boor scheme from Theorem 2.4.5 in arbitrary dimensions, quadrature rules, and tensor products. This means, section 2.4 is completely sufficient in order to understand how the code works. For more details we again refer to [Sch07, dB01]. All implementations are tested on Matlab versions 2015b and 2016a.

The implementation aims at having a rather broad functionality. It allows to discretize with tensor product B-spline bases $SP^n_{\Delta_T,\mathbf{k}}$ according to (2.4.8) with arbitrary order in each coordinate direction $\mathbf{k} = (k_1, \dots, k_n)$ and array of extended knot sequences $T = (T_1, \dots, T_n)$. Although section 2.4 is restricted to simple knots as in Proposition 2.4.2 for simplicity, most implementations can handle more general non-uniform knot sequences with coinciding inner knots as well. Moreover, the discretization in solution and test space can be chosen independently of each other. This functionality is essential in order to handle stable subspaces according to chapter 4. All bilinear forms of the form

$$\langle v^{(\boldsymbol{\alpha})}, q^{(\boldsymbol{\beta})} \rangle, \quad \text{with multiindices } \boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_n),\ \boldsymbol{\beta} = (\beta_1, \dots, \beta_n)$$

can be discretized directly, where $v^{(\boldsymbol{\alpha})}$ and $q^{(\boldsymbol{\beta})}$ denote the derivatives in several dimensions according to Definition 2.2.1. Due to the tensor product structure, the discretization of such kind of bilinear forms decomposes into Kronecker products of system matrices in one coordinate direction. Since B-splines are polynomial between two knots, the entries are calculated exactly with piecewise Gauss-Legendre quadrature of sufficiently high order, where the required point evaluations are performed with de Boor's algorithm from Theorem 2.4.5.

The main functions are implemented to be user friendly in a way that, for example, the suitable quadrature order is chosen automatically and the assembly of the Kronecker products only require the spline and derivative order. There is even a graphical user interface implemented to treat parabolic problems in any of the three formulations introduced in this thesis with arbitrary right hand side in tensor format. In addition to the `StiffGen` routine which explicitly sets up the stiffness matrices, there are also some more functions to simplify the handling of the code, as, for example, `InitUniNodes` which generates a uniformly distributed extended knot sequence of arbitrary order and grid size. The code is arranged to rely on cell arrays in order to treat the different dimensions. In this way, the required knot sequences defining the B-splines in each coordinate direction can be passed to the functions quite easily and component-wise.

If no cells are necessary, but a vector suffices even for the most arbitrary consideration, as it is the case for the different orders of B-splines, then vectors are used instead. The data format of the input and output is specified for each function in this chapter.

We organize the documentation from top to bottom, orientating towards the functions called by the example scripts collected in the subfolder `/Examples` as good as possible. This does not necessarily mean that the most important routines are presented first. But it is the most convenient way for a user to see how the code is structured and how one could write own codes by making use of the routines from this package. The precalculated data is collected in the `/Examples/Data` folder and its subfolders.

- `Main_UI`: This script yields a graphical user interface (GUI) to construct and solve parabolic problems in first formulation, its homogenization or second formulation. The user can specify the following:
  - Dimension of the problem,
  - spatial boundary conditions for the test and solution space (the temporal boundary conditions are specified automatically depending on the type of formulation),
  - order of B-splines for test and solution space independent in each variable,
  - discretization level for test and solution space independent in each variable,
  - number of term of the PDE,
  - order of differentiation w.r.t. test and solution space of the spatial differential operator in each term,
  - regularity of the spatial operator,
  - arbitrary right hand side as tensor product,
  - plot resolution.

  Depending on the user specified input, the script calls the corresponding subroutines automatically and solves the PDE numerically with a CG-method `Normal_CG`.
- `Examples/ExampleName`: The Examples used for the numerical results from chapter 6 are collected in the subfolder `Examples/`. Similar to `Main_UI` these scripts call the routines specified in the particular example.
- `InitUniNodes`: Used to generate a uniformly distributed extended knot sequence for B-spline bases.
  Input:
  - Refinement level `j`,
  - order of B-spline `k`.

OUTPUT:

- Corresponding extended set of knots `T`.

- `StiffGen`: Helpfunction to initialize sums of Kronecker product matrices automatically. It builds a multidimensional system matrix containing different sums of Kronecker product matrices. The stiffness matrices w.r.t. one coordinate direction are assembled with `StiffMat`. It can only assemble matrices without coefficients/coefficient functions in this form.
  INPUT:
  - Cell array `Tsol` of extended knots for B-spline bases with respect to solution space; each component stands for the coordinate direction, e.g., `Tsol`$\{1\}$ for the knots of the B-spline basis in the first variable in the solution space,
  - cell array `Ttest` of extended knots for B-spline bases with respect to test space test space,
  - vector of order `ksol` of B-splines in the solution space; each component stands for the order of one-dimensional spline in a coordinate direction,
  - vector of order `ktest` of B-splines in the test space,
  - a cell matrix `Matind` which keeps the information about each stiffness matrix; the information is encoded in a cell matrix with rows representing each coordinate direction and columns representing the term in the sum of matrices. The entries itself consist of two values, namely the order of differentiation in the solution and test space, i.e., `Matind`(i,k)= $(d_1, d_2)$ corresponds to $\langle \phi_{i,1}^{(d_1)}, \phi_{i,2}^{(d_2)} \rangle$, the stiffness matrix in the $i$-th variable and the $k$-th term in the sum,
  - classification of boundary conditions of the solution space indicated by `offsetsol`= $(bs_1, bs_2)$; first value classifies the left boundary offset and second value the right boundary offset, where a value 1 means zero boundary condition and a value of 0 naturally boundary condition,
  - classification of boundary conditions of the test space indicated by `offsettest`,
  - an *optional* flag `secForm` to set up discretizations of parabolic PDEs in second form; activated flag `secForm`= 1 includes a factor $(-1)$ in the first term.

  OUTPUT:
  - System matrix `B` arranged according to the inputs specified above.

- `RieszGen`: Function to automatically assemble discrete Riesz matrices, i.e., Gram matrices. It covers intersections of tensor product spaces.
  INPUT:

  - Cell array `T` of extended knots for B-spline bases,
  - vector of order `k` of B-splines,
  - classification of boundary conditions indicated by `offset`; cf. `StiffGen`,
  - matrix `Matind` with information about the structure of spaces; `Matind`$(d,t)$ encodes the order of Sobolev space w.r.t. to $d$-th coordinate variable and $t$-th intersection, e.g., `Matind`$= \begin{pmatrix} 0 & 1 \\ 2 & 0 \end{pmatrix}$, yields the Gram matrix w.r.t. $L_2 \otimes H^2 \cap H^1 \otimes L_2$.

  OUTPUT:

  - Gram matrix `R` arranged according to the inputs as specified above.

- `Ndiff`: Helpfunction for point evaluations of derivatives of B-splines. That means, it calculates the value $N_{i,k}^{(q)}(x_j)$ for any order of derivative $q$ of a B-spline $N_{i,k}$ at given point $x_j$.
  INPUT:

  - Expanded knot sequence `T` for the B-spline basis,
  - order `k` of B-spline,
  - order of differentiation `diff`,
  - position index `i` of the B-spline,
  - evaluation point `t`.

  OUTPUT:

  - Point value `Nx` of $N_{i,k}^{(q)}(x_j)$.

- `Normal_CG`: Algorithm to solve the minimal residual Petrov-Galerkin problem (4.1.5) via the Gauss normal type equation (4.1.14) using conjugate gradient (CG) method. This is a standard CG-method where the inverse Gram matrix $\mathbf{R}_Y^{-1}$ is not calculated exactly, but a system of equation is solved in each step. In this way one can save massively memory and also speed up the computation in comarison with a naive implementation. Nevertheless, notice that the system is *not* preconditioned and therefore it might be slow for high resolutions.
  INPUT:

  - System matrix `B`,
  - Gram matrix `RY`,
  - right hand side `f`,
  - maximum number of iterations `kmax`,
  - desired tolerance `tol` of the relative residual measured in the $\ell_2$-norm,

– initial vector `x`.

OUTPUT:

– Solution `x`,
– number of required iterations `counter`,
– relative residual of the final output `res`.

- `Plot_Spline`: Routine for plotting a tensor product spline of the form

$$S(x_1, \ldots, x_n) = \sum_{i_1=1}^{N_1} \cdots \sum_{i_n=1}^{N_n} c_{i_1,\ldots,i_n} \; N_{i_1,k_1}(x_1) \cdots N_{i_n,k_n}(x_n),$$

given by the expansion vector of coefficients $c_{i_1,\ldots,i_n}$ in lexicographical order and B-spline bases $N_{i_j,k_j}$ for $i_j = 1, \ldots, N_j$ with $j = 1, \ldots, n$, for each coordinate. The dimension $n$ is restricted to maximum three in order to produce a reasonable graphical output. Dimension one yields a two dimensional plot of the function, dimension two a three dimensional plot and dimension three a movie, where its first variable is taken as time variable. The routine makes use of the multidimensional evaluation scheme, cf. Theorem 2.4.5 and (2.4.9). The dimension does not need to be specified, but will be determined by the routine automatically depending on the input data. The plot resolution can be chosen independently of the refinement level of the B-splines.
INPUT:

– Expansion coefficient vector `c` according to $c_{i_1,\ldots,i_n}$ in lexicographical order,
– set of nodes `T` of B-splines in each coordinate given in a Matlab cell format,
– oder `k`=$(k_1, \ldots, k_n)$ of B-splines in each coordinate,
– level `lev`= $(l_1, \ldots, l_n)$ of uniform grid in each direction to set the plot resolution; level, e.g., $l_1 = 5$ and $l_2 = 6$ would result on a uniform grid of grid size $2^{-5}$ in the first coordinate and $2^{-6}$ in the second coordinate.

OUTPUT:

– Function values `Val` as vector, matrix or movie frames depending on the dimension.

- `StiffMat`: Function to assemble system matrices of Petrov-Galerkin discretization with B-splines on unit interval $[0, 1]$. Works for uniform or non-uniform grids as well as for extended knot sequences with coinciding inner knots.
INPUT:

– Set of extended knots `Tsol` and `Ttest` for the B-spline basis with respect to the solution and test space,

  - order `ksol` and `ktest` of B-splines in the solution and test space,
  - classification `diff` of derivatives; `diff`= $(d_1, d_2)$ yields system matrix corresponding to $\langle \phi_1^{(d_1)}, \phi_2^{(d_2)} \rangle$, e.g., `diff`=$(0,0)$ for mass matrix and `diff`= $(1,1)$ for Laplacian,
  - classification of boundary conditions indicated by `offsetsol`= $(bs_1, bs_2)$ and `offsettest`= $(bt_1, bt_2)$; first value classifies the left boundary offset and second value the right boundary offset, where a value 1 means zero boundary condition and a value of 0 naturally boundary condition.

  OUTPUT:

  - System matrix `A` according to the input, cf. `diff`.

- `ColMat`: Function to set up system matrices of a collocation method with B-splines on unit interval $[0, 1]$.
  INPUT:

  - Extended knot sequence `T` for the B-spline basis,
  - order `k` of B-spline,
  - classification of boundary conditions indicated by `offset`= $(bs_1, bs_2)$; first value classifies the left boundary offset and second value the right boundary offset, where a value 1 means zero boundary condition and a value of 0 naturally boundary condition,
  - collocation points `x`,
  - classification `diff` of derivatives; `diff`= $(d_1, d_2)$ yields system matrix corresponding to $\langle \phi_1^{(d_1)}, \phi_2^{(d_2)} \rangle$, e.g., `diff`=$(0,0)$ for mass matrix and `diff`= $(1,1)$ for Laplacian.

  OUTPUT:

  - Collocation matrix according to the input.

- `RieszMat`: Function to set up discretizations of Riesz mappings. The function can handle Riesz mappings $R_{H^m} \colon H^m \to \dot{H}^{-m}$ (with possibly additional boundary conditions) but also $R_{\dot{H}^{-m}} \colon \dot{H}^{-m} \to H^m$. $R_{H^m}$ results in a Gram matrix according to $(\cdot, \cdot)_{H^m}$ as a Petrov-Galerkin discretization with B-splines on unit interval $[0, 1]$, whereas $R_{\dot{H}^{-m}}$ results in a discretization according to (6.1.17). Since Gram matrices are particular system matrices, they are assembled by using the routine `StiffMat` with specified input.
  INPUT:

  - Set of extended knots `T` for the B-spline basis,
  - order `k` of B-splines,

– classification of boundary conditions `offset`, see `StiffMat`,

– order `order` of Sobolev space; positive value yields corresponding Gram matrix and negative value a matrix according to (6.1.17).

OUTPUT:

– Gram matrix or matrix according to (6.1.17), respectively, for the specified basis and Riesz mapping $R_{H^{\texttt{order}}(0,1)}\colon H^{\texttt{order}}(0,1) \to (H^{\texttt{order}}(0,1))'$.

- `Nev`: Performs a spline evaluation at a point `x` according to a multidimensional version of the Neville-like de Boor scheme from Theorem 2.4.5, see also (2.4.9) and the explanations there. That means, given a tensorized spline

$$S(\mathbf{x}) = \sum_{i_1=1}^{N_1} \cdots \sum_{i_d=1}^{N_n} c_{i_1,\ldots,i_n} N_{i_1,k_1}(x_1) \ldots N_{i_n,k_n}(x_n),$$

the routine yields the value of $S(\bar{\mathbf{x}})$ at a given point $\bar{\mathbf{x}}$. This is one of the most important schemes since almost everything relies on these point evaluations.
INPUT:

– Multidimensional vectorized expansion coefficient `c` in lexicographical order,

– cell array of expanded knot sequences `T` for the B-spline bases; each component corresponds to one coordinate direction, e.g., `T{1}` is the expanded knot sequence w.r.t. the B-spline basis in the first variable,

– vector `k` of B-spline order for each coordinate direction,

– evaluation point `x`$= (x_1, \ldots, x_n)$.

OUTPUT:

– Value `Val` of the given multidimensional spline classified by the inputs at point `x`.

- `Utilities/Gaussq`: Performs a Gauss-Legendre quadrature of arbitrary order $k \leq 5$. A Gauss quadrature is known to be exact for polynomials up to order $2k$. It is used for, e.g., evaluating the inner products in the stiffness matrices and right hand side. With smooth coefficient functions, stiffness matrices are assembled *exactly* for B-splines up to order $k_1$ and $k_2$ with $k_1 k_2 \leq 10$ by using this routine piecewise, where $k_1$ denotes the order of B-spline in the solution space and $k_2$ in the test space. But the Gauss-quadrature also yields a very accurate approximation for $k_1 k_2 > 10$. It can also be called in higher space dimensions. This is useful, e.g., for calculating norms of multidimensional splines.
INPUT:

  - Left integration limit (vector) `a`; e.g., `a`(1) corresponds to the first coordinate direction,
  - right integration limit (vector) `b`,
  - integrand `f` given as a Matlab function handle; it needs to be in a nested format, i.e., @($y$)@($x$)$g$,
  - quadrature order `k`.

  OUTPUT:

  - Gauss-Legendre approximation `val` of $\int_a^b f(\mathbf{x})\mathrm{d}\mathbf{x}$, with $a$=`a`, $b$=`a` and $f$ given by `f`.

- `RefineMat`: Assembles a refinement matrix $M_j$ on level $j$ according to Proposition 2.4.7. It only works on a uniform grid.
  INPUT:

  - Level `lev` of the B-spline basis set which shall be refined,
  - order `k` of the B-spline basis,
  - (*optional*) boundary conditions indicated by the vector `offset`; first value classifies the left boundary offset and second value the right boundary offset, where a value 1 means zero boundary condition and a value of 0 naturally boundary condition.

  OUTPUT:

  - (Rectangular) refinement matrix `M`.

- `Utilities/GaussHermite`: Helpfunction to calculate a Gauss-Hermite quadrature of order 13, 14 or 15 for given evaluations at the corresponding Gauss points. The quadrature is adapted to weight function $w(x) := \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right)$ according to normal distribution.
  INPUT:

  - Values `f` given as a vector of function values at Gauss-Hermite points depending on the order,
  - order of quadrature `order`.

  OUTPUT:

  - Gauss-Hermite approximation `val` via the weighted sum $\sum_{i=1}^{\texttt{order}} \alpha_i \texttt{f}_i$, with (adapted) weights $\alpha_i$.

- `Utilities/GaussLegendre`: Helpfunction to calculate a Gauss-Legendre quadrature of order 4 for given evaluations at the corresponding Gauss points. The quadrature is adapted to arbitrary finite intervals $[a, b]$.
  INPUT:
  - Values `f` given as a vector of function values at Gauss-Legendre points of order 4,
  - left boundary `a` of the interval,
  - right boundary `b` of the interval.

  OUTPUT:
  - Gauss-Legendre approximation `val` via the weighted sum $\sum_{i=1}^{\texttt{order}} \alpha_i \texttt{f}_i$, with (adapted) weights $\alpha_i$.
- `Utilities/HelpLegendrePoints`: Helpfunction to allocate Gauss-Legendre points of order 4 on each subinterval of a uniform grid with given grid size.
  INPUT:
  - Level of refinement `ref` corresponding to grid size $2^{-\texttt{ref}}$,
  - left boundary `l` of the interval,
  - right boundary `r` of the interval.

  OUTPUT:
  - Gauss-Legendre points `points` on all subintervals $[\texttt{l}, 2^{-\texttt{ref}}\texttt{l}] \cup \ldots \cup [\texttt{r} - 2^{-\texttt{ref}}, \texttt{r}]$.
- `Utilities/HelpGauss`: Helpfunction to calculate a Gauss-Legendre quadrature for given evaluations at corresponding Gauss points at all subintervals of the form as in `Utilities/HelpLegendrePoints`.
  INPUT:
  - Values `u` given as a vector of function values at all Gauss-Legendre points according to `Utilities/HelpLegendrePoints`,
  - level of refinement `ref` of the underlying grid,
  - left boundary of the interval `l`,
  - right boundary of the interval `r`.

  OUTPUT:
  - Gauss-Legedre approximation `val` via summation of all weighted sums $\sum_{i=1}^{\texttt{order}} \alpha_i \texttt{f}_i$, with (adapted) weights $\alpha_i$ on each subinterval given by `l`, `r` and `ref`.

# B Symbols

The following collection shall be seen a list of the most important and frequently used notations throughout this thesis. It is separated according to the chapters in this thesis where they were introduced.

## B.1 General Notation

| | |
|---|---|
| $\lesssim, \gtrsim$ | less/greater or equal to except for a positive constant independent of all parameters, i.e., $u \lesssim v \ :\Leftrightarrow$ there exist a constant $m > 0$ such that $u \leq m\, v$ and analog for $\gtrsim$ |
| $\sim$ | if both $\lesssim$ and $\gtrsim$ hold |
| $V'$ | dual space of a Hilbert space $V$ |
| ${}_{V'}\langle \cdot, \cdot \rangle_V$ | duality pairing on $V' \times V$ |
| $(\cdot, \cdot)_H$ | inner product on $H$ |
| $\| \cdot \|_H$ | norm on $H$ |
| $\|A\|_{X \to Y'}$ | operator norm of a linear operator $A \in \mathcal{L}(X, Y')$ w.r.t. Hilbert spaces $X$ and $Y$, cf. Definition 2.1.2 |
| $R_X$ | Riesz isomorphism $R_X \colon X \to X'$, cf. Theorem 2.1.4 |
| $D^\alpha$ | weak derivative $D^\alpha f := \frac{\partial^{\lvert\alpha\rvert}}{\partial \mathbf{x}^\alpha} f$ |
| $H_0^m(D)$ | Sobolev space $H^m(D)$ on domain $D$ with zero boundary conditions (2.2.3) |
| $H^m(I; X)$ | Bochner space or vector valued Sobolev space on domain $I$ and Banach space $X$, cf. Definition 2.2.5 |
| $X \cap Y$ | intersection space of two Hilbert spaces $X$ and $Y$, (2.2.11) |
| $[V, \tilde{V}]_\theta$ | interpolation space of order $\theta$ between separable Hilbert spaces $V \hookrightarrow \tilde{V}$, cf. Definition 2.2.13 |
| $A'$ | dual operator $A' \in \mathcal{L}(Y', X')$ of a linear operator $A \in \mathcal{L}(X, Y)$ on reflexive Hilbert spaces $X$ and $Y$, cf. Definition 2.3.1 |
| $A_{\max}$ | boundedness/continuity constant of operator $A$, cf. Corollary 2.3.8 |
| $A_{\min}$ | coercivity or also inf-sup constant of operator $A$, cf. Corollary 2.3.8 |
| $\Pi_k$ | space of polynomials of order $k \in \mathbb{N}$ |

| | |
|---|---|
| $\Delta_T$ | inner knot sequence w.r.t. to extended knot sequence $T$, cf. Proposition 2.4.2 |
| $SP_{\Delta,k}$ | space of splines of order $k \in \mathbb{N}$ w.r.t. knot sequence $\Delta$, cf. Definition 2.4.1 |
| $SP_{j,k}$ | space of splines of order $k \in \mathbb{N}$ w.r.t. a uniform knot sequence on $[0,1]$ with grid size $2^{-j}$, (2.4.6) |
| $N_{i,k}$ | B-spline of order $k \in \mathbb{N}$ with respect to knots $\theta_i, \ldots, \theta_{i+k}$ of an extended knot sequence $T := \{\Theta\}_{j=1,\ldots,N+k}$, cf. Proposition 2.4.7 |
| $SP^n_{\Delta,\mathbf{k}}$ | space of tensor-product splines $SP^n_{\Delta_T,\mathbf{k}} := \bigotimes_{j=1}^n SP_{\Delta_{T_j},k_j}$ of order $(k_1,\ldots,k_n)$ on knot sequence $(\Delta_{T_1},\ldots,\Delta_{T_n})$, (2.4.8) |
| $SP_{\mathbf{j},\mathbf{k}}$ | space of tensor-product splines on uniform grid analog to $SP_{j,k}$ |
| $H^{\mathbf{r}}(D)$ | tensor Sobolev space of order $\mathbf{r} \in \mathbb{R}^n$ on domain $D \subset \mathbb{R}^n$, cf. Theorem 2.4.12 |

## B.2 Space-Time Weak Formulation

| | |
|---|---|
| $D$ | spatial domain |
| $I := (0,T)$ | temporal domain |
| $n$ | space dimension |
| $T$ | final time |
| $H$ | spatial pivot space |
| $V$ | spatial space on which $A(t)\colon V \to V'$ is defined |
| $W_+ \hookrightarrow W_0 \hookrightarrow W_-$ | shifted regularity spatial spaces, cf (3.3.10) |
| $\mathcal{X}$ | solution space of first formulation, (3.1.8) |
| $\mathcal{Y}$ | test space of first formulation, (3.1.7) |
| $a(t;\cdot,\cdot)$ | spatial bilinear form, (3.1.1) resp. (3.1.2) |
| $\lambda$ | Gårding shift, (3.1.2) |
| $A(t)$ | spatial differential Operator, (3.1.3) |
| $g(t)$ | right hand side of strong formulation, (3.1.4) |
| $u_0$ | initial value function, (3.1.4) |

| | |
|---|---|
| $b(\cdot,\cdot)$ | space-time bilinear form of first formulation, (3.1.9) |
| $B$ | space-time operator of first formulation, (3.1.11) |
| $\mathcal{F}$ | right hand side of first formulation, (3.1.10) |
| $\dot{v}$ | temporal derivative of $v$ |
| $\rho$ | constant according to $\mathcal{Y} \hookrightarrow \mathcal{C}(\overline{I}; H)$, (3.1.14) |
| $\varrho$ | embedding constant of $V \hookrightarrow H$, cf. (3.1.16) |
| $u_0^*$ | auxiliary function for homogenization such that $u_0^* \in \mathcal{X}$, $u_0^*(0) = u_0$, cf. Theorem 3.2.7 |
| $\mathcal{X}_0$ | solution space of homogenized first formulation, (3.2.1) |
| $\mathcal{Y}_0$ | test space of homogenized first formulation, (3.2.2) |
| $H_{0,\{0\}}(I; V')$ | Bochner space $H(I; V')$ with zero initial condition, cf. (3.2.1) |
| $H_{0,\{T\}}(I; V')$ | Bochner space $H(I; V')$ with zero final time condition, cf. (3.3.2) |
| $b_0(\cdot,\cdot)$ | space-time bilinear form of homogenized first formulation, (3.2.4) |
| $B_0$ | space-time operator of homogenized first formulation, (3.2.6) |
| $\mathcal{F}_0$ | right hand side of homogenized first formulation, (3.2.5) |
| $\bar{\mathcal{X}}_0$ | shifted solution space of homogenized first formulation, (3.2.8) |
| $\bar{\mathcal{Y}}_0$ | shifted test space of homogenized first formulation, (3.2.8) |
| $\tilde{\mathcal{X}}$ | solution space of second formulation, (3.3.3) |
| $\tilde{\mathcal{Y}}$ | test space of second formulation, (3.3.3) |
| $\tilde{b}(\cdot,\cdot)$ | space-time bilinear form of second formulation, (3.3.5) |
| $\tilde{B}$ | space-time operator of second formulation |
| $\tilde{\mathcal{F}}$ | right hand side of second formulation, (3.3.6) |
| $\bar{\tilde{\mathcal{X}}}$ | shifted solution space of second formulation, (3.3.9) |
| $\bar{\tilde{\mathcal{Y}}}$ | shifted test space of second formulation, (3.3.9) |

## B.3 Petrov-Galerkin Discretization

| | |
|---|---|
| $S_j$ | discrete subspaces of solution space with refinement level $j$, cf. (4.1.3) and (4.3.2) |
| $Q_\ell$ | discrete subspaces of test space with refinement level $\ell$, cf. (4.1.3) and (4.3.2) |
| $S_j^x$ | discrete subspaces w.r.t. variable $x$ of solution space with refinement level $j$, (4.3.2) |
| $Q_\ell^x$ | discrete subspaces w.r.t. variable $x$ of test space with refinement level $\ell$, (4.3.2) |
| $u_j$ | minimal residual Petrov-Galerkin solution, (4.1.5) |
| $\beta_{j,\ell}$ | discrete inf-sup condition, (4.1.6) |
| $\mathbf{B}_{j,\ell}$ | system matrix w.r.t. $B$ on $S_j$ and $Q_\ell$, (4.1.10) |
| $\mathbf{u}_j$ | expansion coefficients of $u_j$ w.r.t. $S_j$, cf. (4.1.10) |
| $\mathbf{w}_\ell$ | expansion coefficients of $w_\ell$ w.r.t. $Q_\ell$ |
| $\mathbf{f}_\ell$ | discrete right hand side, cf. (4.1.10) |
| $L,\ \beta,\ C_{J,X'},\ C_{CS},$ $C_{B,X'},\ C_{J,Y},\ C_+,$ $c_s,\ c_X,\ C_X,\ C_{B,S}^{(x)},$ $c_s^{(t)},\ c_s^{(x)},\ c_{Y_+},\ C_Y$ | all relevant constants according to Table 4.3.27 |
| $\nu$ | base of Bernstein and Jackson estimates on test and (dual) solution space (4.2.3), (4.2.4) |
| $P_{S_j^x}$ | (bi-)orthogonal projector from $L_2(D)$ to $S_j^x$, cf. Proposition 4.3.7 |
| $\mu$ | base of Bernstein and Jackson estimates w.r.t. $L_2$, (4.3.10), (4.3.11) |
| $d_F$ | range of Bernstein inequality (4.3.11) on space $F$ |
| $d_F$ | range of Jackson inequality (4.3.10) on space $F$ |
| $d_t,\ d_x$ | regularity shift in $t$ and $x$, respectively, cf. Table 4.3.29 |

## B.4 Random PDEs

| | |
|---|---|
| $U(t,\omega)$ | solution of parabolic random PDE, (5.1.1) |
| $U_0(\omega)$ | initial value, cf. (5.1.1) |
| $(\Omega,\Sigma,\mathbb{P})$ | probability space |
| $\tilde{b}_\omega(\cdot,\cdot),\ \tilde{\mathcal{F}}_\omega,\dots$ | random parameter dependent pendants of its deterministic counterparts, cf. e.g. (5.1.2) |
| $\|\cdot\|_{\tilde{\mathcal{X}}_\omega},\ \|\cdot\|_{\tilde{\mathcal{Y}}_\omega}$ | $\omega$-dependent norm on $\tilde{\mathcal{X}}$, resp. $\tilde{\mathcal{Y}}$, (5.1.13) |
| $\mathcal{N}(\mu,\sigma^2)$ | normal distribution with parameter $\mu,\ \sigma^2$ |
| $\mathcal{LN}(\mu,\sigma^2)$ | log-normal distribution $\mu,\ \sigma^2$ |
| $\mathcal{U}(a,b)$ | uniform distribution on $[a,b]$ |
| $c_S$ | subspace dependent constant $\|\cdot\|_{V'}\lesssim\|\cdot\|_{S'}$, (5.2.1) |
| $P_S$ | $H$-orthogonal projector onto $S$, (5.2.2) |
| $\tilde{\mathcal{X}}_S$, $\tilde{\mathcal{Y}}_S$ | semidiscrete solution and test spaces, (5.2.3) |
| $b_{S,\omega}(\cdot,\cdot),\ B_{S,\omega}$ | semidiscrete random bilinear form (5.2.4) and corresponding operator |
| $U_S$ | semidiscrete Galerkin solution of (5.2.4) |
| $U_j$ | Petrov-Galerkin solution of random PDE (5.3.1) w.r.t. $S_j$ |
| $\partial_t S_j$ | space of time derivatives w.r.t. $S_j$, (5.3.8) |

## B.5 Numerical Results

| | |
|---|---|
| `function` | typewriter fonts indicate functions or variables implemented in Matlab as well as folders |
| $\Psi^{\tilde{\mathcal{X}}},\ \Psi^{\tilde{\mathcal{Y}}}$ | Riesz basis of $\tilde{\mathcal{X}}$ resp. $\tilde{\mathcal{Y}}$, cf. (6.1.1) |
| $\Delta_{\tilde{\mathcal{X}}}$ | (infinite) index set of Riesz basis $\Psi_{\tilde{\mathcal{X}}}$, cf. (6.1.1) |
| $\Delta^{\tilde{\mathcal{X}}}_{(j)}$ | index set of Riesz basis $\Psi_{\tilde{\mathcal{X}}}$ up to level $j$, cf. after (6.1.1) |
| $r_{\tilde{\mathcal{X}}},\ R_{\tilde{\mathcal{X}}}$ | lower and upper Riesz constant w.r.t. $\Psi_{\tilde{\mathcal{X}}}$, (6.1.1) |
| $\Psi^{\tilde{\mathcal{X}}}_j,\ \Psi^{\tilde{\mathcal{Y}}}$ | set of Riesz basis functions $\Psi^{\tilde{\mathcal{X}}}$ up to level $j$, cf. after (6.1.1) |

| | |
|---|---|
| $\lambda_{\min}(\mathbf{B}),\ \lambda_{\max}(\mathbf{B})$ | absolute smallest and largest eigenvalue of matrix/operator $\mathbf{B}$, cf. (6.1.4) |
| $\sigma_{\min}(\mathbf{B}),\ \sigma_{\max}(\mathbf{B})$ | smallest and largest singular value of matrix/operator $\mathbf{B}$, (6.1.4) |
| $S^t_{j,k}$ | abbreviation for spline spaces spanned by B-splines on uniform grid with spacing $2^{-j}$ on $[0,1]$ and order $k$ in coordinate $t$ w.r.t. the solution space, (6.1.13) |
| $Q^t_{\ell,k}$ | abbreviation for spline spaces spanned by B-splines on uniform grid with spacing $2^{-\ell}$ on $[0,1]$ and order $k$ in coordinate $t$ w.r.t. the test space |
| $S_{\mathbf{j},\mathbf{k}}$ | Kronecker product of $S^t_{j_1,k_1}$ and $S^x_{j_2,k_2}$ w.r.t. the solution space, (6.1.14) |
| $Q_{\boldsymbol{\ell},\mathbf{k}}$ | Kronecker product of $Q^t_{\ell_1,k_1}$ and $Q^x_{\ell_2,k_2}$ w.r.t. the test space, (6.1.14) |
| $(S^t_{j,k})_{0,\{0\}}$, $(S^t_{j,k})_{0,\{1\}}$, $(S^t_{j,k})_0$ | boundary adapted spline space spanned by B-splines, (6.1.15) |
| $f_{\mathcal{N}(\mu,\sigma^2)},\ f_{\mathcal{LN}(\mu,\sigma^2)}$ | probability density function of normal resp. log-normal distribution with parameter $\mu,\ \sigma^2$, cf. (6.2.8) and (6.3.5) |
| $b_{0,\omega}(\cdot,\cdot),\ \mathcal{F}_{0,\omega}$ | homogenized random bilinear form and corresponding right hand side, cf. Example 6.3.1 |

# References

[ABHN01] W. Arendt, C. J. K. Batty, M. Hieber, and F. Neubrander. *Vector valued Laplace transforms and Cauchy problems*. Monographs in mathematics. Birkhäuser, 2001.

[AF03] R.A. Adams and J.J.F. Fournier. *Sobolev spaces*. Academic press, second edition, 2003.

[And12] R. Andreev. *Stability of space-time Petrov-Galerkin discretizations for parabolic evolution problems*. PhD thesis, ETH Zürich, 2012.

[And13] R. Andreev. Stability of sparse space-time finite element discretizations of linear parabolic evolution problems. *IMA J. Numer. Anal.*, 33:242–260, 2013.

[And14] R. Andreev. Space-time discretization of the heat equation. *Numer. Algor.*, 67:713–731, 2014.

[And16] R. Andreev. Wavelet-in-time multigrid-in-space preconditioning of parabolic evolution equations. *SIAM J. Sci. Comput.*, 38(1):216–242, 2016.

[AT07] R. J. Adler and J. E. Taylor. *Random Fields and Geometry*. Springer, 2007.

[Aub00] J.-P. Aubin. *Applied functional analysis*. Wiley, second edition, 2000.

[Bab71] I. Babuška. Error-bounds for finite element method. *Numerische Mathematik*, 16(4):322–333, 1971.

[Bau95] H. Bauer. *Probability Theory*. de Gruyter, 1995.

[BCH05] Y. Bazilevs, J. Cottrell, and T. Hughes. Isogeometric analysis: CAD, finite elements, NURBS, exact geometry and mesh refinement. *Comput. Methods. Appl. Mech. Engrg.*, 194:4135–4195, 2005.

[BCH09] Y. Bazilevs, Cottrell, and T. Hughes. *Isogeometric Analysis: Toward Integration of CAD and FEA*. Wiley, 2009.

[BJ89] I. Babuška and T. Janik. The h-p version of the finite element method for parabolic equations. Part 1. The equations in time. *Numer. Methods Partial Differential Equations*, 5:363–399, 1989.

[BJ90] I. Babuška and T. Janik. The h-p version of the finite element method for parabolic equations. Part 2. The h-p version in time. *Numer. Methods Partial Differential Equations*, 6:343–369, 1990.

[BL76] J. Bergh and J. Löfström. *Interpolation Spaces*. Springer, 1976.

[Bra07] D. Braess. *Finite Elements: Theory, Fast Solvers, and Applications in Solid Mechanics*. Cambridge University Press, third edition, 2007.

[Bre11] H. Brezis. *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Springer, 2011.

[BS88] C. Bennett and R. Sharpley. *Interpolation of Operators*, volume 129. Academic Press, 1988.

[CD15] A. Cohen and R. DeVore. Approximation of high-dimensional parametric PDEs. *Acta Numerica*, 24:1–159, 2015.

[CDD01] A. Cohen, W. Dahmen, and R. DeVore. Adaptive wavelet methods for elliptic operator equations: Convergence rates. *Mathematics of Computation*, 70(233):27–75, 2001.

[CDD02] A. Cohen, W. Dahmen, and R. DeVore. Adaptive wavelet methods ii — Beyond the elliptic case. *Found. Comput.Math.*, 2(3):203–246, 2002.

[CDD03a] A. Cohen, W. Dahmen, and R. DeVore. Adaptive wavelet schemes for nonlinear variational problems. *SIAM J. Numer. Anal.*, 41(5):1785–1823, 2003.

[CDD03b] A. Cohen, W. Dahmen, and R. DeVore. Sparse evaluation of compositions of functions using multiscale expansions. *SIAM J. Math. Anal.*, 35(2):279–303, 2003.

[Cha12] J. Charrier. Strong and weak error estimates for elliptic partial differential equations with random coefficients. *SIAM J. Numer. Anal.*, 50(1):216–246, 2012.

[CS11] N. Chegini and R. Stevenson. Adaptive wavelet schemes for parabolic problems: Sparse matrices and numerical results. *SIAM J. Numer. Anal*, 49(1):182–212, 2011.

[Dah94] W. Dahmen. Some remarks on multiscale transformations, stability and biorthogonality. In P.-J. Laurent, A. Le Méhauté, and L. L. Schumaker, editors, *Wavelets, images and surface fitting*, pages 157–188. A K Peters, Ltd., 1994.

[Dah96] W. Dahmen. Stability of multiscale transformations. *J. Fourier Anal. Appl.*, 2:341–362, 1996.

[Dah97] W. Dahmen. Wavelet and multiscale methods for operator equations. *Acta Numerica*, 6:55–228, 1997.

[dB01] C. de Boor. *A practical guide to splines*. Springer, 2001.

[DHSW12] W. Dahmen, C. Huang, C. Schwab, and G. Welper. Adaptive Petrov-Galerkin methods for first order transport equations. *SIAM J. Numer. Anal.*, 50(5):2420–2445, 2012.

[DK01] W. Dahmen and A. Kunoth. Appending boundary conditions by Lagrange multipliers: General criteria for the LBB condition. *Numer. Math.*, 88:9–42, 2001.

[DKO12] S. V. Dolgov, B. N. Khoromskij, and V. Oseledets. Fast solution of parabolic problems in the tensor train/quantized tensor train format with initial application to the Fokker-Planck equation. *SIAM J. Sci. Comput.*, 34(6):A3016–A3038, 2012.

[DKU99] W. Dahmen, A. Kunoth, and K. Urban. Biorthogonal spline-wavelets on the interval — Stability and moment conditions. *Applied and Comp. Harmonic Analysis*, 6(2):132–196, 1999.

[DL88] R. Dautray and J.-L. Lions. *Functional and Variational Methods*, volume 2 of *Mathematical Analysis and Numerical Methods for Science and Technology*. Springer, second edition, 1988.

[DL90] R. Dautray and J.-L. Lions. *Spectral Theory and Applications*, volume 3 of *Mathematical Analysis and Numerical Methods for Science and Technology*. Springer, 1990.

[DL92] R. Dautray and J.-L. Lions. *Evolution Problems I*, volume 5 of *Mathematical Analysis and Numerical Methods for Science and Technology*. Springer, 1992.

[DPW14] W. Dahmen, C. Plesken, and G. Welper. Double greedy algorithm: Reduced basis methods for transport dominated problems. *ESAIM: Mathematical Modelling and Numerical Analysis*, 48(3):623–663, 2014.

[DR08] W. Dahmen and A. Reusken. *Numerik für Ingenieure und Naturwissenschaftler*. Springer, second edition, 2008.

[DS99] W. Dahmen and R. Stevenson. Element-by-element construction of wavelets satisfying stability and moment conditions. *SIAM J. Numer. Anal.*, 37(1):319–352, 1999.

[EG04] A. Ern and J.-L. Guermond. *Theory and Practice of Finite Elements*, volume 159. Springer, 2004.

[EHR+13] J. Evans, T. Hughes, A. Reali, D. Schillinger, and M. Scott. Isogeometric collocation: Cost comparison with Galerkin methods and extension to adaptive hierarchical NURBS discretizazions. *Comput. Methods. Appl. Mech. Engrg.*, 267:170–232, 2013.

[Eva10] L. C. Evans. *Partial differential equations*. Graduate studies in mathematics. American Math. Soc., second edition, 2010.

[GAS14] C. J. Gittelson, R. Andreev, and C. Schwab. Optimality of adaptive Galerkin methods for random parabolic partial differential equation. *J. Comput. Appl. Math.*, 263:189–201, 2014.

[GK11] M.D. Gunzburger and A. Kunoth. Space-time adaptive wavelet methods for optimal control problems constrained by parabolic evolution equations. *SIAM J. Contr. Optim.*, 49(3):1150–1170, 2011.

[GO95] M. Griebel and P. Oswald. Tensor product type subspace splittings and multilevel iterative methods for anisotopic problems. *Adv. Comput.Math.*, 4(1):171–206, 1995.

[GO07] M. Griebel and D. Oeltz. A sparse grid space-time discretization scheme for parabolic problems. *Computing*, 81:1–34, 2007.

[GS91] R. G. Ghanem and P. D. Spanos. *Stochastic Finite Elements: A Spectral Approach*. Springer, 1991.

[HÖ3] K. Höllig. *Finite Element Methods with B-Splines*. Society for Industrial and Applied Mathematics, 2003.

[Hac81] W. Hackbusch. Optimal $H^{p,p/2}$ error estimates for a parabolic Galerkin method. *SIAM J. Numer. Anal.*, 18(4):681–692, 1981.

[JK16] V. Jovanovic and S. Koshkin. The Ritz method for boundary problems with essential conditions as constraints. *Adv. Math. Phys.*, 2016. Article ID 7058017.

[Kac86] J. Kacur. Method of Rothe in evolution equations. In *Lecture Notes in Mathematics*, volume 1192. Teubner-Texte zur Mathematik, 1986.

[KM15] A. Kunoth and C. Mollet. Space-time approximations of parabolic PDE-constrained control problems — Variations on a theme. in revision, 2015.

[KS13] A. Kunoth and C. Schwab. Analytic regularity and GPC approximation for control problems constrained by parametric elliptic and parabolic PDEs. *SIAM J. Contr. Optim.*, 51(3):2442–2471, 2013.

[KS15] A. Kunoth and C. Schwab. Sparse adaptive tensor Galerkin approximations of stochastic PDE-constrained control problems. Preprint 2015-37, Seminar for Applied Mathematics, ETH Zürich, to appear in: SIAM/ASA Journal on Uncertainty Quantification, 2015.

[Lan01] J. Lang. *Adaptive multilevel solution of nonlinear parabolic PDE systems*, volume 16 of *Lecture Notes in Computational Science and Engeneering*. Springer, 2001.

[LM72] J. L. Lions and E. Magenes. *Non-Homogeneous Boundary Value Problems and Applications*, volume 1. Springer, 1972.

[LM16a] S. Larsson and M. Molteni. Numerical solution of parabolic problems based on weak space-time formulation. Preprint, arXiv:1603.03210, 2016.

[LM16b] S. Larsson and M. Molteni. A weak space-time formulation for the linear stochastic heat equation. *International Journal of Applied and Computational Mathematics*, 2016.

[LMM16] S. Larsson, C. Mollet, and M. Molteni. Quasi-optimality of Petrov-Galerkin discretizations of parabolic problems with random coefficients. Preprint, arXiv:1604.06611, 2016.

[LMN15] U. Langer, S.E. Moore, and M. Neumüller. Space-time isogeometric analysis of parabolic evolution equations. Report 2015-19, RICAM, 2015.

[LS15] S. Larsson and C. Schwab. Compressive space-time Galerkin discretizations of parabolic partial differential equations. Preprint, arXiv:1501.04514, 2015.

[LW13] U. Langer and M. Wolfmayr. Multiharmonic finite element analysis of a time-periodic parabolic optimal control problem. *J. Numer. Math.*, 21(4):265–300, 2013.

[MB97] C. G. Makridakis and I. Babuska. On the stability of the discontinuous Galerkin method for the heat equation. *SIAM J. Numer. Anal.*, 34(1):389–401, 1997.

[MKM13] C. Mollet, A. Kunoth, and T. Meier. Excitonic eigenstates of disordered semiconductor quantum wires: Adaptive wavelet computation of eigenvalues for the electron-hole Schrödinger equation. *Commun. Comput. Physics*, 14(1):21–47, 2013.

[Mol11] C. Mollet. *Excitonic Eigenstates in Disordered Semiconductor Quantum Wires: Adaptive Computation of Eigenvalues for the Electronic Schrödinger Equation Based on Wavelets.* Shaker–Verlag, DOI: 10.2370/OND000000000098, 2011.

[Mol13a] C. Mollet. Adaptive wavelet methods for calculating excitonic eigenstates in disordered quantum wires. In *Oberwolfach Report 56/2013*, Numerical Solution of PDE Eigenvalue Problems, pages 3242–3245, 2013.

[Mol13b] C. Mollet. Stability of Petrov-Galerkin discretizations: Application to the space-time weak formulation for parabolic evolution problems. *Comput. Methods. Appl. Math.*, 14(2):231–255, 2013.

[Mol13c] C. Mollet. Stability of Petrov-Galerkin discretizations: Application to the weak space-time formulation for parabolic PDEs. In *Oberwolfach Report 39/2013*, Multiscale and High-Dimensional Problems, pages 2222–2225, 2013.

[MP13] C. Mollet and R. Pabel. Efficient application of nonlinear stationary operators in adaptive wavelet methods — the isotropic case. *Numer. Algor.*, 63(4):615–643, 2013.

[Nol04] W. Nolting. *Grundkurs Theoretische Physik 5/1*. Springer, 6. edition, 2004.

[Nol12] W. Nolting. *Grundkurs Theoretische Physik 5/2*. Springer, 7. edition, 2012.

[NSV09] R. H. Nochetto, K. G. Siebert, and A. Veeser. Theory of adaptive finite element methods: An introduction. In A. Kunoth and R. DeVore, editors, *Multiscale, Nonlinear and Adaptive Approximation*, pages 409–542. Springer, 2009.

[Pab15] R. Pabel. *Adaptive Wavelet Methods for Variational Formulations of Nonlinear Elliptic PDEs on Tensor-Product Domains*. Logos-Verlag Berlin, 2015. PhD thesis.

[Pis03] G. Pisier. *Introduction to Operator Space Theory*. Cambridge University Press, 2003.

[RR04] M. Renardy and R. C. Rogers. *An introduction to partial differential equations*. Texts in applied mathematics. Springer, second edition, 2004.

[Rud91] W. Rudin. *Functional Analysis*. International Series in Pure and Applied Mathematics. McGraw-Hill, second edition, 1991.

[Sch91] W. E. Schiesser. *The Numerical Method of Lines: Integration of Partial Differential Equations*. Academic press, 1991.

[Sch07] L. L. Schumaker. *Spline Functions: Basic Theory*. Cambridge University Press, third edition, 2007.

[Sch15] L. L. Schumaker. *Spline Functions: Computational Methods*. Society for Industrial and Applied Mathematics, 2015.

[SS09] C. Schwab and R. Stevenson. Space-time adaptive wavelet methods for parabolic evolution problems. *Mathematics of Computation*, 78(267):1293–1318, 2009.

[SS16] C. Schwab and R. Stevenson. Fractional space-time variational formulation of (Navier-) Stokes equation. Preprint, 2016.

[Sta11] F. Stapel. Space-time tree-based adaptive wavelet methods for parabolic PDEs. Diploma thesis, Institut für Mathematik, Universität Paderborn, 2011.

[Ste10] O. Steinbach. *Numerical Approximation Methods for Elliptic Boundary Value Problems*. Springer, 2010.

[Ste15] O. Steinbach. Space-time finite element methods for parabolic problems. *Comput. Methods Appl. Math.*, 15(4):551–566, 2015.

[Tan13] F. Tantardini. *Quasi Optimality in the Backward Euler-Galerkin Method for Linear Parabolic Problems*. PhD thesis, Universita' degli Studi di Milano, 2013.

[Tec13] A. L. Teckentrup. *Multilevel Monte Carlo Methods and Uncertainty Quantification*. PhD thesis, University of Bath, 2013.

[Tho06] V. Thomée. *Galerkin finite element methods for parabolic problems*, volume 25 of *Springer series in Computational Mathematics*. Springer, second edition, 2006.

[TOS01] U. Trottenberg, C.W. Oosterlee, and A. Schüller. *Multigrid*. Academic press, 2001.

[UP12] K. Urban and T. Patera. A new error bound for reduced basis approximation of linear parabolic differential equations. *C. R. Acad. Sci. Paris, Ser I*, 350(3-4):203–207, 2012.

[UP14] K. Urban and T. Patera. An improved error bound for reduced basis approximation of linear parabolic problems. *Math. Comp.*, 83:1599–1615, 2014.

[Urb09] K. Urban. *Wavelet Methods for Elliptic Partial Differential Equations*. Oxford University Press, 2009.

[Ver95] R. Verführt. The stability of finite element methods. *Numer. Methods Partial Differential Equations*, 11(1):93–109, 1995.

[Wlo82] J. Wloka. *Partielle Differentialgleichungen*. Sobolevräume und Randwertaufgaben. B.G. Teubner, 1982.

[XZ03] J. Xu and L. Zikatanov. Some obervations on Babuška and Brezzi theories. *Numer. Math.*, 94:195–202, 2003.

[Zei95a] E. Zeidler. *Applied Functional Analysis: Applications to Mathematical Physics*. Springer, 1995.

[Zei95b] E. Zeidler. *Applied Functional Analysis: Main Principles and Their Applications*. Springer, 1995.

# Eidesstattliche Erklärung

Ich versichere, dass ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie – abgesehen von unten angegebenen Teilpublikationen – noch nicht veröffentlicht worden ist, sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde.
Die Bestimmungen der Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Dr. Angela Kunoth/Prof. Dr. Ulrich Trottenberg betreut worden.

Christian Mollet Köln, 2016

## Teilpublikationen


[1] S. Larsson, C. Mollet, and M. Molteni. Quasi-optimality of Petrov-Galerkin discretizations of parabolic problems with random coefficients. Preprint, arXiv:1604.06611, 2016.

[2] C. Mollet. Stability of Petrov-Galerkin discretizations: Application to the space-time weak formulation for parabolic evolution problems. *Comput. Methods. Appl. Math.*, 14(2):231-255, 2013.

[3] C. Mollet. Stability of Petrov-Galerkin discretizations: Application to the weak space-time formulation for parabolic PDEs. In *Oberwolfach Report 39/2013*, Multiscale and High-Dimensional Problems, pages 2222–2225, 2013.